

Clinical Decision Making Part II: Why Diagnostic Procedures vary in Performance

D. Nanan, F. White

Department of Community Health Sciences, The Aga Khan University, Karachi.

Abstract

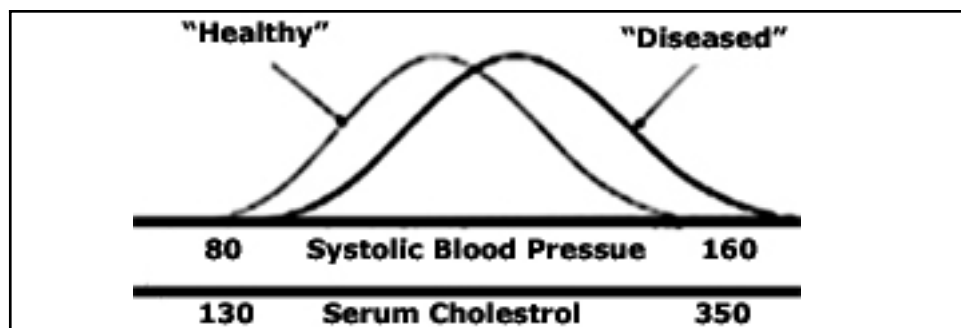
As outlined in Part I of this two-part series, the consequences arising from Type 1 and Type 2 errors in clinical decision making may be serious. Thus, it is helpful to consider what is the predictive value of a positive result, and also a negative result, regardless of whether the method of diagnosis is clinically based (signs and symptoms) or laboratory based. The positive predictive value (PPV) and negative predictive value (NPV) of a diagnostic method can vary depending on the background prevalence of the condition. This overview of probability, errors and consequences in clinical decision making also briefly introduces Bayes' Theorem, which is the mathematical basis for evaluating the "conditional probability" of an event, e.g what is the probability that the individual has the disease given that they have tested positive (PPV)? Similarly, what is the probability that the individual does not have the disease given that they have tested negative (NPV)? The theorem is a useful tool for physicians and researchers wishing to study the process and performance of medical diagnoses and related decisions (JPMA 53:194;2003).

Appying Type 1 and Type 2 Errors to Populations

Following on from our discussion in Part I¹, let us now visit the vexing problem of Type 1 errors when the indicator used is endemic in the population. In recent years, epidemiologists have demonstrated the widespread presence in apparently "well populations" of findings associated with increased risk of disease, commonly recognized examples being elevated blood pressure and cholesterol levels. When individuals present with elevations in these measures, clinical importance is attached to them. But what are the implications on a population based level? A classic analysis conducted by the late Geoffrey Rose, on the topic of "sick individuals and sick populations", revealed theoretically that there is more to be gained by shifting the entire population curve to the left than to attend to the small proportion of individuals at ostensibly higher risk above a pre-determined cutpoint on the curve (Figure).² This approach is supported by randomized trials that demonstrate a reduction in individual risk by lowering either blood pressure or cholesterol at virtually any

point on the curve. In this example (in the clinical decision making context), lack of action based on an individual's value being below a clinical cutpoint is similar to a Type 1 error (where the null hypothesis is taken as disease is present): the individual is considered "well" and remains untreated when in fact some measure of intervention could have a beneficial impact. When taken across an adult population, the error magnifies and indeed accounts for the majority of related cardiovascular disease.

From the above, it seems logical therefore to direct primary intervention efforts towards entire populations for particular risk factors than to solely target medical interventions to those deemed to be at "high risk". It may even be argued that there is more to be gained by doing away with individual testing and simply treating entire populations presumptively, especially in developing countries. In fact, the recommended daily use of low dose aspirin from middle age onwards is an example of such a mass action, even in the absence of individual risk assessment. In advocating this, we are committing a Type 2 error: there will be those taking such prophylaxis who are truly at no elevated risk of heart disease,



Adapted from: Rose G. Sick individuals and sick populations. International Journal of Epidemiology 11985,14::32-38

Figure. Simulated distributions for "Healthy" and "Diseased" populations: BP and Cholesterol.

Nonetheless, like fluoridation, this illustrates in the context of chronic diseases a rationale for public health action: to apply population approaches more emphatically, and not to focus so exclusively on medical diagnosis and treatment, both of which (by definition) come too late.

In case such a proposal seems too radical, or somehow not applicable to Asia, a recent editorial in the *New England Journal of Medicine* notes that: "recent estimates from populations in east Asia suggest that a reduction of just 3 percent in average blood pressure levels in such populations (as might be achieved, for example, by sustained reductions in dietary sodium or caloric intake) would be expected to reduce the incidence of disease (largely among non-hypertensive persons) almost as much as would hypertensive therapy targeted to all hypertensive persons in the population".³

Clinical and Laboratory Diagnosis and the Use of Cutpoints

In all laboratory tests, false negatives and false positives may occur, both of which have implications, and must be considered when interpreting results. For example, using a polymerase chain reaction (PCR) method for diagnosing malaria can detect up to 12% more positive samples than diagnosis by microscopy alone.⁴ Similarly, in an evaluation of three different diagnostic methods in a low endemic area, malaria prevalence was 5.9% by thick blood smear, 7.4% by a quantitative buffy coat method and 21.9% by PCR.⁵ In that study, differences between PCR results and those from other methods were attributed to low level parasitemias, which clinically might be associated with transient, subclinical, mild or early disease. These examples reveal the relationship of the technology used to the likelihood of diagnosis, as well as error rates, and the increasing potential for false negatives as one crosses the spectrum from symptomatic to mild and subclinical disease.

The same applies to clinical algorithms. In a clinic based study of 451 children with signs and symptoms consistent with malaria in rural Sindh, malaria slide positivity rate was only 5.9%.⁶ If slide parasitemia is to be taken as the "gold standard", up to 94.1% of these clinical diagnoses may have been invalid (false positives or Type 2 errors in relation to the clinical null hypothesis). In the study area, persons suspected of having malaria were usually treated on a symptomatic basis, exposing the individual to potential drug toxicity needlessly. While the use of more sensitive but expensive methods such as PCR for malaria diagnosis is not feasible at a local level, clinical algorithms tailored to a specific context and used in conjunction with microscopy can reduce Type 2 errors. Furthermore, in addition to the hazards of misdiagnosis and inappropriate treatment, the widespread practice of diagnosing malaria on signs and symptoms alone, then

reporting "cases" in rudimentary surveillance systems, may underlie a major error in over-estimating disease burdens. The impact on priority setting, and on the cost of wasted medications, should not be ignored.

To illustrate how the choice of test can affect Type 1 and Type 2 error rates obtained, we can extend the malaria example. If we apply the prevalence estimate of approximately 6% from the Sindh study to a theoretical population of 10,000, this implies that 600 persons would have the disease and 9,400 would be disease free (Table 1). Using a malaria diagnostic test with sensitivity 90% and specificity 90%, 540 of the 600 persons with the disease would be correctly identified as disease positive (0.90×600); however, 60 persons would be falsely labeled as disease negative ($600 - 540 = 60$). Similarly, 8,460 persons would be identified correctly as disease negative, but 940 would be falsely labeled as having the disease. One can also calculate across the columns to derive the positive and negative predictive values (PPV and NPV, respectively). The probability that a person has the disease given that the test is positive, or the test's PPV, is 36.5%; the probability that a person does not have the disease given that the test is negative, or the NPV, is 99.3%.

Table 1. Predictive values for a test with sensitivity 90%, specificity 90% and disease prevalence 6%.

	Disease Present	Disease Absent	
Test Positive	540	940	PPV=540/1480 =36.5%
Test Negative	60	8460	NPV=8460/8520 = 99.3%

For a low prevalence setting, the results for the NPV are not an issue. However, a PPV of 36.5% is low by normally desirable diagnostic standards, implying that only one in three persons who test positive would actually have the disease. If sensitivity and specificity were both increased to 95% (not shown), the PPV will move to slightly better than 50%, or a one in two chance that the person who tests positive will have the disease. To go further, if sensitivity and specificity were both increased to 99% (not shown), PPV would be 86.3%, implying that 17 out of every 20 persons who test positive do have the disease, definitely more satisfactory from the standpoint of community application. While this illustration is based on arbitrary levels of sensitivity and specificity, the results of recently published validation studies for rapid diagnostic tests for malaria indicate that these levels are achievable.⁷ We explore the effect of prevalence itself on PPV and NPV later in this paper.

Table 2. Predictive values for a test with sensitivity 90%, specificity 90%, and disease prevalence 30%.

	Disease present	Disease absent	
Test Positive	2700	700	PPV = 2700/3400 = 79.4%
Test Negative	300	6300	NPV = 6300/6600 = 95.5%

Bayes' Theorem

As in the above example, constructing a 2x2 table offers a direct way to compute the PPV and NPV of a test. Another way gives identical results, and is based on a theorem named after Thomas Bayes, an 18th century English mathematician. Bayes' Theorem is a method for estimating the "conditional probability" of an outcome given that certain events have occurred, that is the "pre-test likelihood" (referred to as the "prior probability" in the example of hospital statistics, Part I). In Bayesian inference, the investigator takes into account prior probability in arriving at a new probability, called "posterior probability" or "post-test likelihood".^{8,9}

In statistical language, Bayes' Theorem is as follows:

If the probability of B occurring, or P(B), is not = 0, then

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A-) \cdot P(A-)}$$

In the example of malaria, all relevant characteristics related to the problem of determining PPV can be restated by substituting T (test) for A, and D (disease) for B. Although this may look very different, all the elements are captured here:

Test sensitivity = P(T+ | D+) and test specificity = P(T- | D-)

Positive Predictive Value = P(D+ | T+) and Negative Predictive Value = P(D- | T-)

Disease prevalence = P(D+) and Disease absence = P(D-).

Error rates can also be expressed in this manner:

False positive rate = P(T+ | D-), which is the same as 1-Specificity or 1- P(T- | D-)

False negative rate = P(T- | D+), which is the same as 1-Sensitivity or 1- P(T+ | D+).

Applying Bayes' Theorem, we can therefore derive PPV from the following equation:

$$P(D+ | T+) = \frac{P(T+ | D+) \cdot P(D+)}{P(T+ | D+) \cdot P(D+) + P(T+ | D-) \cdot P(D-)}$$

In more literary terms, the formula could be restated:
PPV = $\frac{\text{Sensitivity} \times \text{Disease Prevalence}}$

$$\frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + \{(1-\text{Specificity}) \times 1-\text{Prevalence}\}}$$

Similarly, Bayes' Theorem may be used to determine the NPV. Thus:

$$\text{NPV} = \frac{\text{Specificity} \times (1-\text{Prevalence})}{\text{Specificity} \times (1-\text{Prevalence}) + \{(1-\text{Sensitivity}) \times \text{Prevalence}\}}$$

$$\frac{\text{Specificity} \times (1-\text{Prevalence})}{\text{Specificity} \times (1-\text{Prevalence}) + \{(1-\text{Sensitivity}) \times \text{Prevalence}\}}$$

In summary therefore, sensitivity, P(T+ | D+) and specificity, P(T- | D-) and their reciprocals (false negative and false positive rates), measure the inherent accuracy of a test, independent of the probability of disease. However, the predictive values of a test {ie, P(D+ | T+) and P(D- | T-)}, depend not only on both these measures of test validity, but also the prior probability of disease, in other words, disease prevalence. For any given prevalence therefore, either Bayes' Theorem or the 2x2 table approach provides a method by which the performance of a test in terms of its predictive values can be computed, taking into account test sensitivity and specificity as well as disease prevalence.

Using the earlier example of malaria testing, if the prevalence of malaria were 30% rather than 6%, then a test with only 90% sensitivity and specificity would deliver a more acceptable PPV (79.4%). Table 2 uses the 2x2 table method to demonstrate, and we invite the reader to apply Bayes' Theorem to obtain the same results.

Thus, a test or clinical algorithm with less than perfect sensitivity and specificity may perform well in a referral setting, where the prior probability of disease is high by virtue of a referral process (i.e., high prevalence). However, the same method of diagnosis may be of marginal utility in a community clinic where the condition being investigated is of low prevalence.

Bayes' Theorem also has other useful applications such as: to articulate the basis for prognostic stratification, to estimate potential for therapeutic or drug toxicity changes, to model cancer risks, to predict clinical outcomes including cost-effectiveness, to assist in forensic investigations, and to design decision pathways. Run a Medline search, and you will discover many more examples of this aspect of clinical epidemiology.

Acknowledgements

This review and formulation of the issues introduced in Part I and the accompanying exercise on the effect of varying disease prevalence on predictive value of positive test results detailed in Part II, were developed for the Post-Graduate Medical Education Conference, Complexity Science and Health Care, May 31 and June 1, 2002, The Aga Khan University, Karachi.

References

1. White F, Nanan D. Probability, errors and consequences in clinical decision making: Part I. *J Pak Med Assoc* 2003;53:157-9.
2. Rose G. Sick individuals and sick populations. *Int. J Epidemiol* 1985;14:32-8.
3. MacMahon S. Blood Pressure and the risk of cardiovascular disease. *N Engl J Med* 2000;342:50-2.
4. Rubia JM, Benito A, Berzosa PJ, et al. Usefulness of seminested multiplex PCR in surveillance of imported malaria in Spain. *J Clin Microbiol* 1999;37: 3260-4.
5. Carrasquilla G, Banguero M, Sanchez P, et al. Epidemiological tools for malaria surveillance in an urban setting of low endemicity along the Colombian Pacific coast. *Am J Trop Med Hyg* 2000;62:132-7.
6. Hozhabri S, Akhtar S, Rahbar MH, et al. Prevalence of plasmodium slide positivity among children treated for malaria, Jangara, Sindh. *J Pak Med Assoc* 2000;50:401-5.
7. Lee MA, Aw LT, Singh M. A comparison of antigen dipstick assays with polymerase chain reaction (PCR) technique and blood film examination in the diagnosis of malaria. *Ann Acad Med Singapore* 1999;28:498-501.
8. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. *Biostatistics in Clinical Medicine*. Toronto. Collier MacMillan Canada, Inc. 1983.
9. Hirsch RP, Riegelman RK. *Statistical operations: analysis of health research data*. Oxford: Blackwell, 1996.

