Applying Epidemiology to Diagnostic Findings

A. H. Meek, *D.V.M.*
S. W. Martin, *D.V.M.*
Ontario Veterinary College
University of Guelph, Ontario
Canada N1G 2W1

A clinical diagnosis can be considered a probability statement. How good that statement is depends on a number of factors related to the examiner, the examined and the examination (Sackett et al, 1985). In addition to this, the ability to extrapolate ones findings from a few examined animals to the population eg. herd, will depend on how representative the examined animals are of the population and/or the number of animals sampled. In this paper, primary emphasis will be given to the examination, particularly to the use of tests, for detecting subclinical disease and to the extrapolation of results to the population.

Detecting Subclinical Diseases with Screening Tests

Tests can be classified as either pathognomonic or surrogate. Pathognomonic tests are ones for which the detection of a sign or substance is an absolute predictor of the disease or agent. In contrast to this, surrogate tests detect secondary changes which it is hoped will predict the presence or absence of the disease. For example, a positive culture of *Brucella abortus* from a cow's milk sample is pathognomonic for brucella infection. However, testing the milk for antibodies is a surrogate test, since it is not measuring the presence of the agent directly but rather the body's reaction to it. Surrogate tests, may produce false positive reactions, while pathognomonic tests do not. Both types of tests can have false negatives. Such false results and interpreting screening tests leads to the topic of sensitivity and specificity (Martin, 1977; Martin, 1983).

Sensitivity and Specificity

Let us assume that it is possible to correctly classify animals into two categories, those with disease X and those without it, using available methods. A new test has been developed and its ability to discriminate between diseased and nondiseased animals is to be evaluated.

The initial step is to select a sample of animals known to have disease X and a sample without it. It is important that the new test is biologically independent of the methods used to establish the true health status of the animals. Once selected, the animals are tested and classified as being positive, or negative, on the basis of the new test. The resultant cross classification, according to their true health status and the test results, can be displayed as follows:

	Actual Health Status (Disease X)		
Test Result	**Present (D+)**	**Absent (D−)**	**Total**
Positive (T+)	a	b	a + b
Negative (T−)	c	d	c + d
Total	a + c	b + d	n

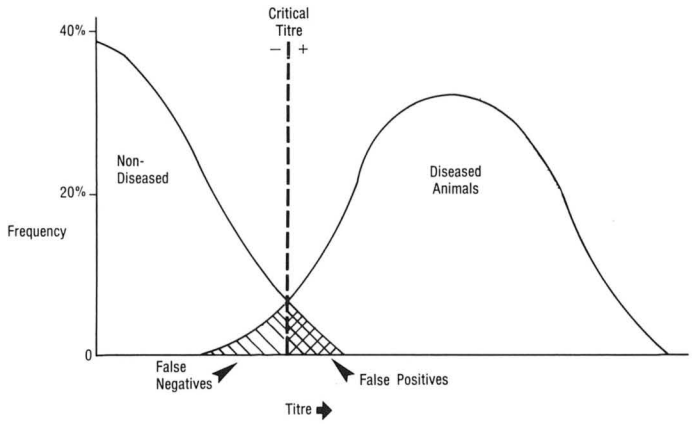
The sensitivity of the test is its ability to detect diseased animals and is defined as the proportion of the diseased animals that test positive, that is, a/(a+c). The specificity of the test is its ability to detect nondiseased animals and is defined as the proportion of nondiseased animals that test negative, that is, d/(b+d). Together, these two statistics describe how well a test can differentiate between nondiseased and diseased animals.

In a random sample of the overall population, the true prevalence proportion of disease P(D+), would be estimated by p(D+), that is, (a+c)/n. However, this is almost always unknown with only the test results being available, and hence the estimate of P(D+) is the apparent prevalence proportion, namely, (a+b)/n. The true and apparent prevalence proportions are equal only if b = c. Usually b tends to be larger than c and thus the apparent prevalence is usually higher than the actual prevalence.

For most surrogate tests there is usually an inverse relationship between sensitivity and specificity. That is, if the critical value of the test is adjusted so that its sensitivity is increased, then its specificity will be decreased. This results from the fact that the substances being measured may be present in nondiseased as well as diseased animals, and their respective distributions overlap. For example, Figure 1 displays the distribution of antibody titres to agent A in samples of nondiseased and diseased animals. Note that many nondiseased animals do not have a titre, while some have low titres and few have high titres. In diseased animals, the distribution is somewhat bell-shaped with only a few diseased animals with low titres and the rest with moderate to very high titres. It can be seen, that although the diseased animals have higher titres on average, the two distributions overlap, thus producing an inverse relationship between the sensitivity and specificity of tests measuring this response.

In application, a critical titre is selected so that animals with titres above that point are considered positive, and those with titres equal to or below it are considered negative.

FIGURE 1. The Distribution of Titres to Agent A in a Sample of Nondiseased and Diseased Animals.



With respect to the previous 2x2 table, diseased animals with titres above the critical titre are the true positives, their number being represented by 'a'; nondiseased animals with titres below the critical titre are the true negatives, represented by 'd'; nondiseased animals with titres above the critical titre are false positives, represented by 'b', and diseased animals with titres equal to or less than the critical titre are false negatives, represented by 'c'.

If the critical titre is altered to increase the sensitivity ie. lowered or moved to the left, the number of false positives will increase, and specificity will be lowered. If the critical titre is altered by moving it to the right, then the sensitivity of the test will decrease, and there will be a larger number of false negatives. An example of the effect of changing the critical titre when testing for traumatic reticuloperitonitis, is shown in Table 1 (Dubensky and White, 1983).

In general, sensitivity and specificity describe the discriminatory ability of a test, based on a single sample at a point in time. They do not describe how well the test will function if applied at various times in the disease process.

The Predictive Value of Screening Test Results

The predictive value of a positive test is defined as the proportion of diseased animals among those that test positive, that is, p(D+/T+), calculated as a/(a+b). Predictive value is important, as it reflects the way test results are applied in the field. That is, given that an animal has a positive test, what is the probability that it has the disease or infection under study? This question arises because the true state of health is unknown, hence, the clinician must argue from test results to the likelihood of disease, not the other way around.

The predictive value of a test, is influenced by the sensitivity and specificity of the test, as well as, by the true prevalence of disease in the population. Since the latter is usually unknown, it makes the selection of the 'best' test difficult. In general, as the true prevalence of the disease decreases, so will the predictive value of the test. The example in Table 1 illustrates the relationship between predictive value, and sensitivity and specificity, with the prevalence of disease being constant (37.3%).

TABLE 1. Sensitivity, Specificity and Predictive Value of Total Plasma Protein in the Diagnosis of Traumatic Reticuloperitonitis.

Total Plasma Protein cutoff value (g/L)	Sensitivity (%)	Specificity (%)	Predictive Value[1] Positive (%)	Negative (%)
65	97.0	11.3	39.4	85.7
70	95.2	28.3	44.1	90.9
75	88.9	46.2	49.6	87.5
80	74.6	64.2	55.3	81.0
85	61.9	74.5	59.1	76.7
90	41.3	87.7	66.7	71.5
95	30.2	92.5	70.4	69.0
100	20.6	96.2	76.5	67.1
105	15.9	98.1	83.3	66.2
110	6.3	99.1	80.0	64.0

[1] calculated at a prevalence of 37.3%

Reference: Dubensky, R. A. and M. E. White. The sensitivity, specificity and predictive value of total plasma protein in the Diagnosis of Traumatic Reticuloperitonitis. Can. J. Comp. Med. 1983, 47: 241-244.

Accuracy and Precision

Unlike sensitivity and specificity, which relate to the discriminatory power of a test, accuracy and precision relate more to "quality control". However, if a test is inaccurate and lacks precision, the results will influence its sensitivity and specificity. For present purposes, however, accuracy and precision will be treated independently of sensitivity and specificity.

An accurate test is one that on average yields a true measure of the substance of concern, eg. level of blood sugar. Precision is the ability of a test to give a consistent reading upon repeated testing of the same sample. The precision and accuracy of a test are influenced by the variability of the test itself, variability of the person who performs the test, and of differences between laboratories.

The results of a study (Reif et al, 1970) of intra and inter-individual variation (precision) in the interpretation of canine chest X-rays is shown in Table 2. The agreement between the two radiologists was 74% and, on average, they agreed with their own findings approximately 82% of the time. The average sensitivity and specificity of chest radiographs for detecting pulmonary disease, assuming histologic diagnosis is correct, was 72.4% and 87.1% respectively. Given the low specificity, radiography would not appear to be an appropriate method of screening dogs for respiratory disease if the true prevalence was low; as the predictive value of positive radiographs would be extremely low.

TABLE 2. Findings on the Sensitivity, Specificity and Precision of Radiography Techniques Used to Determine Pulmonary Disease in Dogs.

		Histological Diagnosis	
		Diseased	Non diseased
Radiographic Interpretation	Positive	100	8
	Negative	38	54
		138	62

Sensitivity = 100/138 = 72.4%

Specificity = 54/62 = 87.1%

In reexamining 130 of the above radiographs, the researchers disagreed with themselves 24 times and with each other 34 times resulting in:

Intra-individual Precision = 81.5% (18.5% error)

Inter-individual Precision = 73.9% (26.1% error)

Reference: Reif, J. S., Rhodes, W. H., and D. Cohen. Canine pulmonary disease and the urban environment I. The validity of radiographic examination for estimating the prevalence of pulmonary disease. Arch. Environ. Hlth. 20: 676-683, 1970.

Detecting the Presence of a Disease in a Population

Often during disease control or eradication programs, herds are tested to determine if the specified disease is present, or alternatively to ensure that it is absent. However, testing entire herds is expensive, and the veterinarian may have to accept the results of testing only a sample of the population.

When sampling is used for this purpose, a frequently asked question is what sample size is required so that one can be 95% or 99% confident that the herd, is disease free if no animals in the sampling give a positive result?

Infectious diseases tend to spread and would be expected to cluster somewhat within a herd. Thus, if the disease is present, the herd will likely contain more than one diseased individual. This knowledge may be utilized when sampling to detect disease. In fact, the sampling strategy is designed to detect disease if more than a specified number have the disease. The actual number or percentage of diseased animals to specify when calculating the sample size should be based on knowledge of the biology of the disease under investigation.

Table 3 presents the sample size required to be 95% or 99% certain that disease is present at or below the specified prevalence if no diseased animals are observed. For this purpose, the minimum number of diseased animals assumed to be present in a population is one, and for populations of greater than 100 individuals, the number of diseased animals is based on estimated prevalences ranging from 1 to 50 percent. It is important to note, that a formal random sampling method, with individuals as the sampling units, is required if the desired confidence level is to be attained. If no random selection is done, then the confidence one has in the result is unknown, at least quantitatively.

The formula, used to derive the numbers in Table 3 may be solved for D, rather than n, with the result being:

$D = (1-(1-a)^{1}/n)\ (N-((n-1)/2))$

TABLE 3. Sample sizes [1]required to be 95 and 99 percent confident that disease is present at/or below the specified prevalence D/N, if no diseased animals are observed.

Population Size	Prevalence of Disease: (D/N) × 100			
	1%	5%	10%	50%
30	29/30	23/27	19/23	5/7
60	57/60	38/47	23/31	5/7
100	95/99	45/59	25/36	5/7
300	189/235	54/78	28/41	5/7
500	225/300	56/83	28/42	5/7
1,000	258/368	58/86	29/43	5/7
10,000	294/448	59/90	29/44	5/7

[1] Derived using the following formula for Cannon and Roe, 1982:

$n = (1-(1-a)^{1/D})\ (N-((D-1)/2))$

where n is the required sample size,

'a' is the probability (confidence level) of observing at least one diseased animal in the sample when the disease affects at least D/N in the population,

D is the number of diseased animals in the population and

N is the population size.

The above formula is useful in that in provides the maximum number of diseased animals (D) expected in a population, with confidence a, when n individuals are examined and found to be free of disease.

Table 4 gives the probability of failure to detect diseased animals from an 'infinite' population, with the specified

TABLE 4. Probability of Failure to Detect Diseased Animals.

Prevalence	Number of Animals in Sample Tested					
	5	10	25	50	75	100
1%	0.951	0.904	0.778	0.605	0.471	0.366
2%	0.904	0.817	0.603	0.364	0.220	0.133
3%	0.859	0.737	0.467	0.218	0.102	0.048
4%	0.815	0.665	0.360	0.130	0.047	0.017
5%	0.774	0.599	0.277	0.077	0.021	0.006
6%	0.734	0.539	0.213	0.045	0.010	0.002
7%	0.696	0.484	0.163	0.027	0.004	0.001
8%	0.659	0.434	0.124	0.015	0.002	0.000
9%	0.624	0.389	0.095	0.009	0.001	0.000
10%	0.590	0.349	0.072	0.005	0.000	
12%	0.528	0.279	0.041	0.002	0.000	
14%	0.470	0.221	0.023	0.001	0.000	
16%	0.418	0.175	0.013	0.000		
18%	0.371	0.137	0.007	0.000		
20%	0.328	0.107	0.004	0.000		
24%	0.254	0.604	0.001	0.000		

proportion of positives and number of animals tested. For example, a series of random samples of 10 animals in a large population in which 5% are positive would fail to detect any positives in 59.9% of such groups.

References

1. Sackett, D.L., Haynes, R.B., and P. Tugwell. Clinical epidemiology: A basic science for clinical medicine. Little, Brown and Co., Toronto, 1985. 2. Martin, S.W. The evaluation of tests. Can. J. Comp. Med., 4:19-25, 1977. 3. Martin, S.W. Estimating disease prevalence and the interpretation of screening test results. Prev. Vet. Med., 2:463-472, 1984. 4. Dubensky, R.A. and M.E. White. The sensitivity, specificity and predictive value of total plasma protein in the diagnosis of traumatic reticuloperitonitis. Can. J. Comp. Med., 47:241-244, 1983. 5. Reif, J.S., Rhodes, W.H., and D. Cohen. Canine pulmonary disease and the urban environment I. The validity of radiographic examination for estimating the prevalence of pulmonary disease. Arch. Environ. Hlth., 20:676-683, 1970. 6. Cannon, R.M. and R.T. Roe. Livestock disease surveys: A field manual for veterinarians. Australian Bureau of Animal Health, Australian Government Publishing Service, Canberra, 1982.