



A novel Bayesian Latent Class Model (BLCM) evaluates multiple continuous and binary tests: A case study for *Brucella abortus* in dairy cattle

Yu Wang^{a,*}, Emilie Vallée^a, Chris Compton^a, Cord Heuer^a, Aizhen Guo^b, Youming Wang^c, Zhen Zhang^d, Matthieu Vignes^e

^a EpiCentre, School of Veterinary Science - Tawharau Ora, Massey University, Palmerston North, New Zealand

^b National Key Laboratory of Agricultural Microbiology, Hubei Hongshan Laboratory, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China

^c China Animal Health and Epidemiology Center, Qingdao, China

^d Henan Dairy Herd Improvement Center, Zhengzhou, Henan, 450045, China

^e School of Mathematical and Computational Sciences, Massey University, Palmerston North, New Zealand

ARTICLE INFO

Keywords:

Bovine brucellosis
Bayesian Latent Class Model (BLCM)
Cut-off calibration
Receiver Operating Characteristic (ROC)
Diagnostic performance
Serological tests

ABSTRACT

Bovine brucellosis, primarily caused by *Brucella abortus*, severely affects both animal health and human well-being. Accurate diagnosis is crucial for designing informed control and prevention measures. Lacking a gold standard test makes it challenging to determine optimal cut-off values and evaluate the diagnostic performance of tests. In this study, we developed a novel Bayesian Latent Class Model that integrates both binary and continuous testing outcomes, incorporating additional fixed (parity) and random (farm) effects, to calibrate optimal cut-off values by maximizing Youden Index. We tested 651 serum samples collected from six dairy farms in two regions of Henan Province, China with four serological tests: Rose Bengal Test, Serum Agglutination Test, Fluorescence Polarization Assay, and Competitive Enzyme-Linked Immunosorbent Assay. Our analysis revealed that the optimal cut-off values for FPA and C-ELISA were 94.2 mP and 0.403 PI, respectively. Sensitivity estimates for the four tests ranged from 69.7% to 89.9%, while specificity estimates varied between 97.1% and 99.6%. The true prevalences in the two study regions in Henan province were 4.7% and 30.3%. Parity-specific odds ratios for positive serological status ranged from 1.2 to 2.2 for different parity groups compared to primiparous cows. This approach provides a robust framework for validating diagnostic tests for both continuous and discrete tests in the absence of a gold standard test. Our findings can enhance our ability to design targeted disease detection strategies and implement effective control measures for brucellosis in Chinese dairy farms.

1. Introduction

Bovine brucellosis, primarily caused by *Brucella abortus* (*B. abortus*), is a significant disease of both economic and public health concern globally, especially in the Middle East and parts of Asia, including China (Pappas et al., 2006; Zhang et al., 2018; Wang and Jiang, 2020). In dairy cattle, brucellosis mainly affects sexually mature cows and manifests through late-trimester abortions, infertility, and reduced milk yield (Neta et al., 2010). Alarming findings from recent studies indicate an upward trajectory in the animal-level prevalence of this disease within the Chinese dairy industry (Ran et al., 2019). Such an escalating prevalence potentially poses severe consequences for affected herds and has profound implications for production within the Chinese dairy industry,

ultimately affecting public health in China. Transmission of the disease to humans mainly occurs through consumption of undercooked meats or raw dairy products originating from infected animals (Bercovich, 1998), direct contact with contaminated aborted products and inhalation of contaminated aerosols (Pappas, 2022). Consequently, implementing comprehensive control measures on animal sectors has been highlighted as a valuable and economically rewarding strategy, serving to prevent the dissemination of the pathogen to consumers and the public (Zinsstag et al., 2007).

Diagnostic tests are essential tools for decision-makers to confirm or inform the infection status of individual animals or farms, estimate disease prevalence, demonstrate disease freedom, and conduct epidemiological studies. Direct detection of *B. abortus* can be achieved

* Corresponding author.

E-mail address: y.wang22@massey.ac.nz (Y. Wang).

<https://doi.org/10.1016/j.prevetmed.2024.106115>

Received 8 August 2023; Received in revised form 3 January 2024; Accepted 4 January 2024

Available online 12 January 2024

0167-5877/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

through bacterial culture and biochemical characteristics (e.g., CO₂ requirement and H₂S production) or polymerase-chain-reaction (PCR) tests (WOAH, 2022b). While bacterial culture and biochemistry are considered the ‘gold standard’ for confirming *B. abortus* infection, they are influenced by factors including bacterial viability, tissue concentration, environmental contamination, shedding period, and the need for an accredited biosafety level 3 laboratory (Mantur et al., 2007; Rahman et al., 2019; WOA, 2022b). Furthermore, these methods are also imperfect, expensive and time-consuming, and hence not suitable for routine disease surveillance (Ducrottoy et al., 2018; WOA, 2022b). Indirect testing of anti-*B. abortus* antibodies in serum and milk samples are preferred in routine diagnosis and surveillance (Ducrottoy et al., 2018; Wang et al., 2020; WOA, 2022b). Serological tests are advantageous due to their low cost, rapidity, ease of operation, and high diagnostic sensitivity (*Se*) and specificity (*Sp*). In dairy cattle, detectable levels of IgM antibodies against *B. abortus* are typically reached one-week post-infection, while IgG antibodies become detectable by two weeks post-infection (Baldi et al., 1996) and are considered to persist lifelong (Baldi et al., 1996; Bercovich, 1998). Serological tests prove particularly valuable in diagnosing animals that display clinical signs, such as abortion, as well as in screening apparently healthy individuals. By employing these tests, it is possible to identify exposed animals and conduct surveillance efficiently and effectively.

Serological tests such as the Rose Bengal test (RBT) and standard agglutination test (SAT) are commonly used in epidemiological surveys to assess the seroprevalence of *B. abortus* in dairy cattle (Ran et al., 2019). However, only a few studies have adjusted these estimates to account for the imperfect *Se* and *Sp* of these tests (Greiner and Gardner, 2000a). Although RBT and SAT are considered accurate, their results are acquired by naked-eye observation and are therefore subject to the judgment and experience of operators, particularly for suspicious samples (Ducrottoy et al., 2018; Dong et al., 2021). Fluorescence polarization assay (FPA) and competitive enzyme-linked immunosorbent assay (C-ELISA) have been developed more recently and can offer the advantage of providing quantitative and automated result readings, making them suitable for large-scale investigations. While FPA and C-ELISA kits are commercially available in China, their thresholds, *Se*, and *Sp* have not been extensively calibrated. Many existing diagnostic evaluation studies have used SAT as reference tests to obtain a relative *Se*, *Sp*, and agreement with other tests; however, these studies often failed to account for the imperfect *Se* and *Sp* of the reference test itself (Dong et al., 2021; Yao et al., 2022). Alternatively, some studies have utilized serum samples from naturally infected animals or clinical challenge trials, as well as disease-free populations, but these settings cannot fully replicate natural infection scenarios (Greiner and Gardner, 2000a; b).

Bayesian latent class models (BLCMs) have emerged as a valuable and recommended approach for estimating the accuracy of diagnostic tests when a ‘gold standard’ test is unavailable (Gardner et al., 2019; Cheung et al., 2021; WOA, 2022a). In BLCMs, disease status is treated as unknown or ‘latent’, and *Se* and *Sp* estimates are derived from the cross-classified test results being evaluated in one or more population(s) (Kostoulas et al., 2017; Cheung et al., 2021). Bayesian analysis allows for the incorporation of prior knowledge or existing information about parameters, facilitating the derivation of posterior distributions of *Se* and *Sp* (Kostoulas et al., 2017). Many studies have utilized BLCMs to estimate the diagnostic *Se*, *Sp* of serological tests, and prevalence of *B. abortus* exposure in dairy cattle using dichotomous data (Arif et al., 2018; Rahman et al., 2019; Wang et al., 2020; Zhu et al., 2020). Moreover, the application of modeling continuous test outcome data in veterinary studies has been well-developed, as it avoids the loss of information resulting from dichotomization and enables the estimation of optimal cut-off values and associated *Se* and *Sp* (Choi et al., 2006; Jones et al., 2009; Yang et al., 2022). However, there are no studies combining several continuous and discrete tests for joint analysis in a BLCM. Covariates such as abortion history and age have been extensively

identified to be associated with *B. abortus* infection (Matope et al., 2011), and Jones et al. (2009) demonstrated incorporating covariates could contribute to accurate disease diagnosis.

Our study aimed to develop a novel BLCM for modeling dichotomous and continuous test outcome data to estimate: (1) optimal thresholds of FPA and C-ELISA; (2) *Se* and *Sp* of four commonly used serological tests in China – RBT, SAT, FPA, and C-ELISA; (3) the animal-level prevalence of brucellosis in selected dairy farms; (4) odds ratio for age (parity) on serological status of brucellosis. We followed the Standards for Reporting of Diagnostic accuracy studies that use Bayesian latent class models (STARD-BLCM) guidelines in reporting this study (Kostoulas et al., 2017).

2. Materials and methods

2.1. Study design and samples

The study was conducted in the Pingdingshan region (PR) and Zhengzhou region (ZR) of Henan province, China, as shown in Fig. 1. The geographic coordinates of the study area lie between latitudes 112°14′ and 114°13′E and longitudes 33°07′ and 34°59′N. In both regions, dairy cattle were managed intensively, coupled with crop-based management systems. The Holstein-Friesian breed was exclusively raised in these areas. Previous investigations have reported apparent prevalences of antibodies against *B. abortus* at the cow level as 2.2% (95% CI: 2.0, 2.5) in PR and 7.8% (95% CI: 4.6, 12.2) in ZR (Wan, 2017; Liu et al., 2020).

A cross-sectional study was undertaken between August 2021 and December 2021 in PR and ZR regions of Henan Province. Blood samples were collected as part of routine early pregnancy testing, taken 28 days after artificial insemination, by farmers or their veterinarians. These samples, originally intended for pregnancy diagnosis, were repurposed for this study. Twenty-four dairy herds in two regions were initially enrolled in the selection of potential herds. Herd selection criteria included factors such as the absence of vaccination history, availability of blood samples, and willingness to participate. In total, three dairy herds from each region were selected based on the above criteria, and a random simple sampling approach was then used to obtain at least 100 animals from each farm. Duplicate samples from the same animal were identified and removed by unique earmark identifiers. Due to data availability limitations, there was no pre-selection process for candidate covariates, and only age (parity) information was available and included in the analysis. Blood samples (approximately 5 ml) were collected from each animal using disposable needles and plain blood tubes via coccygeal vein venipuncture. The samples were properly labeled, and transported to the Henan Dairy Herd Improvement (Henan DHI) Centre in Zhengzhou within 24 h in an ice box. The serum samples were separated by centrifugation at 3000 rpm for 10 min and stored at – 20 °C for a maximum period of 12 months.

The sample size calculation was performed using the PropMiscSampleSize software (Version 6.0, available at <http://www.medicine.mcgill.ca/epidemiology/Joseph>), following the detailed procedure outlined in previous reports (Dendukuri et al., 2004; Dendukuri et al., 2010). We adopted the average coverage criterion, a highest posterior density (HPD) length of 0.15 and a 95% HPD coverage, and uniform prior information for prevalence and informative priors for the *Se* and *Sp* of these tests (Table 1). These prior information and assumptions resulted in a minimum sample size of 502 totally from two populations. Due to the limitation of this calculation method to independence between tests and a maximum of three tests (Dendukuri et al., 2004), we manually increased the final sample size ($n = 603$) by 20% to adjust for potential dependencies between tests as the rationales of these tests are based on the detection of antibodies against *Brucella* spp. Considering possible sample loss and insufficient volume during the sampling and testing process, the final number of samples collected was 651.



Fig. 1. Map of Henan Province, China displaying the areas included in this diagnostic performance evaluation study.

2.2. Diagnostic tests and latent condition

2.2.1. Definition of the latent condition

RBT, SAT, FPA, and C-ELISA were performed on each sample. The testing procedures were performed in a blinded manner, with identifiable numbers removed from the samples, and the results were kept confidential until all tests were completed. The rationale of these tests is based on detecting antibodies against *B. abortus*, with RBT, C-ELISA, and FPA tests primarily targeting IgG1 antibodies (along with some IgM and IgA) (Ducrotoy et al., 2016; Rahman et al., 2019). SAT primarily detects IgM antibodies, while also capable of detecting some IgG and IgA antibodies. The latent condition modeled jointly by the four tests is determined by the overlap of the biomarker distributions detected by each test in the host. In our case, the latent condition pertains to *B. abortus* seropositivity, rather than specific antibody type or active infection. It is worth noting that *B. abortus* has been reported to elicit the production of lifelong antibodies in cattle (Bercovich, 1998). Therefore, our hypothesis posits that antibodies will persist throughout the lifetime of cattle after infection with *B. abortus*. However, it is possible that some animals exposed to *B. abortus* may not produce detectable antibody titers, and thus, they might not be captured by this latent condition. These animals constitute a fraction of the exposed population that cannot be accounted for in our analysis. A diagram illustrating the latent condition is presented in Fig. 2, adapted from Ducrotoy et al. (2016); Marquetoux et al. (2023).

2.2.2. Rose Bengal test (RBT)

RBT was performed as recommended by the World Organisation for

Animal Health (WOAH, 2022b). Briefly, all sera, antigen, and positive and negative controls were brought to room temperature at least 30 min in advance. 25 μ l of RBT antigen and the same volume of serum sample were added to a transparent glass slide and mixed thoroughly. The reaction was observed within 4 min post-mixing as positive if agglutination was observed and negative if there was no reaction between serum and antigen. If a doubtful agglutination occurred, the sample was retested. The antigen, positive, and negative controls were purchased from the Harbin Pharmaceutical Group Bio-vaccine Co., Ltd, China.

2.2.3. Standard Agglutination Test (SAT)

SAT was performed following the protocols outlined in the Chinese National Standard Diagnostic Techniques for Animal Brucellosis (GB/T 18646–2018) (Li et al., 2021). The SAT antigen, positive, and negative controls were purchased from the National Animal Brucellosis Reference Laboratory at the China Institute of Veterinary Drug Control. All sera, antigen, and positive and negative controls were brought to room temperature 30 min in advance. Sera were subjected to serial 2-fold dilutions using normal saline containing 0.5% carbolic acid, ranging from 1:25 to 1:200. 100 μ l of diluted serum was thoroughly mixed with 100 μ l SAT antigen solution, resulting in a final serum dilution range of 1:50 to 1:400. The reaction mixture was sealed and placed in a 37 $^{\circ}$ C incubator for 18–24 h. Interpretation of the results followed the criteria established by the Chinese National Standard Diagnostic Techniques for Animal Brucellosis (GB/T 18646–2018) (Li et al., 2021). Sera with endpoint titers of \leq 1:50 were classified as seronegative, while titers of \geq 1:100 were considered positive.

Table 1

Priors used for sensitivity, specificity, and prevalence estimates in the analysis of the sensitivity and specificity of four serological tests for the detection of antibodies against *Brucella abortus* in dairy cattle.

Priors	Descriptions (mode + 95% of certainty of lower limit)	Parameters ^c	References
RBT - Sensitivity	0.812 (0.60)	B(13.58, 3.91)	Gall and Nielsen (2004)
RBT - Specificity	0.863 (0.60)	B(9.92, 2.42)	Gall and Nielsen (2004)
SATc[i] - Sensitivity	0.759 (0.60)	B(21.40, 7.48)	Gall and Nielsen (2004)
SATc[i] - Specificity	0.863 (0.60)	B(9.92, 4.42)	Gall and Nielsen (2004)
PR - Prevalence	0.020 (0.20)	B(1.30, 15.80)	Liu et al. (2020)
ZR - Prevalence	0.080 (0.01)	B(1.32, 3.67)	Wan (2017)
μ _{C-ELISA, D-}	expected range (0.1 to 0.35 PI) ^a	U(-2.3, -0.7)	Kit instruction
μ _{C-ELISA, D+}	expected range (0.4 to 1.0 PI) ^a	U(μ _{C-ELISA, D+} , 0.0)	Kit instruction
τ _{C-ELISA, D-}	Diffuse distribution ^a	G(0.01, 0.01)	
τ _{C-ELISA, D+}	Diffuse distribution ^a	G(0.01, 0.01)	
μ _{FPA, D-}	expected range (75 to 95 mP) ^b	U(-0.3, 0.0)	Kit instruction
μ _{FPA, D+}	expected range (120 to 180 mP) ^b	U(μ _{FPA, D+} , 0.6)	Kit instruction
τ _{FPA, D-}	Diffuse distribution	G(0.01, 0.01)	
τ _{FPA, D+}	Diffuse distribution	G(0.01, 0.01)	
β _{parity}	Diffuse distribution	N(0, 0.01)	
σ _{farm}	Diffuse distribution	U(0, 1)	

^a : log-transformation
^b : log-transformation of 0.01 times mP
^c : B means Beta distribution; U means Uniform distribution; G means Gammar distribution

2.2.4. Fluorescence polarization assay (FPA)

FPA was performed following the instructions provided in the kit and as described elsewhere (Higgins et al., 2017). The FPA kit was purchased from Ellie LLC, USA. Briefly, all reagents and samples were allowed to return to room temperature for at least 30 min. 20 μl of serum sample was added to a glass tube containing 1 ml of dilution solution. The mixture was thoroughly mixed and allowed to stand for 2–5 min at room temperature. An initial reading was obtained on a fluorescence polarisation analyzer (FPM, portable Sentry 300 instrument, Ellie LLC, USA) to measure the light scatter value. Next, 10 μl of antigen labeled with fluorescein isothiocyanate was added to the glass tube and mixed thoroughly. After standing for an additional 2–5 min, a second reading

was taken using the FPM. The final reading, expressed in millipolarisation units (mP), had to fall within the range of 70–95 mP units for negative control and 120–250 mP units for positive control, as specified by the kit instruction. The FPA results were interpreted twice independently, respectively (1) using the kit-provided cut-off value where samples were judged as positive if a mP was ≥ 100 mP and negative if < 100 mP; (2) employing the optimized cut-off value provided by BLCM. This involved subjecting the continuous values (mP) to a subsequent Bayesian Receiver Operating Characteristic (ROC) analysis that maximized the sum of *Se* and *Sp*.

2.2.5. Competitive enzyme-linked immunosorbent assay (C-ELISA)

The C-ELISA was performed with *Brucella* Compac 2.0 C-ELISA kit purchased from INGEZIM, Spain using the procedure described by the manufacturer and other studies (Zeng et al., 2017; Zhu et al., 2020). Percent inhibition (PI) was calculated to measure the antibody titers by using the formula suggested by the manufacturer of the kit:

$$PI = \left[1 - \left(\frac{\text{Optical density value of each test sample}}{\text{Mean optical density value of negative control}} \right) \right] \times 100\%$$

The C-ELISA results were interpreted twice independently, (1) using the kit-provided cut-off value where samples were judged as positive if PIs were ≥ 40%, and negative if PI (%) < 40%; (2) employing the optimized cut-off value provided by BLCM. This involved subjecting the continuous values (PI) to a subsequent Bayesian Receiver Operating Characteristic (ROC) analysis that maximized the sum of *Se* and *Sp*.

2.3. Statistical analyses

2.3.1. Apparent test results

Following the cut-off values provided by kit manufacturers, we first dichotomized the raw test results of FPA and C-ELISA into binary outcomes to calculate the apparent prevalence. Apparent prevalence obtained by each serological test was calculated for each region with the corresponding 95% confidence interval (CI) using Wilson binomial confidence limits. The kappa statistics and its 95% CI were estimated among four serological tests. These calculations were achieved through ‘epiR’ package (version: 2.0.62) (Stevenson et al., 2022).

2.3.2. Bayesian Latent Class Model (BLCM) framework

We adopted a “two binary tests and two continuous tests in two populations” approach with each region, namely PR and ZR, treated as independent populations. Since the detection rationale of four tests is

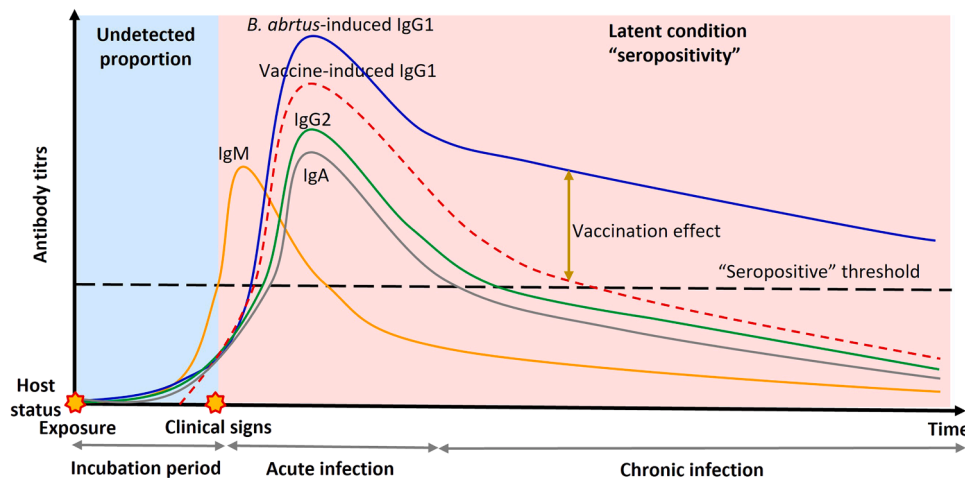


Fig. 2. A diagram illustration of *Brucella abortus* physiopathology and possible dynamics of biomarkers on specific antibodies, with a corresponding latent condition for BLCM.

grounded in the same biological principle of detection of antibodies against *B. abortus* (Fig. 2), correlations between tests exist and need to be accounted for in the diagnostic evaluation (Johnson et al., 2019). Therefore, the four serological tests under evaluation were assumed to be conditionally dependent, given the true disease status.

The outcomes of RBT and SAT were considered dichotomous, whereas the FPA and C-ELISA outcomes were treated as continuous. FPA (mP) results were log-transformed with 0.01 times the original value, and C-ELISA (PI) result values were log-transformed to achieve approximate normality in the analysis and converted back in the presentation of the results (Choi et al., 2006). Bayesian ROC curves were created to estimate the optimal cut-off values by maximizing the Youden Index (sum of Se and Sp) (Youden, 1950). The Bayesian ROC analysis procedure in the absence of a gold standard has been developed and outlined in previous reports (Choi et al., 2006; Jones et al., 2009; Yang et al., 2022).

In BLCM, the true disease status of individual cattle ($z_{[i]}$) was modeled using a Bernoulli sampling with the probability of being seropositive. A hierarchical model structure was adopted to estimate the probability of *B. abortus* seropositivity ($P_{[i]}$) by incorporating fixed effects for region and age, and a random effect for farm (Yang and Laven, 2022). The true disease status ($z_{[i]}$) was then incorporated into Bernoulli sampling for both RBT and SAT results. More specifically, the apparent RBT results depended on Se_{RBT} , Sp_{RBT} , and true disease status ($z_{[i]}$). In the next step, SAT conditional Se and Sp were separately modeled as $Se_{SATc[i]}$ and $Sp_{SATc[i]}$ depending on RBT outcomes ($i = 1$ or 2 , 1 means RBT- or 2 means RBT+), which could be expressed as $Se_{SATc[1]} = Se_{SAT|D+,RBT-}$, $Se_{SATc[2]} = Se_{SAT|D+,RBT+}$, $Sp_{SATc[1]} = Sp_{SAT|D-,RBT-}$, $Sp_{SATc[2]} = Sp_{SAT|D-,RBT+}$. If SAT is independent of RBT, then SAT conditional Se should remain constant regardless of the outcome of RBT; if they are dependent, then the SAT conditional Se should change given the outcome of RBT. Furthermore, Gaussian mixture models were used to model for log-transformed FPA and C-ELISA outcomes, with separate mean (μ) and precision (τ , inverse of the variance = $1/\sigma^2$) parameters depending on the true disease status (z_i) and RBT, and SAT outcomes (Choi et al., 2006; Coupe, 2021; Yang and Laven, 2022). Random correlation coefficients were incorporated to account for the dependence between FPA and C-ELISA. The correlation coefficients between binary tests were calculated based on Se and Sp of RBT and SAT, given the true disease status. The correlation coefficients between binary tests (RBT & SAT) and continuous tests (FPA & C-ELISA) were determined using a mixture mean and variance of continuous tests, along with Se and Sp of binary tests, conditional on the true disease status. This methodology was initially introduced by Coupe (2021). The reproducible R code for the model was provided at <https://github.com/YuWang1009/BLCM.git>.

2.3.3. Definition of prior distributions

The model assuming conditional dependence was over-parameterized because it had more parameters than the number of degrees of freedom in the data, and hence required prior information to be elicited for disease prevalence, Se and Sp of RBT and SAT, and parameters related to Gaussian mixture model to achieve model identifiability (Jones et al., 2010; Cheung et al., 2021). The prior distributions used in the model were outlined in Table 1. To elicit prior information, beta distributions were used for prevalence (π), Se_{RBT} , Sp_{RBT} , $Se_{SAT(RBT-)}$, $Sp_{SAT(RBT-)}$, $Se_{SAT(RBT+)}$, $Sp_{SAT(RBT+)}$. These prior distributions were based on published reports on cattle or previous test results (Gall and Nielsen, 2004; Wan, 2017; Liu et al., 2020), independent of the current study population. For example, the Se estimate for RBT was assumed to be 0.812, with a 95% certainty that the estimate was greater than 0.60, generating a beta distribution Beta(13.58, 3.91). The *epi.betabuster* function from the 'epiR' package (version: 2.0.62) was used to generate the parameters of the beta distribution (Stevenson et al., 2022). Uniform priors were specified for the means of the log-transformed FPA and C-ELISA values for the non-diseased ($\mu_{FPA, D-}$, $\mu_{ELISA, D-}$) and diseased cattle ($\mu_{FPA, D+}$, $\mu_{ELISA, D+}$) based on possible maximum and minimum

limits provided by kit suppliers. For example, the negative control of FPA must read between 75 mP and 95 mP, the corresponding log (0.01 *mP) transformation ranged between -0.29 and -0.05 . Therefore, we set the distribution of log-transformed mean FPA values as $\mu_{FPA, D-} \sim \text{Uniform}(-0.3, 0.0)$ to achieve sufficient diffusion. The prior distributions for parameters of means of log-transformed FPA and C-ELISA, conditioned on the true disease status, were derived in the same manner. It was assumed that diseased animals would exhibit higher mean values for FPA and C-ELISA than healthy animals to ensure model identifiability (Choi et al., 2006). The prior distributions were finally set as follows $\mu_{FPA, D-} \sim \text{Uniform}(-0.3, 0.0)$; $\mu_{FPA, D+} \sim \text{Uniform}(\mu_{FPA, D-}, 0.6)$; $\mu_{ELISA, D-} \sim \text{Uniform}(-2.3, -0.7)$; $\mu_{ELISA, D+} \sim \text{Uniform}(\mu_{ELISA, D-}, 0.3)$. Priors for $\mu_{FPA, D-}$, $\mu_{ELISA, D-}$, $\mu_{FPA, D+}$, and $\mu_{ELISA, D+}$ remained constant regardless of the RBT and SAT outcomes as there was no information about the mean values of FPA and C-ELISA, given RBT and SAT outcomes. Because individual animals were clustered within two regions, we added a farm-level random effect and assumed it to be normally distributed, which can be modeled using a logit-normal distribution with additional covariates (prevalence by region, parity). Diffuse prior distributions, such as $\tau \sim \text{Gamma}(0.01, 0.01)$, $\beta_{\text{parity}} \sim \text{Normal}(0.0, 0.1)$, $\sigma_{\text{farm}} \sim \text{Uniform}(0, 1)$, were specified for the precision of normal distribution of FPA and ELISA results, parity effect and farm-level random effect. The region covariate coefficient parameter was shown as α_i ($i = 1, 2$), which was obtained by logit-transforming region-specific prevalence based on previous reports (Wan, 2017; Liu et al., 2020).

2.3.4. Sensitivity analysis on the prior distribution models

Scenario 1 was presented as the primary model in which prior information was incorporated. Additionally, we changed prior distributions one-by-one to conduct a sensitivity analysis, intending to explore the effects of prior choices on posterior estimates for animal-level prevalence, Se , and Sp . A flat prior (Beta (1, 1)) distribution was used in the sensitivity analysis to replace elicited priors for prevalence, Se and Sp of RBT and SAT, while other parameters (e.g., means and precision for FPA and C-ELISA, and standard deviation for farm random effect) remain constant across the additional seven scenarios given that they are diffuse prior distributions. The specific details of these scenarios can be found in Appendix file 1.

2.3.5. Model implementing and monitoring

Bayesian latent class analyses were implemented using JAGS software version 4.30 (Plummer, 2003) via the R package 'R2jags' (version: 0.7-1) (Su and Yajima, 2021) in R version 4.2.1. Monte Carlo Markov Chain (MCMC) techniques were used to derive the full joint posterior distribution of parameters from prior distributions and test results. Two independent MCMC chains were run in parallel. For each chain, a burn-in period of 5000 iterations was discarded; another 20,000 iterations were used for the posterior distribution. Regarding each parameter, a point estimate was defined as the median of its marginal posterior distribution, and a 95% credible interval was defined by the 2.5 and 97.5 percentiles of this marginal posterior distribution (termed 95% posterior credibility interval, 95% PCI). The convergence was checked by monitoring MCMC chain traces and autocorrelation plots, and computing Gelman and Rubin's statistics (Brooks and Gelman, 1998).

3. Results

3.1. Apparent test results

The apparent test prevalence for each region is presented in Table 2, with values ranging from 0.5% to 1.5% for PR, and 16.6% to 22.7% for ZR. The cross-classified counts, as shown in Appendix file 2 provide the frequency distribution of each of the two regions. Overall, 71 out of 651 samples tested positive on all four tests, 548 tested negative on all four tests, and 32 samples had 'differing' test results. The *kappa* agreement values between RBT & SAT, RBT & FPA, RBT & C-ELISA, SAT & FPA,

Table 2

Apparent test prevalence of detection of antibodies against *Brucella abortus* by each of four tests in PR and ZR regions.

Test	Region	Apparent prevalence (%)	95% Confidence interval
RBT	PR	2.2	1.1, 4.4
	ZR	25.8	21.3, 30.7
SAT	PR	1.2	0.5, 3.2
	ZR	20.6	16.6, 25.3
FPA	PR	2.8	1.5, 5.2
	ZR	27.3	22.7, 32.3
C-ELISA	PR	2.5	1.3, 4.8
	ZR	24.2	19.9, 29.1

SAT & C-ELISA, and FPA & C-ELISA were 0.847 (95% CI: 0.771, 0.923), 0.933 (95% CI: 0.893, 0.972), 0.910 (95% CI: 0.863, 0.957), 0.806 (95% CI: 0.736, 0.875), 0.872 (95% CI: 0.814, 0.930), 0.919 (95% CI: 0.875, 0.963), respectively, indicating there is good agreement among four tests.

3.2. Determination of optimal cut-off values

The ROC curves for C-ELISA and FPA are depicted in Fig. 3. Sensitivities and specificities of the FPA and C-ELISA were calculated at different cut-off values, ranging from 74 mP to 122 mP for FPA and 0.22 PI to 0.60 PI for C-ELISA. FPA test demonstrated a slightly higher Youden index (ranging from 1.85 to 1.88) for cut-off values between 90.5 mP and 100.5 mP. The optimal cut-off value for FPA was 94.2 mP,

resulting in *Se* of 89.9% (95% PCI: 84.3, 94.2) and *Sp* of 98.7% (95% PCI: 98.0, 99.3). At the kit-recommended cut-off, the FPA exhibited a *Se* of 86.1% (95% PCI: 79.1, 91.3) and *Sp* of 99.6% (95% PCI: 99.1, 99.9). C-ELISA test demonstrated a similar Youden index at around 1.78 for cut-off values between 0.37 and 0.44 PI. The optimal cut-off value of 0.403 PI was determined to maximize the Youden index (1.78, 95% PCI: 1.70, 1.85) for C-ELISA, resulting in a *Se* of 81.6% (95% PCI: 72.7, 88.3) and *Sp* of 97.1% (95% PCI: 96.0, 97.9). At the kit-recommended cut-off of 0.40 PI, the C-ELISA exhibited a *Se* of 81.8% (95% PCI: 72.9, 88.5) and *Sp* of 96.9% (95% PCI: 95.7, 97.8).

3.3. Sensitivity and specificity estimates of four tests

Table 3 shows the median of the posterior distribution along with 95% PCI for *Se* and *Sp* of the four tests for Scenario 1. This scenario represents the final model, wherein informative priors were incorporated. Model convergence was checked by plotting traceplots (Appendix file 3). Among the tests, FPA has the highest sensitivity (89.9% at optimal cut-off), and outperforms RBT (79.9%), SAT (69.7%), and C-ELISA (81.6% at optimal cut-off). All four tests also have good specificities, ranging from 97.1% to 99.6%. The posterior distributions of *Se* estimates of RBT and SAT are slightly shifted relative to the priors, whereas the posterior distributions of *Sp* of RBT and SAT are shifted to the right and highly concentrated (Appendix file 4). Notably, the conditional sensitivity and specificity of SAT indicate a high dependence on the RBT test results. SAT has significantly higher *Se* in RBT-positive

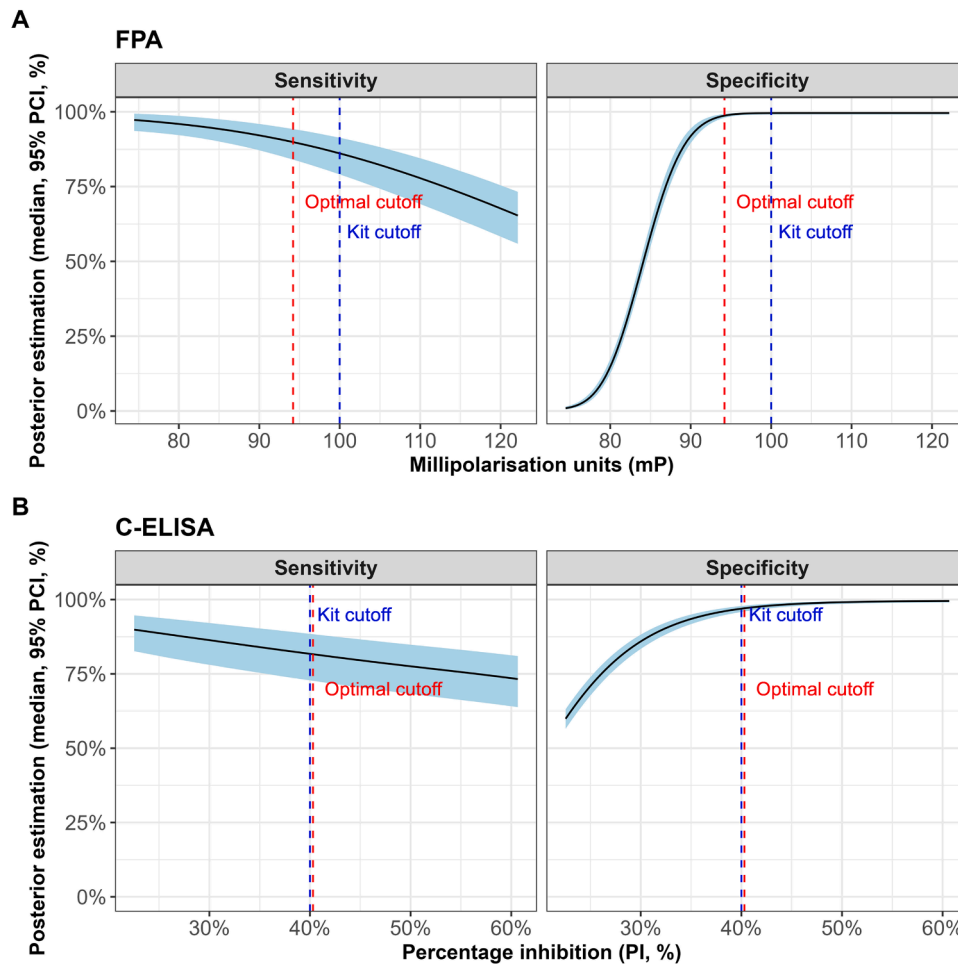


Fig. 3. Bayesian receiver operating characteristic (ROC) curves for the results of FPA (Panel A) and C-ELISA (Panel B) for detecting antibodies against *Brucella abortus* in dairy cows. The optimal cut-off value was determined by maximizing the Youden Index. Solid black line represents the median estimate with a 95% posterior credible interval (blue shade areas).

Table 3

Sensitivity (*Se*) and specificity (*Sp*) estimates of the four tests and animal-level prevalence of antibodies against *Brucella abortus* of each region and farm with their 95% posterior credible intervals (PCIs) according to the model with informative priors.

Parameters	Median (%)	95% PCI (%)
<i>Se</i> RBT	79.9	69.8, 87.5
<i>Sp</i> RBT	99.6	98.8, 99.9
<i>Se</i> SAT	69.7	61.4, 77.1
<i>Sp</i> SAT	99.5	98.7, 99.9
<i>Se</i> SAT _{C[1]}	42.8	28.1, 58.9
<i>Sp</i> SAT _{C[1]}	99.6	98.8, 99.9
<i>Se</i> SAT _{C[2]}	76.6	68.6, 83.5
<i>Sp</i> SAT _{C[2]}	82.1	55.4, 96.4
<i>Se</i> FPA (kit cut-off)	86.1	79.1, 91.3
<i>Sp</i> FPA (kit cut-off)	99.6	99.1, 99.9
<i>Se</i> FPA (optimal cut-off)	89.9	84.0, 94.2
<i>Sp</i> FPA (optimal cut-off)	98.7	98.0, 99.3
<i>Se</i> C-ELISA (kit cut-off)	81.8	72.9, 88.5
<i>Sp</i> C-ELISA (kit cut-off)	96.9	95.7, 97.8
<i>Se</i> C-ELISA (optimal cut-off)	81.6	72.7, 88.3
<i>Sp</i> C-ELISA (optimal cut-off)	97.1	96.0, 97.9
Prevalence PR	4.7	4.0, 5.9
Prevalence ZR	30.3	28.5, 33.0
Prevalence farm 1	0.9	0.9, 1.9
Prevalence farm 2	1.8	1.8, 4.5
Prevalence farm 3	10.7	9.7, 13.6
Prevalence farm 4	29.2	27.4, 32.7
Prevalence farm 5	24.5	21.8, 29.1
Prevalence farm 6	37.4	35.5, 40.2

C[1]: conditional on RBT-negative; C[2]: conditional on RBT-positive

conditions ($Se_{SATC[2]} = 76.6\%$, 95% PCI: 68.6, 83.5) than in RBT-negative conditions ($Se_{SATC[1]} = 42.8\%$, 95% PCI: 28.1, 58.9), while the *Sp* of SAT is conversely higher in RBT-negative conditions ($Sp_{SATC[1]} = 99.6\%$, 95% PCI: 98.8, 99.9) than in RBT-positive conditions ($Sp_{SATC[2]} = 82.1\%$, 95% PCI: 55.4, 96.4). We also find that the FPA test results are correlated with C-ELISA as evidenced by correlation coefficients (*rhos* varying from 0.30 to 0.78) (Appendix file 5). The binary tests (RBT and SAT) are correlated with continuous tests (FPA and C-ELISA) in the case of diseased animals (*rhos* varying from 0.33 to 0.51), but there is no correlation between them for non-diseased animals (Appendix file 5). The posterior distributions of the parameters used in the Gaussian mixture model are summarized in Appendix file 6.

3.4. Prevalence and parity effect

The estimated true animal-level prevalence of antibodies against *B. abortus* was found to be 4.7% (95% PCI: 4.0, 5.9) in PR and 30.3% (95% PCI: 28.5, 33.0) in ZR. Furthermore, when examining specific farms, the animal-level prevalence varied widely, ranging from 0.9% to 37.4% (Table 3). The effect of parity on the detection of antibodies against *B. abortus* was assessed. The details about the distribution of individual parity for each farm and region were summarized in Appendix file 7. The results indicated that the odds of detecting these antibodies increased with parity, with estimates ranging from 1.2 to 2.2 for different parity groups compared to primiparous cows (Fig. 4).

3.5. Sensitivity analysis on priors

The sensitivity analysis conducted in an additional seven scenarios reveals the relative robustness of the model. When certain priors, such as prevalence, *Se* and *Sp* of the RBT test, are changed to a flat distribution (Beta(1, 1)) in Scenarios 2, 3, and 5, the posterior estimates of animal-level prevalence, as well as the *Se* and *Sp* of the four tests, remain nearly identical to those obtained in Scenario 1 (fully informative priors). This indicates that the model is not heavily influenced by changes in these specific priors, as depicted in Figs. 5 and 6. However, notable changes in the posterior estimates of animal-level prevalence, sensitivity, and specificity occur when the conditional *Se* and *Sp* of the SAT ($Se_{SATC[i]}$ and $Sp_{SATC[i]}$) are altered to a flat distribution. These changes are particularly prominent in the performance of SAT, as demonstrated in Figs. 5 and 6. This indicates that the model is sensitive to variations in the specified priors for SAT conditional *Se* and *Sp*.

4. Discussion

In this study we have successfully developed a novel BLCM that addresses the challenges of analyzing both continuous and discrete testing data with additional covariates, allowing us to estimate the optimal cut-off values of FPA and C-ELISA and assess the test performance of four serological tests for the diagnosis of *B. abortus* exposure in Henan dairy farms. The developed BLCM framework offers remarkable flexibility and applicability in handling one or multiple continuous or discrete testing data, making it a valuable tool in diagnostic evaluation and prevalence estimation for animal or human diseases. The cut-off

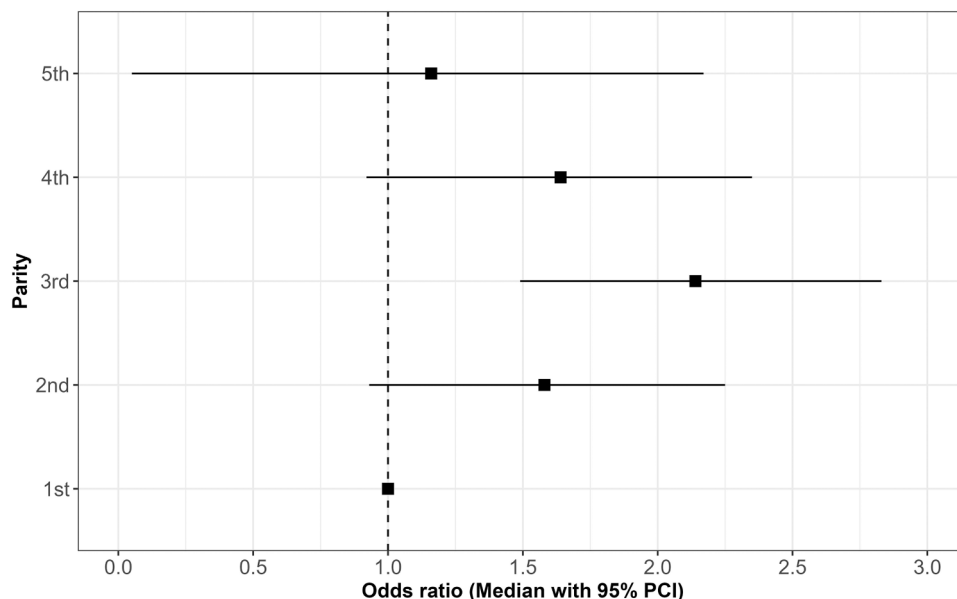


Fig. 4. Effects of parity (median, 95% posterior credibility interval - PCI) on the detection of antibodies against *Brucella abortus* in Henan Province.

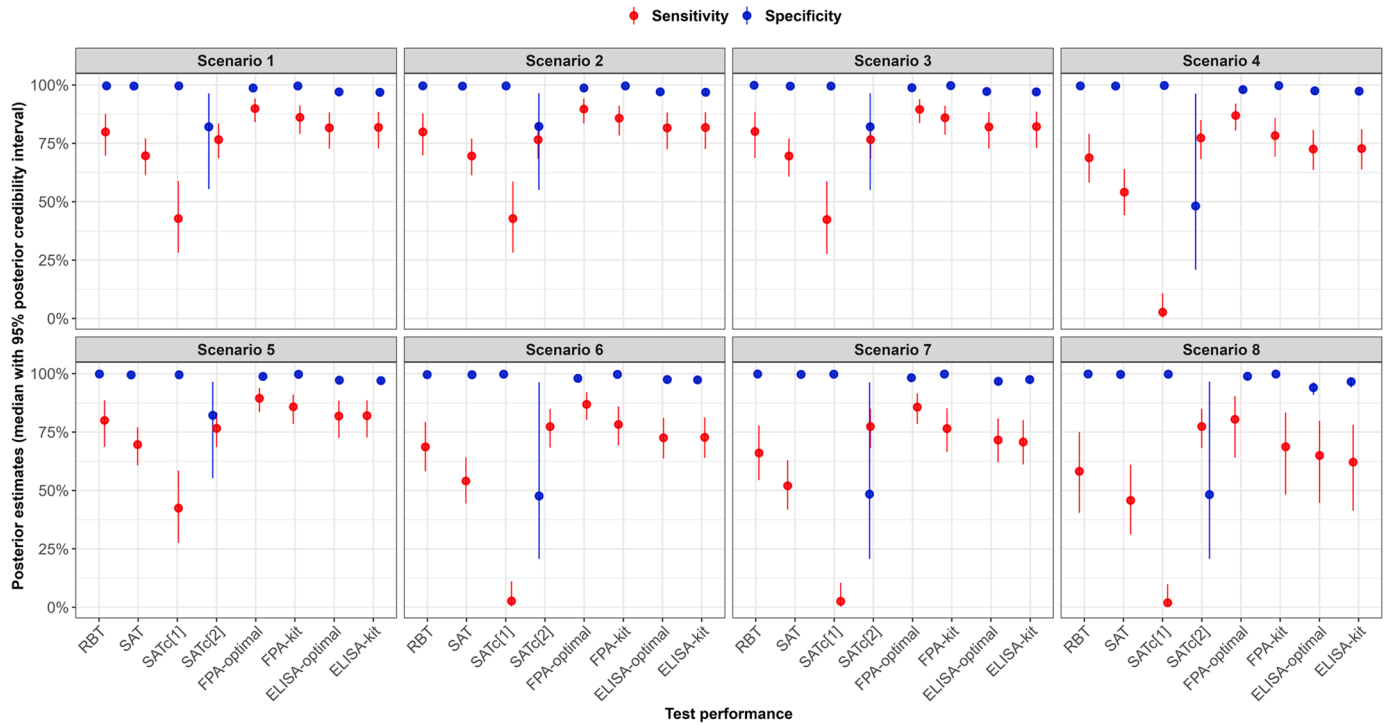


Fig. 5. A dot plot with bar (median, 95% posterior credibility interval - PCI) showing the impact of changing prior information on sensitivity and specificity estimates for eight scenarios. $c_{[1]}$: conditional on RBT-negative; $c_{[2]}$: conditional on RBT-positive.

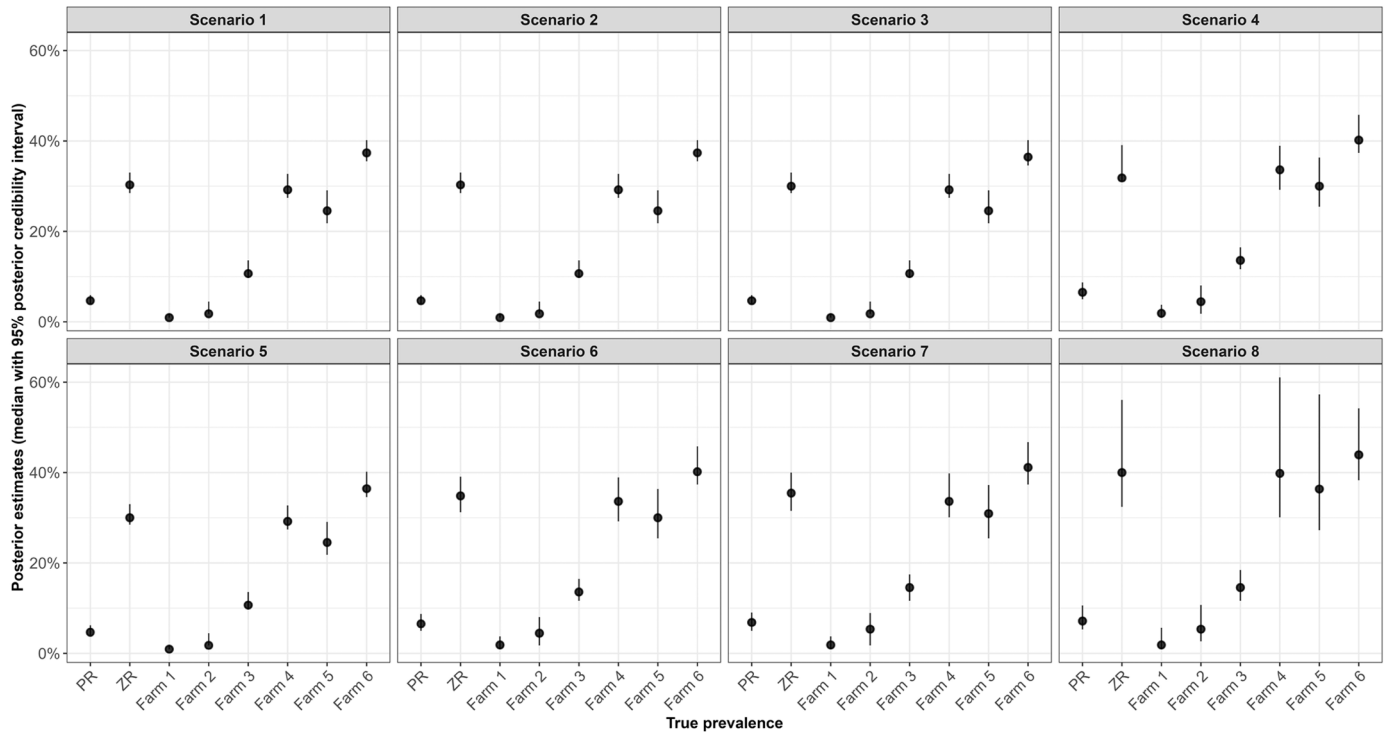


Fig. 6. A dot plot with bar (median, 95% posterior credibility interval - PCI) showing the impact of changing prior information on animal-level prevalence estimates within specific regions (PR and ZR) and farms for eight scenarios.

values, Se , and Sp estimates obtained from our model are essential information for various stakeholders, including farmers, decision-makers, and researchers, in the clinical diagnosis and screening of brucellosis. Furthermore, these findings can be utilized in quantitative risk assessments and prevalence estimations, providing a solid foundation for

designing informed *B. abortus* control programs in Chinese dairy industries and indirectly benefiting human health.

4.1. Sensitivity and specificity estimates and their epidemiological implications

At the optimal cut-off value of 94.2 mP, our estimates for Se of 89.9% (95% PCI: 84.3, 94.2) and Sp of 98.7% (95% PCI: 98.0, 99.3) of FPA differ from a traditional validation study using a cut-off of 95 mP (relative Se = 98.6%, relative Sp = 100%) (Nielsen et al., 2001), particularly for Se . This may be because they used samples that were positive based on other tests (e.g., C-ELISA). Other validation studies conducted in Canada and Mexico indicated that FPA has 99.0% Se and 96.9%–99.9% Sp at 90 mP cut-off (Nielsen et al., 1996; Dajer et al., 1999). For C-ELISA, our sensitivity estimate (Se = 81.6%, 95% PCI: 72.7, 88.3) was lower than the previously reported estimates from a Bayesian latent class model (Se = 98.1%) (Wang et al., 2020). This may be due to the difference in modeling approach, we used continuous data for modeling, while the previous study used binary data (Wang et al., 2020). Another study conducted in Tanzania yielded a lower sensitivity estimate (Se = 62.3%) than this study for C-ELISA (Bodenham et al., 2021), which may be related to the unique study population and epidemiological context. SAT had the lowest sensitivity (69.7%) among the four tests evaluated, but it had a high specificity like other tests. The Se of RBT was estimated to be 79.9%, which is lower than one estimate (Se = 89.7%) (Rahman et al., 2019) but comparable to another (Se = 77.9%) (Bodenham et al., 2021).

The Se/Sp ratio can provide insights into the overall test performance of tests, particularly when balancing the trade-off between Se and Sp , but does not provide a complete picture on its own. Depending on the specific epidemiological question and context, Se or Sp may be prioritized. In high-prevalence farms, high sensitivity ensures accurate detection of infected animals, reducing transmission risk. In low-prevalence farms, high specificity avoids unnecessary measures for healthy animals but may have limited impact due to the small proportion of true positives. Additional measures might be needed in low-prevalence populations to optimize disease control efforts. However, tests with high specificity are commonly employed as a ‘confirmatory’ test following a positive RBT result. This diagnostic practice has been commonly used in dairy cattle and human studies in China and is also recommended by the Chinese National Standard Diagnostic Techniques for Animal Brucellosis (GB/T 18646–2018) (Li et al., 2021). This practice has been argued that the concept of a ‘confirmatory’ test is not absolute and can be misleading (Ducrottoy et al., 2018). In the absence of vaccination and a high-prevalence context, a positive RBT does not require further retesting by a ‘confirmatory’ test such as SAT and ELISAs as the Se/Sp ratios of others are not better than that of RBT, like in ZR region. It is in the context of vaccination that SAT or C-ELISA becomes the ‘confirmatory’ test for positive RBT results, aiming to mitigate the false-positive problems associated with vaccinal antibodies and unnecessary culling, as advocated by WOAHP rules (WOAH, 2022b). Samples from non-vaccinated and high-prevalence areas such as ZR may be exempt from ‘confirmatory’ test if they are positive on screening test (e.g., RRT) given the high specificities assessed above (all > 97%), but if samples are from vaccinated or low-prevalence areas like PR ‘confirmatory’ tests should be used to mitigate the problem of false positives, especially when the true prevalence is close to or below 1- Sp . When farms are at the last stage of the brucellosis elimination program with a low prevalence (e.g., <1%), testing in series may be a priority to identify all positive animals, rather than considering cost. Other factors, such as cost, availability, and ease of use, should also be considered when selecting a test for use in epidemiological studies or clinical practice.

As the BLCM iterated through the observed data, we found interesting shifts in the posterior distributions of Se and Sp of RBT and SAT (Appendix file 4). Specifically, the posterior distribution of Se of RBT and SAT experienced a marginal shift, suggesting that our initial assumptions about this parameter were relatively robust and aligned well with the empirical data. Conversely, the posterior distribution of Sp of RBT and SAT exhibited a notable shift towards higher values, which

supported a higher Sp than the original assumptions. This adjustment highlights the profound impact of empirical evidence in refining our understanding of test performance. These shifts in the posterior distributions of Se and Sp provide crucial insights into the performance characteristics of the diagnostic tests. Similar shifts in the posterior distribution of parameters have been found in many studies (Arif et al., 2018; Rahman et al., 2019) as prior assumptions can only provide certain information and cannot perfectly fit the empirical data.

To the best of our knowledge, the Se and Sp of four serological tests may vary depending on factors such as parity and other relevant factors, such as species, stage of infection, and sample type (Ducrottoy et al., 2016; Bodenham et al., 2021). The immune response can differ among animals of different ages due to variations in immune maturity and previous exposure to *B. abortus*. Older cows that have been previously exposed to *B. abortus* may have a higher likelihood of developing detectable antibody titers, while younger animals, particularly those that have not yet been exposed to the bacteria or are in incubation stage before occurrence of abortion, may not have reached detectable levels of antibodies (Ducrottoy et al., 2016; Ducrottoy et al., 2018). This age-related difference can lead to higher sensitivity in older animals but lower sensitivity in younger animals. In this study, we integrated parity as a proxy for age into our Bayesian model to obtain a posterior marginal distribution of Se , Sp and parity-specific odds ratios. This approach simplifies the model and facilitates the practical application and generalization of the posterior estimates. On the other hand, the Sp estimates of serological tests are generally high, but false-positive cases can still occur due to cross-reactivity with other bacteria, such as *Yersinia enterocolitica* or *Escherichia coli* (Jones et al., 2009; Jones et al., 2010; Ducrottoy et al., 2016; Ducrottoy et al., 2018). In the presence of concurrent health conditions caused by these pathogens, the Sp estimates may be overestimated. The choice of serological tests for diagnosis and screening of *B. abortus* should be based on a comprehensive understanding of the population under investigation, the existence of cross-reacted pathogens, the specific objectives of the testing, plausible prevalence, and performance characteristics. Interpreting test results in conjunction with other available clinical and epidemiological information contributes to an accurate classification based on these tests.

4.2. Optimal cut-off values for FPA and C-ELISA

Our BLCM in Henan dairy populations shows that the optimal cut-off values for the FPA and C-ELISA kits are 94.2 mP and 0.403 PI respectively. The cut-off value for FPA is lower than the kit-recommended while that for C-ELISA is almost consistent with the recommendation (Fig. 3). Our model indicates that lowering the cut-off value for FPA leads to improved Se estimates, while the decrease in Sp estimates is negligible. This observation is supported by the minimal change observed in the Youden Indexes for FPA within the range of 90.5–100.5 mP. However, it is important to note that test results falling within this range should be regarded as suspicious and controversial. In such cases, we recommend retesting these animals with another supplementary test with high specificity to enhance diagnostic discrimination (WOAH, 2022b). Alternatively, if a decision is made solely based on a single test result falling within this range, the interpretation of test results should take into account the purpose of testing, the likely prevalence and/or pre-test likelihood of disease, and economic and public health consequences of misclassification of disease status. For example, in the case of introducing animals to a *B. abortus*-free farm, Se may be prioritized over Sp to minimize the chance of misclassifying positive animals as negative through adjusting to a lower cut-off value. This consideration is crucial during an emerging outbreak, as the potential cost of false negatives is typically higher than that of false positives.

Comparing the current study with other reports, researchers have reported various optimal cut-off values for FPA and C-ELISA tests based on maximizing the Youden Index, which is the most commonly used criterion to strike a balance between sensitivity and specificity in ROC

analysis (Greiner et al., 2000). For instance, Ren et al. (2016) reported a cut-off value of 95 mP based on ROC analysis, while many studies used the kit-recommended cut-offs (Eurofins-INGENASA, 2022). However, it is important to note that cut-off values can be adjusted for various reasons. The relatively higher cut-off value used by other studies may improve the Sp and thus are suitable for use in low prevalence and test-and-slaughter contexts to reduce false positive misclassifications from financial considerations (Greiner and Gardner, 2000a). In our study, an optimal cut-off value of FPA lower than that recommended by the kit could improve Se with a negligible sacrifice of Sp . Another interesting point is that in the Youden index calculation, adjusting the cut-off values of both FPA and C-ELISA within certain ranges (74–122 mP for FPA, 0.22–0.60 PI for C-ELISA) has a weighted effect on Se than on Sp (Fig. 3). We assume that this may be related to the nature of *B. abortus* and the high specificity of serological tests. Therefore, we recommend the need for validation of cut-off values when testing in new populations or species, and the choice of cut-off values should be based on empirical evidence, considering the purpose of testing, the specific epidemiological context and potential implications on result interpretation to ensure the accuracy and reliability of the tests (Greiner and Gardner, 2000b). The range of possible cut-off values provided by our model, along with the associated Se and Sp for FPA and C-ELISA, offers users the flexibility to adjust cut-off values according to their specific requirements and objectives, taking into account factors such as economic justifications, transmission risks, and public health implications.

4.3. Justification of Bayesian latent class model

Given the absence of a gold standard test, Bayesian latent class approaches have been developed to evaluate the test performance of one or multiple test(s), to determine optimal cut-off values for continuous test outcomes, and to estimate population prevalence along covariate effects. These approaches are considered the most appropriate for obtaining unbiased estimates when evaluating tests without a gold standard (Gardner et al., 2010; Johnson et al., 2019; Cheung et al., 2021; Yang et al., 2022). Following the STARD-BLCM (Kostoulas et al., 2017), careful consideration should be given to the definition of the latent class when interpreting the results. In our study, we used four serological tests to detect *Brucella*-specific antibodies, so the latent class estimated in this study was the seropositivity of *Brucella*-specific antibodies instead of *B. abortus* itself. However, we did not intend to explore more fundamental correlations between IgM and IgG status, as all the tests under evaluation were unable to differentiate between specific types of antibodies. Furthermore, considering the nature of *B. abortus* infection, which is typically chronic and characterized by persistent antibodies (Bercovich, 1998; Ducrottoy et al., 2016; Rahman et al., 2019), it is reasonable to assume that modeling the presence of antibodies closely represents the optimal latent class of *B. abortus* infection (Rahman et al., 2019). Nevertheless, some animals may be in the incubation period (e.g., before the occurrence of the first abortion) or immunocompromised, resulting in undetectable levels of antibodies (Ducrottoy et al., 2016), so the true Se of the four tests may be lower than the current estimates in these animals. Limmathuratsakul et al. (2010) defined the latent class as the true infection status in their latent class model with the addition of culture, which differed from our latent class definition (presence of specific antibodies) (Arif et al., 2018; Rahman et al., 2019; Wang et al., 2020), so their Se and Sp estimates may be more accurate as discussed before. Prospective research endeavors could incorporate bacterial culture and PCR tests to address this knowledge gap and explore the potential relationship between bacterial shedding and antibody dynamics.

Existing studies have primarily focused on utilizing either binary test results (Arif et al., 2018; Rahman et al., 2019; Wang et al., 2020) or continuous test results (Choi et al., 2006; Jones et al., 2009; Yang et al., 2022), with limited research exploring the integration of both binary and continuous data within a Bayesian latent class model. In our study,

we built upon the work of Coupe (2021), who used two binary assays and one continuous assay to detect specific antibodies against *Toxoplasma gondii* in a single population. We expanded upon this framework by incorporating two binary tests and two continuous tests across two distinct populations. One of the major strengths of our model lies in its ability to seamlessly integrate discrete and continuous testing data within a Bayesian framework without any loss of information, all in a single step. A recent study used the posterior positive probability, conditional on Se , Sp , and prevalence estimates, to determine the cut-off value that maximized the sum of sensitivity and specificity for ELISA (Olsen et al., 2022). Although this approach avoids data loss due to dichotomization, it requires multiple steps to obtain the final posterior estimates (Olsen et al., 2022). Furthermore, additional covariates can also be incorporated into the model to adjust the effect of host characteristics on the cut-off, Se and Sp .

4.4. Prevalence estimates

The overall apparent seroprevalence in the PR and ZR regions varied across the four tests, ranging from 1.2% to 2.8% in PR and 20.6% to 27.3% in ZR by four serological tests. This disparity reflects the distinct prevalence of antibodies against *B. abortus* in these two dairy populations. However, the apparent seroprevalence in previous studies was 2.2% and 7.8% in PR and ZR, respectively (Wan, 2017; Liu et al., 2020). Several factors may account for these discrepancies, such as variations in sample size and the characteristics of the sampled populations. In this study, only six farms were included, which may limit the representativeness compared to the previously reported studies (Wan, 2017; Liu et al., 2020). Additionally, the animal-level prevalence in our study was estimated by adjusting for the Se and Sp of four evaluated tests, parity, and farm random effects, which were not accounted for in the aforementioned reports (Wan, 2017; Liu et al., 2020). The potential rapid transmission of *B. abortus* in sampled farms may also occur in the gap between investigations, leading to an increase in prevalence. In this study, the posterior distributions of prevalence in two regions were shifted to the right, which may be related to the relatively conservative and diffuse prior distributions and be driven by observed test results. This alteration in the distribution of prevalence underscores the critical role of empirical evidence in refining our understanding and also indicate potential dynamics of disease in the studies populations. Other studies have also found shifting relative a prior information (Branscum et al., 2004). We highlight that the prevalence estimates may not be appropriate to other regions and populations, especially for young animals and other population groups other than PR and ZR regions. However, the Se and Sp estimates of the four diagnostic tests are robust and reliable, making these Se and Sp estimates potentially applicable in disease screening and surveillance efforts in commercial dairy farms in Henan and potentially other contexts.

4.5. Conditional dependence among tests

Our analysis revealed the presence of conditional dependence among all evaluated tests. Specifically, Se and Sp of SAT are highly dependent on the RBT results, meaning the dependence must be incorporated to calculate the combined sensitivity and specificity when performing serial or parallel testing. In our study, we observed that the combined sensitivity in the RBT-SAT series was 60.9% when accounting for the dependence, whereas it was 55.6% when the dependence was not considered. Therefore, in the presence of conditional dependence, omission of conditional dependence in the calculation of joint specificity in series will lead to underestimated Se and overestimated Sp in serial testing, while the opposite bias occurs in parallel testing. Unfortunately, this independence assumption has often been disregarded in most studies in China, which can introduce bias into the results and have potentially significant consequences in prevalence estimation and risk analysis. Georgiadis et al. (2003) suggested that when the conditional

correlations are low (≤ 0.2) the conditional independence assumption can be ignored with limited influences. On the opposite, when the correlations are high (> 0.2), the conditional dependence model should be preferred. The eight correlation coefficient estimates between FPA and C-ELISA are all larger than 0.3, indicating that there is also a high dependence between FPA and C-ELISA regardless of RBT, SAT results, and disease status. However, the conditional dependence between binary tests (RBT and SAT) and continuous tests (FPA and C-ELISA) is only evident in the presence of disease, not in non-disease states (**Appendix file 5**). Hence, a semi-dependent model that only incorporated a correlation between binary tests and continuous tests given the disease status could be developed in future studies (Kostoulas et al., 2006). Our sensitivity analysis also captured the conditional dependence between FPA and C-ELISA tests, where the model was only sensitive to SAT conditional Se and Sp . This observation may arise from the use of the identical SAT conditional Se and Sp given RBT results and flat prior distributions in sensitivity analysis, however; there are currently no published reports on SAT conditional Se and Sp given RBT. This potential concern could be addressed in future empirical studies by constraining prior distribution given another test result (e.g., SAT $Se_{C[2]} > SAT Se_{C[1]}$, 2 means given RBT-positive, 1 means given RBT-negative) or by using different parameters that remove implausible prior values rather than flat priors (Johnson et al., 2019).

5. Conclusion

We built a novel Bayesian Latent Class Model that integrates both binary and continuous testing outcomes and additional fixed and random effects to calibrate the optimal cut-off values by maximizing Youden Index. Our findings show that sensitivity estimates of each of the four tests vary between 69.7% and 89.9%, and specificity estimates range from 97.1% to 99.6%. The optimal cut-off values for FPA and C-ELISA kits are 94.2 mP and 0.403 PI, respectively. Our modeling approach is broadly applicable to determining an optimal cut-off value and estimating diagnostic performance in the absence of a gold standard test. It has the potential to enhance diagnostic accuracy and contribute to more reliable disease detection and management strategies in animal disease control programs.

Ethics approval and consent to participate

The study protocol was approved by a local ethics committee – Animal Experimental Ethical Inspection of Laboratory Animal Centre, Huazhong Agricultural University (Protocol Number: HZAUBU-2021–0003). The study involving human participants was reviewed and approved by peer review at Massey University Human Ethics Committee (Protocol Number: SOA 21/60). Written consent of farm owners was obtained before the collection of blood samples from dairy farms.

Funding

This work was financially supported by the China Agriculture Research System of MOF and MARA (Grant Number: CARS-37) and the 2023 Massey University Doctoral Conference Grants. YW is also supported by EpiCentre, School of Veterinary Science, Massey University regarding the doctoral tuition and stipend.

CRedit authorship contribution statement

Compton Chris: Conceptualization, Methodology, Resources, Software, Supervision, Validation, Writing – review & editing. **Vallée Emilie:** Data curation, Formal analysis, Methodology, Supervision, Validation, Writing – review & editing. **Wang Yu:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Guo Aizhen:** Conceptualization, Funding acquisition, Investigation, Methodology,

Supervision, Writing – review & editing. **Heuer Cord:** Conceptualization, Methodology, Project administration, Resources, Supervision. **Wang Youming:** Investigation, Methodology, Supervision, Writing – review & editing. **Vignes Matthieu:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – review & editing. **Zhang Zhen:** Data curation, Investigation, Methodology, Resources, Supervision.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request. The reproductive code is available at <https://github.com/YuWang1009/BLCM.git>.

Acknowledgments

The authors thank the participating farms and Henan DHI Centre staff for their warm help and support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.prevetmed.2024.106115](https://doi.org/10.1016/j.prevetmed.2024.106115).

References

- Arif, S., Heller, J., Hernandez-Jover, M., McGill, D.M., Thomson, P.C., 2018. Evaluation of three serological tests for diagnosis of bovine brucellosis in smallholder farms in Pakistan by estimating sensitivity and specificity using Bayesian latent class analysis. *Prev. Vet. Med.* 149, 21–28.
- Baldi, P.C., Giambartolomei, G.H., Goldbaum, F.A., Abdón, L.F., Velikovskiy, C.A., Kittelberger, R., Fossati, C.A., 1996. Humoral immune response against lipopolysaccharide and cytoplasmic proteins of *Brucella abortus* in cattle vaccinated with *B. abortus* S19 or experimentally infected with *Yersinia enterocolitica* serotype 0:9. *Clin. Diagn. Lab. Immunol.* 3, 472–476.
- Bercovich, Z., 1998. Maintenance of *Brucella* Abortus-free herds: A review with emphasis on the epidemiology and the problems in diagnosing brucellosis in areas of low prevalence. *Vet. Q.* 20, 81–88.
- Bodenham, R.F., Mazeri, S., Cleaveland, S., Crump, J.A., Fasina, F.O., de Glanville, W.A., Haydon, D.T., Kazwala, R.R., Kibona, T.J., Maro, V.P., Maze, M.J., Mmbaga, B.T., Mtui-Malamsha, N.J., Shirima, G.M., Swai, E.S., Thomas, K.M., Bronsvort, B.M., Halliday, J.E.B., 2021. Latent class evaluation of the performance of serological tests for exposure to *Brucella* spp. in cattle, sheep, and goats in Tanzania. *PLOS Negl. Trop. Dis.* 15, e0009630.
- Branscum, A.J., Gardner, I.A., Johnson, W.O., 2004. Bayesian modeling of animal- and herd-level prevalences. *Prev. Vet. Med.* 66, 101–112.
- Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7, 434–455.
- Cheung, A., Dufour, S., Jones, G., Kostoulas, P., Stevenson, M.A., Singanallur, N.B., Firestone, S.M., 2021. Bayesian latent class analysis when the reference test is imperfect. *Rev. Sci. Et. Tech. (Int. Off. Epizoot.)* 40, 271–286.
- Choi, Y.-K., Johnson, W.O., Collins, M.T., Gardner, I.A., 2006. Bayesian inferences for receiver operating characteristic curves in the absence of a gold standard. *J. Agric., Biol., Environ. Stat.* 11, 210.
- Coupe, A., 2021. Investigating *Toxoplasma gondii* in the marine environment in New Zealand: from cats to kai moana (shellfish): a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Veterinary Science at Massey University, School of Veterinary Science, Massey University, Palmerston North, New Zealand. Massey University.
- Dajer, A., Luna-Martinez, E., Zapata, D., Villegas, S., Gutiérrez, E., Peña, G., Gurría, F., Nielsen, K., Gall, D., 1999. Evaluation of a fluorescence-polarization assay for the diagnosis of bovine brucellosis in México. *Prev. Vet. Med.* 40, 67–73.
- Dendukuri, N., Rahme, E., Bélisle, P., Joseph, L., 2004. Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test. *Biometrics* 60, 388–397.
- Dendukuri, N., Bélisle, P., Joseph, L., 2010. Bayesian sample size for diagnostic test studies in the absence of a gold standard: comparing identifiable with non-identifiable models. *Stat. Med.* 29, 2688–2697.
- Dong, S.-B., Xiao, D., Liu, J.-Y., Bi, H.-M., Zheng, Z.-R., Wang, L.-D., Yang, X.-W., Tian, G.-Z., Zhao, H.-Y., Piao, D.-R., Xing, Z.-F., Jiang, H., 2021. Fluorescence polarization assay improves the rapid detection of human brucellosis in China. *Infect. Dis. Poverty* 10, 46.

- Ducrotay, M.J., Conde-Álvarez, R., Blasco, J.M., Moriyón, I., 2016. A review of the basis of the immunological diagnosis of ruminant brucellosis. *Vet. Immunol. Immunopathol.* 171, 81–102.
- Ducrotay, M.J., Muñoz, P.M., Conde-Álvarez, R., Blasco, J.M., Moriyón, I., 2018. A systematic review of current immunological tests for the diagnosis of cattle brucellosis. *Prev. Vet. Med.* 151, 57–72.
- Eurofins-INGENASA, 2022. INgezim® BRUCELLA Compac 2.0 Detection of specific antibodies to LPS of *Brucella* spp., in bovine, ovine, caprine and porcine sera samples.
- Gall, D., Nielsen, K., 2004. Serological diagnosis of bovine brucellosis: a review of test performance and cost comparison. *Rev. Sci. Et. Tech. (Int. Off. Epizoot.)* 23, 989–1002.
- Gardner, I.A., Greiner, M., Dubey, J.P., 2010. Statistical evaluation of test accuracy studies for toxoplasma gondii in food animal intermediate hosts. *Zoonoses Public Health* 57, 82–94.
- Gardner, I.A., Colling, A., Greiner, M., 2019. Design, statistical analysis and reporting standards for test accuracy studies for infectious diseases in animals: progress, challenges and recommendations. *Prev. Vet. Med.* 162, 46–55.
- Georgiadis, M.P., Johnson, W.O., Gardner, I.A., Singh, R., 2003. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *J. R. Stat. Soc. Ser. C: Appl. Stat.* 52, 63–76.
- Greiner, M., Gardner, I.A., 2000a. Application of diagnostic tests in veterinary epidemiologic studies. *Prev. Vet. Med.* 45, 43–59.
- Greiner, M., Gardner, I.A., 2000b. Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3–22.
- 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.
- Higgins, J.L., Gonzalez-Juarrero, M., Bowen, R.A., 2017. Evaluation of shedding, tissue burdens, and humoral immune response in goats after experimental challenge with the virulent *Brucella melitensis* strain 16M and the reduced virulence vaccine strain Rev. 1. *PLOS ONE* 12, e0185823.
- Johnson, W.O., Jones, G., Gardner, I.A., 2019. Gold standards are out and Bayes is in: implementing the cure for imperfect reference tests in diagnostic accuracy studies. *Prev. Vet. Med.* 167, 113–127.
- Jones, G., Johnson, W.O., Vink, W.D., 2009. Evaluating a continuous biomarker for infection by using observed disease status with covariate effects on disease. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)* 58, 705–717.
- Jones, G., Johnson, W.O., Hanson, T.E., Christensen, R., 2010. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* 66, 855–863.
- Kostoulas, P., Leontides, L., Billinis, C., Florou, M., 2006. Application of a semi-dependent latent model in the Bayesian estimation of the sensitivity and specificity of two faecal culture methods for diagnosis of paratuberculosis in sub-clinically infected Greek dairy sheep and goats. *Prev. Vet. Med.* 76, 121–134.
- Kostoulas, P., Nielsen, S.S., Branscum, A.J., Johnson, W.O., Dendukuri, N., Dhand, N.K., Toft, N., Gardner, I.A., 2017. STARD-BLCM: standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. *Prev. Vet. Med.* 138, 37–47.
- Li, Y., Tan, D., Xue, S., Shen, C., Ning, H., Cai, C., Liu, Z., 2021. Prevalence, distribution and risk factors for brucellosis infection in goat farms in Ningxiang, China. *BMC Vet. Res.* 17, 39.
- Limmathurotsakul, D., Jansen, K., Arayawichanon, A., Simpson, J.A., White, L.J., Lee, S.J., Wuthiekanun, V., Chantratita, N., Cheng, A., Day, N.P.J., Verzilli, C., Peacock, S.J., 2010. Defining the True Sensitivity of Culture for the Diagnosis of Melioidosis Using Bayesian Latent Class Models. *PLOS ONE* 5, e12485.
- Liu, Y., Ban, F., Hu, W., Cheng, Y., Zhang, H., Xie, C., Zhao, X., Wang, D., Li, J., Shen, C., Yan, R., 2020. Cross-sectional study on the prevalence and risks of spreading brucellosis among cow farms in pingdingshan city of Henan Province. *China Anim. Health Insp.* 37, 22–26. +38.
- Mantur, B.G., Amarnath, S.K., Shinde, R.S., 2007. Review of clinical and laboratory features of human brucellosis. *Indian J. Med. Microbiol.* 25, 188–202.
- Marquetoux, N., Vignes, M., Burroughs, A., Sumner, E., Sawford, K., Jones, G., 2023. Evaluation of the accuracy of the IDvet serological test for *Mycoplasma bovis* infection in cattle using latent class analysis of paired serum ELISA and quantitative real-time PCR on tonsillar swabs sampled at slaughter. *PLOS ONE* 18, e0285598.
- Matope, G., Bhebhe, E., Muma, J.B., Oloya, J., Madekurozwa, R.L., Lund, A., Skjerve, E., 2011. Seroprevalence of brucellosis and its associated risk factors in cattle from smallholder dairy farms in Zimbabwe. *Trop. Anim. Health Prod.* 43, 975–982.
- Neta, A.V.C., Mol, J.P.S., Xavier, M.N., Paixão, T.A., Lage, A.P., Santos, R.L., 2010. Pathogenesis of bovine brucellosis. *Vet. J.* 184, 146–155.
- Nielsen, K., Gall, D., Jolley, M., Leishman, G., Balsevicius, S., Smith, P., Nicoletti, P., Thomas, F., 1996. A homogeneous fluorescence polarization assay for detection of antibody to *Brucella abortus*. *J. Immunol. Methods* 195, 161–168.
- Nielsen, K., Gall, D., Smith, P., Kelly, W., Yeo, J., Kenny, K., Heneghan, T., McNamara, S., Maher, P., O'Connor, J., Walsh, B., Carroll, J., Rojas, X., Rojas, F., Perez, B., Wulff, O., Buffoni, L., Salustio, E., Gregoret, R., Samartino, L., Dajer, A., Luna-Martinez, E., 2001. Fluorescence polarization assay for the diagnosis of bovine brucellosis: adaptation to field use. *Vet. Microbiol.* 80, 163–170.
- Olsen, A., Nielsen, H.V., Alban, L., Houe, H., Jensen, T.B., Denwood, M., 2022. Determination of an optimal ELISA cut-off for the diagnosis of *Toxoplasma gondii* infection in pigs using Bayesian latent class modelling of data from multiple diagnostic tests. *Prev. Vet. Med.* 201, 105606.
- Pappas, G., 2022. The Lanzhou *Brucella* leak: the largest laboratory accident in the history of infectious diseases? *Clin. Infect. Dis.* 75, 1845–1847.
- Pappas, G., Papadimitriou, P., Akritidis, N., Christou, L., Tsianos, E.V., 2006. The new global map of human brucellosis. *Lancet Infect. Dis.* 6, 91–99.
- Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*, 1–10.
- Rahman, A.K.M.A., Smit, S., Devleeschauwer, B., Kostoulas, P., Abatih, E., Saegerman, C., Shamsuddin, M., Berkvens, D., Dhand, N.K., Ward, M.P., 2019. Bayesian evaluation of three serological tests for the diagnosis of bovine brucellosis in Bangladesh. *Epidemiol. Infect.* 147, e73.
- Ran, X., Cheng, J., Wang, M., Chen, X., Wang, H., Ge, Y., Ni, H., Zhang, X.-X., Wen, X., 2019. Brucellosis seroprevalence in dairy cattle in China during 2008–2018: A systematic review and meta-analysis. *Acta Trop.* 189, 117–123.
- Ren, Y., Zhang, H., Sun, X., Li, Z., Cai, J., Fan, W., 2016. Comparative study of FPA and ELISA method in the diagnosis of animal brucellosis. *Chin. J. Zoonoses* 32, 806–809.
- Stevenson, M., Sergeant, E., Nunes, T., Heuer, C., Marshall, J., Sanchez, J., Thornton, R., Reiczgel, J., Robison-Cox, J., Sebastiani, P., Solyomos, P., Yoshida, K., Jones, G., Pirikahu, S., Firestone, S., Kyle, R., Popp, J., Jay, M., Reynard, C., Cheung, A., Singanallur, N., Szabo, A., Rabiee, A., 2022. epiR: Tools for the Analysis of Epidemiological Data. R package version 2.0.62.
- Su, Y.-S., Yajima, M., 2021. R2jags: Using R to run 'JAGS'. R package version 0.7–1.
- Wan, J., 2017. A investigation on prevalence of brucellosis in dairy cows in one area of Henan Province. *China Herbivore. Science* 37, 75–76.
- Wang, X., Jiang, H., 2020. Global prevalence of human brucellosis. *Zhonghua Liu Xing Bing. Xue Za Zhi* 41, 1717–1722.
- Wang, Y., Robertson, I.D., Cheng, S., Wang, Y., Hou, L., Wang, G., Wu, X., Li, X., Chen, Y., Guo, A., 2020. Evaluation of a milk ELISA as an alternative to a serum ELISA in the determination of the prevalence and incidence of brucellosis in dairy herds in Hubei Province, China. *Prev. Vet. Med.* 182, 105086.
- WOAH, 2022a. Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2022 - Chapter 2.2.5. Statistical approaches to validation.
- WOAH, 2022b. Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2022 - Chapter 3.1.4. - Brucellosis (infection with *Brucella abortus*, *B. melitensis* and *B. suis*).
- Yang, D.A., Laven, R.A., 2022. Performance of the StaphGold ELISA test in determining subclinical *Staphylococcus aureus* infections in dairy cows using a Gaussian mixture model. *Vet. Med. Sci.* 8, 1632–1639.
- Yang, D.A., Xiao, X., Jiang, P., Pfeiffer, D.U., Laven, R.A., 2022. Keeping continuous diagnostic data continuous: application of Bayesian latent class models in veterinary research. *Prev. Vet. Med.* 201, 105596.
- Yao, M., Guo, X., Wu, X., Bai, Q., Sun, M., Yin, D., 2022. Evaluation of the combined use of major outer membrane proteins in the serodiagnosis of Brucellosis. *Infect. Drug Resist* 15, 4093–4100.
- Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer* 3, 32–35.
- Zeng, J., Duoqi, C., Yuan, Z., Yuzhen, S., Fan, W., Tian, L., Cai, C., Robertson, I., 2017. Seroprevalence and risk factors for bovine brucellosis in domestic yaks (*Bos grunniens*) in Tibet, China. *Trop. Anim. Health Prod.* 49, 1339–1344.
- Zhang, N., Huang, D., Wu, W., Liu, J., Liang, F., Zhou, B., Guan, P., 2018. Animal brucellosis control or eradication programs worldwide: a systematic review of experiences and lessons learned. *Prev. Vet. Med.* 160, 105–115.
- Zhu, X., Zhao, Z., Ma, S., Guo, Z., Wang, M., Li, Z., Liu, Z., 2020. *Brucella melitensis*, a latent “travel bacterium,” continual spread and expansion from Northern to Southern China and its relationship to worldwide lineages. *Emerg. Microbes Infect.* 9, 1618–1627.
- Zinsstag, J., Schelling, E., Roth, F., Bonfroh, B., de Savigny, D., Tanner, M., 2007. Human benefits of animal interventions for zoonosis control. *Emerg. Infect. Dis.* 13, 527.