

Diagnostic Versus Classification Criteria

A Continuum

Hasan Yazici, M.D.

Abstract

The current understanding that disease criteria are different than classification criteria is not well founded. In fact, they are a continuum. The arithmetic behind the two are the same and is built on a clear understanding of the concepts of sensitivity and specificity. Diagnosis is nothing different than classification in the individual patient. The main element that makes a set of criteria diagnostic is the pretest odds. We should question our current practice of making universal disease criteria and perhaps design criteria tailored to subspecialties.

As is true for some other disciplines, many of the diseases we have to deal with in rheumatology do not have specific clinical, laboratory, histological, or radiological features to tell them apart from other conditions with similar disease presentations. To this end, formal “disease criteria” have been formulated. Initially, all disease criteria were called “diagnostic.” A well known example has been the historic Jones criteria for rheumatic fever. Later, it was realized that such disease criteria often were not useful in making the correct diagnosis in an individual patient. Their place was in making classifications mainly for research purposes. This, I propose, led to the current understanding that *disease criteria* are, in essence, different from *classification criteria*.

I suggest this is a misconception. The two, in fact, represent a continuum. Every set of disease criteria is created as a classification and has the potential of becoming diagnostic if it has sufficient internal and, especially, external validity. The “mental arithmetic” behind both disease criteria and

classification criteria is the same. In other words, there is no other cerebral process in making a diagnosis other than that of classification. A diagnosis is, in fact, making a classification in an individual patient.

For many years, the usual method for formulating such criteria had been ad hoc, where experts, without any formal data collection and validation, told us how we should diagnose diseases. As mentioned above, a notable example has been Jones’ criteria for diagnosing rheumatic fever. It is noteworthy how, as late as 2002,¹ this ad hoc approach to criteria formulation prevailed among pediatricians about Jones’ criteria, originally proposed in 1944.² As a reminder, a combination of arthritis, fever, and a high sedimentation rate, in the presence of a recent Group A streptococcal infection, fulfills the Jones’ criteria. Needless to say, very many children with juvenile arthritis also present with the exact same features.

Over the years, the American College of Rheumatology has published many sets of classification criteria, including those for vasculitides.³ Although these criteria were said to be intended for classification only, physicians continued to use them for diagnosis. To emphasize this point, a formal sensitivity and specificity study was done,⁴ and it was shown that these criteria had quite poor (17% to 29%) positive predictive values when applied to 198 patients with various vasculitides and connective tissue diseases. Two years later, a further study⁵ showed that CHC (Chapel Hill Consensus Conference) criteria, another widely recognized vasculitis criteria set, correctly identified only 8 of 27 patients with Wegener’s granulomatosis and 4 of 12 patients with microscopic polyarteritis. The standard response to these observations remained the same; they were intended only for classification.⁶ The creation of diagnostic criteria required other considerations, among which was the frequency of the disease condition being sought. Of course, this is clearly true, as we will further discuss below. On the other hand, I can

Hasan Yazici, M.D., is from the Department of Medicine, Division of Rheumatology, University of Istanbul, Turkey.

Correspondence: Hasan Yazici, M.D., Sefa Sokak, Sen Apt 17/7, Kadikoy, Istanbul 81310, Turkey; hyazici@attglobal.net.

think of only one other consideration in making diagnostic criteria different than the mere disease frequency. This has to do with what the physician wants to do with his or her diagnosis.

At this point, it is important to be reminded about the all important definitions of sensitivity and specificity. Sensitivity is, in short, the percentage of *true positives*. If 80% of patents with rheumatoid arthritis (RA) have a positive test for rheumatoid factor (RF), the sensitivity of RF for RA is 80%. The specificity, on the other hand, is the percentage of *true negatives*. If 80% of patients without RA do not have RF, then the specificity of RF for RA is 80%. In dealing with these terms over the years, it has been my impression that physicians find the concept of specificity somewhat harder to grasp and remember. I think there are two reasons for this, and I suggest that the process of reviewing these reasons through points A and B below will help us to remember better what specificity is.

A. In defining sensitivity, which is usually easier to recall than the definition of specificity, we first make two consecutive positive assertions: 1. Among patients with RA----// and 2. RF is present in----//. These assertions are followed by a third, which is also a positive assertion: Then the sensitivity of RF for RA is----//. On the other hand, to define specificity, our first two assertions are negative: 1. Among patients who do not have RA----// and 2. The frequency of not having RF is----//. This time, our third assertion is positive: 3. The specificity of RF for RA is----//. I strongly suspect that this mental "change in sign" in assertions between the two concepts is one aspect of what makes specificity more difficult to grasp and remember.

B. The second reason for the definition of specificity being harder to remember than sensitivity lies in the completeness of their definitions as we usually state them. When we state that "RF has an 80% sensitivity for RA," we have given most of the information for its utility. On the other hand, when we say "RF has a 5% specificity for RA," this is incomplete information. Remembering that the definition of specificity is all about "not having RA," we always have to specify what makes up the group "not having RA" in our definition. So the correct way to state specificity is, "RF has a specificity of 75% in the diagnosis and classification of RA when tested among n_1 number of patients with systemic lupus, n_2 number of patients with sarcoidosis, n_3 number of patients with viral arthritis, etc., along with n_n healthy controls. Only then, will our definition of specificity have any true meaning. I regret to say this important aspect of specificity is still sometimes neglected, even in standard textbooks across medical disciplines.

The last point important point to be discussed about sensitivity and specificity is that there is always an inverse relationship between them. The graphic description of this relationship is the so-called ROC (receiver operating curve).⁷

Having reviewed the concepts of sensitivity and specificity, let us go back to the issue of what to do with the diagnosis.

In a malaria endemic area, the health authority can declare that a fever of more than a few days and an enlarged spleen is diagnostic of malaria and would require therapy with medication in the form of certain pills for a limited number of days. In most instances, the good that will come out of that health authority's definition of malaria will outweigh its harm. On the other hand, if the same authority announced that a combination fever and enlarged spleen indicated Hodgkin's disease and had to be treated accordingly, that would make a different story. In the first instance, when we give away relatively harmless pills to end an epidemic, we can be very sensitive but not specific, while in the second situation specificity is all important. Few of us realize that to form diagnoses in the interests of our patients this trade-off between sensitivity and specificity is, or ought to be, always present. We always want to, or rather ought to, rule-in a lung infection (much more treatable) than a carcinoma of the lung when confronted with a pathology on a chest radiograph.

The Importance of Disease Frequency in Formulating Diagnostic and Classification Criteria

It is apparent that the more frequently a condition occurs, the more likely it will be diagnosed. Bayes theorem is simply a mathematical way of expressing this phenomenon. Clinicians only seldom appreciate the importance of disease frequency (pretest probability in Bayesian terms) is in making a diagnosis.⁸ Bayes theorem as it applies to classification and diagnosis simply says that the odds of having a disease is equal to the pretest odds (PreO) "times" the likelihood ratio (LR).⁹ In this equation, the information for LR comes from the specificity and sensitivity of our diagnosis and classification criteria set. The source for PreO can be two-fold. It is either the actual disease prevalence or it is based on clinical prediction rules.⁹ To make such rules, we need to know the frequency of the clinical and laboratory findings of the diseases that come into the differential diagnosis. From these frequencies we calculate, by means of 2x2 tables, how each disease feature better defines (how it "weighs" in statistical jargon) the disease condition we seek among other conditions that come into the differential. James F. Fries, M.D., a dedicated student of quantitative medicine, aptly pointed out some years ago that to develop diagnostic criteria invariably involved circular logic.¹⁰ I mentioned above that the LR in the original Bayes formula is derived solely from our diagnostic and classification criteria. On the other hand, such criteria can only be formulated from clinical prediction rules if they are not ad hoc in the first instance. The clinical prediction rules, in turn, are arithmetically dependent on the frequency of the condition sought and the frequencies of the conditions that come into the differential diagnosis in the first instance. This is what makes the entire exercise logically circular.

How does one escape from this circular logic affecting the external validity of the classification and diagnostic criteria

we make, if at all? One approach is to make the patient and the control groups with which we prepare our criteria as representative as possible of the settings (the medical practices) in which these criteria will be used.¹⁰ Experience shows this has been mostly unsuccessful. For example the current International Behçet's syndrome criteria set (with ~90% sensitivity and ~95% specificity) is rather unique among the vasculitis criteria in that it was originally prepared with due consideration as to what is discussed above.¹¹ As such, it is rather useable as a diagnostic tool in a rheumatology practice in a Behçet's disease-endemic area like Turkey (1/250 of the adult population).¹² However, the same set of criteria merits even more the designation of classification when used in settings where Behçet's is rare. In other words, it can almost never be diagnostic in a setting where the PreO is small.

I suggest it is time to reconsider our continuous effort to make "universal" classification and diagnoses for identifying rare conditions like Wegener's or Behçet's syndrome. I envisage that subspecialty tailored criteria will have the potential to be much more useful,¹³ in that with such criteria, first, the PreO will remarkably increase, and, second, the number of conditions that come into the differential diagnosis will decrease, increasing the specificity of the criteria set. To give an example to back-up my first point, we do know that the frequency of Behçet's disease is, at minimum, 1000-fold greater in Japan, as compared to North America. On the other hand, if you go to a dedicated uveitis clinic in either country, you will see that the proportion of Behçet patients that present differs only by several fold, 2.5% in North America and 6.2% in Japan.^{14,15} Furthermore, the clinical prediction rules for a gastroenterologist to distinguish Behçet's disease from inflammatory bowel disease one of the very few conditions that Behçet's has to be separated from in a gastroenterology practice are different from those of a rheumatologist trying to do the same among a more exhaustive list of connective tissue diseases and vasculitides. This is also why I say criteria should be tailored to the practice setting.

It is to be remembered that this approach would not in any way exclude classification criteria for research purposes or presumptive diagnostic criteria for practice purposes. In addition, we should always ask ourselves why we want to diagnose, which, in turn, depends on what we want to do with that diagnosis.

Disclosure Statement

This work has been partially supported by the Turkish Academy of Sciences. The author has no further financial or pro-

prietary interest in the subject matter or materials discussed, including, but not limited to, employment, consultancies, stock ownership, honoraria, and paid expert testimony.

References

1. Ferrieri P. Jones Criteria Working Group. Proceedings of the Jones Criteria workshop. *Circulation*. 2002 Nov 5;106:2521-3.
2. Jones TD. The diagnosis of rheumatic fever. *JAMA*. 1944;126:481-4.
3. Hunder GG, Arend WP, Bloch DA, et al. The American College of Rheumatology 1990 criteria for the classification of vasculitis: introduction. *Arthritis Rheum*. 1990;33:1065-7.
4. Rao JK, Allen NB, Pincus T. Limitations of the 1990 American College of Cardiology classification criteria in the diagnosis of vasculitis. *Ann Intern Med*. 1998;129:345-52.
5. Sorensen SF, Slot O, Tvede N, Petersen J.A prospective study of vasculitis patients collected in a five year period: evaluation of the Chapel Hill nomenclature. *Ann Rheum Dis*. 2000;59:478-82.
6. Hunder GG. The use and misuse of classification and diagnostic criteria for complex diseases. *Ann Intern Med*. 1998;129:417-8.
7. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. *Ann Intern Med*. 1981;94:555-600.
8. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med*. 1998;104:374-80.
9. Max MB, Lynn J (eds). *Symptom Research: Methods and Opportunities*. Bethesda, MD: National Institutes of Health, Department of Health and Human Services. Available at <http://symptomresearch.nih.gov/tablecontents.htm>. Accessed April 4, 2009.
10. Fries JF. Disease criteria for systemic lupus erythematosus. *Arch Intern Med*. 1984;144:252-3.
11. International Study Group for Behçet's Disease. Criteria for diagnosis of Behçet's disease. *Lancet*. 1990 May 5;335(8697):1078-80.
12. Yurdakul S, Günaydin I, Tüzün Y, et al. The prevalence of Behçet's syndrome in a rural area in northern Turkey. *J Rheumatol*. 1988;15:820-2.
13. Yazici H, Seyahi E, Yurdakul S. Behçet's syndrome is not so rare: why do we need to know? *Arthritis Rheum*. 2008;58:3640-3.
14. Rodriguez A, Calonge M, Pedroza-Seres M, et al. Referral patterns of uveitis in a tertiary eye care center. *Arch Ophthalmol*. 1996;114:593-9.
15. Goto H, Mochizuki M, Yamaki K, et al. Epidemiological survey of intraocular inflammation in Japan. *Jpn J Ophthalmol*. 2007;51:41-4.