Tuberculosis xxx (xxxx) xxx

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: www.elsevier.com/locate/tube

A comparative study between milk- and serum-based antibody detection assays for Johne's disease in New Zealand dairy cattle

Venkatesh K.M.[a], Nicolas Lopez-Villalobos[a,*], Sandeep K. Gupta[b], Garry B. Udy[a,c], Richard Laven[d], Shih-Jiuan Chiu[e], Piyush Bugde[e], Yoichi Furuya[e], Venkata Sayoji Rao Dukkipati[a,f,**]

[a] School of Agriculture & Environment, Massey University, Palmerston North 4442, New Zealand
[b] AgResearch Limited, Hopkirk Research Institute, Palmerston North 4442, New Zealand
[c] Ministry for Primary Industries, Charles Fergusson Building, 34-38 Bowen Street, Wellington, 6011, New Zealand
[d] School of Veterinary Science, Massey University, Palmerston North 4442, New Zealand
[e] Pictor Ltd., 24 Balfour Road, Parnell, Auckland, 1052, New Zealand
[f] Shreiber School of Veterinary Medicine, Rowan University, 1000 Gilbreth Parkway, Harrison Township, NJ, 08062, United States

ARTICLE INFO

Keywords:
Dairy cattle
Enzyme-linked immunosorbent assay
Johne's disease
Kappa
Mycobacterium avium subsp. *paratuberculosis*

ABSTRACT

Dairy cattle are affected by Johne's disease. It is caused by *Mycobacterium avium* subspecies *paratuberculosis* (MAP). Suboptimal diagnostic tests add more to the productivity loss resulting from this disease. Agreement between and within different commercial kits is crucial in the decision-making process of disease surveillance programmes. This study compared two ELISAs, that is, Johne's disease commercial antibody detection kits (A and B), using milk and serum samples from New Zealand dairy cattle. These results were also compared with a subset of faecal PCR results. Five scenarios were considered for the comparison of ELISA tests. The point estimates of kappa coefficients (k) between the serum (0.84–0.94) assays were higher than the milk assays (0.59–0.82). The point estimates of kappa coefficients between serum and milk ELISA outcomes were higher for kit B (k = 0.79–0.86) than for kit A (k = 0.55–0.79). The point estimates of kappa coefficients between the ELISA and faecal PCR outcomes varied between 0.43 and 0.74. ELISA tests had point estimates of sensitivity ranging from 0.67 to 0.88 and specificity from 0.62 to 0.93, relative to the faecal PCR test. Results suggest that serum provides a better choice of sample type when both commercial kits A and B are used for Johne's disease surveillance of dairy cattle in New Zealand. Milk assays can be cost-effective to diagnose MAP-positive animals; kit B can be best suited for New Zealand conditions, provided the repeatability of the results is validated.

1. Introduction

Johne's disease, which is also called paratuberculosis, is an infectious gastrointestinal disease of ruminants. It is caused by an intracellular bacterium called *Mycobacterium avium* subspecies *paratuberculosis* (MAP). It is an endemic disease of dairy cattle in most countries [1]. The New Zealand dairy industry is predicted to lose US$54 million annually due to Johne's disease [2]. Suboptimal diagnostic tests, complex epidemiology of MAP, and lack of effective treatments or vaccines are all reasons why Johne's disease is difficult to eradicate in dairy cattle [1]. Hence, disease control programmes are aimed at reducing the prevalence of the disease in dairy cattle rather than eradicating it [1]. For disease preventive measures at any level, the foremost step is screening and/or diagnosing animals as being infected. Diagnostic tests for Johne's disease are either direct (targeting MAP) or indirect (targeting the host response to MAP) [3]. While detection of MAP in faeces by either culture or real-time (quantitative) polymerase chain reaction (qPCR) is the most commonly used direct test, screening serum or milk for anti-MAP antibodies using an enzyme-linked immunosorbent assay (ELISA) is the most widely used indirect diagnostic test for Johne's disease [3]. Direct diagnostic tests are either laborious (culture) or expensive (qPCR), while indirect tests (ELISA) are cost-effective, high-throughput, and quicker. Thus, ELISA is the preferred method for screening herds or individual animals. For ELISA, milk is more

* Corresponding author.
** Corresponding author. School of Agriculture & Environment, Massey University, Palmerston North 4442, New Zealand.
E-mail addresses: n.lopez-villalobos@massey.ac.nz (N. Lopez-Villalobos), dukkipativ@rowan.edu (V.S.R. Dukkipati).

https://doi.org/10.1016/j.tube.2025.102679
Received 13 January 2025; Received in revised form 22 August 2025; Accepted 26 August 2025
Available online 27 August 2025

extensively used than serum, as it is logistically easier [3] and minimises sampling costs for the farmers.

The diagnostic accuracy of ELISA tests is relative to the stage of Johne's disease, which is categorised into the silent infection, subclinical, clinical, and advanced clinical phases. While the clinical phases can be effectively diagnosed using commercial ELISA tests, the early stages are more difficult to diagnose [4,5]. Also, milk and serum ELISA differ in economic terms for the farmers. Additionally, as milk ELISA has a higher variability compared to serum ELISA [6], it might influence the diagnostic accuracy of the tests. Further, the diversity of MAP strains in different countries [1] may sometimes lead to differences in the accuracy of the diagnostic tests [7,8]. In this context, a few studies have reported variable agreement (good, moderate, or poor) between different ELISA tests [7,9–11]. While the superiority of a test can be primarily assessed by the diagnostic accuracy (sensitivity and specificity), it does not reflect the degree of agreement between them. This is particularly important when different diagnostic tests are used in the control measures to decide the fate of the animal. Also, it is useful for knowing the combining potential (serial or parallel testing) of these tests for screening and confirmation of Johne's disease. Thus, it is necessary to know the agreement within and between the different diagnostic tests [9]. To the best of our knowledge, there is no study on the agreement analysis of ELISA tests for Johne's disease in New Zealand dairy cattle. Thus, this study aimed to compare two commercial ELISA kits that are used for Johne's disease diagnosis in dairy cattle using milk and serum samples. We first compared the agreement between the milk and serum ELISA within and between commercial kits using Kappa statistics. Secondly, the sensitivity and specificity of the milk and serum ELISA relative to the faecal PCR results were estimated using a subset of samples. Lastly, milk ELISA tests were compared to serum ELISA results as a pseudo-gold standard for agreement, sensitivity, and specificity.

2. Material and methods

2.1. Study animals and sampling of biological specimens

It is to be noted that the data used in this study were originally collected as part of another study targeting 'in-house paratuberculosis diagnostics.' Four herds in the Manawatu region of New Zealand were selected for this cross-sectional study. These herds were selected based on the farmers' willingness to participate in the study among many that were inquired to participate. At the time of sampling, the true prevalence of Johne's disease was unknown at the animal and herd levels. New Zealand has no bovine paratuberculosis control programme or disease-free herd certification for trading animals. For this reason and farmers' unwillingness to participate in studies, sourcing disease-free herds in New Zealand is a difficult task, as it requires longitudinal monitoring of herds. Based on interaction with farmers, it was expected that both paratuberculosis-suspect and non-suspect herds would be part of the study. The herds in the study had different age and breed groups. Where possible, animals were randomly selected in each herd. Mostly, cows that were in milk were selected; however, a limited number of young heifers were also selected (only blood and faecal samples were obtained from them). In total, 1169 animals were sampled from four herds: 342 from herd A, 325 from herd B, 296 from herd C and 206 from herd D. The selected animals were not tested for bovine tuberculosis at least 60 days before this study. These dairy herds followed a typical spring-calving New Zealand pasture-based production system.

From each animal, milk (10 ml), blood (5 ml in vacutainers with no coagulant), and faecal (10 g) samples were collected at a single timepoint, with the help of the trained research assistants. Sampling of cows in the four herds was undertaken on multiple days in either early summer or late autumn, during the 2022/2023 production season. The collected samples were labelled with the herd and animal number and were transported on ice to the laboratory (Massey University, New Zealand). The serum from the clotted blood samples and skim milk from milk samples were separated by centrifugation of the samples in a refrigerated centrifuge at 1000×*g* for 10 min and were stored at −80 °C until used. Faecal samples were also stored at −80 °C until they were processed for qPCR.

2.2. Diagnostic tests used

The ELISA kits from Idexx Laboratories Inc., USA (IDEXX Paratuberculosis Screening Ab Test) and IDVet Innovative Diagnostics, France (ID Screen® Paratuberculosis Indirect Screening Test), along with a real-time PCR test (VetMAX™ MAP IS900-F57 Kit - catalogue number TMPT), were used in the study. Milk and/or serum samples from a few animals were not processed for ELISA due to errors in sampling or labelling. ELISAs were performed as per the manufacturers' guidelines. Corrected (negative control readings subtracted) antibody responses in the case of each sample (S) and corrected optical density readings at 450 nm wavelength (OD_{450}) were expressed as a percentage of positive (P) control readings (S/P ratio).

A subset of 144 samples, used for ELISA, were subjected to faecal PCR analysis. These samples were included from each of the disease categories (positive, suspect, and negative) determined by MAP-ELISA results for milk and/or serum. The extraction of DNA from faeces and the subsequent real-time PCR assays were conducted by the TB Molecular Diagnostics, AgResearch, Hopkirk Research Institute, Palmerston North, New Zealand. DNA from about 3.5 g of faeces from each animal was extracted using the MagMAX™ CORE Nucleic Acid Purification Kit (Thermo Fisher Scientific, Waltham, MA) with a mechanical lysis module. Quantitative PCRs, using the extracted DNA, were performed with the TaqMan real-time LSI VetMax MAP IS900-F57 Kit (Thermo Fisher Scientific, Waltham, MA, USA), as per the protocol recommended by the manufacturer.

2.3. Test scores interpretation and case definitions

The values of the ELISA S/P ratio were interpreted based on the respective kit manufacturer's description (Table 1). All the ELISA interpretations had three outcomes (positive, suspect, and negative), except for IDVet milk ELISA, which had two outcomes (positive and negative). Thus, for comparing the trichotomous outcomes between the ELISAs, the IDVet milk ELISA was reclassified into three outcomes (positive, suspect, and negative). This was done with a slight adjustment to the manufacturer's interpretation of the S/P ratio (Table 1). Furthermore, as per the manufacturer's guidelines, Idexx milk ELISA S/P ratio interpretations for testing bovine milk samples in countries other than Germany were initially used. However, there were relatively more positives compared to those from IDVet milk ELISA. Subsequently, higher thresholds of S/P ratio (Table 1), as applicable to Germany, were adopted for the present study. The results from these interpretations showed consistently higher agreements with IDVet milk ELISA results (data not shown), compared to those from the initial interpretations. Based on the cycle threshold (ct) value, the faecal real-time PCR results were initially categorised into four outcomes (Table 1); however, the strong positive and positive statuses were considered together as positive in further analyses. Also, suspect outcomes of the PCR tests were reclassified as positive if any ELISA test was positive for the corresponding animal (Table 2); if not, they were considered negative.

Five different scenarios were considered for the comparison of ELISA tests. In the first scenario, the trichotomous outcomes for each ELISA test were considered as such. As the faecal PCR test had very few or no outcomes in subsets, the comparison of ELISA tests with PCR tests was not made in this scenario. For scenarios two to five, trichotomous outcomes of ELISAs were converted into dichotomous classification as shown in Table 2.

For easy understanding, as an example, consider that Idexx milk ELISA has three outcomes (positive, negative, and suspect). In the second scenario, positive and negative outcomes were considered as such.

Table 1
Interpretation of the disease outcome of the animals based on the diagnostic scores.

Diagnostic test	Disease classification based on test scores		
	Positive[f]	Suspect[f]	Negative[f]
IDVet_Milk[a] $	>30	>25 and ≤ 30	≤25
Idexx_Milk[b,g]	≥40	>30 and < 40	≤30
IDVet_Serum[c]	≥70	>60 and < 70	≤60
Idexx_Serum[d]	≥55	>45 and < 55	≤45
Faecal PCR[e]	≤26 (strong positive) and >26 and ≤ 32 (positive)	>32 and ≤ 40	>40

$ As per the manufacturer's interpretation, IDVet milk ELISA has only positive (>30) and negative outcomes (≤30). However, for analysis purposes (trichotomous outcomes agreement only), outcomes were classified into three outcomes by adjusting S/P ratios.

[a] IDvet Innovative Diagnostics, France (milk sample).

[b] Idexx Laboratories Inc., USA (milk sample).

[c] IDvet Innovative Diagnostics, France (serum sample).

[d] Idexx Laboratories Inc., USA (serum sample).

[e] Faecal PCR using TaqMan real-time LSI VetMax MAP IS900-F57 Kit, Thermo Fisher Scientific, USA.

[f] Disease classification of animals by IDVet_Milk, Idexx_Milk, IDVet_serum and Idexx_Serum enzyme-linked immunosorbent assays (ELISAs) were based on corrected sample-to-positive control OD_{450} ratios (%S/P) for respective kits, and faecal PCR was based on cycle threshold (ct) value cutoff points for TaqMan real-time LSI VetMax MAP IS900-F57 Kit.

[g] As per the manufacturer's guidelines, Idexx milk ELISA S/P ratios interpretation (≥30 is positive, >20 and < 30 is suspect, and ≤20 is negative) applies to countries other than Germany. However, S/P ratio interpretations as applicable to the analysis of milk samples in Germany were adopted for the present study.

Table 2
Various scenarios considered for enzyme-linked immunosorbent assay (ELISA) test agreement and diagnostic accuracy comparison.

Scenario	Condition to reclassify suspect results	Final classification
1	"Suspects" were reported independently without reclassification.	Results were categorised as positive, negative, or suspect.
2	"Suspects" were reclassified as positive if any ELISA or PCR test was positive.	Results were categorised as positive or negative.
3	"Suspects" were reclassified as positive only if confirmed by the concurrent ELISA test.	Results were categorised as positive or negative.
4	"Suspects" were reclassified as negative.	Results were categorised as positive or negative.
5	"Suspects" were excluded from the analysis.	Results were categorised as positive or negative, excluding "suspects".

Scenarios 1, 2, 3, 4 and 5 were the scenarios considered for comparison of ELISA tests. Suspect results were reclassified as positive, negative or omitted as shown in the table.

The suspect result of an animal was reclassified as a positive outcome if the corresponding animal had at least one positive result in the remaining three ELISA tests or PCR tests; similarly, in the third scenario, the suspect result was considered positive if the concurrent ELISA result (Idexx serum ELISA) was positive. If the conditions in scenarios 2 and 3 were not met, the suspect result was reclassified as negative. In the fourth scenario, all suspect ELISA results were considered negative, while they were omitted in the fifth scenario.

As faecal PCR is a direct diagnostic test that targets the genetic component (IS900 and F57 genes) of MAP, it was considered a reference test to estimate the diagnostic sensitivity and specificity of the ELISA tests. Two true disease status scenarios were considered as a proxy of true disease status. In the first true disease status scenario, an animal was considered a true positive if the faecal PCR result (irrespective of the ELISA outcome) was either positive or strongly positive; otherwise, it was negative. In the second true disease status scenario, an animal was considered a true positive if its two serum ELISA tests (Idexx and IDVet) were positive, else negative. We acknowledge that the estimated sensitivity and specificity in the present study were not true sensitivity and specificity but were relative to the proxy scenarios considered. However, for simplicity, they were referred to as sensitivity and specificity.

2.4. Statistical analysis

After data editing (deleting duplicates), 930 test scores for IDVet and Idexx milk tests, 935 for IDVet serum tests, 1121 for Idexx serum tests, and 144 test scores for faecal PCR tests were included in the analyses. Data editing and analyses were carried out using Microsoft Excel version 2408 and SAS version 9.4 software, respectively. The agreement analysis between the ELISA and/or faecal PCR statuses by Kappa statistics [12], and the significance level of discordant pairs by McNemar's test [13] or Bowker's symmetry test [14] were performed using the agree function in the SAS version 9.4 software (SAS Institute Inc., Cary, NC, USA). The interpretation of the Kappa values [15] was carried out as follows: <0 (poor agreement), 0.01–0.20 (slight agreement), 0.21–0.40 (fair agreement), 0.41–0.60 (moderate agreement), 0.61–0.80 (substantial agreement), and 0.81–1.00 (almost perfect agreement). The diagnostic sensitivity and specificity of ELISA tests were estimated along with their confidence limits by the Clopper-Pearson method using a web tool called Epitools (https://epitools.ausvet.com.au/) [16].

3. Results

The names of the kits in the results were anonymised (kit A and kit B) to avoid conflicts. The disease classification of the animals based on each ELISA test and faecal PCR test is shown in Table 3. The within-herd apparent prevalence of Johne's disease in this study was less than or equal to 10 % using the two commercial ELISA tests (Supplementary Tables 1–4). The MAP-infection status of herds using faecal PCR was given in Supplementary Table 5. All the cross-tabulations used to estimate the kappa statistics values for different scenarios considered in this study are given in Supplementary Tables 6–59.

3.1. Kappa statistics

The Kappa statistics, with 95 % confidence intervals, from the agreement analysis between the ELISA tests for the five different scenarios are presented in Table 4. When scenarios were considered as a comparison, scenarios three (k = 0.79, CI: 0.70–0.88) and five (k = 0.77, CI: 0.66–0.87) had higher Kappa estimates than the first scenario for within-kit comparison of Kit A (k = 0.55, CI: 0.46–0.65).

For the first scenario, the within-kit (milk vs. serum) comparison of Kit B (k = 0.79, CI: 0.71–0.88) had higher Kappa estimates than that of Kit A (k = 0.55, CI: 0.46–0.65). Similarly, for scenarios two to five, the

Table 3
ELISA and faecal PCR outcomes of animals based on manufacturer's guidelines.

Test	Test outcomes			Total
	Negative	Suspect	Positive	
IDVet milk	880	NA	50	930
IDVet serum	885	4	46	935
Idexx milk	836	38	56	930
Idexx serum	1072	7	42	1121
Faecal PCR	72	32	40 (32 positive + 8 strong positive)	144

NA: As per the manufacturer's guidelines, IDVet milk ELISA does not have a suspect outcome classification.

Table 4
Agreement measures and symmetry tests for within and between enzyme-linked immunosorbent assay (ELISA) tests.

Comparison type	Kappa coefficients with 95 % CI (in parentheses) and significance value of symmetry tests (in italics)				
	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Kit A milk vs.	0.55 (0.46–0.65)	0.72 (0.62–0.81)	0.79 (0.70–0.88)	0.72 (0.62–0.83)	0.77 (0.66–0.87)
Kit A serum	*p < 0.0001*	*p < 0.0001*	*p = 0.0003*	*p = 0.0027*	*p = 0.0016*
Kit B milk vs.	0.79 (0.71–0.88)	0.84 (0.76–0.92)	0.84 (0.76–0.92)	0.83 (0.75–0.91)	0.86 (0.77–0.95)
Kit B serum	*p = 0.5473*	*p = 0.2850*	*p = 0.2850*	*p = 0.1967*	*p = 0.32*
Kit A milk vs.	0.59 (0.50–0.68)	0.78 (0.70–0.87)	0.78 (0.69–0.86)	0.76 (0.67–0.85)	0.82 (0.73–0.91)
Kit B milk	*p < 0.0001*	*p = 0.0002*	*p = 0.0606*	*p = 0.2207*	*p = 0.0075*
Kit A serum vs.	0.84 (0.76–0.92)	0.91 (0.85–0.97)	0.90 (0.83–0.96)	0.88 (0.81–0.95)	0.94 (0.88–1.00)
Kit B serum	*p = 0.3417*	*p = 0.4795*	*p = 0.3173*	*p = 0.2059*	*p = 1.0000*
Kit A milk vs.	0.55 (0.45–0.65)	0.70 (0.60–0.80)	0.74 (0.65–0.84)	0.71 (0.61–0.81)	0.74 (0.63–0.85)
Kit B serum	*p < 0.0001*	*p = 0.0002*	*p = 0.0093*	*p = 0.0343*	*p = 0.0029*
Kit B milk vs.	0.75 (0.65–0.84)	0.82 (0.73–0.90)	0.80 (0.71–0.89)	0.78 (0.68–0.87)	0.86 (0.77–0.95)
Kit A serum	*p = 0.2153*	*p = 0.1336*	*p = 0.0896*	*p = 0.0389*	*p = 0.3173*

Scenarios 1, 2, 3, 4 and 5 were the scenarios considered for comparison of ELISA tests. In Scenario 1, the trichotomous outcomes for each ELISA test were considered as such. In scenario 2, the ELISA suspect result for an animal was reclassified as a positive outcome if a corresponding animal had at least one positive in the remaining three ELISA tests or PCR test (includes strong positive outcome); similarly, in scenario 3, ELISA suspect result was considered positive if the concurrent ELISA result was positive. In scenario 4, all suspect ELISA results were considered negative, while they were omitted in scenario 5.

point estimate values of Kappa coefficients were higher for the within-kit comparison of Kit B (k = 0.83–0.86) than those of Kit A (k = 0.72–0.79).

Likewise, for the first scenario, the within-sample (e.g., Kit A milk vs. Kit B milk) comparison showed a higher Kappa estimate for serum (k = 0.84, CI: 0.76–0.92) samples than that of milk samples (k = 0.59, CI: 0.50–0.68). Similarly, for the fifth scenario, the within-sample comparison showed a higher Kappa estimate for serum (k = 0.94, CI: 0.88–1.00) samples than that of milk samples (k = 0.82, CI: 0.73–0.91).

The Kappa statistics, with the 95 % confidence intervals, for comparisons between ELISA tests and faecal PCR and milk ELISAs with serum true disease status scenarios are shown in Table 5. When the four ELISA test results were compared to faecal PCR results (dichotomous scenarios), the point estimates of kappa values showed either moderate (k = 0.43–0.59) or substantial agreement (k = 0.62–0.74) between them. When the two milk ELISA results were compared with the true disease status serum scenario (dichotomous scenarios), Kit A milk ELISA had point estimate kappa values of substantial agreement (k = 0.69–0.76), while Kit B milk ELISA had point estimate kappa values of substantial or perfect agreement (k = 0.79–0.87).

Table 5
Agreement measures and symmetry tests for between enzyme-linked immunosorbent assay (ELISA) tests and pseudo gold standard scenarios.

Comparison type	Kappa coefficients with 95 % CI (in parentheses) and significance value of symmetry tests (in italics)			
	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Kit A milk vs. PCR	0.62 (0.48–0.77) *p = 0.0495*	0.51 (0.35–0.67) *p = 0.8474*	0.43 (0.26–0.60) *p = 0.1441*	0.49 (0.27–0.71) *p = 0.0707*
Kit A serum vs. PCR	0.66 (0.53–0.79) *p = 0.0495*	0.64 (0.51–0.78) *p = 0.033*	0.55 (0.40–0.71) *p = 0.8415*	0.74 (0.59–0.89) *p = 0.5271*
Kit B milk vs. PCR	0.59 (0.44–0.74) *p = 0.2008*	0.59 (0.44–0.74) *p = 0.2008*	0.54 (0.37–0.70) *p = 0.4142*	0.66 (0.47–0.85) *p =1.0000*
Kit B serum vs. PCR	0.63 (0.49–0.78) *p = 0.3711*	0.63 (0.49–0.78) *p = 0.3711*	0.54 (0.38–0.70) *p = 0.2207*	0.63 (0.44–0.83) *p =0.763*
Kit A milk vs. Serum true status[a]	0.69 (0.59–0.79) *p < 0.0001*	0.76 (0.67–0.86) *p < 0.0001*	0.72 (0.61–0.82) *p = 0.0007*	0.75 (0.64–0.85) *p = 0.001*
Kit B milk vs. Serum true status[a]	0.84 (0.75–0.92) *p = 0.0325*	0.82 (0.73–0.91) *p = 0.0201*	0.79 (0.70–0.89) *p = 0.0076*	0.87 (0.79–0.96) *p = 0.1573*

Scenarios 2, 3, 4 and 5 were the scenarios considered for comparison of ELISA tests. In scenario 2, the ELISA suspect result for an animal was reclassified as a positive outcome if a corresponding animal had at least one positive in the remaining three ELISA tests or PCR test (includes strong positive outcome). Also, the PCR ELISA suspect result for an animal was reclassified as a positive outcome if a corresponding animal had at least one positive in the remaining three ELISA tests; similarly, in scenario 3, the ELISA suspect result was considered positive if the concurrent ELISA result was positive. Also, the PCR ELISA suspect result for an animal was reclassified as a positive outcome if a corresponding animal had at least one positive in the remaining three ELISA tests. In scenario 4, all suspect ELISA results were considered negative, while they were omitted in scenario 5.

[a] True status is positive if both the serum ELISAs (Idexx and IDVet) outcomes are positive.

3.2. Symmetry tests

The McNemar or Bowker symmetry test for contingency Table (2x2 or 3x3) revealed the significance of discordance between the tests (Tables 4 and 5). The within-kit comparison (milk vs. serum) of Kit A was significant ($p < 0.05$) for all five scenarios, while that for Kit B was non-significant for all five scenarios. Between-kit milk comparisons were significant ($p < 0.05$) for scenarios 1, 2, and 5, while those for serum were non-significant for all five scenarios.

When comparing the ELISA results with the faecal PCR results (dichotomous scenarios), all four ELISA tests (Kit A milk, Kit A serum, Kit B milk, and Kit B serum) were non-significant ($p > 0.05$), except Kit A serum (scenarios 2 and 3) and Kit A milk (scenario 2). When true disease status was defined by serum ELISA results of the two kits (dichotomous scenarios), both Kit A and Kit B (except the fifth scenario) milk ELISAs were significant ($p < 0.05$).

3.3. Sensitivity and specificity

Sensitivity and specificity analysis results are shown in Table 6. When faecal PCR results were considered to represent the true disease status (Table 6), all four ELISA tests had a point estimate of sensitivity ranging from 0.67 to 0.88 and specificity from 0.62 to 0.93. When the true status is defined by the results of both serum ELISAs (Table 6), both milk ELISAs exhibited high sensitivity (0.89–0.95) and specificity (0.97–0.99).

4. Discussion

The present study adopted the STRADAS-paraTB criteria, which are standards for reporting of animal diagnostic accuracy studies for paratuberculosis [17]. The primary aim of our study was to compare the agreement within and between two ELISA tests using Kappa statistics in the four dairy herds in New Zealand. Although the agreement for different scenario definitions varied in this study, the more practical approach of omitting the suspect cases in the analysis showed

Table 6
Estimated sensitivity and specificity of milk and serum enzyme-linked immunosorbent assay (ELISA) with their 95 % confidence intervals relative to faecal polymerase chain reaction (PCR) results and true status defined by both the serum ELISA.

PCR vs.	Scenario 2		Scenario 3		Scenario 4		Scenario 5	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp
Kit A milk	0.88 (0.75–0.95)	0.76 (0.63–0.86)	0.73 (0.59–0.85)	0.77 (0.65–0.87)	0.72 (0.55–0.85)	0.74 (0.62–0.83)	0.87 (0.69–0.96)	0.62 (0.42–0.79)
Kit A serum	0.70 (0.55–0.82)	0.93 (0.86–0.98)	0.68 (0.53–0.80)	0.93 (0.86–0.98)	0.67 (0.50–0.81)	0.88 (0.81–0.94)	0.81 (0.63–0.93)	0.93 (0.82–0.98)
Kit B milk	0.71 (0.57–0.83)	0.87 (0.76–0.94)	0.71 (0.57–0.83)	0.87 (0.76–0.94)	0.74 (0.58–0.87)	0.81 (0.70–0.89)	0.83 (0.65–0.94)	0.83 (0.64–0.94)
Kit B serum	0.76 (0.61–0.87)	0.87 (0.77–0.94)	0.76 (0.61–0.87)	0.87 (0.77–0.94)	0.76 (0.60–0.89)	0.80 (0.69–0.88)	0.83 (0.65–0.94)	0.80 (0.61–0.92)
True status defined by serum ELISAs[a]								
Kit A milk	0.93 (0.80–0.98)	0.97 (0.95–0.98)	0.95 (0.83–0.99)	0.98 (0.96–0.99)	0.89 (0.75–0.97)	0.98 (0.96–0.98)	0.94 (0.79–0.99)	0.98 (0.97–0.99)
Kit B milk	0.93 (0.80–0.98)	0.99 (0.98–0.99)	0.93 (0.80–0.98)	0.99 (0.98–0.99)	0.92 (0.79–0.98)	0.98 (0.97–0.99)	0.94 (0.79–0.99)	0.99 (0.98–1)

Se and Sp represent diagnostic sensitivity and specificity, respectively.
Scenarios 2, 3, 4 and 5 were the scenarios considered to estimate the sensitivity and specificity of milk and serum ELISAs. In scenario 2, the ELISA suspect result for an animal was reclassified as a positive outcome if a corresponding animal had at least one positive in the remaining three ELISA tests or PCR test (includes strong positive outcome); Also, PCR ELISA suspect result for an animal was reclassified as positive outcome if corresponding animal had at least one positive in the remaining three ELISA tests. Similarly, in scenario 3, the ELISA suspect result was considered positive if the concurrent ELISA result was positive. Also, the PCR ELISA suspect result for an animal was reclassified as a positive outcome if a corresponding animal had at least one positive in the remaining three ELISA tests. In scenario 4, all suspect ELISA results were considered negative, while they were omitted in scenario 5.
[a] True status is positive if both the serum ELISAs (Idexx and IDVet) outcomes are positive.

consistently higher point estimates of kappa coefficients, like other scenarios, i.e., when they were reclassified as positive or negative outcomes. However, it is advisable to retest the suspect outcomes for clinical or diagnostic purposes.

Further, higher kappa coefficients were observed for within-sample type comparison of the serum ELISAs than the milk ELISAs. This suggests that the two commercial ELISAs can be used interchangeably for serum-based screening of herds to make decisions in Johne's disease surveillance programmes in dairy cattle, especially in New Zealand. However, milk ELISAs had substantial (k = 0.76 to 0.78, scenarios 2 to 4) to almost perfect agreement (k = 0.82, scenario 5) in dichotomous classification. Thus, we propose that the cut-off points for milk ELISA tests could be standardised using an appropriate reference test (e.g., MAP shedding scores). When trichotomous classification was considered, the within-kit (milk vs. serum) comparison of Kit B (k = 0.79, CI: 0.71–0.88) had higher Kappa estimates than that of Kit A (k = 0.55, CI: 0.46–0.65). Considering these findings, Kit B ELISA has a slight advantage over Kit A ELISA for the diagnosis and/or screening of Johne's disease in New Zealand dairy cattle.

Another practical implication of our findings is that the observed agreement between the two ELISA kits provides insight into their potential combined use for diagnosing Johne's disease in dairy cattle. For example, consider the kappa coefficients when the suspect outcomes were omitted from the analysis. The comparison between kits showed almost perfect agreement for serum and milk samples, in which the former had the greater point estimates. These results suggest that parallel testing with serum samples could offer limited additional benefit relative to parallel testing using milk samples.

The observed slight differences in the agreement between the tests can be due to the following reasons: The four herds studied in the present study have different Johne's disease prevalences. A few animals may react strongly, while a few react only weakly, if at all, when they are exposed to MAP infections. This would have resulted in different disease spectra in the herds (animals), causing disparity in ELISA tests' outcomes, considering their varying diagnostic efficiencies [7,11]. Alternatively, the low prevalence herds may result in low positive predictive values of the ELISA test [9], influencing the agreement between them. Milk and serum samples react differently to ELISA tests [6], as observed in a not-so-perfect agreement between them, especially in the case of the trichotomous outcome of Kit A ELISA. As MAP has a prolonged incubation period, it is not known how equally the different ELISAs capture the different immune response phases of the host. This was also reflected in the results of the previous studies. A past study [11] compared the Idexx ELISA with the IDVet ELISA using the serum samples and found a kappa agreement of 0.84, which is in the range seen in the present study.

When faecal PCR results were compared with the ELISA results, the point estimates of kappa coefficients indicated either moderate or substantial agreement (0.43–0.74). These results were comparable but higher than the previous study [7] that found kappa values for faecal PCR with Idexx ELISA (k = 0.28) and IDVet ELISA (k = 0.07). The discrepancy between the faecal PCR and the ELISA outcomes may be because these two tests would be effective at different Johne's disease stages, depending on the age and immune spectrum of the host. This is especially true as ELISAs do not detect most early disease stages [4,5]. We acknowledge that this may be one of the reasons that qPCR as a reference test would underestimate the potential of ELISA tests. Alternatively, ELISA and faecal PCR could have higher specificities that might vary across different infected populations, resulting in lower kappa values [18]. These results suggest that ELISA and faecal PCR tests might complement the different disease spectra of paratuberculosis. This practical application was demonstrated by a study on New Zealand dairy cattle, where the interpretation of four ELISAs in parallel, followed by a confirmatory faecal PCR test recorded high specificity but a moderate sensitivity across herds with different prevalences [19].

The sample and test of choice for disease screening should be based on the diagnostic accuracy of the respective tests. Since Johne's disease has no gold standard test [3], estimating the diagnostic sensitivity and specificity of indirect diagnostic tests, such as ELISAs, against direct diagnostic tests, such as faecal culture or PCR, should be approached cautiously. These reference tests may potentially have false outcomes, resulting in miscalculations of the diagnostic accuracies of ELISA tests. For this reason, the estimated sensitivity and specificity of ELISAs in the present should be considered as rough estimates rather than the true estimates *per se*. For instance, if faecal PCR had resulted in false negative outcomes, the sensitivity of the ELISA tests would have been underestimated. Moreover, New Zealand has no Johne's disease certification for trading animals. Thus, it is hard to select the disease-free herd for the specificity estimation of diagnostic tests. A similar situation of inclusion of MAP suspect herds to estimate the diagnostic specificity of a test was also observed in a past study on Johne's disease on French Sheep [20]. Further, large-scale application of high-specificity tests (e.g., 99.5 %) can still yield considerable false positives in paratuberculosis disease-free farms, prompting further investigations [20]. Alternatively, in infected farms, false positives are less problematic due to higher predictive value and less need for further confirmation [20]. Nevertheless, the estimated sensitivity and specificity of the ELISA tests relative to the faecal PCR test in the present study are comparable to the sensitivity of milk (0.21–0.61) and serum (0.24–0.94) and the specificity of milk (0.83–0.99) and serum (0.40–1.00) ELISA, respectively, for infectious cattle reported in a review [21].

It should be noted that methods such as Bayesian latent class models (BLMCs) can serve as an alternative in the absence of a perfect reference test for Johne's disease [19]. These BLMCs are sensitive to test independence, different prevalences, and sensitivity and specificity across target populations to achieve a better estimate of diagnostic accuracy [19,20,22]. We assumed that the ELISA tests in our study had conditional dependencies, and our target population had no sharp contrast in disease prevalence. Moreover, the less informative prior information available for these ELISA tests, especially in the context of New Zealand dairy cattle, might be sensitive to the BLMCs, as seen in sensitivity estimates in a more recent paratuberculosis study from Saudi Arabia [23]. Besides, the low to moderate disease prevalence might affect the sensitivity of the diagnostic test in BLMCs [19]. We acknowledge these as a limitation in our data, which could form the basis for another study. Nevertheless, sensitivity and specificity estimates of ELISA tests relative to the faecal PCR test in the present study are comparable to the estimates found by Bayesian models in past studies on New Zealand dairy cattle [19,24]. One such study found that when a faecal qPCR cutoff was set at $\geq 1 \times 10^4$ genomes/mL, the Idexx serum ELISA had a median value of sensitivity estimates of 0.36–0.51 and specificity of 0.96–1.00 in conditionally independent (fixed) and dependent (fixed and random) BLMCs [19].

The major limitation of this study was the selection of herds. We could not account for disease prevalence, disease spectra, and geographies. This herd limitation has an impact on the study outcomes and thus test agreement here is context-dependent. For example, based on all the ELISA test results, the within-herd prevalence in this study ranged from 1 to 10 %. Assuming ELISA tests have low sensitivity, especially in low-prevalence herds (e.g., Herd D), the uncertainty increases, and agreement between tests decreases. This was the case in another paratuberculosis study on cattle that compared the ELISA test with the faecal culture test [25]. That study demonstrated that within-herd prevalence affected the performance of the ELISA tests, in which prevalences of <5 % and ≥21 % had lower agreement than that of the intermediary prevalences. Thus, the herds selected could have a limitation in disease prevalence and stages, which might influence the agreement between and within the ELISA tests. Furthermore, the herds represented in this study are limited to the North Island of New Zealand. We suggest future studies to account for these limitations. Nevertheless, Johne's disease is endemic among dairy cattle in New Zealand, with a >40 % herd-level prevalence and within-herd prevalence of 10–15 % [26]. Moreover, as of the 2023/24 production season [27], the most common herd size in New Zealand is 200–249 (11.9 % of herds) cows, followed by 250–299 cows (10.2 % of herds), which is comparable to the herd size in our study with similar production systems. Thus, we believe that although we had a limitation in the selection of herds in our study, the overall results of the within- and between-ELISA tests comparison can be useful if extrapolated with a contextual basis. Another limitation of this study is the usage of a limited number of samples for the faecal PCR test. Additionally, as the herds in the study are not apparently free from Johne's disease, we suggest cautious interpretation of the diagnostic accuracy of the ELISA tests.

5. Conclusions

Because of the almost perfect agreement, the two commercial kits can be used interchangeably for serum-based screening of herds in Johne's disease control programmes. However, milk-based ELISAs need to be standardised for cutoff values (especially Kit A milk ELISA) to increase the agreement between the two commercial kits prior to their preferential use over serum tests. In the intra-kit (serum vs. milk) agreement analysis, Kit B performed better than Kit A, indicating that Kit B serum and milk ELISA outcomes are more identical *per se* than the Kit A ELISA outcomes in the trichotomous classification. When the goal of the diagnostic test is the detection of the MAP-positive cases, milk ELISA, compared to serum ELISA, could be a better option, as it is a more cost-effective option considering their overall performance in this study. Future studies need to validate the results of the present study using different disease definitions, age and immune spectrum of the host, and a larger sample size of faecal PCR/culture results.

CRediT authorship contribution statement

Venkatesh K.M.: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nicolas Lopez-Villalobos:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization. **Sandeep K. Gupta:** Writing – review & editing, Validation, Supervision. **Garry B. Udy:** Writing – review & editing, Validation, Supervision. **Richard Laven:** Writing – review & editing, Resources, Methodology. **Shih-Jiuan Chiu:** Writing – review & editing, Methodology. **Piyush Bugde:** Writing – review & editing, Resources, Methodology. **Yoichi Furuya:** Writing – review & editing, Resources, Methodology. **Venkata Sayoji Rao Dukkipati:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Ethics approval

The ethical permission for this study is approved by the Animal Ethics Committee, Massey University, New Zealand, under protocol number 22/42.

Funding sources

This study was funded by Massey Ventures Ltd, Palmerston North, New Zealand, and Pictor Ltd, Auckland, New Zealand.

Declaration of competing interest

The authors declare that there is no conflict of interest.

Acknowledgements

The first author acknowledges the Doctoral Scholarship and Pūtea Tautoko Doctoral Financial Support Grant provided by Massey University, New Zealand. The authors would like to thank the farmers for providing access to samples for this study. We thank the two anonymous reviewers for their comments that helped us to improve the quality of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tube.2025.102679.

Data availability

The data of the analyses in the present study are given in the form of a cross-tabulation in the supplementary file. ELISA scores (data) are available from the corresponding author with a reasonable request.

References

[1] Barkema HW, Orsel K, Nielsen SS, Koets AP, Rutten VPMG, Bannantine JP, et al. Knowledge gaps that hamper prevention and control of Mycobacterium avium subspecies paratuberculosis infection. Transbound Emerg Dis 2018;65:125–48.
[2] Rasmussen P, Barkema HW, Mason S, Beaulieu E, Hall DC. Economic losses due to Johne's disease (paratuberculosis) in dairy cattle. J Dairy Sci 2021;104:3123–43.
[3] Field NL, McAloon CG, Gavey L, Mee JF. Mycobacterium avium subspecies paratuberculosis infection in cattle – a review in the context of seasonal pasture-based dairy herds. Ir Vet J 2022;75:12.

[4] Whitlock RH, Buergelt C. Preclinical and clinical manifestations of paratuberculosis (including pathology). Vet Clin North Am Food Anim Pract 1996;12:345–56.
[5] Whitlock RH, Wells SJ, Sweeney RW, Van Tiem J. ELISA and fecal culture for paratuberculosis (Johne's disease): sensitivity and specificity of each method. Vet Microbiol 2000;77:387–98.
[6] Sharma K, Sharma S, Dhanda S, Bangar Y, Kumar N, Chaubey KK. Meta-analysis of prevalence of paratuberculosis in cattle using published estimates under serum and milk ELISA. Res Vet Sci 2024;178:105366.
[7] Park H-T, Ha S, Park H-E, Shim S, Hur TY, Yoo HS. Comparative analysis of serological tests and fecal detection in the diagnosis of Mycobacterium avium subspecies paratuberculosis infection. Korean J Vet Res 2020;60:117–22.
[8] Singh SV, Singh AV, Singh PK, Sohal JS, Singh NP. Evaluation of an indigenous ELISA for diagnosis of Johne's disease and its comparison with commercial kits. Indian J Microbiol 2007;47:251–8.
[9] McKenna SLB, Barkema HW, Keefe GP, Sockett DC. Agreement between three ELISAs for Mycobacterium avium subsp. paratuberculosis in dairy cattle. Vet Microbiol 2006;114:285–91.
[10] Collins MT, Wells SJ, Petrini KR, Collins JE, Schultz RD, Whitlock RH. Evaluation of five antibody detection tests for diagnosis of bovine paratuberculosis. Clin Vaccine Immunol 2005;12:685–92.
[11] Fry MP, Kruze J, Collins MT. Evaluation of four commercial enzyme-linked immunosorbent assays for the diagnosis of bovine paratuberculosis in Chilean dairy herds. J Vet Diagn Invest 2008;20:329–32.
[12] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960; 20:37–46.
[13] McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12:153–7.
[14] Bowker AH. A test for symmetry in contingency tables. J Am Stat Assoc 1948;43: 572–4.
[15] Thrusfield M. Diagnostic testing. Vet. Epidemiol. fourth ed. Oxford: John Wiley and Sons Ltd.; 2018. p. 421–56.
[16] Sergeant ESG. Epitools Epidemiological Calculators. Ausvet. Available at: 2018. http://epitools.ausvet.com.au. (accessed October 12, 2024).
[17] Gardner IA, Nielsen SS, Whittington RJ, Collins MT, Bakker D, Harris B, et al. Consensus-based reporting standards for diagnostic test accuracy studies for paratuberculosis in ruminants. Prev Vet Med 2011;101:18–34.
[18] Pinedo PJ, Rae DO, Williams JE, Donovan GA, Melendez P, Buergelt CD. Association among results of serum ELISA, faecal culture and nested PCR on milk, blood and faeces for the detection of paratuberculosis in dairy cows. Transbound Emerg Dis 2008;55:125–33.
[19] Bates A, Laven R, O'Brien R, Liggett S, Griffin F. Estimation of the sensitivity and specificity of four serum ELISA and one fecal PCR for diagnosis of paratuberculosis in adult dairy cattle in New Zealand using Bayesian latent class analysis. Prev Vet Med 2020;185:105199.
[20] Mathevon Y, Foucras G, Falguières R, Corbiere F. Estimation of the sensitivity and specificity of two serum ELISAs and one fecal qPCR for diagnosis of paratuberculosis in sub-clinically infected young-adult French sheep using latent class Bayesian modeling. BMC Vet Res 2017;13:1–11.
[21] Nielsen SS, Toft N. Ante mortem diagnosis of paratuberculosis: a review of accuracies of ELISA, interferon-γ assay and faecal culture techniques. Vet Microbiol 2008;129:217–35.
[22] Kostoulas P, Leontides L, Enøe C, Billinis C, Florou M, Sofia M. Bayesian estimation of sensitivity and specificity of serum ELISA and faecal culture for diagnosis of paratuberculosis in Greek dairy sheep and goats. Prev Vet Med 2006;76:56–73.
[23] Elsohaby I, Kostoulas P, Fayez M, Elmoslemany A, Alkafafy ME, Bahhary AM, et al. Bayesian estimation of diagnostic accuracy of fecal smears, fecal PCR and serum ELISA for detecting Mycobacterium avium subsp. paratuberculosis infections in four domestic ruminant species in Saudi Arabia. Vet Microbiol 2025;301:110377.
[24] Norton S, Johnson WO, Jones G, Heuer C. Evaluation of diagnostic tests for Johne's disease (Mycobacterium avium subspecies paratuberculosis) in New Zealand dairy cows. J Vet Diagn Invest 2010;22:341–51.
[25] Donat K, Schlotter K, Erhardt G, Brandt HR. Prevalence of paratuberculosis in cattle and control measures within the herd influence the performance of ELISA tests. Vet Rec 2014;174:119.
[26] Whittington R, Donat K, Weber MF, Kelton D, Nielsen SS, Eisenberg S, et al. Control of paratuberculosis: who, why and how. A review of 48 countries. BMC Vet Res 2019;15:1–29.
[27] DairyNZ. New Zealand dairy statistics 2023-24. DairyNZ; 2024. p. 1–63. https://www.dairynz.co.nz/resources/resource-list/new-zealand-dairy-statistics-2023-24/. [Accessed 12 December 2024].