J Clin Epidemiol Vol. 41, No. 5, pp. 491–494, 1988
Printed in Great Britain.

0895-4356/88 $3.00 + 0.00

II-8

EVALUATION OF DIAGNOSTIC TESTS WITH MULTIPLE DIAGNOSTIC CATEGORIES

N. J. BIRKETT

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

(*Received in revised form* 18 *August* 1987)

Abstract—The evaluation of diagnostic tests attempts to obtain one or more statistical parameters which can indicate the intrinsic diagnostic utility of a test. Sensitivity, specificity and predictive value are not appropriate for this use. The likelihood ratio has been proposed as a useful measure when using a test to diagnose one of two disease states (e.g. disease present or absent). In this paper, we generalize the likelihood ratio concept to a situation in which the goal is to diagnose one of several non-overlapping disease states. A formula is derived to determine the post-test probability of a specific disease state. The post-test odds are shown to be related to the pre-test odds of a disease and to the usual likelihood ratios derived from considering the diagnosis between the target diagnosis and each alternate in turn. Hence, likelihood ratios derived from comparing pairs of diseases can be used to determine test utility in a multiple disease diagnostic situation.

Diagnostic tests   Likelihood ratios   Bayes' Theorem   Pre-test odds   Post-test odds

INTRODUCTION

A number of statistical parameters have been proposed as guides to evaluate the utility of a diagnostic test (e.g. sensitivity, specificity and predictive value) [1, pp. 41–57]. These parameters were initially developed for a test with two diagnostic levels and a dichotomous outcome (e.g. diseased vs normal) and have been generalized for situations with multiple diseases and multiple diagnostic levels [2]. However, all of these parameters have limitations when clinicians attempt to apply them to patient care. A clinician is primarily interested in the impact of a diagnostic test on the probability that a patient has a particular disease (i.e. the relationship between pre- and post-test disease probability). While this can be calculated from sensitivity and specificity, it requires the use of formulae with which most clinicians are not at ease. Although predictive values provide the direct post-test probability, such predictive values vary with disease prevalence [3] and thus the clinician cannot use a single value to quantify the discriminative ability of a diagnostic test.

An alternate approach to quantifying diagnostic test utility uses the likelihood ratio (odds that a given test result would be expected in a patient with, as opposed to without, the target disease) [4, pp. 108–126] to yield a single number to characterize the test. The likelihood ratio is invariant to changes in pre-test prevalence unless the sensitivity and/or specificity of the test vary as the prevalence changes. The post-test odds of disease can be obtained by multiplying the pre-test odds of disease by the likelihood ratio. Previous work has developed likelihood ratios for the two disease models with either dichotomous or multi-leveled diagnostic tests [4, pp. 108–126]. In this paper we will review the derivation of the standard likelihood ratio and provide an extension to the multiple disease situation.

Reprint requests and correspondence should be addressed to: Dr N. Birkett, The Irene Eleanor Smythe Pain Clinic, Room 3-635, The Bell Wing, Toronto General Hospital, 200 Elizabeth Street, Toronto, Ontario, Canada M5G 2C4.

TWO-DISEASE LIKELIHOOD RATIO

Consider a diagnostic test with multiple levels ($E_1, E_2, \ldots, E_n$) which is being used to diagnose the presence or absence of a disease ($D_1, \bar{D}_1$) in a population with a pre-test probability of disease equal to π_1 and of non-disease being π_2. We can present this situation in a $2 \times n$ table (Table 1).

Direct application of Bayes' theorem can yield the post-test probability of disease D_1 conditional on the test giving result E_1:

$$P(D_1 | E_1) = \frac{P(E_1 | D_1)P(D_1)}{P(E_1 | D_1)P(D_1) + P(E_1 | \bar{D}_1)P(\bar{D}_1)}$$

$$= \frac{P_1\pi_1}{P_1\pi_1 + P_2\pi_2}$$

where

$$P_1 = P(E_1 | D_1), \quad P_2 = P(E_1 | \bar{D}_1),$$

$$\pi_1 = p(D_1) \quad \text{and} \quad \pi_2 = p(\bar{D}_1).$$

Similarly, the post-test probability that the patient does not have D_1 is given by:

$$P(\bar{D}_1 | E_1) = \frac{P_2\pi_2}{P_1\pi_1 + P_2\pi_2}.$$

Therefore, the post-test odds of D_1 compared to $\bar{D}_1$ (or D_2) is:

$$\frac{P(D_1 | E_1)}{P(\bar{D}_1 | E_1)} = \left(\frac{P_1}{P_2}\right) \times \left(\frac{\pi_1}{\pi_2}\right).$$

But π_1/π_2 is the pre-test odds of D_1 compared to $\bar{D}_1$ and P_1/P_2 is independent of the pre-test rates. Therefore, knowledge of P_1/P_2 for a diagnostic test permits a simple conversion from pre-test odds to post-test odds. The quantity P_1/P_2 is called the likelihood ratio for the test.

THREE-DISEASE SITUATION

Suppose we now wish to diagnose the patient as having one of three diseases. Analogous to the two-disease situation, we can generate a $3 \times n$ table containing pre-test disease probabilities (Table 2). For this model we will assume that the three diseases are mutually exclusive (i.e. a patient has only one disease).

If we proceed as in the two-disease situation we have:

$$P(D_1 | E_1) = \frac{P(E_1 | D_1)P(D_1)}{P(E_1 | D_1)P(D_1) + P(E_1 | D_2) \times P(D_2) + P(E_1 | D_3)P(D_3)}$$

$$= \frac{P_1\pi_1}{P_1\pi_1 + P_2\pi_2 + P_3\pi_3}$$

and

$$P(\bar{D}_1 | E_1) = 1 - P(D_1 | E_1)$$

$$= \frac{P_2\pi_2 + P_3\pi_3}{P_1\pi_1 + P_2\pi_2 + P_3\pi_3}.$$

Therefore, the post-test odds are:

$$\frac{P(D_1 | E_1)}{P(\bar{D}_1 | E_1)} = \frac{P_1\pi_1}{P_2\pi_2 + P_3\pi_3}.$$

Finally, we find that the ratio of post-test odds to pre-test odds is given by:

$$\left[\frac{P(D_1 | E_1)}{P(\bar{D}_1 | E_1)}\right] \Big/ \left[\frac{\pi_1}{\pi_2 + \pi_3}\right] = \frac{P_1(\pi_2 + \pi_3)}{P_2\pi_2 + P_3\pi_3}$$

$$= P_1\left[\frac{1 + \pi_2/\pi_3}{P_3 + P_2(\pi_2/\pi_3)}\right].$$

This ratio is not independent of the pre-test probabilities. Rather, it depends directly on π_2/π_3, the pre-test odds of disease two compared to disease three conditional on disease one being ruled out. Since this ratio will vary depending on the composition of the target population, the ratio of pre- and post-test odds is not a productive way to produce a likelihood ratio when there are multiple potential disease states.

An alternate approach to determining post-test odds is obtained by rearranging terms in the equation for the post-test odds:

$$\frac{P(D_1 | E_1)}{P(\bar{D}_1 | E_1)} = \frac{P_1\pi_1}{P_2\pi_2 + P_3\pi_3}$$

$$= \frac{1}{\dfrac{1}{\dfrac{P_1}{P_2}\dfrac{\pi_1}{\pi_2}} + \dfrac{1}{\dfrac{P_1}{P_3}\dfrac{\pi_1}{\pi_3}}}.$$

Expression (1) is one half of the harmonic mean of the quantities $(P_1/P_2)(\pi_1/\pi_2)$ and

Table 1. Two-disease model

	D_1	$D_2 = \bar{D}_1$	
E_1	P_1	P_2	
	π_1	π_2	

Table 2. Three-disease model

	D_1	D_2	D_3	
E_1	P_1	P_2	P_3	
	π_1	π_2	π_3	

$(P_1/P_3)(\pi_1/\pi_3)$. Now, if we consider the conditional probabilities of diseases one and two if disease three is assumed to be absent, then the prior probability of D_1 is $[\pi_1/(\pi_1+\pi_2)]$ and the prior probability of D_2 is $[\pi_2/(\pi_1+\pi_2)]$. Using the results from the previous section, we have the likelihood ratio of diagnosing D_1 vs D_2 conditional on the absence of D_3 is $LR_{12} = (P_1/P_2)$. Similarly, the likelihood ratio of D_1 vs D_3 conditional on the absence of D_2 is $LR_{13} = (P_1/P_3)$. Therefore, the terms in the denominator of formula (1) represent the post-test odds of disease D_1 conditional on the absence of disease D_3 and disease D_2 respectively. We can represent the overall post-test odds of D_1 as:

$$\frac{P(D_1|E_1)}{P(\bar{D}_1|E_1)} = \frac{1}{\dfrac{1}{LR_{12}\dfrac{\pi_1}{\pi_2}} + \dfrac{1}{LR_{13}\dfrac{\pi_1}{\pi_3}}}. \quad (1)$$

Formula (2) provides a simple method of converting pre-test odds into post-test odds. Analogously to the two-disease situation, it requires knowledge of the pre-test odds of the target disease compared to each of the alternate diseases. The diagnostic ability of the test can then be summarized in two likelihood ratios, both based on a two-disease comparison.

Formula (2) can also be manipulated to give a direct method of calculating the post-test probability of disease:

$$P(D_1|E_1) = \frac{1}{1 + \dfrac{1}{LR_{12}\dfrac{\pi_1}{\pi_2}} + \dfrac{1}{LR_{13}\dfrac{\pi_1}{\pi_3}}}. \quad (2)$$

MULTIPLE DISEASE SITUATION

The preceding analysis can be generalized to the situation with n mutually exclusive disease states, $D_1, D_2, \ldots, D_n$. In this case we find that the post-test odds is given by:

$$\frac{P(D_1|E_1)}{P(\bar{D}_1|E_1)} = \frac{1}{\displaystyle\sum_{i=2}^{n} \frac{1}{LR_{1i}\dfrac{\pi_1}{\pi_i}}} \quad (3)$$

where LR_{1i} is the likelihood ratio for the test when diagnosing D_1 vs D_i conditional on the absence of all other diseases. An alternate way of describing LR_{1i} is as the ratio of the sensitivity of the test for disease 1 to the sensitivity for disease i. The ratio (π_1/π_i) is the pre-test odds of D_1 vs D_i conditional on the absence of all other diseases. As before, we note that formula (3) is closely related to the harmonic mean of the quantities $LR_{1i}(\pi_1/\pi_i)$. Further, we can calculate the post-test probability of disease by:

$$P(D_1|E_1) = \frac{1}{1 + \displaystyle\sum_{i=2}^{n} \frac{1}{LR_{1i}\dfrac{\pi_1}{\pi_i}}}.$$

EXAMPLE

Consider an emergency room physician attending a patient presenting with acute abdominal pain. He is reviewing a new diagnostic test which will be used to classify patients into one of three diagnostic categories: non-specific abdominal pain (NS), appendicitis (A) or cholecystitis (C). The paper describing the test reports that a positive test result gives a likelihood ratio for diagnosing NS vs A of 0.3; a ratio for diagnosing NS vs C of 0.5 and a ratio for diagnosing A vs C of 3.0. A study of patients in the emergency room [5] revealed that the prevalence (pre-test probability) of the three diseases were: NS (0.57), A (0.33) and C (0.10). How would a positive test result change the probability of disease? We can calculate the post-test probability of the three diseases from formula (2).

First, we convert the pre-test disease probabilities to conditional pre-test odds. For NS vs A the pre-test odds are: 0.57/0.33 = 1.73. Similarly, the pre-test odds for NS vs C is 5.7 and for A vs C is 3.3. Now, use formula (2) to calculate the post-test odds:

$$\frac{P(\text{NS}|\text{test + ve})}{P(\text{no NS}|\text{test + ve})} = \frac{1}{\dfrac{1}{0.3 \times 1.73} + \dfrac{1}{0.5 \times 5.7}} = 0.44$$

Thus the post-test probability of NS is $0.44/(1+0.44) = 0.31$. Similarly, the post-test probability of A is 0.62 and for C it is 0.07. Therefore a positive result on the new test is helpful in identifying this patient as a surgical candidate with appendicitis being the most likely diagnosis.

DISCUSSION

This paper has presented formulae for the post-test odds of a target disease in a situation in which there are multiple diseases under diagnostic consideration. The main feature of formulae (2) and (3) is that they depend only on data obtained from two disease comparisons. Hence likelihood ratios used in evaluating dichotomous diagnostic decisions can be used directly in the multiple disease situation, provided that the alternative disease studied in the dichotomous situation is clearly defined and is not simply the absence of the target disease. Thus evaluation of diagnostic tests should clearly define the disease alternatives being considered. This recommendation is similar to one proposed by Ransohoff and Feinstein [6].

The formulae presented here depend on the assumption that the diseases are mutually exclusive (i.e. a patient cannot have two or more diseases). This assumption is required to apply Bayes' theorem to obtain the post-test probability of disease. Unfortunately, this assumption is somewhat restrictive since clinical decisions often involve overlapping diagnostic categories (e.g. myocardial infarction vs gastric reflux). A model of diagnostic test utility in the presence of non-exclusive diagnostic categories is more complex since it depends on the correlation between the multiple, overlapping disease states. One approach would be to construct new diagnostic categories for all possible states (e.g. no disease, myocardial infarction only, gastric reflux only, both myocardial infarction and gastric reflux). With this approach, Bayes' theorem could be applied. However, the resulting likelihood ratios would be of questionable utility in evaluating the diagnostic test, since the addition of a new possible disease outcome would mean that the outcome disease categories would all be redefined and hence that the likelihood ratios would need to be recalculated. Further work is needed to develop methods of characterizing diagnostic tests to be used with multiple, non-exclusive disease states.

Formulae (2) and (3) require estimates of the pretest odds of D_1 relative to each other disease state, conditional on the absence of the other diseases. However, it does not require an estimate of the conditional pre-test disease probability. Expanding the list of potential diagnoses must decrease the pre-test disease probability for each of the original diseases. However, in most cases it would not be expected to significantly alter the pre-test odds of the disease compared to the target disease. Therefore we can obtain usable estimates of pre-test odds from situations which did not consider all diagnostic categories of relevance.

Examination of formulae (2) and (3) readily shows that the post-test odds of disease D_1 compared to all other diseases must be smaller than each of the individual conditional post-test odds. That is:

$$\frac{P(D_1|E_1)}{P(\bar{D}_1|E_1)} \leqslant \frac{\pi_1}{\pi_i}\frac{P_1}{P_i} \quad i = 2, \ldots, n.$$

Thus the diagnostic ability of a test in the multiple disease situation with common diseases is largely determined by the pairwise diagnostic situation with the lowest likelihood ratio. Therefore inclusion in the list of diagnostic states of any disease for which the pairwise diagnostic ability of the test is low will produce a test which has poor overall diagnostic ability. Similarly, inclusion of rare diseases in the list of diagnostic states will compromise the ability of the test to achieve high post-test odds for any of the diagnoses.

Acknowledgements—This work was conducted while Dr Birkett was a National Research Scholar of Health and Welfare Canada. The author thanks Dr D. L. Sackett and Professor R. S. Roberts for bringing this problem to his attention and for helpful comments and Mrs M. Taylor for typing parts of the manuscript.

REFERENCES

1. Fletcher RH, Fletcher SW, Wagner EH. **Clinical Epidemiology—The Essentials.** Baltimore: Williams and Wilkins; 1982.
2. Goldsmith CH. Generalized measures of sensitivity, specificity and predictive value in epidemiologic studies. **Am J Epidemiol** 1977; 106: 236 (abstract).
3. Baldessarini RJ, Finkelstein S, Arana GW. The predictive power of diagnostic tests and the effect of prevalence of illness. **Arch Gen Psychiat** 1983; 40: 569–573.
4. Sackett DL, Haynes RB, Tugwell P. **Clinical Epidemiology: A Basic Science for Clinical Medicine.** Boston: Little, Brown and Co.; 1985.
5. de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. **Br Med J** 1972; 2: 9–13.
6. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. **N Engl J Med** 1978; 299: 926–930.