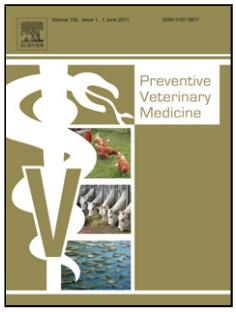
Meta-analyses of the sensitivity and specificity of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland



Title: Meta-analyses of the sensitivity and specificity of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland

Authors: Javier Nuñez-Garcia, Sara H. Downs, Jessica E. Parry, Darrell A. Abernethy, Jennifer M. Broughan, Angus R. Cameron, Alasdair J. Cook, Ricardo de la Rua-Domenech, Anthony V. Goodchild, Jane Gunn, Simon J. More, Shelley Rhodes, Simon Rolfe, Michael Sharp, Paul A. Upton, H. Martin Vordermeier, Eamon Watson, Michael Welsh, Adam O. Whelan, John A. Woolliams, Richard S. Clifton-Hadley, Matthias Greiner

PII: S0167-5877(16)30296-3
DOI: http://dx.doi.org/doi:10.1016/j.prevetmed.2017.02.017
Reference: PREVET 4209

To appear in: *PREVET*

Received date: 27-8-2016
Revised date: 7-1-2017
Accepted date: 26-2-2017

Please cite this article as: Nuñez-Garcia, Javier, Downs, Sara H., Parry, Jessica E., Abernethy, Darrell A., Broughan, Jennifer M., Cameron, Angus R., Cook, Alasdair J., de la Rua-Domenech, Ricardo, Goodchild, Anthony V., Gunn, Jane, More, Simon J., Rhodes, Shelley, Rolfe, Simon, Sharp, Michael, Upton, Paul A., Vordermeier, H.Martin, Watson, Eamon, Welsh, Michael, Whelan, Adam O., Woolliams, John A., Clifton-Hadley, Richard S., Greiner, Matthias, Meta-analyses of the sensitivity and specificity of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland.Preventive Veterinary Medicine http://dx.doi.org/10.1016/j.prevetmed.2017.02.017

Meta-analyses of the sensitivity and specificity of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland

Javier Nuñez-Garcia[a]*, Sara H. Downs [a]*, Jessica E. Parry[a], Darrell A. Abernethy[b,c], Jennifer M. Broughan[a,] Angus R. Cameron[d], Alasdair J. Cook[a,e], Ricardo de la Rua-Domenech R.[f], Anthony V. Goodchild[a], Jane Gunn[a], Simon J. More[g], Shelley Rhodes[a], Simon Rolfe[h], Michael Sharp[a], Paul A. Upton[a], H. Martin Vordermeier[a], Eamon Watson[a,i], Michael Welsh[j,k], Adam O. Whelan[a,l] , John A. Woolliams[m], Richard S. Clifton-Hadley[a], Matthias Greiner[n]

*Joint first authors

[a]Animal and Plant Health Agency (APHA), Weybridge, Surrey KT15 3NB, United Kingdom

[b]Veterinary Service, Department of Agriculture and Rural Development, Belfast BT4 3SB, United Kingdom

[c]Faculty of Veterinary Science, University of Pretoria, South Africa

[d]AusVet Animal Health Services Pty Ltd, PO Box 3180, South Brisbane, Qld 4101, Australia

[e]Department of Veterinary Epidemiology, School of Veterinary Medicine, University of Surrey, GU2 7AL, United Kingdom

[f]Advice Services, APHA, and Bovine Tuberculosis Programme, Department for Environment, Food and Rural Affairs, London SW1P 3JR, United Kingdom

[g]Centre for Veterinary Epidemiology and Risk Analysis, UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin 4, Ireland

[h]Office of the Chief Veterinary Officer, Welsh Government, Cardiff CF10 3NQ, United Kingdom

[i]National Milk Laboratories, Wiltshire SN15 1BN, United Kingdom

[j]Veterinary Sciences Division, Agri-Food & Biosciences Institute (AFBI), Belfast BT4 3SD, United Kingdom

[k]CSO SISAF Ltd, Northern Ireland Science Park, Unit 15A The Innovation Centre, Belfast BT3 9DT, United Kingdom

[l]Microbiology, Dstl, Porton Down, SP4 0JQ, United Kingdom

[m]The Roslin Institute, Roslin Biocentre, Roslin, Midlothian EH25 9PS, United Kingdom

[n]Federal Institute for Risk assessment (BfR), D-10589 Berlin, and Veterinary University Hannover, Foundation, Germany

Corresponding author: Sara Downs

Email: sara.down@apha.gsi.gov.uk

Postal address:

Dr Sara Downs Animal and Plant Health Agency (APHA) Weybridge, Woodham Lane, New Haw, Addlestone, Surrey, KT15 3NB, United Kingdom

Highlights

- First comparison of bovine tuberculosis test performance estimated by meta-analyses
- Sensitivity estimate of single comparative cervical tuberculin test moderate or low
- Wide credible intervals for estimates suggests heterogeneity in performance

Abstract

Bovine Tuberculosis (bTB) in cattle is a global health problem and eradication of the disease requires accurate estimates of diagnostic test performance to optimize their efficiency. The objective of this study was, through statistical meta-analyses, to obtain estimates of sensitivity (Se) and specificity (Sp), for 14 different ante-mortem and post-mortem diagnostic tests for bTB in cattle. Using data from a systematic review of the scientific literature (published 1934-2009) diagnostic Se and Sp were estimated using Bayesian logistic regression models adjusting for confounding factors. Random effect terms were used to account for unexplained heterogeneity. Parameters in the models were implemented using Markov Chain Monte Carlo (MCMC), and posterior distributions for the diagnostic parameters with adjustment for covariates (confounding factors) were obtained using the inverse *logit* function. Estimates for Se and/or Sp of the tuberculin skin tests and the IFN-γ blood test were compared with estimates published 2010-2015. Median Se for the single intradermal comparative cervical tuberculin skin (SICCT) test (standard interpretation) was 0.50 and Bayesian credible intervals (CrI) were wide (95% CrI 0.26, 0.78). Median Sp for the SICCT test was 1.00 (95% CrI 0.99, 1.00). Estimates for the IFN-γ blood test Bovine Purified Protein Derivative (PPD)-Avian PPD and Early Secreted Antigen target 6 and Culture Filtrate Protein 10 (ESAT-6/CFP10) ESAT6/CFP10 were 0.67 (95% CrI 0.49, 0.82) and 0.78 (95% CrI 0.60, 0.90) respectively for Se, and 0.98 (95% CrI 0.96, 0.99) and 0.99 (95% CrI 0.99, 1.00) for Sp. The study provides an overview of the accuracy of a range of

contemporary diagnostic tests for bTB in cattle. Better understanding of diagnostic test performance is essential for the design of effective control strategies and their evaluation.

Keywords: bovine tuberculosis, diagnostic tests, meta-analysis, performance, sensitivity, specificity

Introduction

Bovine tuberculosis (bTB) is endemic in cattle in many parts of the world resulting in substantial social and economic costs (Cousins, 2001, OIE, 2012, Defra, 2014). The disease is difficult to eradicate particularly where it has become endemic in a wildlife population, as in the case of the United Kingdom (UK) and the Republic of Ireland (RoI) (Abernethy et al., 2013, Godfray et al., 2013, More and Good, 2015). Mathematical modelling suggests that a substantial impact on the current epidemic in Great Britain (GB) could be achieved within a national eradication and control strategy through improvements in the detection of infected cattle and herds, with consequential reduction in cattle-to-cattle transmission (Cox et al., 2005, Defra, 2014, Brooks-Pollock and Wood, 2015, Moustakas and Evans 2015). The development of new and improved diagnostic tests for bTB as well as better targeting of tests is an important part of bTB research (de la Rua-Domenech et al., 2006, Schiller et al., 2010, Bezos et al., 2014, Brooks-Pollock and Wood, 2015).

The choice of diagnostic test and cut-off value used for defining an animal as infected is a trade-off between the need for sensitivity (Se) and specificity (Sp) within the local epidemiological context whilst meeting statutory requirements (EEC.,1964, de la Rua-Domenech et al., 2006, ESFA, 2012). High Se will maximise the likelihood of detecting infected animals but in low prevalence disease situations, high Sp is important to reduce the

number of false positive test results and therefore the number of healthy animals culled. Se and Sp are not only functions of the active diagnostic component of tests and the choice of diagnostic cut-off value, but may also be affected by the immunological profile of the host animal such as stage of infection, and local factors such as availability of laboratory facilities, training and experience of testers (EFSA, 2012). Concurrent infections and immunosuppression due to previous tests of the host can also influence the test results by reducing the sensitivity (Coad et al., 2010, Broughan et al., accepted). The presence of other mycobacteria in the environment may reduce specificity because of cross-reactivity with the test diagnostics (Aagaard et al., 2010, Coad et al., 2012). Se and Sp estimates also depend on the accuracy of the reference test (the evidence or "gold standard" used to determine whether an animal is truly infected) unless Latent Class Analysis (LCA) is conducted.

Probability distributions are used to describe the uncertainty produced in Se and Sp estimates due by factors such as the characteristics of the host animal, the test characteristics (and those of a reference standard if used) and local conditions. Statistical meta-analysis combining estimates of Se or Sp from different studies offers an approach to evaluate test performance whilst explicitly adjusting for heterogeneity caused by the factors described above. When conducted within a Bayesian framework, it also allows the inclusion of prior knowledge into the statistical inference. Furthermore, the Markov chain Monte Carlo (MCMC) simulation technique generates the empirical posterior distributions explicitly, which is another advantage since no distributional assumptions about Se and/or Se are required.

Reviews of diagnostic tests for bTB in cattle exist (de la Rua-Domenech et al., 2006, Schiller et al., 2010, Bezos et al., 2014). However, there has only been one published attempt to summarise data from the literature and generate summary estimates through meta-analysis

and this was limited to the caudal fold skin test, which is used in field surveillance in USA, New Zealand and Australia (Farnham et al., 2012). The objective of the current study was to estimate Se and Sp for 14 different ante-mortem and post-mortem diagnostic tests for bTB in cattle through a meta-analysis of data extracted as part of a systematic review of the literature (Downs et al., 2017b).

Materials and methods

Data source for estimates of sensitivity and specificity

The data for the meta-analysis were derived from references with estimates of Se and Sp of ante-mortem and post-mortem tests that had been identified in the systematic review (described in detail in Downs et al., 2017b). Briefly, a standardised two-stage process of review with extraction of relevant data was developed and agreed by an expert Working Group (WG). In stage 1, a comprehensive search strategy was developed to identify studies reporting the performance (Se and/or Sp) of diagnostic tests in cattle naturally exposed to bTB. Abstracts to references were reviewed against predetermined inclusion and exclusion criteria. Stage 2 comprised a more detailed review of references selected at stage 1 and involved reading the entire reference to determine whether the study was eligible for inclusion based on agreed criteria. If a study was eligible for inclusion in the meta-analysis, numerator (number that tested positive or negative) and denominator (number of true positives or negatives) data for the estimation of Se and Sp and covariate information were extracted and stored in a bespoke database. After resolving differences between decisions by reviewers who reviewed the same references, two datasets, one containing Se data and the other containing Sp data from the eligible studies were compiled. Thus, each record in the datasets included: test type, a numerator and denominator from which Se (or Sp) could be calculated and covariate data. The range of covariates for which data were extracted from references had been identified *a priori* by the WG before the review process commenced. Se

was calculated as the ratio of the number of the test positives and number of the true positives, and Sp was calculated as the ratio of the number of the test negatives and true negatives.

Some references reported more than one estimate of Se (or Sp) for a particular test type in the same population due to implementing different test modifications or cut-offs. In order to compensate for the fact that the same population may have been be used to calculate several performance estimates, an artificial covariate to weight the records according to the number of performance estimates reported using the same population was added to the database. For example, if a reference reported three Se estimates for the same test type over the same population, three records would have been added to the database. The three records would show the same population size covariate value, different test modification covariate value and the artificial covariate equal to 1/3. Introducing the artificial covariate allowed the use of the estimates reported within a reference without the population size being over-represented.

Removal of outliers

The data set for each diagnostic test type was explored for outliers using an iterative algorithm based on the chi-square distribution of the squared differences between the Se (or Sp) values and their mean value (Edwards et al., 1963). A very low threshold p-value equal to 0.001 for significance was set to exclude the most extreme values.

Selection and adaptation of covariates

The selection of covariates had been made by the WG prior to the review of potential references with performance estimates (Downs et al., 2017b). However, the covariates

included in the modelling were limited to those reported in all references with eligible performance estimates within each test-type (Table 3).

The modelling approach taken in this meta-analysis study, where Se and Sp were estimated through implementing two independent models, has been used previously (EFSA, 2006, 2008, Greiner et al., 2000). In this study however, both models were related by a *counter-parameter* defined as a covariate (pertaining to the same study and diagnostic test) that contained the corresponding Sp estimates for the Se modelling, and the corresponding Se estimates for the Sp modelling. For estimates where a *counter-parameter* was missing, i.e. estimates within the Se data set for which the Sp estimate was not reported in the reference (or vice versa), the median, calculated from all estimates from references for the test-type, was imputed.

All covariates, except the continuous *counter-parameter*, were transformed into binary variables with values 0 or 1. Value 1 represented the most preferred category or the category most applicable to current use of the test in UK and Ireland (baseline category). Value 0 represented the remaining complementary categories). For example, for the year of publication covariate, records corresponding to papers published between 2000 and 2009 were given a value equal 1, as the WG preferred the final estimate being based on the most recent publications. The remaining records were then set to 0. As a consequence of the binary coding, information associated to categories coded to 0 would be lost. However, this approach was justified by the gain in power and more robust analysis for the relatively small number of studies in each meta-analysis.

Covariates with low variability, where the proportion between the two dichotomous values was more extreme than 10% - 90%, were not included in the modelling. The threshold range of proportions of 10% - 90% was chosen as a compromise in order to only exclude covariates presenting very low variability. As a consequence, if a covariate is thus excluded, the results for the meta-analysis need to be interpreted as "mainly" applicable for the condition represented by the highest proportion. More severe threshold proportions (e.g. 30% - 70%) might produce more parsimonious models albeit with the penalty of a greater loss of information related to the excluded covariates. The opposite effect would apply with a less severe threshold.

Logistic regression modelling within a Bayesian Framework

The model used in this study was based on the logistic regression model with a random effect to account for the heterogeneity not explained by the model covariates. This is; log$it(Se) = z + \beta \boldsymbol{X}$, where the random effect term $z = N(\mu,\tau)$ and $\mu = N(0, \sigma)$. Parameters τ and σ were the variance within and between studies. Parameter β was the vector of parameters and $\boldsymbol{X}$ the matrix of covariates. To measure the effect of a baseline category for a particular covariate X_i in the final estimate, the odds ratio $exp(\beta_i)$ was used (the ratio between the products $Se\ (1\text{-}Se)$ and $Se_{\beta i}(1\text{-}Se_{\beta i})$ where $Se_{\beta i}$ is the Se estimate using the same model but excluding covariate X_i).

The modelling was run within a Bayesian framework by implementing the MCMC method for parameter estimation using WinBugs in the R environment (Lunn et al., 2000) essentially as described elsewhere (Hornick, 2011, Thomas et al., 2006). Two approaches were considered for estimating priors: first, requesting experts to provide performance estimates that could be used as priors, and second, generating priors from references where

performance could be estimated using LCA. The first approach was discarded after discussion within the WG; previous work (EFSA, 2008) had shown that relatively few experts are willing to commit to estimates of diagnostic test performance. With regard to the second approach, only six references were identified where test performance could be estimated through LCA (Downs et al., 2017b). The effect of using LCA estimates as priors was explored. However, with the removal of these references from the main data-set, the sample size for specific test-types became small and the priors had an excessive influence on the posterior distributions. For these reasons, it was agreed that non-informative or flat priors in the context of the regression parameter would be used and these were set to dnorm (0,0.001) in all models.

For each MCMC iteration, a logit(*Se*) value was obtained. The corresponding Se estimate was then calculated by applying to this value the inverse logit function (given by equation $\exp(x)/(1+\exp(x)$ where x = logit(Se)). The posterior distribution for Se was calculated as the density function of the values obtained from the inverse logit for all the MCMC iterations. The median value of the posterior distribution for Se was chosen as the point estimate and the Bayesian credible interval was also reported. Using the posterior distribution for the model parameters, a covariate was considered significant if its corresponding 95% credible interval did not contain zero.

Based on our experiences running these MCMC models, the burn-in and model update iterations were set depending on the number (p) of parameters in the model as 200*p and $800*p^2$, respectively. This approach provides an automated method to achieve convergence within a known time frame. Nevertheless, convergence of the process was monitored visually for three independent MCMC chains.

There were several cases where it was desirable to calculate estimates for Se based on different baseline categories for a particular covariate, for example where test performance had been estimated for different diagnostic antigens. In these cases, the binary coding and the modelling process were repeated for each of the different conditions. These were: for ELISA test-type, for covariate *Purified Protein Derivative* (PPD, tuberculin extracted from specific strains of *Mycobacteria*) the baseline category was set to *Bovine,* then to *Bovine-Avian,* and then to *MPB70*; for IFN-γ test-type, for covariate *PPD*, baseline categories were set to *Bovine*, *Bovine-Avian*, *ESAT6-CFP10* and *MPB70*; for the necropsy test-type baseline categories were set to *laboratory* and *meat inspection* and for SICCT test-type, baseline categories were set to *severe* and *standard* interpretation. See Tables 4 and 5.

Two modelling approaches

In modelling approach (A), a set of covariates was selected a priori by the WG as influential to the final performance estimates. Then, a preliminary analysis to investigate the associations between Se and these covariates using a stepwise logistic regression (with fixed effects) was carried out. Covariates that remained in the models were selected for a Bayesian-MCMC logistic modelling with a random effect (REM) as described above. Table 1 and Table 2 show the covariates included in the models and whether or not they are significant. For test-types with few estimates, convergence was not always achieved. For the latter cases, a basic estimate for Se was generated by applying the same modelling approach but using a simplified fixed effect model (FEM) with no covariates included.

The second modelling approach (B) was applied to the test-types with a successful REM in approach A. This consisted of an iterative algorithm to compute all the models for all

possible combinations of covariates. The model with the highest number of significant covariates was then selected. If a tie occurred, the model with the lowest deviance information criterion (DIC) (Spiegelhalter et al., 2002) was selected. DIC values were also used to compare performance between models A and B.

Note that the same modelling approach described above for Se was also applied for the Sp.

Results

Denominator and numerator data for estimating Se and/or Sp for 14 different diagnostic test types were extracted from 119 references identified as having eligible data during the systematic review (Downs et al., 2017b). The number of Se and Sp estimates varied between test-types and ranged from 1 to 166 (Tables 4 and 5). Two and six estimates that were identified as outliers for Se in the caudal fold test and IFN-γ blood test respectively were removed from respective datasets before the statistical modelling was conducted.

Covariates

The initial number of covariates extracted from each reference varied from 30 to 51 and from 27 to 42 per Se and Sp estimate respectively. Only covariates consistently measured within all references for a test type were included in the models. Up to 17 and 12 covariates for Se and Sp respectively were included in initial models but many were eliminated during the logistic regression analysis (Tables 1, 2 and 3). Eighty percent of references that reported at least one Sp estimate for IFN-γ or ELISA and 76.6% of references that reported at least one Se estimate for IFN-γ or ELISA did not indicate a preferred diagnostic cut-off value for a positive response to the test. Choice of reference standard, sampling strategy (random or

census-based compared to other) and the counter estimate of Sp (when modelling Se) and the counter estimate of Se (when modelling Sp) were influential in virtually all models.

Estimates for sensitivity and specificity by test type

Pooled unadjusted estimates for Se and Sp and estimates weighted by population size and posterior distributions from logistic regression with MCMC modelling with covariates set for UK and Irish conditions, are reported in Tables 4 and 5 (further detail in the online supplement).

Test performance estimates from modelling procedure A

The posterior distributions for estimates of Se had wide credible intervals due to the large amount of heterogeneity unexplained by the models and it was not possible to demonstrate statistically that the estimates from different tests were significantly different from one another (Table 4). Credible intervals were narrower for the Sp estimates (Table 5).

Ante-mortem tests

The highest median sensitivity in the tuberculin skin tests was for the SIT test (0.94 for Model A and 0.81 for Model B). The median sensitivity of the SICCT test at standard interpretation was 0.50 (Model A) and 0.63 (model B). This difference is a consequence of including covariates year of publication, PPD manufacturer and number of reviewers in model A and not in model B (See Table 1, Table 2 and online supplement). Sensitivity of the caudal fold test was lower than that for the SIT test but higher than for the SICCT test, both at standard and severe interpretation (Table 4).

None of the median Se estimates for the IFN-γ or ELISA blood tests was higher than the estimate for the SIT test. The median estimates for Se of the IFN-γ blood tests were higher and the width of the credible intervals narrower than for the serological ELISA with equivalent diagnostic antigens. The median Se for IFN-γ Bovine PPD - Avian PDD was higher than the Se for the SICCT test at standard interpretation. The median estimates for Se of the IFN-γ ESAT6/CFP10 was slightly lower than that for the IFN-γ Bovine, but higher than both comparative PPD tests (the SICCT test and IFN-γ Bovine PPD – Avian PPD). In general, the width of the credible intervals were narrower for the IFN-γ blood tests than for the skin tests but there was considerable overlap of the credible interval between test-types except for IFN-γ using MPB70 which had the lowest Se of all the tests.

The SICCT test at standard interpretation and the caudal fold test had the highest median Sp of all the ante-mortem tests (Table 5). Sp distributions for the IFN-γ blood tests were narrower than those estimated for the ELISA and the median estimates slightly lower. The median estimate for Sp of ESAT/CFP10 was the highest of all the IFN-γ tests and higher than the SIT. The estimates for the Sp of both the LBAA and the rapid test had wide credible intervals. The only estimate for Sp of LBAA was lower than all the median estimates for Sp of the IFN-γ blood tests apart from with the MPB70 diagnostic antigen.

Post-mortem tests

Detailed laboratory necropsy had the highest median Se of all the post-mortem tests (0.96) and ante-mortem tests (Table 4). The median Se for meat inspection in the slaughterhouse was substantially lower and estimated Se for PCR was very low. Sp was estimated to be perfect for meat inspection, PCR and microscopic examination and 0.99 for culture.

Test performance estimates from modelling procedure B

The models generated through automated selection of covariates were more parsimonious than the models for the equivalent tests achieved through procedure A. The distribution of the credible intervals were generally narrower for model B compared to model A for the equivalent test types. Covariates sampling method, region, year of publication and post-mortem type were included in model A but not in model B. While most estimates using procedures A and B were similar, there were a few that were considerably different. All Se for the comparative PPD tests (SICCT test, IFN-γ Bovine PPD – Avian PPD, ELISA Bovine PPD – Avian PPD) were higher with model B. As an extreme example, Model A generated an estimation of Se of PCR of 0.14 whereas the median estimates using model B was 0.86. This difference was due to the inclusion of production type in model A (not selected during automated procedure B) in model A in which dairy cattle were set to baseline. There were two Se estimates in dairy cattle with strong leverage, although they had not been identified as outliers.

Discussion

Meta-analyses were conducted to summarise estimates of performance for diagnostic tests for bTB in cattle obtained through a comprehensive systematic review (Downs et al., 2017b). The analyses used similar methodology to earlier work evaluating bTB tests in deer (EFSA, 2008), and the study is the first to evaluate and compare the performance of a range of ante-mortem and post-mortem diagnostic tests for bTB in cattle using a meta-analysis approach. In estimating the Se and Sp from studies published between 1934 and 2009, the study provides an overview of the accuracy of different tests used to diagnose bTB in cattle.

The SICCT test at standard interpretation was estimated through the meta-analysis to have a median Se of 0.50 (95% CrI 0.26, 0.78) by model A and 0.64 (95% CrI 0.48, 0.78) by model B (Table 4). These estimates can only be described as low or moderate, which is of concern given the widespread use of SICCT test and its official status as a standalone test (EEC, 1964). However, the credible intervals on both estimates were wide. Since the meta-analysis was conducted, the Se of the SICCT test has been estimated as 0.81 (95% 0.70-0.89) using data from herds that were subject to depopulation in GB between 1988 and 2008 (Karolemeas et al., 2012 and Table 6) and six Se estimates have been reported in studies using LCA (Figure 1). The Se estimate from the whole herd slaughters was considerably higher than the estimates from the meta-analysis (Tables 4 and 6). In these circumstances, infected animals are likely to have developed significant pathology, which means post-mortem confirmation of infection (and therefore estimated Se) is likely to have been higher than in the SICCT test positive cattle detected through routine surveillance. In fact, inclusion of the whole herd slaughter data, kindly made available to us by the authors, made minimal difference to the Se estimates. The revised median Se estimates at standard interpretation from modelling procedures A and B were 0.49 (95% CrI 0.27, 0.74) and 0.65 (95% CrI 0.52, 0.77) respectively. Comparison with the LCA studies show that the meta-analysis estimates are at the lower end of the spectrum for reported Se of the SICCT test although not statistically significantly different (Figure 1).

It is possible that UK and RoI conditions give rise to lower Se in tests based on the response to PPD. PPD consists of an ill-defined mixture of proteins, and some of its antigenic components are present in non-pathogenic environmental bacteria (Dunn et al., 2005, Tameni et al., 1998). Subtle changes in *M. bovis* related to intense surveillance and control strategies may have favoured survival of strains of the bacterium better able to evade the test (Smith et

al., 2006). Between 20-50% of reactors to the SICCT test detected during surveillance in GB have post mortem evidence of infection (de la Rua-Domenech et al., 2006). However, the high specificity of the SICCT test suggests that a large proportion of SICCT test positive cattle detected are truly infected. The specificity of SICCT test estimated through the meta-analysis was consistent with a recent estimation of Sp using GB surveillance data, in which freedom from *M. bovis* infection in cattle was based on geographical distance from confirmed infected herds (Goodchild et al., 2015) (Table 7); and also four other estimates from studies using LCA (Figure 2). Another LCA by Hartnack et al., 2012 has reported a specificity of less than 0.70 for the SICCT test (at standard interpretation) but the study population included a group of non-reactor cattle purposively selected from herds that had persistent recurring bTB incidents (Downs et al., 2008).

A meta-analysis of the performance of the CF skin test in the USA reports summary Se (weighted mean 0.859) and Sp (weighted mean 0.926) estimates that are broadly consistent with the meta-analysis distributions (Farnham et al., 2012, Tables 4 and 5).

The meta-analysis results suggested that the IFN-γ blood test using PPD has similar or higher Se than most other tests including the tuberculin skin tests and the ELISA antibody tests using PPD. Se estimates reported for the IFN-γ blood test using Bovine PPD -Avian PPD since 2009 are slightly higher than the median estimates reported in the meta-analysis but within the observed credible intervals (Tables 4 and 6). The same pattern is seen when the meta-analysis results are compared to estimates calculated through LCA since 2009 (Figure 1).

The Sp of the IFN-γ blood test with PPD estimated by meta-analysis was over 0.90 and similar to estimates reported since 2009 (Table 7 and Figure 2) although not as high as the Sp of the SICCT test. A number of studies have shown that the tuberculin skin test and IFN-γ test using PPD detect overlapping yet partly distinct populations (Neill et al., 1994, Coad et al., 2008). Since 2009, point estimates for Sp for the IFN-γ blood test with PPD of less than 0.70 has been calculated through LCA in two studies (Figure 2). A possible explanation for the lower Sp estimates in the LCA studies compared to the meta-analysis is differences in infection prevalence between studies.

The IFN-γ blood test with early secretory antigen target 6kDa (ESAT6) and culture filtrate protein 10 (CFP10) showed as good or slightly better Se than the IFN-γ using PPD suggesting that this defined antigen on its own or in combination with other antigens could have a role in bTB surveillance and control strategies (see also EFSA, 2012). There is also evidence in the literature that ESAT6/CFP10 may have considerable advantages over PPD based tests where cross-reactivity with other mycobacteria such as *M. avium ssp paratuberculosis* could be a problem (Aagard et al., 2010). Small modifications in test platform (such as whether the antigens are presented as a protein and peptide cocktail) can affect performance (Schiller et al., 2009, Casal et al., 2014) although we were not able to control for these differences in the meta-analysis. Skin tests have now been developed with ESAT6 and CFP10 (Table 6 and 7).

The anti-body detection based ELISAs generally had lower accuracy than the tuberculin skin tests and the IFN-γ test. Considerable variability in estimates from 2010-2015 was observed (Tables 6 and 7). Since the ELISAs detect a serological response, sensitivity is likely to be

higher when evaluated in animals with later stage disease with macroscopic lesions, which is also when the animals are likely to be their most infective (Casal et al., 2014). It should also be noted that given that PPD-based skin tests are known to boost antibody responses (Lightbody et al., 2000), the use of antibody tests in the absence of skin-test surveillance would likely reduce their Se estimates further.

The hierarchy of the Se estimates for post-mortem tests was generally in the direction anticipated. Detailed laboratory post-mortem had a higher Se than the other post-mortem tests. Unsurprisingly microscopic examination was less sensitive than meat inspection or detailed post-mortem since all the reference standards for microscopy were laboratory necropsy, meat inspection or culture (Downs et al., 2017b). Correlation between the post-mortem tests and reference standards commonly used in their evaluation is likely to be higher than between the ante-mortem tests and reference standards because these reference standards are predominantly post-mortem tests (Downs et al., 2017b). This may have biased estimates of Se of post-mortem tests upwards compared to ante-mortem tests and may explain why the Se of detailed laboratory post-mortem was higher than that of all ante-mortem tests and Se of meat inspection higher than that of the SICCT test.

Meta-analysis has been more commonly used for summarising effects from therapeutic interventions than for estimating diagnostic test performance. A specialised approach is required because of the non-normal distribution of Se and Sp and the need to build separate regression models (Harbord et al., 2007). Separate regression modelling procedures for the two performance characteristics have also been conducted in other meta-analyses for diagnostic tests (Midgette et al. 1993, Greiner et al., 2009, Jones and Athanasiou 2005, Rosman and Korsten 2007, EFSA, 2006, EFSA, 2008). The regression approach, in contrast

to other models (Moses et al., 1983; Irwig et al., 1995), can accommodate studies that contribute more than one parameter (Se, Sp) estimate. Over 10% of references measuring the performance of IFN-γ blood test with Bovine PDD - Avian PPD diagnostic antigens reported at least seven estimates for Se and 19 estimates for Sp. Factors, for which there was available information and that might explain differences in performance estimates were included in the models and the estimates for Se and Sp were modelled to be as relevant as possible to test conditions in the UK and the RoI.

As well as extracting performance data, reviewers assessed whether studies were methodologically sound (Downs et al., 2017a). This investigation showed that most of the studies of diagnostic test performance identified in the review had some methodological deficiencies, regardless of test-type. Common problems across all test types, which could bias assessments included awareness of the index diagnostic test results when performing the reference test (and vice-versa) and absence of information about animal withdrawals and un-interpretable study results (Downs et al., 2017a). Se and Sp estimates were not weighted by a quality score in the meta-analysis because of the difficulty in designing an informative and neutral score across different test types (Whiting et al., 2005, Leeflang et al., 2008). Publication bias is another issue that could have affected estimates. However, we conducted a comprehensive and systematic search of both published and grey literature and standardised our data extraction protocol (Irwig et al., 1994, Stroup et al., 2000, Bossuyt et al., 2004, Downs et al., 2017b). The summary estimates from the meta-analyses may still be more representative of the diagnostic test performance than estimates from individual studies by providing a better guide of uncertainty or variability in performance.

Two modelling procedures were used. Modelling procedure A included more variables considered *a priori* as influential and that were associated with performance in the models but did not necessarily maximise model fit. Modelling procedure B used automated selection of variables with the best model selected by DIC. Procedure B resulted in narrower credible intervals but procedure A may be less biased (Altman and Andersen, 1989), and more indicative of the heterogeneity in test performance as it truly exists. Commonly, both procedures produced similar distributions but on some occasions the posterior distributions varied significantly. For example, the estimation of the Se of PCR was much higher using modelling procedure B compared to A due to the inclusion of production type in A. Modelling using procedure A suggested that PCR is a less sensitive test for bTB in dairy cattle compared to non-dairy cattle whereas using procedure B did not. Dairy reactor cattle from infected herds are less likely to have visible lesions than non-dairy (O'Hagan et al., 2015, Downs et al., 2017). This may affect the sensitivity of PCR, which is conditional on the Se of necropsy and volume of DNA available (Cardosa et al., 2009). However, differences in PCR primer and other methodological factors not included in the models could also explain the differences in estimates. Similar estimates for Se were reported in five references where performance of PCR was measured in dairy and non-dairy cattle using the same target antigen gene sequence (IS6110) (Liebana et al., 1995, Cornejo et al., 1998, Zanini et al., 2001, Zumarraga et al., 2002, Meikle et al., 2007).

In different studies, performance of the IFN-γ and ELISA blood tests was estimated using different algorithms and cut-off values for a positive response (see the footnotes to Tables 6 and 7 for examples from recent references). It was not possible to determine, in simple comparisons, whether differences in performance estimates were related to the test or affected by the cut-off value chosen, the population or some other factor. Different research

organisations had developed different scales to meet requirements of the specific conditions of their target surveillance population. A counter-parameter making use of the fact that Se and Sp are approximately inversely related was included in our models to adjust for differences in cut-off values, although this will not have eliminated all confounding (EFSA, 2006, 2008, Greiner et al., 2009).

All estimates for performance in the meta-analyses were relative to a reference standard and were therefore influenced by the accuracy of the reference test. The over-riding weakness to evaluating a test against a reference standard is that the reference standard itself will be imperfect and the test being evaluated can equal but will never be calculated as better than the reference standard. Additionally, different tests often perform optimally at different stages in the natural history of the disease within an animal (TDR, 2010). Most screening tests use immunological markers correlated with early stage disease whereas reference tests are commonly based on visible disease that occurs at later stages (Welsh et al., 2005, Pollock et al., 2005). This bias was less likely with the estimation of Sp compared to Se because Sp analyses were limited to studies with a population sample reported to be bTB free.

Latent class analysis (LCA) would have allowed the estimation of Se and Sp without reliance on a reference standard but requires cross-tabulated data for the performance of at least two diagnostic tests in two distinct populations or population subgroups with different levels of disease prevalence or, alternatively, results from one population and three diagnostic tests (Toft et al., 2005). Also LCA generally requires larger sample sizes to achieve similar precision. In the systematic review, Bayesian latent class models could be fitted to the results of six references compared to 119 references where test performance was evaluated against a reference standard (Downs et al., 2017b). Furthermore, LCA is not without methodological

limitations. For example, it has been reported that Se estimated through LCA tends to be biased towards test Se in the population with the highest disease or infection prevalence (Toft et al., 2005). The Se estimates from the meta-analysis for the SICCT test and IFN-γ test using PPD were mostly consistent with Se estimates from LCA studies published between 2009 and 2015 (Figure 1). There was more heterogeneity between Sp estimated through the meta-analysis and Sp estimated through LCA (Figure 2).

The use of LCA data as a prior in the Bayesian regression analysis was also explored. However, non-informative priors were preferred because of the paucity of performance estimates for most test-types which meant that the studies yielding LCA data were very influential. Additionally, data from studies informing the regression priors could not be used in the posterior estimation of test performance limiting power.

In conclusion, this study based on a comprehensive systematic literature review provides an overview of the empirical evidence for the accuracy of a wide range of contemporary diagnostic tests for bTB in cattle. We attempted to control for many factors that may influence Se and Sp, but acknowledge that the adjustment will not be perfect. The wide credible intervals, particularly for the Se estimates, are likely to be due to the limited number of studies with eligible data for some tests and differences in study methodology but may also reflect true heterogeneity in performance as Se varies with local conditions. Understanding performance characteristics in the context of local conditions is essential for the design of effective control strategies and their evaluation.

Funding: The SE3238 project "Meta-analysis of diagnostic tests and modelling to identify appropriate testing strategies to reduce *M. bovis* infection in GB herds" was funded by the UK Department for Environment, Food and Rural Affairs (Defra).

Authors' contributions

All authors contributed to the design of a systematic review and selection of data to be extracted from references for the statistical meta-analysis. PU with advice from SD and JP created bespoke databases. JP, DA, JB, AC, RdlR, AG, JG, SM, JNG, SR, MS, MV, EW, MW AW, JW, RCH and SD conducted the reviews of reference papers about diagnostic tests. PU, JP and SD undertook components of data-cleaning. MG and JNG with advice from SD determined the approach to be taken in the statistical meta-analysis. JNG conducted the statistical meta-analysis. SD and JNG drafted the first version of the manuscript circulated to co-authors. All authors contributed to the discussion of results and read, commented on and approved the final manuscript. SD was the project leader.

Acknowledgements.

We thank Sir David Cox FRS and Professor William Brown for helpful advice during the development of the modelling procedure. We thank Dirk Pfeiffer for helpful comments on an earlier draft of the paper.

Additional Material

Summary tables and graphs for the sensitivity and specificity from the meta-analysis for different diagnostic tests and a description of the search for additional references conducted in 2015 are provided in the supplementary material.

References

Aagaard, C., Govaerts, M., Meikle, V., Gutiérrez-Pabello, J.A., McNair, J., Andersen, P., Suárez-Güemes, F., Pollock, J., Espitia, C., Cataldi, A., 2010. Detection of bovine tuberculosis in herds with different disease prevalence and influence of paratuberculosis infection on PPDB and ESAT-6/CFP10 specificity. Prev. Vet. Med. 96, 161-9.

Abernethy, D.A., Upton, P., Higgins, I., McGrath, G., Goodchild, A.V., Rolfe, S.J., Broughan, J.M., Downs, S.H., Clifton-Hadley, R., Menzies, F.D., de la Rua-Domenech, R., Blissit, M., Duignan, A., More, S.J., 2013. Bovine tuberculosis trends in the United Kingdom and Republic of Ireland, 1995 to 2010. Vet. Rec. 172, 312.
Altman, D.G., Andersen, P.K., 1989. Bootstrap investigation of the stability of a Cox regression model. Stat. Med. 8, 771-783.

Antognoli, M.C., Remmenga, M.D., Bengtson, S.D., Clark, H.J., Orloski, K.A., Gustafson, L.L., Scott, A.E., 2011. Analysis of the diagnostic accuracy of the gamma interferon assay for detection of bovine tuberculosis in U.S. herds. Prev. Vet. Med. 101, 35-41.

Bezos, J., Casal C., Romero, B., Schroeder, B., Hardegger, R., Raeber, A.J., Lopez, L., Rueda, P., Dominguez, L., 2014. Current ante-mortem techniques for diagnosis of bovine tuberculosis. Res. Vet. Sci. 97 Suppl, S44-S52.

Bossuyt, P.M., Reitsma J.B., Bruns D.E., Gatsonis C.A., Glasziou P.P., Irwig L.M., Lijmer J.G., Moher D., Rennie D., Henrica C.W. de Vet. For the STARD group., 2004. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Fam Pract. 21, 4-10.

Brooks-Pollock, E., Wood, J.L., 2015. Eliminating bovine tuberculosis in cattle and badgers: insight from a dynamic model. Proc Biol Sci. B. 282, 1808. DOI:10.1098/rspb.2015.0374.

Broughan, J.M., Judge, J., Ely, E., Bishop, H., Delahay, R. Clifton-Hadley, R.S., Parry, J., Wilson, G., Goodchild, A.V., Downs, S.H., Risk factors for bovine tuberculosis infection in cattle: A review of relevant literature. Epidemiol. Infect., Accepted.

Buddle, B.M., Wilson, T., Luo, D., Voges, H., Linscott, R., Martel, E., Lawrence, J.C., Neill, M.A., 2013. Evaluation of a commercial enzyme-linked immunosorbent assay for the diagnosis of bovine tuberculosis from milk samples from dairy cows. Clin. Vaccine Immunol. 2013. 20, 1812-1816.

Cardoso, M.A., Cardoso, R.F., Hirata, R.D., Hirata, M.H., Leite, C.Q., Santos, A.C., Siqueira, V.L., Okano, W., Rocha, N.S., Lonardoni, M.V., 2009. Direct detection of *Mycobacterium bovis* in bovine lymph nodes by PCR. Zoonoses Public Hlth. 56, 465-470.

Clegg, T.A., Duignan, A., Whelan, C., Gormley, E., Good, M., Clarke, J., Toft, N., More, S.J., 2011. Using latent class analysis to estimate the test characteristics of the γ-interferon test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions. Vet. Microbiol. 151, 68-76.

Casal, C., Diez-Guerrier, A., Alvarez, J., Rodriguez-Campos, S., Mateos, A., Linscott, R., Martel, E., Lawerence, J.C., Whelan, C., Clarke, J., O'Brien, A., Dominguez, L., Aranaz, A.,

2014. Strategic use of serology for the diagnosis of bovine tuberculosis after intradermal skin testing. Vet. Microbiol. 170, 342-351.

Coad, M., Clifford, D., Rhodes, S.G., Hewinson, R.G., Vordermeier, H.M., Whelan, A.O., 2010. Repeat tuberculin skin testing leads to desensitisation in naturally infected tuberculous cattle which is associated with elevated interleukin-10 and decreased interleukin-1 beta responses.Vet. Res. 41, 14.

Coad, M., Clifford, D.J., Vordermeier, H.M., Whelan, A.O., 2012. The consequences of vaccination with Johne's disease vaccine, Gudair, on diagnosis of bovine tuberculosis. Vet. Rec. 171, 448.

Coad, M., Downs, S.H., Durr, P.A., Clifton-Hadley, R.S., Hewinson, R.G., Vordermeier, H.M., Whelan, A.O., 2008. Ancillary blood based assays detect *Mycobacterium bovis* infected cattle missed by tuberculin skin-testing. Vet. Rec. 162, 382-384.

Cornejo, B.J., SahagunRuiz, A., SuarezGuemes, F., Thornton, C.G., Ficht, T.A., Adams, L.G., 1998. Comparison of C-18-carboxypropylbetaine and glass bead DNA extraction methods for detection of *Mycobacterium bovis* in bovine milk samples and analysis of samples by PCR. Appl. Environ. Microb. 64, 3099-3101.

Cousins, D.V., 2001. *Mycobacterium bovis* infection and control in domestic livestock. Rev. Sci. Tech. OIE. 20, 71-85.

Cox, D.R., Donnelly, C.A., Bourne, F.J., Gettinby, G., McInerney, J.P., Morrison, W.I., Woodroffe, R., 2005. Simple model for tuberculosis in cattle and badgers. P. Natl. Acad. Sci. USA. 102, 17588-17593. DOI: 10.1073/pnas.0509003102.

da Silva, E.B., Silva, B.D., Leon, J.R., Kipnis, A., Santos, I.K., Junqueira-Kipnis, A.P., 2011. Using BCG, MPT-51 and Ag85 as antigens in an indirect ELISA for the diagnosis of bovine tuberculosis. Vet. J. 187, 276-278.

Defra (Department for the Environment and Rural Affairs), 2014. The Strategy for achieving Officially Bovine Tuberculosis Free status for England. April 2014 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/300447/pb14088-bovine-tb-strategy-140328.pdf (accessed 3 December 2015).

de la Rua-Domenech, R., Goodchild, A.T., Vordermeier, H.M., Hewinson, R.G., Christiansen, K.H., Clifton-Hadley, R.S., 2006. Ante mortem diagnosis of tuberculosis in cattle: a review of the tuberculin tests, gamma-interferon assay and other ancillary diagnostic techniques. Res. Vet. Sci. 81,190-210.

Downs S.H., Broughan J.M., Goodchild A.V., Upton P.A., Durr P.A., 2016. Responses to diagnostic tests for bovine tuberculosis in dairy and non-dairy cattle naturally exposed to *Mycobacterium bovis* in Great Britain. Vet. J. 216, 8-0.

Downs, S.H., Durr, P., Edwards, J., Clifton-Hadley, R., 2008. Trace micro-nutrients may affect susceptibility to bovine tuberculosis in cattle. Prev. Vet. Med. 87, 311-326.

Downs S.H., More S.J., Goodchild A.V., Whelan A.O., Abernethy D.A., Broughan J.M., Cameron A., Cook, A.J., de la Rua-Domenech R., Greiner M., Gunn J., Nunez-Garcia J., Rhodes S., Rolfe S., Sharp M., Upton P.A., Vordermeier H. M., Watson E., Welsh M., Woolliams J.A., Clifton-Hadley R.S., Parry J.E., 2017a. Evaluation of the methodological quality of studies of the performance of diagnostic tests for bovine tuberculosis in cattle, using QUADAS adapted for the veterinary field. Unpublished, submitted to Prev. Vet. Med.

Downs, S.H., Parry, J. E., Broughan, J.M., Goodchild, A.V., Upton, P.A., Nunez-Garcia, J., Abernethy, D.A., Cameron, A.R., Cook, A.J., de la Rua-Domench, R., Greiner, M., Gunn, J., Pritchard, E., Rhodes, S., Rolfe, S., Sharp, M., Vordermeier, H. M., Watson, E., Welsh, M., Whelan, A.O., Woolliams, J.A., More, S.J., Clifton-Hadley, R.S., 2017b. Systematic review to identify primary research estimating the performance of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in cattle. Unpublished, submitted to Prev. Vet. Med.

Dunn, J.R., Kaneene, J.B., Grooms, D.L., Bolin, S.R., Bolin, C.A., Bruning-Fann, C.S., 2005. Effects of positive results for *Mycobacterium avium* subsp paratuberculosis as determined by microbial culture of faeces or antibody ELISA on results of caudal fold tuberculin test and interferon-gamma assay for tuberculosis in cattle. J. Am. Vet. Med. Assoc. 226, 429-435.
Edwards, W., Lindman, H., Savage, L.J., 1963. Bayesian statistical inference in statistical research. Psychol. Res. 70, 193-242.

EEC (European Economic Community), 1964. Council Directive 64/432/EEC, of 26 June 1964 on animal health problems affecting intra-Community trade in bovine animals and swine. Official Journal 121, 1977-2012 (as amended).
EFSA (European Food Safety Authority), 2006. Scientific Opinion on "Performance of Brucellosis Diagnostic Methods for Bovines, Sheep, and Goats". The EFSA Journal. 432, 1-44.
http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/432.pdf (accessed 20 March 2016).

EFSA (European Food Safety Authority), 2008. Scientific Opinion of the Scientific Panel on Animal Health and Welfare on request from the European Commission on "Tuberculosis testing in deer". The EFSA Journal. 645, 1-34.
http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/645.pdf (accessed 20 March 2016).

EFSA (European Food Safety Authority), 2012. EFSA Panel on Animal Health and Welfare (AHAW); Scientific Opinion on the use of a gamma interferon test for the diagnosis of bovine tuberculosis. The EFSA Journal. 10, 2975.
http://www.efsa.europa.eu/en/efsajournal/pub/2975.htm (accessed 20 March 2016).
Farnham, M.W., Norby, B., Goldsmith, T.J., Wells, S.J., 2012. Meta-analysis of field studies on bovine tuberculosis skin tests in United States cattle herds. Prev. Vet. Med. 102, 234-242.

Faye, S., Moyen, J.L., Gares, H., Benet, J.J., Garin-Bastuji, B., Boschiroli, M.L., 2011. Determination of decisional cut-off values for the optimal diagnosis of bovine tuberculosis with a modified IFNγ assay (Bovigam®) in a low prevalence area in France. Vet. Microbiol. 151, 60-67.

Flores-Villalva, S., Suárez-Güemes, F., Espitia, C., Whelan, A.O., Vordermeier, M., Gutiérrez-Pabello, J.A., 2012. Specificity of the tuberculin skin test is modified by use of a

protein cocktail containing ESAT-6 and CFP-10 in cattle naturally infected with *Mycobacterium bovis.* Clin. Vaccine Immunol. 19, 797-803.

Godfray, H.C., Donnelly, C.A., Kao, R.R., Macdonald, D.W., McDonald, R.A., Petrokofsky, G., Wood, J.L., Woodroffe, R., Young, D.B., McLean, A.R., 2013. A restatement of the natural science evidence base relevant to the control of bovine tuberculosis in Great Britain. P. Roy. Soc. B- Biol. Sci. 280, 1634.

Goodchild, A.V., Downs, S.H., Upton, P., Wood, J.L.N. de la Rua-Domenech R., 2015. Specificity of the comparative skin test for bovine tuberculosis in Great Britain. Vet. Rec., 177, 258.

Greiner, M., Pfeiffer D., Smith, R.D., 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev. Vet. Med. 45., 23-41.

Greiner, M., Verloo D., de Massis F., 2009. Meta-analytical equivalence studies on diagnostic tests for bovine brucellosis allowing assessment of a test against a group of comparative tests. Prev. Vet. Med. 92, 373-381.

Harbord, R.M., Deeks, J.J., Egger, M., Whiting, P., Sterne, J.A., 2007. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics. 8, 239-251.

Hartnack, S., Torgerson, P.R., 2012. The accuracy of the Single Intradermal Comparative Skin Test for the diagnosis of Bovine Tuberculosis-estimated from a systematic literature search. J.Mycobac. Dis., 2, 120. DOI: 10.4172/2161-1068.1000120.

Irwig, L., Tosteson, A.N., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T.C., Mosteller, F., 1994. Guidelines for meta-analyses evaluating diagnostic tests. Ann. Intern. Med. 120, 667-676.

Irwig, L., Macaskill, P., Glasziou, P., Fahey, M., 1995. Meta-analytic methods for diagnostic test accuracy. J. Clin. Epidemiol. 48, 119-130.

Jones, C.M., Athanasiou T., 2005. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann. Thorac. Surg., 79, 16-20.

Jones, G.J., Whelan, A., Clifford, D., Coad, M., Vordermeier, H.M., 2012. Improved skin test for differential diagnosis of bovine tuberculosis by the addition of Rv3020c-derived peptides. Clin. Vaccine Immunol. 2012 Apr;19(4):620-2.

Karolemeas, K., de la Rua-Domenech, R., Cooper, R., Goodchild, A.V., Clifton-Hadley, R.S., Conlan, A.J., Mitchell, A.P., Hewinson, R.G., Donnelly, C.A., Wood, J.L., McKinley, T.J., 2012. Estimation of the relative sensitivity of the comparative tuberculin skin test in tuberculous cattle herds subjected to depopulation. PLoS One.7, e43217.

Leeflang, M.M.G., Deeks, J.J., Gatsonis C., Bossuyt P.M.M. on behalf of the Cochrane Diagnostic Test Accuracy Working Group., 2008. Systematic reviews of diagnostic test accuracy. Ann. Intern. Med. 149, 889-897.

Liebana, E., Aranaz, A., Mateos, A., Vilafranca, M., Gomez-Mampaso, E., Tercero, J.C., Alemany, J., Suarez, G., Domingo, M., Dominguez, L., 1995. Simple and rapid detection of *Mycobacterium tuberculosis* complex organisms in bovine tissue samples by PCR. J. Clin. Microbiol. 33, 33-36.

Lightbody, K. A., R. A. Skuce, S. D. Neill, and J. M. Pollock. 1998. Mycobacterial antigen-specific antibody responses in bovine tuberculosis: an ELISA with potential to confirm disease status. Vet. Rec. 142:295–300.

Lunn D.J., Thomas A., Best N., Spiegelhalter D.A., 2000. Bayesian modelling framework: Concepts, structure, and extensibility. Stat. Comput. 10, 325-337.

Lopes, L.B., Alves, T.M., Stynen, A.P.R., Mota, P.M.P.C., Leite, R.C., Lage, A.P., 2012. Parameter estimation and use of gamma interferon assay for the diagnosis of bovine tuberculosis in Brazil. Pesq. Vet. Bras. 32, 279-283.

Marassi, C.D., Medeiros, L., Lilenbaum, W., 2010. The use of a Gamma-Interferon assay to confirm a diagnosis of bovine tuberculosis in Brazil. Acta Trop. 113, 199-201.

Marassi, C.D., Zarden, C., Oelemann, W., Lilenbaum, W., 2014. Detection of specific antibodies in cows after injection of PPD. Pesq Vet Bras. 34, 241-243.

Meikle, V., Schneider, M., Azenzo, G., Zumarraga, M., Magnano, G., Cataldi, A., 2007. Individual Animals of a Cattle Herd Infected With the Same *Mycobacterium bovis* Genotype Shows Important Variations in Bacteriological, Histopathological and Immune Response Parameters. Zoonoses Public Hlth. 54, 86-93.

Midgette, A.S., Stukel, T.A., Littenberg, B., 1993. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Med. Decis. Making. 13, 253-257.

More S.J., Good M., 2015. Understanding and managing bTB risk: Perspectives from Ireland. Vet. Microbiol. 176, 209-218.

Moses, L.E., Shapiro, D., Littenberg, B., 1983. Combining independent studies of a diagnostic test into a summary ROC curve. Data-analytic approaches and some additional considerations. Stat. Med. 12, 1293-1316.

Moustakas, A., Evans, M.R., 2016. Regional and temporal characteristics of bovine tuberculosis of cattle in Great Britain. Stochastic Environmental Research and Risk Assessment 30, 989-1003.

Müller, B., Vounatsou, P., Ngandolo, B.N., Diguimbaye-Djaïbe, C., Schiller, I., Marg-Haufe, B., Oesch, B., Schelling, E., Zinsstag, J., 2009. Bayesian receiver operating characteristic estimation of multiple tests for diagnosis of bovine tuberculosis in Chadian cattle. PLoS One. 4, 12.

Neill, S.D., Cassidy, J., Hanna, J., Mackie, D.P., Pollock, J.M., Clements, A., Walton, E., Bryson, D.G., 1994. Detection of *Mycobacterium bovis* infection in skin test-negative cattle with an assay for bovine interferon-gamma. Vet. Rec. 135, 134-135.

O'Hagan, M.J., Courcier, E.A, Drewe, J.A., Gordon, A.W., McNair, J., Abernethy, D.A., 2015. Risk factors for visible lesions or positive laboratory tests in bovine tuberculosis reactor cattle in Northern Ireland. Prev. Vet. Med. 120, 283-290.

OIE (World Organisation for Animal Health), 2012. Terrestrial Manual 2012. Chapter 2.4.7 – Bovine Tuberculosis http://www.oie.int/fileadmin/Home/fr/Health_standards/tahm/2.04.07_BOVINE_TB.pdf (accessed 3 December 2015).

Pollock, J.M., Welsh, M.D., McNair, J., 2005. Immune responses in bovine tuberculosis: towards new strategies for the diagnosis and control of disease. Vet. Immunol. Immunop. 108, 37-43.

Praud, A., Boschiroli, M.L., Meyer, L., Garin-Bastuji, B., Dufour, B., 2015. Assessment of the sensitivity of the gamma-interferon test and the single intradermal comparative cervical test for the diagnosis of bovine tuberculosis under field conditions. Epidemiol. Infect. 143, 157-166.

Rosman, A.S., Korsten M.A., 2007. Application of summary receiver operating characteristics (sROC) analysis to diagnostic clinical testing. Adv. Med. Sci. 52, 76-82.

Schiller, I., Waters, W.R., Vordermeier, H.M., Nonnecke, B., Welsh, M., Keck, N., Whelan, A., Sigafoose, T., Stamm, C., Palmer, M., Thacker, T., Hardegger, R., Marg-Haufe, B., Raeber, A., Oesch, B., 2009. Optimization of a whole-blood gamma interferon assay for detection of *Mycobacterium bovis*-infected cattle. Clin. Vaccine Immunol. 16, 1196-1202.

Smith, N.H., Gordon, S.V., de la Rua-Domenech, R., Clifton-Hadley, R.S., Hewinson, R.G., 2006. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nat. Rev. Microbiol. 4, 670-681.

Spiegelhalter, D.J., Best N.G., Carlin B.P., Ven Der Linde A., 2002 Bayesian Measures of Model Complexity and Fit (with Discussion). J. Roy. Stat. B. 64, 583-616.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B., 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group.

Souza, I.I., Melo, E.S., Ramos, CA, Farias, T.A., Osório, A.L., Jorge, K.S., Vidal, CE, Silva, AS, Silva, MR, Pellegrin, AO, Araújo, FR., 2012. Screening of recombinant proteins as antigens in indirect ELISA for diagnosis of bovine tuberculosis. Springerplus. 1, 77.

Tameni, S., Amadori, M., Scaccaglia, P., Quondam-Gianndomenico, R., Tagliabue, S., Archetti, I., Adone, R., Ciuchini, E., 1998. Quality controls and in-vitro diagnostic efficiency of bovine PPD tuberculins. Biologicals. 26, 225-235.

TDR Diagnostics Evaluation Expert Panel, Banoo S., Bell D., Bossuyt P., Herring, A., Mabey, D., Poole, F., Smith, P.G., Sriram, N., Wongsrichanalai, C., O'Brien, T., Cunningham, J., Matsoso, P., Nathanson C.M., Olliaro, P., Peeling, R.W., Ramsay, A., 2010.

Evaluation of diagnostic tests for infectious diseases: general principles. Nat. Rev. Microbiol. 8(12 Suppl), S17-S29.

Thomas, A., O'Hara, B., Ligges, U., Sturtz, S., 2006. Making BUGS open. R. News. 6, 12-17.

Toft, N. Jorgensen, E., Hojsgaard, S., 2005. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev. Vet. Med. 68, 19-33.

VLA (Veterinary Laboratories Agency) 2010. Meta-analysis of diagnostic tests and modeling to identify appropriate testing strategies to reduce *M. bovis* infection GB herds. Final Report.
Welsh, M.D., Cunningham, R.T., Corbett, D.M., Girvin, R.M., McNair, J., Skuce, R.A., Bryson, D.G., Pollock, J.M., 2005. Influence of pathological progression on the balance between cellular and humoral immune responses in bovine tuberculosis. Immunology.114, 101-111.

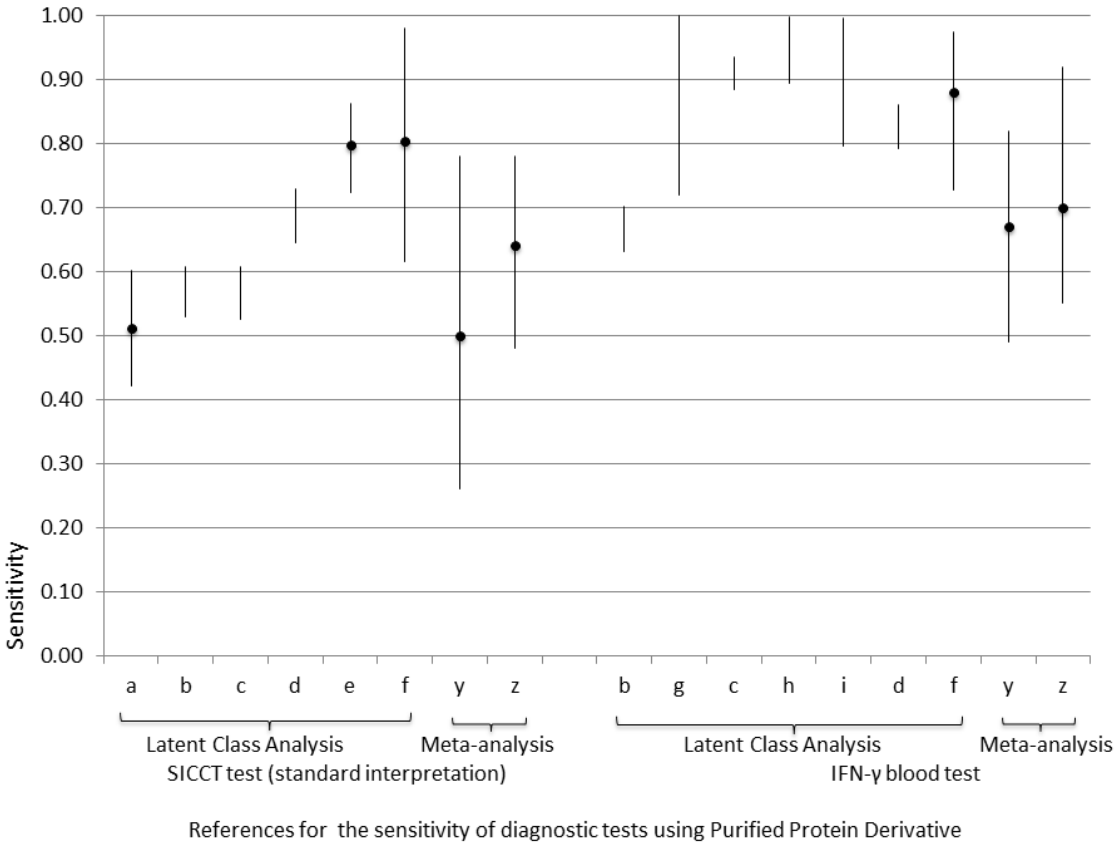
Whelan, A.O., Clifford, D., Upadhyay, B., Breadon, E.L., McNair, J., Hewinson, G.R., Vordermeier, M.H., 2010. Development of a skin test for bovine tuberculosis for differentiating infected from vaccinated animals. J. Clin. Microbiol. 48, 3176-3181.

Whiting, P., Harbord, R., Kleijnen, J., 2005. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med. Res. Methodol. 5, 19.

Zanini, M., Moreira, E.C., Lopes, M.T.P., Oliveira, R.S., Leao, S.C., Fioravanti, R.L., Roxo, E., Zumarraga, M., Romano, M.I., Cataldi, A., Salas, C.E., 2001. *Mycobacterium bovis*: Polymerase chain reaction identification in bovine lymphonode biopsies and genotyping in isolates from southeast Brazil by spolygotyping and restriction fragment length polymorphism. Mem. I. Os. Cr. 96, 809-813.

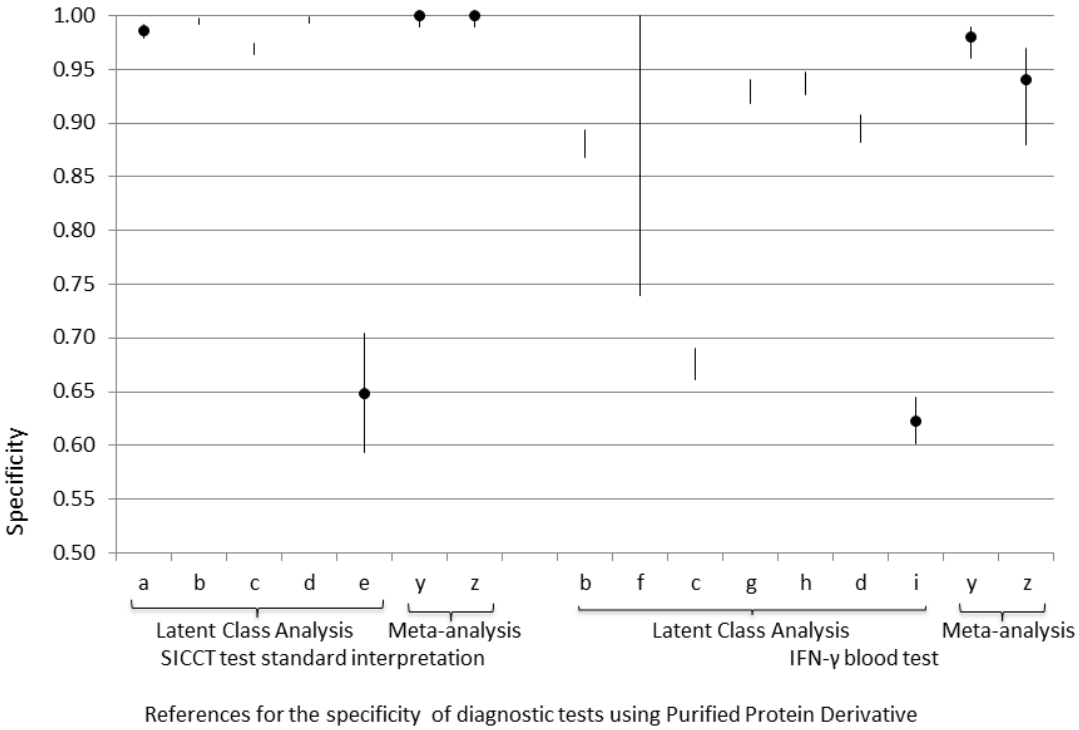
Zumarraga, M.J., Paolicchi, F., Garbaccio, S., Gioffre, A., Caimi, K., Bigi, F., Alito, A., Romano, M.I., Cataldi, A., 2002. Application of the PCR for the Detection of *Mycobacterium bovis* in Tissue Samples. Rev. Med. Vet. Agent. 83, 139-141.

Fig. 1. Sensitivity estimated using Latent Class Analysis in studies published 2009-2015 and from the meta-analysis



Footnote to Fig. 1.
a Muller et al., 2009 (Chad) 95% Confidence Interval
b Clegg et al., 2011 (Republic of Ireland) 95% Credible Intervals
c EFSA, 2012 (Northern Ireland) 95% Credible Intervals
d EFSA, 2012 (Republic of Ireland) 95% Credible Intervals
e Hartnack et al., 2012 (England) 95% Confidence Intervals
f Praud et al., 2015 (France) 95% Credible Intervals
g Lopes et al., 2012 (Brazil) 95% Confidence Intervals
y Meta-analysis, modelling procedure A 95% Credible Intervals
z Meta-analysis, modelling procedure B 95% Credible Intervals

Fig. 2. Specificity estimated using Latent Class Analysis in studies published 2009-2015 and from the meta-analysis



Footnote to Fig. 2.
a Muller et al., 2009 (Chad) 95% Confidence Intervals
b Clegg et al., 2011 (Republic of Ireland) 95% Credible Intervals
c EFSA, 2012 (Northern Ireland) 95% Credible Intervals
d EFSA, 2012 (Republic of Ireland) 95% Credible Intervals
e Hartnack et al., 2012 (England) 95% Confidence Intervals
f Lopes et al., 2012 (Brazil) 95% Confidence Intervals
g Praud et al., 2015 (France) 95% Credible Intervals
y Meta-analysis, modelling procedure A 95% Credible Intervals
z Meta-analysis, modelling procedure B 95% Credible Intervals

Table 1 Variables included in models for estimation of sensitivity

Test Name	Ref. test	Sampling strategy	World region	Pub. Year	Prod. class	PPD manu.	Skin test interp.	Skin test PPD	Diagnostic antigens	Culture media	Post mortem	Sp counter parameter	SICCT test	Rev. no.	Lang.	Time since SICCT test
Skin tests with PPD																
Single intradermal skin	N/N	A/N	N/N	R	R	N/B	A	N/N	NA	NA	R	NA	R	R	N/N	NA
SICCT test severe	A/B	N/N	N/N	A/N	N/N	A/N	N/B	N/N	NA	NA	R	NA	R	N/N	A/N	NA
SICCT test standard	A/B	N/N	N/N	A/N	N/N	A	A/B	N	NA	NA	R	NA	R	A/N	N/N	NA
Caudal fold	A/N	A/N	R	R	R	R	R	A/B	NA	NA	R	NA	R	R	R	NA
IFN-γ blood tests																
IFN-γ Bovine PPD	A/B	A/N	A/B	N/N	N/N	NA	NA	NA	A/B	NA	NA	A/B	A/N	R	R	A/N
IFN-γ Bovine PPD-Avian PPD	A/B	A/B	N/N	N/N	N/N	NA	NA	NA	A/B	NA	NA	A/B	N/N	R	R	A/N
IFN-γ ESAT6/CFP10	A/B	A/B	A/N	N/N	N/N	NA	NA	NA	A/B	NA	NA	A/N	N/B	A/N	R	A/N
IFN-γ MPB70	A/N	A/N	N/N	N/N	A/B	NA	NA	NA	A/B	NA	NA	A/B	A/B	R	R	A/N
Antibody detection tests																
ELISA Bovine PPD	A/B	A/N	A/N	N/N	A/N	NA	NA	NA	A/B	NA	NA	N/B	A/N	R	R	R
ELISA Bovine PPD-Avian PPD	A/B	A/N	A/N	N	A/N	NA	NA	NA	A/B	NA	NA	N/B	A/N	R	R	R
ELISA MPB70	A/B	A/N	A/N	N	A/N	NA	NA	NA	A/B	NA	NA	N/B	A/N	R	R	R
LBAA	X	X	X	X	X	NA	NA	NA	NA	NA	X	X	X	X	X	X
Multiplex immunoassay	X	X	X	X	X	NA	NA	NA	NA	NA	X	X	X	X	X	NA
Glutaraldehyde	X	X	X	X	X	NA	NA	NA	NA	NA	X	X	X	X	X	X
Post mortem																
Meat inspection	N/N	A/B	N/N	N/N	R	NA	NA	NA	NA	NA	A/B	NA	A/B	R	A/B	NA
Detailed necropsy	N/N	A/B	N/N	N/N	N/N	NA	NA	NA	NA	NA	A/B	NA	A/B	R	A/B	NA
Microscopic examination	N/N	R	A/N	R	NA	NA	NA	NA	NA	NA	A/N	NA	A/B	R	A/N	NA
Culture	A/B	R	N	A/B	R	NA	NA	NA	NA	A/B	A/B	NA	A/B	R	A/B	NA
PCR	A/B	A/N	A/N	A/N	NA	NA	NA	NA	NA	NA	A/N	NA	R	R	N/N	NA

Footnote to Table 1

Ref test: reference test; Lang: Language; PPD Manu: PPD manufacturer; Skin Test Interp.: Skin Test Interpretation; LBAA: Latex bead agglutination assay; PCR: Polymerase Chain Reaction; A/B: covariate significant for model A and model B, A: covariate significant for model A, B: covariate significant for model B, N: covariates not included in the final model due to: for model A it did not show association in the preliminary analysis and for model B it was not significant; NA: covariate not applicable to the test-type; R: covariate removed before the modelling due to low variability; X: non included in the model since a FEM with no covariates was used due to scarce data. See Table 3 for further explanation of variables.

Table 2 Variables included in models for estimation of specificity

Test Name	Ref. Test	Sampling strategy	World region	Pub. Year	Prod. class	PPD manu.	Skin test interp.	Skin test PPD	Diagnostic antigens	Cross reactivity	Freedom from infection	Se counter parameter
Skin tests with PPD												
Single intradermal test	X	X	X	X	X	X	X	X	NA	X	X	NA
SICCT test	X	X	X	X	X	X	X	X	NA	X	X	NA
Caudal fold	X	X	X	X	X	NA	NA	NA	NA	X	X	NA
IFN-γ blood tests												
IFN-γ Bovine PPD	A	A	N/N	N/N	N/N	NA	NA	NA	A/B	N/N	B	A/B
IFN-γ Bovine PPD-Avian PPD	A	A	A	N/N	N/N	NA	NA	NA	B	N/N	B	A/B
IFN-γ ESAT6/CFP10	A	A	A	N/N	N/N	NA	NA	NA	A/B	N/N	B	A/B
IFN-γ MPB70	A	A	A	A	N/N	NA	NA	NA	A/B	N/N	B	A/B
Antibody detection tests												
ELISA Bovine PPD	N/N	R	N/N	N/N	R	NA	NA	NA	A/B	A/B	N/N	A/B
ELISA Bovine PPD-Avian PDD	X	A	X	X	X	NA	NA	NA	B	A/B	X	A/B
LBAA	X	X	X	X	X	NA	NA	NA	NA	X	X	NA
Multiplex immunoassay	X	X	X	X	X	NA	NA	NA	X	X	X	NA
Serological Rapid	X	X	X	X	X	NA	NA	NA	NA	X	X	NA
Glutaraldehyde	X	X	X	X	X	NA	NA	NA	NA	X	X	NA
IFA	X	X	X	X	X	NA	NA	NA	NA	X	X	NA
Post Mortem												
Necropsy	X	X	X	X	X	NA	NA	NA	NA	X	X	NA
Microscopic examination	X	X	X	X	X	NA	NA	NA	NA	X	X	NA

PCR	X	X	X	X	X	NA	NA	NA	NA	X	X	NA
Culture	X	X	X	X	X	NA	NA	NA	NA	X	X	NA

Footnote to Table 2.
Ref test: reference test; Lang: Language; PPD Manu: PPD manufacturer; Skin Test Interp.: Skin Test Interpretation; LBAA: Latex bead agglutination assay; IFA: Indirect Fluorescent Antibody Test; PCR: Polymerase Chain Reaction; A/B: covariate significant for model A and model B, A: covariate significant for model A, B: covariate significant for model B, N: covariates non included in the final model due to: for model A it did not show association in the analysis and for model B it was not significant; NA: covariate not applicable to the test-type; R: covariate removed before the modelling due to low variability; X: not included in the model since a Fixed Effects Model (FEM) with no covariates was used due to scarce data. See Table 3 for further explanation of variables.

Table 3 Covariates with description and binary categories in regression modelling

Covariate	Description	UK & Ireland category (baseline category, set to 1)	Other complementary category set to 0
Cross reactivity	Cross reactivity with other mycobacteria probable	Yes	No or unknown
Culture Media	Used to isolate *M. bovis*	Lowenstein-Jensen-Middlebrook	Other or unknown
Diagnostic antigen	Diagnostic antigens in IFN-γ or ELISA blood tests	Bovine PPD	Other diagnostic antigen
Diagnostic antigen	Diagnostic antigens in IFN-γ or ELISA blood tests	Bovine PPD-Avian PPD	Other diagnostic antigen
Diagnostic antigen	Diagnostic antigens in IFN-γ or ELISA blood tests	ESAT6/CFP10	Other diagnostic antigen
Freedom from Infection	Level of evidence that population sample free from infection (for estimation of specificity)	From Officially Tuberculosis free herd, country or region	Other evidence cattle are free from *M. bovis* infection
Language	Language in which the reference was published	English	Other
Post mortem	Diagnostic test conducted after a post mortem	Yes	No
Production class	Cattle production class	Dairy	Non-Dairy
PPD Manufacturer	Name of manufacturer of purified protein derivative given	Yes	No
Publication year	Year reference was published	2000-2010	Before 2000
Reference Test	Reference test (for estimating specifity)	Necropsy or culture	Other
Reviewer Number	Number of reviewers who reviewed and extracted data from the reference	2	1
Sampling strategy	Sampling method to select cattle population	Probability samping frame (random/census)	
Counter parameter	Estimate of Sp extracted from reference when estimating Se and estimate of Se extracted from reference when estimating Sp . In Random Effects Models only		
SICCT test	Selection of population sample using SICCT test (for estimating Se)	No selection using SICCT test	All animals in population sample are SICCT positive
Skin test interpretation	Skin test wheal size 4mm or more	Standard	Other
Skin test interpretation	Skin test wheal size of 2mm or more	Severe	Other
Time since skin test	Conduct of index blood test in relation to tuberculin skin test	Before skin test	Other
World region	Location of test	EUUKI	Other

Footnote to Table 3

Variables selected a-priori and included in one or more regression models.

Table 4 Meta-analysis results for sensitivity of diagnostic tests for bovine tuberculosis in cattle

Test Name	References	Sensitivity Estimates	Pooled Sensitivity	Models A & B	Model A			Model B		
	n	N	Estimate	Type	Median	95% Credible Interval	DIC	Median	95% Credible Interval	DIC
Skin tests with PPD										
Single intradermal skin test	7	16	0.92	REM	0.94	0.49, 1.0	88	0.81	0.53, 0.94	85
SICCT test standard interpretation	14	38	0.78	REM	0.50	0.26, 0.78	159	0.64	0.48, 0.78	156
SICCT test severe interpretation	14	38	0.84	REM	0.63	0.40, 0.84	183	0.75	0.61, 0.86	183
Caudal fold	15	69	0.92	REM	0.76	0.56, 0.89	405	0.96	0.88, 0.98	403
IFN-y blood tests										
IFN-γ Bovine	27	166	0.84	REM	0.87	0.72, 0.95	796	0.87	0.76, 0.94	790
IFN-γ Bovine PPD- Avian PPD	27	166	0.83	REM	0.67	0.49, 0.82	845	0.70	0.55, 0.92	822
IFN-γ ESAT6/CFP10	27	166	0.84	REM	0.78	0.60, 0.90	842	0.79	0.64, 0.89	816
IFN-γ MPB70	27	166	0.42	REM	0.10	0.03, 0.25	800	0.16	0.08, 0.30	728
Antibody detection blood tests										
ELISA Bovine	22	59	0.62	REM	0.76	0.06, 0.99	362	0.79	0.52, 0.93	328
ELISA Bovine PPD-Avian PPD	22	59	0.44	REM	0.36	0.01, 0.97	351	0.60	0.31, 0.86	329
ELISA MPB70	22	59	0.80	REM	0.20	0.01, 0.94	337	0.25	0.05, 0.66	295
Latex bead agglutination assay	2	3	0.91	FEM	0.91	0.60, 0.98	[a]	[b]	[b]	[b]
Multiplex immunoassay	1	5	0.75	FEM	0.74	0.31, 0.95	[a]	[b]	[b]	[b]
Glutaraldehyde	1	1	0.85	FEM	0.84	0.42, 0.98	[a]	[b]	[b]	[b]
Post Mortem										
Necropsy (detailed/laboratory)	6	11	0.79	REM	0.96	0.82, 1.00	39	0.96	0.82, 1.00	39
Necropsy (meat inspection)	6	11	0.46	REM	0.71	0.37, 0.92	39	0.71	0.37, 0.92	39
Microscopic examination	13	21	0.74	REM	0.63	0.15, 0.93	159	0.66	0.41, 0.84	158
Culture	8	16	0.55	REM	0.74	0.46, 0.94	101	0.74	0.46, 0.94	101

PCR	12	25	0.51	REM	0.14	0.0, 0.98	129	0.86	0.65, 0.96	128

Footnote for Table 4.

Estimates are at the animal level. REM: Random Effects Model, FEM: Fixed Effect Model. FEM were fitted for tests with a very sparse data set. For these tests covariates were not included in the model and models A and B are identical. DIC: Deviance Information Criterion for comparing models. Severe interpretation: Reaction to Bovine tuberculin is positive and the reaction to Avian tuberculin is negative or animals show a positive Bovine reaction more than 2 mm, greater than a positive Avian reaction. Standard interpretation in GB: Reaction to Bovine tuberculin is both positive and exceeds the reaction to Avian tuberculin by more than 4 mm. [a]: Not Applicable. [b]: Model B was only performed for test types that had a REM successfully applied.

Table 5 Meta-analysis results for specificity of diagnostic tests for bovine tuberculosis in cattle

Test Name	References	Specificity Estimates	Pooled Specificity	Models A & B	Model A			Model B		
	n	N	Estimate	Type	Median	95% Credible Interval	DIC	Median	95% Credible Interval	DIC
Skin tests										
Single Intradermal skin test	4	10	0.89	REM	0.91	0.70, 1.00	29	0.91	0.63, 1.00	30
SICCT test standard	7	13	1.00	REM	1.00	0.99, 1.00	[a]	1.00	0.99, 1.00	[a]
Caudal fold	2	3	0.99	FEM	1.00	0.92, 1.00	[a]	[b]	[b]	[b]
IFN-γ blood tests										
IFN-γ Bovine PPD	19	137	0.91	REM	0.97	0.94, 0.98	647	0.94	0.86, 0.98	649
IFN-γ Bovine PPD-Avian PPD	19	137	0.96	REM	0.98	0.96, 0.99	645	0.94	0.88, 0.97	621
IFN-γ ESAT6/CFP10	19	137	0.98	REM	0.99	0.99, 1.00	597	0.99	0.98, 1.00	601
IFN-γ MPB70	19	137	0.93	REM	0.94	0.85, 0.98	639	0.91	0.81, 0.96	641
Antibody detection blood tests										
ELISA Bovine PPD	12	27	0.96	REM	0.89	0.80, 0.94	142	0.89	0.80, 0.94	142
ELISA Bovine PPD-Avian PPD	12	27	0.98	REM	0.93	0.84, 0.98	141	0.93	0.84, 0.97	141
LBAA	1	1	1.00	FEM	0.94	0.39, 1.00	[a]	[b]	[b]	[b]
Multiplex immunoassay	1	4	0.88	FEM	0.88	0.34, 0.99	[a]	[b]	[b]	[b]
Serological Rapid	2	3	0.97	FEM	0.97	0.66, 1.00	[a]	[b]	[b]	[b]
Post Mortem										
Necropsy	1	3	1.00	FEM	1.00	0.99, 1.00	[a]	[b]	[b]	[b]
Culture	1	1	1.00	FEM	0.99	0.73, 1.00	[a]	[b]	[b]	[b]
Microscopic examination	1	1	1.00	FEM	1.00	0.95, 1.00	[a]	[b]	[b]	[b]
PCR	4	5	1.00	FEM	1.00	1.00, 1.00	[a]	[b]	[b]	[b]

Footnote for Table 5.

Estimates are at the animal level. REM: Random Effects Model, FEM: Fixed Effect Model. FEM were fitted for tests with a very sparse data set. For these tests covariates were not included in the model and models A and B are identical. DIC: Deviance Information Criterion for comparing models. Standard interpretation in GB: Reaction to bovine tuberculin is both positive and exceeds the reaction to Avian tuberculin by more than 4 mm. [a]: Not Applicable. [b]: Model B was only performed for test types that had a REM successfully applied.

Table 6 Sensitivity reported for skin tests, IFN-γ and ELISA blood tests 2010-2015

Reference	Diagnostic reagents	Reference Standard (bTB infection positive)	bTB infected cattle (n)	Sensitivity
Skin tests				
SIT - severe interpretation				
Casal 2014	Bovine PPD[a]	Visible lesions and/or positive culture	27[b]	0.59
SICCT test - severe interpretation				
Karolemeleas 2012	Bovine PPD-Avian PPD[c]	Visible lesions and/or positive culture	[d]	0.85[e]
Casal 2014	Bovine PPD-Avian PPD[c]	Visible lesions and/or positive culture	27[b]	0.37
SICCT test - standard interpretation				
Karolemeleas 2012	Bovine PPD-Avian PPD[f]	Visible lesions and/or culture positive	[d]	0.81[e]
Defined antigen skin test				
Whelan 2010	ESAT6, CFP10, MPB70, MPB83[g]	Positive SICCT test[f]	34	0.74
Flores-Villalva 2012	ESAT6/CFP10[h]	Positive SICCT test[f]	63	0.76
Jones 2012	ESAT6/CFP10+Rv3615c[h]	Visible lesions with or without positive culture	16	0.75
Jones 2012	ESAT6/CFP10 + Rv3615c+RV3020c[h]	Visible lesions with or without positive culture	16	0.88
IFN-γ blood test				
Purified protein derivative IFN-γ tests				
Jones 2010	Bovine PPD[i]	Positive SICCT test[f]	35	0.80
Schiller 2009	Bovine PPD - Avian PPD[j]	Positive culture and/or lesions	431	0.91
Marassi 2010	Bovine PPD - Avian PPD[j]	Positive SICCT test[f] (with or without) positive culture or	21	0.95

		PCR		
Antognoli 2011	Bovine PPD - Avian PPD[j]	Positive histopathology and PCR or positive culture	87	0.84
Faye 2011	Bovine PPD - Avian PPD[k]	Visible lesions with or without positive culture and/or PCR and/or histology	56	0.93
Flores-Villalva 2012	Bovine PPD - Avian PPD[l]	Positive SICCT test[f]	63	0.89
Casal 2014	Bovine PPD - Avian PPD[m]	Visible lesions or positive culture	32	0.88
Defined antigen IFN-γ blood tests				
Jones 2010	ESAT6/CFP10[n]	Positive SICCT test[f]	35	0.69
Faye 2011	ESAT6/CFP10[o]	Visible lesions with or without positive culture and/or PCR and/or histology	58	0.97
Flores-Villalva 2012	ESAT6/CFP10[p]	Positive SICCT test[f]	63	0.70
Antibody Detection blood tests (ELISA)				
Marassi 2014	MPB70[q]	Positive SICCT test[f]	9	0.11
Marassi 2014	MPB83[r]	Positive SICCT test[f]	9	0.67
Buddle 2013	MPB70 and MPB83[s]	Visible lesions or positive culture	38	0.53
Casal 2014	MPB70 and MPB83[t]	Visible lesions and/or positive culture	27[b]	0.70
Casal 2014	MPB70, MPB83, ESAT6, CFP-10, PPD bovine, MPB70[t]	Visible lesions and/or positive culture	27[b]	0.85
Silva 2011	MPT-51[u]	Positive SICCT test[f]	208	0.55
Silva 2011	Ag8[5v]	Positive SICCT test[f]	208	0.48
Silva 2011	BCG[w]	Positive SICCT test[f]	208	0.82

Footnote
Positive cut-off:

[a]: Change in skin fold thickness in response to Bovine PPD >2mm or clinical signs
[b]: Herd a in Casal et al. 2014
[c]: Change in skin fold thickness in response to Bovine PDD- response to Avian PPD > 2mm and/or clinical signs
[d]: Not reported in Karolemeleas et al. 2012; approx 500 cattle from 16 herds based on source data from Animal and Plant Health Agency
[e]: Estimated from logistic regression modelling
[f]: Change in skin fold thickness in response to Bovine PDD- response to Avian PPD > 4mm
[g]: (Optical Density (OD) Bovine PPD - OD Avian PPD)/(positive control OD - negative control OD)>0.1
[h]: Change in skin fold thickness $\geq$ 1mm above response in negative control
[i]: Response to Bovine PPD - response to negative control>8.33ng/ml
[j]: (OD bovine PPD - OD negative control) AND (Bovine PPD - OD Avian PPD)$\geq$0.1
[k]: (OD Bovine PPD - OD Avian PPD)/(positive control OD - negative control OD)$\geq$0.02
[l]: (OD Bovine PPD- OD Avian PPD)/(OD negative control)$\geq$2
[m]:(OD sample - OD negative control$\geq$0.05) AND ((OD Bovine PPD-negative control)>(OD Avian - negative control))
[n]:Response to ESAT6/CFP10- Response to negative control>8.33ng/ml
[o]: (OD ESAT6/CFP10 - OD negative control)/(positive control OD - negative control OD)$\geq$0.01
[p]: (OD ESAT6/CFP10 - OD negative control)/(positive control OD - negative control OD)$\geq$2
[q]: OD cut off=0.21
[r]: OD cut off=0.25
[s]:(OD sample - OD negative control)/(positive control - negative control)$\geq$ 3
[t]: Sample responses to 2 antigens above the individual's threshold
[u]:OD cut off=1.301
[v]:OD cut off=0.898
[w]:OD cut off=1.287

Table 7 Specificity reported for skin tests, IFN-γ and ELISA blood tests 2010-2015

Reference	Diagnostic reagent	Reference Standard (bTB infection negative)	Uninfected cattle	Specificity
Skin tests				
SICCT test - severe interpretation				
Goodchild 2015	Bovine PPD - Avian PPD[a]	OTF herds>8 km from herds with OTFW	1,379,634	1.00
SICCT test - standard interpretation				
Goodchild 2015	Bovine PPD - Avian PPD[b]	OTF herds>8 km from herds with OTFW	1,379,634	1.00
Defined antigen skin test				
Whelan 2010	ESAT6, CFP10, MPB70, MPB83[c]	bTB free farms	14	1.00
Flores-Villalva 2012	ESAT6/CFP10[c]	bTB free herd	10	0.90
Jones 2012	ESAT6/CFP10+Rv3615c[c]		7	1.00
Jones 2012	ESAT6/CFP10 + Rv3615c+RV3020c[c]		7	1.00
IFN-γ blood test				
Purified protein derivative IFN-γ tests				
Antognoli 2011	Bovine PPD - Avian PPD[d]	Acredited bTB free	4123	0.91

		herds		
Faye 2011	Bovine PPD - Avian PPD[e]	Herds bTB free>10 years and from bTB free region	482	0.98
Defined antigen IFN-γ blood tests				
Faye 2011	ESAT6/CFP10[f]	Herds bTB free>10 years and from bTB free region	472	0.96
Antibody detection blood tests (ELISA)				
Casal 2014	MPB70 and MPB83[g]	Officially Tuberculosis Free > 4 years	60	1.00
Casal 2014	MPB70, MPB83, ESAT6, CFP-10, PPD bovine, MPB70[h]	Officially Tuberculosis Free > 4 years	60	0.98
Silva 2012	MPT-51[i]	bTB free herds from an area bTB is endemic	54	0.52
Silva 2012	AG85[j]	bTB free herds from an area bTB is endemic	54	0.89
Silva 2012	BCG[k]	bTB free herds from an area bTB is endemic	54	0.91

Positive cut-off:

[a]: Change in skin fold thickness in response to Bovine PDD- response to Avian PPD > 2mm

[b]: Change in skin fold thickness in response to Bovine PDD- response to Avian PPD > 4mm

[c]: Change in skin fold thickness > 1mm above response in negative control

[d]: (Optical Density (OD) bovine PPD - OD negative control) AND (Bovine PPD - OD Avian PPD)$\geq$0.1

[e]: (OD Bovine PPD - OD Avian PPD)/(positive control OD - negative control OD)$\geq$0.02

[f]: (OD ESAT6/CFP10 - OD negative control)/(positive control OD - negative control OD)≥0.1
[g]:(OD sample - OD negative control)/(positive control - negative control)≥ 3
[h]: Sample responses to 4 antigens above the individual's threshold
[i]:OD cut off=1.301
[j]:OD cut off=0.898
[k]:OD cut off=1.287