Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Preventive Veterinary Medicine 75 (2006) 189–205
PREVENTIVE VETERINARY MEDICINE
www.elsevier.com/locate/prevetmed

Likelihood ratio estimation without a gold standard: A case study evaluating a brucellosis c-ELISA in cattle and water buffalo of Trinidad

G.T. Fosgate [a,*], A.A. Adesiyun [b], D.W. Hird [c], S.K. Hietala [d]

[a] *Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, United States*
[b] *Faculty of Medical Sciences, School of Veterinary Medicine, University of the West Indies, Champs Fleurs, Trinidad and Tobago*
[c] *Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, CA 95616, United States*
[d] *California Animal Health and Food Safety Laboratory, Davis, CA 95616, United States*

Received 17 April 2005; received in revised form 25 November 2005; accepted 27 February 2006

Abstract

The likelihood ratio (LR) is a measure of association that quantifies how many more times likely a particular test result is from an infected animal compared to one that is uninfected. They are ratios of conditional probabilities and cannot be interpreted at the individual animal level without information concerning pretest probabilities. Their usefulness is that they can be used to update the prior belief that the individual has the outcome of interest through a modification of Bayes' theorem. Bayesian analytic techniques can be used for the evaluation of diagnostic tests and estimation of LRs when information concerning a gold standard is not available. As an example, these techniques were applied to the estimation of LRs for a competitive ELISA (c-ELISA) for diagnosis of *Brucella abortus* infection in cattle and water buffalo in Trinidad.

Sera from four herds of cattle ($n = 391$) and four herds of water buffalo ($n = 381$) in Trinidad were evaluated for *Brucella*-specific antibodies using a c-ELISA. On the basis of previous serologic (agglutination) test results in the same animals, iterative simulation modeling was used to classify animals as positive or negative for *Brucella* infection. LRs were calculated for six categories of the c-ELISA proportion inhibition (PI) results pooled for cattle and water buffalo and yielded the following

* Corresponding author. Tel.: +1 979 845 3203; fax: +1 979 847 8981.
E-mail address: gfosgate@cvm.tamu.edu (G.T. Fosgate).

0167-5877/$ – see front matter
doi:10.1016/j.prevetmed.2006.02.007

estimations (95% probability intervals): <0.10 PI, 0.05 (0–0.13); 0.10–0.249 PI, 0.11 (0.04–0.20); 0.25–0.349 PI, 0.77 (0.23–1.63); 0.35–0.499 PI, 3.22 (1.39–6.84); 0.50–0.749 PI, 17.9 (6.39–77.4); ≥0.75 PI, 423 (129–∞). LRs are important for calculation of post-test probabilities and maintaining the quantitative nature of diagnostic test results.

Keywords: Likelihood ratio; LR; No gold standard; Bayesian; Brucellosis; ELISA; Trinidad; Cattle; Water buffalo

1. Introduction

Brucellosis has recently been recognized as a problem of cattle and domestic water buffalo of Trinidad (Fosgate et al., 2002a) and the identification of this important zoonotic disease necessitated implementation of an appropriate control program. The buffered plate agglutination test (BPAT) was chosen to screen individual domestic animals and a brucellosis competitive ELISA (c-ELISA) was used to confirm positive reactions on the BPAT (Fosgate et al., 2003).

The sensitivity (Se) and specificity (Sp) of the BPAT and c-ELISA were not determined for domestic water buffalo until recently (Fosgate et al., 2002b, 2003). The evaluation of these tests was performed without knowledge of the true *Brucella* infection status because of the study design and the fact that a procedure to determine this true status without error (gold standard) is not readily available (Cheville et al., 1998). Recovery of *Brucella* organisms from an animal is an unerring method to confirm infection, but bacterial culture results are often negative for infected animals (Rao, 1970), making it a poor criterion for uninfected animals. Animals with positive bacterial culture results for *Brucella* organisms, however, might not adequately represent the population in which the diagnostic test will be used because acutely infected animals are more likely to have positive culture results than chronically infected animals (Cheville et al., 1998). Results of serologic tests might also vary according to stage of infection, causing incorrect Se estimates. If the successful isolation of *Brucella* organisms is used as the gold standard method, incorrect Sp estimates might occur because of the misclassification of some infected animals as uninfected (erroneously increases denominator and maybe numerator of Sp estimate). Another limitation of bacterial culture is that it necessitates lymph node biopsy or the collection of specimens at slaughter in animals that are not lactating. Culling of animals as a method of diagnosis might not be appropriate or feasible under all economic and agricultural conditions.

Some of the problems associated with the lack of a perfect reference test for brucellosis have been addressed through development of statistical methods for the evaluation of diagnostic tests in the absence of a gold standard (Hui and Walter, 1980; Vacek, 1985; Joseph et al., 1995; Enøe et al., 2000; Dendukuri and Joseph, 2001). The principal benefit of these statistical models is the reduction in selection bias that might develop in usual test evaluation studies when there is difficulty in determining gold standard status. Selection bias, by definition, occurs when the studied subjects do not accurately represent the true population of interest (Rothman and Greenland, 1998, pp.

119–120). For example, selection bias might develop when the gold standard negative definition is animals residing in herds known to be uninfected because of repeated serologic testing. A gold standard positive definition based on the successful isolation of an organism in clinically affected patients might also be a selection bias if the purpose of the test would be the screening of individuals without clinical signs. Diagnostic test evaluations that do not rely on the assumption of a perfect reference test are most appropriate in situations when the true status of individuals cannot be determined without the introduction of selection bias.

Bayesian analytic techniques have been used for the evaluation of diagnostic tests when information concerning a gold standard is not available (Joseph et al., 1995; Enøe et al., 2000; Dendukuri and Joseph, 2001). Bayesian methods are a formal process to statistically incorporate prior information with the collected data to create a single summary value for the measurement of interest. These methods rely on the premise that prior knowledge about test characteristics (Se and Sp) and infection prevalence (Pr) is available before the initiation of the current experiment or data collection process. The prior knowledge, which can be obtained from previously collected data or expert opinion, is then used as a starting point that is subsequently modified, or updated, using the newly collected data. The updated results, or posterior summary information, are obtained through use of Bayes' theorem (Bayes, 1763), which is the defining mathematical component of a Bayesian analysis. Bayesian methods have been previously used to evaluate the BPAT and c-ELISA using no gold standard methods (Fosgate et al., 2002b, 2003).

The c-ELISA is a quantitative test and the optical density (OD) is usually converted to a measurement of inhibition expressed by the test sample compared to a conjugate-only control free of competing antibodies (Nielsen et al., 1994). Quantitative tests are often dichotomized into positive and negative results to aid in diagnostic interpretation; however, this neglects the magnitude of the test result. Different levels of test results in a quantitative assay usually lead to different likelihoods of having the outcome of interest (e.g. infection). The effect of the magnitude of a result can be evaluated through the plotting of receiver-operating characteristic (ROC) curves (Greiner et al., 2000) or the calculation of likelihood ratios (LRs) (Dujardin et al., 1994).

The LR is a measure of association similar to an odds ratio that quantifies how many more times likely the test result is from an infected animal compared to one that is uninfected. It is calculated as the ratio of the probability that an infected individual will have that test result to the probability that an uninfected individual would have that same result (Simel et al., 1993). This is a general definition that includes the calculation of likelihood ratio positive (LR+) and likelihood ratio negative (LR−), which are often defined by formulas involving Se and Sp (Simel et al., 1991; Giard and Hermans, 1996; Joseph and Gyorkos, 1996; Lee, 1999; Collins, 2002; Johnson, 2004). Calculation of LRs for tests with quantitative outcomes (e.g. titers and optical densities) necessitates creating result categories that ensure adequate numbers for reliable estimation and renders their computation no longer easily defined by Se and Sp. Likelihood ratios are also mathematically related to ROC curves as the slope of the non-parametric ROC curve connecting points corresponding to adjacent test result categories (Choi, 1998).

The usefulness of the LR is that it can be used to update the prior belief that the individual has the outcome of interest (Simel et al., 1993; Giard and Hermans, 1996; Collins, 2002) through a modification of Bayes' theorem. The adjustment of prior probabilities has important clinical implications for diagnosis of infection (Connell and Koepsell, 1985) and the calculation of LRs avoids the limitation of defining a single arbitrary cutoff and discarding the quantitative nature of test results. Individual LRs must be estimated through an appropriate epidemiologic study design and often requires information concerning the true infection status of evaluated animals. However, newer statistical methods do not require this gold standard information and these have been employed for estimating LRs from dichotomized test results (Joseph and Gyorkos, 1996). The objective of this study was to estimate LRs for multiple cutoffs without a gold standard using as an example a brucellosis c-ELISA to screen for *Brucella abortus* infection in cattle and domestic water buffalo of Trinidad.

2. Materials and methods

2.1. Data collection

Animal selection has been described in a previous report (Fosgate et al., 2002b). Briefly, eight herds of domestic livestock in Trinidad were selected for evaluation of brucellosis serologic tests during 1998. Two herds of domestic water buffalo (*Bubalus bubalis*), one herd of *Bos indicus* cattle, and one herd, a mixture of *Bos taurus* and *B. indicus* cattle were considered to be affected with brucellosis because of the results of previous serologic testing performed by the government of Trinidad and Tobago. Two herds each of cattle and water buffalo that were believed to be free of brucellosis based on lack of reported abortions in historically closed herds were also selected for evaluation. The two cattle herds believed to be free of brucellosis were also serologically tested for *Brucella* infection by the government prior to sample collection for this study.

The selection of animals for study, collection of specimens for testing and descriptions of all brucellosis testing protocols have been described elsewhere (Fosgate et al., 2002b, 2003). Briefly, blood samples were obtained from animals on selected farms during 1999 in conjunction with the Trinidad and Tobago government brucellosis-screening program. A minimum of 100 animals over 1 year of age, or the entire herd if there were fewer than 100 qualifying animals, were sampled from each premises. Serologic testing consisted of four traditional agglutination tests (standard plate agglutination test, SPAT; card test; buffered plate agglutination test; standard tube agglutination test, STAT) and a brucellosis c-ELISA. For the c-ELISA, test sera and controls were run in duplicate on each 96-well microtiter plate and the proportion inhibition (PI) was calculated as follows:

$$\mathrm{PI} = \frac{\text{mean OD of conjugate control} - \text{mean OD of test sera}}{\text{mean OD of conjugate control}}$$

where OD stands for optical density measured at a wavelength of 410 nm.

2.2. No gold standard methods introduction

The basic diagnostic test model will be described for a simple situation involving only two tests, without a gold standard. The actual model that forms the basis of this present paper incorporates the results from four diagnostic tests (SPAT, card, BPAT and STAT) as discussed in a previous report (Fosgate et al., 2003). The two test model can be extended to incorporate results of four tests by modifying the mathematical equations, but the underlying theory is the same irrespective of the number of tests involved.

Lack of knowledge concerning the true status of evaluated animals (no gold standard) creates a situation where Se and Sp of diagnostic tests cannot be estimated directly. Available information includes the number of animals with each of the four possible positive (+) and negative (−) test result patterns as follows: +Test 1, +Test 2; +Test 1, −Test 2; −Test 1, +Test 2; −Test 1, −Test 2. The total count for each test pattern is the sum of the unobserved (latent) number of infected and uninfected animals for that category. The probability of belonging to a particular test result category is calculated by summing the two possible situations corresponding to infected and uninfected animals (Fig. 1).

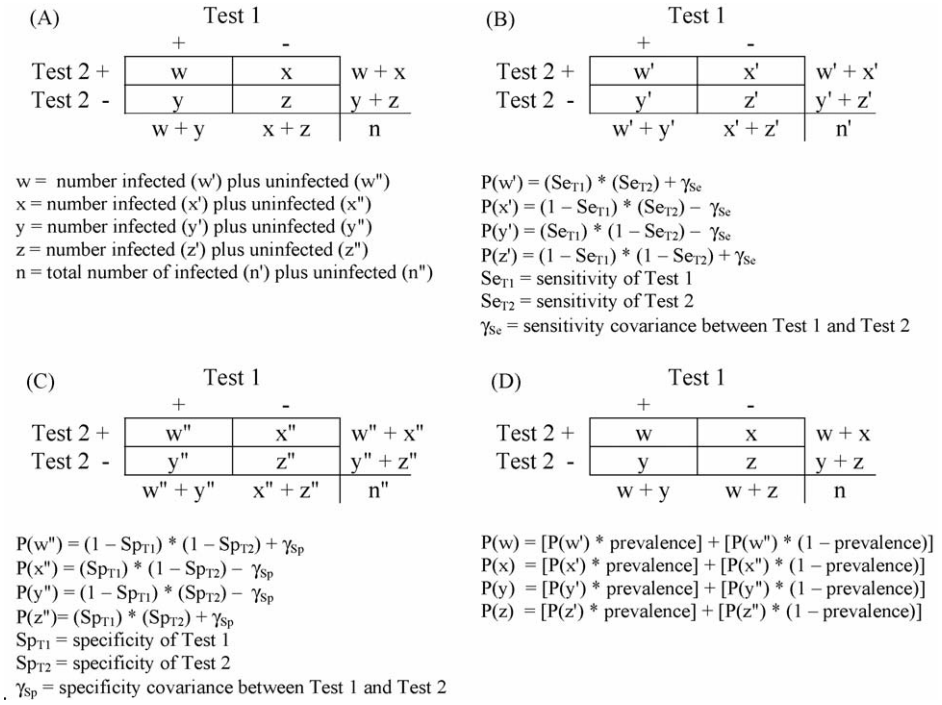

Fig. 1. Description of a statistical model (Fosgate et al., 2002b) to estimate sensitivity and specificity of two diagnostic tests in the absence of a gold standard. The steps involved with determining the probability associated with each cell of the multinomial model include: step 1, observed numbers of animals cross-classified by diagnostic test results are the sum of the unobserved (latent) numbers of infected and uninfected animals (A); step 2, probability statements for infected animals (B); step 3, probability statements for uninfected animals (C); step 4, probability statements for observed data are the weighted sum of probabilities for infected and uninfected animals (D).

For a single animal population, the observed 2×2 table of total counts (cross-classified test results) follows a multinomial distribution with unknown probabilities for belonging in each of the test result categories. The probability of each 2×2 cell (e.g. positive on both tests), although unknown, is dependent on the infection prevalence and diagnostic test parameters (two sensitivities, two specificities and covariance terms). Probability for any given cell is a weighted average with contributions corresponding to infected animals (weighted by prevalence) and uninfected animals (weighted by $1 -$ prevalence).

Tests that measure similar analytes (e.g. antibodies) might not yield independent results and there might be little gain in diagnostic certainty using multiple dependent tests when compared to a single test. Specifically, this is termed conditional dependence and can lead to incorrect Se and Sp estimates if this possibility is ignored in the statistical models. Covariance terms can be added to the probability statements for each diagnostic test result category to investigate and subsequently adjust for conditional dependence. A model that does not include covariance terms assumes conditional independence among diagnostic tests. Separate covariance terms are necessary for infected animals (Se covariance) and uninfected animals (Sp covariance) (Vacek, 1985; Gardner et al., 2000). Sensitivity covariance is the probability observed when test results agree in infected animals above what would be expected if the tests were conditionally independent and is calculated as the difference between the probability of both tests having positive results for an infected animal and the product of the corresponding test sensitivities. Specificity covariance is similarly defined for uninfected animals on the basis of negative results on both tests. The range of possible values for these covariance terms is restricted because all test result probabilities must be between 0 and 1. The limits are based on Se and Sp of the tests (Vacek, 1985; Gardner et al., 2000) and appropriate constraints can be added into the statistical models.

2.3. General Bayesian methods

The collection of variables (Se, Sp and Pr) to be estimated by the study can be denoted by the Greek letter θ. Prior scientific information about θ can be incorporated by using a statistical distribution (density function), $P(\theta)$ that represents expert opinion or results of prior studies as discussed previously. The posterior knowledge concerning the variables of interest is a conditional probability that is obtained by using Bayes' theorem (Bayes, 1763) as:

$$P(\theta|\text{data}) = \frac{P(\text{data}|\theta) \times P(\theta)}{P(\text{data})}$$

where $P(\text{data}|\theta)$ is the likelihood function, $L(\theta)$ and $P(\text{data})$ is the probability of the data.

The denominator is constant for the analysis and is only important to ensure that the total probability will add to 1. Therefore, the formula can be rewritten as:

$$P(\theta|\text{data}) \propto P(\text{data}|\theta) \times P(\theta)$$

The posterior knowledge (represented by a statistical density function), therefore, is proportional to the product of the likelihood of the data, $L(\theta)$ and the prior scientific information, $P(\theta)$. The mean or median of the posterior density can be used as a point

estimate and corresponding intervals can be generated to reflect variability around this estimate. A probability interval is obtained by finding the lower and upper values where the area under the graph of $P(\theta|\text{data})$ is equal to 0.95, or any other preselected probability level. A probability interval, sometimes also referred to as a credibility interval, serves the same function as a confidence interval as to report the variability inherent in the point estimates. The true value is interpreted as being between the upper and lower bounds with 95% probability based on the investigator's prior knowledge combined with the collected data. These calculations can be performed using calculus when $P(\theta|\text{data})$ corresponds to a normal distribution or another recognizable density function.

Computer simulations can be used to solve complicated statistical problems, including situations with unrecognizable posterior densities for which analytic calculus is not possible. Markov chain Monte Carlo (MCMC) methods have been developed to solve problems when calculus is not possible (Gilks et al., 1998; Carlin and Louis, 2000). The MCMC procedure selects data points iteratively and values for each variable of interest (iterates) are dependent on previous selections, thus creating a chain of values. These MCMC values are considered a random sample from the true distribution of the variables of interest (e.g. Se, Sp and Pr). When a large sample is obtained, medians and percentiles of the MCMC iterates can serve as point estimates and probability intervals for the variables, respectively. A typical MCMC procedure will sample several thousand points to form an estimate of the true distribution of the variable of interest.

2.4. Likelihood ratio definition

A likelihood by the usual definition is the probability of an event occurring. A likelihood ratio is simply the ratio of two likelihoods that are formed in the same manner and assumed to have the same proportional relationship to the true probability (Berry and Lindgren, 1996). A ratio of probabilities could be considered a LR with the proportionality constant equal to 1. Likelihood ratios are commonly used for statistical inference (Rothman and Greenland, 1998, pp. 218–220) and are often based on maximum likelihood algorithms (Kleinbaum, 1998).

Likelihood ratios for diagnostic tests meet the previously mentioned criteria and are related to Bayes' theorem. They are primarily used to update the prior odds of the outcome (e.g. infection) conditional on the test results to form post-test odds. The general Bayes' theorem can be modified for the situation of diagnostic testing and is written as follows:

$$P(\text{I}|\text{TR}) = \frac{P(\text{TR}|\text{I}) \times P(\text{I})}{P(\text{TR})} \quad (1)$$

where I = infected and TR = particular test result.

The above formula (1) states that the conditional probability an individual is infected given the test result is equal to the probability of that test result given infection (diagnostic Se for dichotomized test results) multiplied by the probability of infection (Pr) divided by the probability of that particular test result in the source population. This equation can then be modified slightly so that it becomes clear that the post-test probability of infection is equal to the prior probability (i.e. Pr) multiplied by an updating factor. The updating factor

is the amount of added information provided by the diagnostic test result and the modified formula is:

$$P(\mathrm{I}|\mathrm{TR}) = \frac{P(\mathrm{TR}|\mathrm{I})}{P(\mathrm{TR})} \times P(\mathrm{I}) \tag{2}$$

The probability of the test result in the source population is often unknown and it is therefore not possible to use the above formula (2) for updating the probabilities for an individual animal.

The general Bayes' theorem can also be applied for calculation of the post-test probability that the individual is not infected given the particular test result. This formula is:

$$P(\text{not-I}|\mathrm{TR}) = \frac{P(\mathrm{TR}|\text{not-I})}{P(\mathrm{TR})} \times P(\text{not-I}) \tag{3}$$

The above formulas (2) and (3) can be divided and the probability of the test result will then cancel out from the equation removing the limitation caused by the unknown probability of the test result in the source population. The equation can be considered the odds version of Bayes' theorem (Gelman et al., 1997) and the first term to the right of the equality is the usual mathematical definition of the LR. This equation is included below and is interpreted as the posterior odds being equal to the product of the LR and the prior odds (Simel et al., 1993).

$$\frac{P(\mathrm{I}|\mathrm{TR})}{P(\text{not-I}|\mathrm{TR})} = \frac{P(\mathrm{TR}|\mathrm{I})}{P(\mathrm{TR}|\text{not-I})} \times \frac{P(\mathrm{I})}{P(\text{not-I})} \tag{4}$$

2.5. Likelihood ratio calculation

The estimation of Se and Sp for the four agglutination tests generates the data necessary to calculate predictive value positive (PVP) information for all possible agglutination test result patterns (probability of infection conditional on specific test results). The probability of an animal having a specified agglutination test result pattern (e.g. positive results on all four tests) is equal to the probability of having an infected animal with that pattern plus the probability of having an uninfected animal with that same test pattern. In general, the PVP can then be defined as the conditional probability that an animal has the infection given that it has a particular test result pattern and is calculated as the proportion of the total probability contributed by infected animals. The PVP can be calculated from Se, Sp and Pr information and can be determined for all possible test results. The situation involving two tests (T1 and T2) both with positive results is calculated by using the following equation:

$$\mathrm{PVP} = \frac{\mathrm{Se_{T1}} \times \mathrm{Se_{T2}} \times \mathrm{Pr}}{[\mathrm{Se_{T1}} \times \mathrm{Se_{T2}} \times \mathrm{Pr}] + [(1 - \mathrm{Sp_{T1}}) \times (1 - \mathrm{Sp_{T2}}) \times (1 - \mathrm{Pr})]}$$

The numerator is the probability that the animal (or group of animals with same test results) was truly infected and the denominator is the sum of the probabilities that the animal was infected and uninfected. This equation also assumes conditional independence of test

results, but it could be modified to include covariance terms. Predictive value positive was modeled for the agglutination test results of all animals under study and used for the simulation of infection status, which was necessary for the subsequent evaluation of the c-ELISA as discussed in a previous report (Fosgate et al., 2003).

The modeling of the predictive value positive functions was performed without knowledge of the true brucellosis status of the animals in the study. The PVP was then used as a replacement for the knowledge of the true infection status of the animals. For example, if we know that the PVP = 0.5, despite not knowing the infection status of each individual animal, we do know that 50% of the animals with those test results will be infected. The PVP can be iteratively sampled to probabilistically assign infection status to all animals under study.

The c-ELISA results were combined for both cattle and water buffalo in a single analysis because of previous findings that suggested equal accuracy of this test in both species (Fosgate et al., 2003). The test results were divided into six categories using cutoffs believed to be diagnostically important while maintaining adequate sizes of observations in each category. Evaluated cutoffs were <0.10, 0.10–0.249, 0.25–0.349, 0.35–0.499, 0.50–0.749 and ≥ 0.75 proportions inhibition.

Simulation methods based on the agglutination test-based PVP functions were used to assign infection status to each animal under study by use of MCMC procedures. The MCMC techniques were used with available software (WinBUGS Version 1.4, MRC Biostatistics Unit, Cambridge, UK; code available upon request to Dr. Fosgate) to assign infection status and iteratively estimate all LRs of interest. On the basis of the simulated infection status at each iteration, LRs for the diagnosis of *Brucella* infection using the c-ELISA were calculated for the six selected categories. Probabilistic assignment of infection status based on the PVP does not account for variability associated with c-ELISA values for animals under study and therefore a re-sampling approach, similar to the use of a bootstrap method of confidence interval estimation, was employed to account for this variability. This procedure selects a random sample of the original c-ELISA values (i.e. pseudosample), estimates PVP functions and assigns infection status for each animal at each iteration of the MCMC procedure.

The probabilistically assigned infection status was treated as if this were the true status for calculation of the LR at each iteration. The LR for each category was then calculated as the proportion of simulated infected animals within that test result category divided by the proportion of simulated uninfected animals with that same result category. This method of calculation can cause the numerator or denominator of this ratio to be equal to 0 if there are no simulated infected animals (numerator equal to 0) or uninfected animals (denominator equal to 0). A category that did not have any simulated uninfected animals at a particular iteration was adjusted to have one uninfected animal to prevent the denominator of the LR from being equal to 0. In this manner LRs calculated when the numerator or denominator would have been equal to 0 were maintained in the analysis. In these instances the LRs were not expected to be accurate, but were retained to represent LRs that would be either very small or very large.

Plots of the estimated LRs were monitored for trends in successive iterations to determine when convergence was achieved. Convergence occurs when LR values (iterates) no longer have a discernible trend and appear to be fluctuating randomly. Convergence was assessed by calculating the Gelman–Rubin statistic (WinBUGS Version 1.4, MRC

Biostatistics Unit). Values obtained prior to reaching convergence, termed the burn-in phase of the analysis, were not used for making inferences. The complex nature of the model necessitated discarding the first 200,000 iterations as the burn-in phase and inferences were based on the subsequent 400,000 iterations. Median values and percentiles were used as point estimates and probability intervals, respectively, of the LRs for the c-ELISA.

2.6. Use of likelihood ratios

The theoretical benefit of the c-ELISA for diagnosis of *Brucella* infection in cattle and water buffalo of Trinidad was evaluated by using the odds version of Bayes' theorem (4). In this manner the prior odds of infection was multiplied by the point estimate of the LR (median value of posterior distribution) to form the posterior odds of infection. This evaluation was performed over all possible prior probabilities (i.e. Pr) from 0 to 1 at 0.01 intervals and converted to prior odds. Odds can be calculated from probabilities by the formula, odds = probability/(1 − probability). Posterior odds were then converted back to probabilities by using the relationship, probability = odds/(odds + 1).

3. Results

Sera were evaluated for *Brucella*-specific antibodies from 391 cattle and 381 domestic water buffalo from four herds of each species. c-ELISA values in the evaluated water buffalo tended to be higher than the cattle sampled as part of this study (Fig. 2). A PVP cutoff of 0.5 was used to separate the animals in the study as being likely infected or uninfected. This classification was used for descriptive purposes only and the distribution of c-ELISA categories is presented in Fig. 3. The values tended towards extremes of the distribution with fewer observations within middle categories.

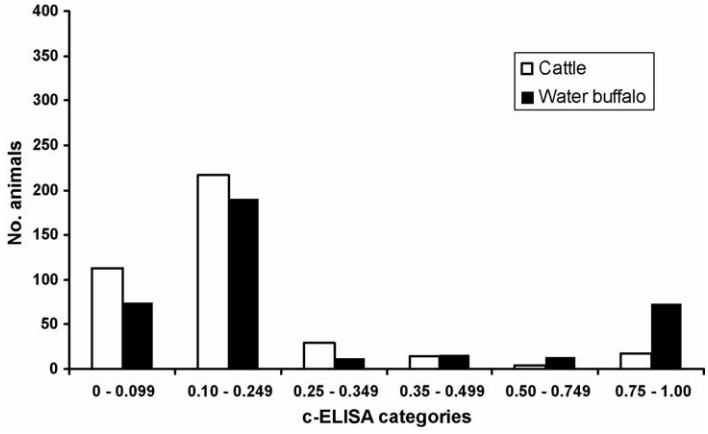

Fig. 2. Distribution of results for six categories of a competitive ELISA used to diagnose *Brucella* infections in 391 cattle and 381 domestic water buffalo of Trinidad.

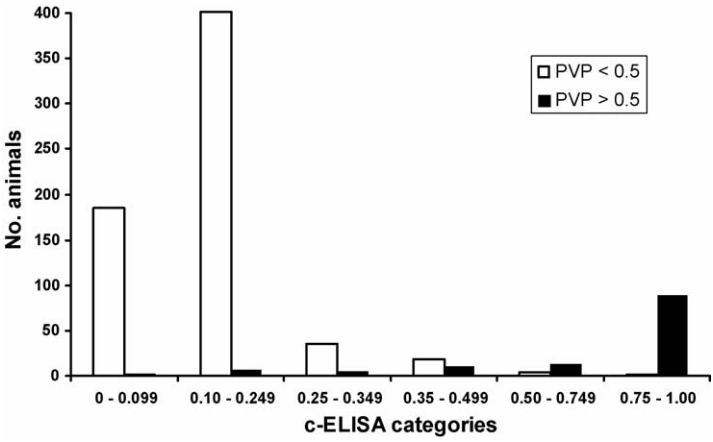

Fig. 3. Distribution of results for six categories of a competitive ELISA used to diagnose *Brucella* infections in 391 cattle and 381 domestic water buffalo of Trinidad. Animals were classified by a predictive value positive (PVP) function on the basis of brucellosis agglutination test results and Markov chain Monte Carlo simulations.

The MCMC procedure simulated the infection status of animals that allowed for estimation of LRs for the six c-ELISA categories. The estimated LRs formed in this manner were less than one for the three lowest result categories of the c-ELISA and greater than one for the others (Table 1). Upper and lower bounds for intervals were not estimable for the highest and lowest categories, respectively, because a relatively large proportion of iterations in these categories simulated zero infected or uninfected animals. The posterior distributions of the LRs were relatively smooth and Fig. 4 demonstrates this for the 0.50–0.749 result category. The histograms formed by results of all iterations for the number of simulated uninfected (Fig. 5) and infected (Fig. 6) animals in this result category were relatively symmetrical. The MCMC procedure was determined to have converged based on the Gelman–Rubin statistic demonstrating convergence of the pooled and within results and their ratio being equal to 1.

The relative usefulness of the c-ELISA for the diagnosis of *Brucella* infection in cattle and water buffalo was determined and the result category of $\geq$0.75 proportion inhibition

Table 1
Likelihood ratio (LR) results and probability intervals for evaluation of a brucellosis c-ELISA in cattle and water buffalo of Trinidad

Result category (PI)	No. infected[a]	No. uninfected[a]	LR[a]	95% interval[b]
<0.10	2	220	0.05	0–0.13
0.10–0.249	8	362	0.11	0.04–0.20
0.25–0.349	6	37	0.77	0.23–1.63
0.35–0.499	11	17	3.22	1.39–6.84
0.50–0.749	13	4	17.9	6.39–77.4
≥0.75	89	1	423	129–∞

PI, proportion inhibition; LR, likelihood ratio.
[a] Median value of MCMC posterior distribution.
[b] Interval formed as the 2.5 and 97.5 percentiles of MCMC posterior distribution.

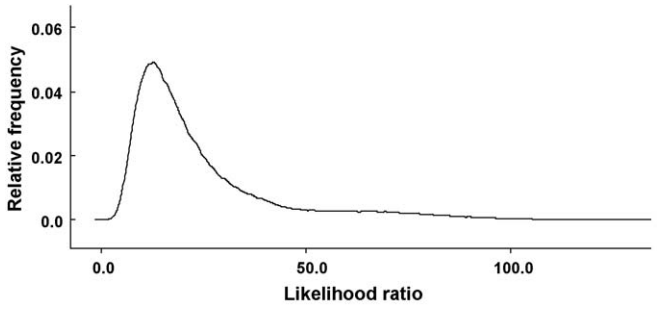

Fig. 4. Posterior distribution of the likelihood ratio for the 0.50–0.749 proportion inhibition category of a competitive ELISA used to diagnose *Brucella* infections in 391 cattle and 381 domestic water buffalo of Trinidad. Results based on MCMC simulation and evaluation without knowledge of true infection status.

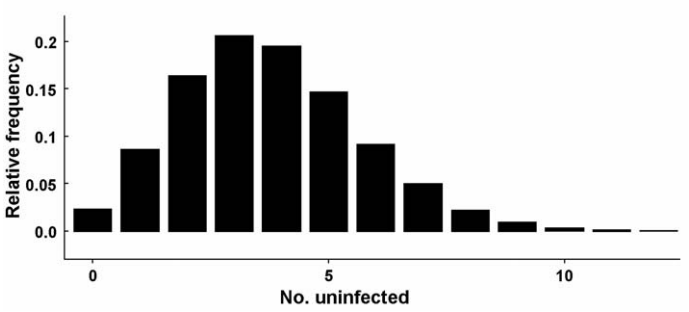

Fig. 5. Posterior distribution of the number of simulated uninfected animals for the 0.50–0.749 proportion inhibition category of a competitive ELISA used to diagnose *Brucella* infections in 391 cattle and 381 domestic water buffalo of Trinidad. Results based on MCMC simulation and evaluation without knowledge of true infection status.

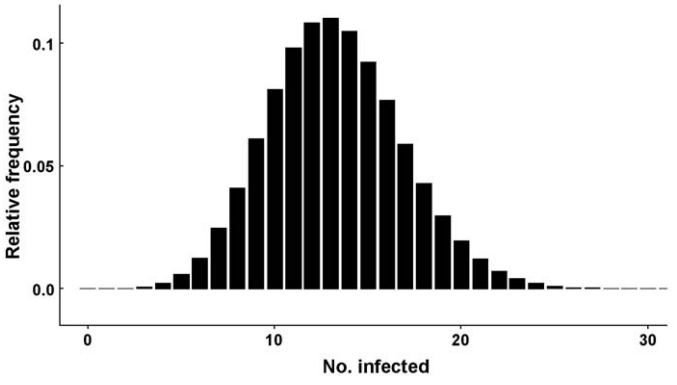

Fig. 6. Posterior distribution of the number of simulated infected animals for the 0.50–0.749 proportion inhibition category of a competitive ELISA used to diagnose *Brucella* infections in 391 cattle and 381 domestic water buffalo of Trinidad. Results based on MCMC simulation and evaluation without knowledge of true infection status.

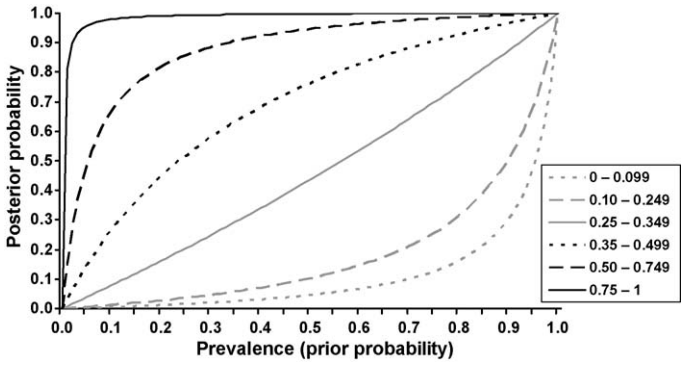

Fig. 7. Usefulness of a competitive ELISA for modifying the prior probability of *Brucella* infection for six evaluated test result categories used in cattle and domestic water buffalo of Trinidad. Likelihood ratio results based on MCMC simulation and evaluation without knowledge of true infection status.

had the greatest effect on converting a low prior probability of infection into a higher probability (Fig. 7). The result category of 0.25–0.349 had the least amount of influence on updating the prior belief. The lowest result category, <0.10, had a strong effect of decreasing the prior probability up until relatively high values of Pr.

4. Discussion

Likelihood ratios can be used to update prior probability of infection (or other outcome) through use of the odds form of Bayes' theorem. In this manner the quantitative nature of the diagnostic test can be maintained rather than dichotomizing results based on an arbitrarily defined positive threshold. Test results are often reported as categories, such as negative, suspect, low positive and strong positive. The assumption is that animals with strong positive results are expected to have a higher probability of infection compared to all other categories, but numerical values associated with this increased likelihood are seldom used for management decisions. Reporting LRs for each test result category would provide more information that could be used to calculate probability of infection when combined with information related to the prior probability in the particular animal being tested.

The LR is interpreted as the number of times it is more likely for an animal with that test result to be infected. It is therefore conditional on the test result and does not necessarily apply to a particular animal within a herd. For example, the LR for a cow with a c-ELISA proportion inhibition of 0.80 was estimated to be 423, but this does not mean that this cow is 423 times more likely to be infected with *Brucella*. What it does mean is that 423 times more animals within that test result category are infected. Information concerning the average prevalence on the farm would be necessary to use the LR to calculate how many more times it is likely that the individual cow is infected. For a herd-level prevalence of 0.01 assigned as the prior probability of infection in the tested cow, then Bayes' theorem can be used to compare the cow's probability of being infected (probability = 0.81) to not being infected (probability = 0.19) giving a ratio of 4.3. A prior probability of 0.10 would

make this ratio 47 (probability infected = 0.98 and not infected = 0.02). A 0.50 prevalence of infection will cause this post-test probability ratio to be equal to the LR, and higher prevalences give even higher probability ratios. Therefore, it is not appropriate to simply consider the LR as a measure of how many more times it is likely that the animal is infected (compared to not).

The common practice of defining LRs in relationship to diagnostic Se and Sp has caused confusion concerning their calculation and interpretation when used for more than two categories (Collins, 2002, 2003; Naugle et al., 2003). Published literature also includes discussions concerning changing the definition of the LR− (Delgado-Rodriguez et al., 1997; Simel and Holleman, 1997), which would produce an inconsistency with the general definition of a LR being the ratio of the probability of the test result in an affected (e.g. infected) individual compared to that same result in an unaffected. These examples highlight the importance of understanding the concept of the LR and how it is derived from Bayes' theorem.

Bayesian statistical methods have received more attention in recent years as a technique to evaluate diagnostic tests in absence of knowledge concerning the true status (gold standard) of individuals. These procedures have also been applied to the estimation of LRs in the situation of two dichotomous tests (Joseph and Gyorkos, 1996). The estimation of LRs for multiple test result categories in absence of a gold standard can also be performed when information independent of the test being evaluated is provided. The present study used information from four brucellosis agglutination tests to estimate predictive distributions (PVP) to permit estimation of LRs for the c-ELISA. These methods are a straightforward extension of previously described procedures to estimate diagnostic accuracy of tests in the absence of knowledge concerning true infection (or disease) status.

The results obtained from studies incorporating Bayesian and other no gold standard approaches to evaluation of diagnostic tests can be as clinically useful as the usual (gold standard) methods when performed appropriately. The results reported here suggest that the c-ELISA is a good test for ruling in or out brucellosis in cattle and water buffalo of Trinidad when results fall within the high or low extremes of the possible values, respectively. A proportion inhibition of 0.30 is recommended as the positive cutoff for use in cattle (Nielsen et al., 1994) and for these data a cutoff of 0.35 was found to be most effective (Fosgate et al., 2003). The findings of this study suggest that test results falling within the range around 0.30 (0.25–0.349) provide little evidence to change the prior belief concerning infection status of a tested animal. Test categories above 0.35 proportion inhibition, however, have a greater effect on modifying the prior probability of infection suggesting more clinical usefulness. The fact that 0.35 appears to be a better positive threshold based on this study is consistent with the previous report using these same data, and should not be interpreted as independent verification of 0.35 as being better than 0.30.

There are a number of potential problems using the methods described in this study and LRs might not be estimable in all situations. The technique relies on the availability of other information for estimation of predictive distributions (i.e. PVP) and probabilistic assignment of infection status. Without this information a model cannot be built. Model convergence can also be a problem, especially in situations involving complicated likelihood functions. A model that does not converge is unstable and should not be used to make inferences. This will occur if the statistical model does not accurately represent the

true state of nature and suggests that assumptions concerning independence of test results, equal sensitivity and specificity across populations or different infection prevalences among populations have been violated. In such instances it becomes necessary to include or exclude conditional covariance terms and investigate effects of pooling or stratification of data. Lack of convergence can also occur if the data do not contain enough information to estimate all unknown parameters. If the number of degrees of freedom is less than the number of parameters to estimate then the model will not be identifiable using maximum likelihood techniques, which do not incorporate prior information in the model. This problem can be reduced using the described Bayesian approach because inclusion of informative prior probabilities act to ''guide'' the analysis in the correct direction. Use of non-informative, or flat, priors in a Bayesian framework will limit the number of estimable parameters to the number of degrees of freedom. It is not always possible to correct convergence problems and it is extremely important to assess because the validity of model estimates depends upon it.

The reporting of valid estimates and providing post-test probabilities of infection based on LRs and prior probabilities would be an improvement over the simple reporting of results as positive and negative. A range of prior infection probabilities could be used or herd-level disease prevalence could be estimated from the data when a number of animals are tested from the same herd. Another alternative would be the generation of graphs similar to Fig. 7 from which veterinarians and producers could visualize the probability of infection for all result categories and pretest probabilities. The relative accessibility of the internet for retrieval of test results in combination with advances in computer technology would also allow for presentation of results in an interactive manner for submitting veterinarians. In such a situation the veterinarian or farm manager could input their estimate of the prevalence and post-test probabilities of infection would then be produced for each animal in the report based on Bayes' theorem and previously estimated LRs.

The LR can provide clinically useful information and should be promoted as a method of reporting results from quantitative diagnostic tests. Newer methods of estimating LRs should be developed that do not rely on the assumption of a perfect reference test. The procedure reported in this paper is the first step in developing more advanced models that will efficiently evaluate diagnostic tests of interest. More sophisticated techniques might be able to model the LR function as a continuous distribution so that each unique test result will have a different value rather than relying on cruder and arbitrarily chosen categories.

5. Conclusions

Likelihood ratios are important for calculation of post-test probabilities while maintaining the quantitative nature of diagnostic test results. These measures are ratios of conditional probabilities that cannot be interpreted at the individual animal level without information concerning pretest probabilities. It is important to recognize derivation of the likelihood ratio from Bayes' theorem rather than the simplified definitions relating it to Se and Sp. Bayesian statistical techniques can be employed for the estimation of likelihood ratios in the absence of a perfect reference test.

References

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Philos. Trans. R. Soc. Lond. 53, 370–418.

Berry, D.A., Lindgren, B.W., 1996. Statistics: Theory and Methods, second ed. Duxbury Press, Belmont, CA, p. 312.

Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis, second ed. Chapman & Hall/CRC, Boca Raton, FL, pp. 121–191.

Cheville, N.F., McCullough, D.R., Paulson, L.R., 1998. Brucellosis in the Greater Yellowstone Area. National Academy Press, Washington, DC, pp. 1–41.

Choi, B.C., 1998. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. Am. J. Epidemiol. 148, 1127–1132.

Collins, M.T., 2002. Interpretation of a commercial bovine paratuberculosis enzyme-linked immunosorbent assay by using likelihood ratios. Clin. Diagn. Lab. Immunol. 9, 1367–1371.

Collins, M.T., 2003. Calculation method for likelihood ratios dictates interpretation. Clin. Diagn. Lab. Immunol. 10, 729–730.

Connell, F.A., Koepsell, T.D., 1985. Measures of gain in certainty from a diagnostic-test. Am. J. Epidemiol. 121, 744–753.

Delgado-Rodriguez, M., Almaraz, A., Farinas-Alvarez, C., 1997. Should the definition for the negative likelihood ratio be changed? J. Clin. Epidemiol. 50, 641–642.

Dendukuri, N., Joseph, L., 2001. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 57, 158–167.

Dujardin, B., Van den Ende, J., Van Gompel, A., Unger, J.P., Van der Stuyft, P., 1994. Likelihood ratios: a real improvement for clinical decision making? Eur. J. Epidemiol. 10, 29–36.

Enøe, C., Georgiadis, M.P., Johnson, W.O., 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Prev. Vet. Med. 45, 61–81.

Fosgate, G.T., Adesiyun, A.A., Hird, D.W., Hietala, S.K., Ryan, J., 2002a. Isolation of *Brucella abortus* biovar 1 from cattle and water buffaloes on Trinidad. Vet. Rec. 151, 272–273.

Fosgate, G.T., Adesiyun, A.A., Hird, D.W., Johnson, W.O., Hietala, S.K., Schurig, G.G., Ryan, J., 2002b. Comparison of serologic tests for detection of *Brucella* infections in cattle and water buffalo (*Bubalus bubalis*). Am. J. Vet. Res. 63, 1598–1605.

Fosgate, G.T., Adesiyun, A.A., Hird, D.W., Johnson, W.O., Hietala, S.K., Schurig, G.G., Ryan, J., 2003. Estimation of receiver-operating characteristic curves to determine accuracy of a competitive enzyme-linked immunosorbent assay for the serodiagnosis of *Brucella* infection in domestic water buffalo (*Bubalus bubalis*) and cattle. Am. J. Vet. Res. 64, 57–64.

Gardner, I.A., Stryhn, H., Lind, P., Collins, M.T., 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Prev. Vet. Med. 45, 107–122.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 1997. Bayesian Data Analysis, Reprinted 2000 ed. Chapman & Hall, Boca Raton, FL, pp. 8–10.

Giard, R.W.M., Hermans, J., 1996. The diagnostic information of tests for the detection of cancer: the usefulness of the likelihood ratio concept. Eur. J. Cancer 32A, 2042–2048.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1998. Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, Boca Raton, FL.

Greiner, M., Pfeiffer, D., Smith, R.D., 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev. Vet. Med. 45, 23–41.

Hui, S.L., Walter, S.D., 1980. Estimating the error rates of diagnostic tests. Biometrics 36, 167–171.

Johnson, N., 2004. Advantages to transforming the receiver operating characteristic (ROC) curve into likelihood ratio co-ordinates. Stat. Med. 23, 2257–2266.

Joseph, L., Gyorkos, T.W., 1996. Inferences for likelihood ratios in the absence of a ''gold standard''. Med. Decis. Making 16, 412–417.

Joseph, L., Gyorkos, T.W., Coupal, L., 1995. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am. J. Epidemiol. 141, 263–272.

Kleinbaum, D.G., 1998. Applied Regression Analysis and Other Multivariable Methods, third ed. Duxbury Press, Pacific Grove, CA, pp. 649–652.

Lee, W.C., 1999. Selecting diagnostic tests for ruling out or ruling in disease: the use of the Kullback–Leibler distance. Int. J. Epidemiol. 28, 521–525.

Naugle, A.L., Shulaw, W.P., Saville, W.J., Wittum, T.E., Byrum, B., 2003. Calculation method for likelihood ratios dictates interpretation. Clin. Diagn. Lab. Immunol. 10, 729.

Nielsen, K., Kelly, L., Gall, D., Bosse, J., Pulkkinen, W., 1994. A Competitive Enzyme Immunoassay for the Detection of Serum Antibody to *Brucella abortus*, Version 2.0 ed. Animal Diseases Research Institute, Nepean, Ont., Canada.

Rao, C.K., 1970. Epizootiology, diagnosis and control of brucellosis in India. Bull. Off. Int. Epizoot. 73, 5–7.

Rothman, K.J., Greenland, S., 1998. Modern Epidemiology, second ed. Lippincott-Raven, Philadelphia, PA.

Simel, D.L., Holleman, D.R., 1997. Should the definition for the negative likelihood ratio be changed? J. Clin. Epidemiol. 50, 641–642.

Simel, D.L., Samsa, G.P., Matchar, D.B., 1991. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J. Clin. Epidemiol. 44, 763–770.

Simel, D.L., Samsa, G.P., Matchar, D.B., 1993. Likelihood ratios for continuous test results—making the clinicians' job easier or harder? J. Clin. Epidemiol. 46, 85–93.

Vacek, P.M., 1985. The effect of conditional dependence on the evaluation of diagnostic tests. Biometrics 41, 959–968.