大韓獸醫學會誌(2004) 第44卷 第4號
Korean J Vet Res(2004) 44(4) : 669~676

Diagnostic performance of enzyme-linked immnosorbent assays for diagnosing paratuberculosis in cattle: a meta-analysis

Son-Il Pak*

Department of Veterinary Medicine, Kangwon National University, Chunchon 200-701, Korea
(Accepted: August 23, 2004)

Abstract : To evaluate the diagnostic accuracy of two commercial ELISA tests (Allied- and CSL-ELISA) for the diagnosis of *Mycobacterium paratuberculosis* in cattle, Meta-analysis using English language papers published during 1990-2001 was performed. Diagnostic odds ratios (DOR) were analyzed using regression analysis together with summary receiver operating characteristic (ROC) curves. The difference in diagnostic performance between the two ELISA systems was evaluated by using linear regression. Publication bias was assessed by funnel plot and linear regression. The pooled sensitivity and specificity were 44% (95% CI, 38 to 51) and 98% (95% CI, 96 to 99) for the random-effect model. The DOR between studies was heterogeneous. The area under the fitted ROC curve (AUC) was 0.72 for the unweighted and 0.77 for the weighted model. Maximum joint sensitivity and specificity for the unweighted and weighted model from their summary ROC curve were 70% and 75%, respectively. Based on the fitted model, at a specificity of 95%, sensitivity was estimated to be 52% for the unweighted and 57% for the weighted model. From the final multivariable model study characteristic, the country was the only significant variable with an explained component variance of 13.3%. There were no significant differences in discriminatory power, sensitivity, and specificity between the two ELISA tests. The overall diagnostic accuracy of two commercial ELISA tests was moderate, as judged by the AUC, maximum joint sensitivity and specificity, and estimates from the fitted model and clinical usefulness of the tests for screening program is limited because of low sensitivity and heterogeneous of DOR. It is, therefore, recommended to use ELISA tests as a parallel testing with other diagnostic tests together to increase test sensitivity in the screening program.

Key words : meta-analysis, *Mycobacterium paratuberculosis*, Diagnostic test

Introduction

The Enzyme Linked Immune System Assay (ELISA) as well as some other diagnostic tests (i.e., agar gel immunodiffusion) is used to detect humoral antibodies against *Mycobacterium avium paratuberculosis*, the agent for bovine paratuberculosis (PTB). The sensitivity and specificity of the ELISA tests have been reported [15, 21-22, 27]. These published articles, however, presented somewhat conflicting results in both diagnostic indices. In the screening program, the sensitivity of a diagnostic test is of great importance, and generally, the use of serology in the diagnosis of PTB is considered of limited value mainly due to low sensitivity. The wide ranges of reported diagnostic indices may be due partly to differences in study size and variations in the subject's characteristics, prevalence of disease in a specific region or region of the country in which the study was conducted, variation in the within-herd prevalence, clinical stages of animals studied, or random variation. To author's knowledge, no published report thus far has attempted to explain the difference in both indices for PTB diagnostic ELISA performance taking into account all published literature.

Statistical methods for evaluating diagnostic tests by meta-analysis have been developed [14, 16]. Meta-analysis can be used to summarize the overall diagnostic accuracy of tests after combining or integrating the results of several independent studies, to increase statistical power of a statistical test, and to assess the conclusions from studies through improvement of the estimates of effect size from various studies. As an outcome measure for meta-

*Corresponding author: Son-Il Pak
Department of Veterinary Medicine, Kangwon National University, Chunchon 200-701, Korea
[Tel: +82-33-250-8672, Fax: +82-33-244-2367, E-mail: paksi@kangwon.ac.kr]

analysis of diagnostic test data, receiver operating characteristic (ROC) curve, which plots sensitivity against 1-specificity, has been applied [10, 23].

The objective of this meta-analysis was to determine the overall diagnostic accuracy of two commercial ELISAs (Allied- and CSL-ELISA) using a summary ROC curve, specifically, "what is the overall discriminatory power of ELISA in diagnosing PTB?". In addition, this study was aimed at updating or summarizing the available evidence and giving researchers further insights into the diagnostic value of the ELISA for the detection of PTB, taking into account some of the potential sources of heterogeneity between studies.

Materials and Methods

Literature search and criteria for article selection

An electronic database produced by the National Library of Medicine (MEDLINE) was searched for relevant articles using paratuberculosis, Johne's disease, cattle, dairy cattle, dairy cows, diagnosis, serology, ELISA, and diagnostic performance as keywords. Also examined were reference lists of identified articles to procure additional studies. To prevent selection bias of articles for analysis, two criteria were predefined. Criteria for inclusion included: 1) original studies published during 1990-2001, 2) clinical studies using either one of two ELISA tests (Allied ELISA (Allied Laboratories) or the CSL ELISA (IDEXX)), 3) studies satisfying the sample size criteria (see sample size), 4) if not field-derived samples, studies used more than 200 reference standard or repository specimens of being microbiologically proven PTB cases, and 5) studies reporting both sensitivity and specificity. Criteria for exclusion included: 1) reviews, abstracts, editorials, letters and comments, 2) studies not performed on cattle, 3) studies which focused on experimental challenge, 4) studies that did not use culture method for PTB confirmation, 5) samples other than sera (i.e., milk), 6) non-English studies, 7) studies unable to derive the true positive and true negative from the data that were given, 8) unpublished data and conference proceedings, and 9) diagnostic methods other than the two commercial ELISA tests.

Using the above-mentioned inclusion and exclusion criteria, studies were selected during the review by checking the titles and abstracts downloaded from MEDLINE. Those studies that were unclear whether they were eligible or for which disagreements persisted, were retrieved and the final decision was made based on the review of the full article.

Sample size

Sample size was considered in the selection of eligible studies. The sample size criteria were based on

Table 1. Diagnostic accuracy of commercial ELISA for the detection of paratuberculosis (PTB) in cattle

First author (year, country)	PTB status			Sensitivity (95% CI)	Specificity (95% CI)	Diagnostic odds ratio (95% CI)
	ELISA	+	-			
Cox (1991, Australia)[a]	CSL	170	997	0.73 (0.66, 0.80)	0.99 (0.96, 1.00)	1341.1 (321.6, 5592.7)
Collins (1991, USA)	Allied	150	196	0.47 (0.39, 0.56)	0.99 (0.96, 1.00)	87.2 (20.9, 364.1)
Ridge (1991, USA)[a]	CSL	150	1000	0.47 (0.39, 0.55)	1.00 (1.00, 1.00)	448.5 (108.0, 1862.6)
Yokomizo (1991, Japan)[a]	CSL	156	3880	0.68 (0.61, 0.75)	1.00 (1.00, 1.00)	1643.0 (642.2, 4203.3)
Sockett (1992, USA)	CSL	177	196	0.45 (0.37, 0.52)	0.99 (0.96, 1.00)	78.2 (18.8, 324.9)
Sockett (1992, USA)	Allied	177	196	0.59 (0.51, 0.66)	0.95 (0.91, 0.98)	29.6 (14.2, 61.6)
Collins (1994, USA)	CSL	177	196	0.51 (0.44, 0.59)	0.95 (0.91, 0.97)	19.7 (9.8, 39.7)
Sweeney (1995, USA)	CSL	408	485	0.45 (0.40, 0.50)	0.99 (0.98, 1.00)	98.8 (36.2, 269.4)
Whitlock (2000, USA)	CSL	107	315	0.23 (0.16, 0.33)	0.95 (0.92, 0.97)	5.7 (2.9, 11.2)
Muskens (2000, Netherlands)	CSL	339	1396	0.35 (0.30, 0.40)	0.99 (0.98, 1.00)	60.8 (33.0, 112.0)
Dargatz (2001, USA)	CSL	590	723	0.50 (0.46, 0.54)	0.97 (0.95, 0.98)	30.4 (19.5, 47.5)
Pooled estimate after excluding 3 outliers [4, 19, 28]						
Random-effect				ND	0.98 (0.96, 0.99)	ND
Fixed-effect				ND	0.97 (0.97, 0.98)	ND
x^2 of heterogeneity				23.1 (df=7, P=0.002)	0.5 (df=7, P=0.99)	203.2 (df=7, P<0.001)*

*Test statistic was computed after adjusting for covariates, using by PROC CATMOD in SAS.

[a]Studies were excluded from the further analysis because of outlier effect.

ND=Not determined; because of heterogeneities among studies no pooled estimate was calculated.

the reported sensitivity (range, 0.23-0.73; Table 1) and specificity (range, 0.95-1.0; Table 1) of ELISA tests. Single binomial parameter calculations were used to determine minimum required sample size [12]. Sample size was 69 for sensitivity and 35 for specificity, assuming a 90% of specificity. Therefore, studies with 69 or more animals with disease and 35 or more animals without disease were required to be included in meta-analysis.

Data extraction and summary ROC curve

The extracted data from each study included the rates of true-positive (TP) or sensitivity, false-negative (FN), or 1-sensitivity, true-negative (TN) or specificity, and false-positive (FP) or 1-specificity according to standard definitions. If absolute numbers were not available from the original papers, the author derived the numbers using sensitivity and specificity and rounded-off to the nearest integers. The 95% confidence intervals (CIs) of the TP and TN for individual studies are calculated according to Diamond's formula [7].

Spearman correlation was used to examine whether monotonically increasing the relationship between TP and FP is present [14]. The statistical homogeneity of TP and FP across studies was tested by usual x^2 test of independency with k-1 degree of freedom (k= number of studies). When the relationships are determined to be correlated (>0.5) or homogeneous (i.e., x^2 test was not significant), statistical pooling was performed using the sum of test positives and false negatives for pooled sensitivity and the sum of test negatives and false positives for pooled specificity. Both fixed-effect model (FEM) and random-effect model (REM) [6, 16] were used to obtain combined estimates.

Simply averaging sensitivity and false positive from various studies would not represent correctly the diagnostic performance because of heterogeneity across studies. Thus, a summary ROC (SROC) curve was constructed using the data combined from different studies. Both weighted and unweighted regression models were fitted. The weights were proportional to the inverse of the variance of difference between logit TP and logit FP of each study, representing the within-study variation. This approach means sample size of each study was taken into account by weighting each observation. Although the validity of this weighing is still under debate for meta-analysis of diagnostic studies [13], this approach has widely been used by many investigators [25]. In this study the results of the two methods were presented together for comparison purpose. In both models, the following logit-difference-sum model [13] was used: D=a+bS, where D is the difference between logit TP and logit FP and is a measure of discriminatory power (i.e., how well the test discriminates between the two populations of diseased and non-diseased animals). S is the sum of the two transforms and is a measure of the threshold for classifying a test as positive, which has a value of 0 when sensitivity equals specificity. The regression slope, b, represents the variation of diagnostic discrimination of the test across the individual studies due to threshold differences. If b=0 or is near zero, the odds ratio is constant across studies and the resulting SROC curve is nearly symmetrical. The intercept, a, can be interpreted as the overall log odds ratio combined over studies. The higher the intercept the closer the curve will be to the upper-left corner. Once the intercept and slope are estimated by the regression analysis, data were back-transformed into an SROC curve to get predicted sensitivity (TPR) for a given FP, according to the following formula [13]:

$$TPR = \frac{1}{1 + \frac{1}{\left(e^{\frac{a}{1-b}}\right) \times \left(\frac{FP}{1-FP}\right)^{\frac{1+b}{1-b}}}}$$

The area under the ROC curve (AUC) was computed as an overall measure of diagnostic accuracy of the test.

Measure of diagnostic performance

The diagnostic odds ratio (DOR), a quantitative measure of the power of the test to discriminate diseased from non-diseased [26], was calculated for each individual study. The DOR calculates the ratio of the odds of a positive ELISA result in the animals with PTB and the odds of positive ELISA test result in the non-PTB animals: (TP/FN)/(FP/TN). A DOR of 1 means no discriminative power at all, and a DOR of greater than 1 indicates that the odds of a positive ELISA test result are higher in the PTB-diseased population. Heterogeneities of DORs were tested using the logistic regression, adjusting for covariates (PROC CATMOD, SAS Institute).

Based on the common ROC curve from each individual study, also defined were the Q-statistic (Q^*), maximum joint sensitivity and specificity that is intersected by a diagonal line that runs from the top left corner to the

bottom right corner of the ROC curve [18]. This point, at which sensitivity and specificity are equal, is the maximum attainable value of sensitivity and specificity for ELISA. This index is interpreted similar to the AUC and calculated by using the formula [6]: $Q^* = (1+e^{-A/2})^{-1}$, where A is the summary log DOR.

Comparison of Allied- and CSL-ELISA

The null hypothesis was examined so that there would be no difference in diagnostic performances between the two ELISA tests, by using a dummy variable in the regression model with the term S. The resulting model is $D=a+b_S \times S+b_e \times E$, where E is a dummy variable indicating the type of ELISA (E=0 for CSL and E=1 for Allied-ELISA), b_S and b_e regression slopes for S and E, respectively.

Potential sources of heterogeneity and meta-regression

The presence of overall heterogeneity of the dataset was assessed by x^2 test. This statistic is approximately distributed as x^2 distribution on k-1 *df* (k=number of studies) under the null hypothesis [1]. Also investigated was the extent to which the results of the current meta-analysis may have been biased as a result of the selective inclusion of articles with positive findings (publication and selection bias) using funnel plots. This plot was constructed by plotting the natural logarithm of the individual summary odds ratio (ln DOR) against sample size [24]. From this plot, large studies tend to lie in a narrow band at the top of the scatter plot, while the smaller studies, with more variation in results, fan out over a larger area at the bottom, indicating that for studies with no publication bias (i.e., studies included both positive and negative results), the points will be symmetrically distributed around the overall effect. Linear regression was used to test the hypothesis that a funnel plot is symmetrical [11]. To take into account the sources of variability in DOR between studies (i.e., if heterogeneity of DOR is significant), an evaluation was made of the relationship between each source of variation as independent variables (year of publication [entered the integer year when papers were published], country in which the study was performed [USA=0 vs other countries=1], type of ELISA [CSL-ELISA=0 vs Allied-ELISA=1], type of sample [repository=1 vs field=1]) and logarithmic DOR as a dependent variable in the multivariable meta-regression. A test was employed for blocks of regression coefficient [8] to test the null hypothesis that the number of p study characteristics explains a significant amount of variation in DOR and to test whether an additionally included study characteristic explains any additional variation in DOR when a block of p-1 study characteristic has already been entered in the regression model. Test statistics, Q_{change} for the two hypothesis tests were compared with x^2 distribution of df=p and df=1, respectively. The components of variance (i.e., how much of the variance in the study is due to random error) explained by each covariate were estimated using REM (PROC MIXED, SAS Institute). Values of $P<0.05$ were considered significant and statistical software programs (The SAS system version 8.2, SAS Institute Inc, Cary, NC and Meta-test version 0.6, New England Medical Center, Boston) were used to perform all analyses.

Results

For the 11 studies retrieved, (9 studies for CSL-ELISA [3-5, 17, 19, 21-22, 27-28] and 2 for Allied-ELISA [2, 20]) the homogeneity test was highly significant for sensitivities ($x^2 = 44.6$; 10 *df*; $P<0.0001$) but was not significant for specificities ($x^2=0.96$; 10 *df*; $P=0.99$). The spearman correlation between the TP and FP was 0.05 (P=0.88). The heterogeneity of sensitivity was due to its great variation in ranges of 23-73% (Table 1). Three studies [4, 19, 28] were outliers with a DOR range of 448.5-1643.0, compared with DORs ranging from 5.7 to 98.8 for the other 8 studies (6 studies for CSL-ELISA [3, 5, 17, 21-22, 27] and 2 for Allied-ELISA [2, 20]). These outliers were further confirmed by plotting the ratio of the logarithmic DOR and its standard error on the y-axis against the reciprocal of its standard error on the x-axis (so-called Galbraith plot), and the resultant SROC curve for 8 studies were shifted upward, compared with when all 11 studies were included. When 3 outliers were excluded, the homogeneity was rejected for sensitivities ($x^2 = 23.1$; 7 *df*; $P=0.002$), but not for specificities ($x^2 = 0.5$; 7 *df*; $P=0.99$). For eight studies, the logit-difference-sum weighted model had an estimated intercept of 2.2 (SE=1.72) and estimated slope -0.41 (SE=0.36). The resulting fitted model is: $D=2.2-0.41 \times S$, which corresponds to logit $TP=1.54+0.42 \times$ logit FP. The hypothesis of a constant odds ratio was not rejected because the slope was not significantly different from zero (t=-1.12; P=0.31), indicating a common log-odds

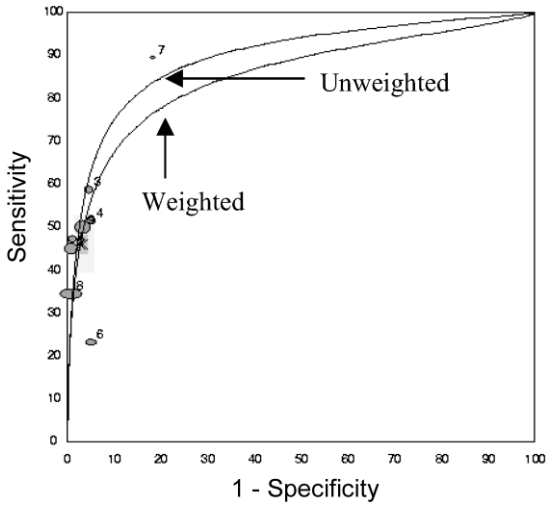

Fig. 1. Weighted- and unweighted summary receiver operating characteristic (SROC) curve for diagnostic accuracy of the 8 pooled studies [2-3, 5, 17, 20-22, 27]. The dot size of the study on the graph is proportional to the square root of the size in the disease and non-disease arms of the study.

ratio was 2.17. A constant odds ratio suggests that the outcome of all studies can be represented by point on a single symmetric logistic ROC curve. The spearman correlation between the TP and FP was 0.55 (P=0.16), suggesting that it was appropriate to combine results using SROC analysis. The DOR was not homogeneous (x^2= 203.2; 7 df; P<0.001), suggesting DORs were still too heterogeneous to be combined.

The intercept (coefficient=1.7; P=0.26) and the slope (coefficient=-0.46; P=0.20) for unweighted regression model were similar to those of weighted model. The resulting AUC was 0.72 for unweighted model and 0.77 for weighted model (Fig. 1). Based on eight studies, the maximum joint sensitivity and specificity rate for ELISA in the unweighted model was 70%, and the corresponding estimates for the weighted model was 75%.

For the model to compare diagnostic performance between the two ELISA tests, the estimated values for the regression parameters were as follows: a=0.98, b_s=-0.61, and b_e=0.78. The coefficient slope was not significantly different from zero (two-tailed t-test=1.03; P=0.35), indicating no significant difference in discriminatory power between the two tests. Table 2 compares TPR for the two tests for given FP.

An inverted funnel plot and regression analysis did not suggest evidence of publication bias (regression coefficient for intercept=2.03; two-tailed t-test=0.81; P=0.45).

Table 2. Predicted sensitivity (TPR) of CSL- and Allied-ELISA from eight studies for a given false-positive (FP) in regression model[a] including the term (S), difference between logit TP and logit FP

	TPR for ELISA			TPR for ELISA	
FP	CSL	Allied	FP	CSL	Allied
0.005	0.34	0.45	0.05	0.47	0.59
0.01	0.38	0.50	0.10	0.52	0.64
0.02	0.42	0.54	0.15	0.55	0.66
0.03	0.44	0.56	0.20	0.57	0.68

[a]The regression model: D=0.98−0.61×S+0.78×(type of ELISA), where D=sum between logit TP and logit FP.

Table 3. Predicted sensitivity of ELISA at a given ranges of specificity from eight studies, using weighted and unweighted fitted model

Specificity	Weighted model	Unweighted model
0.95	0.57	0.52
0.85	0.69	0.63
0.75	0.75	0.68
0.65	0.78	0.71
0.55	0.81	0.74
0.45	0.84	0.77
0.35	0.86	0.79
0.25	0.88	0.82

The null hypothesis that year of publication, type of sample and type of ELISA study characteristics are not related to DOR after controlling for the country characteristic was not rejected (Q_{change}=1.19; 3 df; P= 0.55). Thus, based on the final multivariable model the study characteristic country explained a significant amount of variation (13.3%) in DOR (Q_{change}=6.6; 1 df; P=0.01) and the other covariates were not significant.

From the fitted model, at a specificity of 95%, sensitivity was estimated to be 52% for unweighted model and 57% for weighted model (Table 3).

Discussion

A summary was made of the discriminatory performance of ELISA in diagnosing PTB through the use of an SROC curve for diagnostic meta-analysis, DOR, and maximum joint sensitivity and specificity. Summary ROC curves, like conventional ROC curves, graphically represent sets of sensitivity and specificity rates for a

diagnostic test as the threshold of a positive value for a test is varied. The difference between conventional ROC curves and SROC curves is that the former summarizes the performance of a test in a single study population and the latter describes a single set of operating characteristics for a test across multiple studies.

The sensitivities were not homogeneous even after excluding three outlier studies and thus differences among estimates of TP are unlikely to be due to random variation alone. If correlation is poor and the TP and FP are both homogeneous, then random binomial variation is one explanation for the differences among studies [14]. The author also explained that when estimates of TP and FP are poorly correlated and heterogeneous neither a summary ROC curve nor point estimates are appropriate. In this study, the estimates of sensitivities were not homogeneous but the medium correlation for the eight studies (coefficient=0.55; P=0.16) were found in between TP and FP, suggesting that it was appropriate to combine results using SROC analysis.

The maximum joint sensitivity and specificity was 75% for weighted model and 70% for the unweighted model. This point is an overall measure of the discriminatory power of a test (i.e., a perfect test would have a maximum joint sensitivity and specificity of 100% because of FP=TP=0). This point does not indicate the only, or even the best combination of sensitivity and specificity for a particular clinical setting. The AUC was 0.77 for the weighted model, indicating a relatively good discriminatory power of the ELISA tests. However, no significant difference in discriminatory power was detected between the Allied- and CSL-ELISA tests. In addition, as shown in Table 1, the sensitivity and specificity of the Allied-ELISA were almost identical to those of the CSL-ELISA. Overall, the two ELISAs showed a low sensitivity but high specificity. The lower sensitivities of the two ELISAs may be the inherent characteristics of the test, but the differences in clinical setting between studies may contribute this in part; because animals that shed the organism in their feces are regarded as being in advanced stage of clinical course and, thus are more likely to be serologically positive [21]. This means that as the percentage of nonshedders in a herd increase, the sensitivity of the test will probably decrease. The low sensitivity also suggests that ELISA test alone, particularly in herds with a high prevalence of PTB, may not be appropriate for use in diseases control such as a test-and-slaughter program, even though serological tests may be most convenient screening tests.

When outlier studies were excluded, only two countries (USA and Netherlands) were included for variance estimate. The variables of country and year of publication explained 13.3% and 13.2% of the between studies variance, respectively. The explained variance was only 45.1% when all four covariates were included, representing further covariates that need to be explored to explain the heterogeneity of DORs. This was also confirmed in the final multivariable FEM: The statistic Q_{change} for the error term was 32.5 with 6 df (P<0.001), indicating that extra-variation in DOR estimates cannot be explained using a few study characteristics.

This study has some limitations, some of which are inherent in meta-analysis. First, in the current study, the spectrum of disease (i.e., fecal shedders and non-shedders, clinical and subclinical stages of PTB) was not considered because of few relevant studies. Secondly, an evaluation was made of the year of publication as a source of variation on DOR but there may be many potential sources of variability such as a English-language bias [9]. Some other covariates (i.e., age of animals studied, prevalence of PTB in study area) were not considered in this study, because information on these covariates was not available across all studies. Thirdly, the current literature search method focusing only MEDLINE may not cover all potentially eligible studies. Thus, it is possible that these conclusions are not entirely representative of the diagnostic accuracy of the ELISA for the detection of PTB.

Studies were included in which the fecal culture method was employed as a gold standard for PTB confirmation. However, because this method has some known disadvantages such as poor sensitivity, long incubation time, frequent contamination, cost, and variability in technique and media preparation [27], information bias may be created inherently when reporting sensitivity of an ELISA in comparison to fecal culture method. In addition, an ELISA is not beneficial for use in animals less than two years of age because ELISAs typically have not been validated in young animals that are in the early stages of infection [15, 22]. Therefore, future studies should be directed to assess the correlation between level of fecal shedding and ELISA result to give researchers some hints about the potential use of an ELISA as a method of diagnosing

PTB-infected cattle for control purposes.

Acknowledgement

This study was supported by a grant (No. 0904001-1-1) from the High-Technology Development Project for Agriculture and Forestry, Ministry of Agriculture and Forestry, Republic of Korea.

References

1. **Cochran, W. G.** The combination of estimates from different experiments. Biometrics. 1954, **10,** 101-129.
2. **Collins, M. T., Sockett, D. C., Ridge, S. and Cox, J. C.** Evaluation of a commercial enzyme-linked immunosorbent assay for Johne's disease. J. Clin. Microbiol. 1991, **29**, 272-276.
3. **Collins, M. T., Sockett, D. C., Goodger, W. J., Conrad, T. A., Thomas, C. B. and Carr, D. J.** Herd prevalence and geographic distribution of, and risk factors for, bovine paratuberculosis in Wisconsin. J. Am. Vet. Med. Assoc. 1994, **204,** 636-641.
4. **Cox, J. C., Drane, D. P., Jones, S. L., Ridge, S. and Milner, A. R.** Development and evaluation of a rapid absorbed enzyme immunoassay test for the diagnosis of Johne's disease in cattle. Aust. Vet. J. 1991, **68,** 157-160.
5. **Dargatz, D. A., Byrum, B. A., Barber, L. K., Sweeney, R. W., Whitlock, R. H., Shulaw, W. P., Jacobson, R. H. and Stabel, J.R.** Evaluation of a commercial ELISA for diagnosis of paratuberculosis in cattle. J. Am. Vet. Med. Assoc. 2001, **218,** 1163-1166.
6. **DerSimonian, R. and Laird, N.** Meta analysis in clinical trials. Control Clin. Trials. 1986, **7**, 177-188.
7. **Diamond, G. A.** Limited assurance. Am. J. Cardiol. 1989, **63**, 99-100.
8. **Egger, M., Davey-Smith, G., Schneider, M. and Minder, C.** Bias in meta-analysis detected by a single, graphical test. Brit. Med. J. 1997, **315**, 629-634.
9. **Ferrer, R. L.** Graphical methods for detecting bias in meta-analysis. Fam. Med. 1998, **30**, 579-583.
10. **Hanley, J. A. and McNeil, B. J.** The meaning and use of the area under an ROC curve. Radiology. 1982, **143**, 29-36.
11. **Hedges L. V.** Fixed effect models. In: Cooper H, Hedges LV. The handbook of research synthesis, pp. 285-299. Russell Sage Foundation, New York, 1994.
12. **Kahn, H. A. and Sempos, C. T.** Statistical methods in epidemiology. Oxford University Press, New York, 1989.
13. **Littenberg, B. and Moses, L. E.** Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med. Decis. Making. 1993, **13**, 313-321.
14. **Midgette, A. S., Stukel, T. A. and Littenberg. B.** A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Med. Decis. Making. 1993, **13,** 253-257.
15. **Milner, A. R., Mack, W. N., Coates, K. J., Hill, J., Gill, I. and Sheldrick, P.** The sensitivity and specificity of a modified ELISA for the diagnosis of Johne's disease from a field trial in cattle. Vet. Microbiol. 1990, **25**, 193-198.
16. **Moses, L. E., Shapiro, D. and Littenberg. B.** Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat. Med. 1993, **12**, 1293-1316.
17. **Muskens, J., Barkema, H. W., Russchen, E., van Maanen, K., Schukken, Y. H. and Bakker, D.** Prevalence and regional distribution of paratuberculosis in dairy herds in The Netherlands. Vet. Microbiol. 2000, **77**, 253-261.
18. **Owens, D. K., Holodniy, M., McDonald, T. W., Scott, J. and Sonnad, S.** A meta-analytic evaluation of the polymerase chain reaction for the diagnosis of HIV infection in infants. JAMA. 1996, **275**, 1342-1348.
19. **Ridge, S. E., Morgan, I. R., Sockett, D. C., Collins, M.T., Condron, R. J., Skilbeck, N. W. and Webber, J. J.** Comparison of the Johne's absorbed EIA and complement fixation test for diagnosis of Johne's disease in cattle. Aust. Vet. J. 1991, **68**, 253-257.
20. **Sockett, D. C., Carr, D. J., Richards, W. D. and Collins, M. T.** A repository of specimens for comparison of diagnostic testing procedures for bovine paratuberculosis. J. Vet. Diagn. Invest. 1992, **4**, 188-191.
21. **Sockett, D. C., Conrad, T. A., Thomas, C. B. and Collins, M. T.** Evaluation of four serological tests for bovine paratuberculosis. J. Clin. Microbiol. 1992, **30**, 1134-1139.
22. **Sweeney, R. W., Whitlock, R. H., Buckley, C. L. and Spencer, P. A.** Evaluation of a commercial enzyme-linked immunosorbent assay for the diagnosis of

paratuberculosis in dairy cattle. J. Vet. Diagn. Invest. 1995, **7**, 488-493.

23. **Swets, J. A.** Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. Psychol. Bullet. 1986, **99**, 181-198.
24. **Tang, J. L. and Liu, J. L. Y.** Misleading funnel plot for detection of bias in meta-analysis. J. Clin. Epidemiol. 2000, **53**, 477-484.
25. **Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H. and Thompson, S. G.** A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Stat. Med. 2000, **19**, 3417-3432.
26. **Walter, S. D., Irwig, L. and Glasziou, P. P.** Meta-analysis of diagnostic tests with imperfect reference standards. J. Clin. Epidemiol. 1999, **52**, 943-951.
27. **Whitlock, R. H., Wells, S. J., Sweeney, R. W. and van Tiem, J.** ELISA and fecal culture for paratuberculosis (Johne's disease): sensitivity and specificity of each method. Vet. Microbiol. 2000, **77**, 387-398.
28. **Yokomizo, Y., Kishima, M., Mori, Y. and Nishimori, K.** Evaluation of enzyme-linked immunosorbent assay in comparison with complement fixation test for the diagnosis of subclinical paratuberculosis in cattle. J. Vet. Med. Sci. 1991, **53**, 577-584.