

THE PROBLEM OF VALIDITY OF A DIAGNOSTIC TEST FOR MASSIVE USE AS STATISTICAL SCREENING PROCEDURES

Vicente M. Astudillo¹ & Isabel N. Kantor²

ABSTRACT

The effectiveness of an indirect diagnostic test of massive use in discriminating procedures is assessed by means of conditional probabilities and application of the Bayes theorem. Examples are given of estimates of diagnostic errors and of the true prevalence rate.

For studies on populations where the presence of positive animals can be regarded as a rare event, the relationship existing between the size of the sampling and the probability of failure to detect positives is discussed.

INTRODUCTION

Whenever an indirect test is utilized to screen the animals of a population into positives and negatives with respect to a disease, the validity of the test must be known. That is, does the test "measure" what it proposes to "measure"? For example: does the bovine tuberculin test actually evaluate the presence or absence of tuberculosis? This question cannot be answered by a simple 'yes' or 'no'. Rather, the answer lies on a proportional or percentage scale that indicates the test's degree of validity.

Two parameters are utilized to study specifically the validity of an indirect diagnostic classification test applied to populations: sensitivity and specificity. These tests almost never produce an exact identification of all the true positives and negatives. For example: the tuberculin test produces positives as a result of exposure to *Mycobacterium*. Whereas tuberculosis has been

taken as an example, it should be understood that immunologically, the positive reaction of bovines to the bovine PPD tuberculin is the manifestation of the animals previous exposure to *Mycobacterium bovis* and thus means infection, not necessarily disease. Although due to this germ's pathogenicity and the cattles' susceptibility to it, the infection is soon followed by lesions indicative of the disease. On other occasions, positive tuberculin reactions may occur through infections due to micobacteria other than *M. bovis* because of the existence of common antigens that are not necessarily followed by disease in the animal. In practical situations, these para-specific reactions take on major relative importance in the stages of very low prevalence of tuberculosis in cattle (false positives). On the other hand, the opposite diagnosis error -false negatives- may occur in bovines in the first phases of infection (preallergic period), during other intercurrent infections (virosis) or in animals in the cachectic state.

The false results, called classification errors, indicate the indirect test's fallibility level. They also produce difficulties in both the estimation of tuberculosis prevalence and in the decision-making relative to the application of sanitary measures for the eradication of tuberculosis.

EVALUATION OF THE VALIDITY OF THE INDIRECT TEST

The performance of indirect tests (tuberculinization, in this example) can be assessed through sensitivity and specificity. In practice, this can be done by means of a study of the data relative to the animals in a given region or country (R) having a bovine tuberculosis control program. The following information is required for the assessment: i) tuberculosis test results for a given

¹Pan American Foot-and-Mouth Disease Center. (PAHO/WHO), Caixa Postal 589, 20000 Rio de Janeiro-RJ, Brasil.

²Pan American Zoonoses Center. (PAHO/WHO), Casilla 3092 - Correo Central, 1000 Buenos Aires, Argentina.

period (I); and ii) results of the post-mortem macroscopic examination in slaughterhouses, histopathological examination and/or laboratory culture isolation tests (D).

In the past, in Region R, animals positive to the tuberculin test were sacrificed in sanitary service slaughterhouses and then inspected, and samples were taken and analyzed in the laboratory in order to isolate the agent. On the other hand, animals that were always negative to the tuberculin test in Region R, upon reaching a certain age, are discarded and set off for slaughter. Such data can be used insofar as the slaughterhouse maintains an individual record with identification of the source. These retrospective data on the Region R subject population enable us to determine the frequency distribution shown (5) in Table 1 and Figure 1.

The validity of a diagnostic test should reflect its capability to provide true results about what is being evaluated. To measure a test's validity, its sensitivity and specificity must be estimated.

In the bovine tuberculosis example presented above, sensitivity (U) would be the tuberculin test's ability to identify as positive every animal in the herd that is actually infected by *M. bovis*. Likewise, specificity (V) would be the tuberculin test's ability to identify as negative every animal in the herd that is actually not infected

by *M. bovis*. The characteristics of specificity enable it to be assessed in areas with herds free of bovine tuberculosis.

TABLE 1. Cross classification of direct & indirect tests. Denomination of cellular and marginal frequencies

Indirect test situation	Real situation (direct)		
	D+	D-	Total
I+	a	b	(a+b)
I-	c	d	(c+d)
Total	(a+c)	(b+d)	(n)

where:

a = Number of animals considered positive by the tuberculin test and which are found to be infected; these are the true positives.

b = Number of animals also found positive to the tuberculin test, but in which infection is not substantiated by post-mortem inspection or by organ cultures; these are the false positives.

c = Number of animals considered negative to the tuberculin test and found to be actually infected; these are the false negatives.

d = Number of animals considered negative to the tuberculin test and in which no infection is found; these are the true negatives.

a+b = Total number of infected animals.

b+d = Total number of uninfected animals.

n = Total animals.

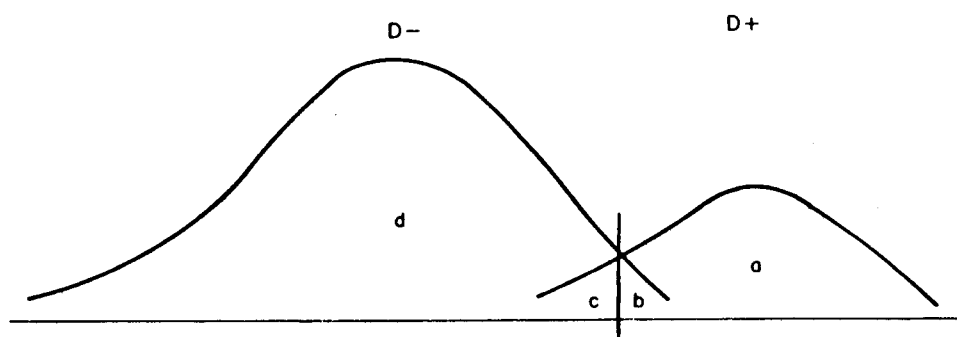


FIGURE 1. Distribution of overlapping frequencies of positive and negatives. Uncertainty zone: classification errors by indirect testing.

These characteristic values can be estimated from the data shown in Table 1:

$$U = a/(a+c)$$

$$V = d/(b+d)$$

The simple tuberculin test on the flat of the neck, using high-potency bovine PPD tuberculin, reaches optimum sensitivity of 91% and specificity of 89% (3).

When these indicators of validity are used, it should be remembered that in regions where the percentage of infection by *M. bovis* is relatively high, there may be cases of animals with recent infection for which the tuberculin test is positive. Nevertheless, the direct tests may fail in the confirmation due to the possibility that no macroscopic lesion may yet have developed. For this reason, the specimen of ganglia and organs for culture may not include tissue wherein the primary infection is present. In this case, the lesion will still be only microscopic. Consequently, the direct test for confirmation of the validity of the indirect test is liable not to provide total security about the presence of tuberculosis infection (4).

PROBABILITY THAT A BOVINE POSITIVE TO THE TUBERCULIN TEST IS REALLY INFECTED

By way of example, let us suppose that in a given region the veterinary service periodically tests a mass of 100,000 milk cows. If we want to know what percentage (P) of test-positive (I+) animals is actually infected (D+) with *M. bovis*, that is

$$P^*(D+/I+)$$

which means that we are concerned about the diagnostic efficacy of the tuberculin test. Proceeding with the example given, in Table 2, and applying the Bayes theorem (2):

$$P(D+/I+) = \frac{P(D+) P(I+/D+)}{P(D+) P(I+/D+) + P(D-) P(I+/D-)}$$

the diagnostic efficacy of the tuberculin test can be assessed. For the given example it would be

$$P(D+/I+) = \frac{(0.01) (0.95)}{(0.01) (0.95) + (0.99) (0.1)} = 0.08756.$$

*P = Probability

TABLE 2. Distribution of frequencies for 1% prevalence, 95% sensitivity and 90% specificity

Indirect test situation	Real situation (direct)		
	D+	D-	Total
I+	950	9,900	10,850
I-	50	89,100	89,150
Total	1,000	99,000	100,000

Therefore, in this example, only 8.8% of the test-positive animals would be really affected with tuberculosis. The $950/10,850 = 0.088$ rate was not utilized because, as a general rule, the number of animals in the "a" cell of Table 1 is unknown. According to the data in the example, it is 950. For this reason the marginal data must be utilized to estimate the diagnostic efficacy of an indirect test such as the tuberculin test (5).

Whereas the infection prevalence is 1% and the cattle population is $n = 100,000$ head, the number of positive or infected animals is 1,000. In Table 2 this corresponds to the first column total (a+c), i.e., the total of true positives [(a+c) = 1%(n)]. Whereas the indirect test sensitivity is 95%, its application will detect 950 of the 1,000 infected animals [a = 95% (a+c)].

The value of 950 corresponds to "a" in Table 2, i.e., truly positive animals also detected by the indirect diagnostic test. The difference of $1,000 - 950 = 50$ makes cell "c", which corresponds to actually positive animals not detected by the indirect test [c = (a+c) - a]. This quantity represents the false negatives. The second column of Table 2 corresponds to the real negatives, whose total (b+d) is the difference between the size of the population (n) and the total quantity of positive individuals (a+c), i.e., [n - (a+c) = (b+d)]. The quantity (b+d) is equal to $100,000 - 1,000 = 99,000$ negative cattle.

Whereas the specificity of the indirect diagnostic test is 90%, then it is possible to detect 89,100 of the 99,000 negative animals [d = 90%(b+d)]. This quantity corresponds to cell "d" in Table 2, which represents the real negative cattle also

identified as such by the indirect test. The difference of $99,000 - 89,100 = 9,900$ constitutes cell "b", which is the real negative animals not identified as such through the indirect test [$b = \{(b+d) - d\}$]. This quantity represents the false positives.

The total of animals in the first row (a+b) corresponds to the total of infected animals classed as such by the indirect diagnostic test, whereas the total in the second row (c+d) represents the total of noninfected animals classed in this category by the indirect diagnostic test. In the last two totals there is a certain amount of misclassification because the first total (a+b) includes false positives and the second total (a+d) includes false negatives.

PREDICTIVE VALUE OF THE DIAGNOSTIC TEST

In order to study all the elements that may affect the results of applying a massive diagnostic test, it is desirable to consider some situations in which both the sensitivity and the specificity of a test remain constant, while prevalence of infection presents different values. Let us suppose that an indirect diagnostic test, like the tuberculin test, is applied to three bovine populations of 100,000 head each.

Sensitivity and specificity are taken as 91% and 89% respectively. The first population has a prevalence of 1%, the second is 10%, and the third is 20%. The results are in Tables 3, 4 and 5.

The predictive capability of an indirect test (G) can be indicated by two indicators: the positive predictive value (G) and the negative predictive value (H), both of which can be expressed from the frequencies shown in Table 1:

$$G = a/(a+b)$$

$$H = d/(c+d)$$

For the data in Table 3, the indicators attain the following values:

$$G = \frac{910}{11,800} \times 100 = 7.8\%$$

$$H = \frac{88,110}{88,200} \times 100 = 99.9\%$$

TABLE 3. Cattle population with 1% prevalence rate
91% sensitivity and 89% specificity

	D+	D-	Total
I+	910	10,890	11,800
I-	90	88,110	88,200
Total	1,000	99,000	100,000

For the data in Table 4, the tuberculin test predictive capability indicators attain the following values:

$$G = \frac{9,100}{19,000} \times 100 = 47.9\%$$

$$H = \frac{80,100}{81,000} \times 100 = 98.9\%$$

TABLE 4. Cattle population with 10% prevalence rate
91% sensitivity and 89% specificity

	D+	D-	Total
I+	9,100	9,900	19,000
I-	900	80,100	81,000
Total	10,000	90,000	100,000

It can be seen that the change of +9 percentage points in prevalence produces a change of +40 percentage points in the positive predictive capability and -1 percentage point in the negative predictive capability. This last aspect is practically unchanged when compared to the increase in the positive predictive capability.

For the data in Table 5, the indicators attain the following values:

$$G = \frac{18,200}{27,000} \times 100 = 67.4\%$$

$$H = \frac{71,200}{73,000} \times 100 = 97.5\%$$

A marked increase in prevalence is again seen to produce a noticeable increase in positive predictive capability (practically 20 percentage points in relation to the situation of 10% prevalence). However, the growth in this indicator did not

TABLE 5. Cattle population with 20% prevalence rate
91% sensitivity and 89% specificity

	D+	D-	Total
I+	18,200	8,800	27,000
I-	1,800	71,200	73,000
Total	20,000	80,000	100,000

reach the magnitude observed when the prevalence went from 1% to 10%. On the other hand, H also declined slightly, but the magnitude of change was in this case somewhat greater than when prevalence rose from 1% to 10%.

These examples enable us to conclude that when an indirect diagnostic test is applied, its positive predictive capability declines if the bovine population has a very low level of infection.

According to Ranney (4), the application of these results to a bovine tuberculosis control and eradication program would indicate that in low prevalence situations -whether as a result of the natural history of the disease or because of the effect of control action- exclusive use of the tuberculin test as the sole method of detecting infection is not advisable. In these cases there should be a dynamic detection activity that uses procedures which permit tracing back from the slaughterhouse (discovery of lesions) to the herds where the agent is active. Once the problem herds are identified, the tuberculosis test can be applied every six months until the herd is declared 'clean'.

In epidemiological situations wherein the prevalence is very high, the positive predictive value of the tuberculin test is also very high. Therefore, the tuberculin test method and elimination of positive reactors can give good results.

ESTIMATING THE REAL PREVALENCE OF A DISEASE LIKE TUBERCULOSIS BASED ON THE RESULTS OF AN INDIRECT TUBERCULIN TEST

Table 6 illustrates a possible tuberculosis situation in a region:

TABLE 6. Cattle herd with prevalence of 1/1000
91% sensitivity and 89% specificity

Indirect test situation	Real situation (direct)		
	D+	D-	Total
I+	91	10,989	11,080
I-	9	88,911	88,920
Total	100	99,900	100,000

In this case the apparent prevalence (PAP) is: $11,080/100,000 = 0,11080$, that is, approximately 11%. Nevertheless, of the 11,080 positives to the indirect test, 10,989 are false positives. This result should be interpreted carefully considering the economic implications occasioned by sending to the slaughterhouse (sacrificed) animals reacting positively to the indirect test, as is the case with tuberculosis (5). The available knowledge about the test's specificity and sensitivity indicators helps to resolve this problem. This way the true prevalence rate (VEP) can be determined:

$$VEP = \frac{PAP + V - 1}{V + U - 1} = \frac{0,00080}{0,80000} = 0,001$$

Observation of the data confirms that the true prevalence is 1 per 1000.

PROBABILITY OF FAILURE TO DETECT INFECTED ANIMALS WHEN USING AN INDIRECT TEST LIKE THE TUBERCULIN TEST

The purpose is to conduct a series of diagnoses to detect tuberculosis in a herd of animals with an indirect test having known sensitivity and specificity. The aim is to assess the probability of failure to detect animals infected by *M. bovis*, i.e., to ascertain the probability of yielding false negatives. The variables in this problem are:

U, V and VEP

Working through the formula in the preceding section, we get:

$$VEP (V + U - 1) - V + 1 = PAP$$

which can be applied to the last example above as follows:

$$PAP = 0,001 (0,89 + 0,91 - 1) - 0,89 + 1 = 0,11080$$

$$PAP = p = 0,11080$$

$$(1 - PAP) = q = 1 - p = 1 - 0,11080 = 0,88920$$

Under these conditions the geometric distribution equation for assessing the probability of failure to detect positives with the indirect test may be applied to the probability of obtaining false negatives (7).

$$\left. \begin{array}{l} \text{Probability of} \\ \text{failure to} \\ \text{detect positives} \end{array} \right\} = \text{Pr (Failure +)} = \alpha$$

$$= P(\text{Failure +}) = (1-p)^n = q^n = \alpha$$

$$\alpha = q^n$$

$$\log \alpha = n \log q \quad \text{where } n = \frac{\log \alpha}{\log q}$$

This is the number of individuals required, or minimum size of the sampling, to have a probability equal to or less than " α " of failure to detect positives through the indirect test, i.e., the probability of detecting false negatives. This equation is applicable when the purpose is to conduct a study to detect the presence of positive individuals and this can be done only by screening. In this case, when a hypothetical " p " is very low, " q " approaches to 1.00 and in logarithmic terms the denominator of the equation approaches zero. On the other hand, when the purpose is to ascertain if a zone is "clean" of a disease, then " α " will have to be very slight, insofar as it is a question of the probability of failure to detect positives. This is why the formula's numerator will become smaller and smaller on the logarithmic scale (-3, -5, -6, etc.) while the minimum number of observations in the sampling will become all the greater.

DISCUSSION

In animal health, the detection of infection or disease in an animal is frequently not ascertained with full certainty. The highly reliable diagnostic procedures, although desirable, are nonetheless frequently impracticable—as in the case of tuberculosis— or are either too costly or too cumbersome for massive application in cattle populations.

When the epidemiological situation of a disease in an animal population of a certain magnitude

is studied, it is common to have to apply some not very sophisticated screening type of diagnostic test usually having a level of error in classifying the epidemiological state of the individual animals.

In any case, to facilitate the interpretation of the results when a screening type indirect test is applied, its validity must be assessed taking as reference the responses to a direct test. The levels of error, false positives and false negatives can be specifically estimated when the screening test is applied on a group of animals whose true epidemiological state is known.

The sensitivity and specificity related to the false negatives and false positives, respectively, are characteristic values of a test and do not depend on the disease's prevalence. Nevertheless, the predictive values are not stable and change according to the prevalence. When the prevalence of a disease declines, the positive predictive value also declines; this situation makes it an unreliable measure.

In the case of some chronic cattle diseases such as bovine tuberculosis, the application of screening tests is obviously always related to its control and eradication. Attention should thus be directed to some of the aspects stressed in this paper.

In the tuberculosis control programs, one of the preoccupations may be to prevent the disease from entering regions or groups of herds free of *M. bovis*. In this case, a high-sensitivity test is especially necessary because it is important not to let any false negatives "pass". In these circumstances, the specificity plays a secondary role.

Another situation that may occur in programs to control diseases like bovine tuberculosis is the need to eliminate infected animals from a region or group of herds. This circumstance also primarily requires a high-sensitivity test; otherwise the program would run the risk of not detecting many infected animals that would then remain among the herds and spread the disease.

Another different situation that the tuberculosis control programs could confront involves monitoring areas free of the disease in order to detect its presence. Specificity is much more

important than sensitivity in this case. In a free area well protected by the control program, it commonly happens that the isolated cases of reactors involve false positives.

In summary, a good screening test is one in which the frequency distributions of positives and negatives have the smallest possible area of overlap. The test will then have both high sensitivity and high specificity. When the area of overlap of the frequency distributions in an indirect test is defined, any attempt to increase its sensitivity will be in detriment of the specificity and vice versa.

REFERENCES

1. FELLER, W. An introduction to probability theory and its applications. Vol. 1. Wiley. New York. 1968.
2. FLEISS, J.L. Statistical Methods for Rates and Proportions. Wiley. New York. 1973.
3. FRANCIS, J., SEILER, R.J., WILKIE, I.W. O'BYLE, D., LUMSDEN, M.J., FROST, A.J. The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* 103: 420-435, 1978.
4. RANNEY, A.F. Los servicios de inspección de carne como ayuda en la erradicación de la tuberculosis. I. Seminario Internacional sobre Tuberculosis bovina para las Américas. Santiago, Chile. OPS, *Pub. Cient.* 258, 1972.
5. SCHWABE, C.W., RIEMANN, H.P., FRANTI, C.E. Epidemiology in Veterinary Practice. Lea & Febiger. Philadelphia, 1977.
6. TOMAN, K. Sensibilidad, especificidad y valor predictivo de los tests diagnósticos. Boletín Unión Internacional contra la Tuberculosis 56, No. 1-2, 1981.