



ELSEVIER

Preventive Veterinary Medicine 52 (2002) 199–211

PREVENTIVE
VETERINARY
MEDICINE

www.elsevier.com/locate/prevetmed

Sensitivity and specificity of pooled faecal culture and serology as flock-screening tests for detection of ovine paratuberculosis in Australia

E.S.G. Sergeant^{a,b,*}, R.J. Whittington^c, S.J. More^b

^aNSW Agriculture, Locked Bag 21, Orange 2800, NSW, Australia

^bSchool of Veterinary Science, The University of Queensland, P.O. Box 125, Kenmore 4069, Qld, Australia

^cNSW Agriculture, Elizabeth Macarthur Agricultural Institute, PMB 8, Camden 2570, NSW, Australia

Received 20 March 2001; accepted 25 October 2001

Abstract

The flock-level sensitivity of pooled faecal culture and serological testing using AGID for the detection of ovine Johne's disease-infected flocks were estimated using non-gold-standard methods. The two tests were compared in an extensive field trial in 296 flocks in New South Wales during 1998. In each flock, a sample of sheep was selected and tested for ovine Johne's disease using both the AGID and pooled faecal culture. The flock-specificity of pooled faecal culture also was estimated from results of surveillance and market-assurance testing in New South Wales.

The overall flock-sensitivity of pooled faecal culture was 92% (95% CI: 82.4 and 97.4%) compared to 61% (50.5 and 70.9%) for serology (assuming that both tests were 100% specific). In low-prevalence flocks (estimated prevalence <2%), the flock-sensitivities of pooled faecal culture and serology were 82% (57 and 96%) and 33% (19 and 49%), respectively, compared to 96% (85 and 99.5%) and 85% (72 and 93%), respectively, in higher-prevalence flocks (estimated prevalence \geq 2%). A Bayesian approach incorporating prior knowledge on flock-specificity of pooled culture produced similar estimates and probability intervals. These estimates assume conditional independence of the two tests, and therefore might have over-estimated the true flock-sensitivities of the tests if the flock-sensitivities of pooled faecal culture and serology were correlated.

The estimated minimum flock-specificity of pooled culture when used for surveillance and assurance testing was 99.1% (96.9 and 99.9%). Surveillance and assurance programs in Australia are designed to provide a flock-sensitivity of 95% for an assumed prevalence of 2%. Pooled faecal culture is performing at close to this level—whereas the flock-sensitivity of serology appears to be lower than expected, particularly in lower prevalence flocks. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Sheep; Paratuberculosis; Johne's disease; Diagnostic tests; Bayesian methods; Sensitivity; Specificity; Pooled faecal culture; Flock tests

* Corresponding author. Present address: AusVet Animal Health Services, 69 Turner Crescent, Orange 2800, NSW, Australia. Tel.: +61-2-6362-1598; fax: +61-2-6369-1473.

E-mail address: evan@ausvet.com.au (E.S.G. Sergeant).

1. Introduction

Ovine Johne's disease (OJD) due to infection with sheep strains of *Mycobacterium avium* subsp. *paratuberculosis* first was diagnosed in Australia in 1980, in central New South Wales (Seaman et al., 1981). Since that time, it has spread (either through movement of sheep or from separate introductions) and by December 2000, more than 800 infected flocks had been reported in five of Australia's seven States (Sergeant, 2001). Despite this, there are still large areas of Australia where the disease has never been diagnosed, or was seen only in introduced sheep. A national ovine Johne's disease-control program has been implemented to limit further spread of the disease while research evaluates possible control or eradication strategies (Allworth and Kennedy, 2000).

Critical elements of the national program include surveillance to detect infected flocks, and a market-assurance program to identify flocks with a low risk of being infected (both as a source of infection-free sheep for commercial producers and as replacements for infected flocks undergoing eradication). Both of these programs depend on the availability of economic and reliable diagnostic tests to detect or exclude paratuberculosis at a flock level (Allworth and Kennedy, 2000; Sergeant, 2001).

Traditionally in Australia, testing for paratuberculosis relied on serology using the agar-gel immunodiffusion test (AGID), with any reactors submitted to postmortem histological examination to confirm or exclude infection (Whittington and Sergeant, 2001; Sergeant, 2001). It is only recently that techniques have been developed to allow reliable culture of ovine strains of *M a paratuberculosis* from tissues and faeces of infected animals (Whittington et al., 1998, 1999). This technology was extended by the development of pooled faecal culture (PFC) as a flock-screening test for paratuberculosis (Whittington et al., 2000). PFC has been used routinely for OJD surveillance in New South Wales since 1999.

Whittington et al. (2000) described the development of PFC and its comparison with the existing strategy of serology followed by histopathology on reactors. In the original evaluation, PFC appeared more sensitive than serology, but quantitative estimates of flock-sensitivity and flock-specificity were not made (Whittington et al., 2000). Because of the pooling of samples, PFC also provides a much-cheaper test for screening at the flock level (Whittington et al., 2000).

Although PFC is regarded as 100% specific (at least, at the flock level), it is essential that this is confirmed in field usage of the test to provide confidence in the performance of the test. Similarly, quantitative estimates of the flock-sensitivity of PFC are required to support its use as the preferred (that is, most-sensitive) screening test for OJD in Australia and to allow determination of appropriate sample sizes for testing programs.

Estimation of the sensitivity and specificity of flock-screening tests for paratuberculosis is further complicated by the fact that there is not a reliable 'gold standard' for flock status against which to compare results and evaluate performance of the test. Although both PFC and serology/histology are highly specific flock tests, both have imperfect sensitivity (Stehman, 1996; Whittington et al., 2000; Whittington and Sergeant, 2001). Thus, use of traditional methods dependent on a gold standard for determining flock-sensitivity and flock-specificity can result in estimates that are importantly biased (Staquet et al., 1981; Greiner and Gardner, 2000; Enøe et al., 2000). Enøe et al. (2000) recently reviewed techniques for estimating test sensitivity and specificity in the absence of a gold standard.

Although these techniques have been developed mainly for use with tests applied to individuals, the methods are equally applicable to flock-level tests such as PFC and serology.

The aims of this study were to evaluate the flock-sensitivity for both tests, to compare test performance in flocks with high and low prevalence, to confirm the adequacy or otherwise of current sample sizes for detection of OJD using these tests and to estimate the flock-level specificity of PFC. Because there was no gold standard available for flock-infection status, several non-gold-standard methods were used for this evaluation.

2. Methods

2.1. Estimation of flock-specificity

2.1.1. Estimation of flock-specificity of pooled faecal culture

The flock-specificity of PFC was estimated using the results of surveillance and assurance testing in areas of New South Wales where OJD was thought to be at a low or zero flock-prevalence, using a method adapted from Seiler (1979). Results from routine use of PFC in New South Wales between 1 January 1999 and 31 October 2000 were used for this analysis. Only submissions for which culture had been completed and which were from districts with <20 known-infected flocks (equivalent to a prevalence of known-infected flocks of $\leq 2\%$) on 30 September 2000 were included in the analysis. For all positive results, the flocks of origin were investigated further either by serology or by repeated faecal culture to determine whether or not infection could be confirmed. Any cases where infection could not be confirmed were regarded as false-positive results. The minimum flock-specificity of PFC in the field was then estimated as 100 minus the percentage of false-positive submissions.

2.2. Estimation of flock-sensitivity

2.2.1. Flock selection and testing

Whittington et al. (2000) described the selection and testing of the flocks used to estimate the flock-sensitivity of PFC and serology. These flocks were tested between April and December 1998 as part of an ongoing surveillance program for OJD in Australia.

Briefly, flocks were selected for testing either because there was some suspicion of infection due to tracing of sheep movements to/from an infected flock, or to demonstrate a flock's low-risk status for entry to the market-assurance program. For flocks with no evidence of disease or an expected low prevalence, sample sizes were chosen to provide 95% confidence of detecting a prevalence of 2%, assuming an animal-level test sensitivity of 30% for the AGID. In these flocks, selection of individual sheep for testing was carried out in a systematic manner across all adult sheep in the flock. Where the disease was likely to be at a higher prevalence, smaller sample sizes were used, and selection of sheep for testing was biased to include the animals most likely to be infected (older or poorer-condition sheep, or those introduced from known-infected flocks). These sample sizes and selection strategies were based on current recommendations for surveillance and assurance

testing for OJD in Australia (Anon., 1998, 2000). In each flock, the selected sheep were tested using both the AGID and PFC, and any seropositive animals were investigated further by postmortem examination and histology to determine their infection status.

Single faecal pellets were collected from the rectum of each sheep and aggregated into pools of up to 50 pellets for PFC testing. In flocks where the sample size was not a multiple of 50, one pool contained pellets from less than 50 sheep. All pooled samples were homogenised, decontaminated and cultured for *M a paratuberculosis* as described by Whittington et al. (2000). A flock was regarded as positive for PFC if *M a paratuberculosis* was isolated either in primary BACTEC medium, and/or in primary or secondary culture on solid medium from one or more pools. All isolates were confirmed as *M a paratuberculosis* by positive PCR for the IS900 sequence and restriction-endonuclease analysis (REA). An inconclusive PFC result was recorded if one or more pools were positive on PCR but insufficient reaction product was available to undertake REA. A flock was positive to serology/histology if one or more animals were identified with histological lesions consistent with paratuberculosis and in which acid-fast bacilli were observed. An inconclusive flock was one in which one or more animals had histological lesions suggestive of paratuberculosis (presence of giant cells and/or accumulations of three or more epithelioid macrophages), but in which no acid-fast organisms could be identified. Flocks in which there either were no seroreactors, or in which all reactors investigated were negative on histology were regarded as negative.

2.2.2. Within-flock prevalence estimation using AGID

Within-flock seroprevalence and true prevalence were estimated for all flocks that were positive to one or both flock tests. Seroprevalence was calculated as the percentage of sheep tested in each flock that were seropositive. The within-flock true prevalence was estimated by adjusting the seroprevalence for an assumed animal-level sensitivity and specificity for the AGID of 30 and 100%, respectively (Rogan and Gladen, 1978). This approach could not be used in seronegative flocks, because it yields a zero prevalence. In these flocks, the true prevalence was estimated as the upper 50% confidence limit of the estimated prevalence for zero positives from the given sample size, divided by the assumed sensitivity of the AGID (30%). An upper limit of 50% was chosen instead of 95% because an estimate of the 'average' within-flock prevalence was required, rather than an estimate of the likely maximum prevalence. An average sensitivity of 30% was used, because this is the assumed value used in sample size calculations for surveillance and assurance testing in Australia.

Flocks were categorised as low or high prevalence if the estimated true prevalence was <2 or $\geq 2\%$, respectively, and by sample size as <350 or ≥ 350 sheep sampled for serology. The cut-off sample-size of 350 was chosen because this was the recommended sample size for use of PFC for market-assurance and surveillance testing. A cut-off of 2% was chosen because sample sizes for market-assurance testing were determined to provide 95% confidence of detecting a prevalence of 2%.

2.2.3. Estimation of flock-sensitivity of both tests

2.2.3.1. *Estimation of flock-sensitivity assuming both tests are 100% specific.* Staquet et al. (1981) described a method for estimating the sensitivities of two tests applied at the

Table 1

Estimation of the flock-sensitivity of two tests, if both tests have 100% flock-specificity and if there is no gold standard

PFC (test 1)	Serology/histology (test 2)		Total
	Positive	Negative	
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$b + d$	n

individual level, where both tests have 100% specificity. This method was adapted to estimate the flock-sensitivities for both PFC and serology (see Table 1 and Eqs. (1) and (2))

$$Se_{pfc} = \frac{a}{a + c} \quad (1)$$

$$Se_{ser} = \frac{a}{a + b} \quad (2)$$

where Se_{pfc} and Se_{ser} were the estimated flock-sensitivities of PFC and serology/histology, respectively (Staquet et al., 1981). Flock-sensitivities for both tests were estimated across all flocks in the study and for low (<2%) and high ($\geq 2\%$) prevalence flocks separately.

A McNemar's test was used to test the statistical significance of differences between flock-sensitivity estimates for the two tests, with a p -value <0.05 (one-tailed) regarded as significant (Thrusfield, 1995, p. 214).

2.2.3.2. Bayesian estimation of flock-sensitivity and specificity. Bayesian analysis using a Gibbs sampler investigated the potential effects of imperfect flock-specificity of PFC, and evaluated the effect of incorporating prior knowledge of the performance of these tests on the resulting flock-sensitivity and flock-specificity estimates (Joseph et al., 1995; Enø et al., 2000).

A Gibbs sampler based on that used by Joseph et al. (1995) was developed using WinBUGS software (Spiegelhalter et al., 1999), and was used to estimate posterior distributions for the flock-sensitivities and flock-specificities of the two tests and the prevalence of infected flocks in the sample (Joseph et al., 1995).

This method requires input of:

- α and β parameters for five prior beta distributions: for flock-prevalence, and for flock-sensitivity and flock-specificity of both tests;
- starting values for flock-prevalence and flock-sensitivity and flock-specificity of both tests; and
- starting values for the estimated number of truly infected animals in each cell of the two-by-two table describing the test comparison.

Parameters and starting values for the prior distributions for flock-sensitivities and flock-specificities and the prevalence of infected flocks were as follows: Flock-specificity for

serology/histology was assumed to be 100%, because the flock-diagnosis was based on confirmation of infection in individual seropositive animals using histology (which is 100% specific at the animal-level). A starting value of 100% and a strong prior distribution assuming 0/1000 false positives for serology/histology were used, with $\alpha = 1001$ and $\beta = 1$. The prior distribution and starting value for PFC flock-specificity were based on the estimated flock-specificity from the analysis of testing results. In this case, $\alpha = (\text{number of submissions} - \text{false positives} + 1)$, $\beta = (\text{false positives} + 1)$ (Vose, 2000, p. 71), and the point estimate of flock-specificity was used as the starting value.

Uninformed (uniform) prior distributions were used for flock-sensitivities for both tests, because these were the parameters of primary interest, and there was considerable uncertainty about the actual flock-sensitivity that would be achieved considering the variation in estimated prevalence and sample size observed in the study flocks. The prevalence of infected flocks in the study was also unknown, so that an uninformed prior distribution and a starting value of 50% also were used for prevalence. These inputs were described as beta distributions with parameters $\alpha = 1$ and $\beta = 1$ (which equates to a uniform distribution with a minimum value of 0 and maximum value of 1) and in each case a starting value of 0.5 was used.

Given the presumed high flock-specificity of both tests, the actual number of flocks observed in each cell of Table 1 were used as starting values for each of cells a–c. For cell d, the number of infected but undetected flocks was estimated from the starting value for flock prevalence, less the estimated number of flocks already identified in cells a–c.

The model was run for 25,000 iterations, with the results from the first 5000 iterations discarded to allow convergence of the model and the results from the subsequent 20,000 iterations used to generate posterior distributions for the flock-sensitivities and flock-specificities of both tests and prevalence of infection. Posterior distributions were estimated using all flocks in the study, as well as in the high and low prevalence subgroups of flocks. The analysis was repeated using uninformed prior distributions for all parameters and the resulting output distributions were compared to those from the original model.

Because this method might be sensitive to changes in starting values and prior distributions, the model for all flocks was re-run with a range of different starting values and prior distributions. Alternative prior distributions that were used assumed that the flock-sensitivities of both tests were either 90 or 95% (because 95% was the target flock-sensitivity of the testing strategy), or that the flock-specificity of PFC was 95 or 99%. Alternative starting values used were generally extreme values within the valid range for the parameter. The analysis also was repeated for 250,000 iterations instead of the original 25,000 to check for convergence and to ensure that the results were repeatable.

2.2.4. Effect of conditional dependence between test sensitivities

Both the above methods used for estimation of flock-sensitivities of the two tests assume that the two tests are independent, conditional on infection status. Briefly, if two tests are applied in an infected population, the covariance of the test flock-sensitivities can be estimated as $\text{Cov}_{\text{Se}} = P_{++} - \text{Se}_{\text{pfc}}\text{Se}_{\text{ser}}$, where Se_{pfc} and Se_{ser} are the flock-sensitivities of the two tests and P_{++} the proportion of infected flocks that test positive to both tests. If $P_{++} = \text{Se}_{\text{pfc}}\text{Se}_{\text{ser}}$, then $\text{Cov}_{\text{Se}} = 0$ and the test flock-sensitivities are independent

(Vacek, 1985; Gardner et al., 2000; Dendukuri and Joseph, 2001). If $\text{Cov}_{\text{Se}} > 0$, then the test flock-sensitivities are correlated positively.

Because the number of infected flocks that were not detected by either test was not known, it was not possible to determine true values for Se_{pfc} , Se_{ser} and Cov_{Se} . However, using the test results for all flocks in the field trial, and assuming that both tests had a flock-specificity of 100%, values for Se_{pfc} , Se_{ser} and Cov_{Se} were estimated separately for assumed numbers of undetected infected flocks ranging between 0 and 50.

3. Results

3.1. Estimation of flock-specificity

3.1.1. Flock-specificity of pooled faecal culture

A total of 392 laboratory submissions from 31 districts were available for analysis, with 227 of these being from 25 districts with <20 known-infected flocks. There were nine faecal culture-positive submissions from these districts, of which seven were from properties that subsequently were confirmed as infected. On both remaining culture-positive properties, a positive PCR result was obtained on growth in BACTEC—but the organism was not isolated on solid media (either on primary culture or on sub-culture from BACTEC), and infection was not confirmed by further investigation using serology and repeated PFC.

There were thus two apparent false positives from 227 eligible submissions, and the estimated minimum flock-specificity of pooled faecal culture in the field was 99.1 (96.9 and 99.9%). α and β parameters for the prior distribution in the Bayesian analysis were 226 and 3, respectively, and a starting value of 99.1% was used.

3.2. Estimation of flock-sensitivity

3.2.1. Prevalence in infected flocks

Of 296 flocks investigated during the field trial, 100 (34%) were identified as infected by one or both tests, with a further 12 flocks inconclusive on either PFC (9) or serology/histology (3). For all additional analyses, these inconclusive results were regarded as being test negative for that test. A total of 162 flocks had <350 sheep tested (72 flocks had ≤ 50 sheep) and 134 had ≥ 350 . About 25% of the flocks with <350 sheep tested were positive to one or both tests compared to 44% of flocks that had ≥ 350 tested.

Sample sizes for serological testing in the positive flocks ranged from 40 to 500 (41 flocks had <350 sheep tested, including 19 flocks with ≤ 50 tested). Seroprevalence in these flocks ranged from 0 to 26%, with 24 flocks seronegative. The estimated true prevalence after adjusting for an assumed sensitivity of 30% for the AGID ranged from 0.5 to 87%. Overall, 46% of flocks had an estimated true prevalence <2%, and 14% of flocks had an estimated prevalence of >10%. Only 17% (7/41) of positive flocks with <350 sheep tested had an estimated prevalence of <2%, compared to 66% (39/59) of those with ≥ 350 tested. Eighteen of the 24 positive flocks that were seronegative had ≥ 350 sheep tested.

Table 2

Test results for PFC and serology in 296 sheep flocks tested in New South Wales during 1998 and in 46 low-prevalence and 54 high-prevalence flocks in the same study, and comparison of flock-sensitivities of PFC and serology assuming both tests were 100% specific

Population	PFC/serology test results				Sensitivities			
	+/+ (a)	+/- (b)	-/+ (c)	-/- (d)	Se _{pfc} (%)	95% CI (%)	Se _{ser} (%)	95% CI (%)
All flocks	58	37	5	196	92	82.4, 97.4	61	50.5, 70.9
Low-prevalence flocks	14	29	3	196	82	56.6, 96.2	33	19.1, 48.5
High-prevalence flocks	44	8	2	196	96	85.2, 99.5	85	71.9, 93.1

3.2.2. Estimation of flock-sensitivity of both tests

3.2.2.1. Estimation of flock-level sensitivity assuming both tests are 100% specific.

Overall, the estimated flock-sensitivity of PFC in the study population was 92% (82 and 97%), compared to 61% (51 and 71%) for serology/histology (Table 2). Serology performed much worse in lower-prevalence flocks. PFC also performed better than serology in higher-prevalence flocks, but the difference was not as great. Differences in flock-sensitivity estimates for the two tests were statistically significant ($p < 0.001$) except in high-prevalence flocks ($p = 0.11$).

3.2.2.2. Bayesian estimation of flock-sensitivity and specificity.

The posterior distributions from the Gibbs sampling model for data from all flocks (Table 3) indicated median flock-specificities of 98.8% (96.8 and 99.7%) for PFC, and 99.9% (99.6 and 100%) for serology/histology. Estimates of the flock-sensitivities (Tables 3 and 4) were comparable to estimates derived using the alternative method (Table 2). Use of uninformed prior distributions for all parameters resulted in a substantial reduction in

Table 3

Posterior distributions for flock-sensitivities and flock-specificities of PFC and serology, from the Bayesian analysis of data from 296 sheep flocks tested in New South Wales during 1998

Simulation	Parameter	Mean (%)	Median (%)	95% PI (%)
1 ^a	Sp _{pfc}	98.7	98.8	96.8, 99.7
2 ^b	Sp _{pfc}	90.4	90.1	81.7, 99.2
1	Sp _{ser}	99.9	99.9	99.6, 100
2	Sp _{ser}	98.2	98.5	95.3, 99.9
1	Se _{pfc}	91.0	91.4	82.9, 96.8
2	Se _{pfc}	94.3	95.0	85.3, 99.7
1	Se _{ser}	62.0	62.2	51.9, 72.2
2	Se _{ser}	78.2	77.5	57, 98.8

^a Simulation 1: prior distributions assume high flock-specificity for both tests.

^b Simulation 2: uniform prior distributions used for flock-specificity for both tests.

Table 4

Comparison of Bayesian estimates of flock-sensitivities for PFC and serology with estimates assuming both tests were 100% specific in 296 sheep flocks tested in New South Wales during 1998

Method	Parameter	Mean (%)	95% interval (%)
<i>All flocks</i>			
Bayesian	Se _{pfc}	91.0	82.9, 96.8
Both tests 100% specific	Se _{pfc}	92.1	82.4, 97.4
Bayesian	Se _{ser}	62.0	51.9, 72.2
Both tests 100% specific	Se _{ser}	61.1	50.5, 70.9
<i>Low-prevalence flocks</i>			
Bayesian	Se _{pfc}	79.6	58.9, 94.7
Both tests 100% specific	Se _{pfc}	82.4	56.6, 96.2
Bayesian	Se _{ser}	34.4	20.8, 50.1
Both tests 100% specific	Se _{ser}	32.6	19.1, 48.5
<i>High-prevalence flocks</i>			
Bayesian	Se _{pfc}	94.0	85.6, 99.1
Both tests 100% specific	Se _{pfc}	95.7	85.2, 99.5
Bayesian	Se _{ser}	87.3	75.0, 97.7
Both tests 100% specific	Se _{ser}	84.6	71.9, 93.1

flock-specificity estimates and a corresponding increase in flock-sensitivity estimates for both tests (Table 3).

Variations in prior input distributions and starting values appeared to have little effect on the estimated flock-sensitivity of PFC: all estimates were within $\pm 2\%$ —despite some quite-extreme changes to input values (data not shown). In contrast, estimates of the flock-sensitivity of serology/histology were highly unstable, varying by up to 23%. Most of these variations were upwards and were associated with either a prior distribution assuming a high flock-sensitivity for serology, or a substantial reduction in the flock-specificity of PFC.

Under most scenarios tested, the median flock-specificity of both tests remained at or above 99% for PFC, and close to 99.9% for serology/histology. Even when the posterior-median flock-specificity of PFC was as low as 90%, its posterior-median flock-sensitivity remained high at 95%, while that for serology/histology increased to 78% (see Table 3).

The model converged rapidly, and results were highly repeatable. Estimates derived from the results of 200,000 iterations after allowing for 50,000 iterations for convergence were no different from the original estimates. Similarly, median estimates for all parameters calculated from consecutive groups of 50,000 iterations were within 0.1% of each other and of the original estimates for all parameters, with similar 95% probability limits.

3.2.3. Effect of conditional dependence between test sensitivities

If there were three infected flocks that were not detected by either test, then Cov_{se} was zero and the flock-sensitivities of the two tests were independent. As the assumed number of undetected flocks increased, Cov_{se} also increased and the true flock-sensitivities for both PFC and serology/histology decreased (flock-sensitivity of PFC dropped faster than that of serology/histology). Despite this, even if there were up to 50 undetected infected flocks, the estimated flock-sensitivity of PFC was still substantially higher than that for serology/

Table 5

Effect of the assumed number of false-negative flocks on flock-sensitivities of PFC and serology, the covariance of test sensitivities and OJD prevalence estimates in 296 sheep flocks tested in New South Wales during 1998

Assumed number of infected flocks negative to both tests	Se _{pfc} (%)	Se _{ser} (%)	Covariance of test sensitivities	Prevalence (%)
0	95	63	-0.02	34
3	92	61	0.00	35
10	86	57	0.03	37
20	79	53	0.07	41
30	73	48	0.09	44
40	68	45	0.11	47
50	63	42	0.12	51

histology (Table 5). Fewer than three additional infected flocks would result in a negative correlation between the tests—a result that is regarded as unlikely (Gardner et al., 2000).

4. Discussion

Both tests evaluated in this study are biologically 100% specific at a flock level. Serological screening relies on the observation of typical lesions containing acid-fast bacilli in tissue samples from serological reactors to confirm infection, whereas PFC depends on identification of *M a paratuberculosis* by PCR for a sequence assumed to be unique to *M a paratuberculosis* or by isolation and identification of the organism on solid medium. Therefore, false-positive results for PFC only should occur through contamination or mis-identification of samples in either the field or laboratory. Strict quality-control measures were used to ensure that contamination did not occur in the laboratory, and guidelines for field collection and submission of samples were provided to minimise the chance of any cross-contamination between samples from different farms. In this context, cross-contamination of samples is most likely to occur between pools for an infected flock, with no effect on the flock-specificity of the test. Cross-contamination between flocks is less likely—but would result in a reduction in flock-specificity.

The flock-specificity of PFC was estimated using the results of general surveillance testing after removal of known-infected animals or flocks. In this case, the true flock-specificity of the test generally will be underestimated—but the estimate will be close to the true level (particularly if disease prevalence is low and known-infected animals or flocks can be identified reliably and excluded from the data) (Seiler, 1979). This analysis supported the assumed high flock-specificity of PFC, with only two apparent false-positive results out of 227 submissions tested. This provides a minimum estimate of flock-specificity, because one or both of these flocks might have been infected (but at an early stage or very low prevalence). Both of these submissions yielded only PCR positives and could not be cultured on solid media—suggesting that the level of growth was low (supporting the interpretation of *minimum* flock-specificity).

The estimated flock-sensitivity of pooled faecal culture suggests that it is performing at an acceptable level, and is likely to be achieving close to the desired level of confidence

when used for surveillance and assurance testing. The guidelines for the market-assurance program require testing a sample size of 350 sheep using PFC in seven pools of 50 each, or up to 500 sheep using serology, to provide 95% confidence of detecting a prevalence of 2% (equivalent to a flock-sensitivity of 95% for flocks with 2% prevalence). In this study, a flock-sensitivity for PFC of 96% (85 and 99.5%) was achieved in flocks with an estimated prevalence $\geq 2\%$, and 82% (57 and 96%) in flocks with a lower prevalence. In contrast, serology/histology failed to achieve the desired level of confidence (even in flocks with higher prevalence). A sample size much larger than currently recommended might be required to provide satisfactory flock-sensitivity for serology in low-prevalence flocks. Alternatively, we might have to lower flock-sensitivity to an unacceptable level.

The use of biased sampling in some flocks might have over-estimated true prevalence in those flocks. Similarly, the method used to estimate true prevalence in seronegative flocks may have over- or under-estimated the true prevalence in those flocks. However, the intention was to categorise flocks simply as high or low prevalence, and these biases are unlikely to have affected the comparison greatly between sub-populations in this study.

Generally, faecal culture is a more-sensitive animal-level test than serology, and faecal shedding begins before humoral responses in most animals (Chaitaweesub et al., 1999; Whittington and Sergeant, 2001). Thus, complete independence of tests is unlikely, and the sensitivities of serology and PFC are likely to be correlated in infected animals and flocks.

More-precise estimates of flock-sensitivity adjusted for any correlation between tests were not possible in this study without knowing the true status of all flocks in the study. Thus it is impossible to know the degree of over-estimation that might occurred and the estimates provided here must be regarded as maximum values rather than true unbiased estimates.

The flocks included in this study were tested as part of ongoing surveillance and market-assurance programs, and therefore are not necessarily representative of the general sheep population in Australia. In addition, flock-sensitivity is affected by within-flock prevalence of disease, and thus varies between flocks (Martin et al., 1992). Thus, the estimates of flock-sensitivity presented here represent a weighted-average for the flocks included in the study and actual sensitivity in individual flocks will vary depending on the characteristics of the flock (especially with respect to the stage of disease in infected sheep).

5. Conclusion

These results confirm that PFC is a highly sensitive and specific flock-test for detection of ovine Johne's disease—and is substantially more sensitive than serology. The difference in performance between the two testing approaches was particularly pronounced in apparently low-prevalence flocks. The current sample size of 350 sheep per flock for surveillance and market-assurance testing using PFC should provide the required 95% flock-sensitivity in flocks with a prevalence of 2% or greater, and a reduced but still-acceptable flock-sensitivity in lower-prevalence flocks. In contrast, the sample sizes used for serology in this study do not provide 95% flock-sensitivity—even in higher-prevalence flocks.

Acknowledgements

The authors thank Dr. David Jordan and Dr. Chris Baldock for their review of the manuscript.

References

- Allworth, M.B., Kennedy, D.J., 2000. Progress in national control and assurance programs for ovine Johne's disease in Australia. *Vet. Microbiol.* 77, 415–422.
- Anon., 1998. National Johne's Disease Program Standard Definitions and Rules for Sheep. Veterinary Committee, Canberra, Australia.
- Anon., 2000. The Australian Johne's Disease Market Assurance Program for Sheep. Animal Health Australia, Canberra, Australia.
- Chaitaweesub, P., Abbott, K.A., Whittington, R.J., Marshall, D.J., 1999. Shedding of organisms and sub-clinical effects on production in pre-clinical Merino sheep affected with paratuberculosis. In: Manning, E.J.B., Collins, M.T. (Eds.), *Proceedings of the Sixth International Colloquium on Paratuberculosis*, Wisconsin, USA, pp. 126–131.
- Dendukuri, N., Joseph, L., 2001. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 57, 158–167.
- Enøe, C., Georgiadis, M.P., Johnson, W.O., 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.* 45, 61–81.
- Gardner, I.A., Stryhn, H., Lind, P., Collins, M.T., 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.* 45, 107–122.
- Greiner, M., Gardner, I.A., 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3–22.
- Joseph, L., Gyorkos, T.W., Coupal, L., 1995. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am. J. Epidemiol.* 141, 263–272.
- Martin, S.W., Shoukri, M., Thorburn, M.A., 1992. Evaluating the health status of herds based on tests applied to individuals. *Prev. Vet. Med.* 14, 33–43.
- Rogan, W.J., Gladen, B., 1978. Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* 107, 71–76.
- Seaman, J.T., Gardner, I.A., Dent, C.H.R., 1981. Johne's disease in sheep. *Aust. Vet. J.* 57, 102–103.
- Seiler, R.J., 1979. The non-diseased reactor: considerations on the interpretation of screening test results. *Vet. Rec.* 105, 226–228.
- Sergeant, E.S.G., 2001. Ovine Johne's disease in Australia—the first 20 years. *Aust. Vet. J.* 79, 484–491.
- Spiegelhalter, D.J., Thomas, A., Best, N.G., 1999. WinBUGS Version 1.2 User Manual. MRC Biostatistics Unit, Cambridge, UK.
- Staquet, M., Rozenzweig, M., Lee, Y.J., Muggia, F.M., 1981. Methodology for the assessment of new dichotomous diagnostic tests. *J. Chron. Dis.* 34, 599–610.
- Stehman, S.M., 1996. Paratuberculosis in small ruminants, deer, and south American camelids. *Vet. Clin. N. Am. Food Anim. Pract.* 12, 441–455.
- Thrusfield, M., 1995. *Veterinary Epidemiology*, 2nd Edition. Blackwell, Oxford.
- Vacek, P.M., 1985. The effect of conditional independence on the evaluation of diagnostic tests. *Biometrics* 41, 959–968.
- Vose, D., 2000. *Risk Analysis—A Quantitative Guide*, 2nd Edition. Wiley, Chichester, UK.
- Whittington, R.J., Sergeant, E.S.G., 2001. Progress towards understanding the spread, detection and control of *Mycobacterium avium* subsp. *paratuberculosis* in animal populations. *Aust. Vet. J.* 79, 267–278.
- Whittington, R.J., Marsh, I., Turner, M.J., McAllister, S., Choy, E., Eamens, G.J., Marshall, D.J., Ottaway, S., 1998. Rapid detection of *Mycobacterium paratuberculosis* in clinical samples from ruminants and in spiked environmental samples by modified BACTEC 12B radiometric culture and direct confirmation by IS900 PCR. *J. Clin. Microbiol.* 36, 701–707.

- Whittington, R.J., Marsh, I., McAllister, S., Turner, M.J., Marshall, D.J., Fraser, C.A., 1999. Evaluation of modified BACTEC 12B radiometric medium and solid media for culture of *Mycobacterium avium* subsp. *paratuberculosis* from sheep. J. Clin. Microbiol. 37, 1077–1083.
- Whittington, R.J., Fell, S., Walker, D., McAllister, S., Marsh, I., Sergeant, E., Taragel, C., Marshall, D.J., Links, I.J., 2000. Use of pooled fecal culture for sensitive and economic detection of *Mycobacterium avium* subsp. *paratuberculosis* infection in flocks of sheep. J. Clin. Microbiol. 38, 2550–2556.