Creation and Validation of Classification Criteria for Discoid Lupus Erythematosus

Scott A. Elman, MD; Cara Joyce, PhD; Kara Braudis, MD; Benjamin F. Chong, MD; Anthony P. Fernandez, MD, PhD; Fukumi Furukawa, MD, PhD; Minoru Hasegawa, MD; Hee Joo Kim, MD; Sara J. Li, BS; Christine G. Lian, MD; Jacek C. Szepeitowski, MD, PhD; Victoria P. Werth, MD; Joseph F. Merola, MD, MMSc

OBJECTIVE To create and validate classification criteria for DLE using 12 previously defined candidate criteria items.

DESIGN, SETTING, AND PARTICIPANTS For this diagnostic study, candidate criteria items were prospectively applied by dermatologists and dermatopathologists at clinical visits of patients with DLE or a condition that could be confused for DLE, termed a DLE mimicker, at academic dermatology practices across the United States, Poland, Japan, and South Korea. Data were collected from December 1, 2017, to February 1, 2019, and analyzed from March 1 to September 19, 2019.

MAIN OUTCOMES AND MEASURES Clinical features among these 2 groups were calculated and compared with χ² or Fisher exact tests. Candidate models were identified using best subsets logistic regression analysis. Improvement tests, fit statistics, and discrimination were considered to choose a final model.

RESULTS Nine sites contributed 215 patients, 15 of whom had missing or incomplete data. The final model for DLE classification criteria includes only clinical variables: atrophic scarring (3 points), location in the conchal bowl (2 points), preference for the head and neck (2 points), dyspigmentation (1 point), follicular hyperkeratosis and/or plugging (1 point), and erythematous to violaceous in color (1 point), with an area under the receiving operating characteristic curve of 0.91 (95% CI, 0.87-0.95). A score of at least 5 points yields a sensitivity of 84.1% and a specificity of 75.9% in the classification of DLE, with increasing scores yielding higher specificity.

CONCLUSIONS AND RELEVANCE These findings provide the initial validation of classification criteria for DLE for use in observational and clinical trials.

Published online June 17, 2020.
ficity, whereas diagnostic criteria reflect a more broad and variable set of features of a given disease and would therefore require a high sensitivity.1,4 Although classification criteria are not synonymous with diagnostic criteria, they typically mirror the list of criteria that are used for diagnosis.4 Diagnostic criteria for DLE were proposed by Fabbri et al5 in 2003; although not formally adopted by the expert community,6 this initial work serves as a good framework for our efforts.

The working group of Elman et al7 has described our efforts to identify a list of candidate criteria items previously. This potential list includes 7 clinical items (erythematous to violaceous in color, atrophic scarring, dyspigmentation, follicular hyperkeratosis/plugging, scarring alopecia, location in the conchal bowl, and preference for the head and neck) and 5 histopathological items (interface/vacuolar dermatitis, perivascular and/or periappendageal lymphohistiocytic infiltrate, follicular keratin plugs, mucin deposition, and basement membrane thickening). The specific objective of this study is to develop a points-based system for classifying DLE and to provide test characteristics (sensitivity and specificity) of this points-based system. Our method was adopted from guidance provided by the American College of Rheumatology guidelines committee and through their work specifically on the American College of Rheumatology-European League Against Rheumatism classification criteria for systemic sclerosis.8,9

Methods

Previous methods to derive a candidate item list of classification criteria are described elsewhere in full.7 To summarize, the Delphi technique is a method of consensus building using a series of iterative questionnaires to collect data from a panel of experts and stakeholders in a given area of interest. The iterative nature of the process, together with controlled anonymous feedback at each questionnaire stage, participant anonymity, and a predefined stop criterion, allow convergence toward a consensus answer. Using this technique, a potential classification criteria item set was narrowed from 48 to 12 items in 2 rounds of voting as well as an intermediary step of nominal group techniques.5,7,10,11 These candidate criteria items were moved forward into this classification development and validation process. This study was approved by the Partners Healthcare System institutional review board, which waived the need for informed consent for use of deidentified data.

Selection of Patients for Validation

Domestic and international experts in the field of CLE were identified. At each expert’s clinical site, each dermatologist was asked to prospectively identify patients with cutaneous morphologic features suggestive of DLE vs a disease mimicker. Disease mimickers constitute a set of conditions from which DLE needs to be distinguished based on morphologic and/or histopathological features to obtain the study population of interest.

Patients were included if they carried a diagnosis of DLE at a single visit, either with the presence of distinct morphologic features that suggested DLE, as determined by an international CLE expert dermatologist at each international academic center (including K.B., B.F.C., A.P.F., F.F., M.H., H.J.K., J.C.S., V.P.W., and J.F.M.), or if a diagnosis of DLE had been made previously by morphologic and/or histopathological features. A virtual investigator meeting was held wherein dermatologists were provided with reference photographs and a collection of images depicting representative clinical images of each morphologic descriptor agreed on by the CLE Steer ing Committee to ensure standardization of definitions.

Patients were also included if they carried a diagnosis of an entity that morphologically could be confused with DLE. These disease mimickers, as defined by the expert panel, included dermatomyositis, subacute CLE, other cutaneous lupus subsets (acute cutaneous lupus, chilblains, lupus panniculitis), psoriasis, lichen planus, lichen planopilaris, rosacea, sarcoidosis, and other scarring alopecias. Individuals were excluded who lacked clinical and/or histopathological findings suggestive of DLE or a mimicker.

After each patient encounter, dermatologists were asked to identify the presence or absence of each candidate item with regard to morphologic characteristics of DLE. Dermatologists then recorded their diagnosis of DLE or another relevant mimicker disease diagnosis. Clinicians were also asked to rank their diagnostic certainty as very certain, certain, neutral, uncertain, or very uncertain.

At each site, if a biopsy specimen of a relevant clinical lesion was available (obtained previously or obtained at this encounter but not for the purposes of the study), 1 central dermatopathologist at each site was asked to review the case. The site dermatopathologist was asked to determine whether the candidate histopathological criteria items were present. As above, dermatopathologists were asked to determine whether the diagnosis was consistent or not consistent with DLE and, if not consistent, to provide an alternative diagnosis. Dermatopathologists were also asked to rank their diagnostic certainty using the same scale from very certain to very uncertain.

Features were compared between groups, and candidate models were identified using best subsets logistic regression analysis to select 1 final model yielding a points-based system for discoid lupus erythematosus classification.

Meaning

This diagnostic study presents the initial validation of classification criteria for discoid lupus erythematosus for use in clinical research.
Table 1. Findings Associated With DLE

<table>
<thead>
<tr>
<th>Item</th>
<th>Consistent with DLE</th>
<th>Not consistent with DLE</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical items associated with DLE, No./total No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erythematous to violaceous in color</td>
<td>87/94 (92.6)</td>
<td>102/120 (85.0)</td>
<td>.09</td>
</tr>
<tr>
<td>Atrophic scarring</td>
<td>78/94 (83.0)</td>
<td>29/120 (24.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Dyspigmentation</td>
<td>79/94 (84.0)</td>
<td>67/121 (55.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Follicular hyperkeratosis/plugging</td>
<td>40/94 (42.6)</td>
<td>13/121 (10.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Scarring alopecia</td>
<td>56/92 (60.9)</td>
<td>25/121 (20.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Location in the concha bowl</td>
<td>45/92 (48.9)</td>
<td>12/120 (10.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Preference for head and neck</td>
<td>82/94 (87.2)</td>
<td>59/121 (48.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Histopathological items associated with DLE, No./total No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interface/vacuolar dermatitis</td>
<td>30/36 (83.3)</td>
<td>26/49 (53.1)</td>
<td>.004</td>
</tr>
<tr>
<td>Perivascular and/or periappendageal lymphohistiocytic infiltrate</td>
<td>35/37 (94.6)</td>
<td>41/49 (83.7)</td>
<td>.18</td>
</tr>
<tr>
<td>Follicular keratin plugs</td>
<td>21/37 (56.8)</td>
<td>10/49 (20.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mucin deposition</td>
<td>27/37 (73.0)</td>
<td>19/49 (38.8)</td>
<td>.002</td>
</tr>
<tr>
<td>Basement membrane thickening</td>
<td>21/37 (56.8)</td>
<td>7/49 (14.3)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Abbreviation: DLE, discoid lupus erythematosus.
* Calculated using $\chi^2$ tests except where expected cell counts were less than 5, which used Fisher exact tests. All items were more associated with DLE except for erythematous to violaceous color and perivascular and/or periappendageal lymphohistiocytic infiltrate.

Statistical Methods

Data were analyzed from March 1 to September 19, 2019. Diagnoses by clinical and dermatopathological features were tabulated and presented as counts and percentages. Clinical features among those with and without diagnoses of DLE were calculated and compared with $\chi^2$ or Fisher exact tests as appropriate. Among those with both clinical and dermatopathological features, agreement was measured with the Cohen $\kappa$ statistic. Candidate models were identified using best subsets logistic regression analysis to predict DLE. Adjusted odds ratios (AORs) and 95% CIs were computed for multivariable models. Differences in the area under the receiver operating characteristic curve (AUC) were tested for statistical significance among models using a nonparametric contrast estimation procedure. Improvement tests, fit statistics, discrimination, and clinical relevance were considered to choose a final model. As a measure of internal validation, the 95% CI for the AUC was calculated using bootstrap simulations, including a deflation factor to account for performance optimism.13 A points-based scoring system was developed using the $\beta$ coefficients from the logistic regression model, with higher scores indicating higher likelihood of DLE diagnosis.14 Cut points in the scoring system were evaluated to identify optimal sensitivity and specificity. Two-sided statistical tests were performed and $P < .05$ indicated significance. Analyses were performed using SAS, version 9.4 (SAS Institute Inc).

Results

DLE Criteria

From December 1, 2017, to February 28, 2019, 9 sites contributed data from 215 patients (178 [82.8%] in the United States, 24 [11.2%] in Japan and South Korea, and 13 [6.0%] in Poland), with 94 patients (43.7%) evaluated as having a diagnosis consistent with DLE. Overall, 189 of 214 patients (88.3%) had findings that were erythematous to violaceous in color; 146 of 215 (67.9%), dyspigmentation; 107 of 214 (50.0%), atrophic scarring; and 81 of 213 (38.0%), scarring alopecia. One hundred forty-one of 215 patients (65.6%) had preference for head and neck, and 57 of 212 (26.9%) had lesions in the conchal bowl. All findings were more prevalent among patients with DLE compared with DLE mimickers, with greatest differences observed for atrophic scarring (78 of 94 [83.0%] vs 29 of 120 [24.2%]; $P < .001$), location in the conchal bowl (45 of 92 [48.9%] vs 12 of 120 [10.0%]; $P < .001$), scarring alopecia (56 of 92 [60.9%] vs 25 of 121 [20.7%]; $P < .001$), and location in the head and neck (82 of 94 [87.2%] vs 59 of 121 [48.8%]; $P < .001$). All comparisons were statistically significant except for differences in erythematous-violaceous in color (87 of 94 [92.6%] vs 102 of 120 [85.0%]; $P = .09$ (Table 1).

Features from dermatopathological evaluation were reported for 86 patients, of whom 76 (88.4%) had perivascular and/or periappendageal lymphohistiocytic infiltrate; 56 of 85 (65.9%), interface/vacuolar dermatitis; 46 of 86 (53.5%), mucin deposition; 31 of 86 (36.0%), follicular keratin plugs; and 28 of 86 (32.6%), basement membrane thickening. Dermatopathological findings were similarly more common among patients with DLE compared with mimickers, with statistically significant differences for basement membrane thickening (21 of 37 [56.8%] vs 7 of 49 [14.3%]; $P < .001$), follicular keratin plugs (21 of 37 [56.8%] vs 10 of 49 [20.4%]; $P < .001$), mucin deposition (27 of 37 [73.0%] vs 19 of 49 [38.8%]; $P = .002$), and interface/vacuolar dermatitis (30 of 36 [83.3%] vs 26 of 49 [53.1%]; $P = .004$) (Table 1).

The most common mimickers included dermatomyositis (n = 31), subacute CLE (n = 15), other forms of cutaneous lupus (n = 8), lichen planopilaris and other lichenoid disorders (n = 21), and psoriasis (n = 10). Clinical findings agreed with histopathological findings in 79 of 86 patients in the sample (91.9%), and the Cohen $\kappa = 0.83$ (95% CI, 0.72-0.95) suggests strong agreement.

Model Development

Fifteen patients who had missing values for any clinical variables, whose diagnoses were uncertain or very uncertain as

jamadermatology.com

© 2020 American Medical Association. All rights reserved.
The final model with β coefficients is provided in the eFigure in the Supplement.

**Model Evaluation**

The internally validated model performance was similar to the apparent model performance measured by the C statistic (AUC, 0.89; 95% CI, 0.85-0.93) via 5000 bootstrap samples. The β coefficients were converted to a points-based system as follows: 3 points for atrophic scarring, 2 points each for location in the conchal bowl or preference for head and neck, and 1 point each for dyspigmentation, follicular hyperkeratosis/plugging, and erythematous to violaceous in color. Based on a cut score of at least 5 points, sensitivity and specificity were 84.1% and 75.9%, respectively. However, as the cut score increases (with a maximum score of 8 points), the specificity of the classification criteria increases. The final model is displayed in Table 3, and the test characteristics for the full point-based system are displayed in Table 4.

**Discussion**

Herein we present the initial validation of classification criteria for DLE. Based on our proposed model, a score of at least 5 yields classification as DLE with sensitivity of 84.1% and specificity of 75.9%, with increasing points yielding higher specificity. These classification criteria can be applied to both localized (lesions above the neck) and generalized (lesions above and below the neck) disease and do not require a biopsy to apply these classification criteria successfully.

A few items, such as scarring alopecia, were not included in the criteria because their inclusion did not substantially change our test characteristics or receiver operating characteristic curve. Furthermore, dermatopathology was not included in the final model. There are several reasons for this. First, only 86 patients (40.0%) had pathological findings available, which limited our ability to meaningfully incorporate these items into classification criteria models. In addition, if a main aim of classification criteria is for enrollment into clinical trials or observational studies, we believe that the feasibility of using our classification criteria is enhanced by not requiring a biopsy for classification. We plan to use our histopathological data to devise a separate DLE histopathological classification criteria set; this work is ongoing.
Abbreviation: DLE, discoid lupus erythematosus.

Dyspigmentation 1
Location in the conchal bowl 2
Preference for head and neck 2
Dyspigmentation 1
Follicular hyperkeratosis/plugging 1
Erythematous to violaceous in color 1

Abbreviation: DLE, discoid lupus erythematosus.

a Specificity increases as the number of points increases, with a maximum score of 8 yielding specificity of 100%.

**Strengths and Limitations**

Strengths of our cohort include relatively large numbers for an uncommon disease, as well as diverse geographic representation and involvement of expert stakeholders from North America, Europe, and Asia. We believe that the classification criteria have good face validity, because nearly all proposed items in the list can be frequently used to diagnose DLE in clinical practice. Furthermore, our data for model development included 9 sites, and we calculated measures of internal validation using bootstrap estimation on the entire sample. This approach was previously demonstrated to be superior to split sample analyses where a proportion of test sites may be used to test independent performance after development on training sites.13

Limitations of our study include that most patients come from specialized referral centers, most of which are in the United States, and that demographic data of patients were not evaluated, because this study sought primarily to assess morphologic and histopathological characteristics alone. These limitations may limit the generalizability of our criteria. Moreover, testing and validating a classification system for DLE is difficult because there is no criterion standard test or criterion apart from expert opinion. We relied on expert opinion from a group of internationally recognized clinical and research leaders in connective tissue disease, including dermatologists, dermatopathologists, and rheumatologists, as our criterion standard, which is similar to the process used in the development of other classification criteria.35,36

**Conclusions**

Importantly, many of the features used in our classification criteria represent features of DLE disease damage rather than disease activity. This is understandable because early active lesions of CLE may be more indistinguishable from other inflammatory dermatoses, with specificity being largely driven by damage characteristics. If one of the purposes of classification criteria is to enroll patients in clinical trials, relying on disease damage may yield a group of patients whose disease is more advanced. It is likely that limiting trials to DLE as opposed to CLE would not allow inclusion of patients with early disease or with more than 1 subtype of CLE. These proposed criteria may be most helpful in defining the subtype of CLE, when possible, for a given patient enrolled in CLE or systemic lupus erythematosus/CLE trials. This would have the added benefit of defining whether an intervention might be effective for DLE relative to early disease or another subtype of CLE. We propose that for clinical trials, additional metrics focused on disease activity, such as the activity score of the Cutaneous Lupus Erythematosus Disease Area and Severity Index, be used to define activity of the underlying DLE as appropriate. It is also worth noting that many patients with CLE may have more than 1 subtype present at any time.

Overall, the importance of DLE classification is highlighted by the need to ensure that patients categorized as having DLE for inclusion in studies do indeed have the disease based on defined characteristics. We hope that classification criteria will provide investigators with a foundation on which to base observational and interventional clinical trials and a common language with which to communicate effectively about this patient population. Classification of DLE is part of a larger process to classify other subsets of CLE as well as other connective tissue diseases, with active efforts under way in dermatomyositis and morphea. All of this is with the aim of advancing knowledge and treatment of these diseases to better care for our patients in the future.
Research Original Investigation

Creation and Validation of Classification Criteria for Discoid Lupus Erythematosus

Pennsylvania (Werth); Department of Dermatology, University of Pennsylvania, Philadelphia (Werth); Division of Rheumatology, Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts (Merola).

Author Contributions: Drs Elman and Joyce served as co-first authors of this article. Drs Werth and Merola served as co-senior authors. Dr Merola had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Elman, Furukawa, Lian, Werth, Merola.

Acquisition, analysis, or interpretation of data: Elman, Joyce, Braudis, Chong, Fernandez, Hasegawa, Kim, Li, Lian, Szepeitowski, Werth, Merola.

Drafting of the manuscript: Elman, Joyce, Fernandez, Furukawa, Li, Merola.

Critical revision of the manuscript for important intellectual content: Braudis, Chong, Fernandez, Furukawa, Hasegawa, Kim, Lian, Szepeitowski, Werth, Merola.

Statistical analysis: Elman, Joyce, Merola.

Obtained funding: Lian.

Administrative, technical, or material support: Elman, Fernandez, Hasegawa, Kim, Li.

Supervision: Elman, Furukawa, Lian, Werth.

Conflict of Interest Disclosures: Dr Chong reported receiving personal fees from ViaBio, Inc, and Beacon Bioscience, Inc, grants from Daavlin Corporation, and other from Pfizer, Inc, and Biogen, Inc, outside the submitted work. Dr Fernandez reported receiving personal fees from AbbVie, Inc, and UCB, grants and personal fees from Novartis International AG and Mallinckrodt Pharmaceuticals, and serving as principal investigator for a phase 2b clinical trial from Pfizer, Inc, and Corbus Pharmaceuticals outside the submitted work. Dr Fernandez reported receiving research support from Mallinckrodt Pharmaceuticals, Novartis International AG, Pfizer, Inc, and Corbus Pharmaceuticals; serving as a consultant for AbbVie, Inc, Novartis International AG, and Mallinckrodt Pharmaceuticals; and serving as a speaker/teacher for AbbVie, Inc, Novartis International AG, and Mallinckrodt Pharmaceuticals. Dr Szepeitowski reported receiving research support from Mallinckrodt Pharmaceuticals, Novartis International AG, Pfizer, Inc, and Corbus Pharmaceuticals; serving as a consultant for AbbVie, Inc, Biogenetica International Laboratories, Leo Pharma A/S, Merck-Serono, Novartis International AG, Pierre Fabre, Sandoz, Inc, and Toray Corporation and a speaker for AbbVie, Inc, Astellas Pharma Inc, Actavis Generics, Adamed Pharma SA, Berlin-Chemie Mennarini, Bioderma Laboratories, Fresenius SE & Co, Jansen-Clag BV, Leo Pharma A/S, Takeda Pharmaceutical Company Limited, and Vichy Laboratories outside the submitted work. Dr Werth reported serving as a consultant for Celgene Corporation, Incorporated, Medimmune LLC, Resolve Therapeutics, LLC, Genentech, Inc, Idera, Inc, Janssen Pharmaceutica, Eli Lilly and Company, Pfizer, Inc, Biogen, Inc, Bristol-Myers Squibb, Gilead Sciences, Inc, Amgen, Inc, Medscape, Nektar Therapeutics, Incyte Corp, EMD Serono, CSL Behring, Princia Biopharma, Inc, Crisalis BioTherapeutics, Viela Bio, Inc, argenx SE, Kyowa Kirin Inc, Regeneron Pharmaceuticals, AstraZeneca plc, Octapharma AG, and GlaxoSmithKline and receiving grants from Celgene Corporation, Incorporated, Janssen Pharmaceutica, Pfizer, Inc, Biogen, Inc, Gilead Sciences, Inc, Corbus Pharmaceuticals, Genentech, Inc, AstraZeneca plc, Viela Bio, Inc, CSL Behring, and Synimmune, Inc. Dr Merola reported serving as a consultant and/or investigator for Merck Research Laboratories, AbbVie, Inc, Dermavant Sciences, Eli Lilly and Company, Novartis International AG, Janssen Pharmaceutica, UCB, Celgene Corporation, Incorporated, Sanofi Regeneron Pharmaceuticals, Almirall SA, Sun Pharmaceutical Industries Ltd, Biogen, Inc, Pfizer, Inc, Incyte Corp, Aclaris Therapeutics Inc, EMD Serono, Avotres Therapeutics LLC, and Leo Pharma A/S outside the submitted work. No other disclosures were reported.

Funding/Support: This study was supported by the Department of Dermatology, Brigham and Women’s Hospital.

Role of the Funder/Sponsor: The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank the International Conference on Cutaneous Lupus Erythematosus, Steering Committee, including Filippa Nyberg, MD, Karolinska University Hospital, Mark Goodfield, MD, Leeds General Infirmary, Branka Marinovic, MD, University Hospital Center Zagren, and Jan Dutz, MD, FRCP, University of British Columbia. We also thank Arash Mostaghighi, MD, MPP, MPH, Brigham and Women’s Hospital, Ruth Ann Vleugels, MD, MPH, Brigham and Women’s Hospital, Meera Tarazi, MD, University of Pennsylvania, and Joyce Okawa, RN, University of Pennsylvania, for their input and assistance with subject recruitment. None of these contributors was compensated for their work. Finally, we thank the Department of Dermatology, Brigham and Women’s Hospital, Boston, Massachusetts, for their administrative assistance.

REFERENCES


© 2020 American Medical Association. All rights reserved.