

UCSF

UC San Francisco Previously Published Works

Title

Multicriteria decision analysis process to develop new classification criteria for systemic lupus erythematosus

Permalink

<https://escholarship.org/uc/item/1bj7d7x8>

Journal

Annals of the Rheumatic Diseases, 78(5)

ISSN

0003-4967

Authors

Tedeschi, Sara K
Johnson, Sindhu R
Boumpas, Dimitrios T
[et al.](#)

Publication Date

2019-05-01

DOI

10.1136/annrheumdis-2018-214685

Peer reviewed

Multicriteria Decision Analysis Process to Develop New Classification Criteria for Systemic Lupus Erythematosus

Sara K. Tedeschi,¹ Sindhu R. Johnson,² Dimitrios Boumpas,³ David Daikh,^{4,5} Thomas Dörner,⁶ Betty Diamond,⁷ Soren Jacobsen,⁸ David Jayne,⁹ Diane Kamen,¹⁰ William McCune,¹¹ Marta Mosca,¹² Rosalind Ramsey-Goldman,¹³ Guillermo Ruiz-Irastorza,¹⁴ Matthias Schneider,¹⁵ Murray Urowitz,² David Wofsy,⁴ Josef S. Smolen,¹⁶ Raymond P. Naden,¹⁷ Martin Aringer,¹⁸ and Karen H. Costenbader¹

¹ Brigham and Women's Hospital and Harvard Medical School, Department of Medicine, Boston, USA

² Toronto Western Hospital, University of Toronto, Department of Medicine, Ontario, Canada

³ University of Crete, Departments of Internal Medicine and Rheumatology, Clinical Immunology and Allergy, Heraklion, Greece

⁴ University of California San Francisco, Department of Medicine, San Francisco, USA

⁵ VA Medical Center, Department of Medicine, San Francisco, USA

⁶ Charite University Hospitals, Department of Medicine/Rheumatology and Clinical Immunology, Berlin, Germany

⁷ Feinstein Institute for Medical Research, Center for Autoimmune, Musculoskeletal and Hematopoietic Diseases, Manhasset, NY

⁸ Copenhagen Lupus and Vasculitis Clinic, Section 4242, Center for Rheumatology and Spine Diseases, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

⁹ University of Cambridge, Department of Medicine, Cambridge, UK

¹⁰ Medical University of South Carolina, Department of Medicine, Charleston, USA

¹¹ University of Michigan, Department of Medicine, Ann Arbor, USA

¹² Department of Clinical and Experimental Medicine, University of Pisa, Italy

¹³ Northwestern University, Department of Medicine/Division of Rheumatology, Chicago, USA

¹⁴ Autoimmune Diseases Research Unit, Department of Internal Medicine, Biocruces Health Research Institute, Hospital Universitario Cruces, University of The Basque Country, Bizkaia, Basque Country, Spain

¹⁵ University Hospital Dusseldorf, Rheumatology Department, Dusseldorf, Germany

¹⁶ Medical University of Vienna, Department of Rheumatology, Vienna, Austria

¹⁷ McMaster University, Department of Medicine, Hamilton, Canada

¹⁸ University Medical Center and Faculty of Medicine Carl Gustav Carus at the TU Dresden, Department of Medicine III, Dresden, Germany

Corresponding author:

Sara K. Tedeschi, MD, MPH

Brigham and Women's Hospital

Division of Rheumatology, Immunology and Allergy

60 Fenwood Road

Boston, MA 02115

Tel: 617-732-5325 Fax: 617-732-5766

Email: stedeschi1@bwh.harvard.edu

Word count: 3092

Keywords: systemic lupus erythematosus, clinical research, methodology

Funding: This work was jointly supported by the European League Against Rheumatism and the American College of Rheumatology. Dr. Tedeschi's work on this project was supported in part by the Lupus Foundation of America Career Development Award and NIAMS L30 AR070514. Dr. Johnson was supported by a Canadian Institutes of Health Research New Investigator Award. Dr. Jacobsen was supported by a grant from the Danish Rheumatism Association (A-3865).

Conflicts of interest: None

ABSTRACT

Objectives: EULAR and ACR are jointly supporting multi-phase development of SLE classification criteria based on weighted criteria and a continuous probability scale. Prior steps included item generation, item reduction, and hierarchical organization of candidate criteria using an evidence-based approach. Our objectives were to determine relative weights using multicriteria decision analysis (MCDA) and to set a preliminary threshold score for SLE classification.

Methods: An SLE Expert Panel (8 European, 9 North American) submitted 164 real, unique cases with a wide range of SLE probability in a standardized format. Using the candidate criteria, experts scored and rank-ordered 20 representative cases. At an in-person meeting, experts reviewed inter-rater reliability of scoring, further refined criteria definitions, and participated in an MCDA exercise. Based on expert consensus decisions on pairwise comparisons of criteria, 1000minds™ software calculated criteria weights and rank-ordered the remaining 164 cases based on their additive scores. The score of the lowest-ranked case for which complete expert consensus was achieved defined the threshold classification score.

Results: Inter-rater reliability of scoring cases with the candidate criteria was good. MCDA involved 74 pairwise decisions and was repeated for the arthritis and mucocutaneous domains when the initial ranking of some cases did not match expert opinion. After criteria weights and additive scores were

re-calculated once, experts reached consensus for SLE classification for all cases scoring >83.

Conclusions: Using an iterative process, the candidate criteria definitions were refined, preliminary weights were calculated, and a preliminary threshold score for SLE classification was determined.

INTRODUCTION

A multinational effort to develop new classification criteria for systemic lupus erythematosus (SLE) for clinical research, jointly supported by the European League Against Rheumatism (EULAR) and American College of Rheumatology (ACR), is underway. The overarching goal is to develop a system that identifies potential participants for clinical research studies, requiring a degree of homogeneity among subjects while simultaneously dealing with the extreme heterogeneity of SLE.[1] The aim was to design a system with the maximum combination of sensitivity and specificity for SLE, retaining face validity. While the classification criteria are not intended for diagnosis or clinical care, it is acknowledged that the only available “gold standard” for the presence of SLE is expert clinician opinion.

A 12 member Steering Committee was formed with input from EULAR and ACR leadership to oversee a four-phase process.[2] In Phase 1, items were collected through a Delphi exercise,[3] early SLE cohort,[4] and SLE patient survey.[5] During Phase 2, the list of potential criteria was narrowed using

nominal group technique.[6, 7] Phase 3 began with a literature review for test performance characteristics of candidate criteria, and data-driven organization of criteria into domains.[1] This report outlines the latter part of Phase 3: criteria weighting and threshold score identification through a consensus-based multicriteria decision analysis (MCDA) approach.[8-10] The goal was to develop a criteria system producing a continuous measure of the relative probability that a case (i.e. particular combination of clinical features) could be characterized as SLE, and a threshold score above which a case could be definitely classified as SLE for clinical research.

METHODS

An international panel of SLE experts collected and rank-ordered patient case scenarios, participated in a 1.5-day in-person consensus meeting, and held post-meeting discussions by email and telephone.

SLE Expert Panel. The Steering Committee invited six additional experts (three European, three North American) to form a 17 person SLE Expert Panel (“SLE experts”) to assist with this phase and establish external validity of the criteria development process.

Development of patient case scenarios. Each of the 17 SLE experts submitted 10 de-identified real cases based on patients from his/her own

cohort in a standardized online form using REDCap (Research Electronic Data Capture), a secure, web-based application for research studies.[11] Each expert was asked to submit five cases with “definite” or “likely” SLE based on clinical judgment, and five cases in which they had considered but ultimately did not diagnose SLE and/or diagnosed a condition mimicking SLE such as rheumatoid arthritis, other inflammatory arthritis, Sjögren’s syndrome, antiphospholipid antibody syndrome, or viral infection.

Rank ordering and scoring of case scenarios. From 164 de-identified cases, three authors of this manuscript (KHC, RPN, SKT) chose a representative sample of 20 reflecting a range of possible SLE cases. Each case was abstracted into standardized paragraph format and SLE experts were asked to rank the cases based on their confidence that the case should be classified as SLE. This exercise introduced SLE experts to the challenge of assessing the relative influence of individual criteria in pointing towards or away from SLE.

SLE experts then scored each of the 20 cases using a standardized online REDCap form containing data fields reflecting the draft SLE classification criteria as of September 2016, based on the Phase 2 nominal group technique exercise[7] and subsequent work by the Steering Committee.[1] The REDCap form included 10 domains; each domain included 2-6 options (e.g. mucocutaneous domain: “none”, “non-scarring alopecia”, “oral ulcers”,

“subacute cutaneous lupus”, “acute cutaneous lupus”, “discoid lupus”).

Experts were provided written instructions for scoring and a list of proposed definitions for each criterion. The instructions specified that within each domain, criteria were ordered from least to most supportive of SLE, and if multiple criteria were present in one domain only the single criterion furthest down the list (i.e. most supportive of SLE) should be scored. The instructions specified that a criterion should not be scored if a cause more likely than SLE existed (e.g. other autoimmune disease, malignancy, medication).

In-person consensus meeting, November 2016. During the 1.5 day in-person meeting, RPN and AH moderated discussions among SLE experts leading to consensus decisions. Goals of this meeting included achieving full consensus on criteria definitions, calculation of criteria weights via a multicriteria decision analysis (MCDA) exercise, and establishment of a threshold score for SLE classification.

- a) **Review of case scoring and refinement of criteria.** Experts reviewed a summary of the REDCap scoring exercise. Discrepancies in scoring individual cases were discussed in depth to understand the underlying reasons. Criteria definitions were discussed in the context of these discrepancies and refined based consensus agreement.
- b) **MCDA to determine weights.** The MCDA exercise is based on the PAPRIKA method,[12] an acronym for **P**otentially **A**ll **P**airwise

RanKings of all possible **Alternatives**, as implemented by 1000minds™ software (www.1000minds.com). This method and software have been used extensively since 2010 for developing disease classification criteria.[8, 9, 13] Experts voted on a series of pairwise decisions about hypothetical cases, each defined by two criteria from two domains. For example: hypothetical case A: “oral ulcers” (mucocutaneous domain) and “acute pericarditis” (serositis domain) vs. hypothetical case B “alopecia” (mucocutaneous domain) and “pleural effusion” (serositis domain). SLE experts were asked to decide whether they would more likely classify hypothetical case A or B as SLE, presuming all else was equal about the cases. Voting was conducted anonymously, but where opinions diverged cases were discussed until full consensus was reached. Such pairwise-ranking questions were repeated with different pairs of hypothetical cases – always involving trade-offs between different combinations of criteria, two at a time – until enough information about expert preferences had been collected to determine relative criteria weights for all criteria. Each time experts ranked a pair, all other cases that could be pairwise ranked via the logical property of ‘transitivity’ were identified and eliminated. For example, if experts ranked hypothetical case A over B and B over C, then, by transitivity, A is also ranked over C (and the method would not ask members to choose between A and C). This procedure ensures the

number of pairwise-ranking questions posed is kept to a minimum, and experts end up having pairwise ranked all possible cases defined on two criteria at a time, explicitly or by transitivity. Consensus decisions were entered into 1000minds™ software, which uses linear programming techniques to derive weights for each criterion.[12]

- c) **Assessment of the face validity of the weights.** Criteria weights were summed to produce an additive score for each case. Only the highest-weighted criterion in each domain was counted toward the additive score, as specified in the written instructions (Table 1). The remainder of the 164 cases were scored and arranged in rank order from highest to lowest score. SLE experts reviewed a spreadsheet listing the criteria present in each of the 164 cases and anonymously voted whether they would classify each as SLE. For cases where expert opinion differed, RPN facilitated discussion to achieve full consensus about case classification. Cases were discussed in descending rank order (confidence that the case should be classified as SLE) until agreement on classifying as SLE could not be reached.
- d) **Determination of an upper threshold score.** The score of the last case for which the group achieved consensus on classification as SLE was assigned as the initial threshold score.

- e) **Review of cases below the threshold.** The cases with scores immediately below the initial threshold were individually reviewed. SLE experts reached consensus that several of these should have been classified as SLE. The experts reviewed the criteria present in those cases and discussed whether the initial criteria weights were consistent with expert opinion about their relative importance in SLE.
- f) **Weighting and upper threshold revision.** The MCDA exercise was repeated once for those criteria whose calculated weights were inconsistent with expert opinion. Weights for all criteria were recalculated using 1000minds™ and additive scores were recalculated. SLE experts again anonymously voted on classifying each case as SLE, followed by a discussion facilitated by RPN to achieve consensus. The score of the last case for which expert consensus was achieved was the provisional full consensus upper threshold score.

Determining a lower threshold score. SLE experts attempted to set not only an upper threshold for definite SLE classification, but also a lower threshold for very low probability for classification. SLE experts discussed that individuals with scores falling between these two thresholds might be candidates for inclusion in observational studies or SLE prevention trials. These patients could be considered as “potential” SLE (also labeled probable,

possible or incomplete lupus, or undifferentiated connective tissue disease in the past). Due to insufficient time to perform this exercise at the November 2016 meeting, the lower threshold was addressed in a series of emails, secondary exercises, and conference calls in the next two months. SLE experts were asked to rate the cases that fell below the upper threshold score as “probable SLE”, “possible SLE”, or “unlikely SLE”. The score of the case for which $\geq 70\%$ indicated “unlikely SLE” was assigned as the lower threshold.

RESULTS

At the in-person meeting, SLE experts agreed that classification as SLE means a patient is appropriate for inclusion in SLE clinical research—and that classification as SLE should *not* guide clinical decisions about SLE diagnosis or treatment. Experts agreed that the threshold score should have high specificity for SLE, ensuring a high degree of homogeneity among classified patients and facilitating comparisons across clinical studies. SLE experts reached consensus that patients with overlap syndromes could be classified as SLE if they met SLE classification criteria, allowing clinical investigators to decide whether to include or exclude patients with overlap syndromes in specific research studies.

Review of scoring and criteria refinement. There was considerable inconsistency between SLE experts using the online REDCap form to score cases. Each expert had scored a total of 200 items (10 domains in each of 20 cases); all 17 experts scored 127/200 (64%) domains exactly the same. Reasons for discrepant data entry were explored in detail through group discussion. These included human error in data entry, not following the instructions, variability in interpreting the candidate criteria based on context, and different interpretations of criteria definitions (see **Supplement 1** for details).

Review of the rank-ordering exercise. There was agreement on the cases that the majority of SLE experts ranked the highest and lowest, but a spectrum of ranking for cases in between (**Figure 1**). This reflected the different relative weights that individual experts attached to particular criteria.

MCDA to determine consensus weights using 1000minds™ software.

SLE experts anonymously voted on 74 pairs of hypothetical cases. When consensus was lacking, members discussed rationale for choosing hypothetical case A vs. case B until consensus was reached. Sometimes it was agreed that hypothetical cases A and B were equally likely to be SLE. For a handful of pairwise comparisons, consensus could not be reached and the decision was to skip that comparison and approach their relativity from other

pairwise comparisons. Significant changes to the classification criteria during this stage included:

- **Mucocutaneous and musculoskeletal domains.** SLE experts decided that observation by a clinician should be required for consistency with other clinical domains. The definition of clinician-observed was broadened to include physical examination or review of a photograph.
- **Neurologic domain.** Due to disagreement over whether seizure or cranial neuropathy was more specific for SLE (the SLICC[15] and ACR 1982[16] manuscripts did not present the specificity of these individual items), and because the prevalence of cranial neuropathy is very low in SLE (and none of the 167 patient cases had cranial neuropathy), the group reached consensus to remove cranial neuropathy.
- **Renal domain.** SLE experts decided that Class VI lupus nephritis was not specific for SLE based on clinical experience and lack of published data, and agreed upon removing Class VI nephritis. Importantly, since historical manifestations are included in the scoring system, previous evidence of class II, III, IV, or V lupus nephritis would be fully accounted for. These steps resulted in the updated definitions depicted in **Table 1**.

Table 1. Provisional SLE classification criteria organization and definitions

Opening Statements:

- A history of a positive ANA by Hep 2 immunofluorescence $\geq 1:80$ is required for consideration of a person for SLE classification.
 - For each criterion, do not score if a cause more likely than SLE exists (such as infection, malignancy, medication, rosacea, endocrine disorder, other autoimmune disease).
 - Occurrence of a criterion on at least one occasion is sufficient.
 - Criteria need not occur simultaneously.
 - At least one clinical criterion must be present.
 - Within each domain, only the highest weighted criterion is counted toward the total score.
-

Clinical domains and criteria

Constitutional

- **Fever:** >38.3 Celsius with no other source identified
-

Hematologic

- **Leukopenia:** WBC $<4,000/\text{mm}^3$
 - **Thrombocytopenia:** Platelets $<100,000/\text{mm}^3$
 - **Autoimmune hemolysis:** (1) evidence of hemolysis, such as reticulocytosis, low haptoglobin, elevated indirect bilirubin, elevated LDH and (2) positive Coomb's (direct antiglobulin) test
-

Neuropsychiatric

- **Delirium:** characterized by (1) change in consciousness or level of arousal with reduced ability to focus, and (2) symptom development over hours to <2 days, and (3) symptom fluctuation throughout the day, and (4) either (4a) acute/subacute change in cognition (e.g. memory deficit or disorientation), or (4b) change in behavior, mood, or affect (e.g. restlessness, reversal of sleep/wake cycle, etc.)
 - **Psychosis:** characterized by (1) delusions and/or hallucinations without insight and (2) absence of delirium
 - **Seizure:** primary generalized seizure or partial/focal seizure, with independent description by a reliable witness. If EEG is performed, abnormalities must be present
-

Mucocutaneous

- **Non-scarring alopecia**, observed by a clinician*
 - **Oral ulcers**, observed by a clinician*
 - **Subacute cutaneous lupus (SCLE) or discoid lupus (DLE):** SCLE is characterized by annular or papulosquamous (psoriasiform) cutaneous eruption observed by a clinician,* usually photodistributed. If skin biopsy is performed, typical changes must be present.[17] DLE is characterized by erythematous-violaceous cutaneous lesions with secondary changes of atrophic scarring, dyspigmentation, often follicular hyperkeratosis/plugging (scalp), observed by a clinician,* leading to scarring alopecia on the scalp. Lesions have a preference for the head and neck, especially the conchal bowl, but may be found in nearly any location. If skin biopsy is performed, typical changes must be present.[17]
 - **Acute cutaneous lupus:** Malar rash (localized) or maculopapular rash (generalized) observed by a clinician,* with or without photosensitivity. If skin biopsy is performed, typical changes must be present.[17]
-

Serositis

- **Pleural or pericardial effusion:** imaging evidence (such as ultrasound, x-ray, CT scan, MRI) of pleural or pericardial effusion, or both
 - **Acute pericarditis:** ≥ 2 of: (1) pericardial chest pain (typically sharp, worse with inspiration, improved by leaning forward), (2) pericardial rub, (3) EKG with new widespread ST-elevation or PR depression, (4) new or worsened pericardial effusion on imaging (such as ultrasound, x-ray, CT scan, MRI)
-

Musculoskeletal

- **Synovitis in ≥ 2 joints:** characterized by joint swelling and tenderness, observed by a clinician*
-

Renal

- **Proteinuria $>0.5\text{g}/24\text{h}$:** on 24 hour urine collection or spot urine protein-to-

Face validity of the weights and initial upper threshold score. The additive score ranged 0-201 for the 164 cases. SLE experts reviewed the cases in rank order from highest to lowest score and reached consensus on classifying the 69 highest-scored cases as SLE. The group was unable to reach full consensus for a case with a score of 70; this patient had oral ulcers, leukopenia, low C3 or C4, and positive anti-dsDNA. The last case for which experts reached consensus (17/17 votes) for classification as SLE had a score of 71, and an initial upper threshold score was set as >70.

Revising criteria weights and provisional upper threshold score. The experts reviewed cases scored 60-70. Many of these cases had arthritis and most experts had voted to classify them as SLE. Therefore, the group felt that the weight assigned to arthritis was too low. After reviewing the specific criteria present in these cases, the mucocutaneous domain was re-organized based on expert consensus: acute cutaneous lupus was assigned the most influential position because it is most specific, and subacute cutaneous lupus and discoid lupus were grouped together and less influential than acute cutaneous lupus. Anonymous voting was repeated for pairwise comparisons including arthritis and mucocutaneous criteria. 1000minds™ software re-calculated relative weights for all criteria and re-scored all cases using the revised weights.

After this second round of MCDA, arthritis received a greater weight than prior, now identical to the weight of pleural or pericardial effusion. Acute cutaneous lupus was assigned the same weight as acute pericarditis and anti-dsDNA (**Table 2**). The group repeated the anonymous voting exercise and reached consensus about the 82 highest-scored cases. Experts were unable to reach full consensus for the same case that determined the initial threshold. As that case now had a score of 83 using the revised criteria weights, a provisional consensus threshold was set as >83. Provisional criteria weights resulting from the MCDA exercise are shown in **Table 2**.

Table 2. Provisional SLE classification criteria weights determined by a multicriteria decision analysis exercise

Clinical domains and criteria	Weight (points)	Immunologic domains and criteria	Weight (points)
Constitutional		Antiphospholipid antibodies	
• Fever	13	• Anticardiolipin IgG >40 GPL units <u>or</u> anti-β2GP1 IgG >40 units <u>or</u> lupus anticoagulant positive	13
Hematologic		Complement proteins	
• Leukopenia	12	• Low C3 <u>or</u> low C4	19
• Thrombocytopenia	26	• Low C3 <u>and</u> low C4	27
• Autoimmune hemolysis	28	SLE-specific antibodies	
Neuropsychiatric		• Anti-dsDNA antibody	38
• Delirium	12	• Anti-Smith antibody	40
• Psychosis	20		
• Seizure	34		
Mucocutaneous			
• Non-scarring alopecia	13		
• Oral ulcers	14		
• Subacute cutaneous <u>or</u> discoid lupus*	29*		
• Acute cutaneous lupus	38		
Serositis			
• Pleural <u>or</u> pericardial effusion	34		

• Acute pericarditis	38
Musculoskeletal	
• Synovitis in ≥ 2 joints	34
Renal	
• Proteinuria $>0.5\text{g}/24\text{h}$	27
• Renal biopsy with Class II or V lupus nephritis	55
• Renal biopsy with Class III or IV lupus nephritis	74

*Subacute cutaneous lupus and discoid lupus each received a weight of 29, thus they were combined as one criterion

Lower threshold score. SLE experts individually rated the 82 cases below the upper threshold score as “probable SLE”, “possible SLE”, or “unlikely SLE” after the ACR 2016 meeting. The distribution of expert opinion is shown in **Figure 2**. The score of the case for which $\geq 70\%$ indicated “unlikely SLE” was 27. Only seven of 52 unique cases (13.5%) included in this exercise would be classified as “unlikely SLE” based on this lower threshold, and the remaining 86.5% would potentially be candidates for inclusion into observational or preventive studies. Through a series of telephone calls and emails, it became clear expert opinion varied considerably concerning the cases below the upper threshold. Additionally, the terms “probable”, “possible” and “unlikely” were not being uniformly interpreted in this exercise and the duration of manifestations had not been specified. The SLE experts decided against assigning a lower threshold because it would exclude only a few cases from clinical studies.

DISCUSSION

In Phase 3 of this SLE classification criteria development project, we applied a consensus-based, data-driven MCDA approach to assign criteria weights and identify a threshold score for SLE classification for clinical research. This exercise resulted in provisional criteria weights that have face validity and are additive, providing a continuous measure of increasing likelihood for SLE based on combinations of criteria. While full consensus of the 17 SLE experts was reached for cases scoring >83 points, it became evident that expert opinions varied for cases with mid-range or low scores. Many cases with scores just under 83 were still considered SLE by the majority of experts, but in an additional exercise focusing cases below the threshold for definite SLE, very few were deemed “unlikely SLE” by $\geq 70\%$ of experts.

This stage was largely based on the items resulting from the Phase 2 nominal group technique exercise[7] and on evidence from our literature review of the sensitivity and specificity of the individual candidate criteria.[1] These efforts followed rigorous data-driven and expert-guided criteria development methodology in order to ensure high face and content validity of the items, and high discriminant validity of the criteria set.[19, 20] However, our literature review also revealed knowledge gaps about the sensitivity and specificity of some of the newly proposed criteria, thus expert consensus opinion was critical for decision making.

Consistent with developing other sets of classification criteria,[21, 22] there were significant discrepancies in ranking 20 cases regarding likelihood of SLE classification. Discussions centered on two aspects: 1) the quality and thus specificity of clinical and serological manifestations, and 2) attribution of manifestations to SLE versus other connective tissue diseases. Some experts expressed concern about misinterpretation of rosacea as acute cutaneous lupus, and about false positive anti-dsDNA via ELISA, each of which would reduce the specificity of the proposed classification system. To address these concerns, SLE experts agreed to include detailed definitions for each criterion to mitigate the risk of misinterpreting clinical signs and symptoms. Because particular laboratory assays (e.g. Farr method for anti-dsDNA) are not uniformly available in all clinical settings, SLE experts decided that the testing method would not be specified, enabling SLE classification in a wide range of clinics.

The attribution of manifestations to SLE was discussed at length. For some of the cases, SLE experts were uncertain about how to interpret particular findings when SLE and another disease, such as primary antiphospholipid syndrome or Sjögren's syndrome, seemed equally likely. It became apparent that not all these decisions could be made with certainty, and that SLE experts from different centers could reach opposing conclusions.

The decision to exclude Class VI lupus nephritis was unanimous, given the lack of specificity of this end-stage finding. The discussions leading to the consensus elimination of mononeuropathy and cranial neuropathy were of greater interest. It was first mentioned that the specificities of these entities differed and that mononeuropathy is not specific for SLE. The group reached full consensus to eliminate mononeuropathy, and cranial neuropathy was initially retained. The group then discussed that cranial neuropathy is a very rare presenting sign in SLE[23] and that none of the 164 cases had cranial neuropathy. Experts reached a unanimous decision that the low prevalence of cranial neuropathy in SLE warranted eliminating it as a criterion.

Using a data-driven approach based on literature review[1] combined with an expert-driven MCDA process based on real patient cases, this third phase of the SLE classification project has led to precisely-defined criteria with individual weights derived through consensus decisions by 17 international SLE experts. The individual criteria weights have face validity, and taken together they depict current expert understanding of SLE. The preliminary classification criteria resulting from this third phase are being refined and validated in a large, distinct set of patient cases to finalize the project.

Acknowledgements: EULAR and ACR jointly provided support for this effort. The authors would like to thank Amy Turner for her outstanding organizational support of this project and Alison Hendry for her expertise and

assistance during the ACR 2016 meeting. This work was previously presented at the EULAR 2017 Annual Congress[24] and the ACR 2017 Annual Meeting.[25]

References

- [1] Tedeschi SK, Johnson SR, Boumpas D, Daikh D, Dorner T, Jayne D, et al. Developing and refining new candidate criteria for systemic lupus erythematosus classification: An international collaboration. *Arthritis Care Res (Hoboken)* 2018;70(4):571-81.
- [2] Aringer M, Dorner T, Leuchten N, Johnson SR. Toward new criteria for systemic lupus erythematosus - a standpoint. *Lupus* 2016;25:805-11.
- [3] Schmajuk G, Hoyer BF, Aringer M, Johnson SR, Daikh DI, Dorner T, et al. Multi-center delphi exercise reveals important key items for classifying systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2017.
- [4] Mosca M, Costenbader KH, Johnson SR, Lorenzoni V, Sebastiani GD, Hoyer BF, et al. How do patients with newly diagnosed systemic lupus erythematosus present? A multicenter cohort of early systemic lupus erythematosus to inform the development of new classification criteria. *Arthritis Rheumatol* 2018.
- [5] Leuchten N, Milke B, Winkler-Rohlfing B, Daikh D, Dorner T, Johnson SR, et al. Early symptoms of systemic lupus erythematosus (sle) recalled by 339 sle patients. *Lupus* 2018:961203318776093.
- [6] Nair R, Aggarwal R, Khanna D. Methods of formal consensus in classification/diagnostic criteria and guideline development. *Semin Arthritis Rheum* 2011;41(2):95-105.
- [7] Johnson SR, Khanna D, Cervera R, Costedoat-Chalumeau N, Gladman DD, Hahn BH, et al. Use of nominal group technique to determine candidate items for sle classification criteria development [abstract]. *Arthritis Rheum* 2016;68(Suppl 10).
- [8] Johnson SR, Naden RP, Fransen J, van den Hoogen F, Pope JE, Baron M, et al. Multicriteria decision analysis methods with 1000minds for developing systemic sclerosis classification criteria. *J Clin Epidemiol* 2014;67(6):706-14.
- [9] Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 gout classification criteria: An american college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis Rheumatol* 2015;67(10):2557-68.
- [10] Neogi T, Aletaha D, Silman AJ, Naden RL, Felson DT, Aggarwal R, et al. The 2010 american college of rheumatology/european league against rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. *Arthritis Rheum* 2010;62(9):2582-91.

- [11] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (redcap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42(2):377-81.
- [12] Hansen P, Omblér F. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis* 2008;15:87-107.
- [13] Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, et al. 2016 american college of rheumatology/european league against rheumatism classification criteria for primary sjogren's syndrome: A consensus and data-driven methodology involving three international patient cohorts. *Ann Rheum Dis* 2017;76(1):9-16.
- [14] Adler Y, Charron P, Imazio M, Badano L, Baron-Esquivias G, Bogaert J, et al. 2015 esc guidelines for the diagnosis and management of pericardial diseases: The task force for the diagnosis and management of pericardial diseases of the european society of cardiology (esc) endorsed by: The european association for cardio-thoracic surgery (eacts). *Eur Heart J* 2015;36(42):2921-64.
- [15] Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum* 2012;64(8):2677-86.
- [16] Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982;25(11):1271-7.
- [17] Elman SA, Joyce C, Nyberg F, Furukawa F, Goodfield M, Hasegawa M, et al. Items for developing classification criteria in discoid lupus erythematosus: Results of a consensus exercise. *Journal of the American Academy of Dermatology* 2017.
- [18] Weening JJ, D'Agati VD, Schwartz MM, Seshan SV, Alpers CE, Appel GB, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *Kidney Int* 2004;65(2):521-30.
- [19] Singh JA, Solomon DH, Dougados M, Felson D, Hawker G, Katz P, et al. Development of classification and response criteria for rheumatic diseases. *Arthritis Rheum* 2006;55(3):348-52.
- [20] Johnson SR, Goek ON, Singh-Grewal D, Vlad SC, Feldman BM, Felson DT, et al. Classification criteria in rheumatic diseases: A review of methodologic properties. *Arthritis Rheum* 2007;57(7):1119-33.
- [21] Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, et al. 2010 rheumatoid arthritis classification criteria: An american college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis Rheum* 2010;62(9):2569-81.
- [22] van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: An american college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis Rheum* 2013;65(11):2737-47.

- [23] Hanly JG, Urowitz MB, Sanchez-Guerrero J, Bae SC, Gordon C, Wallace DJ, et al. Neuropsychiatric events at the time of diagnosis of systemic lupus erythematosus: An international inception cohort study. *Arthritis Rheum* 2007;56(1):265-73.
- [24] Tedeschi SK, Johnson SR, Boumpas DT, Daikh DI, Diamond B, Dorner T, et al. Multicriteria decision analysis for development of new systemic lupus erythematosus classification criteria (abstract). *Ann Rheum Dis* 2017;76(Suppl 2).
- [25] Tedeschi SK, Johnson SR, Boumpas D, Daikh DI, Diamond B, Doerner T, et al. Multicriteria decision analysis for the development of new systemic lupus erythematosus classification criteria (abstract). *Arthritis Rheum* 2017;69(Suppl 10).