Evaluating diagnostic tests for bovine tuberculosis in the southern part of Germany: A latent class analysis

Valerie-Beau Pucken[1]*, **Gabriela Knubben-Schweizer[1]**, **Dörte Döpfer[2]**, **Andreas Groll[3]**, **Angela Hafner-Marx[4]**, **Stefan Hörmansdorfer[4]**, **Carola Sauter-Louis[5]**, **Reinhard K. Straubinger[6]**, **Pia Zimmermann[4]**, **Sonja Hartnack[7]**

1 Clinic for Ruminants with Ambulatory and Herd Health Services, Centre for Clinical Veterinary Medicine, LMU Munich, Oberschleißheim, Germany, **2** Food Animal Production Medicine, School of Veterinary Medicine, University of Wisconsin, Madison, Wisconsin, United States of America, **3** Workgroup Financial Mathematics, Department of Mathematics, LMU Munich, Munich, Germany, **4** Bavarian Health and Food Safety Authority, Oberschleißheim, Germany, **5** Institute for Epidemiology, Friedrich-Loeffler-Institute, Greifswald, Germany, **6** Institute for Infectious Diseases and Zoonoses, Department of Veterinary Science, LMU Munich, Munich, Germany, **7** Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland

* valerie.pucken@vetsuisse.unibe.ch

🔓 OPEN ACCESS

Abstract

Germany has been officially free of bovine tuberculosis since 1996. However, in the last years there has been an increase of bovine tuberculosis cases, particularly in the southern part of Germany, in the Allgäu region. As a consequence a one-time tuberculosis surveillance program was revisited with different premortal and postmortal tests. The aim of this paper was to estimate diagnostic sensitivities and specificities of the different tests used within this surveillance program. In the absence of a perfect test with 100% sensitivity and 100% specificity, thus in the absence of a gold standard, a Bayesian latent class approach with two different datasets was performed. The first dataset included 389 animals, tested with single intra-dermal comparative cervical tuberculin (SICCT) test, PCR and pathology; the second dataset contained 175 animals, tested with single intra-dermal cervical tuberculin (SICT) test, Bovigam® assay, pathology and culture. Two-way conditional dependencies were considered within the models. Additionally, inter-laboratory agreement (five officially approved laboratories) of the Bovigam® assay was assessed with Cohen's kappa test (21 blood samples). The results are given in posterior means and 95% credibility intervals. The specificities of the SICT test, SICCT test, PCR and pathology ranged between 75.8% [68.8–82.2%] and 99.0% [96.8–100%]. The Bovigam® assay stood out with a very low specificity (6.9% [3.6–11.1%]), though it had the highest sensitivity (95.7% [91.3–99.2%]). The sensitivities of the SICCT test, PCR, SICT test, pathology and culture varied from 57.8% [48.0–67.6%] to 88.9% [65.5–99.7%]. The prevalences were 19.8% [14.6–26.5%] (three-test dataset) and 7.7% [4.2–12.3%] (four-test dataset). Among all pairwise comparisons the highest agreement was 0.62 [0.15–1]). In conclusion, the specificity of the Bovigam® assay and the inter-laboratory agreement were lower than expected.

Introduction

Bovine tuberculosis (bTB) which is caused by *Mycobacterium caprae* and *Mycobacterium bovis* is an important public and animal health problem and an international trade issue in Europe and worldwide [1–6]. Therefore, using reliable, fast and cost-effective diagnostic methods is essential for the control of bTB.

National control programs rely on testing of cattle and removal of animals which are diagnosed as positive [7]. Infected animals are detected with tuberculin skin tests or the Bovigam® gamma-interferon (IFN-γ) assay. The tuberculin skin test, the prescribed test for international trade, is used as a single intradermal cervical tuberculin (SICT) test or single intra-dermal comparative cervical tuberculin (SICCT) test in Europe and as caudal fold tuberculin (CFT) test in North America, Australia and New Zealand [7, 8]. Accuracy of the skin tests varies widely due to different factors, affecting the host and the test itself. The exact estimation of the test characteristics in the field is therefore difficult [8–10]. However, the intradermal skin tests normally have a high specificity although sensitivity can be somewhat lower [9]. The low test sensitivity and the logistics of holding cattle for 3 days to read the test has led to the development of the Bovigam® assay, in 1985 [11]. Australia was the first country to officially accept this test for the diagnosis of bovine tuberculosis in 1991 [11]. In comparison to the skin test the Bovigam® assay almost always showed a better sensitivity but an equal or inferior specificity. The Bovigam® assay is supposed to have the ability to detect bTB earlier in the course of infection [11, 12]. In many countries it is used for serial or parallel testing together with the intradermal skin tests [7, 13]. For post-mortem diagnosis of previously positive-tested animals, bacteriological culture and PCR can be used following necropsy [1, 14, 15].

In many countries the test-and-cull regime led to the status Officially Bovine Tuberculosis free (OTF) [9, 16, 17]. Germany reached this status in 1997 [18]. Owing to the OTF status nationwide periodic surveillance using intradermal skin testing was replaced with surveillance by official meat inspection at the abattoir [19, 20]. Remarkably, there is an apparent increase of bTB cases since 2007, many of them detected during routine meat inspections and with a particular high prevalence in the southern part of Germany. These unexpected bTB cases led to a revision of the tuberculosis regulations in 2009, 2012 and 2013 with commencement of the act in 2009, 2013 and 2014. Within this revision the Bovigam® assay and the PCR were added as new diagnostic methods. Furthermore, the increase of bTB cases lead to the implementation of a one-time tuberculosis surveillance program in Germany in 2012 to verify the nation's OTF status [21]. At the beginning of this surveillance program the SICCT test or a serial testing with the SICT test and the Bovigam® assay were used for ante-mortem diagnosis [22]. This was the first time that the Bovigam® assay was used as a field test in Germany. However, as a consequence of irregular test results, the testing regime was changed in March 2013 with the SICCT test as the only ante-mortem test. Moreover, the PCR analysis as described in the national *Official Collection of Methods* was used since then as additional post-mortem method [1, 22–24].

As described in the literature the sensitivities and specificities of bTB diagnostic tests vary widely [12, 13, 25, 26]. This leads to difficulties in identifying truly infected animals as well as in identifying risk factors for bTB [27]. Diagnostic accuracies of bTB diagnostic tests are often estimated using bacteriological culture as the so-called gold standard for confirmation of bTB [26, 28–30]. A gold standard is considered as a test that has known properties with a high sensitivity and specificity. Because bacteriological culture has limitations in sensitivity this may lead to a misclassification of data [13, 31]. By using a latent class approach the test characteristics can be assessed in the absence of a gold standard [32–34]. This latent class approach can be used within a Bayesian model and is based on multiple tests performed on the same animals.

"Latent class" refers to the fact that the true disease state is always hidden [35]. In the *Standard Operating Procedure for OIE Registration of Diagnostic Kits* the Bayesian inference and latent class models are described to use for estimation of diagnostic sensitivities and specificities [36].

The aim of this study was to assess the diagnostic accuracies of the tests used within the bTB surveillance program in Germany between 2010 and 2014. To the best of our knowledge a latent class analysis for the diagnostic tests of bTB has never been applied in Germany.

Material and methods

With the aim to obtain robust diagnostic test accuracy estimates for different pre- and post-mortem tests diagnosing bovine tuberculosis, a Bayesian latent class approach was performed. Regarding the Bovigam® assay agreement between blood samples tested by five different laboratories and between blood samples taken from two different anatomical locations was assessed with Cohen's kappa coefficient.

Ethics statement

The test results which were used for the Bayesian latent class approach were collected within the context of the officially ordered tuberculosis-surveillance program ("Untersuchungsprogramm: Rindertuberkulose in den Landkreisen der Alpenkette; AZ: 42a-G8755-2013/2-450) prior to this study and were not specifically taken for the purpose of this study. The program was conducted according to Directive 64/432/EEC on animal health problems affecting intra-EU trade in bovine animals and swine, Council Directive 80/219/EEC of 22 January 1980 amending Directive 64/432/EEC as regards tuberculosis and brucellosis and Council Directive 97/12/EC of 17 March 1997 amending and updating Directive 64/432/EEC on health problems affecting intra-Community trade in bovine animals and swine [37–39] to verify the OTF status.

The data used for the Bovigam® assay agreement existed prior to our research. The blood samples were taken in the context of the study "Optimierung der Methode Bovigam®—Test beim Rind—vergleichende Untersuchungen an 21 Tieren des Betriebs Spitalhof, Kempten" which was carried out by the Institute for Infectious Diseases and Zoonoses, Department of Veterinary Science, LMU Munich under the direction of Prof. Dr. med. vet. Reinhard K. Straubinger, Ph.D and were approved by the government of Upper Bavaria (approving authority for animal research). According to the approval no. 5.2-1-54-2532.3-26-13 there is no conflict with animal protection law.

Bayesian latent class models

Animal samples. Out of 5736 animals tested between 2012 and 2014 in the districts Ober- and Ostallgäu (Bavaria), two data subsets with test results from multiple tests run in parallel were chosen. The first dataset comprised test results from 175 animals which had been tested from December 2012 to March 2013 by the SICT test, the Bovigam® assay, culture and pathological examination. The second dataset comprised test results from 389 animals which had been tested from April 2013 to February 2014 by the SICCT test, the PCR and which have been examined pathologically. The data was collected as binary data and to some extent, for the SICT test, the SICCT test and the Bovigam® assay, also as metric data.

Diagnostic tests. The SICT test and the SICCT test were performed by field or official veterinarians. Examination of the test results were carried out in accordance to Commission Regulation (EC) No 1226/2002 of 8 July 2002 amending Annex B to Council Directive 64/432/EEC [15]. For the skin tests 0.1 ml of bovine respectively bovine and avian Purified Protein Derivate (PPD) (Wirtschaftsgenossenschaft Deutscher Tierärzte eG) was injected

intradermally in the neck or shoulder of the cattle. For the latent class analysis the inconclusive reactors were assigned twice, once as negative reactors (standard interpretation) and once as positive reactors (severe interpretation). An inconclusive reactor means an increase of skin thickness between 2 and 4 mm (SICT test) respectively 1 and 4 mm (SICCT test) and no occurrence of clinical signs.

For the Bovigam® assay heparinized blood was taken two to 28 days (mean of 8.45 days) after the SICT test by field practitioners. The blood was sent within 6 to 7 hours at room temperature to the laboratory of the Bavarian Health and Food Safety Authority. The Bovigam® assay was carried out according to the manufacturer's instructions. In brief, the blood samples were stimulated overnight with avian and bovine PPD. IFN-γ production of the lymphocytes was then determined by using a sandwich ELISA. Identification of infected animals based on the prescription in the manufacturer's user manual for Germany. This means that the mean optical density (OD) of a sample being stimulated with bovine PPD minus the OD of the same sample stimulated with avian PPD, was greater or equal 0.1.

The pathological examinations were performed at different places (pathology of the Bavarian Health and Food Safety Authority, carcass disposal plants) by veterinarians. Attention was given to the retropharyngeal lymph nodes, lung, gut, spleen, kidneys, liver and the associated lymph nodes as well as organs or lymph nodes with pathological-anatomical changes.

For polymerase chain reaction (PCR) samples were collected during necropsy from the retropharyngeal lymph nodes, lung, gut, spleen, kidneys, liver and the associated lymph nodes. Furthermore, pieces of organs or lymph nodes with pathological-anatomical changes were taken [23]. All samples were investigated in the laboratory of the Bavarian Health and Food Safety Authority. To increase the detection of mycobacteria samples with pathological findings were homogenized. From inconspicuous samples approximately 25 mg were used for DNA extraction. PCR aiming at detecting *Mycobacterium tuberculosis* complex-pathogens was performed for each sample separately according to the *Official Collection of Methods* [1]. The targeted sequences for PCR amplification are a hypothetical helicase and the insertion element (IS) *1081* [40, 41]. According to the official guidelines, results were interpreted as positive if both target sequences were found, as inconclusive if only one target sequence or only weak PCR signals were detected and as negative if no signals were observed. In agreement with the Friedrich-Loeffler-Institute (FLI) single runs were performed for each organ / lymph node.

Bacteriological culture was performed according to the *Official Collection of Methods* [1]. Organs were cultured as aggregate samples, except for organs with macroscopic lesions, which were cultured separately. As liquid media BD BACTEC™ MGIT™ was used. Löwenstein-Jensen and Stonebrink agar slants were used as solid culture media.

Statistical analysis. A Bayesian latent class approach assuming no gold standard, i.e. a perfect diagnostic test without any misclassification, was performed for the three-test dataset (SICCT test, PCR, necropsy) and the four-test dataset (SICT test, Bovigam® assay, necropsy, culture). The skin tests were considered with both their standard and severe interpretation, separately. In total, for the four-test dataset, there were four sensitivities, four specificities, one prevalence and twelve two-way covariances to be estimated, leading in total to 21 unknown parameters [42]. For the three-test dataset there were three sensitivities, three specificities, one prevalence and six two-way covariances to be estimated, leading in total to 13 unknown parameters. Due to the principle of parsimony, higher order terms of covariances were not considered. The specificity of culture was fixed to "1", assuming that no false positive test result exists. This reduces the number of parameters to be estimated for the four-test dataset. For all other estimable parameters, first uninformative beta priors (1,1) were utilized. Second, informative priors basing on expert opinion and published test accuracies [9, 10] were utilized for the sensitivities and specificities of the SICT and the SICCT test. This was done for sensitivity

analysis, respectively, for setting constraints to have still an identifiable model for the three-test dataset with taking the covariances into account [42]. To incorporate the prior information beta distributions (a,b), modeled by beta buster (http://252s-weblive.vet.unimelb.edu.au:3838/users/epi/beta.buster/), were used. For the SICT test we assumed—being 95% sure—that the sensitivity is greater than 50% with a mode at 70% (a = 13,3221; b = 6,2809) and that the specificity is greater than 70% with a mode at 85% (a = 23.903, b = 5.042). Similarly, for the SICCT test we assumed that the sensitivity is greater than 45% with a mode at 65% (a = 12.1979, b = 7.0296) and the specificity is greater than 80% with a mode at 90% (a = 42.5732, b = 5.6192). The presence of conditional dependencies between tests was checked by assessing separately the impact of each covariance term compared to a covariance term set to 0 on the other estimates. Presence of conditional dependencies was assessed graphically (histograms). Additionally, to assess if higher-level conditional dependencies potentially affect the results, random effect models based on the model from Qu et al. 1996 were also explored using the R package randomLCA [43, 44].

Model selection was based on DIC (Deviance Information Criterion) with lower values indicating a better model fit. For a sensitivity analysis of the three-test dataset considering the covariances uninformative priors were used. The best fitting model of the four-test dataset was additionally run with a higher cut off of the Bovigam® assay (OD difference 0.2 instead of 0.1). Due to missing values for continuous Bovigam® assay results, only 171 animals could be included. The models were implemented in JAGS (Just Another Gibbs Sampler) version 3.4.0 for Markov Chain Monte Carlo Simulation (http://mcmc-jags.sourceforge.net/), the software R version 3.0.3 (https://www.r-project.org/) and the package coda [45]. The model code is given in the supplementary online material (S1 Text). For all models the first 20,000 iterations were discarded as burn-in and based on the next 200,000 iterations the posterior distributions of the unknown parameters were derived. Three chains were run from different starting points. Convergence was checked visually by inspecting the density plots of the three chains.

The positive and negative predictive values of the skin tests (standard and severe interpretation), Bovigam® assay (OD difference 0.1 and 0.2), PCR, necropsy and culture were derived based on the estimated prevalence and posteriors obtained from the different models.

Kappa test of agreement

Animal samples and testing. Blood was taken from 21 cows (Braunvieh breed) at six different time points from the *V. jugularis*. All animals belonged to one farm and were tested previously as positive or inconclusive with the SICT test. On two time points the blood was additionally taken from the *V. caudalis mediana* resp. *V. subcutanea abdominis*. Immediately after collection, the blood was sent to five laboratories, all over Germany. After arriving at the laboratories the blood was directly examined with the Bovigam® assay. Due to the fact that the laboratories were distributed all over Germany the time between blood collection and further examination was between 4 to 29 h with a median of 8.0 h. The samples were not blinded.

Statistical analysis. To determine if the laboratories were classifying approximately the same proportion of individuals as positive, first McNemar's test was applied for each given time point between all possible pairwise laboratory comparisons [46]. McNemar's test was performed with the software R version 3.0.3 (https://www.r-project.org/) with the package exact2x2 [47]. For the inter-laboratory agreement the time point with the best accordance of the proportion of positive test results was chosen to determine Cohen's kappa. Also the agreement between the test results of the Bovigam® assay from the differing localizations was estimated using McNemar's test and Cohen's kappa. Cohen's kappa was calculated online with Graphpad software (http://graphpad.com/quickcalcs/kappa2/).

Table 1. Prevalence and diagnostic test accuracies of different models considered from the dataset with 175 animals tested with SICT test, Bovigam® assay, culture [sp = 100%] and necropsy.

Model	Prevalence (95% CI)	SICT test (95% CI)		Bovigam® assay (95% CI)		Culture (95% CI)		Necropsy (95% CI)	
		se	sp	se	sp	se	sp	Se	sp
1	**7.7 (4.2–12.3)**	**70.3 (44.9–90.5)**	**75.8 (68.8–82.2)**	**95.7 (91.3–99.2)**	**6.9 (3.6–11.1)**	**88.9 (65.5–99.7)**	**fixed at 100**	**76.8 (51.6–94.4)**	**99.0 (96.8–100)**
2	7.7 (4.2–12.3)	70.1 (53.5–84.7)	77.1 (70.9–82.8)	95.8 (91.3–99.2)	6.9 (3.6–11.1)	88.9 (65.0–99.7)	fixed at 100	76.9 (51.7–94.4)	98.9 (96.8–99.9)
3	7.9 (4.3–12.6)	70.2 (45.1–90.4)	76.4 (69.7–82.6)	83.3 (74.2–93.5)	23.5 (17.4–30.3)	88.9 (64.8–99.7)	fixed at 100	77.0 (51.2–94.6)	98.9 (96.7–99.9)
4	7.9 (4.3–12.5)	98.2 (94.9–99.9)	4.1 (1.7–7.5)	95.7 (91.2–99.2)	6.9 (3.7–11.2)	87.5(62.8–99.5)	fixed at 100	77.1 (51.4–94.6)	99.1 (97.0–100)

Model 1: SICT test [standard interpretation, uninformative priors], Bovigam® assay [cut-off = 0.1], culture [sp = 100%], no covariances

Model 2: SICT test [standard interpretation, informative priors], Bovigam® assay [cut-off = 0.1], culture [sp = 100%], no covariances

Model 3: SICT test [standard interpretation, uninformative priors], Bovigam® assay [cut-off = 0.2], culture [sp = 100%], no covariances

Model 4: SICT test [severe interpretation, uninformative priors], Bovigam® assay [cut-off = 0.1], culture [sp = 100%] no covariances

se, sensitivity

sp, specificity

Results

Bayesian latent class models

The raw data, comprising the dichotomized test results of the four-test and the three-test dataset, are presented in S1 and S2 Tables.

Four-test dataset. Posterior means and corresponding 95% credibility intervals resulting from Bayesian latent class models are shown in Table 1.

There was no evidence, based on DIC and visual inspection of covariance histograms that including any covariance term led to a better model fit (S3 Table and S1 Fig). Including covariance terms did also not alter the posterior means. Adding random effects to model higher level conditional dependencies did not improve model fit.

If informative instead of flat priors were used for SICT test the DIC decreased slightly (392.6 instead of 393.7) and the posterior means were only marginally affected. If a higher cut-off of 0.2 instead of 0.1 for the Bovigam® assay was applied, then the sensitivity of the Bovigam® assay decreased from 95.7% (91.3–99.2%) to 83.3% (74.2–93.5%) and the specificity increased from 6.9% (3.6–11.1%) to 23.5% (17.4–30.3%). The estimated posteriors of the other tests and the prevalence differed maximally around 0.4%. The dichotomized test results are presented in S4 Table. When interpreting the inconclusive test results of the SICT test as positive the specificity of the SICT test was extremely low with 4.1% (1.7–7.5%).

The positive and negative predictive values for the SICT test (standard and severe interpretation), Bovigam® assay (cut-off 0.1 and 0.2), necropsy and culture are presented in S5 Table.

Three-test dataset. The posterior sensitivities and specificities for the three-test dataset resulting from the Bayesian latent class models are presented in Table 2. With the incorporated informative priors the sensitivity of the SICCT test increased by 2%. The estimated test characteristics of the other tests were only marginally affected. With regard to the histograms (S2 Fig), the posteriors and the DIC (S6 Table) dependence between the sensitivity of the PCR and necropsy seemed to be the most likely. Within this model the sensitivities of the PCR and the necropsy decreased and the prevalence increased. The remaining posteriors range around the same values.

By running this model with flat priors the DIC increased to 754.8 instead of 750.7 and the sensitivities of PCR and necropsy decreased around 5.4% and 6.0%. The other estimated

parameters differed maximal around 2%. The density plots of the estimated probability distributions showed better convergence for the model with the SICCT test as standard interpretation. The specificity for the SICCT test was extremely low with 12.0% (8.7–15.8%) for the severe interpretation.

The positive and negative predictive values for the SICCT test (standard and severe interpretation), PCR and necropsy based on the estimated posteriors and prevalence of the three test data set can be found in S7 Table.

Inter- and intra-laboratory agreement

The raw data utilized for assessing agreement between the different laboratories and the different localizations are presented in S8 and S9 Tables. Based on McNemar's test to assess if the proportions of samples classified as positive differed significantly between the laboratories or the anatomical location, the time point with most non-significant tests was chosen (Table 3) [46]. Estimated p-values for McNemar's test ranged from 0.03 to 1.00. The inter-laboratory agreement between Laboratory 2 and 3 reached a Cohen's kappa value of 0.62 (95% confidence interval from 0.15 to 1.00). The other agreements constituted between -0.16 (95% confidence interval from -0.32 to -0.01) and 0.38 (95% confidence interval from 0.01 to 0.76).

For the agreement between the varying localizations nearly all McNemar's tests are non-significant (Table 4), thus indicating the proportion of positive test results did not significantly differ between the laboratories. One estimated p-value based on McNemar's test was 0.03, giving evidence that there is a disagreement between the two proportions of Bovigam® test results [46]. The best agreement was seen for Laboratory 3 by comparing the Bovigam® assay results between the blood of the *V. jugularis* and *V. subcutanea abdominis* (1.00). Also the agreement between the results of the *V. jugularis* and the *V. caudalis mediana* of this laboratory reached at least a Cohen's kappa value of 0.62 (95% confidence interval from 0.00 to 1.00). Laboratory 5 had a substantial agreement by the comparison between *V. jugularis* and *V. caudalis mediana*. All other agreements were below 0.54 indicating a poor to moderate agreement.

Discussion

Due to the detection of bovine tuberculosis at several occasions during regular abattoir meat inspections in the Allgäu region, a new tuberculosis control program was implemented in

Table 2. Prevalence and diagnostic test accuracies of different models considered from the dataset with 389 animals tested with SICCT test [standard interpretation], PCR and necropsy.

Model	Prevalence (95% CI)	SICCT test (95% CI)		PCR (95% CI)		Necropsy (95% CI)	
		se	sp	se	sp	se	sp
1	17.3 (13.5–21.5)	55.5 (43.3–67.7)	91.7 (88.3–94.6)	80.6 (69.1–90.6)	99.2 (97.6–100)	90.6 (80.6–98.0)	99.1 (97.2–100)
2	17.2 (13.4–21.4)	57.5 (46.5–68.1)	91.5 (88.4–94.2)	80.6 (69.1–90.6)	99.1 (97.5–100)	90.7 (80.7–98.0)	99.1 (97.2–100)
3	**19.8 (14.6–26.5)**	**57.8 (48.0–67.6)**	**92.8 (89.2–96.3)**	**70.6 (52.0–86.0)**	**99.0 (97.4–99.9)**	**78.4 (58.6–93.7)**	**98.9 (96.8–100)**
4	21.8 (15.1–30.9)	56.1 (46.3–66.2)	94.5 (89.7–99.5)	65.2 (44.6–84.5)	99.0 (97.4–99.9)	72.4 (49.8–92.5)	99.0 (96.9–100)
5	15.6 (11.4–20.2)	94.9 (89.5–98.8)	12.0 (8.7–15.8)	87.9 (73.0–99.4)	98.8 (96.9–100.0)	92.2 (81.3–99.6)	97.6 (94.6–99.9)

Model 1: SICCT test [standard interpretation, uninformative priors], no covariances

Model 2: SICCT test [standard interpretation, prior information of se and sp], no covariances

Model 3: SICCT test [standard interpretation, prior information of se and sp], covariance between sensitivity PCR and necropsy

Model 4: SICCT test [standard interpretation, uninformative priors], covariance between sensitivity PCR and necropsy

Model 5: SICCT test [severe interpretation, uninformative priors], no covariances

se, sensitivity

sp, specificity

Table 3. Comparison of the Bovigam® assay test results from five different laboratories by the calculated *p*-value based on McNemar's test, Cohen's kappa values with the 95% confidence interval and the proportions of the test results of one given time point.

Comparison between:	*p*-value	Kappa	CI	Proportions of test results[a]		
				pos/pos	neg/neg	discordant
Lab 1 / Lab 2	0.69	0.31	-0.12 to 0.74	57%	14%	29%
Lab 1 / Lab 3	0.38	0.30	-0.15 to 0.74	63%	11%	26%
Lab 1 / Lab 4	1.00	0.26	-0.23 to 0.75	61%	11%	28%
Lab 1 / Lab 5	0.75	0.00	-0.42 to 0.42	38%	14%	48%
Lab 2 / Lab 3	1.00	0.83	0.50 to 1.00	79%	16%	5%
Lab 2 / Lab 4	1.00	0.26	-0.23 to 0.75	61%	11%	28%
Lab 2 / Lab 5	0.22	0.38	0.01 to 0.76	52%	19%	29%
Lab 3 / Lab 4	1.00	0.09	-0.42 to 0.61	63%	6%	31%
Lab 3 / Lab 5	0.06	0.41	0.05 to 0.77	58%	16%	26%
Lab 4 / Lab 5	0.45	0.11	-0.33 to 0.55	50%	11%	39%

Lab, Laboratory; neg, negative; pos, positive; CI, 95% confidence interval

[a]The number of analyzable test results ranged from 16–21 animals

https://doi.org/10.1371/journal.pone.0179847.t003

November 2012 in Germany. Within this testing regime the Bovigam® assay was performed for the first time as a field test in Germany. The test results gained from this control program were utilized to estimate the sensitivities and specificities of the different tests with a latent class analysis. This was especially of interest as the persons involved in this testing program recognized contradictory test outcomes for the Bovigam® assay which led to distrust and termination of the testing with the Bovigam® assay. These contradictory test outcomes seem to be corroborated by the raw data presented in S1 Table, where out of 175 tested animals 115 were diagnosed positive only with the Bovigam® assay.

The estimated test characteristics of the SICT and the SICCT test, the PCR, necropsy and culture are in line with already published data [9, 10, 13, 28, 31, 48, 49]. For the Bovigam® assay an extremely low specificity was estimated. In this population with an estimated true prevalence of 7.7 the positive predictive values of the Bovigam® assay would be 7.9% (OD difference of 0.1) respectively 8.54% (OD difference of 0.2).

This finding could be corroborated with additional intra- and inter-laboratory testing of agreement.

In this study no-gold-standard-models, relying on Bayesian latent class approaches, which are increasingly used in medical and veterinary sciences, were used [50, 51]. The specificity of the culture was set at 100%, as a positive result is assumed to be truly positive [7]. The best

Table 4. Comparison of the Bovigam® assay test results between differing localizations by calculating *p*-values based on McNemar's test and Cohen's kappa values.

		Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
V.j. / V.s.a.	*p*-value	1.00	0.13	1.00	0.38	0.03
	kappa	0.29	0.41	1.00	0.46	0.37
	CI	-0.12–0.72	0.00–0.86	1.00–1.00	0.03–0.83	0.08–0.74
V.j. / V.c.m.	*p*-value	0.45	0.63	0.50	1.00	0.25
	kappa	0.16	0.54	0.62	-0.03	0.70
	CI	-0.22–0.59	0.03–0.90	0.00–1.00	-0.37–0.43	0.36–1.00

V.j.: *V. jugularis*; *V.s.a.*: *V. subcutanea abdominis*; *V.c.m.*: *V. caudalis mediana*; CI: 95% confidence interval

https://doi.org/10.1371/journal.pone.0179847.t004

fitting model was chosen by DIC. Additionally, as the DIC has its limitations, histograms and posteriors were evaluated [52–54]. In order to comply with good statistical practice all possible two way covariances were taken into account [55]. A conditional dependency could only be seen between the sensitivity of PCR and necropsy. These two examination methods do not rely on similar biological basics, but were related as sample selection and sample size for the PCR were associated with pathological examination. Due to the fact that both datasets were quite small, conditional dependence between other tests could not be excluded. Inclusion of prior information of the SICT test, for a sensitivity analysis of the four-test dataset, did not influence the posteriors. For the three-test dataset the DIC increases by running a sensitivity analysis with flat priors, indicating a worse model fit. As already shown by Álvarez et al. [10] the test characteristics of the skin tests alter with a severe interpretation insofar that the sensitivity increases and the specificity decreases. Within our data a strong shift to lower specificities could be seen for the severe interpretation of the skin tests. This outcome seems to be data driven, as in both datasets most of the animals had an inconclusive skin test result. Therefore and because of the poorer convergence for the models with severe interpretation, which could be due to the small amount of true positive test results, the focus was set on the skin tests standard interpretation.

For the skin tests, PCR, necropsy and culture, the estimated sensitivities and specificities are in accordance of test characteristics from other publications [9, 10, 13, 28, 31, 48, 49]. The wide credibility intervals for the sensitivities (19.6 to 45.6) could be explained by the small data pool of true positive animals. With regard to the estimated test characteristics of the skin tests it has to be considered that these could have been affected by the performance of the skin tests [56]. The test characteristics of the pathological examination were within both datasets 76.8% (51.6–94.4%) and 78.4% (58.6–93.7%), respectively, for the sensitivity and around 99.0% (96.8–100%) for the specificity. The fact that the pathological examination was done in different localizations from different persons as well as the small data pool of true positive animals explains the wide credibility interval for the sensitivity. This spectrum bias appears to be present in both subpopulations. Within our estimated test characteristics the SICCT test is less sensitive although more specific than the SICT test. PCR and necropsy are less sensitive than culture. Therefore, culture is still an essential diagnostic tool.

In the literature the test characteristics of the Bovigam® assay are stated as between 66.9–100% for sensitivity and 70–99.6% for specificity [9, 12, 49]. We estimated a quite high sensitivity of 95.7% (91.3–99.2%), but an extremely low specificity of 6.9% (3.6–11.1%). This stands in line with the experience of the persons involved in the bovine TB testing. With setting the cut-off higher an increase in the specificity was expected, as already reported by others [57, 58]. Indeed, the specificity raised to 23.5% (17.4–30.3%) thereby the sensitivity decreased to 83.3% (74.2–93.5%), This shows again that the model itself is robust. To our knowledge such low specificities were never recognized before for the Bovigam® assay. Although it was already stated that the Bovigam® assay is more sensitive, but less specific than the SICCT test [59]. And it was shown that fewer than 20% of the animals tested positive in the Bovigam® assay were also positive in culture or pathology [60]. Van Dijk [61] showed that the Bovigam® assay is likely to have false positive results and this in a higher amount than the SICCT test. With a decrease of the prevalence the amount of false positive test results even increases [61].

Among the influential factors discussed in literature a previously performed skin test is discussed to have an effect on the specificity of the Bovigam® assay. Within our study the SICT test was performed two to 28 (mean of 8.45 days) days prior to the Bovigam® assay. Several studies discuss the effect of a previous skin test (either CFT test or SICCT test) towards the IFN-γ production in natural or experimental infected cattle. Whereas the CFT test leads to a clear increase of IFN-γ production, this influence is not obvious after the SICCT test [62]. The

previously performed skin test in this study may have had an impact on the estimated specificity within the examined subpopulation. A genetic influence and an association between the breed and the outcome of the SICCT test were reported by Amos et al. [63]. This was not seen for the Bovigam® assay [64]. An influence of breed and genetics might be present in the Allgäu region, but further investigations have to be made to confirm this. The Bovigam® assay was only carried out within the regions Ober- and Ostallgäu during November 2012 until March 2013. This regional and seasonal limitation could have had an impact on the high amount of false positive test results [64, 65]. The correlation between season and occurrence of saprophytic mycobacteria might be associated with this [66, 67]. Moreover, infections with *Mycobacterium avium* subspecies *paratuberculosis* (MAP) may lead to false positive results for the Bovigam® assay [68]. Since the tested animals have not been examined for a concurrent MAP infection this impact could not be excluded. Furthermore, an infection with *Fasciola hepatica* is also reported to influence the IFN-γ response. Although this is until now only stated for the skin test and in context of false negative test results [69]. The specificity of the Bovigam® assay tests varies also with the concentration and potency of PPDs [70], which can differ remarkably [71, 72]. These influences might explain to some extend the estimated low specificities of the Bovigam® assay. Besides, the fact that bovine tuberculosis, in the regions Ober- and Ostallgäu was caused by *Mycobacterium caprae* may have influenced the Bovigam® assay results, too, as bovine tuberculosis in other regions worldwide is predominantly caused by *Mycobacterium bovis*. However, the low inter- and intralaboratory agreements between the Bovigam® assay outcomes could not be fully explained by this. The transportation time as well as the experience seems to influence the Bovigam® assay test outcomes, as between the laboratories with the shortest transportation time (Laboratory 3, data not shown) and the most experience (Laboratory 2 and 3, data not shown) the best, but still only substantial agreement was estimated. By comparing the intralaboratory agreement between the different laboratories again the laboratory with the most experience and the shortest transportation time (Laboratory 3) had the best agreement between the results of the blood taken from the *V. jugularis* and the *V. subcutanea abdominis*. A longer storage or transportation of the blood samples might lead to a decrease in the mean OD or the IFN-γ production [73–75]. With regard to the sensitivity and specificity of the Bovigam® assay several studies state that blood could also be processed 24 h later without statistical significant changes [76, 77]. However, Laboratory 3 reached also only a substantial agreement of 0.62 between the Bovigam® assay test results of the *V. jugularis* and the *V. caudalis mediana*. As for the intralaboratory agreement only the localization of the blood collection altered, much better accordance between the Bovigam® assay test results were expected, as blood taken from differing localizations should not differ [78, 79]. But the smaller diameter of the tail vain could lead to more damage and therefore micro-clotting, resulting in captured lymphocytes and therefore lower IFN-γ release [80]. Despite, there are conflicting views if an equal distribution of all lymphocyte subpopulations all over the body can be assumed in general. Regarding this, it must be taken into account that a blood sample can only give a snapshot. Although all five laboratories were officially approved none of them reported good concordance for the Bovigam® assay test results. It seems that the Bovigam® assay is a diagnostic tool with some disadvantages. Many influences including external factors (MAP, saprophytic mycobacteria, previous skin test and genetic components) and factors directly connected with the test performance, as the concentration of the PPDs, transportation time of the blood, localization of blood collection and also the experience of the laboratories might lead to differing test results. A higher specificity of the Bovigam® assay, especially in low prevalence herds and animals having a co-infection with MAP, can be achieved by using the proteins ESAT6 and CFP10 instead of PPDa and PPDb [68, 81]. Also working out and evaluating an individual test performance (proteins, protein concentration, cut-off etc.) for

each Bovigam® assay application as a field test might lead to better test characteristics [64, 70]. This could be demonstrated in this study in so far as the specificity increased by setting the cut-off to 0.2.

To our knowledge the use of a latent class analysis for the estimation of test characteristics for bTB diagnostic tests was never done before in Germany. An important strength of this study is that the data were gained from surveillance and therefore originates from a special epidemiologic situation. But this means also that only a subpopulation was tested and the animals were not chosen randomly. According to this background information our findings cannot be generalized. Additionally a new version of the Bovigam® assay has been developed since 2013 to which our findings cannot be transferred [82].

Conclusion

With this latent class analysis the test characteristics of different diagnostic tests used in the current bovine TB outbreak in Southern Germany could be estimated. Within this study an extremely low specificity and a low inter- and intralaboratory agreement were estimated for the Bovigam® assay. These findings might be due to influences affecting the environment or the immune system of the cow. Also factors that are associated with the testing procedure and the laboratories chosen might have had an effect. Therefore, the change during the testing regime towards SICCT test as only ante-mortem test was correct and founded. Despite the fact that the Bovigam® assay has been further advanced, a previous test evaluation prior to future surveillance programs is highly recommended. The estimated test characteristics for the other tests were in an acceptable range.

Supporting information

S1 Text. Bayesian latent-class model code for four diagnostic tests.
(DOCX)

S1 Table. Number of test result combinations in the four-test dataset (n = 175), the inconclusive test results of the SICT test once considered as negative (standard interpretation) and once as positive (severe interpretation). [b] cut-off: 0.1.
(DOCX)

S2 Table. Number of test result combinations in the three-test dataset (n = 389), the inconclusive test results of the SICCT test once considered as negative (standard interpretation) and once as positive (severe interpretation).
(DOCX)

S3 Table. DIC, prevalence and diagnostic test accuracies of different models, without and with covariances of the sensitivities between the different tests, considered from the dataset (n = 175) tested with SICT test [standard interpretation, uninformative priors], Bovigam® assay [cut-off = 0.1], culture [sp = 100%] and necropsy. Model 1: no covariances; Model 2: covariance sensitivity SICT test, Bovigam® assay; Model 3: covariance sensitivity SICT test, culture; Model 4: covariance sensitivity SICT test, pathology; Model 5: covariance sensitivity Bovigam® assay, culture; Model 6: covariance sensitivity Bovigam® assay, pathology; Model 7: covariance sensitivity culture, pathology; CI, credibility interval; se, sensitivity; sp, specificity.
(DOCX)

S4 Table. Dichotomized test results of the Bovigam® assay for two different cut-offs.
(DOCX)

S5 Table. Positive and negative predictive values of the SICT test, Bovigam® assay, culture [sp = 100%] and necropsy calculated from the prevalence and diagnostic test accuracies obtained from the models of Table 1. PPV, positive predictive value; NPV, negative predictive value.
(DOCX)

S6 Table. DIC, prevalence and diagnostic test accuracies of different models, without and with covariances of the sensitivities and specificities between the different tests, considered from the dataset (n = 389) tested with SICCT test [standard interpretation; prior information], PCR and necropsy. Model 1: no covariances; Model 2: covariance sensitivity SICCT test, PCR; Model 3: covariance sensitivity SICCT test, pathology; Model 4: covariance sensitivity PCR, pathology; Model 5: covariance specificity SICCT test, PCR; Model 6: covariance specificity SICCT test, pathology; Model 7: covariance specificity PCR, pathology; CI, credibility interval; se, sensitivity; sp, specificity.
(DOCX)

S7 Table. Positive and negative predictive values of the SICCT test, PCR and necropsy calculated from the prevalence and diagnostic test accuracies obtained from the models of Table 2. PPV, positive predictive value; NPV, negative predictive value.
(DOCX)

S8 Table. Test results of the Bovigam® assay from five officially approved laboratories; the results of the time point with the best accordance of the proportion of positive test results is shown. Pos, positive; neg, negative; n.a., not analyzable.
(DOCX)

S9 Table. Test results of the Bovigam® assay from five officially approved laboratories. The assayed blood was taken at two time points respectively from two differing localizations. *V.j.*, *V. jugularis; V.s.a.*, *V. subcutanea abdominis; V.c.m.*, *V. caudalis mediana*; pos, positive; neg, negative;?, inconclusive; n.a., not analyzable.
(DOCX)

S1 Fig. Histograms from the covariances of the sensitivities between the different tests, considered from the four-test dataset (n = 175) tested with SICT test [standard interpretation; no prior information], Bovigam® assay [cut−off = 0.1], culture [sp = 100%] and necropsy.
(PDF)

S2 Fig. Histograms from the covariances of the sensitivities and specificities between the different tests, considered from the three-test dataset (n = 389) tested with SICCT test [standard interpretation; prior information], PCR and necropsy.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: GKS RKS SH.

Data curation: AG DD VBP SH.

Formal analysis: VBP SH AG.

Funding acquisition: GKS.

Investigation: PZ SH AHM RKS VBP.

Methodology: SH VBP.

Project administration: GKS SH VBP.

Resources: GKS CSL.

Software: SH.

Supervision: GKS SH.

Validation: SH RKS VBP.

Visualization: VBP SH.

Writing – original draft: VBP SH.

Writing – review & editing: SH GKS PZ SH AG RKS.

References

1. Bovine tuberculosis (*Mycobacterium bovis* and *Mycobacterium caprae*). 2015 28.05.2015. In: Official Collection of Methods [Internet]. Jena: Friedrich-Loeffler-Institute [1–22]. Available from: https://www.fli.de/de/publikationen/amtliche-methodensammlung/.

2. Sattelmair H. Die Tuberkulose des Rindes-ein Beitrag zur Geschichte der Haustierkrankheiten. Berlin: Freie Universität Berlin, Germany; 2005.

3. Livingstone PG, Hancox N, Nugent G, de Lisle GW. Toward eradication: the effect of Mycobacterium bovis infection in wildlife on the evolution and future direction of bovine tuberculosis management in New Zealand. New Zealand veterinary journal. 2014; 63(sup1):4–18. https://doi.org/10.1080/00480169.2014.971082 PMID: 25273888

4. Miller RS, Farnsworth ML, Malmberg JL. Diseases at the livestock-wildlife interface: status, challenges, and opportunities in the United States. Preventive veterinary medicine. 2013; 110(2):119–32. https://doi.org/10.1016/j.prevetmed.2012.11.021 PMID: 23254245

5. Robinson PA. A history of bovine tuberculosis eradication policy in Northern Ireland. Epidemiology and Infection. 2015:1–14. https://doi.org/10.1017/S0950268815000291 PMID: 25778830

6. Zinsstag J, Schelling E, Roth F, Kazwala R, Thoen CO, Steele JH, et al. Economics of Bovine Tuberculosis. Mycobacterium Bovis Infection in Animals and Humans: Blackwell Publishing Ltd; 2008. p. 68–83.

7. Cousins DV. Bovine tuberculosis. 2009. In: Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2014 [Internet]. Paris: World Organisation for Animal Health; [1–16]. Available from: http://www.oie.int/international-standard-setting/terrestrial-manual/access-online/.

8. Monaghan ML, Doherty ML, Collins JD, Kazda JF, Quinn PJ. The tuberculin test. Veterinary Microbiology. 1994; 40(1–2):111–24. https://doi.org/10.1016/0378-1135(94)90050-7 PMID: 8073619

9. de la Rua-Domenech R, Goodchild AT, Vordermeier HM, Hewinson RG, Christiansen KH, Clifton-Hadley RS. Ante mortem diagnosis of tuberculosis in cattle: a review of the tuberculin tests, gamma-interferon assay and other ancillary diagnostic techniques. Res Vet Sci. 2006; 81(2):190–210. Epub 2006/03/04. https://doi.org/10.1016/j.rvsc.2005.11.005 PMID: 16513150

10. Álvarez J, Perez A, Bezos J, Marqués S, Grau A, Saez JL, et al. Evaluation of the sensitivity and specificity of bovine tuberculosis diagnostic tests in naturally infected cattle herds using a Bayesian approach. Veterinary Microbiology. 2012; 155(1):38–43. https://doi.org/10.1016/j.vetmic.2011.07.034 PMID: 21890284

11. Wood PR, Jones SL. BOVIGAM: an in vitro cellular diagnostic test for bovine tuberculosis. Tuberculosis (Edinburgh, Scotland). 2001; 81(1):147–55. https://doi.org/10.1054/tube.2000.0272 PubMed Central PMCID: PMC11463236. PMID: 11463236

12. Bezos J, Casal C, Romero B, Schroeder B, Hardegger R, Raeber AJ, et al. Current ante-mortem techniques for diagnosis of bovine tuberculosis. Research in Veterinary Science. 2014; 97, Supplement: S44–S52. https://doi.org/10.1016/j.rvsc.2014.04.002

13. Courcoul A, Moyen JL, Brugere L, Faye S, Henault S, Gares H, et al. Estimation of sensitivity and specificity of bacteriology, histopathology and PCR for the confirmatory diagnosis of bovine tuberculosis using latent class analysis. PLoS One. 2014; 9(3):e90334. Epub 2014/03/15. https://doi.org/10.1371/journal.pone.0090334 PMID: 24625670; PubMed Central PMCID: PMCPmc3953111.

14. Corner LA. Post mortem diagnosis of Mycobacterium bovis infection in cattle. Veterinary Microbiology. 1994; 40(1–2):53–63. https://doi.org/10.1016/0378-1135(94)90046-9 PMID: 8073629

15. Commission Regulation (EC) No 1226/2002 of 8 July 2002 amending Annex B to Council Directive 64/432/EEC, (09/07/2002, 2002).

16. Hunermund G, Kropp R. The battle against cattle tuberculosis in Germany. Pneumologie (Stuttgart, Germany). 2006; 60(12). https://doi.org/10.1055/s-2006-944284 PMID: 17163317

17. Official Journal of the European Union, Commission implementing decision of 12 October 2011 amending Decision 2003/467/EC as regards the declaration of Latvia as officially tuberculosis-free Member State and the declaration of certain administrative regions in Portugal as officially enzootic-bovine-leukosis-free regions (notified under document C(2011) 7186), Pub. L. No. OJ L 268(13.10.2011, 2011).

18. Richtlinie 97/12/EG des Rates vom 17. März 1997 zur Änderung und Aktualisierung der Richtlinie 64/432/EWG zur Regelung viehseuchenrechtlicher Fragen beim innergemeinschaftlichen Handelsverkehr mit Rindern und Schweinen, (25.04.1997, 1997).

19. Köhler H, Moser I, Menge C, editors. Tuberkulose des Rindes: wieder ein Thema?! Infektionskrankheiten des Rindes–Neues und Wichtiges für die tierärztliche Praxis; 2012; Stiftung Tierärztliche Hochschule Hannover Akademie für Tiergesundheit e. V.

20. Rolle M, Mayr A, Selbitz H-J, Truyen U, Valentin-Weigand P. Tiermedizinische Mikrobiologie, Infektions und Seuchenlehre. Stuttgart: Enke Verlag; 2011.

21. Moser I, Köhler H, Menge C. Die Tuberkulose des Rindes– überraschend wieder oder immer noch präsent? Tierärztliche Praxis Großtiere. 2014; 42(4):240–9.

22. Heigl F. Rinder-und Rotwild-Tuberkulose im Oberallgäu, Eine aktuelle Übersicht. 2013.

23. Verordnung zum Schutz gegen die Tuberkulose des Rindes, (12.07.2013, 2013).

24. Zellner G. Untersuchungsprogramm "Rindertuberkulose in den Landkreisen der Alpenkette". In: Gesundheit BSfUu, editor. München2013.

25. Corner LA, Melville L, McCubbin K, Small KJ, McCormick BS, Wood PR, et al. Efficiency of inspection procedures for the detection of tuberculous lesions in cattle. Australian veterinary journal. 1990; 67 (11):389–92. https://doi.org/10.1111/j.1751-0813.1990.tb03020.x PMID: 2085291

26. Norby B, Bartlett PC, Fitzgerald SD, Granger LM, Bruning-Fann CS, Whipple DL, et al. The sensitivity of gross necropsy, caudal fold and comparative cervical tests for the diagnosis of bovine tuberculosis. Journal of Veterinary Diagnostic Investigation. 2004; 16(2):126–31. https://doi.org/10.1177/104063870401600206 PMID: 15053363

27. Szmaragd C, Green LE, Medley GF, Browne WJ. Impact of Imperfect Test Sensitivity on Determining Risk Factors: The Case of Bovine Tuberculosis. PLoS ONE. 2012; 7(8):e43116. https://doi.org/10.1371/journal.pone.0043116 PMID: 22912804

28. Gerstmair E-M. Validierung molekularbiologischer und immunologischer Nachweisverfahren für die Tuberkulose bei Rindern und Tuberkulosemonitoring beim Rotwild. München: Ludwig-Maximians Universität München; 2011.

29. Buddle BM, McCarthy AR, Ryan TJ, Pollock JM, Vordermeier HM, Hewinson RG, et al. Use of mycobacterial peptides and recombinant proteins for the diagnosis of bovine tuberculosis in skin test-positive cattle. The Veterinary record. 2003; 153(20):615–20. Epub 2003/12/05. PMID: 14653340.

30. Quirin R, Rasolofo V, Andriambololona R, Ramboasolo A. Validity of intradermal tuberculin testing for the screening of bovine tuberculosis in Madagascar. The Onderstepoort journal of veterinary research. 2001; 68(3):231. PMID: 11769356

31. Hartnack S, Torgerson P. The accuracy of the Single Intradermal Comparative Skin Test for the diagnosis of Bovine Tuberculosis estimated from a systematic literature search. J Mycobac Dis. 2012; 2 (120):2161–1068.1000120.

32. Hui SL, Walter SD. Estimating the Error Rates of Diagnostic Tests. Biometrics. 1980; 36(1):167–71. https://doi.org/10.2307/2530508 PMID: 7370371

33. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. Statistical Methods in Medical Research. 1998; 7(4):354–70. https://doi.org/10.1177/096228029800700404 PMID: 9871952

34. Bermingham ML, Handel IG, Glass EJ, Woolliams JA, Bronsvoort BMDC, McBride SH, et al. Hui and Walter's latent-class model extended to estimate diagnostic test properties from surveillance data: A latent model for latent data. Scientific Reports. 2015; 5. https://doi.org/10.1038/srep11861 PMID: 26148538

35. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. Journal of Clinical Epidemiology. 1988; 41(9):923–37. https://doi.org/10.1016/0895-4356(88)90110-2 PMID: 3054000

36. Standard Operating Procedure for OIE Registration of Diagnostic Kits, Guide and Administrative Forms Paris: World Organisation for Animal Health (OIE); 2012. Available from: http://www.oie.int/doc/ged/D12069.PDF.

37. Directive 64/432/EEC on animal health problems affecting intra-EU trade in bovine animals and swine, (1965).

38. Council Directive 80/219/EEC of 22 January 1980 amending Directive 64/432/EEC as regards tuberculosis and brucellosis (1980).

39. Council Directive 97/12/EC of 17 March 1997 amending and updating Directive 64/432/EEC on health problems affecting intra-Community trade in bovine animals and swine, (1997).

40. Rodriguez JG, Mejia GA, Del Portillo P, Patarroyo ME, Murillo LA. Species-specific identification of Mycobacterium bovis by PCR. Microbiology. 1995; 141(9):2131–8. https://doi.org/10.1099/13500872-141-9-2131 PMID: 7496524

41. Taylor GM, Worth DR, Palmer S, Jahans K, Hewinson RG. Rapid detection of Mycobacterium bovis DNA in cattle lymph nodes with visible lesions using PCR. BMC Veterinary Research. 2007; 3:12–. https://doi.org/10.1186/1746-6148-3-12 PMID: 17567891

42. Berkvens D, Speybroeck N, Praet N, Adel A, Lesaffre E. Estimating Disease Prevalence in a Bayesian Framework Using Probabilistic Constraints. Epidemiology. 2006; 17(2):145–53 https://doi.org/10.1097/01.ede.0000198422.64801.8d PMID: 16477254

43. Qu Y, Tan M, Kutner MH. Random Effects Models in Latent Class Analysis for Evaluating Accuracy of Diagnostic Tests. Biometrics. 1996; 52(3):797–810. https://doi.org/10.2307/2533043 PMID: 8805757

44. Beath K. randomLCA: Random Effects Latent Class Analysis. R package version 1.0–11 ed2017.

45. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006; 6(1):7–11.

46. Dohoo I, Martin W, Stryhn H. Veterinary epidemiologic research. Charlottetown, Canada2009.

47. Fay MP. Two-sided Exact Tests and Matching Confidence Intervals. R Journal. 2010; 2(1):53–8.

48. Koöhler H, Hoffmann B, Moser I, Moöbius P. Novel probe-based real-time PCR for rapid diagnosis of bovine Tuberculosis in tissue samples with suspect lesions. 16th International Symposium of the World Association of Veterinary Laboratory Diagnosticians; Berlin2013.

49. Clegg TA, Duignan A, Whelan C, Gormley E, Good M, Clarke J, et al. 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. Veterinary Microbiology. 2011; 151(1–2):68–76. https://doi.org/10.1016/j.vetmic.2011.02.027 PMID: 21470800

50. Lewis F, Torgerson P. A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. Emerging Themes in Epidemiology. 2012; 9(1):9. https://doi.org/10.1186/1742-7622-9-9 PMID: 23270542

51. van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent Class Models in Diagnostic Studies When There is No Reference Standard—A Systematic Review. American Journal of Epidemiology. 2013. https://doi.org/10.1093/aje/kwt286 PMID: 24272278

52. Ward EJ. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecological Modelling. 2008; 211(1–2):1–10. https://doi.org/10.1016/j.ecolmodel.2007.10.030

53. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014; 76(3):485–93. https://doi.org/10.1111/rssb.12062

54. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64(4):583–639. https://doi.org/10.1111/1467-9868.00353

55. Vacek PM. The Effect of Conditional Dependence on the Evaluation of Diagnostic Tests. Biometrics. 1985; 41(4):959–68. https://doi.org/10.2307/2530967 PMID: 3830260

56. Pucken VB, Götz F, Mansfeld R, Moder S, Sauter-Louis C, Straubinger RK, et al. Der Tuberkulin-Hauttest: Literatur, Richtlinie und Umsetzung in der Praxis. Tierärztliche Praxis Großtiere. 2015;(2015-07-02 00:00:00). https://doi.org/10.15653/TPG-141145

57. Faye S, Moyen J-L, Gares H, Benet J-J, Garin-Bastuji B, Boschiroli M-L. 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. Veterinary Microbiology. 2011; 151(1–2):60–7. https://doi.org/10.1016/j.vetmic.2011.02.026 PMID: 21420258

58. Antognoli MC, Remmenga MD, Bengtson SD, Clark HJ, Orloski KA, Gustafson LL, et al. Analysis of the diagnostic accuracy of the gamma interferon assay for detection of bovine tuberculosis in US herds. Preventive veterinary medicine. 2011; 101(1):35–41.

59. Vordermeier M, Ewer K. Specificity Trial of the BOVIGAM® IFN-Gamma Test in GB Cattle In: TB Research Group VLA, editor. New Haw, Addlestone,: DEFRA; 2006.

60. anonym. VLA Surveillance Project SB4008: Laboratory Testing and Epidemiology Support for the National Gamma Interferon Field Trial, Part 1—Report on the 195 herds recruited into the trial DEFRA, 2006.

61. van Dijk J. Towards Risk-Based Test Protocols: Estimating the Contribution of Intensive Testing to the UK Bovine Tuberculosis Problem. PLoS ONE. 2013; 8(5):e63961. https://doi.org/10.1371/journal.pone.0063961 PMID: 23717517

62. Schiller I, Vordermeier HM, Waters WR, Whelan AO, Coad M, Gormley E, et al. Bovine tuberculosis: effect of the tuberculin skin test on in vitro interferon gamma responses. Veterinary immunology and immunopathology. 2010; 136(1–2):1–11. Epub 2010/03/12. https://doi.org/10.1016/j.vetimm.2010.02.007 PMID: 20219253.

63. Amos W, Brooks-Pollock E, Blackwell R, Driscoll E, Nelson-Flower M, Conlan AJK. Genetic Predisposition to Pass the Standard SICCT Test for Bovine Tuberculosis in British Cattle. PLoS ONE. 2013; 8(3): e58245. https://doi.org/10.1371/journal.pone.0058245 PMID: 23554880

64. Gormley E, Doyle M, Duignan A, Good M, More SJ, Clegg TA. Identification of risk factors associated with disclosure of false positive bovine tuberculosis reactors using the gamma-interferon (IFNγ) assay. Veterinary research. 2013; 44(1):117. https://doi.org/10.1186/1297-9716-44-117 PMID: 24308747

65. Lauzi S, Pasotto D, Amadori M, Archetti IL, Poli G, Bonizzi L. Evaluation of the specificity of the gamma-interferon test in Italian bovine tuberculosis-free herds. Veterinary journal (London, England: 1997). 2000; 160(1):17–24. Epub 2000/08/19. https://doi.org/10.1053/tvjl.1999.0444 PMID: 10950131.

66. Donoghue H, Overend E, Stanford J. A longitudinal study of environmental mycobacteria on a farm in south-west England. Journal of applied microbiology. 1997; 82(1):57–67. PMID: 9113878

67. Cooney R, Kazda J, Quinn J, Cook B, Muller K, Monaghan M. Environmental mycobacteria in Ireland as a source of non-specific sensitisation to tuberculins. Ir Vet J. 1997; 50:370–3.

68. Barry C, Corbett D, Bakker D, Andersen P, McNair J, Strain S. The effect of Mycobacterium avium complex infections on routine Mycobacterium bovis diagnostic tests. Veterinary medicine international. 2011; 2011:7 pages. https://doi.org/10.4061/2011/145092 PMID: 21772961

69. Claridge J, Diggle P, McCann CM, Mulcahy G, Flynn R, McNair J, et al. Fasciola hepatica is associated with the failure to detect bovine tuberculosis in dairy cattle. Nature Communications. 2012; 3:853. https://doi.org/10.1038/ncomms1840 PMID: 22617293

70. Cagiola M, Feliziani F, Severi G, Pasquali P, Rutili D. Analysis of possible factors affecting the specificity of the gamma interferon test in tuberculosis-free cattle herds. Clinical and diagnostic laboratory immunology. 2004; 11(5):952–6. Epub 2004/09/11. https://doi.org/10.1128/CDLI.11.5.952-956.2004 PMID: 15358658; PubMed Central PMCID: PMCPMC515264.

71. Good M, Clegg TA, Murphy F, More SJ. The comparative performance of the single intradermal comparative tuberculin test in Irish cattle, using tuberculin PPD combinations from different manufacturers. Veterinary microbiology. 2011; 151(1):77–84.

72. Bakker D. Practicalities of the immune-based diagnostic assay for the control of bovine tuberculosis. VI International M Bovis conference; Cardiff2014.

73. Gormley E, Doyle MB, McGill K, Costello E, Good M, Collins JD. The effect of the tuberculin test and the consequences of a delay in blood culture on the sensitivity of a gamma-interferon assay for the detection of Mycobacterium bovis infection in cattle. Veterinary immunology and immunopathology. 2004; 102(4):413–20. Epub 2004/11/16. https://doi.org/10.1016/j.vetimm.2004.08.002 PMID: 15541794.

74. Schiller I, Waters WR, Vordermeier HM, Nonnecke B, Welsh M, Keck N, et al. Optimization of a whole-blood gamma interferon assay for detection of Mycobacterium bovis-infected cattle. Clinical and Vaccine Immunology. 2009; 16(8):1196–202. https://doi.org/10.1128/CVI.00150-09 PMID: 19571108

75. Coad M, Hewinson RG, Clifford D, Vordermeier HM, Whelan AO. Influence of skin testing and blood storage on interferon-gamma production in cattle affected naturally with Mycobacterium bovis. The Veterinary record. 2007; 160(19):660–2. Epub 2007/05/15. PMID: 17496274.

76. Lopes LB, Alves TM, Stynen APR, Mota PM, Leite RC, Lage AP. Parameter estimation and use of gamma interferon assay for the diagnosis of bovine tuberculosis in Brazil. Pesquisa Veterinária Brasileira. 2012; 32(4):279–83.

77. Ryan T, Buddle B, De Lisle G. An evaluation of the gamma interferon test for detecting bovine tuberculosis in cattle 8 to 28 days after tuberculin skin testing. Research in veterinary science. 2000; 69(1):57–61. https://doi.org/10.1053/rvsc.2000.0386 PMID: 10924395

78. Sears PM, Paape MJ, Pearson RE, Gwazdauskas FC. Comparison between tail vein and jugular vein cannulation in cattle. Journal of dairy science. 1978; 61(7):974–9. https://doi.org/10.3168/jds.S0022-0302(78)83676-5 PMID: 701538

79. Rothel JS, Corner LA, Wood PR. Bovine tuberculosis: immunodiagnosis. 1993. In: Australian Standard Diagnostic Techniques for Animal Diseases [Internet]. East Melbourne, Vic: CSIRO. Available from: http://www.agriculture.gov.au/SiteCollectionDocuments/animal/ahl/ASDT-bovine_tuberculosis.pdf.

80. Böttcher J, Janowetz B, Moser H, Winter P, Alex M, Meier N. Diagnosis of BHV-1Infection in the presence of maternally derived antibodies. 1st Congress of the European Association of Veterinary Laboratory Diagnosticians; Lelystad2010.

81. Aagaard C, Govaerts M, Meikle V, Gutiérrez-Pabello JA, McNair J, Andersen P, et al. Detection of bovine tuberculosis in herds with different disease prevalence and influence of paratuberculosis infection on PPDB and ESAT-6/CFP10 specificity. Preventive Veterinary Medicine. 2010; 96(3–4):161–9. https://doi.org/10.1016/j.prevetmed.2010.06.007 PMID: 20630607

82. BOVIGAM® 2G TB Kit 2016. Available from: https://www.thermofisher.com/order/catalog/product/63330?ICID=search-product.