

Evaluation of diagnostic tests for Johne's disease (*Mycobacterium avium* subspecies *paratuberculosis*) in New Zealand dairy cows

Solis Norton,¹ Wesley O. Johnson, Geoff Jones, Cord Heuer

Abstract. To quantify the sensitivity and specificity of a serum enzyme-linked immunosorbent assay (ELISA) and fecal culture (FC) tests and to estimate the prevalence of Johne's disease (JD) in New Zealand dairy herds using Bayesian methods, 4 New Zealand dairy herds were tested simultaneously by ELISA and FC 5 times over 3 lactations. Test results were dichotomized. A Bayesian regression model was developed that considered test sensitivity as a function of the covariates parity, lactation stage, and prevalence of JD, which is expected to vary between herds. It was applied to a cross-sectional subset of the data and the full, repeated measures data set. Results were compared with frequentist pseudo gold standard results of the full data. Using the regression model, sensitivity of the ELISA was higher in older animals, but the sensitivity of the FC test showed no trend across age groups. Both FC and ELISA sensitivity were lower in late lactation. Estimated prevalence was lower and FC sensitivity higher when analyzing the complete data. The regression model enabled a more accurate diagnosis of JD to be made because it incorporated cow-specific information in the diagnosis, such as age and lactation stage. The model also enabled the incorporation of previous test results for an individual when diagnosing disease. The trends in results from the regression model support the current understanding of the disease process. The advantage of repeated testing of individuals in the assessment of test performance is discussed in the current study.

Key words: Bayes; cattle; dairy; gold standard; Johne's disease; New Zealand; regression.

Introduction

Johne's disease (JD), caused by *Mycobacterium avium* subspecies *paratuberculosis* (MAP) is a chronic enteritis of ruminant species. It is of particular importance in intensive livestock industries including the dairy industry. In cattle, the disease is characterized by infection during calthood followed by a variable latent period, often of several years. Progressive disease usually follows, demonstrated by worsening diarrhea, weight loss, and a sharp drop in milk production.⁶ The disease has no treatment and is normally fatal. The latent period of infection is a contributing factor in the poor sensitivity of diagnostic tests for JD. It has led to the commonly held dogma that for every cow with clinical signs of JD born in the herd, there are probably 15–20 subclinically infected herd mates.³⁶

A substantial research effort has been dedicated to the improvement and validation of diagnostic tests

for JD.^{2,7,9,15,23,24,26,27,31,32} Research has focused on the tests most commonly used: the enzyme-linked immunosorbent assay (ELISA) for serum antibody and culture of MAP from the feces. However, 2 challenges to evaluating diagnostic tests for JD exist that cannot easily be overcome by sampling methods or laboratory technology.

The first challenge is the absence of a gold standard (i.e., the true infection status of a tested individual remains unknown). Postmortem examination is currently the most accurate method of determining infection status, but it is seldom feasible to conduct postmortem examinations on the large number of trial animals required to validate the results of a diagnostic test. As a result, misclassification rates of JD tests under field conditions remain difficult to determine. One possible, but time-consuming and expensive, alternative is to test an animal for JD throughout its lifetime, which allows infection status to be estimated with greater certainty. ELISA test performance has previously been validated against the fecal culture (FC) test in lieu of a gold standard.^{9,10,34} Since FC has imperfect sensitivity and specificity, this procedure gives biased estimates of ELISA sensitivity and specificity.

The second challenge is that test sensitivity varies with severity of the disease and with lactation stage.²⁵ In general, it is rare for an antibody response or fecal

From the Centre for the Study of Agriculture, Food, and Environment, University of Otago, Dunedin, New Zealand (Norton); the Department of Statistics, University of California, Irvine, CA (Johnson); and the Department of Statistics, Institute of Fundamental Sciences (Jones), and the EPICentre, Institute of Veterinary, Animal, and Biomedical Sciences (Heuer), Massey University, Palmerston North, New Zealand.

¹ Corresponding Author: Solis Norton, Centre for the Study of Agriculture, Food, and Environment, University of Otago, PO Box 56, Dunedin 9054, New Zealand. solis.norton@otago.ac.nz

shedding of MAP to be detectable in the early, preclinical state of disease. During more advanced and clinical states of disease, animals develop high levels of antibody and shed large numbers of MAP in their feces. Reports of sensitivity for the antibody ELISA test ranged from 15.4% to 88.1% and were positively correlated with the progression of disease.¹¹

Bayesian probability models are well suited to address diagnostic test evaluation, and there is a growing body of research in veterinary science describing their application. Initially, a previous study²⁰ developed a Bayesian approach to the model described by Hui and Walter in a previous study¹⁹ in which the performance of 2 independent diagnostic tests was estimated in the absence of a gold standard, using 2 populations each with a distinct prevalence. Subsequently, 2 approaches were developed to adjust for dependence in the 2-test model, one in 2001¹² and another in 2003.^{3,17}

A Bayesian regression model for estimating the performance of diagnostic tests that incorporates cow-level covariates parity and lactation stage, and allows test performance to vary as a function of these covariates, is presented in the current study. The model was applied to both cross-sectional and repeat measure data. Results from a pseudo gold standard (PGS) approach are also presented for comparison.

Materials and methods

Study herds

A longitudinal population study was conducted using a convenience sample of 4 dairy herds with a history of clinical JD in the Manawatu district of North Island, New Zealand (40° 19' north and 175° 30' east). The herds were coded A–D. Herd managers consented to a request for participation, and the herds were considered typical examples of New Zealand's seasonal calving, pasture-based dairy production system.

Data collection

The data were collected from October 2000 to October 2002. Unique animal identification was possible using the dairy herd improvement (DHI)-type monitoring services provided by Livestock Improvement Corporation. Herds were tested for JD twice per milking season, once in early lactation (October) and again in late lactation (May). Three herds were tested 5 times, and 1 herd was tested 3 times during the study.

At each herd visit, a sample of whole blood was taken by caudal venipuncture and a fecal sample collected. The simultaneous collection of a blood sample and fecal sample was defined as a single test-event. All milking cows in herds A–C were sampled at each visit. These herds milked approximately 100 cows. In herd D, which milked about 500 cows, samples were collected from a subset of approximately 100 cows. When culled, members of this subset were replaced with herd mates by random selection

within age group, so that the age distribution of the sampled fraction continued to represent that of the entire herd.

Serology was conducted by a commercial veterinary laboratory using a commercial Johnes's absorbed ELISA kit with an absorption step using soluble *M. phlei* antigen, an in-house antigen, and a single cut-off value recommended by the kit.^a The antigen preparation methodology has been previously described.²⁹ In brief, a New Zealand isolate of the *M. avium* complex, not containing IS 900 (Wag 1001), *M. avium* strain 18, formerly identified as *M. paratuberculosis*,⁵ and *M. paratuberculosis* Teps were used for antigen production. Bacteria were grown on modified Watson–Reid medium and washed in distilled water by centrifugation prior to disruption by sonication and further centrifugation. Culture of feces was conducted by AgResearch Wallaceville Animal Research Centre (Upper Hutt, New Zealand) using the technique described in a previous study.⁸

Data structure

The data had a hierarchical structure with 3 levels. The uppermost level was herd, followed by individual cow, and the lowest level was test-event repeated within cow. A subset of the data was taken to create a cross-sectional data set, which contained only the first test event (simultaneous ELISA and FC) for each cow.

Methods for evaluating the performance of diagnostic tests

A summary of the 3 methods used is presented in Table 1. A detailed description of each method follows.

Method 1: Cross-sectional. In the Hui and Walter model,¹⁹ the data are assumed to come from at least 2 distinct populations, each with a distinct prevalence. In the current case, the populations were the 4 herds. Test accuracy was assumed to be constant across populations. The data then consist of a 2×2 contingency table for each population (Table 2). The cell counts corresponding to the i th row and j th column for population k are modeled with independent multinomial distributions, where

$$y_k \sim \text{multinomial}[n_k(p_{11k}, p_{12k}, p_{21k}, p_{22k})], \quad (1)$$

and

$$\begin{aligned} p_{11k} &= P_k(T_1+, T_2+) = p_{i_k} \times Se_E \times Se_F \\ &\quad + (1 - p_{i_k}) \times (1 - Sp_E) \times (1 - Sp_F) \\ p_{12k} &= P_k(T_1+, T_2-) = p_{i_k} \times Se_E \times (1 - Se_F) \\ &\quad + (1 - p_{i_k}) \times (1 - Sp_E) \times Sp_F \\ p_{21k} &= P_k(T_1-, T_2+) = p_{i_k} \times (1 - Se_E) \times Se_F \\ &\quad + (1 - p_{i_k}) \times Sp_E \times (1 - Sp_F) \\ p_{22k} &= P_k(T_1-, T_2-) = p_{i_k} \times (1 - Se_E) \\ &\quad \times (1 - Se_F) + (1 - p_{i_k}) \times Sp_E \times Sp_F \\ k &= 1, 2, 3, 4, \end{aligned}$$

Table 1. Description of statistical methods used to evaluate the performance of the enzyme-linked immunosorbent assay and fecal culture tests for Johne's disease.

	Method		
	1	2	3
Statistical approach	Bayesian model*	Bayesian model*	Frequentist
Data	Cross sectional	Repeated measures	Cross-sectional
Data set size ($n =$)	779	2,046	779
Sensitivity modeled as function of covariates	Yes	Yes	No

* WinBUGS code included in Appendix 1.

where $Se_E =$ ELISA sensitivity, $Se_F =$ FC sensitivity, $Sp_E =$ ELISA specificity, $Sp_F =$ FC specificity, $T_1 =$ ELISA, $T_2 =$ FC, and $pi_k =$ prevalence in population k .

This approach can be extended to include covariate effects on sensitivities and/or specificities by allowing these parameters to be functions of the covariates. Equation 1 is then replaced by a cow-level equivalent, where the diagnostic test outcome for ELISA and FC (p_{11} [+ , +], p_{12} [+ , -], p_{21} [- , +], or p_{22} [- , -]) for cow j is multinomial with $n = 1$ and corresponding probabilities:

$$\begin{aligned}
 p_{11j} &= Z_j Se_E[j] \times Se_F[j] + (1 - Z_j) \times (1 - Sp_E[j]) \times (1 - Sp_F[j]) \\
 p_{12j} &= Z_j Se_E[j] \times (1 - Se_F[j]) + (1 - Z_j) \times (1 - Sp_E[j]) \times Sp_F[j] \\
 p_{21j} &= Z_j (1 - Se_E[j]) \times Se_F[j] + (1 - Z_j) \times Sp_E[j] \times (1 - Sp_F[j]) \\
 p_{22j} &= Z_j (1 - Se_E[j]) \times (1 - Se_F[j]) + (1 - Z_j) \times Sp_E[j] \times Sp_F[j]
 \end{aligned}
 \tag{2}$$

and the infection status Z_j of cow j has a Bernoulli distribution (i.e., if cow j is from herd k then $P(Z_j = 1) = pi_k$).

In this analysis, the average of the individual sensitivities $Se_E[j]$ and $Se_F[j]$ was monitored to give estimates of overall sensitivity for each test, but the sensitivity of each test was also modeled as a function of the covariates parity and lactation stage. The prevalence, on the other hand, was modeled only as a function of herd. It was assumed that the 2 tests were conditionally independent given the disease status of the animal and that test specificity would not vary as a function of covariates.

Sensitivity of the ELISA and the FC test were subsequently modeled on the logit scale for each cow. The covariate parity, categorized as 1, 2, 3, 4, and >4, was

included as a proxy for age. The covariate lactation stage was categorized as either early or late. It was included to represent stress levels on the cow associated with parturition and the beginning of lactation. The sensitivities for each test and cow were modeled as

$$\begin{aligned}
 \text{logit}(Se_E[j]) &\leftarrow g(\text{parity}[j]) + g[6] \times x[j] \\
 \text{logit}(Se_F[j]) &\leftarrow h(\text{parity}[j]) + h[6] \times x[j],
 \end{aligned}
 \tag{3}$$

where j denotes the j th cow in the data, $\text{parity}[j]$ its parity, and $x[j]$ its lactation stage. In this parameterization, 6 coefficients are used; 1 for the sensitivity at each parity level, in late lactation, and 1 for the effect of lactation stage (Table 3). An informative prior was used for the regression coefficients based on a previously described technique.¹ This technique requires provision of scientific input that is independent of the data about the sensitivity of each test under 6 different parity and lactation stage combinations. For example, in the current study, for a cow in parity 1 and late lactation, the best prior estimate of ELISA sensitivity was 20%, with 95% certainty that it was less than 75%. On the other hand, the best guess for the sensitivity for a late-lactation cow in parity 4 was 50%, with 95% certainty that it was less than 85%, so it was believed there might be a 30% increase in sensitivity for parity 4 over parity 1 in late-lactation cows. Please note that a great deal of uncertainty was attached to these estimates. Observe that, with this parameterization, the effect on sensitivity associated with changing from early to late lactation was modeled to be the same, on the logit scale, across all parities.

Table 2. Cross-tabulated results for an enzyme-linked immunosorbent assay (ELISA) and fecal culture (FC) test applied to 779 dairy cows from 4 New Zealand dairy herds.*

ELISA	Herd A		Herd B		Herd C		Herd D	
	FC+	FC-	FC+	FC-	FC+	FC-	FC+	FC-
+	0	13	4	1	3	0	3	0
-	7	111	5	132	9	223	5	263
Total	131		142		235		271	

* + = positive; - = negative.

Table 3. Prior information for the sensitivity of the enzyme-linked immunosorbent assay (ELISA) and fecal culture (FC) test for Johne’s disease at each parity, for early and late lactation, and for herd level prevalence.

Parameter	Prior estimate	95th percentile	Distribution
ELISA sensitivity in late lactation			
Parity 1	0.20	<0.75	dbeta(1.38, 2.51)
Parity 2	0.30	<0.85	dbeta(1.33, 1.78)
Parity 3	0.50	<0.85	dbeta(2.23, 2.23)
Parity 4	0.50	<0.85	dbeta(2.23, 2.23)
Parity >4	0.40	<0.85	dbeta(1.62, 1.94)
ELISA sensitivity, parity 1, early lactation	0.23	<0.75	dbeta(1.48, 2.59)
FC sensitivity in late lactation			
Parity 1	0.25	<0.80	dbeta(1.38, 2.15)
Parity 2	0.35	<0.85	dbeta(1.46, 1.85)
Parity 3	0.55	<0.90	dbeta(1.75, 1.62)
Parity 4	0.55	<0.90	dbeta(1.75, 1.62)
Parity >4	0.45	<0.85	dbeta(1.87, 2.06)
FC sensitivity in parity 1, early lactation	0.28	<0.80	dbeta(1.47, 2.21)
Herd level prevalence			
Herd A	0.15	<0.60	dbeta(1.55, 4.12)
Herd B	0.15	<0.60	dbeta(1.55, 4.12)
Herd C	0.06	<0.50	dbeta(1.25, 4.88)
Herd D	0.06	<0.50	dbeta(1.25, 4.88)

Priors describing ELISA sensitivity in different parity groups were represented by only slightly informative beta distributions after considering previous studies.^{14,15,18,36} Priors for FC sensitivity were assigned similarly but reflected the assumption that it is more sensitive than the ELISA, based on a study occurring under dairying conditions similar to those used in New Zealand.¹⁴ The prior for the effect on sensitivity of lactation stage reflected the belief that sensitivity would be slightly higher in early lactation, due to the heightened stress associated with parturition and the onset of lactation.

The prior estimate for ELISA specificity was determined after reviewing available literature^{4,6,18,29,31,36} to be 0.95 (and 95% confident that it was greater than 0.65). Fecal culture specificity was 0.98 (and 95% confident that it was greater than 0.70).

Beta distributions used to describe the prevalence in each herd are given in Table 3. Information on the within-herd prevalence was obtained before data collection based on the farmers’ perception of the annual incidence of clinical cases of JD.

Method 2: Repeated test events (complete data). In using the complete data, multiple test events for each cow (repeated measures) are available. Thus the formulae in Equation 2 now apply at the test event level, rather than the cow level, since a particular cow’s covariate values will change over time. Because test events were repeated at 5- to 7-month intervals, it was assumed that observations on the same cow would be “practically” independent, conditional on disease status.

Taking j now to denote the test event rather than the cow, the outcome for test event j is modeled as a multinomial with $n = 1$ and corresponding probabilities:

$$\begin{aligned}
 p_{11j} &= Z_{\text{cow}(j)} \text{Se}_E[j] \times \text{Se}_F[j] + (1 - Z_{\text{cow}(j)}) \\
 &\quad \times (1 - \text{Sp}_E[j]) \times (1 - \text{Sp}_F[j]) \\
 p_{12j} &= Z_{\text{cow}(j)} \text{Se}_E[j] \times (1 - \text{Se}_F[j]) + (1 - Z_{\text{cow}(j)}) \\
 &\quad \times (1 - \text{Sp}_E[j]) \times \text{Sp}_F[j] \\
 p_{21j} &= Z_{\text{cow}(j)} (1 - \text{Se}_E[j]) \times \text{Se}_F[j] + (1 - Z_{\text{cow}(j)}) \\
 &\quad \times \text{Sp}_E[j] \times (1 - \text{Sp}_F[j]) \\
 p_{22j} &= Z_{\text{cow}(j)} (1 - \text{Se}_E[j]) \times (1 - \text{Se}_F[j]) + (1 - Z_{\text{cow}(j)}) \\
 &\quad \times \text{Sp}_E[j] \times \text{Sp}_F[j]
 \end{aligned}
 \tag{4}$$

Recall that the disease status of an individual is presumed to be constant during the study period. $Z_{\text{cow}(j)}$ represents the infection status of the cow tested in event j ; this will be fixed (1 for infected, 0 for noninfected) for all test events on a particular cow. As before, it was assumed that $Z_{\text{cow}(j)}$ has a Bernoulli distribution with probability pi_k , where k denotes the herd to which this cow belongs. The WinBUGS code is given in Appendix 1. Note that the sensitivities and specificities are now allowed to vary for each cow across test events because lactation stage and parity vary with time.

Method 3: Pseudo gold standard. A PGS was created in which any cow with a positive test at any time during the follow-up period was considered infected with JD. The prevalence was defined for each herd as the proportion of cows that had a positive test of either type at any time divided by the total number of cows present during the period that herd was observed.

Table 4. Descriptive information for study farms and diagnostic testing.

	Farm			
	A	B	C	D
Cows sampled	131	142	235	271
Predominant breed	Friesian	Jersey	Friesian	Jersey
Test events*				
Early lactation (2000/01)†	100	110	109	160
Late lactation (2000/01)	65	105	94	152
Early lactation (2001/02)	104	107	175	176
Late lactation (2001/02)	92	101	0	146
Early lactation (2002/03)	77	88	0	85
Total	438	511	378	719

* A test event is the simultaneous collection of a fecal and serum sample from 1 cow.

† The milking season spanning 2000 and 2001.

The ELISA sensitivity was calculated using only the first ELISA for each cow. The number of positive tests in this group was divided by the number of cows that were positive to either test at any point during the study to give ELISA sensitivity. A similar approach was used to determine the sensitivity of FC, using only the first culture result for each cow. Non-Bayesian confidence intervals were calculated using standard methods based on binomial proportions. By definition of this method, specificities were assumed to be perfect. The covariance and confidence interval for the covariance between ELISA and FC test results within the infected population was estimated using a formula from a previous study.¹⁶

Assessment of convergence and sensitivity analysis

The models for methods 1 and 2 were run for 100,000 iterations after discarding an initial 5,000 as a burn-in period. Convergence to the posterior distribution was checked by monitoring 3 sets of starting values and the corresponding Brooks–Gelman–Rubin plots, in WinBUGS. Posterior distributions of interest were described using the median and 95% probability intervals. Sensitivity, specificity, and prevalence values were reported as percentages.

To ensure the models were stable, sensitivity analyses were conducted. The models were re-run after a single prior was changed within biologically sensible limits. This process was repeated for 4 priors in total: 2 for diagnostic test sensitivity, 1 for specificity, and 1 for JD prevalence.

Results

Summary of the data

A summary of the data is given in Table 4. In all, 779 cows were tested. For these cows, there were 2,046 test events, that is, when both ELISA and FC tests were used on the same animal simultaneously. Two hundred and fifty cows had 1 test event, 146 had 2 test events, 151 had 3 test events, 109 had 4 test events, and 123 had 5 test events. In herds A–D, there were 131, 142, 235, and 271 cows tested, respectively.

Sixty-five cows had at least 1 positive test. Ninety-eight of 2,046 test events were positive for 1 or both tests.

Performance of the ELISA test

The point estimate for overall sensitivity of the ELISA (Table 5) was 38.2% using repeat measure data and 41.4% using cross-sectional data. The PGS method estimate was lower at 30.4%. The error bounds for each estimate were wide. Estimates of the specificity of the ELISA (Table 5) were similar when using cross-sectional data (97.7%) and repeat measure data (99.5%).

When considering ELISA sensitivity within covariate levels, the general trend was for sensitivity to be lowest in parities 1 and 2, and higher in parities 3 and above (Table 5). It was higher in early lactation than in late lactation, by a similar magnitude using cross-sectional data (increase of 31.6%) and repeat measure data (increase of 28.1%).

Performance of the fecal culture test

The point estimates for the overall sensitivity of the FC test (Table 5) was 40.4%, using cross-sectional data, and exceptionally high (74.6%), using repeat measure data. The PGS method estimate was 45.6%. Again, the error bounds for all estimates were wide. Estimates of the specificity were lower for the FC test (Table 5) than for the ELISA, at 98% (cross-sectional data) and 98.5% (repeat measure data), respectively.

No trend in the sensitivity of the FC test was evident across parity groups. It was higher in early lactation than in late lactation by 31.9% (cross-sectional data) and 18.1% (repeat measure data). The error bounds associated with these estimates were wide (Table 5).

Prevalence

Estimates of the true herd-level prevalence (Table 5) varied in herd A, being 3.9% using repeat measure data, 21.7% using cross-sectional data, and 19.1% by the PGS method. In the other 3 herds, prevalence estimates were highest by the PGS method, which was to be expected as this method assumed no false-positive results occurred, and similar between the 2 Bayesian models.

Covariance between the ELISA and fecal culture test in the infected population

Covariance between the ELISA and FC test in the positive population was estimated to be very small, but positive using the PGS method (0.011, 95% confidence interval: 0.005–0.020). This finding supports the assumption made in the Bayesian analyses that there was negligible dependence between the 2 tests in the data set in the present study.

Table 5. Estimates of overall and covariate level sensitivity and specificity for enzyme-linked immunosorbent assay (ELISA) and fecal culture, and herd level prevalence, using a Bayesian model and either a single simultaneous application of both tests to 779 dairy cows (method 1) or repeated applications of both tests (method 2).

Statistical method/ parameter	Estimate (95% error bound)		
	Pseudo gold standard	Regression model (cross-sectional data)	Regression model (repeat measures data)
ELISA			
Overall sensitivity	30.4 (20.2–40.5)	41.4 (25.5–65.0)	38.2 (26.5–53.5)
Overall specificity	1.0*	99.7 (98.9–99.9)	99.1 (98.6–99.4)
Sensitivity by parity (late lactation)			
Parity 1		22.9 (5.7–54.3)	25.8 (9.5–50.7)
Parity 2		26.5 (4.8–69.6)	24.7 (4–66.9)
Parity 3		48.3 (17.4–83.1)	44.4 (20.5–74)
Parity 4		33.6 (7.1–76.9)	51.5 (27.5–79.3)
Parity >4		60.3 (25.2–90.6)	35.6 (13.4–65)
Sensitivity in lactation stage†			
Late lactation		22.9 (5.7–54.3)	25.8 (9.5–50.7)
Early lactation		54.5 (23.5–88.4)	53.9 (31.7–76.2)
Fecal culture			
Overall sensitivity	45.6 (34.6–56.6)	40 (26.7–55.4)	74.6 (61.5–85.3)
Overall specificity	1.0*	98 (96.2–99.6)	98.5 (97.7–99)
Sensitivity by parity (in late lactation)			
Parity 1		19.8 (4.7–51.9)	59.2 (30.7–84.2)
Parity 2		45.5 (12–84.6)	46.6 (11.6–89.9)
Parity 3		74.5 (35.6–95.7)	52.2 (24.3–83.9)
Parity 4		32.2 (4.9–78.5)	68.4 (41.2–88.8)
Parity >4		40.2 (12.9–74.2)	74.9 (39.4–95.6)
Sensitivity in lactation stage‡			
Late lactation		19.8 (4.7–51.9)	59.2 (30.7–84.2)
Early lactation		51.7 (21.2–81.8)	77.3 (52.4–92.8)
Herd level true prevalence			
Herd A	19.1 (12.4–25.8)	21.7 (11.5–36.4)	3.9 (1–8.9)
Herd B	14.8 (8.9–20.6)	9.6 (3.6–20.1)	6.7 (3.2–11.9)
Herd C	6.0 (2.9–9.0)	4.9 (1.3–11.7)	3.2 (1.2–6.3)
Herd D	7.0 (4.0–10.0)	2.8 (0.8–6.8)	2.3 (0.9–4.7)

* The pseudo gold standard method requires the assumption of perfect test specificity.

† ELISA test sensitivity in early lactation was 31.6% higher (54.5–22.9 = 31.6) than in late lactation using method 1, and 28.1% higher (53.9–25.8) using method 2.

‡ Fecal culture sensitivity in early lactation was 31.9% higher (51.7–19.8 = 31.7) than in late lactation using method 1, and 18.1% higher (77.3–59.2 = 18.1) using method 2.

Infection status of individuals

In the Bayesian analyses, the probability of infection was predicted for selected cows by monitoring their infection status $Z_{\text{cow}(j)}$ in the WinBUGS output. These cows were chosen because they had both positive and negative test results, either initially or in the subsequent repeated measures. The additional information provided by repeated measurements changed significantly the probability that an individual cow was infected compared with the cross-sectional analysis (Table 6). Cows with 1 initial positive result (cows 1, 2, 4, and 5) tended to have a high predicted probability of infection in the cross-sectional analysis, although this was influenced by lactation stage, parity at testing, and the within-herd

prevalence. Multiple subsequent negative tests reduced this probability drastically (cows 1, 2, and 4), whereas a subsequent positive test (cow 5) gave a much higher probability. When both initial tests were negative (cows 3 and 6), subsequent positive results gave, as expected, a significantly higher probability of infection. These decisions on the status of individual animals have, in turn, an effect on the estimates of sensitivities and specificities. For example, the initial results for cow 2 would be interpreted as a false negative for FC in the cross-sectional analysis but as a false positive for ELISA in the full-data analysis. Thus the assumption that infection status is a constant for each animal provides more information for evaluating test performance.

Table 6. The probability that individual cows with varying test result combinations were infected with Johne's disease estimated by the regression model using cross-sectional data and longitudinal data.

Cow	Herd prevalence	First test event result*	Subsequent test event results†	Predictive value positive	
				Cross-section	Repeated measures
1	4.1	- +	- - / - - / - - / - -	0.67	0.00
2	4.1	+ -	- - / - - / - - / - -	0.89	0.00
3	6.8	- -	- - / - + / - - / - +	0.04	0.32
4	3.2	- +	- -	0.47	0.14
5	2.3	- +	- +	0.26	0.93
6	6.8	- -	- - / - - / - +	0.06	0.12

* The enzyme-linked immunosorbent assay (ELISA) result precedes the fecal culture result, where + = positive and - = negative.

† The ELISA result precedes the fecal culture result, where + = positive, - = negative, and slash (/) = an approximately 6-month period between testing.

Sensitivity analysis

The model remained stable in response to changes in 4 prior values. Only the posterior value associated with the altered prior showed any change. The greatest change was a decline of 6% to the sensitivity of FC for cows in parities greater than 4 of 34.3. This change and the others observed during the sensitivity analyses were small in comparison to the error bounds surrounding the point estimates. All Markov chains converged consistently to the same distributions when initiated from different starting values. Autocorrelation had generally disappeared completely after a lag of between 10 and 20 iterations, and there was no evidence of poor mixing.

Discussion

In the current study, the sensitivity and specificity of the ELISA and FC tests for JD in dairy cattle was estimated. A Bayesian regression model was applied to a repeated test events data set for 779 cows and to a cross-sectional subset of this data set that included only the first test event for each of these cows. Results from a PGS approach were also included for comparison with the results from the Bayesian method.

The strength of the Bayesian model was that it provided an estimate of the sensitivity for the ELISA and for the FC test across the sampled population and for subgroups of that population of different age and lactation stage. The strength of applying this model to a repeated measure data set was that up to 5 test events (i.e., simultaneous ELISA and FC on 5 occasions) were available per cow. It has previously been concluded that only repeated serial testing of cattle will enable detection of the true infection rates of JD.²⁶

For the ELISA test, estimated overall sensitivity was higher by the Bayesian method than by the PGS method. Within covariate levels, sensitivity was lowest in parities 1 and 2 with no clear trend apparent in higher parity groups. This finding was consistent with previous research^{21,25} and the understanding that

animals in the early stages of infection cannot necessarily be identified by ELISA because months or even years may elapse before they seroconvert.²³

For the FC test, estimated overall sensitivity was comparable between the Bayesian method applied to cross-sectional data and the PGS method, given the breadth of their error bounds, but it was unusually high (74.6%, error bound: 61.5–81.3) when estimated using repeat measure data. This high sensitivity interpreted in light of the low prevalence and high specificity estimates suggests, from a biological point of view, that the latently infected population, or the base of the JD “iceberg” discussed in a previous study,³⁶ is smaller than those authors suggest. For the 4 herds in this study, at least, this may be a more accurate reflection of the true situation than offered by the cross-sectional analysis. There was no clear trend in FC sensitivity across parity levels.

Each test was estimated to be more sensitive in early rather than late lactation. The priors for both ELISA and FC sensitivity described them as 3% higher in early lactation. The posterior values were 28.1% and 18.1% higher in early lactation. This finding was consistent with the understanding that the high levels of stress associated with calving and peak milk production may cause infection with JD to worsen and positive tests for JD to be more common as a consequence. However, it conflicts with previous research that found a positive serum ELISA result was more likely at the end than in early lactation.¹⁸

The specificity of FC was estimated to be 98% (cross-sectional data) or 98.5% (longitudinal data), which conflicts with the common assumption that it is 100%.^{4,6,31} Transient infection of cattle with JD has been demonstrated,³⁵ while a previous study³⁶ mentions the pass-through phenomena in which cattle in an environment heavily contaminated with MAP may ingest and excrete the organism without being detectably infected. Under these conditions, the assumption of perfect specificity is called into

question, and while the likelihood of the true specificity being as low as 98% seems low it remains difficult to quantify. Estimates of ELISA specificity in this study were consistent with previous studies in which it was generally 99% or greater.^{29–31,36}

The point estimates of sensitivity must be interpreted with attention to their error bounds, which were wide in some cases. Nevertheless, they illustrate trends that were consistent with the current biological understanding of JD. The wide error bounds were probably due to the very small percentage (4.8%) of positive test results, and a data set with a higher percentage of positive tests may enable test performance to be determined with greater precision. For example, it would be interesting to apply this model to the Australian data set with up to 7 annual test events for cows in over 500 herds in Victoria.²¹

The results of this study stand to benefit both veterinary practitioners and the makers of policy for controlling JD. Veterinarians, when considering the test results for a possible case of JD, can refine their diagnosis depending on the age and lactation stage of the cow. Policy makers can use these results to support a testing program that focuses on older cows in early lactation, to assist in developing a cost-effective approach to large-scale control.

Estimates of prevalence followed a trend for herds B–D. The PGS value was slightly higher than the Bayesian model with cross-sectional data, which was in turn slightly higher than the Bayesian model with repeat measure data. The PGS method was expected to give the highest estimates of prevalence because it assumed that there were no false negative tests. It seems counterintuitive that repeatedly testing a population for a disease results in a lower prevalence than when the population is tested in a cross-sectional manner. However, this result can be readily explained with reference to the data on individual cows in Table 6. Despite high test specificity, the low prevalence meant that cows with a single positive test and multiple negative tests were considered more likely to be uninfected than infected. Most of the cows in this study had predominantly negative test results and consequently a lower probability of infection considering multiple test events than when considering a single test event, leading to a lower herd-level prevalence estimate after repeated testing. The low estimate of prevalence obtained when using repeat measure data was especially evident in herd A. Cows in this herd that tested positive usually did so only once and usually had multiple negative test results. Consequently, the probability that they were infected was higher when using cross-sectional data, than when using the complete data set. These singly positive tests may indicate temporarily elevated anti-

body levels in infected cows, or they may be due to cross-reaction with antigen from sources other than MAP, thus generating genuinely false positive results. It is conceivable that an organism present on farm A but not on the other 3 farms caused cross-reaction of the test. The prevalence results in this study are of value because they illustrate how repeated testing may lead to a lower probability of infection in many individuals within a population (herd).

A weakness of the regression model in the present study was that it did not account for dependence between the ELISA and FC tests. Such a simplification may have been acceptable in this data set because dependence was found to be very low (0.01, 95% confidence interval: 0.005–0.020). Insignificant impact on results has been reported when dependence was of this order of magnitude. Nevertheless, it is acknowledged that analysis of the performance of diagnostic tests with moderate accuracy and measuring the same biological response should, in general, attempt to adjust for dependence between the tests.¹⁷ Ignoring dependence between the 2 tests in the noninfected population is generally considered acceptable because both tests have high specificity.

Dependence may also exist between subsequent tests in an individual.¹⁶ The inclusion of a random effect to account for such dependence impaired Markov chain convergence, possibly owing to the low frequency of positive tests; however, its omission was recognized as undesirable. A data set containing a larger number of observations for each cow or a greater proportion of positive test results may have facilitated addition of the random effect. Given that testing occurred at 6-month intervals, the assumption of independence between test results for an individual cow was perhaps not unreasonable, and when tested, correlation between subsequent tests within cow was slightly positive but not significant (except tests 2 and 3, which were slightly negatively correlated [correlation coefficient = -0.01 , $P = 0.837$]).

It was assumed in this analysis that cows became infected with JD at a young age and that infection persisted for life, based on the experimental evidence available.^{13,22,28,33} While experimental infection studies indicate that, on occasion, the host may eliminate infection³³ or that transient infection may occur,³⁵ it was not believed these cases were sufficiently frequent to warrant representation in the model. If they were frequent, estimates of sensitivity and specificity would be erroneously low and high, respectively, while prevalence estimates would be higher than their true values.

A PGS based on serial testing of each individual (up to 5 test events [i.e., 5 ELISA tests and 5 FC tests] per cow over 3 years) was considered a useful contrast

for the results of the Bayesian model in the present study. Previously it had been concluded that only repeated serial testing of cattle would enable detection of the true infection rates of JD.²⁵ Two aspects of the PGS method require justification. First, it assumed that the tests were perfectly specific (specificity = 1.0), and while this is a simplification of reality, the high specificities (>98%) suggest it was acceptable. Second, all cows tested for JD were included in the analysis, with their infection status defined using all available testing information. This approach maximized analytical power, ensured data were consistent across the analytical approaches, and avoided bias between test-positive and test-negative populations. For example, if cows with 4 or more tests had been selected as the test-negative population, PGS results would have been biased toward older animals.

The performance of the Bayesian model remained stable under sensitivity analysis. There was little autocorrelation between subsequent iterations during the Gibbs sampling process, and posterior distributions remained the same when sampling was initiated from different starting values.

In conclusion, the Bayesian regression model described in the current study offered improved precision in the diagnosis of JD because, when estimating test sensitivity, it incorporated cow-level information such as age, lactation stage, and, when using repeated measure data, previous test results. Results were consistent with the current understanding of the disease process and previous research. Also, results demonstrated that prevalence estimates in a population may decline as test frequency increases, compared with when the population is tested in a cross-sectional manner. This model represents a tool to assist in the diagnosis of JD in individual cows, and to assist development of cost-efficient, large-scale disease control programs in which diagnostic testing plays a role.

Acknowledgements

Staff and students of the EPICentre, Massey University assisted with data collection. Livestock Improvement Corporation and Meat New Zealand provided project funding. Funding for the author's PhD (SN) was provided by the Agricultural Marketing and Research Development Trust, New Zealand (project number 20383).

Sources and manufacturers

- a. CSL ELISA, CSL Pty Ltd., Parkville, Victoria, Australia.

References

1. Bedrick E, Christensen R, Johnson W: 1997, Bayesian binomial regression: predicting survival at a trauma center. *Am Stat* 51:211–218.
2. Billman-Jacobe H, Carrigan M, Cockram F, et al.: 1992, A comparison of the interferon gamma assay with the absorbed ELISA for the diagnosis of Johne's disease in cattle. *Aust Vet J* 69:25–28.
3. Branscum A, Gardner IA, Johnson W: 2005, Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med* 68:145–163.
4. Buergelt C, Hall C, Merkal R, et al.: 1977, Lymphocyte transformation: an aid in the diagnosis of paratuberculosis. *Am J Vet Res* 38:1709–1715.
5. Chiodini R: 1993, Abolish *Mycobacterium paratuberculosis* strain 18. *J Clin Microbiol* 31:1956–1958.
6. Chiodini R, Van Kruiningen H, Merkal R: 1984, Ruminant paratuberculosis (Johne's disease): the current status and future prospects. *Cornell Vet* 74:218–262.
7. Colgrove G, Thoen CO, Blackburn BO, et al.: 1989, Paratuberculosis in cattle: a comparison of three serologic tests with results of fecal culture. *Vet Microbiol* 19:183–187.
8. Collins D, Stephens D, de Lisle G: 1993, Comparison of polymerase chain reaction tests and fecal culture for detecting *Mycobacterium paratuberculosis* in bovine feces. *Vet Microbiol* 36:289–299.
9. Cox J, Drane D, Jones S, Milner A: 1991, Development and evaluation of a rapid absorbed enzyme immunoassay test for the diagnosis of Johne's disease in cattle. *Aust Vet J* 68:157–160.
10. Dargatz D, Byrum B, Barber L, et al.: 2001, Evaluation of a commercial ELISA for diagnosis of paratuberculosis in cattle. *J Am Vet Med Assoc* 21:1163–1166.
11. Dargatz D, Byrum B, Hennager S, et al.: 2001, Prevalence of antibodies against *Mycobacterium avium* subspecies *paratuberculosis* among beef cow-calf herds. *J Am Vet Med Assoc* 219:497–501.
12. Dendukuri N, Joseph L: 2001, Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 57:158–167.
13. Doyle T: 1953, Susceptibility to Johne's disease in relation to age. *Vet Rec* 65:363–365.
14. Eamans G, Whittington R, Marsh I, et al.: 2000, Comparative sensitivity of various fecal culture methods and ELISA in dairy cattle herds with endemic Johne's disease. *Vet Microbiol* 77:357–367.
15. Eda S, Bannantine J, Waters W, et al.: 2006, A highly sensitive and subspecies-specific surface antigen enzyme-linked immunosorbent assay for diagnosis of Johne's disease. *Clin Vaccine Immunol* 13:837–844.
16. Gardner IA, Stryhn H, Lind P, et al.: 2000, Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev Vet Med* 45:107–122.
17. Georgiadis M, Johnson W, Gardner I, et al.: 2003, Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Appl Stat* 52:63–76.
18. Hilbink F, West D, de Lisle G, et al.: 1994, Comparison of a complement fixation test, a gel diffusion test and two absorbed and unabsorbed ELISAs for the diagnosis of paratuberculosis in sheep. *Vet Microbiol* 41:107–116.
19. Hui S, Walter S: 1980, Estimating the error rates of diagnostic tests. *Biometrics* 36:167–171.
20. Johnson W, Gastwirth J, Pearson L: 2001, Screening without a "Gold Standard": the Hui-Walter paradigm revisited. *Am J Epidemiol* 153:921–924.
21. Jubb T, Sergeant E, Callinan APL, et al.: 2004, Estimate of the sensitivity of an ELISA used to detect Johne's disease in Victorian dairy cattle herds. *Aust Vet J* 82:569–573.
22. Larsen A, Merkal R, Cutlip R: 1975, Age of cattle as related to resistance to infection with *Mycobacterium paratuberculosis*. *Am J Vet Res* 36:255–257.

23. Milner A, Lepper A, Symonds W, et al.: 1987, Analysis by ELISA and Western blotting of antibody reactivities in cattle infected with *Mycobacterium paratuberculosis* after absorption of serum with *M. phlei*. Res Vet Sci 42:140–144.
24. Milner A, Mack W, Coates K, et al.: 1990, The sensitivity and specificity of a modified ELISA for the diagnosis of Johne's disease from a field trial in cattle. Vet Microbiol 25:193–198.
25. Nielsen S, Enevoldsen C, Grohn Y: 2002, The *Mycobacterium avium* subsp. *paratuberculosis* ELISA response by parity and stage of lactation. Prev Vet Med 54:1–10.
26. Nielsen S, Grohn YT, Quaas RL, et al.: 2002, Paratuberculosis in dairy cattle: variation of the antibody response in offspring attributable to the dam. J Dairy Sci 85:406–412.
27. Nielsen S, Thamsborg SM, Houe H, et al.: 2000, Bulk-tank milk ELISA antibodies for estimating the prevalence of paratuberculosis in Danish dairy herds. Prev Vet Med 44:1–7.
28. Payne J, Rankin J: 1961, A comparison of the pathogenesis of experimental Johne's disease in calves and cows. Res Vet Sci 2:175–179.
29. Reichel M, Kittelberger R, Penrose M, et al.: 1999, Comparison of serological tests and fecal culture for the detection of *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle and analysis of the antigens involved. Vet Microbiol 66:135–150.
30. Ridge S, Morgan I, Sockett D, et al.: 1991, Comparison of the Johne's disease absorbed EIA and the complement-fixation test for the diagnosis of Johne's disease in cattle. Aust Vet J 68:253–257.
31. Sockett D, Conrad T, Thomas C, et al.: 1992, Evaluation of four serological tests for bovine paratuberculosis. J Clin Microbiol 30:1134–1139.
32. Spangler E, Bech-Nielsen S, Heider L: 1992, Diagnostic performance of two serologic tests and fecal culture for subclinical paratuberculosis, and associations with production. Prev Vet Med 13:185–195.
33. Stewart D, Vaughan J, Stiles P, et al.: 2007, A long-term bacteriological and immunological study in Holstein-Friesian cattle experimentally infected with *Mycobacterium avium* subsp. *paratuberculosis* and necropsy culture results for Holstein-Friesian cattle, Merino sheep and Angora goats. Vet Microbiol 122:83–96.
34. Sweeney RW, Whitlock RH, Buckley C, et al.: 1995, Evaluation of a commercial enzyme-linked immunosorbent assay for the diagnosis of paratuberculosis in dairy cattle. J Vet Diagn Invest 7:488–493.
35. Sweeney RW, Whitlock RH, Rosenberger E: 1992, *Mycobacterium paratuberculosis* isolated from fetuses of infected cows not manifesting signs of the disease. Am J Vet Res 53:477–480.
36. Whitlock RH, Wells SJ, Sweeney RW, et al.: 2000, ELISA and fecal culture for paratuberculosis (Johne's disease): sensitivity and specificity of each method. Vet Microbiol 77: 387–398.

Appendix 1

model;

Sensitivities

{ for (j in 1:n) {

 logit(Se[1,j]) <- g[parity[j]] + g[6]*x[j]

 logit(Se[2,j]) <- h[parity[j]] + h[6]*x[j] }

In late lactation, Se (ELISA) for each parity:

st1[1] ~ dbeta(1.38,2.51) # Parity 1, mode 0.20, 95% sure <0.75

st1[2] ~ dbeta(1.33,1.78) #2 mode 0.30, 95% sure <0.85

st1[3] ~ dbeta(2.23, 2.23) #3 mode 0.50, 95% sure <0.85

st1[4] ~ dbeta(2.23, 2.23) #4 mode 0.50, 95% sure <0.85

st1[5] ~ dbeta(1.62, 1.94) #5+ mode 0.40, 95% sure <0.85

st1[6] ~ dbeta(1.48, 2.59) # Parity 1, early Lactation, mode 0.23, 95% sure <0.75

In late lactation, Se (fecal culture) for each parity:

st2[1] ~ dbeta(1.38, 2.15) # parity 1, mode 0.25, 95% sure <0.80

st2[2] ~ dbeta(1.46, 1.85) # mode 0.35, 95% sure <0.85

st2[3] ~ dbeta(1.75, 1.62) # mode 0.55, 95% sure <0.90

st2[4] ~ dbeta(1.75, 1.62) # mode 0.55, 95% sure <0.90

st2[5] ~ dbeta(1.87, 2.06) # mode 0.45, 95% sure <0.85

st2[6] ~ dbeta(1.47, 2.21) # Parity 1, early lactation, mode 0.28, 95% sure <0.80

Induces priors on the regression coefficients for both tests

for (i in 1:5)

 g[6] <- logit(st1[6]) - g[1] h[6] <- logit(st2[6]) - h[1]

Infection status of selected cows: Test event results as (ELISA, FC)

cow[1] <- z[151] # - + - - - - - -

cow[2] <- z[245] # + - - - - - - -

cow[3] <- z[376] # - - - - - + - - - +

cow[4] <- z[533] # - + - -

cow[5] <- z[206] # - + - +

cow[6] <- z[726] # - - - - - - - +

```

# Specificities
Sp[1] ~ dbeta(8.5, 1.4) # mode 0.95, 95% sure >0.65
Sp[2] ~ dbeta(9.2,1.17) # mode 0.98, 95% sure >0.70

# Prevalence
for (i in 1:ncid){ z[i] ~ dbern(pi[i])
  pi[i] <- pt[hid[off[i]]] }

for (i in 1:n) {
  y[i,1:4] ~ dmulti(p[i,1:4],1)

  p[i,1] <- z[cid[i]]*Se[1,i] * Se[2,i] + (1-z[cid[i]]) * (1-Sp[1]) * (1-Sp[2])
  p[i,2] <- z[cid[i]]*Se[1,i] * (1-Se[2,i]) + (1-z[cid[i]])*(1-Sp[1])*Sp[2]
  p[i,3] <- z[cid[i]]*(1-Se[1,i])*Se[2,i] + (1-z[cid[i]])*Sp[1]*(1-Sp[2])
  p[i,4] <- z[cid[i]]*(1-Se[1,i])*(1-Se[2,i]) + (1- z[cid[i]])*Sp[1]*Sp[2] }
pt[1] ~ dbeta(1.55, 4.12) # True prevalence for herd B, mode: 0.15, 95% sure <0.60
pt[2] ~ dbeta(1.55, 4.12) # True prevalence for herd H, mode: 0.15, 95% sure <0.60
pt[3] ~ dbeta(1.25, 4.88) # True prevalence for herd M, mode: 0.06, 95% sure <0.50
pt[4] ~ dbeta(1.25, 4.88) # True prevalence for herd T, mode: 0.06, 95% sure <0.50

}

# Data

list(ncid=779, n=2046,
off=c(1, 6, 7, 8, 11, 14, ...))

cid[] hid[] y[,1] y[,2] y[,3] y[,4] x[] parity[]
1 2 0 0 1 0 1 4
1 2 0 0 1 0 0 4
1 2 0 0 1 0 1 4
1 2 0 0 1 0 0 4
1 2 0 0 1 0 1 4
2 2 1 0 0 0 1 4
3 2 0 0 0 1 1 4
4 2 0 0 0 1 0 4
4 2 0 0 0 1 1 4
4 2 0 0 0 1 0 4
5 2 0 1 0 0 1 4
5 2 0 0 0 1 0 4
5 2 0 0 0 1 1 4
6 4 0 0 0 1 0 4
.....
END

```