Identification of novel classes for patients with lupus nephritis using two-step cluster model

Ting Huang · Shasha Xie · Liqing Ding · Xiaofeng Zeng · Hui Luo

Received: 12 July 2021 / Accepted: 21 December 2021 / Published online: 17 January 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Summary
Objective To identify and reclassify the patients in the lupus nephritis (LN) cohort, and to further analyze the prominent clinical features and clinical significance of each cluster.

Methods In this retrospective cross-sectional study, we used a two-step clustering method to classify 635 patients with LN into different clusters, then we observed the main differences and analyzed relevant clinical significance between the clusters.

Results Cluster 1 (20.5%) presented with the highest disease severity, patients in this group had the disease for a longer duration and higher systemic lupus erythematosus disease activity index (SLEDAI) score, with multiple positive auto-antibodies and lower complement level. Patients of cluster 2 (20.8%) had lower levels of IgG, IgA and IgM, with renal function being relatively worse in this cluster than in clusters 1 and 3. Cluster 3 was the largest group (58.7%), and the patients in this group showed mild disease severity.

Conclusion This study reclassified LN patients in a large cohort into three clusters. Our classification might be helpful to implement targeted therapy at various stages of systemic lupus erythematosus.

Keywords Erythematous · Immunoglobulin · Cohort · Anti-nuclear antibody · Damage index

Abbreviations
ACR · American College of Rheumatology
ANA · Antinuclear antibodies
BIC · Bayesian Information Criterion
Cr · Creatinine
ds DNA · Antibodies to double-stranded DNA
eGFR · Estimated glomerular filtration rate
Hb · Hemoglobin
IgA · Immunoglobulin A
IgG · Immunoglobulin G
IgM · Immunoglobulin M
ISN/RPS · International Society of Nephrology/Renal Pathology Society
L · Lymphocyte
LA · Lupus anticoagulant
LN · Lupus nephritis
N · Neutrophil
PGA · Physicians’ global assessment
PLT · Blood platelet
SD · Standard deviation
SDI · Damage index
SLE · Systemic lupus erythematosus
SLICC · Systemic Lupus International Collaborating Clinics
SLEDAI · Systemic lupus erythematosus disease activity index
SS-A and B · Sjögren-syndrome (SS)-A and B antibodies
UACR · Urinary albumin/creatinine ratio
WBC · White blood cells
Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease involving multiple organs, with renal involvement significantly increasing the morbidity and mortality of SLE [1]. Therefore, effective management of the clinical manifestations of lupus nephritis (LN) could reduce the burden of SLE and improve patient prognosis [2]. Currently, the Systemic Lupus International Collaborating Clinics (SLICC)/American College of Rheumatology (ACR) damage index (SDI) [3] is used to assess the disease burden, with the 2003 International Society of Nephrology/Renal Pathology Society (ISN/RPS) classification system used to apply targeted treatment strategies [4, 5]. Despite these efforts, 10–30% of patients with LN do progress to end-stage kidney disease within 15 years from the time of diagnosis [6]. The classification of LN could be improved by incorporating clinical indices and pathological features for more effective treatment.

Therefore, the aim of our study was to use a reasonable clustering method to identify clusters of prominent clinical features among patients with LN and to evaluate the clinical significance of the different clusters identified.

Material and methods

Statement of ethics

Our study was approved by the Medical Ethics Committee of Xiangya Hospital of Central South University in Changsha, China. All patients provided written informed consent. The study was conducted in accordance with the Declaration of Helsinki.

Study group

This was a retrospective, cross-sectional study conducted at the Rheumatology Department of Xiangya Hospital of Central South University between November 2006 and September 2020. Newly diagnosed patients and those already undergoing therapy were all included in the study. All patients met the ACR 1997 revised classification criteria for SLE. Data analysis was based on the CSTAR online registry, which is an academic union that was first funded by the Chinese Ministry of Science & Technology in 2009.

Data collection

The following data were retrieved from patients’ medical records for analysis: demographic information, including sex, age and duration of disease at the time of recruitment; laboratory findings, including the white blood cell count (WBC), neutrophil count (N), lymphocyte count (L), hemoglobin (Hb) concentration, blood platelet (PLT) count, urinary albumin/creatinine ratio (UACR), and creatinine (Cr) level; the estimated glomerular filtration rate (eGFR); levels of immunoglobulin G (IgG), immunoglobulin A (IgA), immunoglobulin M (IgM), C3, and C4; presence of antinuclear antibodies (ANA), antibodies to double-stranded DNA (dsDNA), anti-Smith antibodies, Sjögren syndrome (SS) A and B antibodies, nucleosome particles, antihistone antibodies, and antiribosomal P protein. The clinical evaluation included physicians’ global assessment (PGA) of disease activity, systemic lupus erythematosus disease activity index (SLEDAI), and the SDI score [7], both of which were calculated by experienced clinicians. The PGA was scored prior to reviewing the complement and anti-DNA antibody results [8].

Statistical analysis

The analysis was performed using a two-step cluster model to identify groups of patients with LN who had similar patterns of SLE clinical presentation. The analysis was performed as follows.

In the first step, we established a cluster feature tree, with the first record of the data set onto a leaf node at the root of the tree. All variables for this record were included. A logarithmic similarity criterion was then used to measure the distance, or likelihood of similarity between data sets. Data sets with high similarity were located on the same node, while those with low similarity generated new nodes. The similarity criterion assumed that variables obey a certain probability distribution and that the variables in the clustering model are independent of one another. With respect to the probability distribution, it was assumed that continuous variables had a normal (Gaussian) distribution and categorical variables had a multinomial distribution. Empirical internal tests have shown that the clustering process is robust against violations of the assumptions of independence and distribution. We evaluated these assumptions for our own data set, using a bivariate correlation to test the independence between two continuous variables and cross-tabulation to test the independence between two categorical variables. Exploration processes were used to confirm the normal distribution of continuous variables, with the χ²-test used to confirm the multimodal distribution of categorical variables.

In the second step, we used a merging clustering algorithm to combine leaf nodes. A set of clustering schemes was generated for different numbers of clusters of leaf nodes, with the Bayesian information criterion (BIC) used to compare various clustering schemes to select the optimal clustering scheme.

The two-step cluster analysis automatically determined that three clusters were to be generated. Continuous variables within each cluster were reported as the mean and standard deviation (SD), with categorical variables reported as a count and percentage. Differences in continuous variables between the three clusters were evaluated using an analysis of variance...
(ANOVA), with the χ²-test used for categorical variables.

All analyses were performed using SPSS (version 26.0, SPSS Inc., Chicago, IL, USA), with significance defined by a P-value < 0.05.

**Results**

**Demographic and clinical characteristics of patients with LN**

A total of 635 patients with LN were enrolled in the study of them 599 (94.3%) were women and the mean age of the patients was 33.8 ± 10.4 years. The baseline characteristics of our study group are described in Table 1. Since most of the patients had been treated, the average results of blood routine, renal function and immunological test were within the normal range; however, C3 levels were still lower than normal. Both ANA and anti-dsDNA autoantibodies were available for 606 (95.4%) patients. Among them, anti-Sm antibodies were found in 6%, anti-RNP antibodies in 1.4%, anti-SSA antibodies in 5%, anti-SSB antibodies in 2.2%, anti-rRNP antibodies in 0.8%, and anti-nucleosome and antihistone antibodies were detected in 3% and 1.3%, respectively. The PGA score ranged from 0 to 1.5, while mean level of SDI score was 0.3.

**Cluster analysis**

The three clusters identified are shown in Fig. 1, with the characteristics of patients in each of the three clusters summarized in Table 2, including inter-cluster differences. Cluster 1 included 130 (20.5%) patients, with 132 (20.8%) patients in cluster 2 and 373 (58.7%) patients in cluster 3. Cluster 3 was the largest among the three clusters, with patients in this cluster having a lower mean SLEDAI score than patients in clusters 1 and 2, indicative of mild disease intensity. The WBC, neutrophil, and lymphocyte counts were

---

**Table 1** Demographics and clinical characteristics of LN patients

| Patient characteristic | Value |
|------------------------|-------|
| Age at diagnosis, years | 33.8 (10.4) |
| Sex, n (%) | Male: 36 (5.7), Female: 599 (94.3) |
| WBC, (10⁹/L) | 6.7 (2.9) |
| N, (10⁹/L) | 4.4 (2.3) |
| L, (10⁹/L) | 1.8 (2.0) |
| Hb, (g/L) | 126.1 (17.8) |
| PLT, (10⁹/L) | 226.6 (72.6) |
| UACR (g/g) | 1.3 (2.0) |
| Cr, (µmol/L) | 72.9 (34.4) |
| eGFR, mL/min/1.73 m² | 96.0 (24.0) |
| IgG, (g/L) | 13.7 (4.2) |
| IgA, (g/L) | 2.7 (1.2) |
| IgM, (g/L) | 1.0 (0.6) |
| C3, (mg/L) | 712.6 (173.4) |
| C4, (mg/L) | 150.4 (67.4) |
| ANA, n (%) | 606 (95.4) |
| Anti-dsDNA antibodies, n (%) | 606 (95.4) |
| Anti-Sm antibodies, n (%) | 38 (6) |
| Anti-RNP antibodies, n (%) | 9 (1.4) |
| Anti-SSA antibodies, n (%) | 32 (5) |
| Anti-SSB antibodies, n (%) | 14 (2.2) |
| Anti-rRNP antibodies, n (%) | 5 (0.8) |
| Anti-nucleosome antibodies, n (%) | 19 (3) |
| Anti-histone antibodies, n (%) | 8 (1.3) |
| PGA, median (range) | 1.0 (0.5) |
| SDI score, mean (SD) | 0.3 (1.2) |

Data are given as mean (SD), median, or as number and percentage. WBC: white blood cell, N: neutrophils, L: lymphocytes, Hb: hemoglobin, PLT: platelets, UACR: urinary albumin/creatinine ratio, Cr: creatinine, eGFR: estimated glomerular filtration rate, IgG: immunoglobulin G, IgA: immunoglobulin A, IgM: immunoglobulin M, C3: complement 3, C4: complement 4, ANA: antinuclear antibodies, anti-dsDNA antibody: antibodies to double-stranded DNA, anti-Sm antibodies: antibodies to Smith, anti-RNP antibodies: antibodies to ribosomal P protein, PGA: physician’s global assessment, SDI: Systemic Lupus International Collaborating Clinics/ACR damage index.
| Patient characteristic | Cluster 1, \(n = 130\) (20.5%) | Cluster 2, \(n = 132\) (20.8%) | Cluster 3, \(n = 373\) (58.7%) | P-values |
|-------------------------|---------------------------------|---------------------------------|---------------------------------|-----------|
| Age at diagnosis, years | 31.3 ± 10.0c                    | 33.3 ± 11.1                     | 34.8 ± 10.1a                    | 0.003     |
| Sex, n (%)              | –                               | –                               | 0.002                           | 0.002     |
| Male                    | 5 (3.8)                         | 16 (12.1)                       | 15 (4.0)                        | –         |
| Female                  | 125 (96.2)                      | 116 (87.9)                      | 358 (96)                        | –         |
| Duration of disease, years | 9.0 ± 4.9c                    | 5.4 ± 4.4a                      | 6.0 ± 3.3ab                     | 0.008     |
| WBC, (10^9/L)            | 7.1 ± 3.2bc                     | 9.7 ± 2.9ac                     | 5.4 ± 1.6ab                     | 0.000     |
| N, (10^9/L)              | 4.6 ± 2.4bc                     | 6.9 ± 2.5ac                     | 3.5 ± 1.3ab                     | 0.000     |
| L, (10^9/L)              | 2.7 ± 4.1bc                     | 2.0 ± 0.9ac                     | 1.4 ± 0.6ab                     | 0.000     |
| Hb, (g/L)                | 123.3 ± 25.7a                   | 126.3 ± 17.8                    | 126.3 ± 17.8                    | 0.051     |
| PLT, (10^9/L)            | 250.2 ± 74.9f                   | 254.7 ± 75.1c                   | 208.5 ± 65.6ab                  | 0.000     |
| UACR (g/g)               | 1.5 ± 6.1                       | 1.7 ± 3.5                       | 0.3 ± 0.2                       | 0.111     |
| Cr, (μmol/L)             | 65.3 ± 66.0ac                   | 66.4 ± 11.9c                    | 66.4 ± 11.9c                    | 0.000     |
| eGFR, mL/min/1.73 m²     | 103.0 ± 20.4e                   | 81.0 ± 33.1ae                   | 98.8 ± 18.6b                    | 0.000     |
| IgG, (g/L)               | 15.7 ± 4.1f                     | 10.1 ± 3.0bc                    | 15.0 ± 3.9ab                    | 0.000     |
| IgA, (g/L)               | 2.8 ± 1.2bc                     | 2.2 ± 1.0bc                     | 2.9 ± 1.3b                      | 0.000     |
| IgM, (g/L)               | 1.3 ± 0.9bc                     | 0.8 ± 0.5bc                     | 1.0 ± 0.5ab                     | 0.000     |
| C3, (mg/L)               | 685.1 ± 179.5f                  | 805.8 ± 207.8ab                 | 692.6 ± 147.4f                  | 0.000     |
| C4, (mg/L)               | 130.4 ± 56.5f                   | 199.1 ± 81.2bc                  | 140.1 ± 56.8b                   | 0.000     |
| ANA                      | 104 (80)b,c                     | 129 (97.8)a                     | 373 (100)a                      | 0.000     |
| Anti-dsDNA antibodies    | 104 (80)b,c                     | 129 (97.8)a                     | 373 (100)a                      | 0.000     |
| Anti-Sm antibodies       | 38 (29.2)b,c                    | 0 (0)a                          | 0 (0)a                          | 0.000     |
| Anti-RNP antibodies      | 9 (6.9)b,c                      | 0 (0)a                          | 0 (0)a                          | 0.000     |
| Anti-SSA antibodies      | 32 (24.6)b,c                    | 0 (0)a                          | 0 (0)a                          | 0.000     |
| Anti-SSB antibodies      | 14 (10.8)b,c                    | 0 (0)a                          | 0 (0)a                          | 0.000     |
| Anti-rRNP antibodies     | 4 (3)b,c                        | 0 (0)a                          | 1 (0.2)a                        | 0.000     |
| Anti-nucleosome antibodies | 16 (12.3)b,c               | 3 (2.3)a                        | 0 (0)a                          | 0.000     |
| Anti-histone antibodies  | 8 (6.2)b,c                      | 0 (0)a                          | 0 (0)a                          | 0.000     |
| PGA, median (range)      | 1 (0–1.1)                       | 1 (0.9–1.1)                     | 1 (1–1.3)                       | 0.000     |
| SLEDAI score, mean (SD)  | 0.8 ± 1.9bc                     | 0.3 ± 1.3a                      | 0.2 ± 0.8a                      | 0.000     |
| Current medications (%)  | Prednisolone 95.8               | 86.4                            | 80.0                            | –         |
| Hydroxychloroquine       | 70.8                            | 63.6                            | 90.0                            | –         |
| Cyclophosphamide         | 4.2                             | 36.4                            | 10.0                            | –         |
| Mycophenolate mofetil    | 29.2                            | 40.9                            | 40.0                            | –         |
| Methotrexate             | 12.5                            | 0.0                             | 5.0                             | –         |
| Tacrolimus               | 25.0                            | 13.6                            | 10.0                            | –         |
| Cyclosporine             | 4.2                             | 0.0                             | 10.0                            | –         |
| Belimumab                | 12.5                            | 0.0                             | 0.0                             | –         |

Data are given as mean (SD), median, or as number and percentage

WBC White blood cell, N Neutrophils, L Lymphocytes, Hb Haemoglobin, PLT blood platelets, UACR urinary albumin/creatinine ratio, Cr Creatinine, eGFR Estimated Glomerular Filtration Rate, IgG Immunoglobulin G, IgA Immunoglobulin A, IgM Immunoglobulin M, C3 Complement 3, C4 Complement 4, ANA antinuclear antibodies, anti-dsDNA antibody antibodies to double-stranded DNA, anti-Sm antibodies antibodies to Smith, anti-rRNP antibodies antibodies to ribosomal P protein, SDI Systemic Lupus International Collaborating Clinics/ACR Damage Index, SLEDAI systemic lupus erythematosus disease activity index

1Significantly different from cluster 1
2Significantly different from cluster 2
3Significantly different from cluster 3
lower for cluster 3 than those for clusters 1 and 2; however, the levels of IgG and IgA in cluster 3 were the highest among the three clusters. All patients in cluster 3 had positive ANA and anti-dsDNA antibodies. Cluster 1 included the smallest group of patients, with these patients having the highest SLEDAI and SDI scores among all the three clusters, indicative of high disease activity and severe damage. The complement level was the lowest among all the clusters, with evidence of multiple positive auto-antibodies. The duration of disease was obviously longer in cluster 1 (9.0 ± 4.9 years) than in cluster 2 (5.4 ± 4.4 years) and cluster 3 (6.0 ± 3.3 years), \( p = 0.008 \). The treatment effect in this cluster was not desirable, even after treatment with belimumab. However, more follow-up data are needed to support the conclusion. Patients in cluster 2 had the highest levels of granulocytes and lowest levels of IgG, IgA, and IgM among the three clusters, with the renal function of these patients being relatively worse than of patients in clusters 1 and 3.

**Discussion**

Our analysis shows that the 635 patients with LN in our study group could be classified into 3 clusters, with each cluster having a different clinical profile. The severity of the disease was different among the three clusters, with highest disease severity for patients in cluster 1. These patients had a relatively high lymphocyte count, low complement level, multiple positive autoantibodies, higher SLEDAI, SDI scores and longer disease duration. We found that, compared with the other two clusters, there were individuals with significant damage to multiple organs in cluster 1, resulting in a correspondingly higher SDI score; however, the number of patients with multiple organs involved were not high; therefore, we did not list the specific conditions of organ damage in the table.

Differences in the WBC count played an important role in the cluster analysis. According to previous literature, abnormalities in T and B lymphocytes are the major adaptive immune responses in SLE, with anomalous complement activation and autoantibody production contributing to the pathogenesis of SLE and LN [9–12]. Of course, we cannot rule out that high levels of granulocytes might be the consequence of glucocorticoids at the same time. The presence of multiple positive autoantibodies in the ANA profile is the main clinical feature of patient in cluster 1. Previous studies suggest that a variety of positive ANA profile antibodies might be related to the development or organ involvement in SLE. For example, anti-Smith antibodies, associated with SLE disease activity, have also been associated with neurologic disorders and cardiovascular diseases [13–15], anti-SS-A antibodies presenting higher frequencies of musculoskeletal involvement [16, 17] and anti-RNP antibodies are more prevalent among patients with renal involvement [18, 19]. This might be the reason for higher SLEDAI and SDI scores in cluster 1.

We conducted statistical tests on the pathological results of LN of the three clusters of patients, and no statistically significant conclusions were found. Each group of patients included patients with type III, IV, and V lupus nephritis, and some patients were clinically diagnosed with LN, and did not undergo renal biopsy; hence, we did not show this part of the result in Table 2.

The limitations of the study are as follows: firstly, further longitudinal research is required to better explain the classification of these results; secondly, although most of the data types selected were continuous variables, there were several categorical variable indicators, which would have an impact on the clustering results.

In summary, the two-step cluster analysis we used in our study was successful in classifying patients with LN into three clusters, with laboratory indicators being different among these clusters. This classification may assist in providing targeted therapy to patients.

**Conclusion**

By applying a two-step clustering model, we were able to classify a large cohort of patients with LN into three clusters, presenting with different clinical and laboratory profiles and disease severity.

**Funding** Financial support for this study was provided by National Natural Science Foundation of China, China (No. 81771765). The funding agreement ensured the authors’ independence in designing the study, interpreting the data, writing, and publishing the report.

**Declarations**

**Conflict of interest** T. Huang, S. Xie, L. Ding, X. Zeng and H. Luo declare that they have no competing interests.

**Ethical standards** For this article no studies with human participants or animals were performed by any of the authors. All studies performed were in accordance with the ethical standards indicated in each case.

**References**

1. Yu F, Haas M, Glassock R, et al. Redefining lupus nephritis: clinical implications of pathophysiologic subtypes. Nat Rev Nephrol. 2017;13(8):483–95.
2. Meacock R, Dale N, Harrison MJ. The humanistic and economic burden of systemic lupus erythematosus: a systematic review. Pharmacoeconomics. 2013;31(1):49–61.
3. Gladman D, Goldsmith CH, Urowitz MB, et al. The Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index for Systemic Lupus Erythematosus International Comparison. J Rheumatol. 2000;27(2):373–6.
4. Sada KE, Makino H. Usefulness of ISN/RPS classification of lupus nephritis. J Korean Med Sci. 2009;24(Suppl):S7–S10.
5. Markowitz GS, D’Agati VD. The ISN/RPS 2003 classification of lupus nephritis: an assessment at 3 years. Kidney Int. 2007;71(6):491–5.
6. Bertsias GK, Tektonidou M, Amoura Z, et al. Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of adult and paediatric lupus nephritis. Ann Rheum Dis. 2012;71(11):1771–82.
7. Gladman D, Ibanez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol. 2002;29(2):288–91.
8. Lam GK, Petri M. Assessment of systemic lupus erythematosus. Clin Exp Rheumatol. 2005;23(Suppl 39):S120–S32.
9. Yap D, Chan TM. B cell abnormalities in systemic lupus erythematosus and lupus nephritis—role in pathogenesis and effect of immunosuppressive treatments. Int J Mol Sci. 2019;20(24):6231.
10. Devarapu SK, Lorenz G, Kulkarni OP, et al. Cellular and molecular mechanisms of autoimmunity and lupus nephritis. Int Rev Cell Mol Biol. 2017;332:43–154.
11. Yap DY, Lai KN. Pathogenesis of renal disease in systemic lupus erythematosus—the role of autoantibodies and lymphocytes subset abnormalities. Int J Mol Sci. 2015;16(4):7917–31.
12. Bagavant H, Fu SM. Pathogenesis of kidney disease in systemic lupus erythematosus. Curr Opin Rheumatol. 2009;21(5):489–94.
13. Flechsig A, Rose T, Barkhudarova F, et al. What is the clinical significance of anti-Sm antibodies in systemic lupus erythematosus? A comparison with anti-dsDNA antibodies and C3. Clin Exp Rheumatol. 2017;35(4):598–606.
14. Choe JY, Lee S, Kwak SG, et al. Anti-Sm antibody, damage index, and corticosteroid use are associated with cardiac involvement in systemic lupus erythematosus: data from a prospective registry study. J Korean Med Sci. 2020;35(21):e139.
15. Ahn SS, Jung SM, Yoo J, et al. Anti-Smith antibody is associated with disease activity in patients with new-onset systemic lupus erythematosus. Rheumatol Int. 2019;39(11):1937–44.
16. Novak GV, Marques M, Balbi V, et al. Anti-RO/SSA and anti-La/SSB antibodies: association with mild lupus manifestations in 645 childhood-onset systemic lupus erythematosus. Autoimmun Rev. 2017;16(2):132–5.
17. Higuera-Ortiz V, Mora-Arias T, Castillo-Martinez D, et al. Anti-Ro/SSA antibodies are associated with severe mitral valve regurgitation in patients with systemic lupus erythematosus. Mod Rheumatol. 2017;27(3):476–80.
18. ter Borg EJ, Groen H, Horst G, et al. Clinical associations of antiribonucleoprotein antibodies in patients with systemic lupus erythematosus. Semin Arthritis Rheum. 1990;20(3):164–73.
19. Bastian HM, Roseman JM, McGwin GJ, et al. Systemic lupus erythematosus in three ethnic groups. XII. Risk factors for lupus nephritis after diagnosis. Lupus. 2002;11(3):152–60.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.