Concise report

The performance of different classification criteria for systemic lupus erythematosus in a real-world rheumatology department

Brandon C. H. Tan 1,*, Isaac Tang1,*, Julie Bonin1, Rachel Koelmeyer1 and Alberta Hoi 1,2

Abstract

Objective. New classification criteria have been proposed to improve classification of systemic lupus erythematosus (SLE). We aimed to evaluate their performance by determining their sensitivity, specificity and accuracy in a real-world rheumatology department.

Methods. SLE patients who were enrolled in the Australian Lupus Registry and Biobank were included and compared with controls recruited from other rheumatology clinics. Clinical and immunological features were reviewed, according to ACR 1997, SLICC 2012, EULAR/ACR 2019, or Systemic Lupus Erythematosus Risk Probability Index (SLERPI). Performance of each set of criteria was evaluated for the overall cohort and in a subgroup of patients with early SLE.

Results. The study included 394 SLE and 123 control patients with other rheumatological conditions. Sensitivity was highest using SLICC 2012 or SLERPI 2020 criteria. Specificity was highest using ACR 1997 criteria. The SLICC 2012 criteria had the highest overall accuracy at 94.4% (95% CI: 91.7, 97.1%). In the subgroup analysis of SLE patients with early disease, SLICC 2012 performed similarly well.

Conclusions. The sensitivity and specificity of each set of classification criteria vary slightly, with SLICC 2012 and SLERPI 2020 having the highest sensitivities and the ACR 1997 criteria having the highest specificity in our patient cohort. All classification criteria serve as good instructional aids for clinicians to understand SLE manifestations. For the Australian Lupus Registry and Biobank, we will continue to use the ACR 1997 and/or SLICC 2012 as entry to the observational cohort.

Key words: Systemic lupus erythematosus, classification criteria, diagnosis, cohort studies

Introduction

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease characterized by chronic inflammatory manifestations affecting the skin, joints and kidneys in conjunction with classic immunological perturbation [1]. Disease manifestations can be heterogeneous and occur over time. Early and accurate diagnosis of SLE is challenging, as there are many mimics [2]. Like many rheumatological conditions, there is no gold standard test or biomarker that defines the disease [3]. However both delayed diagnosis and over-diagnosis remain...
among the most challenging areas in lupus management and contribute significantly to patient anxiety [4–6].

The use of classification criteria can often aid the diagnostic process, although these criteria are designed primarily for the purpose of research and the objective is to develop standardized definitions that will describe a relatively homogeneous cohort. In accordance with this purpose, classification criteria traditionally have high specificity, potentially at the expense of lower sensitivity. On the other hand, a good set of diagnostic criteria theoretically demand a high sensitivity, so that it can be useful for clinicians to capture a broad range of disease manifestations in conditions such as SLE.

The classification criteria for SLE have been through several iterations. The 1997 American College of Rheumatology (ACR 1997) criteria had been used for over 15 years until a new classification paradigm was proposed by the SLE International Collaborating Clinics (SLICC 2012). The SLICC 2012 criteria defined SLE by the presence of at least one clinical and immunological feature, or the presence of histopathologically confirmed lupus nephritis. Studies have shown that SLICC 2012 criteria have better sensitivity and similar specificity compared with ACR 1997 in their derivation dataset, but slightly lower specificity compared with ACR 1997 in their validation set [7]. To improve the specificity, an international collaboration between the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR) developed new classification criteria (herein referred to as EULAR/ACR 2019) that have a number of fundamental differences from the previous classification criteria. Firstly presence of antinuclear antibodies was considered as an entry criterion, and an additive point system across organ domains and immunological profiles was developed so that only one item of each domain was counted towards the total score, with a score of ≥10 resulting in a classification of SLE [8]. Furthermore, a new algorithm called the SLE Risk Probability Index (SLERPI) was also developed recently using machine learning and artificial intelligence tools, which reports a relatively high sensitivity for early disease [9].

We sought to evaluate these new sets of classification criteria and to compare their performance with other classification criteria using an Australian lupus cohort and comparable non-SLE controls.

Methods

Setting and participants

We performed an audit of patients attending a large metropolitan tertiary teaching hospital located in Melbourne, Australia. Monash Health is the founding site for the Australian Lupus Registry and Biobank (ALRB) and has recruited SLE patients based on inclusion of either ACR 1997 or SLICC 2012 criteria. These patients served as the ‘gold standard’ in our model when we evaluated against different classification criteria. We included all patients unless their medical or laboratory records could not be reviewed to confirm their clinical and immunological features. Controls were age and sex matched non-SLE patients from other rheumatology ambulatory care clinics. We excluded patients with overlap connective tissue disease who also fulfilled SLE criteria.

In the analysis of the performance of EULAR/ACR 2019 classification criteria, if the non-SLE patients had features that could not be conclusively excluded towards SLE, for example the presence of inflammatory arthritis, the clinical manifestation was scored towards the calculation of the EULAR/ACR 2019 score. This specific study was approved by the Monash Health research office as a quality improvement project (Reference: RES-21-0000-212Q-74506).

Data collection, criteria and attribution

To determine the clinical and immunological features outlined by the classification criteria sets, each patient’s electronic medical record (including outpatient clinic notes, pathology and imaging results, inpatient progress notes, and external correspondence letters) were extensively reviewed to determine the presence and absence of relevant features. For a criterion to be positive, the patient must have displayed this feature at least once without a more likely cause (e.g. medications, malignancy, infection) [9]. When attribution to SLE was ambiguous, the determination was arbitrated by the senior author (A.H.). For example, in cases of inflammatory arthritis among the control group, if there was an absence of specific features such as anti-cyclic citrullinated peptide antibodies or dactylitis, the presence of arthritis was counted towards the scoring. The definitions for meeting the classification criteria were in accordance with their respective publications [7–10].

Subset analysis

We also evaluated the performance of the different classification criteria in a subset of patients with early disease, defined by their recruitment into the ALRB within 15 months of their SLE diagnosis. For this subset of patients, data collection was restricted to the first 12 months of observation.

Statistical analysis

Statistical analyses were performed using STATA Version 16 (StataCorp, College Station, TX, USA). General characteristics of cases and controls were compared using the appropriate bivariate test (e.g. Pearson’s chi-square and Wilcoxon rank-sum tests). Diagnostic test characteristics assessed for each set of classification criteria included sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), with 95% confidence intervals. Accuracy was measured as the area under the receiver operating characteristic curve.
Results

Sample description

A total of 394 patients were included, 84.6% were female, 49.1% were of non-Caucasian ethnicity, and the median age of diagnosis was 30.0 (interquartile range 23.0–43.0) years. Controls were recruited sequentially from attendees of other non-SLE rheumatology outpatient clinics and were matched for age, sex and Caucasian (yes/no) ethnicity. A total of 123 controls were included, with a range of rheumatological diagnoses, such as spondyloarthopathies (26%), connective tissue disease excluding overlap lupus syndrome (including scleroderma, dermatomyositis or Sjögren’s syndrome) (22%), undifferentiated arthritis (15%), systemic vasculitis (14%), rheumatoid arthritis (11%), fibromyalgia (6%) and miscellaneous (7%). There was no statistically significant difference between cases and controls with regard to age (P = 0.079), gender (P = 0.53) or proportion of non-Caucasians (P = 0.73).

Comparison of the different sets of classification criteria

Table 1 summarizes the diagnostic characteristics of the different sets of classification criteria. The SLICC 2012 and SLERPI 2020 criteria both had the highest sensitivity out of the four sets of criteria evaluated. Specificity was highest using ACR 1997 criteria, at 95.9% (95% CI: 90.8, 98.7%). The SLICC 2012 criteria has the highest overall accuracy at 94.4% (95% CI: 91.7, 97.1%).

The EULAR/ACR 2019 classification criteria had a slightly lower sensitivity based on a false-negative rate of 5.1% (20/394). The misclassified cases were mostly due to a lack of positive ANA (12/20, 60%), despite having a range of compatible clinical features such as classic malar rash, mouth ulcers, positive anti-dsDNA, biopsy proven lupus nephritis or a number of haematological manifestations. The rest of the misclassified cases were patients with inflammatory arthritis or cutaneous lupus, with some but insufficient serological features to meet criteria.

Conversely the EULAR/ACR 2019 had a false-positive rate of 12.2% (15/123), including patients with an alternative connective tissue disease diagnosis (5/15, 33.3%), inflammatory arthritis (5/15, 33.3%), vasculitis (3/15, 20%), Still’s disease (1/15, 6.7%) and sarcoidosis (1/15, 6.7%). Clinical features such as inflammatory arthritis, which could account for 6 points in the EULAR/ACR 2019 criteria, was present in 11/15 cases (73.3%). A variety of other criteria such as hypocomplementaemia (9/15, 60%), haematological manifestations (2/15, 13%) and presence of antiphospholipid antibodies (2/15, 13%) were also seen in the controls.

Performance of the classification criteria in early disease

In early disease, the sensitivities of all four criteria sets were slightly lower (see Table 2). Both SLICC 2012 and SLERPI 2020 had equal best sensitivity at 97.6% (95% CI: 93.1, 99.5%). The specificity was still highest using ACR 1997 criteria, and overall accuracy was highest with SLICC 2012 criteria at 93.9% (95% CI: 91.0, 96.6%).

Discussion

Classification criteria are used every day by rheumatologists. The intent of developing new classification criteria is to improve the sensitivity and specificity compared with older sets of criteria, particularly in specific subgroups such as in early disease [11]. The ideal tool has both high sensitivity and specificity. Accuracy is one metric for evaluating classification models, taking into account the number of correct predictions from the total number of predictions.

Among the SLE classification criteria, the SLICC 2012 was reported to have a higher sensitivity than older ACR 1997 criteria and introduced the concept of having both clinical and immunological features to support the diagnosis of SLE [7]. The EULAR/ACR 2019 criteria further explored the concept for a requirement to have a positive ANA at entry. In the validation study the EULAR/ACR 2019 had a greater sensitivity and specificity than the ACR 1997, but a lower sensitivity than SLICC 2012 criteria [11].

Our data confirmed the slightly lower sensitivity by EULAR/ACR 2019, predominantly due to the entry criterion of positive ANA. In contrast to the earlier report, we have found that the EULAR/ACR 2019 did not have a higher specificity when compared with the SLICC 2012 criteria in our study. This is likely because of the controls that were used, as we intentionally included patients who had diagnoses such as other non-SLE patients.

Table 1 Overall performance of SLE classification criteria

| Classification system | Sensitivity (95% CI), % | Specificity (95% CI), % | PPV (95% CI), % | NPV (95% CI), % | Accuracy (95% CI) |
|------------------------|------------------------|------------------------|----------------|----------------|-------------------|
| ACR 1997               | 90.6 (87.3, 93.3)      | 95.9 (90.8, 98.7)      | 98.6 (96.8, 99.6) | 76.1 (68.6, 82.6) | 93.3 (91.0, 95.5) |
| SLICC 2012             | 98.5 (96.7, 99.4)      | 90.2 (83.6, 94.9)      | 97.0 (94.8, 98.4) | 94.9 (89.2, 98.1) | 94.4 (91.7, 97.1) |
| EULAR/ACR 2019         | 94.9 (92.3, 96.9)      | 87.8 (80.7, 93.0)      | 96.1 (93.7, 97.8) | 84.4 (76.9, 90.1) | 91.4 (88.3, 94.5) |
| SLERPI 2020            | 98.5 (96.7, 99.4)      | 84.6 (76.9, 90.4)      | 95.3 (92.8, 97.2) | 94.5 (88.5, 98.0) | 91.5 (88.3, 94.8) |

NPV: negative predictive value; PPV: positive predictive value; SLERPI: Systemic Lupus Erythematosus Risk Probability Index.
connective tissue diseases or other systemic vasculitides. The lower specificity of EULAR/ACR 2019 criteria when used against other rheumatological conditions may be due to the attribution rule, but it is still of interest to rheumatologists when trying to apply the criteria in individual cases. Interestingly the ACR 1997 criteria performed well in the differentiation between SLE vs non-SLE in our study.

We found the SLERPI 2020 performed as well as SLICC 2012 in terms of sensitivity (estimated sensitivity: 98.5%; 95% CI: 96.7, 99.4%). SLERPI 2020 criteria, like other recent criteria, placed a greater weighting to the presence of acute malar rash, haematological manifestations (thrombocytopenia or autoimmune haemolytic anaemia), and proteinuria [9]. The model proposed that a point can be deducted if a patient presents with interstitial lung disease, which is more prevalent in other connective tissue diseases such as scleroderma or Sjögren’s syndrome. The ACR 1997 demonstrated the highest specificity at 95.9% (95% CI: 90.8, 98.7%). However, the SLICC 2012 criteria demonstrated the highest accuracy overall at 94.4% (95% CI: 91.7, 97.1%).

The choice of which classification criteria to use will depend on the intent. For research purposes, having a homogeneous population defined by these criteria that will help us understand treatment response and prognosis, based on observational or interventional studies that were performed on these patient cohorts defined by the specific criteria set. The primary objective of classification criteria is to discriminate the target disease from other diseases, as well as from healthy subjects. Generally, the choice of classification criteria usually prioritizes high specificity, over sensitivity, so that they can discriminate between clinically related disorders. Many historic longitudinal cohort studies have utilized the ACR 1997 criteria. In our study, we have found that the ACR 1997 criteria have the highest specificity compared with the other newer sets of classification criteria.

While originally intended for research, classification criteria are frequently used informally as diagnostic criteria. Theoretically, classification criteria that have a high sensitivity could be used to aid clinicians in diagnosing SLE, especially in patients who are early in their disease course. Our study has demonstrated that both the SLICC 2012 criteria and the SLERPI 2020 criteria have the highest sensitivity to correctly classify lupus in the inception cohort. In the absence of endorsed diagnostic criteria for SLE, clinicians can utilize classification criteria to aid the diagnostic process, provided we understand the limitations.

Regardless of whether we are using the classification criteria for research or in practice as the framework of making a diagnosis of SLE, it is important to recognize that care needs to be taken in the interpretation of clinical features. For example, clinical features such as malar rash, photosensitivity, or pleuritic chest pain can have many mimics, and often correct attribution requires good clinical acumen. The danger of over-diagnosis is well recognized and unfortunately these criteria sets cannot simplify the process of making a diagnosis in real-life practice [4, 12].

Classification criteria can be used as a pedagogical aid, as they often describe the breadth of common disease manifestations particularly in heterogeneous disease such as SLE. In the case of the EULAR/ACR 2019 criteria, the concept of SLE being an autoimmune disease characterized by the presence of autoantibodies is shown by its entry criterion. Clinicians less familiar with SLE can use these criteria to understand common or key features of the disease. The notion of ANA negative SLE is contentious but is highlighted by the different sets of classification criteria, and only one has an absolute requirement for positive ANA. Further research on the disease course following patients selected according to the specific criteria will be important to understand the true implication for the ANA negativity in the SLE cohort.

This study is limited by its single centre and retrospective nature, but we believe that the medical notes and investigations were generally adequate given patients were recruited from a tertiary teaching hospital and its specialist rheumatology clinics. Patients with overlap lupus features were excluded and the utility of these criteria in this even more heterogeneous group could not be answered by current study. We also had limited evidence regarding the effects of these classification criteria on patient outcomes. Future studies on the usefulness of these criteria to add value to the patient and clinician in making decisions about effective clinical care should be considered in the primary care and other general clinic settings.

| Classification system | Sensitivity (95% CI), % | Specificity (95% CI), % | PPV (95% CI), % | NPV (95% CI), % | Accuracy (95% CI) |
|-----------------------|------------------------|-------------------------|----------------|----------------|------------------|
| ACR 1997              | 96.4 (79.1, 91.9)      | 95.9 (90.8, 98.7)       | 95.6 (90.0, 98.5) | 87.4 (80.6, 92.4) | 91.1 (87.7, 94.7) |
| SLICC 2012            | 97.6 (93.1, 99.5)      | 90.2 (83.6, 94.9)       | 91.0 (84.8, 95.3) | 97.4 (92.5, 99.5) | 93.9 (91.0, 96.6) |
| EULAR/ACR 2019        | 92.8 (86.8, 96.7)      | 87.8 (80.7, 93.0)       | 88.5 (81.8, 93.4) | 92.3 (88.9, 96.4) | 90.3 (86.7, 94.0) |
| SLERPI 2020           | 97.6 (93.1, 99.5)      | 84.6 (76.9, 90.4)       | 86.3 (79.5, 91.7) | 97.2 (92.0, 99.4) | 91.1 (87.6, 94.6) |

NPV: negative predictive value; PPV: positive predictive value; SLERPI: Systemic Lupus Erythematosus Risk Probability Index.
In conclusion, we have found that the newer classification criteria for SLE have good statistical performance in our cohort, but the older criteria, such as the ACR 1997 and SLICC 2012 criteria, have performed well if not better. The availability of the newer classification criteria is useful for clinicians to understand SLE manifestations and serve as a good instructional aid to assist with making a lupus diagnosis. Their utility as an entry criterion for longitudinal cohort studies needs to be further evaluated.

Acknowledgements

The Australian Lupus Registry has received sponsorship from AstraZeneca, UCB, Janssen and Arthritis Victoria.

Funding: No specific funding was received from any funding bodies in the public, commercial or not-for-profit sectors to carry out the work described in this manuscript

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

Data availability statement

The data underlining this article cannot be publicly shared due to the strict protocols and procedures outlined in Australian Lupus Registry & Biobank Data Access Policy to protect patients’ privacy and to maintain data security and ethical principles.

References

1. Tsokos GC. Autoimmunity and organ damage in systemic lupus erythematosus. Nat Immunol 2020;21:605–14.
2. Cojocaru M, Cojocaru IM, Silosi I, Vrabie CD. Manifestations of systemic lupus erythematosus. Maedica (Bucur) 2011;6:330–6.
3. Yu C, Gershwin ME, Chang C. Diagnostic criteria for systemic lupus erythematosus: a critical review. J Autoimmun 2014;48–49:10–3.
4. Narain S, Richards HB, Satoh M et al. Diagnostic accuracy for lupus and other systemic autoimmune diseases in the community setting. Arch Intern Med 2004;164:2435–41.
5. Lateef A, Petri M. Unmet medical needs in systemic lupus erythematosus. Arthritis Res Ther 2012;14 (Suppl 4):S4.
6. Doria A, Zen M, Canova M et al. SLE diagnosis and treatment: when early is early. Autoimmun Rev 2010;10:55–60.
7. Petri M, Orbai AM, Alarcon GS et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.
8. Aringer M, Costenbader K, Daikh D et al. 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. Ann Rheum Dis 2019;78:1151–9.
9. Adamichou C, Genitsaridi I, Nikolopoulos D et al. Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus. Ann Rheum Dis 2021;80:758–66.
10. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40:1725.
11. Johnson SR, Brinks R, Costenbader KH et al. Performance of the 2019 EULAR/ACR classification criteria for systemic lupus erythematosus in early disease, across sexes and ethnicities. Ann Rheum Dis 2020;79:1333–9.
12. Ugarte-Gil MF, Aralco GS. Incomplete systemic lupus erythematosus: early diagnosis or overdiagnosis? Arthritis Care Res (Hoboken) 2016;68:285–7.