A SIMPLIFIED SPREADSHEET IMPLEMENTATION OF MIXTURE POPULATION MODELLING TO ESTIMATE TEST SENSITIVITY AND SPECIFICITY

Cameron AR[1] and Eamens GJ[2]

[1]AusVet Animal Health Services, 140 Falls Road, Wentworth Falls, Australia, 2782
[2]Elizabeth Macarthur Agricultural Institute, PMB 8, Camden, NSW, Australia, 2570

Traditional approaches to estimating test sensitivity and specificity depend on testing populations of both known diseased and non-diseased animals. The status of animals is based on the use of a 'gold standard' test. Unfortunately, for many diseases, either no true gold standard exists or the use of the gold standard is too expensive to allow large populations to be studied. This study examines the performance of an ELISA to detect antibodies to *Mycobacterium paratuberculosis* in cattle (the bovine Johne's disease or BJD ELISA). This test has a low sensitivity and 'gold standards' to assess the true disease status of animals are slow, expensive or unreliable.

Mixture population modelling offers an approach to determining test sensitivity and specificity in the absence of a gold standard, through an analysis of the distribution of test results from a mixed population of diseased and non-diseased animals. In most situations, this distribution (or a log transformation of it[1]) will be composed of two normal distributions (representing the diseased and the non-diseased populations) which combine to form the typical bi-modal distribution usually observed. Mixture population modelling aims to determine the parameters of the underlying normal distributions. These parameters may then be used to determine the sensitivity and specificity of the test for any given cut-off value. Custom software has been used in the past to perform mixture population modelling[2]. This paper describes a simplified implementation using a computer spreadsheet to determine the parameters of the underlying distributions.

Materials & Methods

The data used in this study is derived from the BJD ELISA developed at the Elizabeth Macarthur Agricultural Institute (EMAI) as described by Eamens *et al* [3]. Raw optical density (OD) results are adjusted by dividing the mean OD of duplicate test wells by the mean OD of sextuplicate negative control wells, to provide the ELISA ratio (ER). The ERs from 8081 tests were used for this analysis. The tests were conducted on herds of mixed and unknown infection status, for both individual diagnostic purposes as well as routine herd tests forming part of a JD market assurance program. No data other than ERs (eg. herd of origin, age, clinical status) was available for analysis.

Analysis was undertaken using the MS Excel spreadsheet program. The ER data was log transformed, and a frequency distribution calculated, using a small interval (0.02)

and large number of categories (130). This frequency distribution was then plotted to detect the expected bimodal distribution. Arbitrary values describing the mean, standard deviation and a scaling factor for two normal distributions (the positive and negative populations) were chosen. These values were used to calculate the predicted counts for each observed category. The sum of the two predicted distributions was calculated and the curve describing this sum was plotted over the observed distribution. The parameters for the predicted distribution were modified by trial and error so that the observed and predicted curves were roughly superimposed. These values provided appropriate starting values for optimisation.

For each category in the frequency distribution, the absolute difference between the observed and predicted distributions was calculated and the sum of these absolute differences determined.

Determination of the best combination of parameters to describe the observed distribution was then achieved using the spreadsheet's 'Solver' function, which uses the Generalized Reduced Gradient nonlinear optimisation algorithm[4]. This is achieved by minimising the sum of absolute differences, by varying the six parameters describing the positive and negative distributions.

The resultant predicted distributions were then used to determine test sensitivity and specificity at a range of cut-off values, as well as producing an ROC curve.

Results

Due to the mixed source population and the low prevalence of infected animals, the magnitude of the right (positive) peak of the distribution was much smaller than the left (negative peak). Log ERs ranged from -0.6 to 1.05. The observed, positive and negative distributions, as well as the sum of the predicted distributions, is shown in Figure 1.

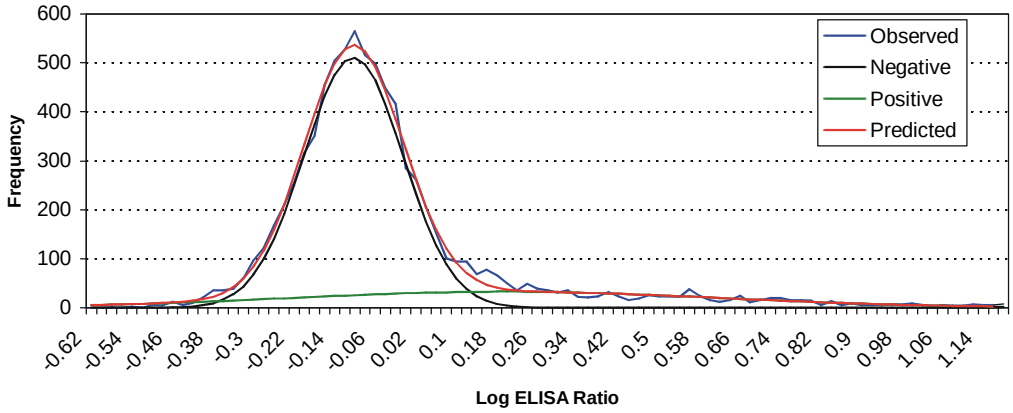

Figure 1. Observed and predicted distributions of ER values

The calculated parameters for the negative and positive distributions are shown in Table 1.

Parameter	Negative	Positive
Mean	-0.083	0.217
Standard Deviation	0.098	0.424
Scaling Factor	125.398	34.900

Table 1: Parameters for predicted distributions

The predicted sensitivities and specificities for a range of ER cut-off values are shown in Table 2 and the ROC curve in Figure 2.

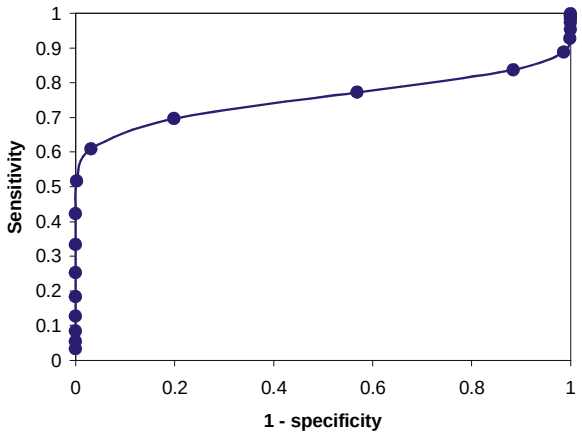

Figure 2: ROC curve for the BJD ELISA

Cut off ER	Sensitivity	Specificity
1.1	0.6607	0.8978
1.2	0.6275	0.9511
1.3	0.5962	0.9778
1.4	0.5665	0.9903
1.5	0.5386	0.9959
1.6	0.5123	0.9983
1.7	0.4876	0.9993
1.8	0.4643	0.9997
1.9	0.4424	0.9999
2	0.4218	1.0000
2.1	0.4023	1.0000
2.2	0.3840	1.0000

Table 2: Sensitivity and specificity values for the BJD ELISA with a range of cut off ERs.

Discussion

The simple approach described allows estimates of test sensitivity and specificity to be made without the need for knowledge of the true disease status of individual animals. The large volumes of data generated by routine laboratory testing or disease control programs can be used to better understand the performance of the test. This analysis would be further enhanced if more data on individual animals (eg herd and age) were available, allowing stratification of estimates.

It is however important to note that, in the case of the BJD ELISA at least, the test is not detecting infection, simply the presence of antibodies, which may appear some time after initial infection. The two populations described are therefore antibody positive and antibody negative, rather than JD positive and JD negative.

References

[1] Crofts N, Maskill W, Gust ID. 1988. Evaluation of enzyme-linked immunosorbent assays: a method of data analysis. Journal of Virological Methods **22**: 51-59.

[2] Vannuffel P, Gilot P, Limbourg B, Naerhuyzen B, Dieterich C, Coene M, Machtelinckx L, Cocito C. 1994. Development of species-specific enzyme-linked immunosorbent assay for diagnosis of Johne's disease in cattle. Journal of Clinical Microbiology **32**: 1211-1216.

[3] Eamens GJ. 1998. Development and assessment of EMAI abosorbed ELISA for bovine paratuberculosis. Report to SCAHLS. 12 pp.

[4] Microsoft Corporation, 1996. Microsoft Excel 97 computer software.