1 **The interpretation of serial Johne's disease milk antibody results is affected by test**

2 **characteristics, pattern of test results and parallel bovine tuberculosis testing**

3

4 M. Barden, R. F. Smith and H. M. Higgins

5

6 Department of Livestock and One Health, Institute of Infection, Veterinary and Ecological

7 Sciences, University of Liverpool, Leahurst Campus, Chester High Road, Neston, Cheshire,

8 United Kingdom, CH64 7TE

9

10 Email address: mbarden@liverpool.ac.uk

11

12 Abbreviations used:

13 bTB, bovine tuberculosis; CIS, Cattle Information Service; df, degrees of freedom; ELISA,

14 enzyme-linked immunosorbent assay; JD, Johne's disease; JD-mELISA, Johne's Disease

15 milk antibody enzyme-linked immunosorbent assay; MAP, *Mycobacterium avium* subsp.

16 *paratuberculosis*; NMR, National Milk Record group; OIE, World Organisation for Animal

17 Health; PPD, purified protein derivative; PPI, posterior probability of infection; QMMS, Quality

18 Milk Management Services; SICCT, single intradermal cervical comparative tuberculin; S/P,

19 sample-to-positive ratio; UK, United Kingdom

20

21

22

23

31

32

33

34

35

36

37

Abstract

Background: In the UK, quarterly Johne's disease milk antibody ELISAs (JD-mELISAs) are commonly used to classify animals which are likely to be infectious, termed "red cows". "Red cows" are classified following two positive results from the previous four tests (e.g. + - - +). All cattle are also regularly screened for bovine tuberculosis using intradermal avian and bovine tuberculin, and it is advised to maintain a 60 day interval between a tuberculosis test and JD-mELISA.

Aims: To evaluate the impact of bovine tuberculosis testing on JD-mELISAs, and to quantify the impact of test specificity and "red cow" classification test pattern on the probability of infection.

Methods: Four years of individual cow milk records with JD-mELISA results were collated from 735 dairy farms and matched to tuberculosis testing records. A two-level multivariable logistic regression model quantified the effect of tuberculosis testing on JD-mELISA result. The specificity and age-dependent sensitivity of a single JD-mELISA were estimated and used to calculate likelihood ratios following each test. Using Bayes' theorem, the posterior probability of infection with Johne's disease was calculated for different specificities, ages of cow, and patterns of test results.

Results: There were increased odds of a positive JD-mELISA if it was ≤30 days (OR: 2.1) or 31-60 days (OR: 1.2) after a tuberculosis test, compared to >90 days. A larger avian skin reaction at the tuberculosis test was also associated with increased odds of a positive JD-mELISA. The proportion of cows which tested exclusively negative after their first positive JD-mELISA was higher if that JD-mELISA was ≤30 days after a tuberculosis test compared to >90 days. The posterior probability of infection reduced substantially when the test specificity was slightly reduced. In "red cows" classified following two consecutive positive tests, if the test specificity was reduced to 0.95, then the posterior probability of infection was only >95% if the prior probability was >13%. If the "red cow" classification was due to two non-consecutive

64   positive tests (+ - - +), the posterior probability of infection was only >95% if the prior probability

65   was >43%.

66   **Conclusions:** Testing for Johne's disease within 60 days of a tuberculosis test is associated

67   with a higher chance of a positive JD-mELISA and this may reflect a reduction in the ELISA

68   specificity. Relatively small reductions in JD-mELISA specificity can markedly reduce the

69   posterior probability of infection which also depends on the pattern of test results which

70   classifies "red cows".

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

1. Introduction

Johne's disease (JD) is a chronic wasting disease of cattle caused by *Mycobacterium avium subspecies paratuberculosis* (MAP). The estimated herd prevalence in the UK is 68%, based on the presence of antibodies in the bulk milk (Velasova et al., 2017), and the average seroprevalence within herds is reported to be 22.4% when adjusted for test sensitivity and specificity (Woodbine et al., 2009). In addition to the welfare concerns of chronic wasting, JD has a significant economic impact on dairy farms and there is increasing concern about MAP contributing to the aetiology of Crohn's disease in humans (Liverani et al., 2014; McAloon et al., 2019).

Although it is not always the case, the typical infection dynamics of MAP are considered to be the infection of young animals followed by a latent period, which can last several years, during which the disease status of the animal is difficult to determine (Fecteau, 2018; Mortier et al., 2013). In some animals, MAP will eventually evade control by the innate immune system, the infection progresses and the animal becomes infectious (Arsenault et al., 2014). At this stage diagnosis may be possible through detection of the organism in faeces or a detectable humoral immune response to the organism (Jenvey et al., 2018; Nielsen, 2008). One common diagnostic approach for JD is to identify this humoral immune response using an enzyme-linked immunosorbent assay (ELISA) which can detect antibodies in both milk and serum. The sensitivity and specificity of a single JD milk antibody ELISA (JD-mELISA) ranges from 0.21 to 0.61, and 0.83 - 1.00 respectively, for animals with detectable MAP in faeces (Nielsen and Toft, 2008). The sensitivity has been reported to reach 0.79 in older animals for a single JD-mELISA based on modelling which used future test results to define truly infected and non-infected animals (Nielsen et al., 2013).

112   Farms which are attempting to control JD require identification of infectious animals before

113   they shed the organism in proximity to susceptible animals. Neonatal calves are considered

114   to be at the highest risk of infection (Mortier et al., 2013; Windsor and Whittington, 2010),

115   therefore animals which are likely to be shedding MAP need to be identified prior to calving.

116   JD-mELISAs are usually conducted repeatedly at regular intervals, often quarterly, on dairy

117   farms in the UK (Geraghty et al., 2014; Whittington et al., 2019). This is partly due to the

118   convenience of testing milk in dairy cows, but serial testing also partially mitigates the poor

119   sensitivity of a single JD-mELISA during the latent phase of infection. This allows the

120   identification of infectious animals at specific, critical timepoints such as prior to breeding or

121   drying off. In the UK, these high-risk cows are often termed "red cows" to highlight their

122   potential shedding status. Nielsen (2008) demonstrated that 35% of cows with consecutive

123   positive JD-mELISAs, approximately one month apart, would shed MAP within one year. This

124   was incorporated into control schemes in the UK so that cows with two consecutive positive

125   tests (2 in 2) were classified as "red cows" and treated as high-risk of being, or becoming,

126   infectious. Once cows are classified as "red cows" their risk status is not downgraded

127   regardless of future test results (Orpin et al., 2020a), therefore there needs to be a high degree

128   of confidence in the validity of this classification.

129

130   One advantage of using quarterly JD-mELISAs is that the infection status of the animal after

131   each test, the posterior probability of infection (PPI), can be continually updated which each

132   consecutive result. A recent study in the UK investigated the probability of infection following

133   serial JD-mELISAs and concluded that the probability of infection after two consecutive

134   positive tests (2 in 2) was similar to the probability following two positive tests that were not

135   consecutive (Meyer et al., 2018). However, the reported standard deviations in this study were

136   large which suggests a wide dispersion of the individual results. In 2018 the largest milk

137   recording organisation in the UK, National Milk Records (NMR), changed their test

138   interpretation and subsequent advice to farmers to classify all cows with two positive tests in

139 the previous four (2 in 4) as "red cows" (Anon., 2018). Therefore there are currently three

140 patterns of test results which can result in a cow being classified as a "red cow": positive-

141 positive (+ +), positive-negative-positive (+ - +), and positive-negative-negative-positive (+ - -

142 +).

143

144 Another mycobacterium, *Mycobacterium bovis* which causes bovine tuberculosis (bTB), is

145 endemic in cattle in the UK and under statutory surveillance with the single intradermal

146 comparative cervical tuberculin (SICCT) test (Nuñez-Garcia et al., 2018). The SICCT test uses

147 two purified protein derivatives (PPD) from *Mycobacterium bovis* and *Mycobacterium avium*

148 injected into the cervical skin. The skin thickness is measured prior to inoculation and again

149 72 h later to detect a delayed-type hypersensitivity immune response to each PPD (de la Rua-

150 Domenech et al., 2006). The testing regime is government mandated under EU legislation

151 (Council Directive 64/432/EEC), and testing frequency is dependent on the regional

152 prevalence and previous test results in the herd. Test frequencies range from four-yearly, if

153 no positive animals have been previously identified and the herd is in a low-risk area, to every

154 60 days if positive animals have been identified during the previous test.

155

156 It is advised to delay use of JD-mELISAs after a SICCT test due the potential immune priming

157 nature of PPD inoculation (Varges et al., 2009). The advice regarding the exact interval

158 between the two tests varies; many countries advise 90 days but in the UK this is often

159 impractical due to the frequency of bTB testing, therefore a 60 day interval is generally

160 recommended. Although it has been demonstrated that there is a rise in MAP antibodies

161 following intradermal inoculation with avian and bovine PPD (Kennedy et al., 2014; May et al.,

162 2016; Roupie et al., 2018), it is not clear whether this represents an elevation due to antibody

163 cross-reaction or an anamnestic response due to stimulation from PPD inoculation in

164 genuinely MAP-infected animals. In other words, the effect of a recent SICCT test could be to

165 decrease the specificity of the JD-mELISA, to increase its sensitivity, or a combination of both.

166

167 Historically, a similar diagnostic approach to bTB has been used for JD: a MAP derived Johnin-

168 PPD was inoculated intradermally and the reaction to it recorded after 72 h in a similar way to

169 the SICCT test (Kalis et al., 2003). This skin test is no longer used for JD diagnosis due to

170 issues with Johnin-PPD production and test specificity because of the homology of MAP and

171 other mycobacteria, particularly other *M. avium* subspecies (Collins, 1996). MAP is a

172 subspecies of *M. avium* and different MAP strains have between 34-62% spectral profile

173 similarity with *M. avium sbsp. avium* assessed with matrix-assisted laser desorption/ionization

174 time-of-flight mass spectrometry (Ravva et al., 2017). There are antigenic similarities between

175 the avian, bovine and Johnin PPDs (Gilot and Cocito, 1993), and both avian and Johnin PPDs

176 stimulate gamma interferon release (Bassey and Collins, 1997; Jungersen et al., 2002).

177 Therefore, reactions at the avian site of the SICCT test may be due to infection with MAP and

178 could be relevant to the interpretation of subsequent JD-mELISAs.

179

180 The aims of this study were to quantify the association between bTB testing and JD-mELISA

181 results, specifically the interval between the two tests and avian skin reactions recorded during

182 the SICCT test. The patterns of JD-mELISA results subsequent to the first time a cow has

183 positive test, and following "red cow" classification, were assessed for trends which may

184 indicate the likely infection status. Additionally, Bayes' theorem was applied to determine the

185 posterior probability of infection following different JD-mELISA patterns which classify "red

186 cows" and using different test specificities.

187

2. Material and methods

The study was conducted following ethical approval from the University of Liverpool Veterinary Research Ethics Committee (VREC567). Farms were recruited if they supplied milk to the British supermarket Tesco (Tesco Sustainable Dairy Group) during the period 2014 - 2017. These farms were required to test for JD four times a year using milk ELISAs as an adjunct to their regular milk recording. The JD-mELISA results, with accompanying milk records, were requested directly from milk recording organisations: National Milk Recording group (NMR); the Cattle Information Service (CIS) and Quality Milk Management Services (QMMS). Records included cow and herd identification, date of birth, milk recording date and the sample-to-positive ratio (S/P) of the JD-mELISA. All three milk recording organisations used the same indirect ELISA (IDEXX Laboratories, Maine, USA) and reported the same S/P cut-off for positive results (≥30%).

SICCT testing details were requested from the Animal Plant Health Agency (APHA) for all farms included in the dataset for the period 2012 to 2017. This extended period was used to include SICCT tests for two years preceding the first JD-mELISA. The data from the APHA included cow identification, herd identification, breed, date of SICCT test and the difference in skin thickness between day one and two of the SICCT test at the avian and bovine PPD inoculation sites. These data were merged by cow ear tag with the milk recording data (Figure 1). The final dataset had one record for each JD-mELISA, with each cow having one or more records and the interval from the most recent SICCT test, with associated avian and bovine skin reaction details, included for each record.

2.1. Association between bTB test and JD-mELISA

A two-level multivariable logistic regression model was fitted using Stata (StataCorp LP, 2016), the binary dependent variable was JD-mELISA result using a S/P ratio ≥30% as the threshold

which is common practice in the UK dairy industry. The interval between SICCT test and JD-mELISA was transformed into a categorical variable with bins chosen to reflect the specific time intervals of interest: ≤30 days, 31-60 days, 61-90 days and >90 days. The difference in skin thickness at both the avian and bovine PPD between day one and two of the SICCT test inoculation sites were also binned to minimise the effects of different testers and small measurement variabilities between the day one and two of the test. An increase of 4mm is advised when using a single intradermal tuberculin test for bTB diagnosis (OIE, 2019) and has been used for JD diagnosis with intradermal Johnin-PPD (Kalis et al., 2003); so the specific bands used for these categorical variables were ≤2mm, 2 - 4mm, and >4 mm.

The purpose of the model was to quantify the effects of SICCT to JD test interval and SICCT test skin reaction on JD-mELISA result. To accurately estimate these effects, all potential confounders present in the dataset, such as age of cow and stage of lactation, were considered as covariables and explored in the univariable analysis. Linearity was checked between dependent and independent variables and continuous variables were discretised when necessary. If the univariable analysis was statistically significant, at p-value <0.1, the variable was retained in the multivariable analysis. All covariables were forced into the multivariable model and retained if they remained statistically significant at p-value <0.05. To avoid the effects of clustering within cow, and the difficulties of incorporating autocorrelation into multi-level modelling of repeated binary data, one test was selected randomly from each cow for inclusion in the final model which took the form:

$$logit(y_{ij}) = \beta_0 + \boldsymbol{\beta_i} \boldsymbol{x_i^T} + \mu_j \ (1)$$

238 where $logit(y_{ij})$ is the natural log of the odds of cow i on farm j having a positive JD-mELISA,

239 β_0 is a constant, x_i^T is the vector of covariables for each cow and β_i is the vector of coefficients

240 for these covariables, μ_j is a random intercept for each farm. All two-way interactions between

241 covariables which included at least one of the primary independent variables of interest

242 (SICCT to JD test interval and SICCT test skin reaction) were explored using a Wald test and

243 by plotting the predicted values. The final covariables were checked for collinearity which was

244 minimal.

245

246 *2.2. Analysis of JD-mELISA results subsequent to the first positive result and "red cow"*

247 *classification*

248 Firstly, the number of "red cows" (2 in 4) was calculated from the test results in the whole

249 dataset (Figure 1). The age at which cows became a "red cow" and the proportion of positive

250 tests afterwards were also calculated. Secondly, to focus on future test results in more detail,

251 cows which had at least one positive test, and at least one more test subsequently, were

252 extracted from the full dataset (Figure 1). Cows were categorised by the interval between

253 SICCT test and JD-mELISA (≤30d, 31-60d, 61-90d or >90d) at the first positive test and then

254 further divided by the reaction at the avian tuberculin site using a 4mm threshold. The

255 proportion of subsequent tests which were positive was calculated and categorised as 0%, 1-

256 99% or 100% as the main outcomes of interest were cows which had no more positive tests

257 and those which had only positive tests. Chi-squared tests were used to assess the differences

258 in frequency across categories and to assess comparisons between specific categories. The

259 median age at first positive test and median number of all subsequent tests were calculated

260 and compared across SICCT test to JD-ELISA intervals (bTB-JD interval) using the Kruskal-

261 Wallis test. Finally, the process was repeated after excluding cows which tested positive for

262 the first time during the final year of the study window (2017), therefore ensuring all cows had

263 the potential to be tested for at least one more year after their first positive test.

264

265

2.3. Sensitivity and specificity estimation of a single JD -mELISA

The case definition was intended to identify cows that were infected with MAP and progressing towards becoming infectious and clinically affected; this progression is associated with a shift to a humoral immune response (Koets et al., 2015; Smith et al., 2016). A non-case definition was chosen to include cows which were either truly not infected with MAP or had a non-progressing infection which did not result in faecal shedding or persistent MAP antibodies (Arsenault et al., 2014; Smith et al., 2016). Cows in the full dataset were therefore classified as JD cases or non-cases based on the pattern of test results, defined using the same criteria as Meyer et al. (2018). Cows with a minimum of three tests were classified as cases if the final two test results were positive, and cows with a minimum of nine tests were classified as non-cases if the final eight test results were negative (Figure 1).

277

Specificity and age-dependent sensitivity of the milk JD-mELISA were estimated using the approach described by Nielsen et al. (2013). In order to estimate specificity, a randomly selected test result was selected from each non-case and the specificity was calculated as the probability of the selected test being negative; confidence limits were calculated using the Agresti-Coull interval method (Brown et al., 2001) using the *Hmisc* package in R (Harrell, 2020; R Development Core Team 3.6.1, 2019). To estimate sensitivity, a randomly selected test result was selected from each case and the sensitivity of the test at age t (*Se(t)*) was calculated as the probability of the selected test being positive test at a given age. *Se(t)* was estimated using a non-linear logistic regression model (Equation 2) with the NLMIXED procedure in SAS (version 9; SAS Institute Inc., Cary, NC), where a is the upper limit of the logit of sensitivity at maximum age (t), b is the scaling factor and c is the coefficient of decay as age (t) increases.

$$logit\left(Se(t)\right) = a - b * e^{-c*t} \quad (2)$$

290 The estimated test sensitivity as a function of age at testing (Equation 2) was plotted in R,

291 confidence intervals were estimated using the delta method of approximating prediction

292 variance (SAS Institute Inc., 2015). For comparison and context, the same equation was

293 plotted alongside using parameters estimated in studies with a comparable methodology:

294 Nielsen et al. (2013) and Meyer et al. (2018).

295

296 *2.4. Conditional probability using serial Johne's disease test results*

297 To calculate the posterior probability of infection (PPI) after each JD-mELISA, the likelihood

298 ratio given the test result was calculated using Equation 3 for positive (LR_+) and negative

299 results (LR_-); where T is the test result and D is the true disease status of the cow (Dohoo et

300 al., 2012).

301

302 $$LR_+ = \frac{P(T+|D+)}{P(T+|D-)} = \frac{P(T+|D+)}{1- P(T-|D-)} \ , \ LR_- = \frac{P(T-|D+)}{P(T-|D-)} = \frac{1- P(T+|D+)}{P(T-|D-)} \qquad (3)$$

303

304 The estimates for test specificity ($\widehat{Sp}$) and age-dependent sensitivity ($\widehat{Se}(t)$) calculated in

305 section 2.3 were substituted into Equation 3 to calculate the likelihood ratio for each test result

306 given the age (t) of the animal at the test:

307 $$Se(t) = P(T+|D+) \ , \qquad Sp = P(T- | D-)$$

308 $$\qquad\qquad\qquad (4)$$

309 $$LR_+(t) = \frac{\widehat{Se}(t)}{1- \widehat{Sp}} \ , \ LR_-(t) = \frac{1- \widehat{Se}(t)}{\widehat{Sp}}$$

310

311 Bayes' theorem was applied whereby the posterior odds of infection following each test result

312 were calculated as the product of the likelihood ratio and prior odds. The posterior odds were

14

divided by one plus the odds to calculate the posterior probability, termed the posterior probability of infection (PPI). The PPI following different test results were calculated for any given prior probability assuming quarterly testing using R This allowed the effect of test pattern and test specificity to be quantitatively evaluated for different prior probabilities and with different ages of cow.

In the UK, dairy cows which are considered to be high risk of being infectious for Johne's disease are termed "red cows". This irreversible classification is based on a defined pattern of test results. Previously, the "red cow" classification required two consecutive positive tests (2 in 2), but the current industry definition is two positive tests in the previous four (2 in 4). Therefore, there are fundamentally three different test patterns which can define a cow as "red" using the current 2 in 4 "red cow" criteria:

1.   + +

2.   + - +

3.   + - - +

The first pattern would have also met the previous "red cow" definition (2 in 2), but patterns two and three would only classify a cow as "red" following the change in definition. Compared to pattern two (+ - +), pattern three (+ - - +) would result in a cow meeting the "red cow" classification with a higher number of recent negative tests. Therefore to explore the differences in probability of infection using different "red cow" test result patterns, the "+ - - +" test pattern was chosen to highlight the greatest difference in PPI compared to cows with two consecutive positive tests (+ +). Consequently, the PPI at the point of a cow being classified as "red" after the "+ +" and "+ - - +" result patterns were modelled, including the effect of using different test specificities to the one calculated in Section 2.3.

3. Results

3.1. Association between bTB test and JD-mELISA

The dataset was collated from tests that were conducted between 01/01/2014 – 31/12/2017 and consisted of 1,257,354 JD-mELISAs from 225,296 cows on 735 farms. In total, 3.3% of the JD-mELISAs were positive and 9.4% of cows had at least one positive test result. There were 928,471 JD-mELISAs which could be matched by cow ear tag to the most recent SICCT test.. Of these, 2.2% of JD-mELISAs (20,599 tests) were conducted within 30 days of a SICCT test, and a further 12.5% (115,642) were conducted within 60 days. The proportion of JD-mELISAs which were positive was 5.8% when the SICCT test to JD-mELISA interval was <30 days, 4.0% when it was 30 - 60 days, 3.3% when it was 60 - 90 days and 3.1% when it was >90 days (X^2 = 672.9, df = 3, p-value <0.001).The majority of SICCT tests recorded no or minimal changes in skin thickness (avian and bovine skin reactions ≤2mm), but 4.4% (41,471 tests) recorded an avian skin reaction greater than 4mm and 1.3% (11,725 tests) recorded a bovine skin reaction greater than 4mm.

Univariable analysis was conducted for the following factors: bTB-JD test interval, difference in skin thickness at both the avian and bovine injection site between days one and two of the SICCT test, age in years at JD test, days since calving, breed, geographical region of the UK and milk recording organisation (fixed effects), and farm as a random effect. Each of these factors had a statistically significant (p-value <0.05) influence on the odds of a positive JD-mELISA (Table 1) and consequently these were initially all forced into the multivariable model. The only variable which was no longer significant in the multivariable model was bovine skin reaction (p-value = 0.754), which was highly correlated with the avian skin reaction (r = 0.71), and this was dropped from the final model. Two-way interactions which included bTB-JD test interval or avian skin reaction were explored. The only interaction which was significant in the two-level multivariable model was between bTB-JD test interval and avian skin reaction. The

364 model was fitted using only complete records and if the JD test record was matched to a bTB

365 test record, hence the final model used records from 158,287 cows on 664 farms. All

366 covariables were all statistically significant (p-value <0.05); the coefficients and odds ratios for

367 these parameters are presented in Table 2.

368

369 *3.2. Analysis of JD-mELISA results subsequent to the first positive result and "red cow"*

370 *classification*

371 Table 3 displays characteristics and test results of "red cows", based on the test pattern that

372 classified them as such. In the full dataset, 7,308 cows met the classification criteria for "red

373 cows" using the current definition of 2 in 4 positive tests, which represents 32.6% of all cows

374 which had at one or more positive tests in the study period (N = 22,401). Of cows which had

375 one or more positive tests in the study period and also recorded at least four more subsequent

376 tests (N = 9,296), 35.3% became "red cows" (2 in 4). The majority of "2 in 4 red cows" (80.0%,

377 5,846) were classified as "red" after two consecutive positive tests, with the remainder

378 classified following either the "+ - +" (867) or "+ - - +" (595) result patterns. Although 30.7%

379 (2,241) of "red cows" did not record any further tests after meeting the criteria to be classified

380 as a "red cow" (2 in 4), 31.2% (2,282) recorded at least four more tests. Of those cows which

381 recorded at least four more tests, 22.8% (520) had no more positive tests and 21.5% (491)

382 had only positive tests (Table 3). Regardless of the test pattern that classified cows as "red"

383 (2 in 4), 87.3% cows (6,382) had two consecutive tests at some stage and therefore would

384 have also met the previous "red cow" definition (2 in 2), albeit at a later test. Therefore, the

385 change in "red cow" definition equated to a 14.5% increase in the overall number of "red cows"

386 within the limits of this dataset.

387

388 The proportion of tests that were positive after the first positive test for each cow is displayed

389 in Table 4, sub-divided by SICCT test to JD-mELISA interval (bTB-JD interval) and avian skin

17

390   reaction at the SICCT test. The general trends indicated that the proportion of cows which had

391   either only positive or only negative subsequent tests varied depending on bTB-JD interval (X^2

392   = 141.2, df = 3, p-value <0.001). For example, a greater proportion of cows did not record any

393   further positive tests if the bTB-JD interval was ≤30 days compared to >90d, and particularly

394   if the avian skin reaction was ≤4mm at the SICCT test (X^2 = 65.0, df = 1, p-value <0.001).

395   Additionally, regardless of bTB-JD interval, an avian skin reaction >4mm was associated with

396   more cows having only positive tests after their first positive test than cows having only

397   negative tests (X^2 = 60.1, df = 1, p-value <0.001). For example if the bTB-JD interval at the

398   first positive test was >90 days, 30.4% of cows (146) which had an avian skin reaction >4mm

399   recorded only positive tests and 42.1% (202) recorded only negative tests, whereas if the

400   avian was reaction ≤4mm 20.8% of cows (1,444) recorded only positive tests and 56.6%

401   (3,924) recorded only negative tests (X^2 = 41.1, df = 2, p-value <0.001). This approach was

402   repeated after excluding cows which recorded their first positive test in the final year of the

403   study window (2017), displayed in Supplementary Table 1. It was also applied to the test

404   results subsequent to "red cow" (2 in 4) classification, displayed in Supplementary Table 2.

405   The general trends were consistent and statistically significant in both cases (X^2 = 249.7 and

406   60.9 respectively, p-value <0.001 in both cases), but the number of cows eligible for analysis

407   was smaller in these sub-sets.

408

409   *3.3. Sensitivity and specificity estimation of a single JD-mELISA*

410   From the full dataset, 3,904 cows were defined as cases and 42,704 cows were defined as

411   non-cases (Figure 1). The characteristics of cases and non-cases, including number of tests,

412   number of positive tests, and age at final test are displayed in Table 5. The specificity of a

413   single test was estimated to be 0.996 (95% CI: 0.995 - 0.996). The age-dependent sensitivity

414   was estimated from the non-linear logistic regression model (Equation 2) with parameters: a

415   =1.551 (SE: 0.257), b = 4.870 (SE: 0.598) and c = 0.329 (SE: 0.080). It was therefore

416   estimated that the test sensitivity was 0.28 (95% CI: 0.22 - 0.33), 0.56 (95% CI: 0.54 - 0.58)

18

417   and 0.71 (95% CI: 0.69 - 0.73) at two, four and six years old, respectively. Figure 2 displays

418   the estimated test sensitivity as a function of age at testing (Equation 2) plotted alongside the

419   same equation using parameters estimated by Nielsen et al. (2013) and Meyer et al. (2018).

420

421   *3.4. Conditional probability using serial Johne's disease test results*

422   The relationship between prior and posterior probability of infection with different test result

423   patterns and different test specificities were modelled, accounting for changing test sensitivity

424   with age as described in Section 3.3. The PPI in "red cows" classified after the "+ +" and "+ -

425   - +" test result patterns are displayed with different test specificities in Figure 3, if the cow was

426   four-years old at the first of those tests. If the test specificity was 0.996, following two

427   consecutive positive tests ("+ +", Figure 3a), the prior probability needed to be at 0.1% or more

428   to produce a PPI of at least 95%. If the specificity was reduced to 0.986 then the prior

429   probability needed to be at least 1.1% to produce the same result, and 12.6% if the specificity

430   was 0.95. The effects of reducing the test specificity had more substantial effects with the "+ -

431   - +" test pattern (Figure 3b), for example if the test specificity was 0.95, a prior probability of

432   at least 42.9% was required to produce a PPI of at least 95%.

433   The effect of negative test results on the PPI increased with the age of the animal due to the

434   age-dependent sensitivity and are displayed for two and six year old animals in Figure 4 using

435   the specificity estimate from this dataset (0.996). The PPI following consecutive negative tests

436   can be read from this graph for any given prior probability of infection; for example ten

437   consecutive negative quarterly tests are required to produce a PPI of less than 5% from a

438   starting prior probability of 95% in a two year old animal, but only five tests in a six year old

439   animal.

440

441

442

443

Discussion

4.1. Association between bTB test and JD-mELISA

The odds of a positive JD-mELISA were highest when the bTB-JD interval was ≤30d days, but also elevated when the bTB-JD interval was 31 - 60 days compared to >90 days. There was not an increase in the odds of a positive JD-mELISA if the bTB-JD interval was 61 - 90 days compared to >90 days. A recent Spanish study demonstrated an increased odds of a positive JD test if the interval between bTB test (using bovine PPD only) and JD test was less than 90 days compared to greater than 90 days, although only 3.6% of the JD tests in this study were <90 days after a bTB test so shorter bTB-JD intervals were not explored (Picasso-Risso et al., 2019). Unfortunately, due to the current situation in the UK, it can be challenging to avoid a short bTB-JD interval as many herds are often required to have frequent SICCT tests parallel to regular JD testing. Therefore, despite contrary advice 14.7% of JD-mELISAs in this dataset were within 60 days of a SICCT test. As the greatest effect on JD-mELISA was present when bTB-JD interval was 30 days or less, every effort should be made to avoid testing in this window if it is not possible to wait the recommended interval of 60 days. A limitation of this study was the reliance on farm records which are not always complete, and this resulted in cows being excluded from the regression model. It is not known, but neither is it likely, if data were missing completely at random because well managed farms often have good records and vice versa.

In addition to bTB-JD interval and avian skin reaction, the other covariables in the multivariable model all had a statistically significant association with the odds of a positive JD-mELISA (Table 2). Milk yield has been demonstrated to be negatively correlated with antibody concentration in milk (Eisenberg et al., 2015), which corroborates the negative association observed between yield and odds of a positive JD-mELISA. There was a positive association between age and odds of a positive JD-mELISA which has been previously reported (Beaver et al., 2017; Eisenberg et al., 2015), and one which is consistent with current understanding about the immune response to MAP infections (Koets et al., 2015). The relationship between

21

471 odds of a positive JD-mELISA and stage of lactation indicated that the odds were greatest in

472 the first 30 days after calving, compared to later in lactation; previous studies observed that

473 MAP antibodies were highest at either end of the lactation (Eisenberg et al., 2015; Jakobsen

474 et al., 2000; Nielsen et al., 2002). The only breed that were significantly different to Holstein-

475 Friesians were Jerseys, which a previous study also indicated have an increased probability

476 of testing positive for JD (Jakobsen et al., 2000). Scottish herds had lower odds of a positive

477 JD-mELISA than other regions, although the differences between Scotland, East England and

478 South-East England were not statistically significant. There are no recent studies which report

479 the JD prevalence in all regions of the UK, but these results broadly reflect the regional density

480 of dairy cattle (Anon, n.d.), and may suggest that disease risk is associated with population

481 density. There was also a statistically significant difference between the individual milk

482 recording organisations and odds of a positive JD-mELISA, this is likely to be confounded by

483 farms electing to use specific organisations for the other services offered such as pedigree

484 classification and registration, or infectious disease screening.

485

486 *4.2 Analysis of JD-mELISA results subsequent to the first positive result and "red cow"*

487 *classification*

488 In the absence of auxiliary diagnostic testing, it is only possible to speculate whether the

489 association between SICCT test and JD-mELISA is driven by a change in the specificity or

490 sensitivity of the JD-mELISAs. Casal et al. (2014) reported improved performance of an *M.*

491 *bovis* antibody ELISA 15 days after intradermal inoculation with bovine PPD, suggesting that

492 an anamnestic response could be triggered by intradermal PPD inoculation. Roupie et al.

493 (2018) demonstrated an increase in MAP antibodies in animals which had prior inoculation

494 with intradermal bovine PPD and were previously seronegative to MAP, although exposure to

495 MAP was assumed. The authors suggest this could be due to a cross-reactive immune

496 response as control animals, which had not been experimentally infected with *M. bovis,* had

497 minimal increases in MAP antibodies following the intradermal bovine PPD inoculation. A

Brazilian study demonstrated an increase in MAP antibodies (using an in-house ELISA) following a SICCT test in just 2 out of 17 cows in a herd that was considered free from bTB and MAP. In this case the S/P ratio for MAP antibodies did not exceed the positive threshold until 60 days or more after the SICCT test (Varges et al., 2009). A study in a 139-cow Irish dairy herd, which were initially seronegative for MAP antibodies, demonstrated 30% were seropositive ten days after a SICCT test, and this dropped to 12% after 42 days. This herd was bTB free but JD had been previously diagnosed and most cows were expected to be exposed to MAP. Two years after this study, 3 of the 70 cows which were still present in the herd tested positive for MAP shedding using faecal PCR, only one of which had been ELISA positive in the original study (Kennedy et al., 2017). Authors of both these studies concluded that their results most likely suggested a decrease in JD-mELISA specificity following a SICCT test.

An attempt was made to quantify the possible effect of bTB-JD interval on JD-mELISA specificity by excluding JD-mELISA results within 30 days of a SICCT test, and then estimating the test specificity again from a sub-set of non-cases, as previously described (Section 2.3). This resulted in 9,784 fewer cows meeting the non-case definition but there was little change in the proportion of positive test results from non-cases and hence only a negligible change in the estimated specificity (0.997, 95% CI: 0.996 – 0.997). It is likely that this limited change reflects the method of specificity calculation rather than the true relationship between SICCT test and JD-mELISA specificity, discussed in more detail in Section 4.3.

The trends in JD-mELISA results following the first positive test could be considered to be crudely representative of the true infection status, in other words a cow which had only positive test results could be considered truly infected and vice versa. The trends evident in JD-mELISA results after the first positive test suggested that, in some cows, the positive result

524   may have been a false positive because they subsequently had exclusively negative results.

525   This occurred more frequently if the bTB-JD interval was ≤30 days when the cow first tested

526   positive and if the avian skin reaction recorded was ≤4mm (Table 4). Although there are

527   limitations to inferring true infection status of cows from serial JD-mELISAs, the consistent

528   trends lend support to the hypothesis that the SICCT test may decrease the specificity of the

529   JD-mELISA, particularly if a large skin reaction at the avian site was not recorded. As cows

530   which tested positive for the first time in the final year of the study period (2017) would have

531   had fewer subsequent tests than other cows these cows were excluded in case they biased

532   the results, but the same trends were present (Supplementary Table 1).

533

534   The size of avian reaction recorded at the SICCT influenced the odds of a positive JD-mELISA.

535   However, there was an interaction with the bTB-JD interval suggesting this relationship is

536   dependent on how recently it was recorded, in other words the influence of avian reaction on

537   JD-mELISA decreased as the interval between the two tests increased. Furthermore, the

538   highest proportion of cows which recorded only positive tests after their first ever positive JD-

539   mELISA, and therefore could be considered to be truly infected, had a bTB-JD interval >60

540   days and an avian skin reaction of >4mm. One possible explanation of this trend is that the

541   avian skin reaction is a response to MAP infection (Kalis et al., 2003), and as such these cows

542   are more likely to have a positive JD-mELISAs in the future, although not necessarily until

543   disease progression has resulted in a humoral immune response (Stabel, 2000). The

544   alternative explanation is that the skin reaction indicates exposure to *M. avium* subspecies,

545   other than MAP, and these bacteria have stimulated antibodies which cross-react with the JD-

546   mELISA. Therefore, although an avian skin reaction may be an early indication of JD infection

547   status, these cows should not be considered to be infected with MAP unless this is confirmed

548   with additional, more specific, diagnostic tests. Further research is needed to clarify the

549   relationship between avian skin reaction during a SICCT test and JD infection.

550

24

In this dataset, 30.7% (2,241) of all "red cows" (2 in 4) recorded no further tests (Table 3). Of the 5,067 "red cows" which were tested at least once more, 46.3% (2,349) recorded only positive tests and 22.5% (1,138) had only negative tests. Of "red cows" which could be matched to a SICCT test, 160 cows had a bTB-JD interval ≤30 days for their first positive test (Supplementary Table 2), and of these 38.1% (61) had only positive tests, and 34.4% (55) had only negative tests. Despite the small number of cows in this specific sub-set, these skewed proportions compared to all "red cows" are again supportive of the hypothesis that short bTB-JD intervals result in some cows having falsely positive JD-mELISAs, and therefore the validity of some "red cow" classifications could be questioned.

4.3 Sensitivity and specificity estimation of a single JD-mELISA

The estimated specificity in this study was 0.996 and similar to the specificities of 0.995 and 0.987 which were reported by Meyer et al. (2018) and Nielsen et al. (2013) respectively; the plots of the age-dependent sensitivities (Figure 2) also suggest similar results. Both studies used serial JD-mELISA analysis and the same ELISA kit, but it is worth noting that different case and non-case definitions were used in Nielsen et al. (2013): cases were defined as cows with at least two tests if the final test was positive and non-cases were defined as cows with at least five tests if the final four tests were all negative. Additionally, a different positive threshold (S/P ratio ≥15%) was used and this may explain why 79% of cows had only positive tests following their first positive test compared to 22% in this study (Table 4).

There are fundamental limitations to defining cases and non-cases by serial results from the same test because they are predicated on the assumption that the final test results represent the true disease status of an animal. Firstly, as a defined study window was applied, the final test results may only represent the last test results within the study period, not the lifetime of the animal. Secondly, the reaction of a farmer to a specific test result dictates the interpretation

of the result. For example, there were 134 cows which met the non-case definition which had previously had two consecutive positive tests (Table 5). If these cows had been culled following the second positive test they would have been classified as cases, but because they were not culled, and subsequently tested repeatedly negative, they were considered to be non-cases. Finally, the more robustly case and non-case criteria correctly identify truly infected and non-infected animals, the more the specificity and sensitivity estimates will be biased upwards; this is implicit in defining cases and non-cases with a repeated single test type. For example, only cows which have at least eight consecutive tests, and no subsequent positive tests, can meet the non-case definition. Consequently, although there can be a reasonably high confidence that these animals are truly not infected, the specificity will be inflated as very few positive tests remained in this sub-set which could have reduced the specificity estimate.

There is a balance between having case and non-case definitions, which accurately classify animals, and more stringent classification criteria which are more susceptible to bias. Meyer et al. (2018) reported the sensitivity and specificity estimates increased if more tests results were used to classify cases and non-cases, and vice versa. However, irrespective of the specific criteria, the influence of on-farm culling decisions, and the intrinsic bias of excluding positive test results from the non-case sub-set, cannot be incorporated in these definitions and they will always be susceptible to bias. It is, therefore, important to acknowledge that as PPI is a function of the likelihood ratio, it is dependent on the specificity and sensitivity values and therefore higher test characteristics would increase the PPI and vice versa.

4.4 Conditional probability using serial Johne's disease test results

In 2019, of the 2,923 dairy herds which reported their elected management strategy from the UK National Johne's Management Plan, 77.7% were committed to segregating or culling high risk cows (Orpin et al., 2020b). This requires confidence that "red cows" are truly infected and

603 represent a risk to other stock, this is what is encompassed by the PPI calculated in this study.

604 Although there is no universal figure for what PPI threshold would warrant culling an animal,

605 and culling decisions are rarely based on a single factor, a 95% probability of infection has

606 been advocated (Meyer et al., 2018). The PPI at the point of "red cow" classification, and the

607 effect of different test result patterns and test specificity, were modelled (Figure 3). The

608 specificity estimate in this study was very high (0.996) and two consecutive positive tests (+

609 +) in a four year old animal would result in a PPI over 95% in all cases unless the prior

610 probability was virtually zero. The PPI following the "+ - - +" test pattern, with the same

611 specificity, is greater than 95% unless the prior probability is less than approximately 0.5%.

612 The reduced specificities which were modelled showed greater effects with the "+ - - +" pattern

613 than with the "+ +" pattern. To generate a PPI of at least 95% with "+ - - +" and a test specificity

614 of 0.986 (i.e. a 1% reduction), the prior probability needs to be greater than approximately 6%.

615 If the specificity reduced to 0.95, the PPI is less than 95% if the prior probability is less than

616 43% following "+ - - +". These results demonstrate that a high degree of confidence in the

617 infection status of "red cows" is reliant on an exceptionally high specificity, and the (+ - - +)

618 "red cow" test pattern is very susceptible to small changes in this specificity. Therefore, if

619 anything reduces the specificity by a small proportion, such as a recent SICCT test, the

620 posterior probability of infection can be substantially reduced, and care should be taken if

621 culling decisions are based solely on a "red cow" classification.

622

623 Once cows are classified as "red" they are treated as high risk regardless of future results and

624 this is based partly on a longitudinal study in Denmark that showed these cows have a higher

625 probability of shedding in the future (Nielsen, 2008). Consecutive negative test results will

626 reduce the PPI even if it is initially high, such as in many "red cows". The number of tests

627 required to reduce the PPI to a level with a high confidence of the cow being uninfected, such

628 as a 5% PPI (95% probability of being uninfected), depends primarily on the age-dependent

629 sensitivity of the test. The modelling of PPI after quarterly, consecutively negative tests

demonstrated how significantly the age of the animal influences the value of a negative test (Figure 4). Most cows that become "red cows" are not retained in herds because of the potential risk they represent and therefore it is not known what further tests in these cows would reveal.

The conditional probability approach to serial test interpretation can be applied to specific cows, using the posterior probability from the previous test as the prior probability for the subsequent test. This approach requires an initial prior probability which can be taken as the prevalence of JD in the herd; this can be estimated by calculating the proportion of cows which have at least one positive JD-mELISA each year and adjusting this result for the average age-dependent sensitivity in the herd and the test specificity (Rogan and Gladen, 1978). In this dataset the median age of "red cow" classification was 4.8 years and the upper quartile was 6.2 years (Table 3). Therefore, a typical "red cow" may calve for the first time at 24 months, have three tests each year which are all negative, and then have the second of two consecutive positive tests at 4.8 years old. If the starting prior probability is assumed to be 16.8%, which is the average adjusted prevalence of JD in British dairy herds reported by Meyer et al. (2018), then after seven negative tests and two consecutive positive tests the PPI would be 98.5%. However, given that 25% of cows in this dataset were at least 6.2 years old when they met the "red cow" classification, they could have plausibly had as many as twelve negative tests before having two consecutive positive tests. In these cows the PPI at the point of "red cow" classification would only be 30.0%. Obviously, there are many herds with a lower prevalence of JD than 16.8% and therefore the PPI in "red cows" would be lower than the previous examples. This dispersion in PPI can be inferred by the high standard deviations reported by Meyer et al (2018) and highlights the limitations of defining "red cows" based on serial test results without consideration of the prior probability of infection, and this is especially problematic in low prevalence herds.

657   *4.5 Conclusions*

658   The odds of a positive JD-mELISA are significantly affected by the bTB-JD interval. The trends

659   in results following a positive JD-mELISA suggest that some of these are likely to be false

660   positive results, and more false positive results are more likely if the bTB-JD interval is less

661   than 30 days, and to a lesser extent if the bTB-JD interval is 31 – 60 days. The PPI in "red

662   cows" is largely dependent on the test result pattern, test specificity, and prior probability of

663   infection. Therefore, if the test specificity is reduced, the PPI of "red cows" may not be high

664   enough to confidently advise culling the animal, especially if cows have not had consecutive

665   positive JD-mELISAs. Furthermore, if cows from low prevalence herds do meet the "red cow"

666   criteria they may still have a low PPI, particularly if they have had multiple negative JD-

667   mELISAs previously. Many farms cull "red cows" and therefore it is important that culling

668   decisions include consideration of individual test results and herd prevalence; it may also be

669   beneficial to review the "red cow" classification criteria based on serial JD-mELISAs in low

670   prevalence herds.

671   Finally, the odds of a positive JD-mELISA were increased if a large avian skin reaction was

672   recorded at the previous SICCT test, and more cows had exclusively positive JD-mELISAs

673   subsequently regardless of the bTB-JD test interval. Therefore, avian reactions recorded

674   during a SICCT test could be used highlight cows which may be infected with MAP, or at least

675   cows which are likely to test positive to future JD-mELISAs.

676

677   **Funding sources**

680

681

29

References

691   Anon., 2018. HerdWise rule change [WWW Document]. URL

693   https://www.nmr.co.uk/uploads/files/files/HW Classification change farmer(1).pdf

694   (accessed 12.6.19).

695   Anon, n.d. eFoodChainMap [WWW Document]. URL www.emap.org.uk/HeatMap.aspx

696   (accessed 4.8.20).

697   Arsenault, R.J., Maattanen, P., Daigle, J., Potter, A., Griebel, P., Napper, S., 2014. From

698   mouth to macrophage: Mechanisms of innate immune subversion by Mycobacterium

699   avium subsp. Paratuberculosis. Vet. Res. https://doi.org/10.1186/1297-9716-45-54

700   Bassey, E.O.E., Collins, M.T., 1997. Study of T-lymphocyte subsets of healthy and

701   Mycobacterium avium subsp. paratuberculosis-Infected cattle. Infect. Immun. 65, 4869–

702   4872.

703   Beaver, A., Sweeney, R.W., Hovingh, E., Wolfgang, D.R., Gröhn, Y.T., Schukken, Y.H.,

704   2017. Longitudinal relationship between fecal culture, fecal quantitative PCR, and milk

705   ELISA in Mycobacterium avium ssp. paratuberculosis-infected cows from low-

706   prevalence dairy herds. J. Dairy Sci. 100, 7507–7521. https://doi.org/10.3168/jds.2017-

707   12928

708   Brown, L.D., Cai, T.T., DasGupta, A., 2001. Interval Estimation for a Binomial Proportion.

709   Stat. Sci. 16, 101–133. https://doi.org/10.1214/SS/1009213286

710   Casal, C., Díez-Guerrier, A., Álvarez, J., Rodriguez-Campos, S., Mateos, A., Linscott, R.,

711   Martel, E., Lawrence, J.C., Whelan, C., Clarke, J., O'Brien, A., Domínguez, L., Aranaz,

712   A., 2014. Strategic use of serology for the diagnosis of bovine tuberculosis after

713   intradermal skin testing. Vet. Microbiol. 170, 342–351.

714   https://doi.org/10.1016/j.vetmic.2014.02.036

715   Collins, M.T., 1996. Diagnosis of Paratuberculosis. Vet. Clin. North Am. - Food Anim. Pract.

716   27, 357–371. https://doi.org/10.1016/j.cvfa.2011.07.013

717   de la Rua-Domenech, R., Goodchild, A.T., Vordermeier, H.M., Hewinson, R.G.,

718   Christiansen, K.H., Clifton-Hadley, R.S., 2006. Ante mortem diagnosis of tuberculosis in

719   cattle: A review of the tuberculin tests, γ-interferon assay and other ancillary diagnostic

720   techniques. Res. Vet. Sci. 81, 190–210. https://doi.org/10.1016/j.rvsc.2005.11.005

721   Dohoo, I., Martin, W., Stryhn, H., 2012. Methods in epidemiologic research., Rev. ed. VER

722   Inc., Charlottetown, Prince edward Island.

723   Eisenberg, S.W.F., Veldman, E., Rutten, V.P.M.G., Koets, A.P., 2015. A longitudinal study of

724   factors influencing the result of a Mycobacterium avium ssp. paratuberculosis antibody

725   ELISA in milk of dairy cows. J. Dairy Sci. 98, 2345–2355.

726   https://doi.org/10.3168/jds.2014-8380

727   Fecteau, M.E., 2018. Paratuberculosis in Cattle. Vet. Clin. North Am. - Food Anim. Pract.

728   https://doi.org/10.1016/j.cvfa.2017.10.011

729   Geraghty, T., Graham, D.A., Mullowney, P., More, S.J., 2014. A review of bovine Johne's

730   disease control activities in 6 endemically infected countries. Prev. Vet. Med.

731   https://doi.org/10.1016/j.prevetmed.2014.06.003

732   Gilot, P., Cocito, C., 1993. Comparative analysis of three sensitins used in cutaneous testing

733   for tuberculosis and paratuberculosis in cattle. FEMS Microbiol. Lett. 110, 307–311.

734   https://doi.org/10.1111/j.1574-6968.1993.tb06340.x

735   Harrell, F., 2020. Hmisc: Harrell Miscellaneous.

736   Jakobsen, M.B., Alban, L., Nielsen, S.S., 2000. A cross-sectional study of paratuberculosis

737   in 1155 Danish dairy cows. Prev. Vet. Med. 46, 15–27. https://doi.org/10.1016/S0167-

738   5877(00)00138-0

739   Jenvey, C.J., Hostetter, J.M., Shircliff, A.L., Stabel, J.R., 2018. Relationship between the

740   pathology of bovine intestinal tissue and current diagnostic tests for Johne's disease.

741   Vet. Immunol. Immunopathol. 202, 93–101.

742   https://doi.org/10.1016/j.vetimm.2018.06.012

743   Jungersen, G., Huda, A., Hansen, J.J., Lind, P., 2002. Interpretation of the gamma interferon

744   test for diagnosis of subclinical paratuberculosis in cattle. Clin. Diagn. Lab. Immunol. 9,

745   453–460. https://doi.org/10.1128/CDLI.9.2.453-460.2002

746   Kalis, C.H.J., Collins, M.T., Hesselink, J.W., Barkema, H.W., 2003. Specificity of two tests for

747   the early diagnosis of bovine paratuberculosis based on cell-mediated immunity: the

748   Johnin skin test and the gamma interferon assay. Vet. Microbiol. 97, 73–86.

749   https://doi.org/10.1016/S0378-1135(03)00242-6

750   Kennedy, A.E., Byrne, N., O'Mahony, J., Sayers, R.G., 2017. Investigations and implications

751   of associations between mycobacterial purified protein derivative hypersensitivity and

752   MAP-antibody ELISA in Irish dairy cows. Res. Vet. Sci. 115, 13–16.

753   https://doi.org/10.1016/j.rvsc.2017.01.018

754   Kennedy, A.E., da Silva, A.T., Byrne, N., Govender, R., MacSharry, J., O'Mahony, J.,

755   Sayers, R.G., 2014. The single intradermal cervical comparative test interferes with

756   Johne's disease ELISA diagnostics. Front. Immunol. 5.

757   https://doi.org/10.3389/fimmu.2014.00564

758   Koets, A.P., Eda, S., Sreevatsan, S., 2015. The within host dynamics of Mycobacterium

759   avium ssp. paratuberculosis infection in cattle: Where time and place matter Modeling

760   Johne's disease: From the inside out Dr Ad Koets and Prof Yrjo Grohn. Vet. Res.

761   https://doi.org/10.1186/s13567-015-0185-0

762   Liverani, E., Scaioli, E., Cardamone, C., Dal Monte, P., Belluzzi, A., 2014. Mycobacterium

763   avium subspecies paratuberculosis in the etiology of Crohn's disease, cause or

764   epiphenomenon? World J. Gastroenterol. https://doi.org/10.3748/wjg.v20.i36.13060

765   May, H., Orpin, P., Pearse, H., Jones, P., Clough, H., 2016. The influence of tuberculosis

766   testing in dairy cattle on milk ELISA tests for Johne's disease [WWW Document]. URL

767   https://www.researchgate.net/publication/237445560_The_influence_of_tuberculosis_te

768   sting_in_dairy_cattle_on_milk_ELISA_tests_for_Johne's_disease (accessed 1.12.20).

769 McAloon, C.G., Roche, S., Ritter, C., Barkema, H.W., Whyte, P., More, S.J., O'Grady, L.,

770   Green, M.J., Doherty, M.L., 2019. A review of paratuberculosis in dairy herds — Part 1:

771   Epidemiology. Vet. J. https://doi.org/10.1016/j.tvjl.2019.01.010

772 Meyer, A., Bond, K., Van Winden, S., Green, M., Guitian, J., 2018. A probabilistic approach

773   to the interpretation of milk antibody results for diagnosis of Johne's disease in dairy

774   cattle. Prev. Vet. Med. 150, 30–37. https://doi.org/10.1016/j.prevetmed.2017.11.016

775 Mortier, R.A.R., Barkema, H.W., Bystrom, J.M., Illanes, O., Orsel, K., Wolf, R., Atkins, G., De

776   Buck, J., 2013. Evaluation of age-dependent susceptibility in calves infected with two

777   doses of Mycobacterium avium subspecies paratuberculosis using pathology and tissue

778   culture. Vet. Res. 44. https://doi.org/10.1186/1297-9716-44-94

779 Nielsen, S.S., 2008. Transitions in diagnostic tests used for detection of Mycobacterium

780   avium subsp. paratuberculosis infections in cattle. Vet. Microbiol. 132, 274–282.

781   https://doi.org/10.1016/j.vetmic.2008.05.018

782 Nielsen, S.S., Enevoldsen, C., Gröhn, Y.T., 2002. The Mycobacterium avium subsp.

783   paratuberculosis ELISA response by parity and stage of lactation. Prev. Vet. Med. 54,

784   1–10. https://doi.org/10.1016/S0167-5877(02)00008-9

785 Nielsen, S.S., Toft, N., 2008. Ante mortem diagnosis of paratuberculosis: A review of

786   accuracies of ELISA, interferon-γ assay and faecal culture techniques. Vet. Microbiol.

787   https://doi.org/10.1016/j.vetmic.2007.12.011

788 Nielsen, S.S., Toft, N., Okura, H., 2013. Dynamics of Specific Anti-Mycobacterium avium

789   Subsp. paratuberculosis Antibody Response through Age. PLoS One 8, e63009.

790   https://doi.org/10.1371/journal.pone.0063009

791 Nuñez-Garcia, J., Downs, S.H., Parry, J.E., Abernethy, D.A., Broughan, J.M., Cameron,

792   A.R., Cook, A.J., De La Rua-Domenech, R., Goodchild, A. V., Gunn, J., More, S.J.,

793   Rhodes, S., Rolfe, S., Sharp, M., Upton, P.A., Vordermeier, H.M., Watson, E., Welsh,

794   M., Whelan, A.O., Woolliams, J.A., Clifton-Hadley, R.S., Greiner, M., Nuñez, J., Nuñez-

795   Garcia, N., Downs, S.H., Parry, J.E., Abernethy, D.A., Broughan, J.M., Cameron, A.R.,

796   Cook, A.J., De La Rua-Domenech, R., Goodchild, A. V., Gunn, J., More, S.J., Rhodes,

797   S., Rolfe, S., Sharp, M., Upton, P.A., Martin Vordermeier, H., Watson, E., Welsh, M.,

798   Whelan, A.O., Woolliams, J.A., Clifton-Hadley, R.S., Greiner, M., Nuñez-Garcia, J.,

799   Downs, S.H., Parry, J.E., Abernethy, D.A., Broughan, J.M., Cameron, A.R., Cook, A.J.,

800   De La Rua-Domenech, R., Goodchild, A. V., Gunn, J., More, S.J., Rhodes, S., Rolfe,

801   S., Sharp, M., Upton, P.A., Vordermeier, H.M., Watson, E., Welsh, M., Whelan, A.O.,

802   Woolliams, J.A., Clifton-Hadley, R.S., Greiner, M., 2018. Meta-analyses of the

803   sensitivity and specificity of ante-mortem and post-mortem diagnostic tests for bovine

804   tuberculosis in the UK and Ireland. Prev. Vet. Med. 153, 94–107.

805   https://doi.org/10.1016/j.prevetmed.2017.02.017

806 OIE, 2019. Bovine Tuberculosis, in: Manual of Diagnostic Tests and Vaccines for Terrestrial

807   Animals.

808 Orpin, P., Sibley, D., Bond, K., 2020a. Johne's disease in dairy herds 2. Effective control

809   using the National Johne's Management Plan. In Pract. 42, 159–168.

810   https://doi.org/10.1136/inp.m874

811 Orpin, P., Sibley, D., Bond, K., 2020b. Johne's disease in dairy herds 1. Understanding the

812   disease. In Pract. 42, 39–46. https://doi.org/10.1136/inp.l6924

813 Picasso-Risso, C., Grau, A., Bakker, D., Nacar, J., Mínguez, O., Perez, A., Alvarez, J., 2019.

814   Association between results of diagnostic tests for bovine tuberculosis and Johne's

815   disease in cattle. Vet. Rec. vetrec-2018-105336. https://doi.org/10.1136/vr.105336

816 R Development Core Team 3.6.1, 2019. A Language and Environment for Statistical

817   Computing. R Found. Stat. Comput.

818   Ravva, S. V., Harden, L.A., Sarreal, C.Z., 2017. Characterization and differentiation of

819   mycobacterium avium subsp. Paratuberculosis from other mycobacteria using matrix

820   assisted laser desorption/ionization time-of-flight mass spectrometry. Front. Cell. Infect.

821   Microbiol. 7. https://doi.org/10.3389/fcimb.2017.00297

822   Rogan, W.J., Gladen, B., 1978. Estimating prevalence from the results of a screening test.

823   Am. J. Epidemiol. 107, 71–76. https://doi.org/10.1093/oxfordjournals.aje.a112510

824   Roupie, V., Alonso-Velasco, E., Van Der Heyden, S., Holbert, S., Duytschaever, L., Berthon,

825   P., Van Dosselaer, I., Van Campe, W., Mostin, L., Biet, F., Roels, S., Huygen, K.,

826   Fretin, D., 2018. Evaluation of mycobacteria-specific gamma interferon and antibody

827   responses before and after a single intradermal skin test in cattle naturally exposed to

828   M. avium subsp. paratuberculosis and experimentally infected with M. bovis. Vet.

829   Immunol. Immunopathol. 196, 35–47. https://doi.org/10.1016/j.vetimm.2017.12.007

830   SAS Institute Inc., 2015. SAS/STAT 14.1 User's Guide. Cary, NC.

831   Smith, R.L., Gröhn, Y.T., Pradhan, A.K., Whitlock, R.H., Van Kessel, J.S., Smith, J.M.,

832   Wolfgang, D.R., Schukken, Y.H., 2016. The effects of progressing and nonprogressing

833   Mycobacterium avium ssp. paratuberculosis infection on milk production in dairy cows.

834   J. Dairy Sci. 99, 1383–1390. https://doi.org/10.3168/jds.2015-9822

835   Stabel, J.R., 2000. Transitions in immune responses to Mycobacterium paratuberculosis.

836   Vet. Microbiol. 77, 465–473. https://doi.org/10.1016/S0378-1135(00)00331-X

837   StataCorp LP, 2016. Stata/MP 14.2.

838   Varges, R., Dray Marassi, C., Oelemann, W., Lilenbaum, W., 2009. Interference of

839   intradermal tuberculin tests on the serodiagnosis of paratuberculosis in cattle. Res. Vet.

840   Sci. 86, 371–372. https://doi.org/10.1016/j.rvsc.2008.08.006

841   Velasova, M., Damaso, A., Prakashbabu, B.C., Gibbons, J., Wheelhouse, N., Longbottom,

842   D., Van Winden, S., Green, M., Guitian, J., 2017. Herd-level prevalence of selected

843   endemic infectious diseases of dairy cows in Great Britain. J. Dairy Sci. 100, 9215–

844   9233. https://doi.org/10.3168/jds.2016-11863

845 Whittington, R., Donat, K., Weber, M.F., Kelton, D., Nielsen, S.S., Eisenberg, S., Arrigoni, N.,

846   Juste, R., Sáez, J.L., Dhand, N., Santi, A., Michel, A., Barkema, H., Kralik, P.,

847   Kostoulas, P., Citer, L., Griffin, F., Barwell, R., Moreira, M.A.S., Slana, I., Koehler, H.,

848   Singh, S.V., Yoo, H.S., Chávez-Gris, G., Goodridge, A., Ocepek, M., Garrido, J.,

849   Stevenson, K., Collins, M., Alonso, B., Cirone, K., Paolicchi, F., Gavey, L., Rahman,

850   M.T., De Marchin, E., Van Praet, W., Bauman, C., Fecteau, G., McKenna, S., Salgado,

851   M., Fernández-Silva, J., Dziedzinska, R., Echeverría, G., Seppänen, J., Thibault, V.,

852   Fridriksdottir, V., Derakhshandeh, A., Haghkhah, M., Ruocco, L., Kawaji, S., Momotani,

853   E., Heuer, C., Norton, S., Cadmus, S., Agdestein, A., Kampen, A., Szteyn, J., Frössling,

854   J., Schwan, E., Caldow, G., Strain, S., Carter, M., Wells, S., Munyeme, M., Wolf, R.,

855   Gurung, R., Verdugo, C., Fourichon, C., Yamamoto, T., Thapaliya, S., Di Labio, E.,

856   Ekgatat, M., Gil, A., Alesandre, A.N., Piaggio, J., Suanes, A., De Waard, J.H., 2019.

857   Control of paratuberculosis: Who, why and how. A review of 48 countries. BMC Vet.

858   Res. https://doi.org/10.1186/s12917-019-1943-4

859 Windsor, P.A., Whittington, R.J., 2010. Evidence for age susceptibility of cattle to Johne's

860   disease. Vet. J. 184, 37–44. https://doi.org/10.1016/j.tvjl.2009.01.007

861 Woodbine, K.A., Schukken, Y.H., Green, L.E., Ramirez-Villaescusa, A., Mason, S., Moore,

862   S.J., Bilbao, C., Swann, N., Medley, G.F., 2009. Seroprevalence and epidemiological

863   characteristics of Mycobacterium avium subsp. paratuberculosis on 114 cattle farms in

864   south west England. Prev. Vet. Med. 89, 102–109.

865   https://doi.org/10.1016/j.prevetmed.2009.02.005

866

867

868   Table 1. Univariable logistic regression of factors associated with a single positive Johne's

869   disease milk antibody ELISA in 225,296 cows on 735 dairy farms in the UK

	Frequency of cows[1]	Frequency of cows with a positive test[1]	Coef.	SE	p-value	OR	Lower 95% CI	Upper 95% CI
SICCT test to JD test interval								
≤30 d	3,456	192	0.62	0.08	<0.01	1.85	1.60	2.15
31 – 60 d	19,640	724	0.19	0.04	<0.01	1.21	1.11	1.31
61 – 90 d	36,814	1,169	0.03	0.04	0.34	1.03	0.97	1.11
>90 d	102,004	3,137	Ref.					
Difference in skin thickness at avian site during SICCT test								
≤2 mm	149,097	4,509	Ref.					
2 – 4 mm	5,584	247	0.39	0.07	<0.01	1.48	1.30	1.69
>4 mm	7,233	466	0.79	0.05	<0.01	2.21	2.00	2.44
Difference in skin thickness at bovine site during SICCT test								
≤2 mm	152,073	4,924	Ref.					
2 – 4 mm	2,755	180	0.70	0.08	<0.01	2.02	1.73	2.35
>4 mm	1,864	118	0.67	0.10	<0.01	1.96	1.62	2.36
Age at recording								
≤3 y	79,900	1,574	Ref.					
3 – 5 y	84,113	2,861	0.55	0.03	<0.01	1.73	1.62	1.84
>5 y	53,872	2,276	1.03	0.03	<0.01	2.80	2.64	2.98
Milk yield								
≤15 kg	16,222	1,251	Ref.					
15 – 30 kg	106,039	3,821	-0.76	0.03	<0.01	0.47	0.44	0.50
30 – 45 kg	80,578	1,979	-1.14	0.04	<0.01	0.32	0.30	0.34
>45 kg	15,046	360	-1.17	0.06	<0.01	0.31	0.28	0.35
Days since calving								
≤30 d	22,046	984						
31 – 90 d	49,864	1,227	-0.62	0.04	<0.01	0.54	0.50	0.59
91 – 150 d	40,946	1,079	-0.55	0.05	<0.01	0.58	0.53	0.63
151 – 210 d	36,599	991	-0.52	0.05	<0.01	0.60	0.55	0.65
211 – 270 d	33,848	1,152	-0.28	0.04	<0.01	0.75	0.69	0.82
271 – 330 d	22,724	954	-0.06	0.05	0.17	0.94	0.86	1.03
>330 d	18,887	1,001	0.18	0.05	<0.01	1.20	1.10	1.31

			Coef.	SE		OR	CI	
Breed								
Holstein / Friesian	168,536	5,105	Ref.					
Brown Swiss	910	45	0.51	0.15	<0.01	1.67	1.23	2.25
Shorthorn	703	36	0.55	0.17	<0.01	1.73	1.23	2.42
Ayrshire	1,597	61	0.24	0.13	0.07	1.27	0.98	1.65
Jersey	1,424	79	0.63	0.12	<0.00	1.88	1.50	2.36
Montbeliarde	1,525	59	0.25	0.13	0.06	1.29	0.99	1.67
Norwegian Red	887	30	0.11	0.19	0.54	1.12	0.78	1.62
Swedish Red	1,341	70	0.57	0.12	<0.01	1.76	1.38	2.25
Other	1,248	65	0.56	0.13	<0.01	1.76	1.37	2.26
Milk recording organisation								
1	88,135	3,027	Ref.					
2	25,653	1,037	0.17	0.04	<0.01	1.18	1.10	1.27
3	111,508	3,347	-0.14	0.03	<0.01	0.87	0.83	0.92
UK region								
Scotland	19,793	405	Ref.					
East/East Midlands	12,813	395	0.42	0.07	<0.01	1.52	1.32	1.75
North-west	41,278	1,182	0.34	0.06	<0.01	1.41	1.26	1.58
South-east	8,049	188	0.14	0.09	0.13	1.14	0.96	1.36
South-west	44,593	1,670	0.62	0.06	<0.01	1.86	1.67	2.08
Wales	13,955	465	0.50	0.07	<0.01	1.65	1.44	1.89
West Midlands	37,690	1,245	0.49	0.06	<0.01	1.64	1.46	1.83
			ICC	**SE**				
Random effect (farm)			0.14	0.01				

870

871 [1] The frequency of cows in total and the frequency with a positive Johne's disease milk

872 antibody ELISA from a single randomly selected test result from each cow.

873 *Coef.: Regression coefficient; SE: Standard error of coefficient; OR: Odds ratio; CI:*

874 *Confidence interval of odds ratio; ICC: Intra-class correlation*

875

876

877 Table 2. Two-level multivariable logistic regression model of factors associated with a single

878 positive Johne's disease milk antibody ELISA result in 158,287 cows on 664 dairy farms in the

879 UK

	Coef.	SE	p-value	OR	Lower 95% CI	Upper 95% CI
SICCT test to JD test interval						
≤30 d	0.73	0.09	<0.01	2.08	1.73	2.50
31 – 60 d	0.14	0.05	0.01	1.15	1.04	1.27
61 – 90 d	-0.01	0.04	0.77	0.99	0.91	1.07
>90 d	Ref.					
Difference in skin thickness at avian site during SICCT test						
≤2 mm	Ref.					
2 – 4 mm	0.27	0.09	<0.01	1.31	1.09	1.57
>4 mm	0.62	0.07	<0.01	1.87	1.62	2.16
SICCT test to JD test interval x Difference in skin thickness at avian site during SICCT test interaction						
≤30 d x ≤2 mm	Ref.					
≤30 d x 2 – 4 mm	0.26	0.35	0.45	1.30	0.66	2.57
≤30 d x >4 mm	0.54	0.26	0.04	1.72	1.04	2.86
31 – 60 d x ≤2 mm	Ref.					
31 – 60 d x 2 – 4 mm	0.24	0.19	0.21	1.27	0.87	1.85
31 – 60 d x >4 mm	0.43	0.15	0.01	1.53	1.14	2.06
61 – 90 d x ≤2 mm	Ref.					
61 – 90 d x 2 – 4 mm	0.10	0.17	0.56	1.10	0.79	1.55
61 – 90 d x >4 mm	0.31	0.13	0.01	1.37	1.07	1.76
>90 d x ≤2 mm	Ref.					
>90 d x 2 – 4 mm	Ref.					
>90 d >4 mm	Ref.					
Age at recording						
≤3 y	Ref.					
3 – 5 y	0.71	0.04	<0.01	2.02	1.86	2.2
>5 y	1.22	0.04	<0.01	3.39	3.11	3.70
Milk yield						
≤15 kg	Ref.					
15 – 30 kg	-0.67	0.05	<0.01	0.51	0.47	0.56

40

30 – 45 kg	-1.15	0.06	<0.01	0.32	0.28	0.35
>45kg	-1.15	0.09	<0.01	0.22	0.19	0.27
Days since calving						
≤30 d	Ref.					
31 – 90 d	-0.59	0.05	<0.01	0.56	0.50	0.62
91 – 150 d	-0.69	0.06	<0.01	0.50	0.45	0.56
151 – 210 d	-0.80	0.06	<0.01	0.45	0.40	0.50
211 – 270 d	-0.67	0.06	<0.01	0.51	0.46	0.57
271 – 330 d	-0.62	0.06	<0.01	0.54	0.48	0.61
>330 d	-0.60	0.06	<0.01	0.55	0.49	0.62
Breed						
Holstein/Friesian	Ref.					
Brown Swiss	-0.10	0.19	0.60	0.90	0.62	1.32
Shorthorn	0.07	0.20	0.75	1.07	0.71	1.59
Ayrshire	-0.22	0.18	0.23	0.80	0.56	1.14
Jersey	0.36	0.14	0.01	1.43	1.08	1.90
Montbeliarde	0.29	0.17	0.08	1.33	0.96	1.84
Norwegian Red	0.18	0.20	0.36	1.20	0.81	1.79
Swedish Red	0.20	0.14	0.17	1.22	0.92	1.62
Other	0.31	0.14	0.03	1.37	1.04	1.80
Milk recording organisation						
1	Ref.					
2	-0.01	0.15	0.96	1.00	0.80	1.26
3	-0.25	0.08	<0.01	0.78	0.67	0.90
UK region						
Scotland	Ref.					
East/East Midlands	0.23	0.15	0.14	1.26	0.93	1.69
North-west	0.32	0.13	0.02	1.38	1.06	1.78
South-east	0.11	0.19	0.59	1.12	0.77	1.61
South-west	0.48	0.13	<0.01	1.61	1.24	2.09
Wales	0.49	0.15	<0.01	1.64	1.22	2.18
West Midlands	0.47	0.13	<0.01	1.59	1.23	2.07
Intercept	-3.29	0.14	<0.01	0.04	0.03	0.05
	ICC	**SE**				
Random effect (farm)	0.11	0.01				

880

881   *Coef.: Regression coefficient; SE: Standard error of coefficient; OR: Odds ratio; CI:*

882   *Confidence interval of odds ratio; ICC: Intra-class correlation.*

883　Table 3. The characteristics and subsequent test results of "red cows" (2 in 4) based on the result pattern which resulted in "red cow"

884　classification. (From 225,296 cows on 735 dairy farms in the UK).

	Test pattern that classified cow as "red "			
	+ +	+ - +	+ - - +	Total (all "red cows")
Frequency of cows (%)	5,846 (80.0%)	867 (11.9%)	595 (8.1%)	7,308 (100%)
Frequency of each cows within each "red cow" test pattern which had consecutive positive tests at any point	5,846 (100%)	348 (40.1%)	188 (31.6%)	6,382 (87.3%)
Median number of tests after cow classified as a "red cow" (IQR)	1 (0 - 4)	3 (1 - 6)	3 (1 - 5)	3 (1 - 6)
Median age (years) when classified as a "red cow" (IQR)	4.8 (3.7 - 6.1)	4.9 (3.8 - 6.1)	5.2 (4.0 - 6.8)	4.8 (3.7 - 6.2)
Number of recorded tests after classified as a "red cow"				
Frequency with no more recorded tests (%)	1,921 (32.9%)	204 (23.5%)	116 (19.5%)	2,241 (30.7%)
Frequency with 1 - 3 more recorded tests (%)	2,240 (38.3%)	298 (34.3%)	247 (41.5%)	2,785 (38.1%)
Frequency with ≥4 more recorded tests (%)	1,685 (28.8%)	365 (42.1%)	232 (39.0%)	2,282 (31.2%)

885

Proportion of subsequent tests that were positive after classified as a "red cow" if ≥1 more test recorded				
0% positive	668 (17.0%)	232 (35.0%)	238 (49.7%)	1,138 (22.5%)
1 - 99% positive	1,149 (29.3%)	266 (40.1%)	165 (34.5%)	1,580 (31.2%)
100% positive	2,108 (53.7%)	165 (24.9%)	76 (15.9%)	2,349 (46.3%)
Total	3,925 (100%)	663 (100%)	479 (100%)	5,067 (100%)
Proportion of subsequent tests that were positive after classified as a "red cow" if ≥4 more test recorded				
0% positive	309 (18.3%)	118 (32.3%)	93 (40.1%)	520 (22.8%)
1 - 99% positive	936 (55.6%)	208 (57.0%)	127 (54.7%)	1,271 (55.7%)
100% positive	440 (26.1%)	39 (10.7%)	12 (5.2%)	491 (21.5%)
Total	1,685 (100%)	365 (100%)	232 (100%)	2,282 (100%)

887   *IQR: Inter-quartile range; "Red cow": cow considered high risk of being infectious with Johne's disease based on two positive test results in the*

888   *previous four tests (2 in 4).*

889 Table 4. The proportion of subsequent tests that were positive following a cow's first ever positive test if cows had at least one more test recorded,

890 N = 13,193. (A total of 22,401 cows had at least one positive test, 16,497 of these could be matched to the most recent bTB test, of those 3,304

891 cows had no further tests after their first positive test.)

bTB - JD test interval at first positive JD test (days)		Number of cows with each proportion of positive tests following their first positive test (%)							
		≤30d		31-60d		61-90d		>90d	
Change in skin thickness at avian site during SICCT test		≤4mm	>4mm	≤4mm	>4mm	≤4mm	>4mm	≤4mm	>4mm
Proportion of subsequent tests that are positive after the first positive test [a]	0%	464 (69.3%)	41 (53.2%)	1,136 (57.1%)	78 (44.3%)	1,356 (51.7%)	108 (43.0%)	3,924 (56.6%)	202 (42.1%)
	1-99%	150 (22.4%)	21 (27.3%)	568 (28.6%)	62 (35.2%)	612 (23.4%)	59 (23.5%)	1,561 (22.6%)	132 (27.5%)
	100%	56 (8.4%)	15 (19.5%)	284 (14.3%)	36 (20.5%)	654 (24.9%)	84 (33.5%)	1,444 (20.8%)	146 (30.4%)
Total		670 (100%)	77 (100%)	1,988 (100%)	176 (100%)	2,622 (100%)	251 (100%)	6,929 (100%)	480 (100%)
Median number of tests after first positive test (IQR) [b]		5 (3 - 8)		5 (2 - 7)		3 (2 - 6)		4 (2 - 7)	
Median age in years at first positive test (IQR) [c]		4.0 (2.9 - 5.4)		4.1 (3.0 – 5.4)		4.2 (3.2 – 5.7)		4.5 (3.3 – 5.9)	

892   [a] Pearson's Chi-squared test to compare proportion of tests that were positive following the first positive test for different bTB-JD test intervals

893   and avian reaction details, $X^2 = 249.7$, df=14, p-value <0.001

894   [b] Kruskal Wallis test to compare the median number of tests recorded after the first positive test for different bTB-JD test intervals, $X^2 = 239.0$,

895   df=3, p-value <0.001

896   [c] Kruskal Wallis test to compare the median age at first positive test for different bTB-JD test intervals, $X^2 = 117.5$, df=3, p-value <0.001

897

898 Table 5. Characteristics of cows which met the "case" or "non-case" definitions. (From total
899 dataset of 225,296 cows on 735 dairy farms in the UK).

900

	Cases	Non-cases
Definition	Minimum of 3 tests, final 2 tests were positive	Minimum of 9 tests, final 8 tests were negative
Frequency	3,904	42,704
Median number of tests (IQR)	7 (4 - 7)	11 (10 - 13)
Median number of positive tests (IQR)	3 (2 - 5)	0 (0 - 0)
Range of positive tests	2 - 16	0 - 5
Median age (years) at final test (IQR)	5.5 (4.3 - 6.9)	6.1 (5.2 - 7.3)
Frequency of 2 in 2 "red cows" [1]	3,904	134
Frequency of 2 in 4 "red cows" [2]	3,904	192

901

IQR: Inter-quartile range

903 [1] Cows which also met the previous definition of "red cows": two consecutive positive tests (2
904 in 2)

905 [2] Cows which also met the current definition of "red cows": two positive tests in the previous
906 four tests (2 in 4)

907

Initial data → Sub-set of data → Analysis → Results

Tuberculosis test records (178,171 cows)

- One randomly selected, complete record per cow (158,287 cows) → Two-level multivariable logistic regression of positive Johne's test → Table 2

- At least one positive Johne's test followed by a minimum of one more Johne's test (16,471 cows) → Trends in Johne's test results after the first positive test - in relation to most recent tuberculosis test → Table 4

Matched by cow

Johne's test results (milk ELISAs) and milk records (225,296 cows)

- "Red cows": two positive Johne's tests in the previous four tests (7,308 cows) → Trends in Johne's test results after classified as a "red cow" → Table 3

- Cases: minimum of three Johne's tests, final two tests were positive (3,904 cows) → Estimate specificity and age-dependent sensitivity of a single Johne's test → Figure 2 & Table 5 → Calculate posterior probability of infection for serial Johne's test results → Figures 3 & 4

- Non-cases: minimum of nine Johne's tests, final eight tests were negative (42,704 cows)
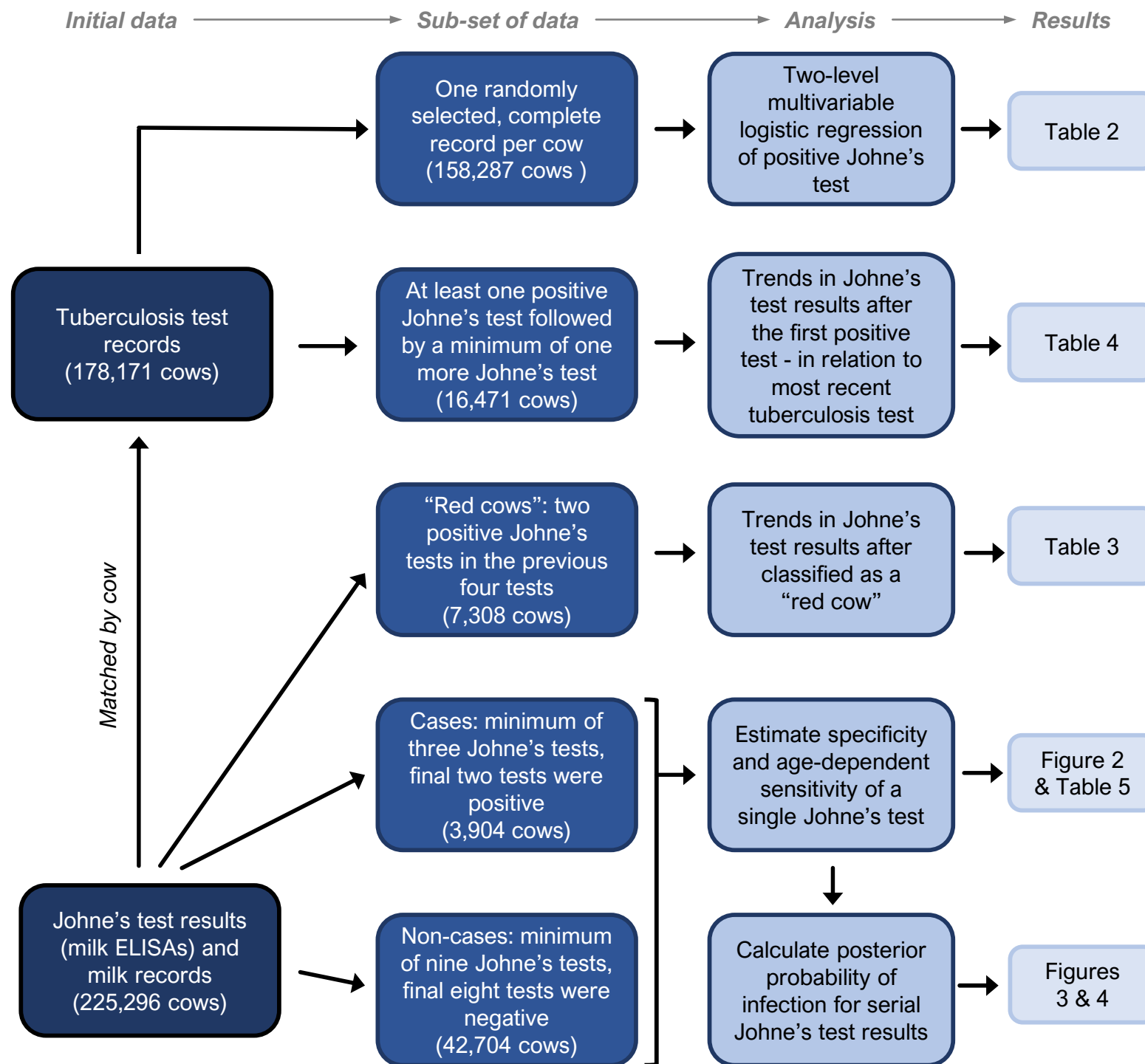
Figure 1.
Flow chart of data management and analysis

Figure 2. The predicted Johne's disease milk antibody ELISA sensitivity at different ages. Age-dependent sensitivity calculated using a non-linear logistic regression model (Eq. (2)) with shaded area showing the 95 % confidence interval.
Recent publications which used the same method are included for comparison (Nielsen et al., 2013; Meyer et al., 2018).
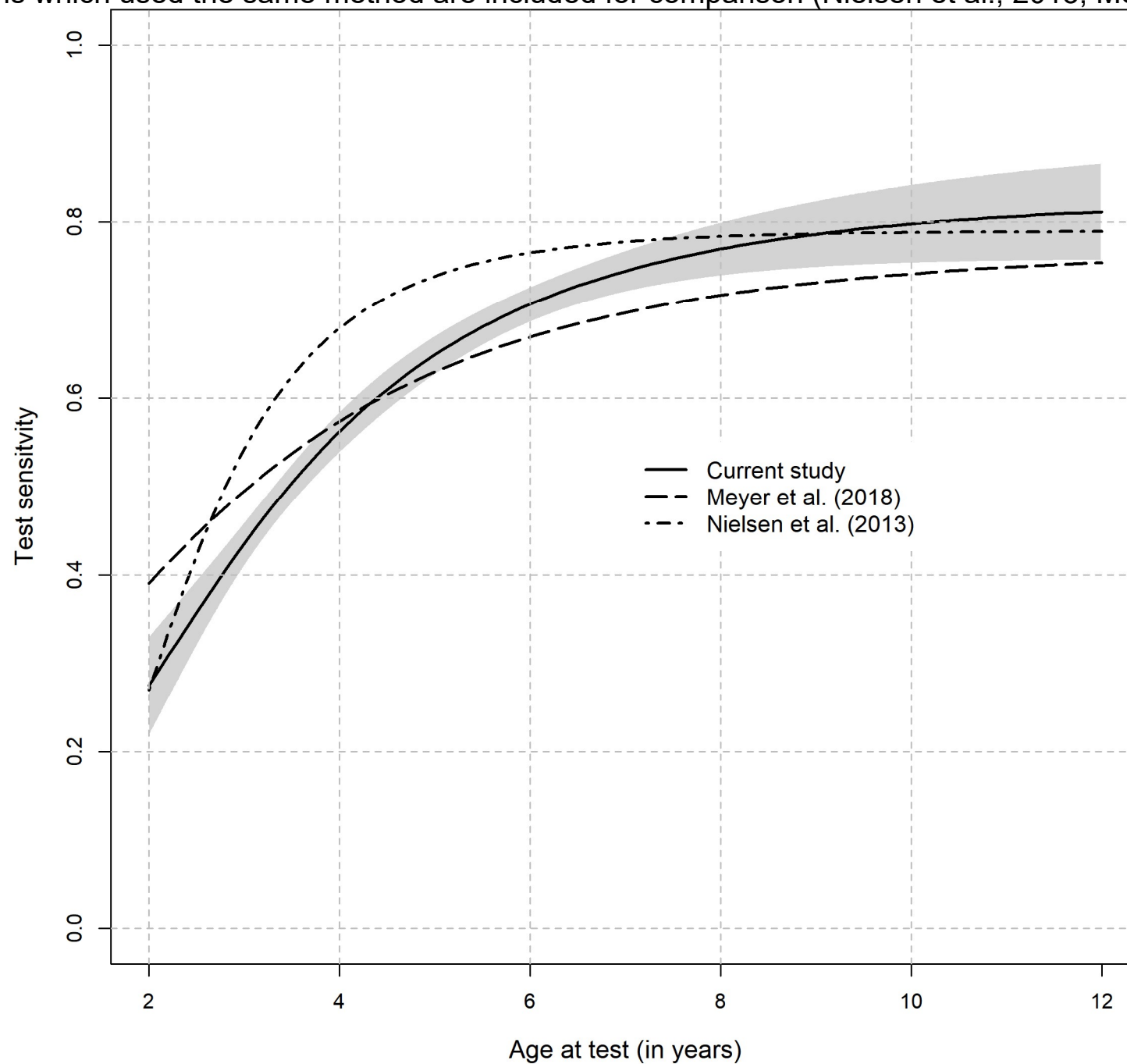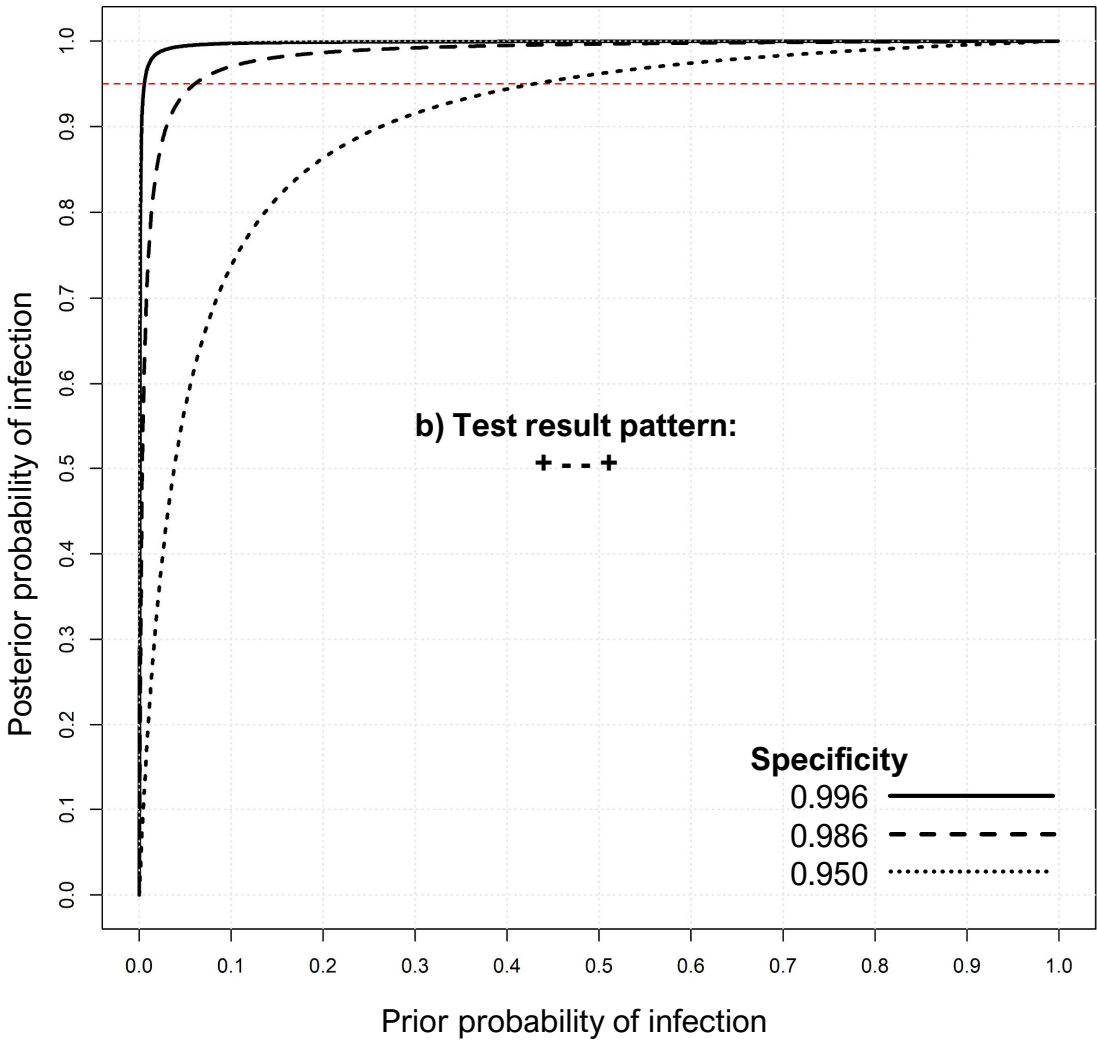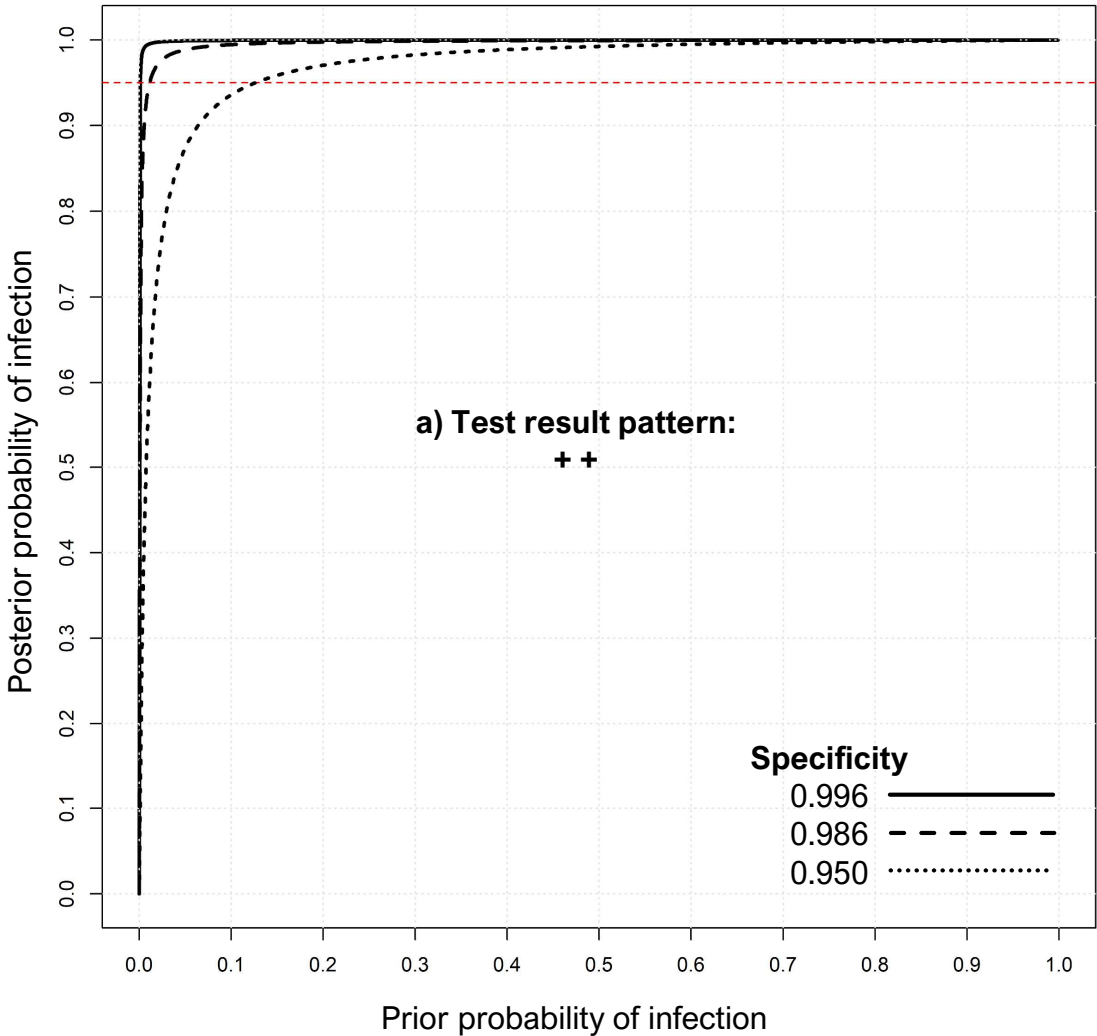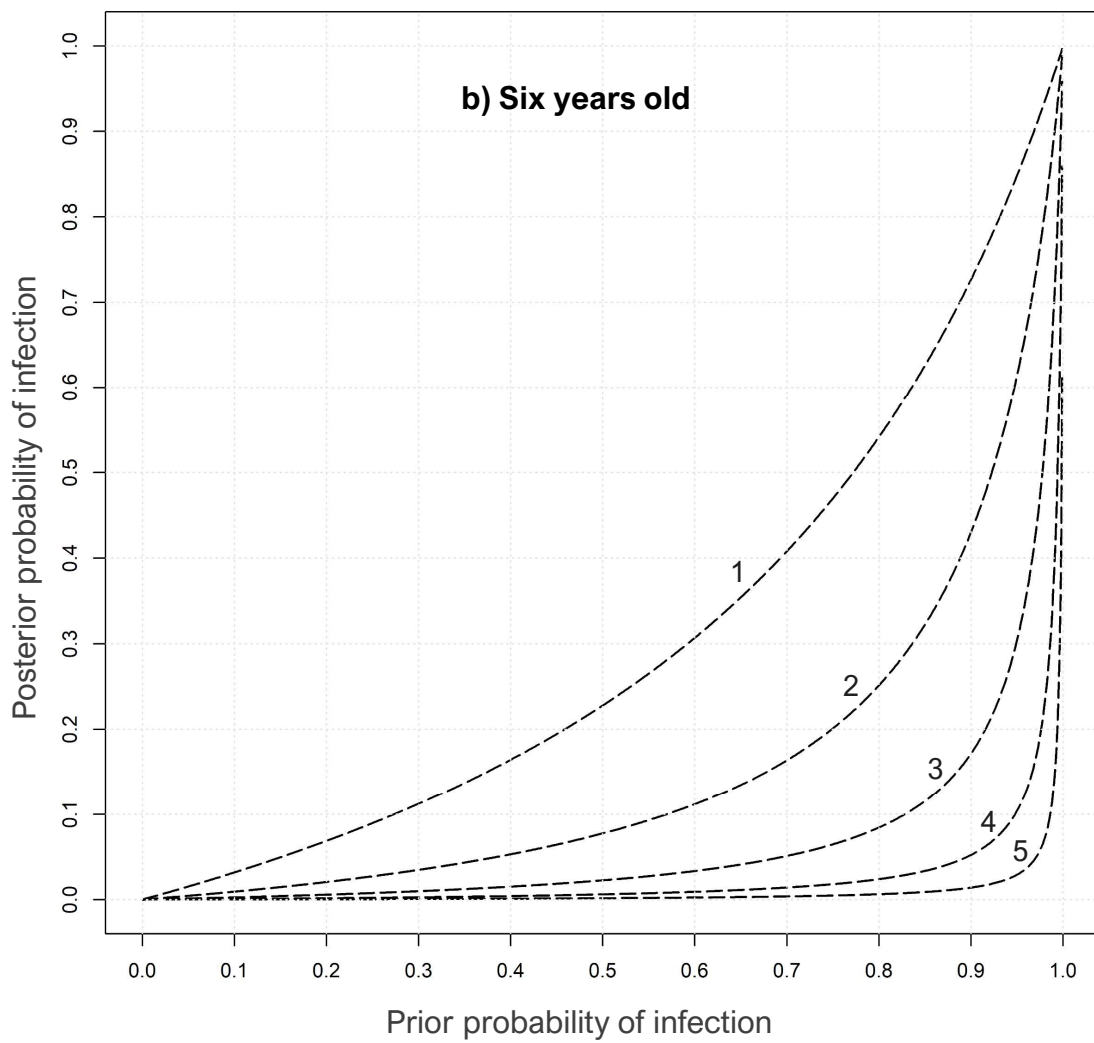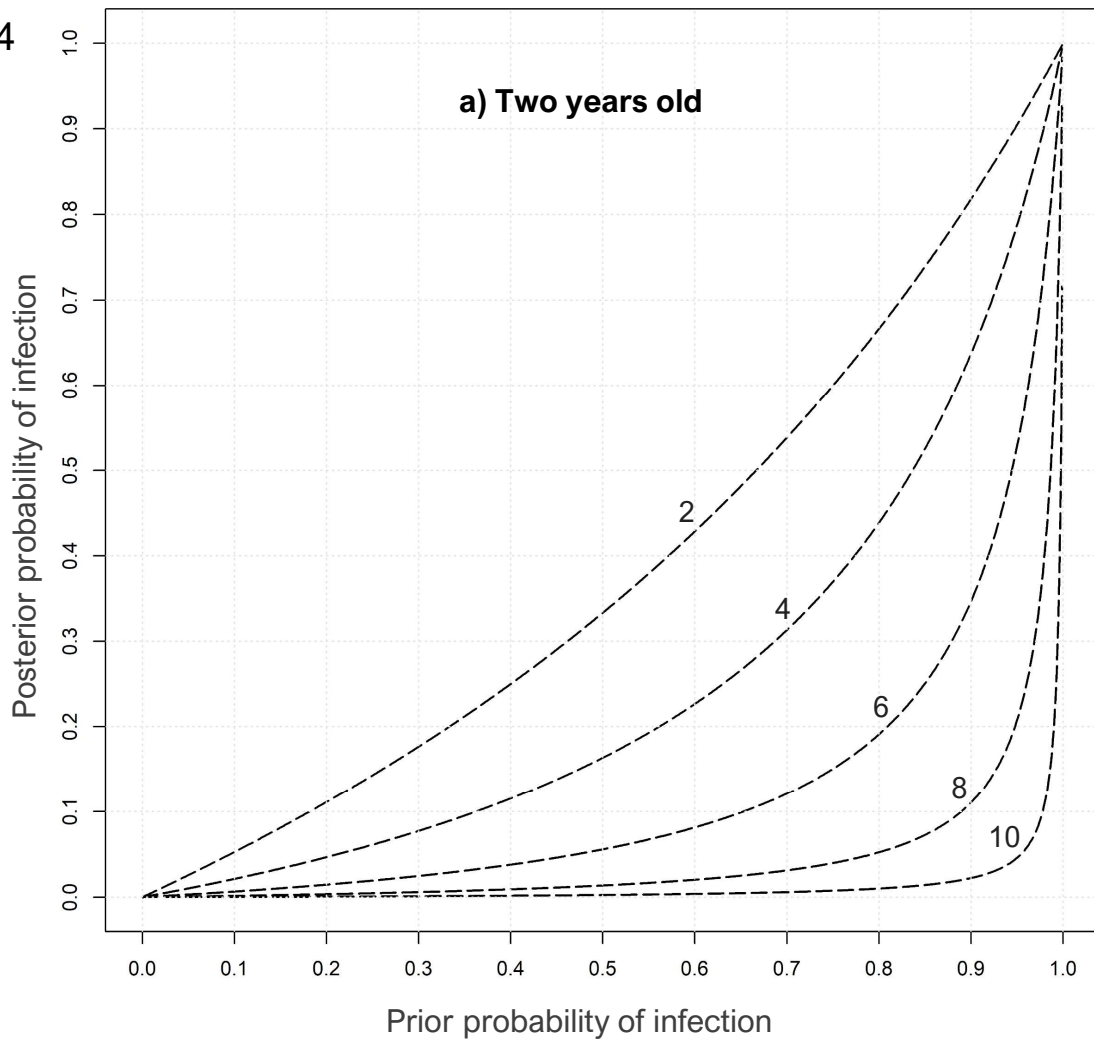
Figure 3

Figure 4

908 Supplementary Table 1. The proportion of subsequent tests that were positive following a cow's first ever positive test if cows had at least one

909 more test recorded, excluding cows which tested positive for the first time in 2017 (the final year of the study window). N = 9,927 (A total of 14,851

910 cows had at least one positive test in 2014 – 2016; 11,024 of these could be matched to the most recent bTB test, 1,097 cows had no further

911 tests after their first positive test.)

bTB - JD test interval at first positive JD test (days)		Number of cows with each proportion of positive tests following their first positive test (%)							
		≤30d		31-60d		61-90d		>90d	
Change in skin thickness at avian site during SICCT test		≤4mm	>4mm	≤4mm	>4mm	≤4mm	>4mm	≤4mm	>4mm
Proportion of subsequent tests that are positive after the first positive test [a]	0%	390 (68.2%)	32 (50.8%)	808 (54%)	50 (41%)	924 (48.5%)	71 (40.1%)	2,819 (53.9%)	139 (38.9%)
	1-99%	142 (24.8%)	20 (31.7%)	520 (34.8%)	55 (45.1%)	549 (28.8%)	54 (30.5%)	1,446 (27.6%)	121 (33.9%)
	100%	40 (7%)	11 (17.5%)	168 (11.2%)	17 (13.9)	432 (22.7%)	52 (29.4%)	970 (18.5%)	97 (27.2%)
Total		572 (100%)	63 (100%)	1,496 (100%)	122 (100%)	1,905 (100%)	177 (100%)	5,235 (100%)	357 (100%)

912

913 [a] Pearson's Chi-squared test to compare proportion of tests that were positive following the first positive test for different bTB-JD test intervals

914 and avian reaction details, $X^2 = 249.7$, df=14, p-value <0.001

915 Supplementary Table 2. The proportion of tests that were positive following a cow being classified as "red" based on two positive tests in the

916 previous four tests (2 in 4), only cows with at least one more test after being classified as "red" are included. The SICCT testing information

917 relates to the most recent SICCT test when cows had their first positive test result. N = 3,684 (A total of 7,308 cows were classified as "red cows",

918 5,277 of these could be matched to the most recent bTB test at their first positive JD test, 1,593 cows had no further tests after being classified

919 as a "red cow".)

bTB - JD test interval at first positive JD test (days)		Number of cows with each proportion of positive tests following their first positive test (%)							
		≤30d		31-60d		61-90d		>90d	
Change in skin thickness at avian site during SICCT test		≤4mm	>4mm	≤4mm	>4mm	≤4mm	>4mm	≤4mm	>4mm
Proportion of subsequent tests that are positive after the first positive test [a]	0%	49 (36.6%)	6 (23.1%)	158 (28.7%)	14 (20.9%)	152 (19.9%)	10 (11.8%)	427 (22.9%)	29 (15.2%)
	1-99%	37 (27.6)	7 (26.9%)	206 (37.4%)	21 (31.3%)	255 (33.4%)	33 (38.8%)	587 (31.4%)	59 (30.9%)
	100%	48 (35.8%)	13 (50%)	187 (33.9%)	32 (47.8%)	356 (46.7%)	42 (49.4%)	853 (45.7%)	103 (53.9%)
Total		134 (100%)	26 (100%)	551 (100%)	67 (100%)	763 (100%)	85 (100%)	1,867 (100%)	191 (100%)

920 [a] Pearson's Chi-squared test to compare proportion of tests that were positive following the first positive test for different bTB-JD test intervals

921 and avian reaction details, X^2 = 60.9, df=14, p-value <0.001