Ann. rheum. Dis. (1973), 32, 171

Testing the 'Preliminary Criteria for Classification of SLE'

JAMES F. FRIES AND ROBERT C. SIEGEL
From the Department of Medicine (Immunology), Stanford University School of Medicine, Stanford, California 94305, U.S.A.

Systemic lupus erythematosus (SLE) is a disease with varied manifestations involving multiple organ systems, accompanied by multiple laboratory abnormalities, and characterized by exacerbations and remissions. Noting that 'uniform classification of defined groups of patients is necessary in order to assemble and compare data from different sources concerning natural history, evaluation of therapy, and epidemologic description', the Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association recently approached the formidable task of assembling criteria for the classification of SLE (Cohen, Reynolds, Franklin, Kulka, Ropes, Shulman, and Wallace, 1971). To increase the utility of the criteria in population studies, the use of 'exclusions' and 'duration of manifestations' was avoided. A distinction between 'major' and 'minor' criteria was also not used. Thus, the form of these criteria differs from that of the Jones Criteria for Rheumatic Fever and the ARA Criteria for Rheumatoid Arthritis (Arthritis Foundation, 1964).

The committee collected a data base of 245 patients with unequivocal SLE, 234 patients with rheumatoid arthritis, and 217 patients with miscellaneous non-rheumatic diseases. Data was contributed by 52 investigators throughout the United States, who were requested to submit patients with 'unequivocal' SLE, and with 'classic' or 'definite' rheumatoid arthritis. Patients whose clinical course was atypical or in whom diagnostic uncertainty remained were excluded from these two groups.

Data was sought on 74 variables of disease considered of potential value for classification. Adequate data was obtained on 57 potential criteria, and the list was reduced to 21 by exclusion of variables which were least useful in distinguishing patients with 'SLE' from 'RA' and 'other' diseases. The final 21 variables were reduced to fourteen 'criteria' by grouping several variables within a single criterion (Table I, overleaf). The criteria were then tested against the data base, by simple reclassification. A level of four or more positive criteria best distinguished the populations and was recommended by the committee. Tested against the same populations, this level included approximately 90 per cent. of the patients with 'SLE' and excluded all but 1 per cent. of the patients with 'rheumatoid arthritis' and all but 2 per cent. of the patients with 'other' diseases.

This study evaluates the proposed criteria in the context of five questions:

(1) How representative was the ARA test patient population?

(2) How adequate are the criteria when tested against an 'outside' patient population?

(3) How much does the sensitivity and specificity of the criteria depend upon the point in the course of the patient's illness at which they are applied?

(4) How much 'interdependence' is present between various criteria and how much effect does this have on the performance of the criteria?

(5) Which diseases might exhibit four or more positive criteria and what percentage of the time may such 'false positives' be expected?

Two concepts useful for the testing of criteria are proposed:

(1) Testing against a computer-based clinical databank,

(2) Testing against 'simulated' patient populations created by computer from known characteristics of the population.

Methods of procedure

The Stanford clinical computer databank of Rheumatic Disease at present contains nearly one million items of clinical information on 319 patients, including 112 with systemic lupus erythematosus (Fries, 1972). Clinical data are listed under 352 variables for each patient and serial

Accepted for publication September 15, 1972.
This work was supported in part by an Arthritis Clinical Research Center Grant from the Arthritis Foundation and in part by a grant from the Biotechnology Resources Branch of The National Institutes of Health (RR00311-04) to the ACME Computing Facility.

Table I *Preliminary criteria for classification of SLE*

1. Facial erythema (butterfly rash)
   Diffuse erythema, flat or raised, over the malar eminence(s) and/or bridge of the nose; may be unilateral.
2. Discoid lupus
   Erythematous raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesions; may be present anywhere on the body.
3. Raynaud's phenomenon
   Requires a two-phase colour reaction, by patient's history or physician's observation.
4. Alopecia
   Rapid loss of a large amount of the scalp hair, by patient's history or physician's observation.
5. Photosensitivity
   Unusual skin reaction from exposure to sunlight, by patient's history or physician's observation.
6. Oral or nasopharyngeal ulceration
7. Arthritis without deformity
   One or more peripheral joints involved with any of the following in the absence of deformity;
   (*a*) Pain on motion,
   (*b*) Tenderness,
   (*c*) Effusion or periarticular soft tissue swelling.
   (Peripheral joints are defined for this purpose as feet, ankles, knees, hips, shoulders, elbows, wrists, metacarpophalangeal, proximal interphalangeal, terminal interphalangeal, and temporomandibular joints)
8. L.E. cells
   Two or more classical L.E. cells seen on one occasion or one cell seen on two or more occasions, using an accepted published method.
9. Chronic false-positive serological test for syphilis
   Known to be present for at least 6 months and confirmed by TPI or Reiter's tests.
10. Profuse proteinuria
    Greater than 3·5 g./day.
11. Cellular casts
    May be red cell, haemoglobin, granular, tubular, or mixed.
12. One or both of the following:
    (*a*) Pleuritis; good history of pleuritic pain; or rub heard by a physician; or *x*-ray evidence of both pleural thickening and fluid,
    (*b*) Pericarditis, documented by EKG or rub.
13. One or both of the following:
    (*a*) Psychosis,
    (*b*) Convulsions, by patient's history or physician's observation in the absence of uraemia or offending drugs.
14. One or more of the following:
    (*a*) Haemolytic anaemia,
    (*b*) Leucopenia, WBC less than 4,000 per cu. mm. on two or more occasions,
    (*c*) Thrombocytopenia, platelet count less than 100,000 per cu. mm.

observations are available on over 90 per cent. of the patients. The population includes all patients with rheumatic diseases seen at Stanford Medical Center since the system was instituted in 1970, with a mean observation period of 18 months. The data include all manifestations from the onset of the first symptom. Diagnoses are updated at each patient visit. Patients in whom the diagnosis is not considered to be secure may be recorded as 'polyarthritis, undiagnosed', or 'connective tissue disease, type unknown'. Such patients are noted as 'other' in the following tabulations. The characteristics of our SLE patient population (Table II, opposite) are similar to published series (DuBois and Tuffanelli, 1964; Estes and Christian, 1971), and fall midrange in a group of twenty personal communications from investigators at University Medical Centers.

LE-cell tests are not available for all our patients and we have modified the eighth criterion to include either a positive LE-cell preparation or a positive fluorescent antinuclear antibody at a titre of 1:20 or greater. This weakens the performance of the criteria, and data are presented both with the original criteria and with the modified eighth criterion.

Programs were written in PL/ACME, a subset of IBM language PL/1, for use with our databank in an IBM 360/50 digital computer. Each patient in the databank was analysed with regard to the number of positive criteria and to the particular criteria which were positive. The population was first surveyed for positive criteria occurring at any time in the patient course. Secondly, a survey was made of findings at the time of the first visit. Thirdly, assessment of one thousand patient-visits (other than the first) was made to ascertain the frequency with which criteria were positive at a random point in the disease course. Computer programs to determine interdependence of variables, and to simulate patient populations, were written with the same system.

Results

Table II compares the 'unequivocal SLE' population of the committee with SLE patients drawn from three single universities—the Stanford databank population used in this study, a large group of 520 patients reported by DuBois and Tuffanelli (1964), and a recent series reported by Estes and Christian (1971). Interestingly, the call for 'unequivocal' patients in the ARA group led to a very high frequency of positive LE-cell preparations (91 per cent.), butterfly rash (64 per cent.), and photosensitivity (37 per cent.). Similarly, the ARA Rheumatoid Arthritis group was characterized by a very high frequency of deformity (90 per cent.) and rather low frequencies of non-arthritic manifestations (leukopenia, positive LE-cell preparations, Raynaud's phenomenon, pleurisy). The majority of our rheumatoid arthritis patients do not *to date* have deforming arthritis, and the other manifestations are less unusual.

The preliminary criteria perform less well in less classic populations, as shown in Table III. Whereas only 10 per cent. of the ARA lupus patients failed to demonstrate four or more positive criteria at some point in their disease course, 27 per cent. of the databank population failed to do so. Review of these patients failed to reveal errors in clinical diagnosis. For example, recorded as having only one positive criterion and a clinical diagnosis of SLE, was a 23-year-old woman with fever, polyarthralgias (but not arthritis), widespread skin rash (but not discoid

Table II *Frequency of fourteen individual criteria in four lupus populations (per cent.)*

Population	*ARA*	*Databank*	*DuBois* (1964)	*Estes* (1971)
1. Malar rash	64	28	56	39
2. Discoid LE	17	10	15	14
3. Raynaud's phenomenon	20	17	18	21
4. Alopecia	43	38	21	37
5. Photosensitivity	37	13	32	33
6. Oral ulcers	15	9	9	7
7. Arthritis, non-deforming	83	65	65	85
8. LE-cells* or FANA†**	91*	98** (72*)	75*	78*
9. False positive STS	12	3	12	24
10. Profuse proteinuria	20	22	20	26
11. Cellular casts	48	47	40	53
12. Serositis	65	48	65	50
13. Psychosis or convulsions	25	12	20	33
14. Haemolytic anaemia, leucopenia, thrombocytopenia	50	43	50	66
Mean number of positive criteria per patient	5·90	4·53	4·97	5·66

† FANA = fluorescent antinuclear antibody at 1:20 or greater.

Table III *Percentage of patients with four or more positive criteria*

Diagnosis	*'Test' group*	*Databank (FANA)*	*Databank (LE)*
SLE	89·8	73·3	71·1
RA	0·8	19·2	9·9
Scleroderma	—	35·3	16·0
'Other'	—	6·7	4·5

or malar), low serum complement, antibody to double-stranded DNA, positive LE-cell preparations, positive antinuclear antibody, lymphadenopathy, and a positive Coombs test. Several patients with the nephritis of systemic lupus erythematosus, with biopsy documentation, anti-DNA antibody, and positive LE-cell preparations manifested only three of the fourteen criteria.

10 per cent. of our patients with rheumatoid arthritis manifested four or more of the suggested criteria at some point in their course, compared with 1 per cent. of the ARA rheumatoid group. Differences in patient populations account for much of this discrepancy, but the special problem of Felty's syndrome deserves comment. All of the three RA patients with seven positive criteria had Felty's syndrome, with leucopenia, non-deforming arthritis, positive antinuclear antibody, oral ulcers, Raynaud's phenomenon, and two other criteria. Patients with Felty's syndrome will frequently satisfy at least four of the suggested criteria.

16 per cent. of our scleroderma patients satisfy the suggested criteria when cumulative manifestations are considered. These patients nearly all have Raynaud's phenomenon, and quite commonly pleurisy, pericarditis, cellular casts, antinuclear antibody, or arthritis are present. Less than 5 per cent. of the remaining 105 patients with miscellaneous rheumatic or immunological diseases meet the suggested criteria.

The computer databank allows an approach to the interesting question of classification at different points in the disease course (Table IV). The percentage of 'false negative' results in our lupus patients rises from 27 per cent. when cumulative findings are used as in the committee evaluation, to 43 per cent. if only the findings available at the first visit are used, and to almost 70 per cent. if only the findings present at the time of a later patient visit are allowed. These differences obtain partly because of a failure of the physician specifically to identify all criteria at a single patient visit, partly because of suppression of some manifestations by therapy, and partly because relatively few of the criteria are present at the same point in time. A similar trend is present in our rheumatoid arthritis and scleroderma patient populations; the number of false positives decreases when more limited observation periods are used. These findings emphasize that the number of 'positives' in truly similar populations can vary widely depending upon the duration of follow-up available, or on whether the reporting institution emphasizes referral evaluations or continued long-term care. The slightly lower frequency of positive findings in our databank population compared with other reported patient populations probably reflects this same phenomenon, since our follow-up period is somewhat shorter.

Table IV *Performance of criteria at different points in patient course. (Percentage of patients with four or more positive criteria)*

Diagnosis	*Cumulative*	*First visit*	*Later visit*
SLE	73·3	57·2	30·5
RA	19·2	10·8	2·2
Scleroderma	35·3	6·5	5·9

In general, individual criteria are most effective if they are completely independent of other criteria; they are less efficient to the extent that several criteria measure the same aspect of the disease. Since a disease may be regarded as the common cause of its symptoms, it might be expected that all manifestations of the disease would be somewhat interrelated, and linked on the basis of disease severity.

Interdependence of parameters may be measured, using the computer databank. Table V lists interdependence between eight selected 'criteria' as measured in the databank SLE population. If pleurisy is found in 50 per cent. of the patient population and arthritis in 60 per cent. of the population, then 30 per cent. would be 'expected' to have both pleurisy and arthritis. This 'expected' number is combined with the number of simultaneous events actually 'observed' in a computer search of the same population to form a ratio. The numerical values in the Table represent the ratio of 'observed' to 'expected' for the occurrence of both 'criteria' in the same patient, and a value of 1·00 represents 'independence' of the criteria.

Two interesting results emerge from Table V:

(1) The criteria are interrelated, although the magnitude of the interrelationship is only moderate. The mean ratio of 'observed' to 'expected' values is 1·14 for all 28 tests, and is statistically significant at the 0·001 level by 4-fold χ^2 test. This results in the fourteen criteria performing approximately as well as would twelve 'independent' criteria. Some criteria, such as proteinuria and granular casts, are significantly interrelated, while others are very close to being independent.

(2) Some criteria are significantly inversely related. For example, the frequency of arthritis was significantly less than expected in the presence of profuse proteinuria. The finding of inversely-related criteria suggests one or both of two interpretations:

(i) The disease 'SLE' contains two separate subgroups, one characterized by proteinuria, granular casts and anaemia, and the other by arthritis, pleurisy, and leucopenia, or

(ii) Therapy, such as corticosteroids, given for certain manifestations may result in the suppression of other manifestations.

A second method of testing criteria is possible if the frequency with which individual criteria are positive in a particular disease is known or can be estimated, and if the criteria are relatively independent. This method may be used even if an actual patient population is not available, as is the case with rare diseases. If these conditions are met, a 'simulated'

Table V *Interdependence between selected parameters in patients with systemic lupus erythematosus* (Cumulative findings)

Values in this Table represent the ratio: observed/expected for the occurrence of both events in the same patient.
* Statistically significant at 0·05 level.
** Statistically significant at 0·01 level (by 4-fold χ^2 test)

	Proteinuria	Granular casts	Leucopenia	FANA 1:20	Anaemia	Arthritis	Pleurisy
Granular casts	1·75**						
Leucopenia (WBC below 4,000)	0·87	0·75					
FANA 1:20	1·05	1·10	1·41*				
Anaemia (Haematocrit below 30)	1·49*	1·36	1·23	1·23			
Arthritis (non-deforming)	0·44**	0·64*	1·92**	1·09	0·78		
Pleurisy	0·49*	0·96	1·51	1·13	1·48	2·46**	
Skin rash	1·06	1·04	1·95**	1·02	1·42	1·55	1·00

patient population may be created by computer and examined for the presence of positive criteria. A matrix is constructed which lists the frequencies with which given criteria are found positive over the cumulative course of given diseases. Difficulties in matrix construction are offset by the ability of this technique to 'test' rare diseases in which a prospective evaluation would be impossible, and to estimate frequencies of variables not often recorded in retrospective records.

To 'simulate' a patient with a given disease, the computer generates a random number between zero and one and compares it with the first frequency in the disease column of the matrix. If the number is less than the frequency, then the criterion is considered positive. The process is repeated with the second and subsequent frequencies until a total number of positive criteria is established for the simulated patient. The procedure is performed for 1,000 simulated patients with a given disease and the results tabulated. The entire procedure may be quickly repeated with any desired changes in the given frequencies.

Table VI compares results with 'observed' and 'simulated' patients in five patient groups, and thus compares the two methods. In each case the frequencies used to create the 'simulated' population were derived from the parallel 'observed' population. The criteria always seem to perform better in the simulated population than in the actual one; this reflects the complete independence of criteria in the simulated population. However, the differences are relatively small. In these populations, the artificial model appears to provide a reasonably accurate approximation of the effectiveness of the criteria. Using simulated populations in SLE, four or more positive criteria are found in 94 per cent. of the ARA patients, 91 per cent. of the Estes series, 83 per cent. of the Dubois series, and 74 per cent. of our databank patients. These values are sufficiently close to those obtained by actual count, so that 'simulated' populations may be used to estimate results which would be obtained in actual populations.

Table VII lists results of tests of simulated populations of patients with other diseases against the suggested criteria. In addition to diseases already discussed, significant numbers of 'false positives' are to be expected in drug-induced SLE, Wegener's granulomatosis, discoid LE, 'lupoid' hepatitis, polyarteritis, juvenile rheumatoid arthritis, psoriatic arthritis, Reiter's syndrome, sarcoidosis, and acute rheumatic fever. Results in actual populations would be expected to show somewhat higher numbers of false positives, because of the interdependence of criteria.

Table VII *Criteria testing with simulated populations of 'other' diseases*

Disease	*Percentage with four or more criteria*
Drug lupus	23·5
Scleroderma	19·9
Wegener's granulomatosis	18·1
Discoid LE	14·8
Felty's syndrome	13·6
'Lupoid' hepatitis	13·4
Polyarteritis	11·1
Rheumatoid arthritis	9·9
Juvenile rheumatoid arthritis	2·4
Psoriatic arthritis	2·2
Reiter's syndrome	1·6
Sarcoidosis	0·9
Acute rheumatic fever	0·6
Gonococcal arthritis	0·3
Ankylosing spondylitis	0·1
Acute gout	0·1
Porphyria cutanea tarda	0·1
Chronic gout	0·0

Discussion

Several intuitive objections to the proposed criteria prompted this study.

(1) The criteria utilized 'simple reclassification' and was tested against 'classic' cases. Testing against a population other than that used to derive the criteria or against a more usual distribution of 'classic' and 'unusual' cases should result in less accurate discrimination. Even in the study population, the reported 90 per cent. 'sensitivity' appeared relatively low.

(2) The 'classic' population tested appeared unusual in several respects. Specifically, the frequency of malar 'butterfly' rash and of positive LE cell preparations appeared high.

(3) The committee did not have sufficient data to allow the use of antinuclear antibody tests as a

Table VI *Comparison of 'observed' and 'simulated' patient populations*
Percentage with four or more positive criteria

Population	*ARA lupus*	*Databank lupus*	*ARA rheumatoid*	*Databank rheumatoid*	*Databank scleroderma*
Observed	89·8	73·3	0·8	19·2	35·3
Simulated	93·7	74·3	0·1	9·9	19·9

criterion, despite widespread belief that these tests are the most sensitive method currently available for detection of patients with SLE.

(4) The wording of some criteria suggests circularity. For example, 'butterfly' malar erythematous rash and 'non-deforming' arthritis are so closely associated with SLE that it is doubtful whether they would be sought for or applied equally to patients thought of as 'SLE' or as 'RA'.

(5) Several groups of criteria appear to be closely related to each other, and therefore somewhat redundant. The first six criteria refer to cutaneous or mucous membrane manifestations. Erythematous rash and photosensitivity are clinically very closely linked. Similarly, a finding of granular casts is to be expected in patients with proteinuria.

(6) The criteria were tested against patients with well-developed disease and many manifestations. We do not know whether the criteria would be effective earlier in the clinical course of these patients, or at a time in which they were in remission.

(7) The criteria could not separate drug-induced SLE from the idiopathic form, a distinction which seems to be clinically important.

(8) In contradistinction to the ARA criteria for rheumatoid arthritis, the sensitivity of the individual criteria is very low, of the order of 30 to 40 per cent. With a disease of varied presentation and many possible features, it is possible for a patient to have many manifestations of the disease but very few 'criteria'. Only one feature of SLE, positive anti-nuclear antibody tests, approaches 100 per cent. in sensitivity.

Evaluation supports these objections. The criteria perform less well in patient populations other than that from which they were derived. They perform less well in patients seen for short periods of time, or early in their course, than when the opportunity to survey cumulative findings is present. Lack of independence of the individual criteria decreases their effectiveness, although only moderately. There are considerable numbers of 'false positives' in a variety of disease states, as well as 'false negatives' in SLE patients. These problems limit the usefulness of the criteria as at present formulated.

Classification criteria must be tested by comparison with clinical diagnoses made without criteria; this inevitably raises questions about the adequacy of the clinical diagnoses. Three indirect tests of the clinical diagnoses are possible:

(1) The characteristics of the population should be consistent with the literature and with experience in other centres; our data were approximately midrange in the populations we compared.

(2) If over-diagnosis were present in our patients, the discrepancies between classification diagnosis and clinical diagnosis should have identified the patients in whom the clinical diagnosis was most suspect. This was not the case.

(3) If findings were not diligently sought and data recording was inadequate, then lower frequencies should be found in all patient groups, and false positives should be rare. We found instead an increase in the number of false positives.

Lack of methodology for criteria-testing has limited the development of clinical criteria. Accumulation of patients for a large-scale field test is a long process, and cannot be frequently repeated. The effectiveness of a proposed set of criteria may be reasonably accurately determined with relative ease by utilizing 'computer-stored' and 'computer-simulated' patient populations. Testing a 'computer-stored' patient population is direct, but is limited by its retrospective nature; it cannot be used if the proposed criteria have not been stored, or in disease states not adequately represented in the databank. Testing a 'computer-simulated' patient population avoids these limitations, but accuracy is hindered by problems in developing adequate symptom frequencies and by the assumption of independence. The techniques complement each other and taken together provide a means of estimating the accuracy of criteria without a prospective field trial. Criteria may be revised and re-tested with the same techniques. Alternative sets of criteria may be competitively evaluated. The effects of varying individual criteria can be measured. The testing of proposed criteria before publication can promote improvement in criteria before their widespread dissemination.

Summary

Preliminary criteria for the classification of systemic lupus erythematosus have been evaluated using computer-stored and computer-simulated test patient populations. The proposed criteria perform less well in these populations than in that from which they were derived. They perform best when the opportunity to survey cumulative findings is present, and less well in patients seen for short periods of time or early in the course of their illness. Lack of independence of the individual criteria somewhat diminishes their effectiveness. There are considerable numbers of 'false positives' in a variety of disease states, as well as 'false negatives' in SLE patients. These problems limit the usefulness of the criteria as at present formulated. The ability of the computer rapidly to search stored patient information and

quickly to create 'simulated' patient populations with desired characteristics allows repetitive revision and re-testing of proposed criteria, with final selection of the best. These capabilities should be exploited in future approaches to the formal definition of complex diseases.

The authors wish to thank Dr. Byron W. Brown, Dr. Halsted R. Holman, Dr. Hugh O. McDevitt, Dr. Alfonse T. Masi, and Dr. Harry Robinson for statistical guidance and constructive comments. Mrs. Helen Page, Mrs. Wendy Cramer, and Miss Bonnie Obrig provided technical assistance.

References

ARTHRITIS FOUNDATION, NEW YORK (1964) 'Primer of the Rheumatic Diseases'

COHEN, A. S., REYNOLDS, W. E., FRANKLIN, E. C., KULKA, J. P., ROPES, M. W., SHULMAN, L. E., AND WALLACE, S. L. (1971) *Bull rheum. Dis.*, **21**, 643 (Preliminary criteria for the classification of systemic lupus erythematosus)

DUBOIS, E. L., AND TUFFANELLI, D. L. (1964) *J. Amer. med. Ass.*, **190**, 104 (Clinical manifestations of systemic lupus erythematosus)

ESTES, D., AND CHRISTIAN, C. L. (1971) *Medicine* (Baltimore), **50**, 85 (The natural history of systemic lupus erythematosus by prospective analysis: computer analysis of 520 cases)

FRIES, J. F. (1972) *J. Amer. med. Ass.*, **222**, 1536 (Time-oriented patient records and a computer databank)