

1 Manuscript accepted for publication by Preventive Veterinary Medicine

2 <https://doi.org/10.1016/j.prevetmed.2017.11.016>

3 © 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

4 <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## 5 **A probabilistic approach to the interpretation of milk antibody results for** 6 **diagnosis of Johne's disease in dairy cattle**

7 Meyer A<sup>1</sup>, Bond K<sup>1,2</sup>, Van Winden S<sup>1</sup>, Green M<sup>3</sup>, Guitian J<sup>1</sup>

8 <sup>1</sup> Veterinary Epidemiology, Economics and Public Health. Department of Pathobiology and Population

9 Sciences. The Royal Veterinary College. Hawkshead Lane, North Mymms, Hatfield, Herts, AL9 7TA, UK.

10 <sup>2</sup> National Milk Records Group. Fox Talbot House, Greenways Business Park, Bellinger Close, Chippenham,

11 Wiltshire, SN15 1BN, UK.

12 <sup>3</sup> The School of Veterinary Medicine and Science, University of Nottingham, Sutton Warwickshire, CV8 2TL

13 Bonington Campus, Sutton Bonington, Leicestershire, LE12 5RD, UK.

14

15 *Corresponding author:*

16 Anne Meyer

17 Details:

18 [annemeyer@live.fr](mailto:annemeyer@live.fr)

19 41, Grande Rue, 52130 BAILLY-AUX-FORGES – FRANCE

20 0033 695 993 639

21

## Abstract

Johne's disease is a serious wasting disease of ruminants that is of high economic importance for the dairy sector in particular. The chronic nature of the disease, the fluctuations in antibody levels and the limited ability of diagnostic tests to identify cows at early stages of infection are huge challenges for the control of the disease. In the United Kingdom, the latter is commonly based on repeated milk ELISA testing of lactating cows, followed by selected culling and improved management practices around calving. In this paper, the dataset built through a large quarterly screening programme conducted in the United Kingdom since 2010 is used to investigate the use of milk ELISA testing for Johne's disease management. Over the study period, 13,509 out of 281,558 cows were identified as high-risk of being infected and shedding mycobacteria in the faeces, based on a case definition of at least two consecutive positive milk ELISA results. Around a third of them were kept in the dairy herd a year or more after being classified as high-risk. However, 16% of these cows did not have any further positive test, suggesting that they might be uninfected animals. The mean specificity and sensitivity of the milk ELISA test were estimated at 99.5 % and 61.8%, respectively. The cows in the dataset are categorised in different result groups according to the number of positive test results and whether they are classified as high-risk according to the programme's case definition. The posterior probability of infection is calculated after each test in order to investigate the impact of repeated testing on the belief in a cow's infection status. The interpretation of the results show that most cows classified as high-risk are very likely to be infected, while some other groups that do not match the case definition could reasonably be considered as infected too. Our results show that there is considerable potential for more targeted use of serological testing, including adjusting the testing frequency and implementing the posterior probability approach.

## Key words

Johne's disease / *Mycobacterium avium* subsp. *paratuberculosis* / dairy cattle / repeated testing / milk ELISA / antibody testing

## 1 Introduction

2 Johne's disease (JD)<sup>1</sup>, caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP), is a disease of serious  
3 economic importance for the dairy sector (Harris and Barletta, 2001, Ott et al., 1999, Stott et al., 2005).  
4 However, reliable impact estimates require precise data on the prevalence of the disease, which have been  
5 elusive largely due to the chronic nature of the disease and the limited ability of diagnostic tests to identify  
6 cows at early stages of infection. Nielsen and Toft in a systematic review carried out in 2009 concluded that  
7 accurate prevalence estimates were largely lacking across European countries (Nielsen and Toft, 2009). In  
8 the United Kingdom (UK), a study conducted in the late 90s reported an animal-level prevalence of infection  
9 of between 2.6 and 3.5% in Southwest England (Çetinkaya et al., 1996). Another study based on postal  
10 surveys reported a prevalence of between 17% and 71% of clinically infected herds in different study areas  
11 across the UK (Cetinkaya et al., 1998; Daniels et al., 2002). In a later study, it was estimated that 75 to 78% of  
12 herds in Southwest England had at least one seropositive animal (Woodbine et al., 2009). More recently,  
13 Velasova et al. (2017) estimated the herd level prevalence of JD across Great Britain as 68% (95% confidence  
14 interval: 59 to 77%) by means of antibody detection in bulk milk samples. Poor sensitivity of the available  
15 ante-mortem diagnostic tests poses a challenge for accurate estimation of JD prevalence and farm-level  
16 decision making with regard to individual cows (Collins et al., 2006; Eamens et al., 2000; Nielsen and Toft,  
17 2008). Among these tests, enzyme-linked immunosorbent assays (ELISA) for the detection of antibodies in  
18 milk are commonly used due to performance, convenience and cost (Nielsen and Toft, 2008). The ability of  
19 these tests to correctly identify JD-infected cows is highly dependent on their age and stage of infection  
20 (Hanks et al., 2013; Nielsen et al., 2002a, 2002b; Nielsen et al., 2013; Sweeney et al., 2006; van Schaik et al.,  
21 2003). Test sensitivity (the proportion of infected cows detected as positive by the test) is low at early stages  
22 of infection and increases as the disease progresses (Nielsen et al., 2013). For this reason, decisions  
23 regarding the management of individual cows are normally made upon consideration of the results of

---

<sup>1</sup> Abbreviations used: ELISA, enzyme-linked immunosorbent assay; JD, Johne's disease; MAP, *Mycobacterium avium* subsp. *paratuberculosis*; MES, mean effective sensitivity; NMR, National Milk Record group; PPI, posterior probability of infection; S/P, sample-to-positive ratio; UK, United Kingdom.

1 repeated tests together with either clinical manifestations and/or other factors such as somatic cell count,  
2 milk yield and fertility. To achieve reduction of the prevalence in the herd, it is commonly recommended to  
3 farmers to cull cows repeatedly testing positive, as they are more likely to be excreting MAP into the  
4 environment (Nielsen et al., 2002b; Nielsen, 2008; Sweeney et al., 2006) and to have reduced milk yield  
5 (Nielsen et al., 2009). A number of studies have formally investigated the control of JD in dairy herds and its  
6 impact, especially in the United States (Aly et al., 2012; Robins et al., 2015; Smith et al., 2017) and in  
7 Denmark (Kudahl et al., 2008, 2007). However, only one study has focused on JD control in the UK dairy  
8 sector so far (Stott et al., 2005), and it only looked at the impact of decreased milk yield and early culling. In  
9 the UK, JD control is usually based on repeated testing of all cows in infected herds associated with selective  
10 culling and improved calving and calf management practices. Anecdotal evidence suggests that the course of  
11 action following a specific set of test results is highly inconsistent. Therefore, individual cow management  
12 could be an area where farmers and veterinarians would benefit from further evidence to inform their  
13 decisions. In this study, a large dataset of individual cow milk ELISA test results was used to investigate  
14 several critical aspects in relation to the use of repeated tests for the purpose of JD management. The  
15 primary objective of the study was to evaluate whether using a probabilistic approach could support  
16 interpretation of repeated test results for individual cows. To this effect, we firstly estimated the test  
17 performances (sensitivity and specificity) based on the test results obtained for cows at different ages. Then,  
18 we calculated the true within-herd seroprevalence and estimated the predictive values of single and  
19 repeated testing. These data were combined to calculate and analyse the posterior probability of infection of  
20 individual cows after each test result. Lastly, we assessed the potential implications of our results on the  
21 current recommendations for management and culling of cows deemed as “high-risk”.

# 1    **Material and methods**

## 2    **Source of data and data management**

3    The study was based on detailed exploration of a dataset provided by the National Milk Records Group  
4    (NMR) containing 1,694,172 complete records representing all JD milk ELISA tests conducted by NMR from  
5    01/01/2010 to 18/05/2015. ELISA test results included in the dataset were performed within a JD screening  
6    programme for dairy herds ('HerdWise') provided by the NMR group in the UK. Within herds enrolled in this  
7    programme, all milking cows are tested for JD antibodies by milk ELISA on a quarterly basis using samples  
8    obtained for milk recording. A single herd can therefore contribute records from multiple cows and an  
9    individual cow can be represented in the dataset with multiple records, representing tests conducted at  
10   different points in time. Each record is one milk ELISA test result, with the following information:  
11   identification of the herd and the cow, birth date of the cow, date of the test and optical density value  
12   obtained from the sample. Milk samples are collected during milking, into pots containing bronopol as a  
13   preservative and delivered to the laboratory within 48 hours of collection. The samples are de-fatted and  
14   tested by means of a commercial ELISA test. The same test has been in use over the entire period (IDEXX  
15   Paratuberculosis Screening Ab Test, IDEXX Laboratories, Maine, United States). The test interpretation was  
16   performed following the manufacturer's instructions using a cut-off of 30: tests with a sample-to-positive  
17   ratio (S/P) of 30 or above were considered positive, whilst others were considered negative. Within the  
18   HerdWise programme, the so-called "red cows" are defined as cows with two consecutive positive milk  
19   ELISA results and as a result are deemed to be at high-risk of being infected and shedding MAP in the faeces.

## 20   **Evaluation of single test performance**

21   We estimated the age-specific specificity and sensitivity of the milk ELISA test from the available data,  
22   following the approach described by Nielsen et al. (2013). The target condition was defined as MAP-infected  
23   cow in which a humoral response would become detectable within the economic life of the cow. The  
24   transition from a cell-mediated response to a detectable humoral response has been associated with the

progression of the disease, development of symptoms and faecal bacterial shedding (Koets et al., 2015; Nielsen et al., 2009; Stabel, 2000). Each cow from the dataset was classified as case or non-case for this condition according to the definitions presented below, or excluded from the dataset if it did not comply with either definition.

To evaluate the specificity of a single test, cows with at least nine test results and for which the last eight tests were negative were classified as non-cases. Non-cases were considered as non-MAP infected or as MAP infected but with no progression of the infection during the cow's lifetime (Mitchell et al., 2015). Eight negative tests correspond to around two years of negative tests in a quarterly programme. Preliminary data analysis showed that most of the cows with at least one positive test had seroconverted by the age of five. As testing starts around two to three years of age, this definition enabled us to retain cows that were likely to be non-MAP infected. Based on this definition, 55,586 cows were selected as non-cases. To avoid dealing with correlated data, one test result from each non-case cow (and the age at which it was obtained) was selected at random. The specificity of the test was estimated as the number of negative tests among the tests obtained from non-cases. The lower and upper limits of the confidence interval of the specificity were calculated with the Agresti-Coull method which is recommended for large samples (Brown et al., 2001).

To evaluate the sensitivity of a single test, we chose a case definition matching the "red cow" definition used within the HerdWise programme. Cows with at least three test results available and where the two last tests conducted yielded a positive result were classified as cases. Based on this definition, 9,553 cows were selected as cases, and, therefore, as MAP-infected cows with progression of the infection. One test result from each case cow (and the age at which it was obtained) was selected at random. The age-specific sensitivity  $Se(t)$  of the test at age  $t$  is the proportion of positive tests at a given age among the tests obtained from cases.  $Se(t)$  was estimated using a non-linear logistic regression model (Nielsen et al., 2013):

$$\text{logit}(Se(t)) = a - b * e^{-c*t}$$

where  $a$  is the upper limit of the logit function when  $t$  increases,  $b$  is a scaling factor and  $c$  is the coefficient for the decay of the age effect. The inverse-logit of  $a$  is the upper limit of the age-specific sensitivity when

age increases. The age-specific sensitivity accounts for both the probability that an infected cow excretes detectable levels of MAP antibodies in the milk and the intrinsic characteristics of the milk ELISA. As age increases, the probability that an infected cow secretes detectable levels of MAP antibodies in the milk increases. Therefore, the upper limit of the age-specific sensitivity is a good estimator of the sensitivity of the ELISA test in an infected cow which is secreting detectable level MAP antibodies in its milk at the time of testing. Parameters  $b$  and  $c$  condition the rate at which the sensitivity increases with age. After the model was fitted, the Pearson's correlation coefficient between observed and predicted values was calculated to estimate the goodness-of-fit.

### Evaluation of true within-herd seroprevalence

The age-specific sensitivity as estimated above was used to calculate the mean effective sensitivity (MES) of the milk ELISA test in dairy herds enrolled in HerdWise, given the age distribution of the cows in these herds (Kirkeby et al., 2015). As this test can only be performed in lactating animals, we only considered cows aged two years and older in this section of the study. We approximated the age distribution of the cows in each herd by the age distribution of the cows tested at least once between 01/01/2015 and 08/05/2015. This may not be an accurate estimate of the true age distribution in these herds as some cows may have left the herd, joined the herd or been present but not tested during the chosen period. However, precise demographic data were not available at the time of analysis. We assumed that the age distribution was constant in time over the study period. For each herd  $k$  enrolled in the HerdWise programme, the mean effective sensitivity of the test ( $MES_k$ ) is the average of the individual sensitivities  $Se(t_i)$  for each cow  $i$  of age  $t_i$ ,  $i = (1, 2, \dots, N_k)$  where  $N_k$  is the number of cows in herd  $k$ :

$$MES_k = \frac{\sum_{i=1}^{N_k} Se(t_i)}{N_k}$$

The mean effective sensitivity  $MES_k$  and specificity  $Sp$  of the test were used to calculate the true seroprevalence  $P_{k,y}$  of the disease in herd  $k$  during year  $y$  (Rogan and Gladen, 1978). The apparent

1 seroprevalence  $\pi_{k,y}$  was estimated as the proportion of cows with at least one positive result (“seropositive  
2 cows”) in herd  $k$  during year  $y$ .

$$3 \quad P_{k,y} = \frac{\pi_{k,y} + Sp - 1}{MES_k + Sp - 1}$$

#### 4 **Evaluation of the predictive values of repeated tests**

5 Within JD screening programmes, cows are tested at regular intervals. The information provided by each  
6 successive test can be accumulated in order to strengthen the evidence of a certain disease status (MAP-  
7 infected or not). To this effect, the posterior probability that a given cow is infected after each test provides  
8 useful information. Such calculation draws on the age-specific sensitivity, specificity and true herd-  
9 seroprevalence estimated in the first part of the study, using likelihood ratios of test results to infer the  
10 posterior probability of infection from the prior probability of infection. The likelihood ratio of a given test  
11 result is the ratio of the likelihood that this result would be expected in an infected cow to the likelihood  
12 that the same result would be expected in a non-infected cow (Dohoo et al., 2003). Likelihood ratios depend  
13 on the age of the tested cow, as the performances of the test are strongly affected by this parameter  
14 (Nielsen et al., 2013). We incorporated the age-specific sensitivity in the calculation of likelihood ratios for  
15 dichotomous results, following the definitions currently used within HerdWise, i.e. only two categories:  $S/P <$   
16  $30$  (negative test) and  $S/P \geq 30$  (positive test). Likelihood ratios of positive and negative tests at age  $t$  ( $LR_+(t)$   
17 and  $LR_-(t)$ , respectively) were calculated with the following equations, where  $T+$  ( $T-$ ) represents the  
18 occurrence of a positive (negative) test and  $C+$  ( $C-$ ) represents a case (non-case) cow:

$$19 \quad LR_+(t) = \frac{P(T+|C+)}{P(T+|C-)} = \frac{P(T+|C+)}{1 - P(T-|C-)} = \frac{Se(t)}{1 - Sp}$$

$$20 \quad LR_-(t) = \frac{P(T-|C+)}{P(T-|C-)} = \frac{1 - P(T+|C+)}{P(T-|C-)} = \frac{1 - Se(t)}{Sp}$$

21



1 We assumed that the prior probability of infection before any test is performed was equal to the true  
2 seroprevalence in the herd to which the tested cow belonged. The posterior odds of infection after each test  
3 is the product of the prior odds of infection and the likelihood ratio corresponding to the result of the test  
4 (positive or negative) and the age of the cow. The odds of an event are obtained by dividing the probability  
5 of the event by its complement. For further tests, the prior probability of infection is equal to the posterior  
6 probability of infection from the previous test. The posterior probability of infection (PPI) after each test was  
7 calculated for every cow in the dataset. Cows in the dataset were categorized into five different “result  
8 groups” depending on the number of positive test results (0, 1, 2, 3 to 7 and 8 or more, for result groups A,  
9 B, C, D, E, respectively). Result group C was sub-divided into groups C1 and C2, according to whether the cow  
10 was diagnosed as “red cow” or not, respectively. Result group D was sub-divided into groups D1 and D2 for  
11 the same reason. All cows in result group E were “red cows” so this group was not divided further. A  
12 summary of the group definitions is provided in the three first rows of Table 2. The distribution of the PPI  
13 was plotted for a typical cow for each result group in order to illustrate how the belief in a cow’s status  
14 evolves over time. Key statistics regarding the distribution of the PPI in different groups were calculated: age  
15 at first positive test, PPI after first, second and third positive test results when applicable, PPI after last test  
16 recorded for each individual cow. The Kruskal Wallis test and the Dunn test for post-hoc comparisons were  
17 used to compare the number of tests, the PPI after the second positive test result and the PPI after the last  
18 test between result groups. We considered PPI below 5% and above 95% to be indicating likely un-infected  
19 and infected cows, respectively. The percentages of cows above or below these thresholds were calculated  
20 at different points in time.

## 1    **Specific investigation of the data related to “red cows”**

2    As classification of a cow as “red cow” (i.e., cow belonging to group C1, D1 or E) is a critical element of  
3    current JD management programmes in the UK, the practices associated with these cows were investigated  
4    further. We extracted data in terms of age at diagnosis (age at the second positive test of the first sequence  
5    of two consecutive positive tests), number of days “red cows” were kept in the herd following their  
6    diagnosis, number of high-risk-cow-years at herd level during the study period and number of tests carried  
7    out (and their results) following identification of a cow as “red cow”.

## 8    **Sensitivity analysis**

9    A sensitivity analysis was conducted to assess the effect of modifying the case definitions on the results. The  
10   age-specific sensitivity, specificity and MES were calculated after increasing or decreasing the number of  
11   negative tests required for non-cases by 4 tests, corresponding to about one year of testing (i.e. 12 and 4  
12   negative tests, respectively). The number of positive tests required for a cow to be considered a case was  
13   also increased or decreased by one test (i.e., 3 and 1 positive tests, respectively).

## 14   **Results**

### 15   **Descriptive statistics**

16   The number of tests conducted as part of the HerdWise programme has been increasing with time, from  
17   39,762 in 2010 to 549,664 in 2014. The dataset contained test results from 281,554 cows belonging to 1,222  
18   herds, and 5.6% of the tests yielded a positive result.

### 19   **Evaluation of single test performances**

20   Based on the test results of non-cases, the mean specificity of the test was estimated at 99.5 % (95% CI: 99.4  
21   - 99.6 %). As for the age-specific sensitivity, the results of the non-linear logistic model are shown in Table 1  
22   and Figure 1.

**Table 1. Results of the non-linear regression model of the age-specific sensitivity of the test.** The regression coefficients were the upper limit of the logit function when age increases (a), the scaling factor (b) and the coefficient for the decay of the age effect (c).

<b>Coefficient estimate (standard error)</b>	<b>a</b>	1.2 (0.17)
	<b>b</b>	3.0 (0.23)
	<b>c</b>	0.30 (0.067)
<b>Pearson's correlation coefficient between observed and predicted values</b>		98.3%
<b>Upper limit of Se(age), mean and 95% confidence interval</b>		77.5% (71.0% - 82.9%)

**Figure 1. Estimated test sensitivity as a function of the age at testing.** The plain line shows the maximum likelihood estimate while the 95% confidence interval of the predictions is displayed as dashed lines. The bold line represents the upper limit of the sensitivity function when the age increases.

### Evaluation of true within-herd seroprevalence

The median age of the 186,846 cows sampled in 2015 was 4.4 years and the interquartile range 3.2 to 5.9 years. Based on this distribution, we calculated the MES in each herd of the dataset. The density distribution of the MES at herd-level and of the true within-herd seroprevalence calculated for the study herds in 2015 are presented in Figure 2. The mean herd-level MES and within-herd true seroprevalence were 61.8% and 16.8%, respectively.

1 **Figure 2. Density distribution of the mean effective sensitivity at herd level and true within-herd**  
2 **seroprevalence for the study herds in 2015.**

### 3 **Evaluation of the predictive values of repeated tests**

4 Example probability patterns for each result group (A to E) are presented in Figure 3 and the distribution of  
5 the cows by result group is presented in Table 2. There was a statistically significant difference between the  
6 total number of tests recorded by cow between result groups, however the effect size was small (Kruskal  
7 Wallis  $\chi^2 = 15,713$ ,  $df = 6$ ,  $p < 0.01$ ). There was also a statistically significant difference between the PPI after  
8 the last test between result groups (Kruskal Wallis  $\chi^2 = 62,137$ ,  $df = 6$ ,  $p < 0.01$ ) and a post-hoc Dunn test  
9 showed that the probabilities were different within each pair of result groups.

10 Most cows in the dataset (83% of all cows, Table 2) belong to result group A, with repeated negative tests  
11 only. In this group, the PPI fell below 5% after a median number of two tests (s.d. 0.8; 95<sup>th</sup> percentile: 3). The  
12 median age of the cows when the PPI fell below 5% was 3.0 years (s.d. 1.6 years, 95<sup>th</sup> percentile: 6.8). The  
13 median PPI after the last test available for an individual was 0.3% (s.d. 3%, 95<sup>th</sup> percentile: 8.5%).

14 Cows in result group B (11% of all cows, 62% of seropositive cows) have repeated negative tests and one  
15 isolated positive test. The median age when the positive test result was obtained was 4.0 years (s.d. 2.1  
16 years). The median PPI after the positive test was 66.3%, with a wide confidence interval (s.d. 37.6%). For  
17 the 89% of cows where the PPI rose above 5%, it fell under this threshold again after a median of 4 further  
18 tests (s.d. 2.1). However, the PPI after the last test remained higher than it was in group A (median 11.4%,  
19 s.d. 36.0%).

1 Cows in result group C have had two positive tests only (3% of all cows or 17% of seropositive cows), while  
2 those with three or more positive tests are in result groups D and E. When the two positive tests are  
3 consecutive, the cow is classified as “red cow” (group C1, 53% of cows with two positive tests). The median  
4 age at 1<sup>st</sup> and 2<sup>nd</sup> positive tests was 4.2 and 4.5 years, respectively. After the second positive test, the median  
5 PPI was 99.6% and it was above 95% for 73% of the cows. Most of cows from result group C1 (73%) are not  
6 tested any further, suggesting early culling. The remainder of the cows in result group C1 only have negative  
7 tests subsequently to their diagnosis as “red cow”. The median PPI at the last test for the cows that were re-  
8 tested was 89.3% (s.d. 40.1%) and 45% of these cows had a last PPI above 95%. In Figure 3, the cow would  
9 have been classified as high-risk early in life (during the 3<sup>rd</sup> year of age). However, no further positive test  
10 occurred for the next four years of her life and the probability of infection slowly decreases towards zero.

11 Almost half of the cows with two positive tests (47%) are not “red cows” as they do not have consecutive  
12 positive tests, and were classified in result group C2. After the second positive test, the median PPI was  
13 92.8% and it was above 95% for 46% of the cows. It was significantly smaller than in result group C1 (Kruskal  
14 Wallis  $\chi^2 = 1,208$ , df = 1,  $p < 0.01$ ). The median PPI at the last test was 67.0% and 30% of the cows had a last  
15 PPI above 95%.

16 Cows in result group D have between three and seven positive tests over time. They represent 3% of all cows  
17 and 18% of seropositive cows. Most of the cows from result group D (92%) are “red cows”, i.e., they have at  
18 least two consecutive positive test results (result group D1). A smaller proportion do not have consecutive  
19 positive tests and are not “red cows” (result group D2). The test results in group D often alternate between  
20 positive and negative, with S/P values close (either above or under) to the accepted threshold. The median  
21 PPI after the 2<sup>nd</sup> positive test was 99.6% and 98.0%, in groups D1 and D2, respectively, and it was above 95%  
22 for 76% and 61% of the cows. After the 3<sup>rd</sup> positive test, the median PPI was over 99.9% (s.d. 14.7%) and it  
23 was over 95% in 92% of the cows, in both sub-groups. The median PPI after the last test was over 99.9% (s.d.  
24 16.2%) and it was above 95% for 92% of the cows. For 99.1% of the cows, the median PPI did not fall under  
25 95% at any one point after the 3<sup>rd</sup> positive test.

1 Lastly, a small number of cows (less than 1% of all cows, 2% of seropositive cows) have had eight or more  
2 positive results in total (result group E). Even if these cows may have a few negative tests intercalated, they  
3 are all “red cows”. An example of these cows is the individual with profile E on Figure 3. The median PPI  
4 after the 2<sup>nd</sup> positive test was 94.4% and it was above 95% for 94% of the cows. The median PPI did not fall  
5 under 95% in any of the cows after the 3<sup>rd</sup> positive test.

6 **Table 2. Key characteristics of the result groups defined according to the number of positive tests and “red**  
7 **cow” status in the study population (N=281,554).**

Group		A	B	C1	C2	D1	D2	E
Number of positive tests		0	1	2		3 – 7		> 7
"Red cow"		No	No	Yes	No	Yes	No	Yes
Number of cows		233,362	30,026	4,435	3,934	7,897	723	1,177
% of all cows		82.9	10.7	1.6	1.4	2.8	0.3	0.4
% of seropositive cows			62.3	9.2	8.2	16.4	1.5	2.4
Number of tests per cow	median	5	7	6	10	8	12	12
	s.d.	3.8	4.2	3.9	3.9	3.7	3.9	4.7
Age at first positive test (years)	median		4.0	4.2	3.9	4.0	4.0	3.8
	s.d.		2.1	1.9	1.9	1.9	1.8	1.7
Posterior probability of infection after the first positive test (%)	median		66.3	69.0	77.9	83.4	86.8	94.4
	s.d.		37.6	38.0	34.7	34.4	31.3	24.4
Posterior probability of infection after the second positive test (%)	median			99.6	92.8	99.6	98.0	99.9
	s.d.			27.7	37.6	25.4	29.5	9.0
Posterior probability of infection after the last test (%)	median	0.3	11.4	99.1	67.0	>99.9	99.1	100
	s.d.	3.1	36.0	33.5	41.4	12.9	32.8	2.9

8

9 **Figure 3. Common test result profiles encountered in the dairy cow population.** Horizontal axes show the  
10 age at testing in years. Left vertical axes (black curve) show the posterior probability of infection. Right  
11 vertical axes show the S/P result of the ELISA test (green and red dots for negative and positive results,  
12 respectively). The dashed lines represent the S/P cut-off.

1 **Specific investigation of the data related to “red cows”**

2 Overall, 13,509 cows were classified as “red cows” (groups C1, D1 and E depending on the total number of  
 3 positive tests), in 1,106 different herds. The median age at diagnosis for these cows was 4.6 years  
 4 (interquartile range: 3.5 – 6.0 years). These high-risk cows were kept in their respective herds an average of  
 5 297 days following their identification as high-risk (interquartile range 0-469 days, median 154 days). Around  
 6 32 % of them were kept in the herd a year or more after being classified as high-risk. The average of high-  
 7 risk-cow-years after detection is 9.9 high-risk-cows-years per herd over the study period. The proportion of  
 8 positive tests after diagnosis of a cow as “red cow” is shown in Table 3. Roughly, one third of the cows in the  
 9 dataset had only subsequent positive tests, one third were not re-tested (or were culled) and one third had  
 10 one or more negative tests after the diagnosis. Out of the cows which were re-tested, 84% had at least one  
 11 further positive test (groups D1 or E).

12 **Table 3. Proportion of positive test results after the “red cow” diagnosis over the study period (N=13,509).**

Proportion of positive tests among the post-diagnosis results	No further tests	0%	>0% and <25%	25 to 50 %	50 to 75 %	> 75% and <100%	100%
Proportion of cows in the category	30 %	11 %	3 %	5 %	9 %	5 %	37 %

13 **Sensitivity analysis**

14 Modifications of the case and non-case definitions had a significant impact on the performance parameters  
 15 of the test. The results of the sensitivity analysis are presented in Table 4.

16 **Table 4. Results of the sensitivity analysis on the test performance parameters. Parameters expressed as**  
 17 **means and 95% confidence intervals.**

Alterations of the case / non-case definitions	Original definitions (two positive tests for cases, eight negative tests for non-cases)	Decrease of the required number of tests (one positive test for cases, four negative tests for non-cases)	Increase of the required number of tests (three positive tests for cases, twelve negative tests for non-cases)
--	---	---	--

<b>Specificity</b>	99.5% (99.4 - 99.6 %)	99.1% (99.0% - 99.2%)	99.8% (99.7% - 99.9%)
<b>Upper limit of the sensitivity</b>	77.5% (71.0% - 82.9%)	54.7% (50.3% - 59.1%)	82.3% (78.8% - 86.5%)
<b>Mean effective sensitivity</b>	61.8% (56.0% - 69.9%)	43.9% (40.6% - 48.5%)	69.5% (61.9% - 78.6%)

## Discussion

The milk ELISA specificity and upper limit of the sensitivity function estimated in our study (99.5% and 77.5%) are slightly lower than previously reported for the Pourquier ELISA in the Netherlands. The values reported in the Dutch study were as follows: the specificity was estimated at 100 and 99.8% for S/P cut-offs of 30 and 20, respectively, while the sensitivity estimates were 80% in cows positive in serum ELISA and 89% in cows shedding MAP in their faeces (van Weering et al., 2007). These values are significantly higher than the sensitivity of a different milk ELISA test, estimated at 28.9% among faecal shedders (Collins et al., 2005). Previous studies have used pseudo-gold standard or latent class analysis methods to estimate the performance of JD ELISA tests (Collins et al., 2010; Lavers, 2013; Wang et al., 2006). The fact that only ELISA antibody test results were available in the study animals was a major challenge for the analysis. To overcome this, we used definitions for cases and non-cases based on combining the serological evidence from numerous tests. Such analysis might have generated dependencies between the calculations of the test performances and the estimations of the probabilities of infection of individual cows. The definitions used for cases and non-cases had a substantial impact on the estimates of the test specificity, sensitivity and MES (Table 4). Further, we could not include the impact of other factors such as the stage of lactation (van Schaik et al., 2003) in the estimation of the sensitivity and specificity. However, our average results were comparable to those of relevant studies published previously as detailed above. Therefore, we believe that they can be used to draw meaningful inferences and conclusions regarding the disease-testing programme under study.

The median age of cows was lower in a similar study based on a JD control programme in Denmark (3.6 years against 4.4 years here) (Kirkeby et al., 2015). Although both screening programs are based on the same



1 milk ELISA test, a cut-off of 30 was used in the present study while the Danish study was based on a cut-off  
2 of 15 (Nielsen et al., 2013). Despite the differences observed in terms of age distribution and test cut-off  
3 between the two studies, the mean MES in the study cited above (60%) was not significantly different from  
4 the value estimated within our study herds (61.8%). We calculated an average within-herd seroprevalence of  
5 17% in the infected herds enrolled on HerdWise. This is comparable to estimates previously published  
6 (Woodbine et al., 2009) from dairy herds in Southwest England where the true seroprevalence was  
7 estimated at 20 to 23% on different rounds of sampling. However, one of the 69 study herds was likely to be  
8 non-infected, while all the herds in our study were MAP-infected.

9 Although farmers and veterinarians can be confident that cows with repeated negative test results (group A  
10 in figure 3) are unlikely to be JD-infected and pose a low risk, the interpretation of the results from other  
11 groups is more complex. Cows with only one positive test (group B) did nonetheless have a relatively high  
12 posterior probability of infection after the positive test result (median 66%), raising suspicion concerning  
13 their true status. Although the posterior probability of infection decreased with further repeated negative  
14 results, a median of 4 additional tests (equivalent to about a year of testing) were needed to see this  
15 probability fall under 5%. Several reasons for false positive results have been described in the literature, e.g.,  
16 comparative intradermal tuberculin testing prior to JD testing (Kennedy et al., 2014; May et al., 2016),  
17 testing during the five first days in milk (Nielsen et al., 2002) and infections with other mycobacteria (Scott et  
18 al., 2010). Although cows with two consecutive positive results are considered as infected within the  
19 HerdWise programme (and within the Danish testing programme), the true status of cows with two positive  
20 results that are not consecutive is more difficult to ascertain. A third of these cows had a PPI over 95% at  
21 their last recorded test. As shown by the likelihood profiles presented in Figure 3, they are almost as likely to  
22 be truly infected as “red cows”. Similarly, cows with three positive tests or more (result groups D and E) are  
23 very likely infected: after their last test, 92% of the cows in group D had a PPI over 95%. After the third  
24 positive test, the PPI stayed above 95% in over 99% of the cows from groups D and E for the remaining of the  
25 test results. Our results suggest that the sensitivity and timeliness of the screening programme would be

1 improved by relaxing the “red cow” definition, for instance to “two positive results within four consecutive  
2 tests or at least three positive test results”.

3 Many cows with a high number of positive test results (eight or more) have been identified in our dataset.  
4 The probability of these cows being infected given their test results is almost 100% and they should probably  
5 have been selected for early culling, but other factors might have interfered with the farmer’s decision such  
6 as milk production or herd management issues. As vertical and post-partum transmissions are frequent,  
7 young calves should be protected from any contact with milk or faeces from MAP-shedding cows (Sweeney,  
8 1996; Windsor and Whittington, 2010) and it is recommended that high-risk cows are prioritised for culling  
9 before their next calving. As the “red cows” remaining in the herd represent a risk of contamination to their  
10 own calves and other calves with which they come into contact, adaptive management is highly  
11 recommended for “red cows” that are allowed to calve. In large herds with a relatively high prevalence of JD  
12 especially, it is often not feasible for farmers to cull large numbers of productive animals. In this case, test  
13 results are essential as a management tool used to target cows for which specific procedures around calving  
14 are necessary. The “red cows” in our study were kept in the herd for a mean of 10 months after diagnosis.  
15 This value suggests that the average “red cow” is milked for a substantial period before culling. A third were  
16 kept for over 12 months in the herd. It is not known whether these cows were allowed to calve again or  
17 were kept solely to milk until the end of their lactation, as no calving data were available. The figures  
18 presented here may underestimate the real values as the last ELISA test was used a proxy of the culling date  
19 (i.e., we assumed that cows that ceased to be tested had been culled).

20 Our analysis showed that a third of the cows diagnosed as “red cows” had at least one negative ELISA test  
21 after their diagnosis. Further, among the cows that were re-tested after their diagnosis, 16% did not have  
22 any further positive tests, suggesting that they may be uninfected animals. This is an important factor in the  
23 farmers’ decision to cull or keep the “red cows”, as subsequent negative tests may weaken their belief that  
24 the cow is truly MAP-infected, this is especially true of an otherwise productive cow. Farmer and vet  
25 confidence in categorising “red cows” as high risk is crucial if they are to be kept in the herd as careful

1 management at calving is necessary to prevent disease transmission. Improving awareness and  
2 understanding of this feature of JD testing is also important to avoid distrust and lack of motivation because  
3 of inconsistent test results. Imperfect specificity of the case definition means that some cows might be  
4 diagnosed as “red cows” and culled although they were not truly infected. Further analysis would be useful  
5 to determine the optimal economic balance between specificity (minimising unnecessary culls and additional  
6 calving management) and sensitivity (keeping infectious animals in the herd without specific management)  
7 of the case definition. The imperfect test characteristics and the chronic nature of the disease render  
8 repeated testing necessary. However, as illustrated by the individual cow-profiles selected as examples, if  
9 the reason for additional testing is to increase the confidence on the status of an individual cow, there is  
10 considerable potential for more targeted use of serological testing. Scenario analysis could provide  
11 quantitative estimates of the costs and benefits associated with alternative testing scenarios. Such analysis  
12 would require further data collection, including economic data. In the meanwhile, more strategic use of  
13 testing by farmers and veterinarians would be facilitated by expressing the combined test results as a  
14 posterior probability of infection.

## 15 **Funding sources**

16 This study was conducted within a project funded by AHDB Dairy, a division of the Agriculture and  
17 Horticulture Development Board.

18

19

## References

- Aly, S.S., Anderson, R.J., Whitlock, R.H., Fyock, T.L., McAdams, S.C., Byrem, T.M., Jiang, J., Adaska, J.M., Gardner, I.A., 2012. Cost-effectiveness of diagnostic strategies to identify *Mycobacterium avium* subspecies paratuberculosis super-shedder cows in a large dairy herd using antibody enzyme-linked immunosorbent assays, quantitative real-time polymerase chain reaction, and bacte. *J. Vet. Diagnostic Investig.* 24, 821–832. doi:10.1177/1040638712452107
- Brown, L.D., Cai, T.T., DasGupta, A., 2001. Interval Estimation for a Binomial Proportion. *Stat. Sci.* 16, 101–133. doi:10.1214/ss/1009213286
- Çetinkaya, B., Egan, K., Harbour, D.A., Morgan, K.L., 1996. An abattoir-based study of the prevalence of subclinical Johne's disease in adult cattle in south west England. *Epidemiol. Infect.* 116, 373. doi:10.1017/s0950268800052705
- Cetinkaya, B., Erdogan, H.M., Morgan, K.L., 1998. Prevalence, incidence and geographical distribution of Johne's disease in cattle in England and the Welsh borders. *Vet. Rec.* 143, 265–269. doi:10.1136/vr.143.10.265
- Collins, M., Djuranovic, N., Estey, L., Leathers, V., 2010. Multi-level interpretation of the new IDEXX M. paratuberculosis ELISA on serum and milk based on likelihood ratio analysis. 43rd Annu. Conf. Am. Assoc. Bov. Pract. Albuquerque, USA. 208 p.
- Collins, M.T., Gardner, I.A., Garry, F.B., Roussel, A.J., Wells, S.J., 2006. Consensus recommendations on diagnostic testing for bovine paratuberculosis in the U.S. *J Am Vet Med Assoc* 229, 14pp. doi:10.2460/javma.229.12.1912
- Collins, M.T., Wells, S.J., Petrini, K.R., Collins, J.E., Schultz, R.D., Whitlock, R.H., 2005. Evaluation of five antibody detection tests for diagnosis of bovine paratuberculosis. *Clin. Diagn. Lab. Immunol.* 12, 685–692. doi:10.1128/CDLI.12.6.685-692.2005

1 Daniels, M.J., Hutchings, M.R., Allcroft, D.J., McKendrick, J., Greig, A., 2002. Risk factors for Johne's disease in  
2 Scotland--the results of a survey of farmers. *Vet. Rec.* 150, 135–139.

3 Dohoo, I., Martin, W., Stryhn, H., 2003. *Veterinary epidemiologic research*. VER, Incorporated.

4 Eamens, G.J., Whittington, R.J., Marsh, I.B., Turner, M.J., Saunders, V., Kemsley, P.D., Rayward, D., 2000.  
5 Comparative sensitivity of various faecal culture methods and ELISA in dairy cattle herds with endemic  
6 Johne's disease. *Vet. Microbiol.* 77, 357–367.

7 Hanks, J.D., Taylor, N.M., Kossaibati, M.A., 2013. Is targeted milk sampling an effective means of detecting  
8 Johne's disease in dairy herds? *Cattle Pract.* 22, 26–34.

9 Harris, N.B., Barletta, R.G., 2001. *Mycobacterium avium* subsp. *paratuberculosis* in Veterinary Medicine. *Clin.*  
10 *Microbiol. Rev.* 14, 489–512. doi:10.1128/CMR.14.3.489-512.2001

11 Kennedy, A.E., Da Silva, A.T., Byrne, N., Govender, R., MacSharry, J., O'Mahony, J., Sayers, R.G., 2014. The  
12 Single Intradermal Cervical Comparative Test Interferes with Johne's Disease ELISA Diagnostics. *Front.*  
13 *Immunol.* 5, 564. doi:10.3389/fimmu.2014.00564

14 Kirkeby, C., Graesboll, K., Halasa, T., Toft, N., Nielsen, S.S., 2015. Mean effective sensitivity for  
15 *Mycobacterium avium* subsp. *paratuberculosis* infection in cattle herds. *BMC Vet. Res.* 11, 190.  
16 doi:10.1186/s12917-015-0512-8

17 Koets, A.P., Eda, S., Sreevatsan, S., 2015. The within host dynamics of *Mycobacterium avium* ssp.  
18 *paratuberculosis* infection in cattle: where time and place matter. *Vet. Res.* 46. doi:10.1186/s13567-015-  
19 0185-0

20 Kudahl, A.B., Nielsen, S.S., Østergaard, S., 2008. Economy, Efficacy, and Feasibility of a Risk-Based Control  
21 Program Against Paratuberculosis. *J. Dairy Sci.* 91, 4599–4609. doi:10.3168/jds.2008-1257

1 Kudahl, A.B., Sørensen, J.T., Nielsen, S.S., Østergaard, S., 2007. Simulated economic effects of improving the  
2 sensitivity of a diagnostic test in paratuberculosis control. *Prev. Vet. Med.* 78, 118–129.  
3 doi:10.1016/j.prevetmed.2006.10.004

4 Lavers, C., 2013. Evaluation of diagnostic tests for detection of *Mycobacterium avium* subsp.  
5 paratuberculosis (MAP) at the herd-level and cow-level. PhD thesis, Univ. Prince Edward Isl.

6 May, H., Orpin, P., Pearse, H., Jones, P., Clough, H., 2016. The influence of tuberculosis testing in dairy cattle  
7 on milk ELISA tests for Johne's disease [WWW Document]. URL  
8 <http://academic.research.microsoft.com/Publication/10575053> (accessed 2.15.17).

9 Mitchell, R.M., Schukken, Y., Koets, A., Weber, M., Bakker, D., Stabel, J., Whitlock, R.H., Louzoun, Y., 2015.  
10 Differences in intermittent and continuous fecal shedding patterns between natural and experimental  
11 *Mycobacterium avium* subspecies paratuberculosis infections in cattle. *Vet. Res.* 46. doi:10.1186/s13567-  
12 015-0188-x

13 Nielsen, S., Enevoldsen, C., Gröhn, Y.T., 2002a. The *Mycobacterium avium* subsp. paratuberculosis ELISA  
14 response by parity and stage of lactation. *Prev. Vet. Med.* 54, 1–10. doi:10.1016/S0167-5877(02)00008-9

15 Nielsen, S., Gröhn, Y., Enevoldsen, C., 2002b. Variation of the Milk Antibody Response to Paratuberculosis in  
16 Naturally Infected Dairy Cows. *J. Dairy Sci.* 85, 2795–2802. doi:10.3168/jds.s0022-0302(02)74366-x

17 Nielsen, S., Toft, N., 2008. Ante mortem diagnosis of paratuberculosis: A review of accuracies of ELISA,  
18 interferon- $\gamma$  assay and faecal culture techniques. *Vet. Microbiol.* 129, 217–235.  
19 doi:10.1016/j.vetmic.2007.12.011

20 Nielsen, S.S., 2008. Transitions in diagnostic tests used for detection of *Mycobacterium avium* subsp.  
21 paratuberculosis infections in cattle. *Vet. Microbiol.* 132, 274–282. doi:10.1016/j.vetmic.2008.05.018

22 Nielsen, S.S., Enevoldsen, C., Gröhn, Y.T., 2002. The *Mycobacterium avium* subsp. paratuberculosis ELISA  
23 response by parity and stage of lactation. *Prev. Vet. Med.* 54, 1–10. doi:10.1016/S0167-5877(02)00008-9

1 Nielsen, S.S., Krogh, M.A., Enevoldsen, C., 2009. Time to the occurrence of a decline in milk production in  
2 cows with various paratuberculosis antibody profiles. *J. Dairy Sci.* 92, 149–155. doi:10.3168/jds.2008-1488

3 Nielsen, S.S., Toft, N., 2009. A review of prevalences of paratuberculosis in farmed animals in Europe. *Prev.*  
4 *Vet. Med.* 88, 1–14. doi:10.1016/j.prevetmed.2008.07.003

5 Nielsen, S.S., Toft, N., Okura, H., 2013. Dynamics of Specific Anti-*Mycobacterium avium* Subsp.  
6 paratuberculosis Antibody Response through Age. *PLoS One* 8, e63009. doi:10.1371/journal.pone.0063009

7 Ott, S.L., Wells, S.J., Wagner, B.A., 1999. Herd-level economic losses associated with Johne's disease on US  
8 dairy operations. *Prev. Vet. Med.* 40, 179–192. doi:10.1016/s0167-5877(99)00037-9

9 Robins, J., Bogen, S., Francis, A., Westhoek, A., Kanarek, A., Lenhart, S., Eda, S., 2015. Agent-based model for  
10 Johne's disease dynamics in a dairy herd. *Vet. Res.* 46. doi:10.1186/s13567-015-0195-y

11 Rogan, W., Gladen, B., 1978. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 107,  
12 71–76.

13 Scott, M.C., Bannantine, J.P., Kaneko, Y., Branscum, A.J., Whitlock, R.H., Mori, Y., Speer, C.A., Eda, S., 2010.  
14 Absorbed EVELISA: A Diagnostic Test with Improved Specificity for Johne's Disease in Cattle. *Foodborne*  
15 *Pathog. Dis.* 7, 1291–1296. doi:10.1089/fpd.2010.0541

16 Smith, R.L., Al-Mamun, M.A., Gröhn, Y.T., 2017. Economic consequences of paratuberculosis control in dairy  
17 cattle: A stochastic modeling study. *Prev. Vet. Med.* 138, 17–27. doi:10.1016/j.prevetmed.2017.01.007

18 Stabel, J., 2000. Transitions in immune responses to *Mycobacterium paratuberculosis*. *Vet. Microbiol.* 77,  
19 465–473. doi:10.1016/s0378-1135(00)00331-x

20 Stott, A.W., Jones, G.M., Humphry, R.W., Gunn, G.J., 2005. Financial incentive to control paratuberculosis  
21 (Johne's disease) on dairy farms in the United Kingdom. *Vet. Rec.* 156, 825–831. doi:10.1136/vr.156.26.825

22 Sweeney, R.W., 1996. Transmission of Paratuberculosis. *Vet. Clin. North Am. Food Anim. Pract.* 12, 305–312.  
23 doi:10.1016/s0749-0720(15)30408-4

1 Sweeney, R.W., Whitlock, R.H., McAdams, S., Fyock, T., 2006. Longitudinal Study of ELISA Seroreactivity to  
2 *Mycobacterium Avium* subsp. Paratuberculosis in Infected Cattle and Culture-Negative Herd Mates. *J. Vet.*  
3 *Diagnostic Investig.* 18, 2–6. doi:10.1177/104063870601800102

4 van Schaik, G., Rossiter, C.R., Stehman, S.M., Shin, S.J., Schukken, Y.H., 2003. Longitudinal study to  
5 investigate variation in results of repeated ELISA and culture of fecal samples for *Mycobacterium avium*  
6 subsp paratuberculosis in commercial dairy herds. *Am. J. Vet. Res.* 64, 479–484.  
7 doi:10.2460/ajvr.2003.64.479

8 van Weering, H., van Schaik, G., van der Meulen, A., Waal, M., Franken, P., van Maanen, K., 2007. Diagnostic  
9 performance of the Pourquier ELISA for detection of antibodies against *Mycobacterium avium* subspecies  
10 paratuberculosis in individual milk and bulk milk samples of dairy herds. *Vet Microbiol* 125, 49–58.  
11 doi:S0378-1135(07)00239-8 [pii]\r10.1016/j.vetmic.2007.05.010

12 Velasova, M., Damaso, A., Chengat Prakashbabu, B., Gibbons, J., Wheelhouse, N., Longbottom, D., Guitian, J.,  
13 2017. Herd-level prevalence of selected endemic infectious diseases of dairy cows in Great Britain. *J. Dairy*  
14 *Sci.* in press.

15 Wang, C., Turnbull, B.W., Gröhn, Y.T., Nielsen, S.S., 2006. Estimating Receiver Operating Characteristic  
16 Curves with Covariates When There Is No Perfect Reference Test for Diagnosis of Johne’s Disease. *J. Dairy*  
17 *Sci.* 89, 3038–3046. doi:10.3168/jds.s0022-0302(06)72577-2

18 Windsor, P.A., Whittington, R.J., 2010. Evidence for age susceptibility of cattle to Johne’s disease. *Vet. J.* 184,  
19 37–44. doi:10.1016/j.tvjl.2009.01.007

20 Woodbine, K.A., Schukken, Y.H., Green, L.E., Ramirez-Villaescusa, A., Mason, S., Moore, S.J., Bilbao, C.,  
21 Swann, N., Medley, G.F., 2009. Seroprevalence and epidemiological characteristics of *Mycobacterium avium*  
22 subsp. paratuberculosis on 114 cattle farms in south west England. *Prev. Vet. Med.* 89, 102–109.  
23 doi:10.1016/j.prevetmed.2009.02.005