SPECIAL ARTICLE

Rigorous New Approach to Constructing a Gold Standard for Validating New Diagnostic Criteria, as Exemplified by the Eosinophilia-Myalgia Syndrome

Phillip A. Hertzman, MD; Daniel J. Clauw, MD; Joseph Duffy, MD; Thomas A. Medsger, Jr, MD; Alvan R. Feinstein, MD

Background: Constructing diagnostic criteria, a common problem in clinical medicine, is particularly difficult for diseases that lack a pathognomonic "gold standard." To develop an improved strategy for constructing such criteria, we used the eosinophilia-myalgia syndrome as an example. The goal, for research classifications, was to construct validated clinically sensible criteria and to develop improved methods that can be used for other disorders.

Methods: Using a "pattern-based" approach with data from several separate sources, a committee of investigators first prepared and informally tested criteria for the diagnosis of eosinophilia-myalgia syndrome. A gold standard challenge set of reports of cases and noncases was independently generated and separately validated by an external panel of clinical experts. The criteria were then tested using the gold standard set, and interobserver variability and diagnostic accuracy were determined.

Results: Interobserver variability showed the following mean proportionate agreements: 98.7% for the presence of specific criteria elements, 99% to 100% for diagnosis, and 97% to 98% for diagnostic pattern. κ Values were correspondingly high. Diagnostic accuracy showed sensitivity at 88%, specificity at 97%, and overall accuracy at 92%.

Conclusions: The proposed criteria are accurate and reproducible, and can be used in future clinical investigations of the eosinophilia-myalgia syndrome. The new strategy and methods developed for this challenge can be valuable for solving analogous problems in constructing criteria for other clinical disorders.

Arch Intern Med. 2001;161:2301-2306

From the Los Alamos Medical Center, Los Alamos, NM (Dr Hertzman); Division of Rheumatology, Immunology, and Allergy, Department of Medicine, Georgetown University School of Medicine, Washington, DC (Dr Clauw); Division of Rheumatology, Department of Internal Medicine, Mayo Clinic, Rochester, Minn (Dr Duffy); Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pa (Dr Medsger); and Department of Internal Medicine and Epidemiology, Yale University School of Medicine, New Haven, Conn (Dr Feinstein).

DESPITE THE importance of a well-delineated method for defining criteria for new diseases,[1] constructing classification or diagnostic criteria is a particularly difficult challenge for conditions, such as rheumatologic or neurologic disorders, that cannot be identified from a "gold standard" feature provided by unique morphologic characteristics or laboratory test results. Criteria for such disorders have been previously developed with either a peremptory or a trial-and-error process. In the peremptory method, the components of the criteria are directly chosen by a single person,[2] a committee,[3-6] or a mathematical algorithm.[7,8] In the trial-and-error approach,[9-13] the initially proposed elements considered important in the diagnoses of eosinophilia-myalgia syndrome (EMS) and related disorders are selected for inclusion only after being individually evaluated for efficacy using diverse statistical tests and/or judgmental consensus.

In confronting the challenge of delineating diagnostic criteria, we wanted to use the trial-and-error approach to develop clinically "sensible" criteria, but we also wanted to improve some of the imperfections in the

EMS Expert Panel

Physicians

Park Nicollet Clinic, Minneapolis, Minn: Caroline Bowles, Scott Glickstein; *University of Pennsylvania School of Medicine, Philadelphia:* Bruce Freundlich; *Minor and James Medical Group, Seattle, Wash:* Philip J. Mease; *University of California School of Medicine, San Francisco:* Kenneth Sack; *Medical University of South Carolina, Charleston:* Richard M. Silver; *Georgetown University School of Medicine, Washington, DC:* Virginia D. Steen; *Cleveland Clinic Foundation, Cleveland, Ohio:* Holly L. Thacker; *University of Illinois School of Medicine, Chicago:* John Varga.

1. Construction and Modification of Preliminary Criteria With Informal Testing

List and Prioritize Important Clinical Elements of the Disease
Identify Clinical Patterns of the Disease
Specify Glossary of Terms
Designate Exclusions
Construct Preliminary "Pattern-Based" Criteria
Informally Test Criteria Using a "Learning Set" of Cases and "Noncases"
Modify Criteria Based on Informal Testing and Consensus

2. Development of a "Gold Standard" Set to Test the Accuracy of the Criteria

Enlist an Independent External Expert Panel
Experts Submit Reports Representing Disease and Nondisease Cases
Experts Blindly Diagnose Each Report
Expert Panel Diagnoses Become the Gold Standard for Testing the Criteria

3. Formal Testing of the Criteria Using the Gold Standard Set

Committee Members Blindly Diagnose Each Gold Standard Report Using Proposed Criteria
Diagnoses Using the Criteria Are Compared With the Gold Standard Diagnoses of the External Experts
Interobserver Variability and Diagnostic Accuracy Are Calculated
If Criteria Performance is Unacceptable, Revise Criteria and Retest With a New Set of Gold Standard Cases

4. Preparation of the Final Version of the Criteria

Final Minor Revisions in Wording Based on Results of the Formal Testing, Review of the Participants' Comments, and Committee Discussion

The stages of construction for the criteria.

prior development of analogous criteria.[14] In addition to our personal interest in EMS,[15] we chose to work with this disease because (1) it had been a diagnostic challenge in clinical practice and epidemiological studies; (2) the original Centers for Disease Control and Prevention surveillance criteria,[16] developed by consensus of a committee, had not been validated or tested for performance; and (3) various experts had urged that an "additional clinical definition" was needed for EMS.[17]

Before this project's inception, a new set of criteria had been proposed by one of us (P.A.H.), based on personal experience and on a review of the clinical reports of 20 patients selected from a published literature of 245 EMS cases.[18] Appearing in a published summary[19] of a meeting of EMS researchers in December 1994, these early criteria recognized EMS as a multisystem disease[14,17,20-22] with an early phase,[23] distinctive but not unique histopathological features, and a more diverse late phase. The early new criteria served as a starting point for this project.

We report the new approach and methods, because they could be particularly valuable, not just for the EMS challenge but also for other disorders in the future.

BASIC STRATEGIES

A crucial consideration in constructing disease criteria is the availability of an acceptable gold standard for testing accuracy. Because EMS lacks discrete pathognomonic features, we thought that the best surrogate gold standard for testing the criteria would be a set of reports of EMS cases and noncases that were generated and validated by an external panel of experts. In the past, when gold standard reports for testing disease criteria have been collected, the common approach has been to accept the reports provided by expert clinicians who are usually members of the same group developing the criteria. Instead, for the present study, we developed a unique approach in which the gold standard reports would come from the consensus of an independent panel of experts who had extensive clinical experience with this disease, but who did not participate in the development or testing of the criteria. These "external" experts would submit reports of cases and noncases and would review reports that they had not submitted. Those reports in which the experts agreed strongly on the diagnosis would become part of the gold standard set. Although any gold standard, regardless of method of construction, would be arbitrary, we believed that this approach would result in the most appropriate objective standard possible. The gold standard reports would then be used separately to test the criteria.

In planning the new criteria, we chose a format containing patterns of disease that combined groups of clinical and laboratory "descriptors" or elements. As exemplified by the modified Jones criteria for rheumatic fever[24] and by the TNM staging system for cancer,[25] such patterns are easy to understand and can readily be appraised for clinical "sensibility." This type of sensibility is not immediately apparent when a diagnosis comes from arbitrary counts of an array of possible elements[10] or from purely mathematical rules for preparing either weighted scores[7,26] or arbitrarily partitioned clusters.[7,26]

To promote reproducibility, the elements of the criteria were defined in an accompanying glossary. The performance of the elements was initially checked for consistent use among observers, and then rechecked after pertinent modifications. As a last step, the modified criteria would be challenged for accuracy against the gold standard set of cases and noncases.

METHODS

Three rheumatologists (D.J.C., J.D., and T.A.M.) and a clinical epidemiologist (A.R.F.) joined one of us (P.A.H.) as a "coordinating committee" to develop the methods described herein.

The project involved 4 stages: (1) construction and modification of preliminary criteria via consensus and informal testing by the coordinating committee, (2) development of a gold standard set of EMS cases and noncases by an independent panel of experts, (3) formal evaluation of the reproducibility and accuracy of the criteria when tested against the gold standard set, and (4) preparation of the final version of the criteria. The 4 stages are outlined in the **Figure** and discussed in the sections that follow.

CONSTRUCTION AND MODIFICATION OF THE NEW CRITERIA

Initial Phase

The committee compiled an initial list of 45 elements considered important in the diagnosis of EMS and related disorders, and then, by consensus, reduced the list to 10 elements. To identify clinical patterns of EMS,

the committee next studied a set of 12 "case scenarios," prepared by the coordinator (P.A.H.), showing diverse presentations of EMS. Each scenario was analyzed to consider various clinical patterns of the illness and to select appropriate elements to include in the criteria.

Development of Preliminary Criteria

The consensus plan was to diagnose EMS according to combinations of elements that fit within any 1 or more of several possible patterns that were believed to represent the typical clinical presentations of patients with the syndrome. Although this basic structure was maintained, the proposed criteria were modified several times before reaching the final version shown in **Table 1**. To ensure consistent use, each of the 10 elements was defined in a glossary of terms (available from the authors). In addition, certain illnesses that would preclude the diagnosis of EMS were designated as exclusions.

Informal Evaluation and Modification of the Preliminary Criteria

To informally test the preliminary criteria, the coordinating committee members, from their own personal records, next prepared a "learning set" of 25 case reports of patients with either EMS or a closely related illness. Each patient in this set was then diagnosed using the preliminary criteria as "having EMS" or "not having EMS." This exercise identified problems that led to further refinement in structure of the criteria and in definition of the elements.

DEVELOPMENT OF A GOLD STANDARD SET OF DISEASE AND NONDISEASE CASES

The next step was to assemble an independently prepared gold standard set of EMS and non-EMS cases for formal testing of the accuracy of the proposed criteria.

Enlistment of Independent External Expert Panelists

Nine physicians with extensive clinical and research experience with EMS were recruited as members of an external panel of experts. Each member was asked to submit 10 summary reports of patients from his or her institution. The reports were to include 5 patients diagnosed as having EMS; 4 with other illnesses resembling EMS; and 1 with possible, but uncertain, EMS. Each report contained a brief narrative, copies of pertinent ancillary tests, and a summary placed on a data template, which included more than 75 clinical and laboratory entries. The data template had been pilot tested by 2 members of the committee (P.A.H. and T.A.M.) for ease and consistency of use.

Of the 9 expert panelists, 7 submitted 10 reports and 2 submitted 11. All 92 reports were reviewed by the coordinator, who contacted the panel members to clarify inconsistencies and to obtain available but unreported missing data. One report was discarded because of inadequate data. The remaining 91 summaries were then edited to remove any potentially biasing information. After preparation in a standard format, these reports were assigned randomized identification numbers to ensure against grouping by submitting physician or by diagnosis.

Independent Definition of the Gold Standard Set

To develop a pertinent gold standard for testing the proposed criteria, the "challenge set" of 91 reports was sent to each member of the panel of experts, omitting any instances submitted by that expert. Each expert was asked to use personal clinical experience and judgment to diagnose each report as "EMS" or "not EMS."

The gold standard diagnostic category for each report was determined from a count of votes by the panel of experts, without regard to the originally submitted diagnoses. We decided that for a report to be accepted as a gold standard, 75% or greater agreement among the expert panel votes would be required. Fifty reports for which the expert panel voted EMS by margins of 8:0, 7:1, or 6:2 were accepted as EMS cases. Thirty-five reports for which the expert panel voted not EMS by the same margins were considered non-EMS cases. The 6 reports with votes of 5:3, 4:4, or 3:5 were considered indeterminate and were, therefore, not used as gold standard cases for testing the diagnostic accuracy of the criteria. These cases, however, were later included for testing interobserver agreement by the 3 committee members for decisions regarding the presence of criteria elements.

FORMAL TESTING OF REVISED CRITERIA

Without knowledge of the gold standard diagnosis for each report, 3 members of the coordinating committee (D.J.C., J.D., and T.A.M.) independently applied the criteria to

Table 1. Proposed Eosinophilia-Myalgia Syndrome (EMS) Criteria

EMS can be diagnosed if either pattern 1 or 2 is satisfied. Elements within each of these patterns must fulfill specified definitions (available from the authors).
Pattern 1
Presence of a documented illness of abrupt or relatively discrete onset accompanied by evidence, in the absence of the exclusions noted below, of all 3 of the following manifestations within 6 mo of onset: (1) eosinophilia; (2) myalgia; and (3) at least one of rash, edema, pulmonary involvement, or neuropathy.
Pattern 2
Presence of an illness, with or without a documented early episode, accompanied by one of the following combinations of manifestations, in the absence of the exclusions noted below, occurring within 24 mo of illness onset: (1) fasciitis, neuropathy, and myalgia or muscle cramps; or (2) any 3 or more of fasciitis, myopathy, neuropathy, or eosinophilia (within 6 mo of onset).
Exclusions
EMS should not be diagnosed in the presence of trichinosis, vasculitis, or any other documented infectious, allergic, neoplastic, connective tissue, or other type of disease that could adequately explain the clinical manifestations.

Table 2. Interobserver Variability Using the Proposed Criteria

Committee Members*	No. of Comparisons†	Agreement, No. (%)	κ
	For Diagnosis		
A vs B	86	85 (99)	0.98
B vs C	81	80 (99)	0.98
A vs C	76	76 (100)	1.00
	For Pattern of Diagnosis		
A vs B	36	35 (97)	0.96
B vs C	40	39 (98)	0.96
A vs C	39	38 (97)	0.97

*The committee members (A, B, C) were 3 of the authors (D.J.C., J.D., and T.A.M.).
†Of the 91 reports, committee members did not review reports that originated from their own institution. Pattern of diagnosis was determined for each pertinent case diagnosed as eosinophilia-myalgia syndrome by each committee member.

Table 3. Diagnostic Accuracy of the Proposed Criteria*

Diagnosis Using the Proposed Criteria	Gold Standard Diagnosis		
	EMS	Not EMS	Total
EMS	44	1	**45**
Not EMS	6	34	**40**
Total	**50**	**35**	**85†**

*The sensitivity was 88% (44/50), the specificity was 97% (34/35), and the overall accuracy was 92% (78/85). EMS indicates eosinophilia-myalgia syndrome.
†Of the 91 reports, the diagnoses in 6 were considered indeterminate and were not used in calculating the diagnostic accuracy.

the 91 summaries from the gold standard set, omitting any reports originating from their own institutions. Working alone and using a standard work sheet, the committee members first determined which elements of the criteria were present in each report. The members next determined whether the case was EMS or not EMS, based solely on the proposed criteria. For those that satisfied the EMS criteria, the pattern or patterns that were fulfilled were noted. The criteria-based diagnosis assigned to each report was based on a majority of the 3 determinations.

Interobserver variability in applying the criteria for the presence of criteria elements, diagnosis, and diagnostic pattern was calculated for pairs of the 3 committee members for all study reports and was expressed in percentage agreement and κ indexes.

The estimation of diagnostic efficacy (sensitivity, specificity, and overall accuracy) was based on the ability of the criteria to correctly diagnose the gold standard reports. This was calculated by comparing the criteria-based diagnoses with the gold standard diagnoses. The *sensitivity* of the criteria was defined as the proportion of the 50 gold standard EMS cases correctly diagnosed by the criteria, and the *specificity* was defined as the correctly diagnosed proportion of the 35 gold standard non-EMS cases. *Overall accuracy* was defined as the percentage of correct diagnoses among all 85 gold standard cases.

PREPARATION OF THE FINAL CRITERIA

After the results of the formal testing, review of the participants' comments, and considerable discussion, the committee then made only minor changes in the wording that appears in the "final" version of the glossary of terms (available from the authors). If analysis had shown poor performance of the criteria, requiring significant revisions, it would have been necessary to generate a new set of gold standard cases and noncases before retesting the revised criteria.

Examination of reports with discrepancies between the submitted diagnosis, the gold standard expert panel diagnosis, and/or the diagnosis by criteria led to identifying criteria elements most often associated with difficult decisions, and to recognition of the illnesses that are most difficult to distinguish from EMS.

RESULTS

The 92 reports submitted by the external panelists included 68 women and 24 men (mean age, 46.8 years). Of the 92 reports, 47 were submitted as instances of EMS; the other diagnoses included eosinophilic fasciitis, fibromyalgia, systemic sclerosis, Churg-Strauss arteritis, generalized vasculitis, systemic lupus erythematosus, undifferentiated connective tissue disease, eosinophilic leukemia, idiopathic thrombocytopenic purpura, coronary artery disease, hypothyroidism, polymyalgia rheumatica, giant cell arteritis, chronic pain syndrome, and indeterminate cause.

INTEROBSERVER VARIABILITY

The 3 committee members (A, B, and C) could be arranged into 3 pairs (AB, AC, and BC) for comparisons. The 3 pairs had 98.7% overall mean agreement regarding the presence of specific criteria elements in each report. The 3781 comparisons contained 48 disagreements (1.3%). In determining the presence or absence of specific criteria elements, the most frequent disagreements involved the following: diagnosing neuropathy when neuropathic symptoms or signs deviated from the usual patterns of peripheral neuropathies; ascribing abnormal histopathological features in the presence of inconsistent terminology in histopathological reports; and interpreting imprecise descriptions of weakness, myalgia, and rashes.

Using the criteria to determine the diagnosis in each report, agreement among pairs of committee members who applied the criteria ranged from 99% to 100% (κ range, 0.98-1.00) for diagnosis, and from 97% to 98% (κ range, 0.96-0.97) for pattern of diagnosis (**Table 2**).

ACCURACY

When compared against the gold standard diagnoses, the criteria had a sensitivity of 88%, a specificity of

97%, and an overall accuracy of 92% (**Table 3**). Eosinophilic fasciitis and Churg-Strauss arteritis were the 2 most difficult diagnoses to distinguish from EMS.

COMMENT

CLINICAL APPLICATIONS

In addition to having a scientific method of development, the new criteria are clinically highly sensitive and specific. Their level of accuracy was similar to or better than levels accepted in prior studies[7,12] of criteria for rheumatic diseases.

We stressed specificity because EMS has a low incidence in the general population and in targeted populations of tryptophan users.[27-31] Consequently, low specificity would lead to many false-positive diagnoses, particularly in the general population, which has a high prevalence of many late symptoms of EMS. In addition, because no therapy for EMS has been shown to be efficacious, false-positive diagnoses may produce needless stress for misdiagnosed patients. The high specificity achieved by the new EMS criteria shows that they are well suited for classifying patients for research studies. Beyond the goal of improving specificity, we also recognized that EMS can occasionally occur without myalgia or documented eosinophilia. The new criteria have corrected this limitation of the original Centers for Disease Control and Prevention surveillance criteria.[16]

Like all other criteria created for research purposes, the new criteria will not identify a few patients in clinical practice. In addition, because the criteria were intended to allow discrimination with a minimum number of elements, the full spectrum of disease manifestations of EMS is not included. However, if more clinical and/or laboratory elements were added to accommodate occasional patients with less common features of EMS, the specificity would be reduced and more false-positive diagnoses would occur.

For guidance in clinical practice with individual patients, the criteria can serve as a starting point. In a separate publication, we will present a supplementary list of features, not included as components in the criteria, that can be potentially useful in diagnosing the conditions of individual patients. We will also (1) describe difficult cases that evoked diagnostic errors and (2) compare the performance of the new criteria with that of the Centers for Disease Control and Prevention surveillance criteria.

Recognizing that determining cause and developing diagnostic criteria involve 2 separate processes and sets of reasoning, and that including a presumed causative agent prematurely might produce biased diagnoses, we, and the Centers for Disease Control and Prevention previously, did not include tryptophan use as an element in the criteria. Our decision was based on the nosologic principle that an alleged causative agent should be avoided in the diagnostic criteria for a disease until scientific evidence supporting the pathogenesis of the disease has been unequivocally demonstrated. Since a single precise cause of EMS has not yet been determined, a requirement for antecedent use of tryptophan would preclude diagnosis of EMS when it occurs, as has been reported,[32,33] in the absence of tryptophan. Our criteria are highly sensitive and specific without the inclusion of tryptophan as a required element.

Analogously, although concomitant group A streptococcus infections have been demonstrated often enough to be included in the most recent version of the modified Jones criteria for rheumatic fever,[24] this element was not specified in the original version,[2] which was proposed before modern antibody tests had shown the necessary streptococcal evidence. Diagnostic bias related to assumed cause has been shown in the past when the identical clinical scenario was often diagnosed as either toxic shock syndrome or something else, depending on whether the use of a tampon was mentioned in the scenario.[34] Similar biases have been demonstrated for the diagnosis of EMS in scenarios that indicate the presence or absence of tryptophan use.[35]

STUDY LIMITATIONS

The main limitation of our study was a dependence on retrospective data that varied substantially in documentary quality and quantity. The problem of missing data, also common in prior attempts to construct disease criteria,[15] can be reduced in the future if data are collected prospectively using a pretested template with precise instructions for the physicians who interview and examine patients.

The low rate of interobserver variability among the committee members who applied the criteria to the challenge cases may not occur for others. Since the members were involved in developing the criteria, they might be more prone to agreement than others who are not previously acquainted with the criteria. Interobserver variability might be better demonstrated if another independent group of clinicians or researchers was enlisted to perform this part of the exercise.

Like the Jones criteria for rheumatic fever,[2,24,36] which were modified after improved causative and diagnostic information became available, the EMS diagnostic criteria are not necessarily final. They will require modification as more knowledge develops about the cause and pathogenesis of this disorder.

NEW APPROACHES AND ACCOMPLISHMENTS

The considerable effort and criteria development during the 3-year course of this project used time-honored techniques,[37,38] but also included the new and probably unique approaches described in the "Basic Strategies" and "Methods" sections.

The various stages of criteria development included the many groups of real or prototype patients described earlier, whereas most previous efforts in criteria construction have relied on only 1 set of patient data. In the process herein, we used different sets of patients for training and testing. A particularly important innovation was the use of an external panel of experts to independently provide and validate gold standard reports for testing the criteria. Although required for any diagnosis, the gold standards chosen for illnesses that lack unique features have seldom been previously independently supported by spe-

cific data. We believe our approach accomplished this goal in an effective unbiased manner. The gold standard case sets were defined and the criteria were interpreted independently, so that the results of one process did not influence the other.

The structure of the criteria led to a reasonably small number of satisfactory combinations, which corresponded to clinically sensible patterns of illness. The template used for data collection reduced bias by including a wide variety of information that might occur in many clinical situations, beyond the particular features chosen as criteria elements.

The proposed criteria can be used in future clinical investigations to study the epidemiological features and natural history of EMS. Together with the cited supplementary information, the criteria can enhance communication about EMS and can provide a starting point for diagnosing the conditions of individual patients. Although recent cases of EMS have not been reported, many[1,39,40] believe that eventually similar syndromes may occur. This, in part, is based on the recognition that 8 years before the outbreak of EMS, there was an outbreak of an almost indistinguishable condition called the toxic oil syndrome in Spain.[41] The proposed criteria for EMS could also be useful for the study of such future outbreaks. Finally, this new approach can be valuable in constructing criteria for other ailments whenever pertinent new challenges arise.

Accepted for publication February 22, 2001.

This study was supported by an unconditional grant from Showa Denko K.K., Tokyo, Japan.

We thank Carolyn K. Wells, MPH, for the data processing; and James Sussman, DO, Peter Lindberg, MD, and Joanna Fair, PhD, for their thoughtful reviews of the manuscript.

Corresponding author and reprints: Phillip A. Hertzman, MD, Los Alamos Medical Center, Suite 130, West Road, Los Alamos, NM 87544 (e-mail: phertz1@aol.com).

REFERENCES

1. Miller FW, Hess EV, Clauw DJ, et al. Approaches for identifying and defining environmentally associated rheumatic disorders. *Arthritis Rheum.* 1999;43:243-249.
2. Jones TD. Diagnosis of rheumatic fever. *JAMA.* 1944;126:481-484.
3. Fox RI, Robinson CA, Curd JG, Kozin F, Howell FV. Sjogren's syndrome: proposed criteria for classification. *Arthritis Rheum.* 1986;29:577-585.
4. Holmes GP, Kaplan JE, Schonberger LB, et al. Chronic fatigue syndrome: a working case definition. *Ann Intern Med.* 1988;108:387-389.
5. Schumacher GA, Beebe G, Kibler RF, et al. Problems of experimental trials of therapy in multiple sclerosis. *Ann N Y Acad Sci.* 1965;122:552-568.
6. De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet.* 1996;62:417-426.
7. Hunder GG, Arend WP, Bloch DA, et al. The American College of Rheumatology 1990 criteria for the classification of vasculitis: introduction. *Arthritis Rheum.* 1990;33:1065-1067.
8. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee: Diagnostic and Therapeutic Criteria of the American Rheumatism Association. *Arthritis Rheum.* 1986;29:1039-1049.
9. Cohen AS, Reynolds WE, Franklin EC, et al. Preliminary criteria for the classification of systemic lupus erythematosus. *Bull Rheum Dis.* 1971;21:643-648.
10. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* 1982;25:1271-1277.
11. Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum.* 1980;23:581-590.
12. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the Multicenter Criteria Committee. *Arthritis Rheum.* 1990;33:160-172.
13. Wallace SL, Robinson H, Masi AT, Decker JL, McCarty DJ, Yu TF. Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum.* 1977;20:895-900.
14. Bloch DA, Michel BA, Hunder GG, et al. The American College of Rheumatology 1990 criteria for the classification of vasculitis: patients and methods. *Arthritis Rheum.* 1990;33:1068-1073.
15. Hertzman PA, Blevins WL, Mayer J, Greenfield B, Ting M, Gleich GJ. Association of the eosinophilia-myalgia syndrome with the ingestion of tryptophan. *N Engl J Med.* 1990;322:869-873.
16. Eosinophilia-myalgia syndrome—New Mexico. *MMWR Morb Mortal Wkly Rep.* 1989;38:765-767.
17. Hertzman PA, Falk H, Kilbourne EM, Page S, Shulman LE. The eosinophilia-myalgia syndrome: the Los Alamos conference. *J Rheumatol.* 1991;18:867-873.
18. Hertzman PA, Graham K. The eosinophilia-myalgia syndrome: follow-up [abstract]. *Arthritis Rheum.* 1991;34(suppl):S131.
19. Hertzman PA. Criteria for the definition of the eosinophilia-myalgia syndrome. *J Rheumatol.* 1996;23(suppl 46):7-12.
20. Varga J, Heinman-Patterson TD, Emery DL, et al. Clinical spectrum of the systemic manifestations of the eosinophilia-myalgia syndrome. *Semin Arthritis Rheum.* 1990;19:313-328.
21. From the Centers for Disease Control. Update: eosinophilia-myalgia syndrome associated with ingestion of L-tryptophan—United States, as of January 9, 1990. *JAMA.* 1990;263:633.
22. Clauw DJ, Nashel DJ, Umhau A, Katz P. Tryptophan-associated eosinophilic connective-tissue disease: a new clinical entity? *JAMA.* 1990;263:1502-1506.
23. Duffy J. The lessons of eosinophilia-myalgia syndrome. *Hosp Pract (Off Ed).* 1992;27:65-69, 73-80, 83-90.
24. Special Writing Group of the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease of the Council on Cardiovascular Disease in the Young of the American Heart Association. Guidelines for the diagnosis of rheumatic fever: Jones criteria, 1992 update. *JAMA.* 1992;268:2069-2073.
25. American Joint Committee for Cancer Staging and End Results Reporting. *Manual for Staging of Cancer.* Chicago, Ill: American Joint Committee on Cancer; 1977.
26. Feinstein AR. *Multivariable Analysis: An Introduction.* New Haven, Conn: Yale University Press; 1996.
27. Swygert LA, Maes EF, Sewell LE, Miller L, Falk H, Kilbourne EM. Eosinophilia-myalgia syndrome: results of national surveillance. *JAMA.* 1990;264:1698-1703.
28. Belongia EA, Hedberg CW, Gleich GJ, et al. An investigation of the cause of the eosinophilia-myalgia syndrome associated with tryptophan use. *N Engl J Med.* 1990;323:357-363.
29. Henning KJ, Jean-Baptiste E, Singh T, Hill RH, Friedman SM. Eosinophilia-myalgia syndrome in patients ingesting a single source of L-tryptophan. *J Rheumatol.* 1993;20:273-278.
30. Kamb ML, Murphy JJ, Jones JL, et al. Eosinophilia-myalgia syndrome in L-tryptophan–exposed patients. *JAMA.* 1992;267:77-82.
31. Slutzger L, Hoesly FC, Miller L, et al. Eosinophilia-myalgia syndrome associated with exposure to tryptophan from a single manufacturer. *JAMA.* 1990;264:213-217.
32. Harvey M, Horwitz RI, Feinstein AR. Diagnostic bias and toxic shock syndrome. *Am J Med.* 1984;76:351-360.
33. Wagner KR, Elmore JG, Horwitz RI. Diagnostic bias in clinical decision making: an example of L-tryptophan and the diagnosis of eosinophilia-myalgia syndrome. *J Rheumatol.* 1996;23:2079-2085.
34. Duffy J, Engel AF, Bowles C. Eosinophilia-myalgia syndrome (EMS) not associated with L-tryptophan (LT) [abstract]. *Arthritis Rheum.* 1991;34(suppl):S132.
35. Clauw DJ, Mullins KW, Blank C. Eosinophilia-myalgia syndrome (EMS) not associated with ingestion of nutritional supplements [abstract]. *Arthritis Rheum.* 1993;36(suppl):R26.
36. American Heart Association. Jones criteria (revised) for guidance in the diagnosis of rheumatic fever. *Circulation.* 1965;32:664-668.
37. Feinstein AR. Clinical biostatistics: the purposes and functions of criteria. *Clin Pharmacol Ther.* 1978;24:479-492.
38. Feinstein AR. Clinical biostatistics: what are the criteria for criteria? *Clin Pharmacol Ther.* 1978;25:108-116.
39. Hertzman PA, Clauw DJ, Kaufman LD, et al. Eosinophilia myalgia syndrome: status of 205 patients and results of treatment two years after onset. *Ann Intern Med.* 1995;122:851-855.
40. Hertzman PA. The eosinophilia-myalgia syndrome: opportunities realized and missed. *J Rheumatol.* 1996;23:1679-1681.
41. Hertzman PA. Are the eosinophilia myalgia syndrome and the toxic oil syndrome clinically distinguishable [abstract]? *Arthritis Rheum.* 1993;36(suppl 9):S115.