SCIENTIFIC OPINION

Scientific Opinion on the use of a gamma interferon test for the diagnosis of bovine tuberculosis[1]

EFSA Panel on Animal Health and Welfare (AHAW) [2, 3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

The procedures for gaining, maintaining, suspending, withdrawing or re-gaining official bovine tuberculosis free herd status and for certification for intra-Union trade are based on the results of tuberculin skin tests. The skin test has a number of drawbacks, therefore the suitability of the gamma interferon test and other tests to be included by EU legislation was assessed. Suitability means that the test has a sensitivity equivalent or superior to the standard test currently used in the European Union and specificity not lower than that of the standard test with the lowest specificity used in the EU. Furthermore, there should be no foreseeable practical difficulties that could compromise test performance. It was concluded that purified protein derivative based gamma interferon tests can be included amongst the official tests for the purpose of demonstrating freedom. However, some results suggest that the specificity of the purified protein derivative based gamma interferon tests may not always be as high as the single intradermal tuberculin test. In case the test is included, the protocols for its use for this purpose should be harmonised in the EU. Based on the reviewed information, other tests should not yet be considered for inclusion in the official tests for the purpose of granting and retaining official tuberculosis free herd status. Further evaluation of the suitability of the gamma interferon tests test should study the influence of factors such as the presence of environmental mycobacteria, prevalence of bovine tuberculosis in the herd, the age type and bovine tuberculosis test history of the animals all of which may affect test specificity and hence the suitability of the test for demonstrating freedom from bovine tuberculosis in different situations.

KEY WORDS

Bovine tuberculosis, diagnostic, gamma-interferon, free status

[1] On request from European Commission, Question No EFSA-Q-2011-01254, adopted on 21 November 2012.

[2] Panel members : Edit Authie, Charlotte Berg, Anette Bøtner, Howard Browman, Ilaria Capua, Aline De Koijer, Klaus Depner, Mariano Domingo, Sandra Edwards, Christine Fourichon, Frank Koenen, Simon More, Mohan Raj, Liisa Sihvonen, Hans Spoolder, Jan Arend Stegeman, Hans-Hermann Thulke, Antonio Velarde, Ivar Vågsholm, Preben Willeberg and Stéphan Zientara. Correspondence: ahaw@efsa.europa.eu

[3] Acknowledgement: The Panel wishes to thank the members of the Working Group on Bovine TB Test, A.J. Stegeman, E. Gormley, S. More, S. Downs, J. Alvarez and E. Authie, for the preparatory work on this scientific opinion and, the data providers A. Räber (Prionics), M. Pacciarini (IZSLER- Italy), S. Rhodes (AHLVLA- UK) S. Strain (Agri-Food and Biosciences Institute- NI-UK), J. Lawrence (Idexx) C. Whelan and J. Clarke (Enferplex) M. Barros (Asturias, Spain), J. Mourelo (Galicia, Spain), R. Diaz and Y. Soriano (Madrid, Spain), O. Mingues and S. Marques (Castilla y Leon, Spain), J. L. Saez (Spanish Ministry of Agriculture, Spain) and EFSA staff, A. Afonso, J. Richardson, D. Verloo and J. Cortinas Abrahantes, for the support provided to this scientific opinion.

Suggested citation: EFSA Panel on Animal Health and Welfare (AHAW); Scientific Opinion on the use of a gamma interferon test for the diagnosis of bovine tuberculosis. EFSA Journal 2012;10 (12):2975 [63 pp.] doi:10.2903/j.efsa.2012.2975. Available online: www.efsa.europa.eu/efsajournal

SUMMARY

Following a request from the European Commission, the Panel on Animal Health and Welfare (AHAW) was asked to deliver a scientific opinion on the use of a gamma interferon test (IFN-γ) for the diagnosis of bovine tuberculosis (bovine TB). The procedures for gaining, maintaining, suspending, withdrawing or re-gaining official bovine TB-free herd status and for certification for intra-Union trade are laid down in Council Directive 64/432/EEC and are based on the results of tuberculin skin tests. The skin test has a number of drawbacks, in terms of test characteristics, limited sensitivity at the level of the individual animal and testing logistics.

In this mandate, the first objective was to assess whether IFN-γ can be added to the official tests as a stand-alone test for demonstration of bovine TB-free herd status and testing for intra-Community trade. The IFN-γ should have a sensitivity equivalent or superior to the standard test currently used in the European Union and have a specificity not lower than that of the standard test with the lowest specificity used in the EU. Furthermore, there should be no foreseeable practical difficulties that could compromise test performance. In addition, any other tests that meet the above requirements should be identified. In the event that no new test could be recommended, the additional objective was to inform the Commission of further studies necessary to evaluate the suitability of the IFN-γ test or any other new test.

Three different data sources were considered: (1) a systematic literature review and meta-analysis of studies on the performance of bovine TB tests carried out by the Animal Health and Veterinary Laboratories Agency (AHVLA); (2) new publications on test performance that had become available since the original searches for the AHVLA review; and (3) data from a public data call. Data from the public data call were analysed by a Bayesian latent class model that enables estimation of sensitivity and specificity in the absence of a gold standard test.

The systematic literature review showed that the sensitivity of the IFN-γ test, based on standard purified protein derivative (PPD), was not significantly different from that of the comparator tuberculin skin test and, although the comparative skin test had the highest specificity, the specificity of the IFN-γ test was not significantly different from that of the standard skin test with the lowest specificity currently used in the EU, the single intradermal test (SIT). New publications did not change this overall picture. Latent class analysis demonstrated a higher sensitivity for the IFN-γ test than for the skin tests but at the cost of a lower specificity. Considerable differences in sensitivity and specificity across populations for both the IFN-γ and the skin test were observed, which may be explained by differences in the way tests are performed and interpreted and differences in the disease prevalence and distribution of stages of infection in the different populations. The probability that an animal or a herd is free of infection when a negative result is obtained from the IFN-γ test is at least as high as when a negative skin test result is obtained. However, considering the specificity estimates obtained, the probability that all animals in a herd will test negative in the IFN-γ test, given that they are free of bovine TB, may be lower than when skin tests are used for diagnosis of infection. The opinion also concluded that, within the EU Member States, the practical requirements for the performance of the IFN-γ test can be met. Information regarding the sensitivity and specificity of IFN-g test based on defined antigens is yet limited and the estimates of sensitivity and specificity of antibody tests are either very imprecise or based on only a few studies.

According to the definition of suitability given above PPD based IFN-γ tests can be included amongst the official tests for the purpose of granting and retaining an officially tuberculosis free herd status. However, the results from the latent class analysis from the public data call suggest that the specificity of the PPD based IFN-γ may not always be as high as the SIT test, the test with the lowest specificity currently used in the EU. The panel recommended that should the PPD based IFN-γ test be included in the official tests for the purpose of granting and retaining official TB-free status, the protocols for its use for this purpose should be harmonised in the EU.

Based on the reviewed information, other tests should not yet be considered for inclusion in the official tests for the purpose of granting and retaining official TB-free herd status.

Further evaluation of the suitability of the IFN-γ test should study the influence of factors such as the presence of environmental mycobacteria, prevalence of bovine TB in the herd, the age type and bovine TB test history of the animals all of which may affect test specificity and hence the suitability of the test for demonstrating freedom from bovine TB in different situations.

TABLE OF CONTENTS

Abstract 1
Summary 2
Table of contents 4
Background as provided by European Commission 5
Terms of reference as provided by European Commission 5
1. Introduction 7
2. Material and methods 8
2.1. Bovine TB diagnostic tests to be considered for evaluation 8
2.2. Data sources 9
2.2.1. Systematic literature review 9
2.2.2. Update from the literature review 9
2.2.3. Public data call 10
2.3. Analysis of data from public data call 10
3. Results 12
3.1. Systematic literature review and meta-analysis 12
3.2. Updated literature review 14
3.3. Analysis of data from public data call 14
4. Discussion 16
4.1. Systematic literature review 16
4.2. Updated literature review 16
4.3. Analysis of data from public data call 17
4.4. Negative predictive value 18
4.5. Probability that a truly bovine TB-free herd will be classified as TB free 20
4.6. Practical issues in the application of the IFN-γ test 20
4.7. Further validation studies on the IFN-γ test 21
4.8. Future perspectives 22
Conclusions and recommendations 23
References 25
Appendices 30
A. Excluded tests and reasons for exclusion 30
B. Diagnostic tests for bovine TB 33
C. Systematic literature review 38
D. Update of the literature review 41
E. Public data call 46
F. Data received – Public data call 48
G. Model details 54
Abbreviations 61
Glossary 63

BACKGROUND AS PROVIDED BY EUROPEAN COMMISSION

The procedures for gaining, maintaining, suspending, withdrawing or re-gaining the officially tuberculosis free (TBOF) herd status are laid down in Annex A to Council Directive 64/432/EEC of 26 June 1964 on animal health problems affecting intra-Community trade in bovine animals and swine ("the Directive") and are based on the results of tuberculin skin tests carried out in bovine herds. In addition Member States or regions thereof may be declared TBOF if certain requirements are fulfilled. Annex B to the Directive sets up details of the diagnosis of bovine tuberculosis (TB). This Annex was thoroughly reviewed in 2002 to incorporate new diagnostic methods and to further align EU requirements to the international standards of the World Organisation for Animal Health (OIE).

The tuberculin skin test in its various forms is the sole test prescribed in EU legislation. While this test has been an effective tool when applied at herd level, a lack of sensitivity at the individual animal level is recognised to be its limitation. An increase in the sensitivity (identifying fewer false test-negative but infected animals that remain in a herd) of the test is achievable by changing the cut-off point of the test. However, specificity (Sp) of the test could be lowered (more false test-positive but healthy animals are removed from a herd) when the test is interpreted in a more stringent way. Few new diagnostic tests have been developed in the last 30 years but the IFN test has showed its value to detect infected animals in a relatively accurate manner under certain circumstances. Therefore, the use of a gamma interferon (IFN-γ) test as an ancillary test carried out concurrently and in parallel to the tuberculin skin test, is currently regulated in Annex B(3) to the Directive. However, the IFN-γ may only be used to detect the maximum number of infected animals in the target population. This limited and strategic use of the IFN-γ test increases the sensitivity of the diagnostic regime. Optimised use of the parallel testing (tuberculin skin test and IFN-γ test) may allow the detection of 2 out of every 3 false tuberculin-negative infected animals that would otherwise be considered negative and thus not removed from the herd, if the tuberculin test alone had been used. In Europe in the last years IFN test has been used in the context of the EU co-financed eradication programmes and valuable information is now available on the performance of this test. The experience gained is not only limited to the use of the IFN-γ test using tuberculin PPD as antigen but also with other antigens such as ESAT-6 and CFP-10 or even other new antigen combinations. In view of the above, the Commission asks the European Food Safety Authority to assess the available scientific data, including the reports provided by the Member States on the outcome of the EU co-finance eradication programmes, and issue a scientific opinion on the suitability of the IFN-γ test to be included by EU legislation as a prescribed test for bovine TB diagnosis, to be used as an alternative to the tuberculin skin test for the establishment and maintenance of an officially tuberculosis free herd status and for certification for intra-Union trade in bovine animals.

TERMS OF REFERENCE AS PROVIDED BY EUROPEAN COMMISSION

The Commission asks the European Food Safety Authority:

1. to issue a scientific opinion on the suitability of the IFN-γ test for inclusion amongst the official tests for the purpose of granting and retaining an officially tuberculosis free herd status as laid down in Annex A to Directive 64/432/EEC and certification for intra Union trade in bovine animals as required in Article 6(2)(a) of that Directive

2. to issue a scientific opinion on the suitability of other, possibly newer, tests, if any, for their inclusion amongst the official tests for the purpose of granting and retaining an officially tuberculosis free herd status as laid down in Annex A to Directive 64/432/EEC and certification for intra-Union trade in bovine animals as required in Article 6(2)(a) of that Directive;

3. in the event of a negative opinion to point (1), to advise the Commission on which further validation studies are necessary to evaluate the suitability of the IFN-γ test, or any other new test, for inclusion amongst the official tests for the purpose of granting and retaining an officially tuberculosis free herd status as laid down in Annex A to Directive 64/432/EEC and certification for intra-Union trade in bovine animals as required in Article 6(2)(a) of that Directive.

ASSESSMENT

1. Introduction

Bovine TB is present in livestock and wildlife populations in a number of regions within the EU. According to EU Directive 64/432/EEC ("the Directive"), bovines can be moved from non-bovine TB-free regions to TB-free regions if they originate from herds that have been declared officially bovine-TB free and if the animals themselves have been tested bovine TB negative before transport. Testing needs to be done by an official test according to the Directive.

Currently, the diagnostic assay known as the intradermal or skin test is the only official test in the EU for the purpose of granting and retaining official TB-free herd status, as laid down in Annex A to the Directive, and certification for intra-Union trade in bovine animals, as required in Article 6(2)(a) of the Directive. Two test procedures are recognised, the single intradermal test (SIT) using purified protein derivative (PPD) bovine tuberculin (a preparation obtained from the heat-treated products of *Mycobacterium bovis*) and the single intradermal comparative cervical test (SICCT) which uses bovine and avian PPDs (avian PPD from *M. avium*). Both SIT and SICCT are performed by intradermal injection of the PPDs, bovine at one site or bovine and avian at adjacent sites, on the neck of the animal and the interpretation is based on observing, measuring and recording 72 hours after inoculation the nature and extent of any increase in skin thickness at the bovine PPD injection site. In the comparative test (SICCT) this response is compared to that observed 72 hours after injection of avian PPD. When the standard interpretation of the tests is applied, inconclusive reactors[4] must be subjected to an additional skin test at least 42 days after the previous one; if they are not negative in the second test, they are deemed positive. However, to maximise the sensitivity of the test a severe interpretation may be applied (particularly in high-prevalence areas), in which all inconclusive reactors in the first skin test are considered positive and removed for slaughter (Anonymous, 2006).

The skin test has a number of drawbacks, in terms of test characteristics, limited sensitivity at the level of the individual animal and testing logistics, as animals need to be examined on two occasions 72 hours apart. Consequently, consideration is being given to the possibility of other tests, in particular the IFN-γ test, being included as official tests for the above-mentioned purpose.

The OIE includes the skin test as the prescribed test for international trade and IFN-γ as the alternative test for international trade. This implies that the skin test is considered optimal and that IFN-γ can be used for international trade where there is mutual agreement between the importing and exporting country (OIE, 2012).

The probability that the herd is truly free of bovine TB, given negative test results, is influenced by a range of factors, including the prior probability of freedom from TB, the herd size, the number of animals sampled, the minimum within-herd prevalence of bovine TB-infected animals to be detected (should the herd be infected) and the sensitivity of the test used. In the case of bovine TB, it is conceivable that only a single infected animal is present in a herd, for example where there is a programme to eliminate infected animals from the herd or where a herd was recently infected. At low within-herd prevalence, and using whole-herd testing (as is the case for granting bovine-TB free status), the sensitivity of the test is the main determinant of the probability that a herd testing negative is truly free of bovine TB. For this reason, this opinion principally considers the diagnostic performance, particularly sensitivity, of tests for bovine TB in individual animals. Specificity is also

[4] The interpretation of the skin test is based on the observation of clinical signs and on the increase in skin-fold thickness at the site of inoculation. A SIT is positive, inconclusive or negative when the increase is greater than 4 mm, between 2 mm and 4 mm or less than 2 mm, respectively; the SICCT is positive, inconclusive or negative when the bovine injection site exceeds the avian site by greater than 4 mm, 1–4 mm and less or 1 mm, respectively.

considered because the rate of false-positive results is relevant with respect to retaining official TB-free herd status.

The current opinion does not include an assessment of the fitness, or lack thereof, of the diagnostic tests for the purpose of control and eradication.

In this mandate, the first objective is to assess whether the IFN-γ test could be added to the official tests as a stand-alone test for demonstrating freedom from bovine TB and with respect to intra-Community trade (ToR 1). After discussion with the European Commission it was agreed that, in order to achieve equivalence (suitability) as a stand-alone test, the IFN-γ test should meet the following requirements:

- have a sensitivity equivalent or superior to the standard test currently used in the EU, and
- have a specificity not lower than that of the standard test with the lowest specificity currently used in the EU.

Furthermore, there should be no foreseeable practical difficulties that could compromise test performance.

In addition, any other tests that meet the above requirements should be identified (ToR 2). In the event that no new test could be recommended, the additional objective was to inform the Commission of further studies required to evaluate the suitability of the IFN-γ test or any other new test (ToR 3).

2. Material and methods

2.1. Bovine TB diagnostic tests to be considered for evaluation

A systematic literature review reported in 2011 identified a large number of tests available for the diagnosis of bovine TB (VLA, 2011). The *ad hoc* working group (WG) considered the list of tests and selected for evaluation the tests included in Table 1. The tests were selected based on their suitability for use in large-scale surveys of live animals and the information available for their evaluation. They are described in Appendix B as well as the skin tests and the post-mortem tests currently used in the EU. The tests that were excluded and the reasons for exclusion are reported in Appendix A.

Table 1: Diagnostic tests for bovine TB considered for evaluation in this opinion

Test name[(a)]	Abbreviation	Long description
PPD-based IFN-γ		
IFN-γ bovine–avian	IFN-γ-BA	Gamma interferon test with bovine PPD and avian PPD diagnostic antigens
IFN-γ bovine	IFN-γ-B	Gamma interferon test with bovine PPD diagnostic antigen
Defined antigens-based IFN-γ		
IFN-γ CFP10 ESAT6	IFN-γ-CE	Gamma interferon test with CFP10 and ESAT6 diagnostic antigens
IFN-γ MPB70	IFN-γ-MPB	Gamma interferon test with MPB70 diagnostic antigen
IFN-γ BACE	IFN-γ-BACE	Gamma interferon test with bovine PPD and avian PPD diagnostic antigens and CFP10 and ESAT6 diagnostic antigens
Antibody detection tests		
ELISA bovine–avian	ELISA-BA	Enzyme-linked immunosorbent assay with bovine PPD and avian PPD diagnostic antigens
ELISA bovine	ELISA-B	Enzyme-linked immunosorbent assay with bovine PPD diagnostic antigen
ELISA MPB70	ELISA-MPB	Enzyme-linked immunosorbent assay with MPB70 diagnostic antigen

Multiplex immunoassay	Multiplex	Enzyme-linked immunosorbent assay based on the use of multiple mycobacterial antigens
Latex bead agglutination assay	LBBA	Latex bead agglutination assay using defined antigens (ESAT-6 and MPB70)
Serological rapid	Rapid	Rapid immunochromatographic assay using defined antigens (MPB83, CFP10/ESAT-6)

(a) The tests have been classified regarding the mechanism of detection test and their antigen composition. It is important to notice that different cut-off values are used for the interpretation of the test results. Currently the IFN-γ assay kit (different antigens can be used with the same basic test kit) the serological rapid test (Rapid) and the multiplex immunoassay (Multiplex) are commercially available.

2.2. Data sources

Data sources used for this opinion were: (1) a systematic literature review and meta-analysis carried out by the Animal Health and Veterinary Laboratories Agency[5] (AHVLA) together with scientists and veterinarians from other research groups and government agencies (Downs et al., 2011; VLA, 2011); (2) new publications about diagnostic test performance that had become available since the original searches for the AHVLA review; and (3) data from a public data call.

Reports provided by the Member States on the outcome of the EU co-financed eradication programmes were evaluated as a possible data source but were not helpful in responding to the mandate owing to a lack of detail on important variables such as test procedures or number of available tests (see Section 2.3 and Appendix E for data requirements).

2.2.1. Systematic literature review

A systematic review of the literature and meta-analysis of the performance of diagnostic tests for bovine TB in cattle was reported in 2011 (Downs et al., 2011; VLA, 2011). This review comprised comprehensive searches of the peer-reviewed and grey literature for relevant studies and a structured approach to the assessment and extraction of data relating to diagnostic test performance. Further details about the methodology of the systematic review of relevant literature and subsequent meta-analysis can be found in Appendix C and the full report (VLA, 2011).

It was agreed that the findings from the AHVLA review should be considered within the response to the mandate. However, for this mandate we considered only studies that comprised the diagnostic tests included in Table 1 and studies in which at least two diagnostic tests (one of which could be used as a reference standard) were used, whereas studies were eligible for the evaluation of specificity in the AHVLA review in which a test was evaluated in a bovine TB-free population, even if another diagnostic test had not been used as a reference standard.

2.2.2. Update from the literature review

Searches were conducted to identify any new data on diagnostic test performance that had become available since the original searches for the AHVLA review. Relevant new results were considered within the response to the mandate but were not incorporated in the meta-analysis previously performed.

The search of electronic databases for the AHVLA systematic review was last performed on 1 December 2008. The update was conducted by EFSA on 13 March 2012 using the same search string without language restrictions on Web of Knowledge (which simultaneously searches Web of Science 1995–, Current contents 1998–, CAB abstracts 1910–, MEDLINE 1950–). Details of the searches and the methodology of the review are presented in the Appendix D.

[5] Formerly known as the Veterinary Laboratories Agency (VLA).

2.2.3. Public data call

To maximise the amount of information available on bovine TB tests, EFSA launched a public data call (Appendix E). Criteria for inclusion or exclusion of data were set. The call was open from 26 March to 26 April 2012. A summary of all received datasets and relevant variables for its evaluation is included in Table 8, Appendix F.

2.3. Analysis of data from public data call

Data from the public data call were analysed by a Bayesian latent class model that enables estimation of sensitivity and specificity in the absence of information on the true bovine TB status of an animal (Toft et al., 2005). The model parameters estimated were: (1) the prevalence in the populations (datasets) considered; (2) the sensitivity and specificity of the various tests in the different populations; and (3) conditional dependence between tests. At a minimum, latent class analysis using results from two tests requires data from two populations with varying prevalence. Unfortunately, owing to the variability of the test protocols (differences in antigen used and cut off-values) it was not possible to combine datasets originating from different countries. This meant that data from a single population sent in response to the call could only be included if at least three test results were available from the population.

The datasets that met these criteria are summarised in Table 2 and they all refer to testing with IFN-γ-BA. Tuberculin skin test (either SIT or SICCT) and different post-mortem tests were used as second and third tests.

Only for the Northern Ireland data sets (ID 3 to 9) were post-mortem results available for all animals in the datasets. For the other studies post-mortem results were missing for animals that were negative for both the IFN-γ and skin test and on some occasions for animals in which both tests were positive or the IFN-γ test was positive and the skin test negative.

Three different scenarios were considered to simulate the missing data:

- Scenario I: The post-mortem result is 98 % negative when the IFN-γ and skin test are negative. The proportions observed in the available data are maintained for the cases when both tests are positive or the IFN-γ test is positive (Table 11, Appendix G). The WG members considered this as the most likely scenario.

- Scenario II: The post-mortem result is in agreement with the IFN-γ test result.

Scenario III: The post-mortem result is in agreement with the skin test result.

The outcomes of the IFN-γ and skin test were considered to be dependent because they both target the cell-mediated immune response, whereas the post-mortem result was assumed to be conditionally independent. In the Bayesian modelling non-informative priors were used. In order to assess the model fit, the predicted frequencies for each of the combination of results for the three tests and the observed frequencies were plotted. The convergence diagnostic proposed by Brooks and Gelman (1998) was used, indicating that convergence was reached in all models. Details of the model are presented in Appendix G.

Table 2: Datasets originating from the public call for data included in the analysis

ID		Year	Country	Cut-off for IFN-γ-BA[a]	Skin test[b]	Post mortem	N[c]
1	Raw data CAM 2010-2011.xlsx	2010	Spain	$PPDb_{OD} - NIL_{OD} > 0.05$ and $PPDb_{OD} > PPDa_{OD}$	SICCT severe	Culture	2 449
2	Raw data CAM 2010-2011.xlsx	2011	Spain	$PPDb_{OD} - NIL_{OD} > 0.05$ and $PPDb_{OD} > PPDa_{OD}$	SICCT severe	Culture or macroscopic lesions	3 690
3	Prob_Net_IFNg	2004	Northern Ireland	$PPDb_{OD} - NIL_{OD} \geq 0.1$ and $PPDb_{OD} - PPDa_{OD} \geq 0.05$	SICCT standard	Culture or macroscopic lesions	367
4	Prob_Net_IFNg	2005	Northern Ireland	$PPDb_{OD} - NIL_{OD} \geq 0.1$ and $PPDb_{OD} - PPDa_{OD} \geq 0.05$	SICCT standard	Culture or macroscopic lesions	1 653
5	Prob_Net_IFNg	2006	Northern Ireland	$PPDb_{OD} - NIL_{OD} \geq 0.1$ and $PPDb_{OD} - PPDa_{OD} \geq 0.05$	SICCT standard	Culture or macroscopic lesions	905
6	Prob_Net_IFNg	2007	Northern Ireland	$PPDb_{OD} - NIL_{OD} \geq 0.1$ and $PPDb_{OD} - PPDa_{OD} \geq 0.05$	SICCT standard	Culture or macroscopic lesions	625
7	Prob_Net_IFNg	2008	Northern Ireland	$PPDb_{OD} - NIL_{OD} \geq 0.1$ and $PPDb_{OD} - PPDa_{OD} \geq 0.05$	SICCT standard	Culture or macroscopic lesions	1 304
8	Prob_Net_IFNg	2009	Northern Ireland	$PPDb_{OD} - NIL_{OD} \geq 0.1$ and $PPDb_{OD} - PPDa_{OD} \geq 0.05$	SICCT standard	Culture or macroscopic lesions	1 469
9	Prob_Net_IFNg	2010	Northern Ireland	$PPDb_{OD} - NIL_{OD} \geq 0.1$ and $PPDb_{OD} - PPDa_{OD} \geq 0.05$	SICCT standard	Culture or macroscopic lesions	1 516
10	Raw data CAM 2010-2011.xlsx	2010	Spain	$PPDb_{OD} - NIL_{OD} > 0.05$ and $PPDb_{OD} > PPDa_{OD}$	SIT severe	Culture	3 649
11	Raw data CAM 2010-2011.xlsx	2011	Spain	$PPDb_{OD} - NIL_{OD} > 0.05$ and $PPDb_{OD} - PPDa_{OD}$	SIT severe	Culture or macroscopic lesions	3 873
12	120326ax1_Ireland(GormleyE)_TC_1st August2012.xlsx	2008	Ireland	$PPDb_{OD} > 0.1$, $PPDb_{OD} - NIL_{OD} > 0.05$ and $PPDb_{OD} > PPDa_{OD}$	SICCT standard	Macroscopic lesions	2 740
13	120326ax1_Ireland(GormleyE)_TC_1st August2012.xlsx	2008	Ireland	PPDbO > 0.1, $PPDb_{OD} - NIL_{OD} > 0.05$ and $PPDb_{OD} > PPDa_{OD}$	SICCT standard	Macroscopic lesions	2 197

(a) Cut off values for interpretation of IFN-γ-BA, the different optical density (OD) readings obtained after the stimulation with each antigen (bovine PPD/avian PPD/PBS) are used to yield a quantitative result: OD obtained after stimulation with PBS (NIL_{OD}) is often subtracted from the OD observed after stimulation with bovine PPD ($PPDb_{OD}$) and avian PPD ($PPDa_{OD}$).

(b) Skin test type and interpretation: severe interpretation in which all inconclusive reactors in the first skin test are considered as positive; and standard in which inconclusive reactors are retested.

(c) Number of records (individual animals).

3. Results

3.1. Systematic literature review and meta-analysis

The performance of the diagnostic tests identified by EFSA as relevant to the mandate that were also estimated in the meta-analysis, conducted by AHVLA and others, is summarised below and in Table 3. Further results can be found in the full report (VLA, 2011).

Estimates of the sensitivity of the diagnostic tests from the meta-analysis had wide credible intervals.[6] In general the credible intervals were narrower for the IFN-γ tests than for the skin test, but there was considerable overlap of the credible intervals between test-types. The sensitivities of IFN-γ-B, IFN-γ-BA and IFN-γ-CE were not significantly different from the sensitivities of SIT and SICCT. IFN-γ using MPB70 had significantly lower sensitivity than both skin tests.

Although the sensitivity of the ELISA tests were also not significantly different from those of the skin tests, the credible intervals were wider than those of the IFN-γ and skin tests, indicating that the variability of the sensitivity across studies is very high. The median sensitivity of the LBBA (0.91) was among the highest of all the tests and the distribution of the credible interval was narrow; however, this was based on the results of only two studies from one research group. There were no eligible data with which to estimate the sensitivity of the rapid test.

The SICCT test at standard interpretation had the highest median specificity of all the diagnostic tests under evaluation and its specificity was significantly higher than that of all the IFN-γ tests except IFN-γ-CE. The median specificity of SIT was lower than that of SICCT and the IFN-γ tests, but the credible interval of SIT was wide and overlapped with those of the other tests.

Specificity distributions for the ELISA tests were wider than those estimated for the IFN-γ tests and the SICCT, and the median estimates were slightly lower. The estimates for the specificity of both the LBBA and the rapid test had wide credible intervals, and the median specificities were lower than those for SICCT test.

Estimates of sensitivity and specificity were weighted in the modelling procedure to account for the varying number of estimates within references for the same test type and population. Adjustment was made for confounding factors and a random effect term was incorporated to account for clustering of errors within references (see Appendix C and VLA, 2011 for further detail).

[6] A 95% Bayesian credible interval states that the estimated probability that the process used to generate the interval includes the correct value of the parameter is 95%.

Table 3: Summary of meta-analysis results for sensitivity and specificity of diagnostic tests for bovine TB on cattle from AHVLA systematic review (VLA, 2011)

Test Name	Sensitivity					Specificity				
	Reference (a)	Estimates(b)	Overall adjusted estimate (c)			References (a)	Estimates(b)	Overall adjusted estimate (c)		
	N	*n*	P025	Median	P975	*N*	*n*	P025	Median	P975
Skin test										
SIT (cervical)	7	16	0.49	0.94	1.00	4	10	0.70	0.91	1.00
SICCT severe(d)	25	57	0.37	0.61	0.82	0	0			
SICCT standard(e)	25	57	0.27	0.49	0.74	7	13	0.99	1.00	1.00
PPD-based IFN-γ										
IFN-γ-B(e)	27	166	0.72	0.87	0.95	19	137	0.94	0.97	0.98
IFN-γ-BA(e)	27	166	0.49	0.67	0.82	19	137	0.96	0.98	0.99
Defined antigen-based IFN-γ										
IFN-γ-CE(e)	27	166	0.61	0.79	0.91	19	137	0.99	0.99	1.00
MPB70(e)	27	166	0.04	0.1	0.25	19	137	0.85	0.94	0.98
Antibody detection tests										
ELISA-B-PPD(e)	22	59	0.06	0.76	0.99	12	27	0.80	0.90	0.95
ELISA-B-PPD-A-PPD(e)	22	59	0.01	0.36	0.97	12	27	0.82	0.93	0.98
ELISA-MPB(e)	22	59	0.01	0.20	0.94	0	0			
LBBA	2	3	0.60	0.91	0.98	1	1	0.39	0.94	1.00
Multiplex	1	5	0.31	0.74	0.95	1	4	0.34	0.88	0.99
Rapid	0	0				2	3	0.66	0.97	1.00
Post mortem										
Meat inspection	6	11	0.38	0.71	0.92	1	3	0.99	1.00	1.00
Detailed necroscopy in laboratory(f)	6	11	0.82	0.96	1.00	0	0			
Culture of *M. bovis*	8	16	0.46	0.74	0.94	1	1	0.73	0.99	1.00

(a) The number of references with at least one estimate of either sensitivity or specificity.
(b) The number of estimates used in the modelling of sensitivity or specificity. This does not equal the number of references because a reference could contain more than one estimate.
(c) Median and Bayesian 95 % credible interval.
(d) Severe interpretation in GB: 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.
(e) Standard interpretation in GB: Reaction to bovine tuberculin is both positive and exceeds the reaction to avian tuberculin by more than 4 mm.
(f) Includes inspection for macroscopic lesions typical of *M. bovis* infection but does not include microscopic examination.

3.2. Updated literature review

The update to the literature review identified 938 research studies published up to March 2012 that had not been included in the systematic literature review (VLA, 2011). After relevance screening, a total of 15 studies was reviewed in order to collect information regarding sensitivity and specificity of bovine TB tests (Table 7, Appendix D). As limited resources did not permit a new meta-analysis including all identified studies (Sections 3.1 and 3.2), it was decided to present only a description of the results reported in the relevant studies.

Sensitivity estimates of IFN-γ-BA were in line with the estimates from the systematic literature review (Alvarez et al., 2009; Clegg et al., 2011) or higher (with point estimates > 0.82 but overlapping confidence intervals: Marassi et al., 2011; Antognoli et al., 2011; Faye et al., 2011; Alvarez et al., 2012). The specificity values were in general lower, with a strong influence of the cut-off value in place in each of the studies. The only study reporting values of IFN-γ with other antigens (ESAT6-CFP10) yielded estimates for both sensitivity and specificity similar to those of the IFN-γ-BA (Faye et al., 2011)

Multiplex assays had high sensitivity values when performed on animals with gross pathology typical of bovine TB (Whelan et al., 2010, 2011), but showed a more limited ability to detect infected animals when estimated using a latent class analysis (0.34–0.72, depending on the cut-off value used (Clegg et al., 2011), which was in agreement with what was obtained in the systematic literature review.

Sensitivity values obtained with other serological techniques (ELISA and lateral flow assays using different antigens from members of the *M. tuberculosis* complex) were in agreement with estimates provided by the systematic literature review. A large variation was observed in the sensitivity estimates reported (median estimates being from a minimum value of 0.344 up to a maximum value of 0.879), which was related to the different antigens/cut-off/comparator test used in each of the studies. When available, specificity estimates of antibody assays were equivalent to or higher than those found in the systematic literature review, although in the case of the Multiplex these were highly influenced by the cut-off value used.

3.3. Analysis of data from public data call

The estimates of the sensitivity and specificity of IFN-γ-BA from the Bayesian latent class analysis are summarised in Table 4. Sensitivity, specificity, covariance and the agreement measure (kappa) for all countries (Scenario I) are shown in Figures 3–10 in Appendix G.

The lower limits of the credible intervals for the sensitivity of IFN-γ-BA, taking Scenario I for the Irish and Spanish data, were higher than the upper limits of the credible intervals for SIT and SICCT, indicating that IFN-γ-BA had a significantly higher sensitivity. The upper limits of the credible intervals of the specificity of IFN-γ-BA were consistently lower than the lower limits of the credible intervals for SIT and SICCT, indicating significantly lower specificity. The differences between the IFN-γ-BA sensitivity estimates from different countries were limited. For specificity, however, the credible interval of the estimates obtained from the Northern Ireland dataset were markedly lower. The estimates resulting from the Spanish data show low sensitivity for SIT and SICCT in comparison with Northern Ireland and Ireland and similar specificity estimates for SIT and SICCT.

Scenarios II and III showed similar results to Scenario I regarding the comparison of IFN-γ and skin tests, indicating that the influence of missing post-mortem data is of minor importance. Only when applying Scenario II to the Irish data did the credible intervals for the specificity of IFN-γ-BA and the skin test grossly overlap, indicating a non-significant difference. Conditional dependency between IFN-γ-BA and the skin test was low.

In the datasets the largest estimated prevalence was approximately 0.4 (0.3707 – 0.4240) and the lowest 0.0034 (0.0005 – 0.0085).

Table 4: 95 % Credible intervals (upper and lower limits) for sensitivity, specificity and conditional dependency as obtained by Bayesian latent class analysis

ID			Sensitivity			Specificity			Conditional dependency between IFN-γ-BA and skin test	
			IFN-γ-BA	Skin test	Post mortem	IFN-γ-BA	Skin test	Post mortem	Infected class	Non-infected class
3 to 9	Northern Ireland SICCT standard		0.885; 0.936	0.525; 0.608	0.543; 0.632	0.661; 0.691	0.964; 0.975	0.988; 0.997	–0.034; –0.009	–0.010; –0.005
1-2	Spain SICCT Severe	Scenario I Most likely	0.864; 0.998	0.298; 0.426	0.620; 0.871	0.918; 0.941	0.988; 0.996	0.977; 0.987	–0.009; 0.036	–0.0003; 0.005
		Scenario II IFN-γ and post-mortem agreement	0.978; 0.998	0.278; 0.379	0.761; 0.995	0.925; 0.948	0.988; 0.996	0.9993; 1	–0.011; 0.0002	-0.0002; 0.005
		Scenario III Skin and post-mortem agreement	0.976; 0.997	0.287; 0.383	0.761; 0.99	0.938; 0.956	0.987; 0.994	0.9995; 1	–0.013; 0.000	0.00; 0.006
10-11	Spain SIT Severe	Scenario I	0.765; 0.996	0.304; 0.460	0.665; 0.981	0.926; 0.947	0.986; 0.995	0.977; 0.992	–0.009; 0.065	0.00; 0.006
		Scenario II	0.976; 0.997	0.287; 0.383	0.761; 0.995	0.938; 0.956	0.987; 0.994	0.9995; 1	–0.013; 0.000	0.00; 0.006
		Scenario III	0.967; 0.996	0.405; 0.525	0.720; 0.988	0.925; 0.941	0.988; 0.995	0.9995; 1	–0.014; 0.001	–0.003; 0.005
12-13	Ireland SICCT standard	Scenario I	0.791; 0.861	0.645; 0.730	0.359; 0.423	0.882; 0.908	0.993; 0.9996	0.974; 0.985	–0.041; 0.0003	0.000; 0.002
		Scenario II	0.846; 0.889	0.480; 0.532	0.438; 0.489	0.996; 1	0.997; 1	0.995; 1	–0.045; –0.024	0.0001; 0.006
		Scenario III	0.788; 0.845	0.651; 0.730	0.351; 0.417	0.878; 0.905	0.993; 0.9996	0.995; 0.9991	–0.030; 0.00	0.0001; 0.005

4. Discussion

4.1. Systematic literature review

The results indicate that the IFN-γ test, as used in the eligible studies, had a sensitivity similar to or higher than the most comparable skin tests. Moreover, although the specificities of the IFN-γ test, apart from IFN-γ-CE, were lower than that of the SICCT test, the median estimates for the IFN-γ test were higher than the median estimates for the specificity of the SIT and the estimates were not statistically significantly different.

Setting aside the results of the IFN-γ test using MPB70, there was no evidence that the performance of the ELISA tests could match that of the IFN-γ test. The estimate for the sensitivity of LBBA was comparable to those for the skin tests and the IFN-γ tests but the data for this test were derived from only two studies.

The wide credible intervals, particularly for the sensitivity estimates, were due to relatively few studies with eligible data, heterogeneity as a result of the variety of cut-off values used to classify a positive response in the blood tests (despite the use of a counter-parameter) and the influence of covariates that were controlled for within the analysis. At the outset of the systematic review a range of potentially influential factors was identified because it was recognised that test performance may vary in different subgroups such as calves and adult cattle, dairy cattle and other cattle. Information, if available, about potentially influential confounding factors such as country of study, year of study, sampling strategy, possible cross-reactivity with environmental mycobacteria, type of animal production system and type of reference standard was extracted from the eligible references in addition to test performance data.

Inclusion criteria for the review required that all the studies from which sensitivity was estimated were from cattle populations naturally exposed to *M. bovis*, and all the studies from which specificity was estimated were from cattle populations reported as officially TB free or reported as free and having been free for several years. This meant that the estimates of sensitivity and specificity were derived from different populations. Furthermore, the estimates of sensitivity relative to the reference test used to classify an animal as truly infected would have been influenced by the accuracy of the reference test and the stage of the disease in the population sample. The probability of misclassification should be lower in the estimation of specificity because the populations on which specificity was measured were selected to be exposure and infection free.

4.2. Updated literature review

The existing differences in the protocols used in the performance of the IFN-γ-BA complicate the joint interpretation of the results of the studies reviewed. The effect of the different cut-off values, as well as other differences in the protocols used in each study (different gold standards, collection of blood samples for the assay 3–10 days after a SICCT test), may explain the width of the confidence intervals reported. The variability in the specificity, ranging from 0.8484 to 0.99, is most likely also affected by the cut-off value applied.

The use of defined antigens (CFP10-ESAT6) provided high sensitivity and specificity, comparable to the specificity estimates obtained for IFN-γ-BA.

Information on the performance of the serological assays (multiplex, ELISAs detecting antibodies elicited against MPB70–83, MPT-51, Ag85 and BCG antigens and the lateral flow assay) was available for samples from two different origins. When performed on samples from animals with macroscopic lesions (and often following isolation of *M. bovis*), consistent with bovine TB, i.e., those animals most likely to be in a progressive stage of infection, or

when the skin test was the only reference test used to define an animal as a true positive, serological assays were able to detect a large proportion of infected animals (> 80 %) when the most sensitive cut-off values were applied. However, all serological tests showed lower sensitivities when a larger sample population was analysed (including animals from which *M. bovis* was isolated but did not present macroscopic lesions). The latter results may provide a more accurate representation of the potential usefulness of these tests when they are used in populations subjected to eradication programmes in which a large proportion of infected animals would be in the early stages of infection.

4.3. Analysis of data from public data call

The results for the different datasets and scenarios show that IFN-γ-BA had higher sensitivity and lower specificity than the skin test. Although the trend is consistent with the results of the systematic literature review, the specificity of IFN-γ-BA was lower here than in the AHVLA literature review, which may (besides differences in the way the tests were performed and interpreted) be explained by the different populations.

The population sampling strategy and prevalence of bovine TB and other environmental mycobacteria in the herd can influence measured test performance (Aagaard et al., 2010; Farnham et al., 2012) and has implications for the extrapolation of the results to other populations. In the meta-analysis specificity was estimated entirely from cattle populations which the authors of the studies had reported as officially TB free or free from infection for several years. This inclusion requirement is likely to have reduced the probability of misclassification of truly infection free cattle. By comparison, the test performance from the data call was estimated from populations in which bovine TB was endemic and was based on comparisons across diagnostic tests of varying accuracy. In addition, the results reported from the meta-analysis were adjusted for a census-based or random population sampling strategy and absence of cross-reactivity with other mycobacteria, based on information available from the reviewed papers. The selection of the cattle populations, the bovine TB testing history of the cattle and the prevalence of environmental mycobacteria (that may infect cattle and elicit responses to PPD), may have influenced performance estimates in the surveillance population samples used in the latent class analysis.

The estimates for specificity obtained from the Northern Ireland dataset are lower that the ones obtained from other datasets. These results are a reflection of the large proportion of positive IFN-γ test results in which both skin test and post-mortem results were negative. The datasets were validated by the data provider and all model checks made, including a test on the possible effect of mixing results from different years. It is difficult to explain because the cut-off values used in Northern Ireland should not have resulted in lower specificity when compared with the results from Ireland where less specific cut-off values are used. Possible explanations for the observed results may be technical issues on the execution of the test (such as time of blood collection) and differences in the cohorts of animals under testing.

The probability that a bovine TB-infected animal is detected will, to a large extent, depend on the stage of the infection, and this applies to all the available bovine TB tests as they target different subpopulations of infected animals, not always overlapping (Pollock et al., 2005). The distribution of this probability across the various infection stages is different for the various tests. When test-positive animals are being removed from a population, the distribution of infected animals across the different infection stages will change and, with that, the sensitivity of the test. This may explain the low sensitivity of the skin test in the Spanish dataset. In these herds the skin test had been performed frequently and, consequently, most of the skin test reactors might have already been removed in previous tests and most of the infected animals remaining in the herds would be in the early stages of infection, when the skin test is known to have more limited sensitivity (Monaghan et al., 1994). Interestingly, specificities of the SIT (Spain) and the SICCT (all three countries) were very similar. In Spain

this could likely be due to a lack of cross-sensitization of cattle by environmental non-tuberculous mycobacteria, which is measured by the response to injection with avian PPD. A low level of cross-sensitization will limit the difference in the test specificity between the SIT and SICCT. In Ireland and the UK, where cross-sensitization is relatively high, the SICCT is routinely used in order to maintain the high specificity of the test, while this test is only used in Spain in OTF herds in which a possible bovine TB infection has been excluded by post-mortem analysis and epidemiological evidence.

The low conditional dependency between the IFN-γ and skin test suggests that, although both tests detect a cell-mediated immune response (Pollock et al., 2005), they target different subpopulations of sensitised lymphocytes (Neill et al., 1994, Pollock et al., 2005, Gormley et al., 2006). Consequently, using both tests in combination is expected to increase sensitivity (parallel use) or specificity (serial use). Serial use (whereby positive test results are confirmed by a second test) is of interest for the purpose of this mandate. As an example, if we take the median test characteristics of IFN-γ-B and SIT shown in Table 3 and assume conditionally independent serial testing, specificity is as high as 0.997, whereas sensitivity is 0.818 (higher than the median estimates of the sensitivity of SICCT or IFN-γ-BA).

4.4. Negative predictive value

When importing animals from a bovine TB-infected region, the negative predictive value (NPV) is of primary importance for the importing country. The NPV) is the probability that a test-negative animal is free from bovine TB. The NPV depends on the accuracy of the test and the prevalence of infected animals in the population. Figure 1(a) and (b) show the NPV for an individual animal testing negative for a prevalence ranging from 0.0 to 0.04 and using the sensitivity and specificity (excluding the value from Northern Ireland) ranges for skin tests and the IFN-γ test shown in Table 4. In the case of a negative test result, and assuming the same prevalence, one can be more certain of having a bovine TB-free animal when using the IFN-γ test than when the skin test is used. In cases in which more animals are being tested, the probability that at least one of them is positive also increases (Figure 1(c)–(f), although to a lesser extent with the IFN-γ test than with the skin test.

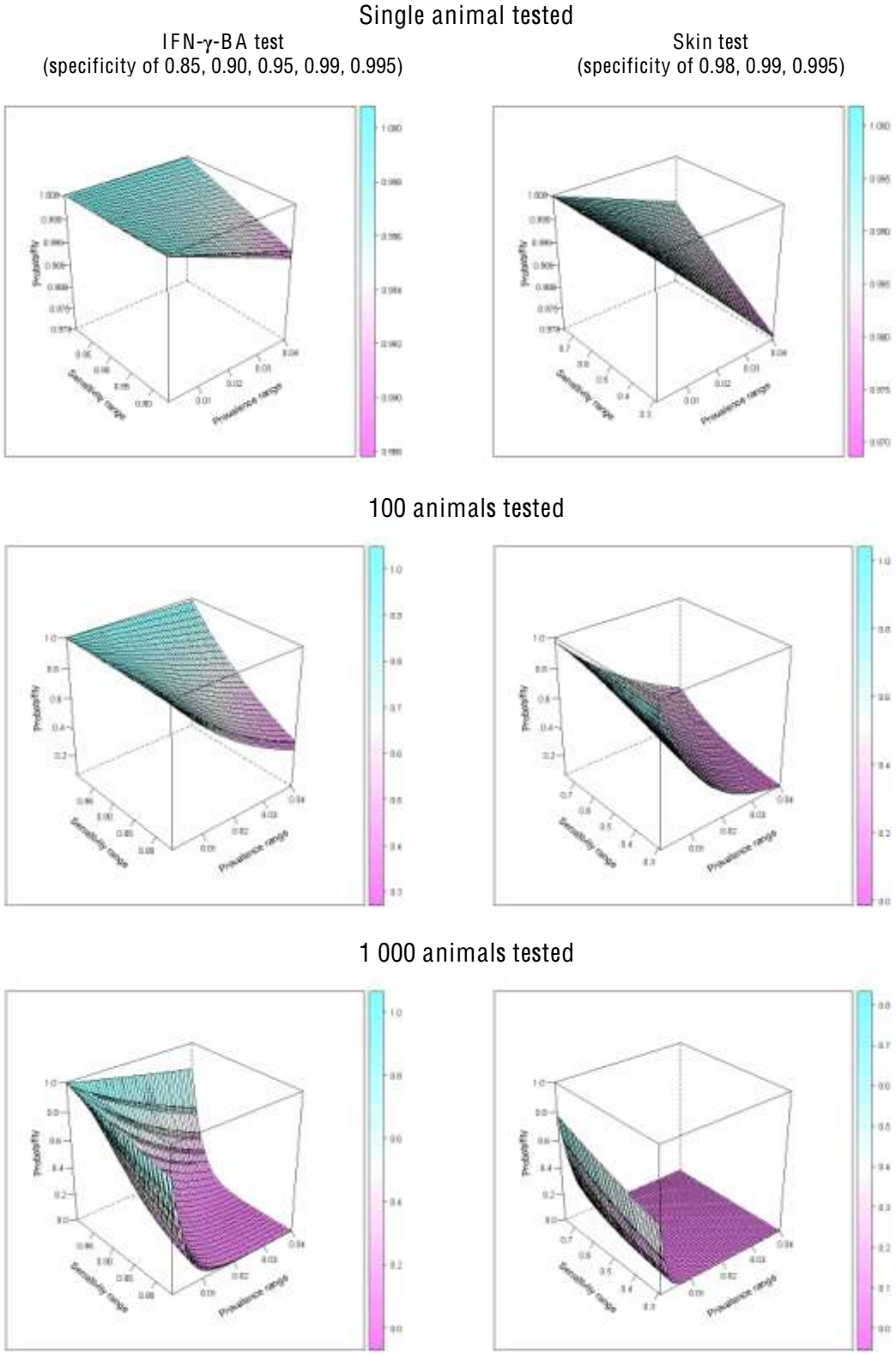

Figure 1: IFN-γ and skin test NPVs when a single animal (a, b), 100 animals (c, d) and 1 000 animals (e, f) are tested. The specificity is represented by layers in the plane.

4.5. Probability that a truly bovine TB-free herd will be classified as TB free

For a herd to be granted bovine TB-free status, all animals will need to test negative during repeated herd tests. Figure 2 shows the probability that all animals in a herd will test negative, given that they are free of bovine TB. At the specificity estimates shown in Table 4, the probability that a herd will not have a single positive result decreases sharply as herd size increases. When the specificity of the test is 98 % (median estimate from the AHVLA for IFN-γ-BA), there is only a 13.3 % probability in a 100-animal herd that all animals will test negative.

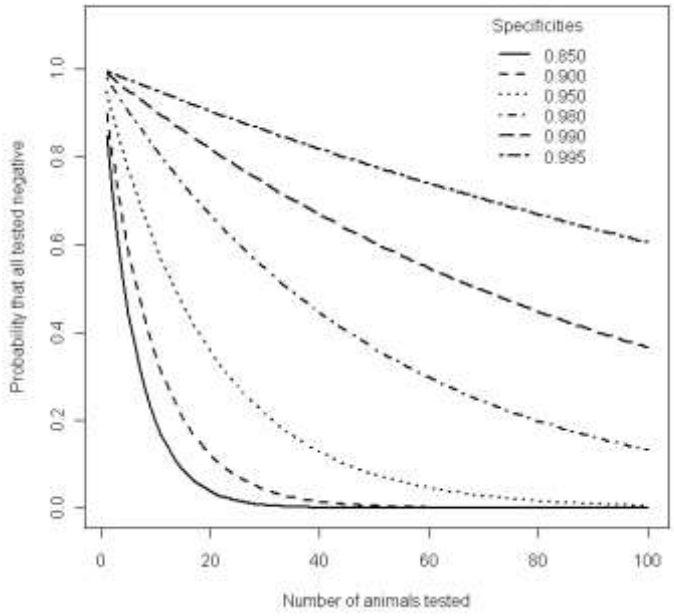

Figure 2: Probability, within a bovine TB-free herd, that all animals will test negative for the number of animals tested and test specificity ranging from 0.850 to 0.995.

4.6. Practical issues in the application of the IFN-γ test

The application of the IFN-γ assay to detect *M. bovis* infection in cattle offers several practical advantages over the skin test. The problems associated with skin testing of animals, including poor facilities, poorly calibrated equipment and the potential for fraud, are reduced when performing the blood-based test. In addition, as it does not require a second farm visit to read the test, this can have a beneficial impact on workload, the safety of personnel and the welfare of animals through reduced handling and minimising stress levels. The IFN-γ assay can be repeated in the laboratory and good laboratory practice (GLP) can readily be applied to ensure accuracy and reliability. In infected herds (containing reactors already disclosed by tuberculin tests) the test can be applied in different ways, depending on the suspected level of infection in the herd.

The test is robust and relatively easy to standardise across different laboratories. The adoption of interpretation criteria based on quantitative values allows for more objective interpretation of the results in contrast to the more subjective interpretation of skin test results. Another advantage of the assay procedure used is that the interpretation criteria can be adjusted to take account of local conditions; this can serve to maximise the detection of infected animals depending on the prevailing circumstances (e.g., in low-incidence areas the interpretation can be adapted to maximise the specificity of the test). The test also allows the incorporation of

defined antigens to enhance the specificity of the test (Buddle et al., 2001, 2003; Waters et al., 2004; Schiller et al., 2010a). Towards the completion of an eradication campaign, the interpretation can be adjusted to optimise the specificity of the test.

There are, however, some disadvantages to the use of the IFN-γ assay. Each of the two stages of the test requires specific equipment. As a result, there can be significant set-up costs involved in carrying out IFN-γ diagnostic tests to GLP standards. In the first stage of the test, the time between collection of the samples and their processing (overnight incubation with antigens) in the laboratory differs between countries and can range from less than 8 hours up to 24 hours post collection, depending on the use to which the test is being put. This can lead to a decrease in the test signal, increasing its specificity but also compromising its sensitivity under certain conditions (Gormley et al., 2004; Coad et al., 2007; Waters et al., 2007; Schiller et al., 2009). Tuberculin potency may be different and relevant differences may lead to difficulties in comparing the sensitivity/specificity of the test across countries (Whipple et al., 2001; Gormley et al., 2006; Schiller et al., 2010b). When a laboratory is fully equipped, the running costs are primarily associated with the consumables required to perform the test, equipment maintenance/servicing and personnel time. The high costs involved may be offset, however, by the fact that there is no need to visit the farm on more than one occasion. As all EU Member States possess the basic necessary infrastructure required to conduct IFN-γ diagnostic assays, there are few, if any, impediments to conducting IFN-γ tests to GLP standards.

Because of the varied and complex nature of the immune response of cattle to infection with *M. bovis* and the fact that *M. bovis*-infected cattle are being detected at a much earlier stage of the disease than formerly, some difficulties in diagnosis can be expected to arise with laboratory-based tests. One factor that might influence the performance of the IFN-γ test is the effect of a prior tuberculin test. Blood for analysis may be collected from animals prior to skin testing or after the skin test is performed (usually the day the test is read, 72 hours after intradermal inoculation of PPDs). This can have an immune modulator effect in terms of the specific IFN-γ release that may be reflected in the test outcome (Whipple et al., 2001; Gormley et al., 2004; Whelan et al., 2004; Schiller et al., 2010c).

4.7. Further validation studies on the IFN-γ test

IFN-γ has predominantly been used as an ancillary test to increase the sensitivity of the testing protocol in order to eliminate infected animals from herds and obtain TB-free herd status. Consequently, there has been limited use of the test as a stand-alone test for the purpose of granting or retaining TB-free status. Nevertheless, data used to estimate specificity in the AHVLA study originated from bovine TB-free regions, indicating that the estimates are valid in that situation. In the AHVLA systematic literature review the specificity estimate of IFN-γ-BA is based on results from 19 different studies, and the credible intervals on the estimates are narrow. It seems that little would be gained from further similar validation studies.

Studying critical factors that may affect the specificity of IFN-γ in different situations could be useful, given the variation in test performance observed in the surveillance populations and the indication of differences in test performance between PPD-based antigens and defined antigens such as ESAT6 CFP10. Specificity may differ from one region to another, for example owing to differences in the distribution of environmental mycobacteria and between cattle populations.

To optimise use of the IFN-γ test in bovine TB-free populations, an optimal cut-off value could be derived from receiver operating characteristic curve (ROC) analysis. (Faye et al., 2011). Furthermore, gains in test specificity may be achieved by including additional antigens or different antigen combinations or other test modifications. In addition, solutions such as the

serial use of the SIT with IFN-γ should be explored as this may result in sensitivities not lower than those achieved with SICCT and specificities closer to 1.

4.8. Future perspectives

The information provided by this opinion indicates that the sensitivity and specificity of a bovine TB test may vary considerably from one population to another, even if the same test protocols are used. A likely reason is that populations differ with respect to infection history resulting in a different distribution of the various stages of infection and a different distribution of infections with other Mycobacteria. As a consequence, the probability that a test negative animal or herd is truly bovine TB free may vary from one population to another.
This problem is not unique for bovine TB, but is general for demonstrating freedom from infection.
In response to this problem, in recent years surveillance frameworks have been developed that define and prescribe the required confidence in the freedom of infection to be obtained by the testing system instead of providing a detailed overview of the testing scheme itself. In this way the heterogeneity in local risk factors can be taken into account and the risk manager can chose the testing scheme (test, sample size and sampling frame) to best obtain the required confidence of freedom (EFSA, 2008, More et al 2009). This approach to surveillance has the potential to provide higher confidence of freedom and is more cost effective than surveillance based on a prescribed testing scheme (Cameron, 2012). Nevertheless, the development of a surveillance program based on confidence of freedom for bovine TB in domestic ruminants still needs more work and, moreover, the practical implementation of such a program is as yet not straightforward.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

ToR 1 To issue a scientific opinion on the suitability of the IFN-γ test for inclusion in the official tests for the purpose of granting and retaining official bovine TB-free herd status, as laid down in Annex A to Directive 64/432/EEC, and certification for intra-Union trade in bovine animals, as required in Article 6(2)(a) of that Directive

In this assessment a test was considered suitable if it 1) has sensitivity equivalent or superior to the current standard tests used in the EU, 2) has specificity not lower than that of the current standard test with lowest specificity used in the EU and 3) is practical in its use. Under this definition of suitability, the majority of the data indicate that PPD-based IFN-γ could be included among the official tests. However, analysis of data obtained for this opinion suggest that in some populations the specificity may not be as high as the SIT the standard test with the lowest specificity currently used in the EU.

There is insufficient evidence upon which to base a definitive conclusion about the suitability of IFN-γ tests using defined antigens, although available information on the ESAT-6 and CFP-10 antigens of this test suggests that they have higher specificity than that of PPD based IFN-γ tests, without marked loss of sensitivity. This remains to be fully evaluated.

ToR 2 To issue a scientific opinion on the suitability of other, possibly newer, tests, if any, for inclusion in the official tests for the purpose of granting and retaining official bovine TB-free herd status, as laid down in Annex A to Directive 64/432/EEC, and certification for intra-Union trade in bovine animals, as required in Article 6(2)(a) of that Directive

Other tests evaluated did not meet the suitability criteria applied in this assessment.

ToR 3 In the event of a negative opinion to point (1), to advise the Commission on which further validation studies are necessary to evaluate the suitability of the IFN-γ test, or any other new test, for inclusion in the official tests for the purpose of granting and retaining official bovine TB-free herd status, as laid down in Annex A to Directive 64/432/EEC, and certification for intra-Union trade in bovine animals, as required in Article 6(2)(a) of that Directive

The conclusion to ToR 1 was not negative. Nonetheless, there is still uncertainty regarding whether the specificity of the PPD-based IFN-g test is always as high as that of the current tuberculin skin test with lowest specificity.

RECOMMENDATIONS

ToR 1

Should the PPD-based IFN-γ test be included in the official tests for the purpose of granting and retaining official bovine TB-free herd status, it is recommended that the protocols for its use for this purpose are harmonised in the EU.

ToR 3

Further validation of the IFN-γ test should evaluate the influence of factors, such as the presence of environmental mycobacteria, the prevalence of bovine TB in the herd, the age and bovine TB test history of the animals, and the type of production system, that may affect the test specificity and hence the suitability of the test for demonstrating freedom from bovine TB in different situations.

Well designed studies to further evaluate the diagnostic performance of IFN-γ test using defined antigens (e.g. ESAT-6 and CFP10) should be conducted. These studies should be carried out in parallel with PPD based IFN-γ tests in a single assay, and conclusions reached on whether the combined use of different antigens can optimise and improve the overall performance of the IFN-γ test.

The potential for the serial use of the IFN-γ test and the skin test (whereby positive test results are confirmed by a second test) to increase the specificity of the test protocol with limited loss of sensitivity should be assessed.

REFERENCES

Aagaard C, Govaerts M, Meikle V, Gutiérrez-Pabello JA, McNair J, Andersen P, Suárez-Güemes F, Pollock J, Espita C, Cataldi A, 2010. Detection of bovine tuberculosis in herds with different disease prevelance and influence of paratuberculosis infection on PPDB and ESAT6/CFP10 specificity. Preventive Veterinary Medicine, 96, 161–169.

Alvarez J, De Juan L, Bezos J, Romero B, Saez JL, Marques S, Dominguez C, Minguez O, Fernandez-Mardomingo B, Mateos A, Dominguez L, Aranaz A, 2009. Effect of paratuberculosis on the diagnosis of bovine tuberculosis in a cattle herd with a mixed infection using interferon-gamma detection assay. Veterinary Microbiology, 135, 389–393.

Alvarez J, Perez A, Bezos J, Marques S, Grau A, Saez JL, Minguez O, De Juan L, Dominguez L, 2012. Evaluation of the sensitivity and specificity of bovine tuberculosis diagnostic tests in naturally infected cattle herds using a Bayesian approach. Veterinary Microbiology, 2012, 155, 38–43.

Andersen, A.B. and Brennan, R., 1994. Protein Antigens of *Mycobacterium tuberculosis*, p. 307-332. In B. R. Bloom (ed.), Tuberculosis: Pathogenesis, Protection and Control. American Society for Microbiology, Washington.

Anonymous, 2006. Working Document on Eradication of Bovine Tuberculosis in the EU, accepted by the Bovine Tuberculosis Subgroup of the Task Force on Monitoring Animal Disease Eradication. SANCO/10200/2006.

Antognoli MC, Remmenga MD, Bengtson SD, Clark HJ, Orloski KA, Gustafson LL, Scott AE, 2011. Analysis of the diagnostic accuracy of the gamma interferon assay for detection of bovine tuberculosis in US herds. Preventive Veterinary Medicine, 101, 35–41.

Bermudez HR, Renteria ET, Medina BG, Hori-oshima S, Valle ADLM, Lopez VG, 2011. Evaluation of a lateral flow assay for the diagnosis of *Mycobacterium bovis* infection in dairy cattle. Journal of Immunoassay and Immunochemistry, 33, 59–65.

Brooks S, Gelman A, 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics, 7, 434–455.

Buddle BM, McCarthy AR, Ryan TJ, Pollock JM, Vordermeier HM, Hewinson RG, Andersen P, de Lisle GW, 2003. Use of mycobacterial peptides and recombinant proteins for the diagnosis of bovine tuberculosis in skin test-positive cattle. Veterinary Record, 153, 615–620.

Buddle BM, Ryan TJ, Pollock JM, Andersen P, de Lisle GW, 2001. Use of ESAT-6 in the interferon-gamma test for diagnosis of bovine tuberculosis following skin testing. Veterinary Microbiology, 80, 37–46.

Buddle, B.M., de Lisle, G.W., Pfeffer, A., Aldwell, F.E., 1995. Immunological responses and protection against *Mycobacterium bovis* in calves vaccinated with a low dose of BCG. Vaccine 13, 1123-1130.

Cameron, A., 2012. The consequences of risk-based surveillance: Developing output-based standards for surveillance to demonstrate freedom from disease. Preventive Veterinary Medicine 105, 280– 286

Clegg TA, Duignan A, Whelan C, Gormley E, Good M, Clarke J, Toft N, More SJ, 2011. Using latent class analysis to estimate the test characteristics of the gamma-interferon test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions. Veterinary Microbiology, 151, 68–76.

Coad M, Hewinson RG, Clifford D, Vordermeier HM, Whelan AO, 2007. Influence of skin testing and blood storage on interferon-gamma production in cattle affected naturally with *Mycobacterium bovis*. Veterinary Record, 160, 660–662.

Costello, E., Quigley, F., Flynn, O., Gogarty, A., McGuirk, J., Murphy, A. & Dolan, L., 1998. Laboratory examination of suspect tuberculous lesions detected on abattoir post mortem examination of cattle from non-reactor herds. Ir. Vet. J. 51, 248-250.

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 SH, Parry J, Nunez-Garcia J, Abernethy DA, Broughan JM, Cameron AR, Cook AJ, Rua-Domenech RDL, Goodchild AV, Greiner M, Gunn J, More SJ, Rhodes S, Rolfe S, Sharp M, Upton P, Vordermeier HM, Watson E, Welsh M, Whelan AO, Woolliams JA, Clifton-Hadley RS, 2011. Meta-analysis of diagnostic test performance and modelling of testing strategies for control of bovine tuberculosis in GB. Proceedings of the Society for Veterinary Epidemiology and Preventative Medicine, Leipzig, Germany, 23–25 March 2011, 139–153.

Du Y, Qi Y, Yu L, Lin J, Liu S, Ni H, Pang H, Liu H, Si W, Zhao H, Wang C, 2011. Molecular characterization of *Mycobacterium tuberculosis* complex (MTBC) isolated from cattle in northeast and northwest China. Research in Veterinary Science, 90, 385–391.

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. O.J.L., 121, pp.1977-2012

EFSA (European Food Safety Agency), 2008. Opinion of the Scientific Panel on Animal Health and Welfare: Tuberculosis testing in deer. The EFSA Journal, 645, 1–34.

Farnham MW, Norby B, Goldsmith TJ, Wells SJ, 2012. Meta-analysis of field studies on bovine tuberculosis skin tests in United States cattle herds. Preventative Veterinary Medicine, 103, 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 IFN gamma assay (Bovigam (R)) in a low prevalence area in France. Veterinary Microbiology, 151, 60–67.

Gormley E, Doyle MB, Fitzsimons T, McGill K, Collins JD, 2006. Diagnosis of *Mycobacterium bovis* infection in cattle by use of the gamma-interferon (Bovigam) assay. Veterinary Microbiology, 112, 171–179.

Gormley E, Doyle MB, McGill K, Costello E, Good M, Collins JD, 2004. 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, 102, 413–420.

Green, L.R., Jones, C.C., Sherwood, A.L., Garkavi, I.V., Cangelosi, G.A., Thacker, T.C., Palmer, M.V., Waters, W.R., Rathe. C.V., 2009. Single-antigen serological testing for bovine tuberculosis. Clin Vaccine Immunol. 16, 1309-1313.

Inwald, J., Hinds, J., Palmer, S., Dale, J., Butcher, P.D., Hewinson, R. G., Gordon, S. V., 2003. Genomic analysis of Mycobacterium tuberculosis complex strains used for production of purified protein derivative. J. Clin. Microbiol. 41, 3929-3932.

Jeon B-Y, Kim SC, Sungmo J, Kwak J, Cho J-E, Woo J-T, Seo S, Shim H-S, Park B-O, Lee S-S, and Cho S-N, 2010. Evaluation of enzyme-linked immunosorbent assay using milk

samples as a potential screening test of bovine tuberculosis of dairy cows in Korea. Research in Veterinary Science, 88, 390–393.

Lepper, A.W., Pearson, C.W., Corner, L.A., 1977. Anergy to tuberculin in beef cattle. Aust. Vet. J. 53, 214-216.

Marassi CD, Medeiros L, Lilenbaum W, 2010. The use of a gamma-interferon assay to confirm a diagnosis of bovine tuberculosis in Brazil. Acta Tropica, 113, 199–201.

Marassi CD, Medeiros L, McNair J, Lilenbaum W, 2011. Use of recombinant proteins MPB70 or MPB83 as capture antigens in ELISAs to confirm bovine tuberculosis infections in Brazil. Acta Tropica, 118, 101–104.

Monaghan, M.L., Doherty, M.L., Collins, J.D., Kazda, J.F., Quinn, P. J., 1994. The tuberculin test. Vet. Microbiol. 40, 111-124.

More, S.J., 2009. What is needed to eradicate bovine tuberculosis successfully: An Ireland perspective. Veterinary Journal 180, 275–278.

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.

OIE (World Organisation for Animal Health), 2012. Terrestrial Animal Health Code. 21th Edition.OIE, Paris, France, www.oie.int.

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

Pollock, J.M. and Neill, S.D., 2002. *Mycobacterium bovis* infection and tuberculosis in cattle. Vet. J. 163, 115-127.

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. Immunopathol. 108, 37-43.

R Development Core Team, 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project.org/.

Ritacco, V., Lopez, B., De Kantor, I.N., Barrera, L., Errico, F., Nader, A., 1991. Reciprocal cellular and humoral immune responses in bovine tuberculosis. Res. Vet. Sci. 50, 365-367.

Rothel, J.S., Jones, S.L., Corner, L.A., Cox, J.C., Wood, P.R., 1990. A sandwich enzyme immunoassay for bovine interferon-gamma and its use for the detection of tuberculosis in cattle. Aust. Vet. J. 67, 134-137.

Schiller I, Oesch B, Vordermeier HM, Palmer MV, Harris BN, Orloski KA, Buddle BM, Thacker TC, Lyashchenko KP, Waters WR, 2010a. Bovine tuberculosis: a review of current and emerging diagnostic techniques in view of their relevance for disease control and eradication. Transboundary and Emerging Diseases, 57, 205–220.

Schiller I, Vordermeier HM, Waters WR, Kyburz A, Cagiola M, Whelan A, Palmer MV, Thacker TC, Meijlis J, Carter C, Gordon S, Egnuni T, Hardegger R, Marg-Haufe B, Raeber A, Oesch B, 2010b. Comparison of tuberculin activity using the interferon-gamma assay for the diagnosis of bovine tuberculosis. Veterinary Record, 167, 322–326.

Schiller I, Vordermeier HM, Waters WR, Whelan A, Coad M, Gormley E, Buddle BM, Palmer M, Thacker T, McNair J, Welsh M, Hewinson RG, Oesch B, 2010c. Bovine tuberculosis: effect of the tuberculin skin test on in vitro interferon gamma responses. Veterinary Immunology and Immunopathology, 136, 1–11.

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

Thom, M., Morgan, J.H., Hope, J.C., Villarreal-Ramos, B., Martin, M., Howard, C.J., 2004. The effect of repeated tuberculin skin testing of cattle on immune responses and disease following experimental infection with Mycobacterium bovis. Vet. Immunol. Immunopathol. 102, 399-412.

Toft N, Jørgensen E, Højsgaard S, 2005. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Preventive Veterinary Medicine, 68, 19–33.

VLA (Veterinary Laboratories Agency), 2011. Meta-analysis of diagnostic tests and modelling to identify appropriate testing strategies to reduce *M. bovis* infection in GB herds. SE3238 Final Report. Available at : http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=16114&FromSearch=Y&Publisher=1&SearchText=SE3238&SortString=ProjectCode&SortOrder=Asc&Paging=10

Waters WR, Buddle BM, Vordermeier HM, Gormley E, Palmer MV, Thacker TC, Bannantine JP, Stabel JR, Linscott R, Martel E, Milian F, Foshaug W, Lawrence JC, 2011. Development and evaluation of an enzyme-linked immunosorbent assay for use in the detection of bovine tuberculosis in cattle. Clinical and Vaccine Immunology, 18, 1882–1888.

Waters WR, Nonnecke BJ, Olsen SC, Palmer MV, 2007. Effects of pre-culture holding time and temperature on interferon-gamma responses in whole blood cultures from *Mycobacterium bovis*-infected cattle. Veterinary Microbiology, 119, 277–282.

Waters WR, Nonnecke BJ, Palmer MV, Robbe-Austermann S, Bannantine JP, Stabel JR, Whipple DL, Payeur JB, Estes DM, Pitzer JE, Minion FC, 2004. Use of recombinant ESAT-6:CFP-10 fusion protein for differentiation of infections of cattle by *Mycobacterium bovis* and by *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis*. Clinical and Diagnostic Laboratory Immunology, 11, 729–735.

Waters, W.R., Buddle, B.M., Vordermeier, H.M., Gormley, E., Palmer, M.V., Thacker, T.C., Bannantine, J.P., Stabel, J.R., Linscott, R., Martel, E., Milian, F., Foshaug, W., Lawrence, J.C., 2011. Development and evaluation of an enzyme-linked immunosorbent assay for use in the detection of bovine tuberculosis in cattle. Clin Vaccine Immunol. 18, 1882-1888.

Waters, W.R., Palmer, M.V., Thacker, T.C., Bannantine, J.P., Vordermeier, H.M., Hewinson, R.G., Greenwald, R., Esfandiari, J., McNair, J., Pollock, J.M., Andersen, P. and Lyashchenko, K.P., 2006. Early antibody responses to experimental Mycobacterium bovis infection of cattle. Clin. Vaccine Immunol. 13, 648-654.

Whelan AO, Coad M, Peck ZA, Clifford D, Hewinson RG, Vordermeier HM, 2004. Influence of skin testing and overnight sample storage on blood-based diagnosis of bovine tuberculosis. Veterinary Record, 155, 204–206.

Whelan C, Shuralev E, Kwok HF, Kenny K, Duignan A, Good M, Davis WC, Clarke J, 2011. Use of a multiplex enzyme-linked immunosorbent assay to detect a subpopulation of mycobacterium bovis-infected animals deemed negative or inconclusive by the single intradermal comparative tuberculin skin test. Journal of Veterinary Diagnostic Investigation, 23, 499–503.

Whelan C, Whelan AO, Shuralev E, Kwok HF, Hewinson G, Clarke J, Vordermeier HM, 2010. Performance of the Enferplex TB assay with cattle in Great Britain and assessment

of its suitability as a test to distinguish infected and vaccinated animals. Clinical and Vaccine Immunology, 17, 813–817.

Whelan, C., Shuralev, E., O'Keeffe, G., Hyland, P., Kwok, H.F., Snoddy, P., O'Brien, A., Connolly, M., Quinn, P., Groll, M., Watterson, T., Call, S., Kenny, K., Duignan, A., Hamilton, M.J., Buddle, B.M., Johnston, J.A., Davis, W.C., Olwill, S. A., Clarke J., 2008. Multiplex immunoassay for serological diagnosis of Mycobacterium bovis infection in cattle. Clin. Vaccine Immunol. 15, 1834-1838.

Whipple D L, Palmer MV, Slaughter RE, Jones SL, 2001. Comparison of purified protein derivatives and effect of skin testing on results of a commercial gamma interferon assay for diagnosis of tuberculosis in cattle. Journal of Veterinary Diagnosis and Investigation, 13, 117–122.

Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J, 2006. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Medical Research Methodology, 6, 9.

Yearsley, D., Egan, J., Costello, E., O'Reilly, P.F., Hewinson, R.G., 1998. An evaluation of an anamnestic ELISA for the detection of tuberculous cattle. Ir. Vet. J. 51, 303-306

APPENDICES

A. EXCLUDED TESTS AND REASONS FOR EXCLUSION

Table 5: Diagnostic tests excluded from the evaluation (adapted from Annex 1 of AHVLA systematic review (VLA, 2011)

Test name	Long description	Large-scale survey of live animals	Sufficient evaluation
Culture	Bacterial culture	NO	YES
Glutaraldehyde	Glutaraldehyde test	YES	NO
Microscopic examination	Microscopic examination	NO	YES
Post mortem (meat inspection)	Meat inspection at the slaughterhouse	NO	YES
Post mortem (detail/laboratory)	Post-mortem examination in a laboratory	NO	YES
Bentonite flocculation test	Bacillary extracts of attenuated *M. bovis* (BCG) are fractionated on diethylaminoethyl cellulose to yield a carbohydrate fraction which sensitises bentonite to react with serum antibodies to *M. bovis*. The degree of flocculation (precipitation) is graded with higher flocculation indicating a positive response. Purified BCG carbohydrate fraction or old tuberculin is used to cause sensitisation	NO	NO
Clinical signs	Ante-mortem diagnosis based upon symptoms such as the following: Body condition: thin, very thin, emaciated Respiratory signs: chronic cough, elicited cough, respiratory distress, swollen mammary gland lymph nodes and superficial lymph nodes Externally visible lesions: head nodes, neck (cervical) nodes, udder tissue, supramammary nodes, skin, prescapular nodes, other nodes	NO	YES
Complement fixation tests	Immunological test that can be used to detect the presence of either specific antibody or specific antigen in serum	YES	NO
Diagnostic anatoxin	Experimental variation of the intradermal skin test	NO	NO
Dot-immunogold silver staining (Dot-IGSS)	The procedure is based on a two-step incubation using a primary antibody and a gold-labelled secondary antibody conjugate	YES	NO
Double intradermal test (see also the Stormont test)	Official test in the UK until 1940. Two injections of mammalian tuberculin (one 48 hours after the other) and measurement of skin thickness 24 hours later	YES	NO
Expression of IL2R, IL4, IL10 or TNF-α	Expression of cytokines in blood in response to infection. PPD antigens are presented to lymphocytes and the production of cytokine from the stimulated T cells is measured in an ELISA format	NO	NO
Flocculation/ Meinicke flocculation reaction	A precipitin test characterised by a flocculent precipitate of antigen and antibody	NO	NO

Fluorescence lamp	Smears from culture of *M. bovis* examined using a fluorescence lamp	NO	NO
Haemagglutination test/passive haemagglutination (PHA) test/Middlebrook/ Middlebrook–Dubos	Serological test based on the agglutination by tuberculous antiserum of sheep red cells sensitised with a carbohydrate extract of tubercle bacilli	NO	NO
Haemolytic/ haemolysis/ haemolysin test	Modification of the haemagglutination test whereby complement is added to the antigen and serum causing haemolysis rather than haemagglutination in the presence of serum containing antibodies to TB	NO	NO
Indirect fluorescent antibody (IFA) test	Serological test to detect circulating antibodies to *M. bovis* using any antigens/conjugates including FITC-anti-bovine IgG and IgM conjugates	YES	NO
LCx amplification assay	Commercial assay using ligase chain reaction to amplify DNA targets in *M. tuberculosis* complex	NO	NO
Leucocyte formation	Blood component count technique		NO
Lymphocyte immunostimulation test	Response of lymphocytes in blood to *M. bovis* antigens compared to response in control samples	YES	NO
M. tuberculosis (complex) direct test	Direct target-amplified nucleic acid probe test for the *in vitro* diagnostic detection of *M. tuberculosis* complex rRNA in acid-fast bacilli smear – positive and negative concentrated sediments from sputum, bronchial specimens, or tracheal aspirates. The tuberculosis complex consists of *M. tuberculosis*, *M. bovis*, *M. bovis* BCG, *M. africanum, M. canetti and M. microti*	NO	NO
Ophthalmic test	Tuberculin is administered into the eye. Congestion of the conjunctiva with a serofibrinous exudate after 6–16 hours is a positive result	NO	NO
Patch test	Ointment containing dead bovine tubercle bacilli and bovine tuberculin are applied to a shaved and washed area on the posterior aspect of the thigh, and covered with sticking plaster. A control area is covered with sticking plaster only. Reactions compared at 48 hours	NO	NO
Radiometric detection	Monitors carbon dioxide released from a broth medium containing radioactively labelled substrate	NO	NO
Stormont test	Variation of the double intradermal skin test. Two injections of tuberculin, the second 7 days after the first. The injection site is measured before reinjection and 24 hours later; an increase of 5 mm or more is considered a positive reaction	YES	NO
Thermal test	A large quantity of tuberculin (e.g., Koch's old tuberculin) is injected into a cow's veins and her temperature measured. If the cow develops a rise in body temperature of at least 1.5 ºF, peaking no lower than 103.2 ºF, the cow is classified as a reactor	NO	NO
Tube gel precipitation technique/gel double diffusion test/precipitation	Serological test demonstrating localised precipitation from the diffusion of antigen to its antibody in gel	YES	NO

test			
Vernes test/Vernes flocculation test	Serological test: Vernes-Bricq-Yvon photometer used to measure flocculation occurring over a 4-hour period in a mixture of serum and resourcinol	NO	NO

BCG, bacille Calmette-Guérin; FITC, fluorescein isothiocyanate; Ig, immunoglobulin; IL, interleukin; LCx, …; TNF, tumour necrosis factor.

B. DIAGNOSTIC TESTS FOR BOVINE TB

Skin test

The detection of early infection is dependent on measuring the cell-mediated immune responses that predominate in the early stages of the disease and which involve recruitment and activation of a variety of T cells to the site of infection (Pollock et al., 2005). These responses, however, can be measured peripherally, and this has been exploited through the development of several immunological diagnostic assays that have proven very effective in diagnosing TB both in cattle and in humans (de la Rua-Domenech et al., 2006).

The most widely used field test for the diagnosis of TB in cattle is the tuberculin skin test, which measures the cell-mediated immune response to *M. bovis* infection (Monaghan et al., 1994). The tuberculins currently in use in the EU contain a crude mixture of predominantly secreted mycobacterial proteins derived from specified strains of *M. bovis* (Andersen et al., 1994; Inwald et al., 2003). Many of these antigens are also found in non-pathogenic environmental mycobacterial species and this cross-reactivity to common antigens can result in reduced specificity of the test (Francis et al., 1978; Monaghan et al., 1997). For this reason, an *M. avium-derived* tuberculin is included in the comparative tuberculin test.

Interpretation of the SIT is based on observing, measuring and recording 72 hours after intradermal inoculation of bovine tuberculin (bovine PPD) the nature and extent of any increase in skin thickness at the site of inoculation. In the comparative test (SICCT) this response is also compared with that observed 72 hours after inoculation of avian tuberculin (avian PPD) at an adjacent site on the neck of the animal. The single test takes only the bovine PPD site into account and, consequently, this test is more sensitive but less specific than the comparative test (Karolemeas at al., 2012). When the standard interpretation of the tests is applied, inconclusive reactors must be subjected to an additional skin test at least 42 days after the previous one; if they are not negative in the second test, they are deemed positive. However, to maximise sensitivity of the test a severe interpretation may be applied (particularly in areas of high prevalence), in which all inconclusive reactors in the first skin test are considered as positive and removed for slaughter (Anonymous, 2006).Variations in the types of tuberculin used and the strictness of the interpretation of the test results, which both modulate sensitivity and specificity, can all directly affect sensitivity and specificity.

Post-mortem and bacteriological examination

Post-mortem examination of cattle, and bacteriological examination of appropriate tissues including lymphatic nodes, are critical steps in the confirmation of the diagnosis of TB in cattle (Costello et al., 1998). In meat plants, the detection of gross lesions of TB on a presumptive basis at routine meat inspection of carcasses is often routinely employed to screen for infected animals. A tentative diagnosis of bovine TB can be made following the finding of typical tuberculous lesions during necropsy. Culturing is rarely required when the disease frequency is high and the cost of misdiagnosis, in terms of the cost of consequential action, is negligible. Conversely, in a disease-free area or one with very low prevalence, culturing is usually needed to ascertain *M. bovis* infection.

The inspection procedure employed to examine cattle that are slaughtered as reactors to a tuberculin test can take one or three forms, namely:

(1) An examination of tissues and organs for macroscopic lesions, conducted either *in situ* at the meat plant at the time of the post-mortem carcass inspection. This is considered adequate and sufficient when bovine TB is endemic and the prevalence of disease is high.

(2) The aseptic collection of tissues displaying lesions for histopathological examination in a laboratory, possibly associated with PCR and/or bacterial culture, especially if microscopic lesions suggestive of TB are observed.

(3) If no macroscopic lesions are observed (no visible lesions), lymph nodes may nevertheless may be sampled in order to carry out PCR and/or bacterial culture.

To determine the true infection status and true TB status of cattle that give a positive reaction in diagnostic tests but have no visible lesions (NVL) reported following routine *post-mortem* inspection in the meat plant, a laboratory-based bacteriological examination is often necessary (Costello et al., 1998). The classification of *M. bovis*-infected animals with NVLs disclosed at slaughter may arise as a result of recent infection, poor *post-mortem* inspection technique or infection with mycobacteria other than *M. bovis*. When successful, and where cross-contamination of the sample during collection at the point of initial inspection can be ruled out, a positive culture for *M. bovis* is considered as the definitive gold standard for the diagnosis of TB in cattle. The degree of sensitivity of the inspection procedure for the detection of lesions at slaughter in meat plants is influenced, *inter alia,* by the time devoted to the inspection and the diligence of the inspectors conducting the inspection.

IFN-γ assay

Arising from the need to increase the detection rate of *M. bovis*-infected animals in exposed herds, the IFN-γ assay was developed in Australia as an ancillary test to improve the sensitivity of testing of cattle when used in parallel with the tuberculin test. The principle of the assay is to use ELISA technology to detect and quantify release of the IFN-γ cytokine when heparinised whole blood is incubated with bovine and avian (PPD) tuberculin within the first 8–24 hours post collection (Rothel et al., 1990). During the first stage of the test blood samples collected from cattle are transported to the laboratory and stimulated overnight with tuberculin. In the second stage of the test, the plasma is harvested from the stimulated blood and is assayed for the presence of IFN-γ. The different optical density (OD) readings obtained after stimulation with each antigen (bovine PPD/avian PPD/PBS, used as a blank) are then used to yield a quantitative result: OD obtained after stimulation with PBS (NIL_{OD}) is often subtracted from the OD observed after stimulation with bovine PPD ($PPDb_{OD}$) and avian PPD ($PPDa_{OD}$); these two figures ($PPDb_{OD} - NIL_{OD}$ and $PPDa_{OD} - NIL_{OD}$) are then compared and, depending on the cut-off in place, an animal is considered a reactor or not. The potential of other more specific antigens (mainly ESAT6/CFP10) for induction of specific release of IFN-γ is also currently under evaluation. Results from experimental and natural infections of cattle indicate that the assay can detect a cell-mediated immune response to infection as early as 14 days post infection, and earlier than the tuberculin test (Buddle et al., 1995).

According to the EU legislation, Member States may authorise the ancillary use of the IFN assay to "enable detection of the maximum number of infected animals in a herd or in a region in addition to the tuberculin test" (Directive 64/432/EEC). Its parallel implementation increases the sensitivity of the diagnostic regime, although it can also cause a decrease in diagnostic specificity. Therefore, its use is not recommended on a routine basis in areas or regions where the herd prevalence is low (Anonymous, 2006). In certain countries/areas (usually free of disease) the IFN assay is used in surveillance programmes following non-negative results to skin tests (serial use of the tests), in order to increase the specificity of the overall diagnostic procedure (Table 6). The EU regulation, however, does not include this serial use of the IFN-γ test.

Antibody detection tests

In countries with managed surveillance systems, the majority of infected animals disclosed by the tuberculin test and, in particular, IFN-γ tend to be in the early stages of the disease and display few, if any, visible lesions upon *post-mortem* examination (Pollock et al., 2005). However, there is a category of *M. bovis*-infected animals that is consistently missed by the tests targeting the cell-mediated immune response (Lepper et al., 1977; Yearsley et al., 1998). Sometimes described as "anergic", these animals may in time develop generalised disease with extensive lung lesions and, as such, pose a serious threat to the health of herd owners and those with whom they come in contact, as well as to the health of other cattle. The reasons for anergy are unclear, although it can be attributed to a number of causes, e.g., the animal may be unable to mount a detectable cellular immune response to infection or the immune response is impaired – possibly owing to intercurrent pathogens with immunosuppressive effect (Claridge et al., 2012), or in some cases rapid progression of the disease may result in suppression of responses (Managhan et al., 1994).

As the disease progresses there is a shift in the balance of the immune response away from the predominant cell-mediated immune response and towards an antibody response (Ritacco et al., 1991; Pollock & Neill, 2002). These antibodies are generally targeted at immunodominant antigens that elicit a humoral response, notably MPB70 and MPB83 released in large amounts by *M. bovis* in the later stages of the disease. A variety of ELISA tests have been developed that depend on the detection of high levels of circulating antibodies to the immunodominant antigens of *M. bovis* (Whelan et al., 2008; Green et al., 2009; Waters et al., 2011*).* It has been reported, however, that the antibody response can be boosted (the anamnestic response) by a prior tuberculin test (Thom et al., 2004). Much of the recent technological effort has focused on developing *in vitro* field tests that can rapidly provide a reliable test result. Recent studies using lateral flow chromatography technology incorporated into a rapid serological test format have shown that the sensitivity of the tests in question, when used in cattle, increases as the disease progresses post infection (Waters et al., 2006).

Table 6: Current use of IFN-γ in the European Union[(a)]

Country	Official status	Antigen (manufacturer)	Tested population	Combined use with the skin test (parallel/serial)	Time between collection of samples and processing in the laboratory	Cut-off point
Austria	OTFC	PPDs (Prionics)	N.A.	N.A.	Processing within the first 24 hours after blood collection	$PPDb_{OD} - NIL_{OD} > 0.1$; $PPDb_{OD}$-PPDa > 0.1
Belgium	OTFC	PPDs (Synbiotics), CFP-10/ESAT-6	N.A.	N.A.	Processing within the first 8 hours post collection	$PPDb_{OD}/NIL_{OD} > 2.0$ and $NIL_{OD} < 0.15$ (Invitrogen kit)
France	OTFC	PPDs (Prionics), CFP-10/ESAT-6	(1) Cattle in infected herds (selective culling) (2) Reinforced surveillance in high-risk areas	Parallel	Processing within the first 8 hours post collection	(PPDb – PPDa)/(PC – NC) ≥ 0.05 (E/C – PBS)/(PC – NC) ≥ 0.015
		PPDs (Prionics), CFP-10/ESAT-6	Surveillance in some areas with low prevalence (nine districts)	Serial (reactors to ST) (blood sampling for IFN 72 hours after ST)	Processing within the first 8 hours post collection	(PPDb – PPDa)/(PC – NC) ≥ 0.05 (E/C- PBS)/(PC-NC) ≥ 0.015
		PPDs (Prionics)	Bullfighting herds in Camargue	In parallel with ST or as stand-alone test	Processing within the first 8 hours post collection	PPDb – PPDa)/(PC – NC) ≥ 0.04
Germany	OTFC	PPDs (Prionics)	Cattle in infected herds	Parallel	Processing within the first 30 hours post collection	$PPDb_{OD} - NIL_{OD} \geq 0.1$; $PPDb_{OD} > PPDa_{OD}$
Netherlands	OTFC	PPDs (Prionics), Peptide cocktail (VLA)	Cattle in positive herds	Serial	Processing within the first 8/24 hours post collection	$PPDb_{OD} - NIL_{OD} \geq 0.05$ (eradication); $PPDb_{OD} - NIL_{OD} \geq 0.1$ (screening); $PPDb_{OD} - PPDa_{OD} > 0.1$
Poland	OTFC	PPDs (Australian Prionics)	Cattle with positive or inconclusive reactions	Serial	Processing 24 hours after blood collection	$PPDb_{OD} - NIL_{OD} \geq 0.05$; $PPDb_{OD} > PPDa_{OD}$

			to skin test			
Greece	Non-OTFC	PPDs (CZ Veterinaria)	Cattle with positive or inconclusive reactions to skin test	Serial	Processing within the first 24 hours post collection	$PPDb_{OD} - NIL_{OD} \geq + 0.1$; $PPDb_{OD} - PPDa_{OD} \geq 0.1$
Hungary	Non-OTFC	N.A.	N.A.	N.A.	Processing within the first 8 hours post collection	$PPDb_{OD} \geq PBSNIL_{OD} + 0.1$; $PPDb_{OD} > PPDa_{OD} + 0.1$
Ireland	Non-OTFC	PPDs (Prionics)	High risk cohorts in infected herds	Parallel	Processing within the first 8 hours post collection	$PPDb_{OD} - NIL_{OD} \geq 0.05$; $PPDb_{OD} \geq PPDa_{OD}$; $PPDb_{OD} > 0.1$
Italy	Non-OTFC/OTFR	PPDs (Prionics)	Cattle in infected herds	Parallel	Processing within the first 8 hours post collection	$PPDb_{OD}/NIL_{OD} \geq 2$; $PPDb_{OD} - PPDa_{OD} \geq 0.05$
Portugal	Non-OTFC	PPDs (CZ Veterinaria)	Cattle in infected herds	Parallel	Processing within the first 8 hours post-collection	$PPDb_{OD} - NIL_{OD} \geq 0.05$; $PPDb_{OD} > PPDa_{OD}$
Romania	Non-OTFC	PPDs (SN Inst. Pasteur SA- Romania) PPDs (Lelystad Biologicals BV-Nederlands)	At the Romanian National Authority's request for expertise, or at the owner's request (with payment)	N.A.	Processing within the first 30 hours post collection	$PPDb_{OD} - NIL_{OD} \geq 0.1$; $PPDb_{OD} > PPDa_{OD} + 0.1$
Spain	Non-OTFC	PPDs (CZ Veterinaria)	Cattle in infected herds	Parallel	Processing within the first 8 hours post collection	$PPDb_{OD} - NIL_{OD} \geq 0.05$; $PPDb_{OD} > PPDa_{OD}$
United Kingdom (England and Wales)	Non-OTFC	PPDs (Prionics) (99.25 %)	Cattle in infected herds	Parallel	Processing 24 hours after blood collection	$PPDb_{OD} - PPDa_{OD} > 0.1$
		ESAT-6/CFP-10 (0.75 %)	Herds with non-specific reactors/suspicions of fraud	Serial	Processing 24 hours after blood collection	ESAT6/CFP10OD–NILOD > 0.1

(a) Source: First technical meeting of the EFSA Scientific Network on Tuberculosis Testing, National Reference Laboratories of the Member States, EU Reference Laboratory for Bovine Tuberculosis.

N.A., information not available; NC, negative control; PC, positive control.

C. SYSTEMATIC LITERATURE REVIEW

In 2008, with funding from Defra,[7] the AHVLA[1] set up a working group (WG) to conduct a systematic review and meta-analysis of the performance of diagnostic tests for bovine TB in cattle. The WG included 18 reviewers, from within the AHVLA and from outside organisations, with a scientific expertise ranging from TB immunology to pathology and laboratory culture of *M. bovis* to veterinary epidemiology and implementation of TB control programmes (Downs et al., 2011, VLA, 2011).

The process of the review was discussed and agreed by the WG. The methodology was adapted from an approach taken previously while reviewing the performance diagnostic tests for TB in deer (EFSA, 2008). Comprehensive search criteria were developed and the process of the review standardised. Sources of references included:

1. Electronic databases including:

- Web of Knowledge (includes Web of Science 1995–, Current Contents 1998–, CAB Abstracts 1910–, Medline 1950–)

- Dialog (includes Embase 1974–, Agricola 1970–, Agris 1975–)

2. Unpublished data and reports identified through contacting research institutions and laboratories (grey literature)

3. References known to members of the WG

4. Review of bibliographies of reports and papers

The final search of electronic databases was carried out on 1 December 2008, with no limits applied by year, language, region or type of diagnostic test, using the following search string:

(bovine tuberc* or mycobacterium bovis*) or ((mycobact* not (paratub* or johne*))

AND

(bovin* or cattle or cow or cows or calf or calves or buffa)

AND

(test* or screen* or diagn* or eia or elisa or pcr or polym* chain react* or lympho* or interferon or skin or rapid or detect* or peptid* or cervical or caudal or sicct or antibody* or necroscopy or necropsy or survei* or sensitivi* or specifici* or perform* or eval* or valid* or accura* or confirmatory)

9 782 potentially eligible references were identified initially by the electronic search and other sources. In order to be included in the final review and meta-analysis, the references identified as potentially eligible had to pass through two stages of review.

The stage 1 review of abstracts, where available, and titles was conducted by two reviewers. The inclusion and exclusion criteria were as follows:

Inclusion criteria:

- the reference related to primary research;

[7] The UK Government Department for Environment, Food and Rural Affairs.

- the reference included either report(s) of sensitivity and/or specificity of a diagnostic test for TB, or provided data enabling the statistics to be calculated;

- the diagnostic test performance was measured on cattle.

Exclusion criteria:

- the sensitivity estimates were from studies in which cattle had been experimentally infected with *M. bovis*.

Entire references of those that passed through the stage 1 review were obtained and randomly allocated to two reviewers[8] for detailed review at stage 2.

Inclusion and exclusion criteria at stage 2:

Inclusion criteria for sensitivity estimates:

- sensitivity could be calculated;

- the bovine population had been naturally exposed to TB;

- each study animal had been individually examined using one of the following reference tests: post-mortem examination (meat inspection or detailed laboratory inspection), culture, microscopic inspection (histology or histopathology), SICCT test.

Exclusion criteria:

- the study population had been experimentally infected with *M. bovis*;

- The definition of "infected" was based on a "group" level inference (such as a sample of animals in the study population being positive for culture of *M. bovis*).

Inclusion criteria for specificity estimates:

- specificity could be calculated;

- there is good evidence that the bovine population was free from infection with, and exposure to, *M. bovis*, including herds with officially TB free (OTF) status, herds from an OTF area or OTF country, herds from a non-endemic TB area where the authors stated that the area has been free of TB for several years, or herds that in the authors' opinion was TB free and had been free of TB for several years.

Exclusion criteria:

- any other evidence of lack of exposure to TB.

Reference papers that appeared to have eligible data were also reviewed by the reviewers using the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) instrument developed by Whiting et al. (2006), adapted for the veterinary use. Each reviewer extracted the agreed range of data from the references they considered met the eligibility criteria and entered the data on to individual copies of the bespoke database. Data entered into the stage 2 databases by the two reviewers were compared using a query system and a hierarchical process was followed to resolve inconsistencies.

[8] Reference papers and reports written in English and Spanish were reviewed by two reviewers. References written in other languages were reviewed by one reviewer.

There were 119 references (published 1934–2009) with eligible estimates of diagnostic test performance for 14 different diagnostic tests.

Pooled unadjusted estimates of performances for the different diagnostic tests and modifications of the tests were calculated. This was followed by modelling to control for confounding factors and the structure of the data, which comprised varying numbers of records per reference, population and test type. The estimation of sensitivity and specificity was carried out separately for each test type because performance estimates were derived from different study populations. Stepwise logistic regression was conducted to identify confounding factors such as country of study, year of study, sampling strategy, evidence of cross-reactivity with environmental mycobacteria, animal production type, type of reference standard, interpretation of tuberculin response (skin test), tuberculin used in skin test, diagnostic antigen (in blood tests), whether the blood test was performed before or after the skin test and others. Relative differences in performance due to cut-off used to define a positive response in the IFN-γ and ELISA blood tests were adjusted for by including a "counter parameter" in the model. The counter parameter was the corresponding estimate of specificity (reported in the reference) where sensitivity was being estimated and the corresponding estimate of sensitivity where specificity was being estimated. Where the corresponding estimate of test performance was not reported within the reference, the median sensitivity or specificity from the range of values estimated for the test was imputed. Covariates remaining in the models after the stepwise procedure were then used in logistic regression modelling with a random effect term to account for reference run in a Bayesian framework implementing the Monte Carlo Markov Chain technique. Final estimates were reported with categories best representing test conditions in GB and in Ireland as baseline. Performance estimates for tests considered for the EFSA mandate are reported in Table 3. Further results and estimates for other tests and details of the AHVLA systematic review and meta-analysis procedure can be found in the final report on the study and its accompanying annexes and appendices (VLA, 2011).

D. UPDATE OF THE LITERATURE REVIEW

Searches were conducted to identify any new data on diagnostic test performance that had become available since the original searches for the AHVLA review (VLA, 2011). The search of electronic databases for the AHVLA systematic review was last performed on 1 December 2008. The update was done by EFSA on 13 March 2012 using the same search string without language restrictions on Web of Knowledge (which simultaneously searches Web of Science 1995 , Current contents 1998–, CAB abstracts 1910–, Medline 1950–).

(bovine tuberc* or mycobacterium bovis*) or ((mycobact* not (paratub* or johne*))

AND

(bovin* or cattle or cow or cows or calf or calves or buffa)

AND

(test* or screen* or diagn* or eia or elisa or pcr or polym* chain react* or lympho* or interferon or skin or rapid or detect* or peptid* or cervical or caudal or sicct or antibody* or necroscopy or necropsy or survei* or sensitivi* or specifici* or perform* or eval* or valid* or accura* or confirmatory)

The search retrieved 946 results for review.

The inclusion criteria considered for stage 1 of the review (relevance screening based on title and abstract) were:

- the reference is related to primary research;
- the reference included either report(s) of sensitivity and/or specificity of a diagnostic test for TB or provides data enabling these statistics to be calculated;
- the diagnostic test performance was measured on cattle;
- the reference includes reports on the performance of any of the following tests: IFN-γ, ELISA, LBBA, multiplex and/or rapid.
- the specificity estimates were from bovine tuberculosis free cattle population unless latent class analysis was used.

The exclusion criteria used was:

- the sensitivity estimates were from studies in which cattle had been experimentally infected.

Each record was reviewed by two reviewers; if either of the reviewers considered the reference was relevant, it was included for stage 2 screening. Where there was insufficient information to determine whether stage 1 criteria were met, the reference automatically passed to stage 2.

The stage 1 review yielded 124 records.

The stage 2 review consisted of a review of titles and abstracts by two reviewers, and all conflicts (non-agreement between reviewers) were discussed and screening of the full text made when necessary to reach agreement. The criteria were the same as in stage 1. The stage 2 review yielded 15 records. Data extracted from these studies are summarised in Table 7.

Table 7: Estimates of sensitivity and specificity and 95% confidence intervals (CI, calculated when possible) from included studies in the updated literature review.

Reference	Test name	Criteria (a)	Population (b)	*N*	Approach (c)	Test used as comparator or reference	Sensitivity	Lower 95% CI	Upper 95% CI	Specificity	Lower 95% CI	Upper 95% CI	Comments
Marassi et al., 2010	ELISA-MPB	12	Other	32	GS	Culture positive or PCR + SICCT	0.344			0.750			Cut-off points based on OD readings were calculated using ROC curves
Waters et al., 2011	ELISA-MPB	12	OTFH	1473	GS	SICCT/Culture				0.98			No measures of spread reported, herds sourced from four different countries
Waters et al., 2011	ELISA-MPB	12	OUT	478	GS	SICCT Standard/ culture	0.630						No measures of spread reported herds sourced from four different countries
Jeon et al., 2010	ELISA-MPB	12	Other	109	GS	Caudal fold	0.818	0.645	0.930				Measured in sera Mean + 3SD of OD in negative controls
Jeon et al., 2010	ELISA-MPB	12	Other	109	GS	Caudal fold	0.879	0.718	0.966				Measured in milk Mean+3SD of OD in negative controls
da Silva et al., 2011	ELISA-MPT-51	12	OUT	262	GS	SICCT test	0.548	0.478	0.617				OD $\geq$ 1.301
Du et al., 2011	ELISA-rMPB70-83-E6		OUT	111	GS	SIT cervical + culture positive	0.378	0.288	0.475				Positive culture = strains of *Mycobacterium* complex isolated from throat swabs from SIT-positive cattle S/P $\geq$0.5 S/P = ((OD of samples – OD of negative controls)/(OD positive controls – OD negative controls))

da Silva et al., 2011	ELISA-Ag85	12	OUT	262	GS	SICCT test	0.480	0.411	0.551				OD ≥ 0.898
Silva et al., 2011	ELISA-BCG	12	OUT	262	GS	SICCT test	82.2	76.3	87.1				OD ≥ 1.287
Alvarez et al., 2012	IFN-γ-BA	1	OUT	6202	LCA	SIT cervical	0.897	0.775	0.972	0.857	0.843	0.88	
Alvarez et al., 2012	IFN-γ-BA	2	OUT	6202	LCA	SIT cervical	0.833	0.719	0.916	0.904	0.891	0.927	
Antognoli et al., 2011	IFN-γ-BA	12	OTFH	4123	GS					0.97	0.965	0.975	Cut-off value ≥ 0.3
Antognoli et al., 2011	IFN-γ-BA	12	OTFH	4123	GS					0.986	0.982	0.989	Cut-off value ≥ 0.5
Antognoli et al., 2011	IFN-γ-BA	6	OTFH	4123	GS					0.907	0.898	0.916	Cut-off value ≥ 0.1
Antognoli et al., 2011	IFN-γ-BA	12	OUT	87	GS	Culture	0.839	0.761	0.916				
Clegg et al., 2011	IFN-γ-BA	1		4937	LCA	Multiplex (cut off 1) and SICCT standard	0.666	0.631	0.701	0.881	0.868	0.894	
Clegg et al 2011	IFN-γ-BA	1		4937	LCA	Multiplex (cut-off 3) and SICCT standard	0.716	0.678	0.742	0.877	0.863	0.89	
Clegg et al., 2011	IFN-γ-BA	1		4937	LCA	Multiplex (cut-off 5) and SICCT standard	0.762	0.728	0.793	0.879	0.864	0.893	
Clegg et al., 2011	IFN-γ-BA	1		4937	LCA	Multiplex (cut-off 1) and SICCT severe	0.641	0.608	0.675	0.887	0.874	0.9	
Clegg et al., 2011	IFN-γ-BA	1		4937	LCA	Multiplex (cut-off 5) and SICCT severe	0.723	0.691	0.754	0.883	0.869	0.897	
Alvarez et al., 2009	IFN-γ-BA	1	OUT	46	GS	Culture	0.783	0.636	0.890				
Marassi et al., 2010	IFN-γ-BA	6	OUT	35	GS	SICCT/culture or macroscopic lesions	0.914	0.776	0.970				
Faye et al., 2011	IFN-γ-BA	12	OUT	60	GS	Culture or macroscopic lesions	0.83	0.72	0.92				Cut-off including ODs from positive controls of the plate. Used 3–10 after skin test

Faye et al., 2011	IFN-γ-BA	12	OTFH	492	GS					0.994	0.982	0.999	Cut-off including ODs from positive controls of the plate. Used 3–10 after skin test
Schiller et al., 2009	IFN-γ-BA	3	OUT	431	GS	Culture and/or lesions	90.9	0.878	0.935				
Schiller et al., 2009	IFN-γ-BA	3	OTFH	874	GS					0.965	0.950	0.976	
Clegg et al., 2011	IIFN-γ-BA	1		4937	LCA	Multiplex (cut off 3) and SICCT severe	0.682	0.651	0.713	0.881	0.867	0.894	
Faye et al., 2011	IFN-γ-CE	12	OUT	60	GS	Culture or macroscopic lesions	0.87	0.75	0.94				Cut-off including ODs from positive controls of the plate. Used 3–10 after skin test
Faye et al., 2011	IFN-γ-CE	12	OTFH	492	GS					0.907	0.879	0.930	Cut-off including ODs from positive controls of the plate. Used 3–10 after skin test
Whelan et al., 2011	Multiplex	12	OUT	60	GS	Culture or macroscopic lesions	0.883	0.774	0.952				Animals negative or inconclusive in SICCT test
Whelan et al., 2010	Multiplex		OUT	96	GS	Culture and macroscopic lesions	0.865	0.780	0.926				Cut-off Enfer 1
Whelan et al., 2010	Multiplex		OUT	96	GS	Culture and macroscopic lesions	0.813	0.720	0.885				Cut-off Enfer 3
Whelan et al., 2010	Multiplex		OUT	96	GS	Culture and macroscopic lesions	0.771	0.671	0.850				Cut-off Enfer5
Whelan et al., 2010	Multiplex		OTFR	93	GS					0.796	0.7	0.8723	Cut-off Enfer 1
Whelan et al., 2010	Multiplex		OTFR	93	GS					0.946	0.879	0.982	Cut off Enfer 3
Whelan et al., 2010	Multiplex		OTFR	93	GS					1.000	0.950	1	Cut-off Enfer5
Clegg et al., 2011	Multiplex			4937	LCA	SICCT severe and IFN	0.642	0.610	0.674	0.922	0.910	0.933	Cut-off Enfer 1

Clegg et al 2011	Multiplex			4937	LCA	SICCT severe and IFN	0.488	0.455	0.522	0.992	0.988	0.995	Cut-off Enfer 3
Clegg et al., 2011	Multiplex			4937	LCA	SICCT Severe and IFN	0.341	0.311	0.372	0.998	0.996	1.00	Cut-off Enfer 5
Clegg et al., 2011	Multiplex			4937	LCA	SICCT standard and IFN	0.684	0.648	0.719	0.921	0.910	0.932	Cut-off Enfer 1
Clegg et al., 2011	Multiplex			4937	LCA	SICCT standard and IFN	0.522	0.485	0.559	0.992	0.988	0.995	Cut-off Enfer 3
Clegg et al., 2011	Multiplex			4937	LCA	SICCT standard and IFN	0.371	0.337	0.406	0.998	0.996	1.00	Cut-off Enfer 5
Bermudez et al., 2011	Rapid		OUT	268		Culture	0.458	0.361	0.557				
Bermudez et al., 2011	Rapid		OUT	268		Direct PCR	0.488	0.403	0.578				

(a) Positive if: value obtained in the blood sample stimulated with PBS, bovine PPD, avian PPD and specific antigens respectively 1-$PPDb_{OD} - NIL_{OD} > 0.05$ and $PPDb_{OD} > PPDa_{OD}$;, 2-$PPDb_{OD} - NIL_{OD} > 0.1$ and $PPDb_{OD} > PPDa_{OD}$; 3-$PPDb_{OD} - NIL_{OD} > 0.1$ and $PPDb_{OD} - PPDa_{OD} > 0.1$; 4-$PPDb_{OD} - PPDa_{OD} > 0.04$; 5-$PPDb_{OD} - PPDa_{OD} > 0.05$; 6-$PPDb_{OD} - PPDa_{OD} > 0.1$; 7-$PPDb_{OD}/NIL_{OD} > 2$ and $NIL_{OD} < 0.15$, 8-Others PPD (please specify); 9-$ANTIG_{OD} - NIL_{OD} > 0.040$; 10-$ANTIG_{OD} - NIL_{OD} > 0.050$; 11-$ANTIG_{OD} - NIL_{OD} > 0.1$; 12-Others ANTIG.

(b) Source population status regarding bovine TB infection: OTFC, officially TB-free country; OTFR, officially TB-free region or province; OTFH, officially TB-free herd; OUT, outbreaks or reactors in the past 2 years in the same herds; NOOUT, not officially free but no outbreaks; OTHER, other.

(c) Methodology used to estimate sensitivity and specificity: LCA, latent class analysis; GS, gold standard.

E. PUBLIC DATA CALL

EUROPEAN FOOD SAFETY AUTHORITY
Call for data on Bovine Tuberculosis testing

Published: 26 March 2012
Deadline: 26 April 2012

Background

The Commission requested the European Food Safety Authority (EFSA) to issue a scientific opinion on the suitability of the IFN-γ test or other, possibly newer, tests for inclusion amongst the official tests for the purpose of granting and retaining an officially tuberculosis free herd status as laid down in Annex A to Directive 64/432/EEC and certification for intra Union trade in bovine animals as required in Article 6(2)(a) of that Directive furthermore.

http://registerofquestions.efsa.europa.eu/roqFrontend/questionLoader?question=EFSA-Q-2011-01254

Objective

To address this question it must be assessed if the performance of the alternative tests, are equivalent or better when compared to the current standard test used in the European Union (EU).

Characteristics of diagnostic accuracy will be considered. Diagnostic sensitivity (DSe) is defined as the probability of a positive test result in an infected animal, and diagnostic specificity (DSp) as the probability of a negative test result in an animal that is not infected.

Call for data

EFSA kindly asks governments, companies, universities, research institutions, other stakeholders and any individuals to submit any available data concerning diagnostic test accuracy (sensitivity and specificity). EFSA will evaluate the received data if data is in compliance with the following inclusion and exclusion criteria:

Inclusion criteria:

- The diagnostic test under evaluation (TUE) is one of the tests included in the Table 1.
- The diagnostic test performance (diagnostic sensitivity and/or Sp) was measured on bovines
- Each study animal had been individually examined using a official intradermal tuberculin test, either Single intradermal test (SIT) or Single intradermal comparative cervical test (SICCT) as the comparator test (CT) or with a reference standard (RT) for confirmation of infection by culture, microscopic examination or identification of macroscopic lesions.

For a study to be included animals must have been tested by at least 2 of the tests considered

Test under evaluation	Comparator test	Reference standard
X	X	
	X	X
X	X	X
X		X

1. Additional data concerning testing of the same animals with other tests for confirmation of infection although valuable, is not essential.

Exclusion criteria:

- Any study where animals are experimentally infected with TB.

Table 1 :Diagnostic tests to be evaluated (TEU)

Test name	Abbreviation	Long description
IFN-γ Bovine-Avian	IFN-γ-BA	Gamma - interferon test with bovine PPD and avian PPD diagnostic antigens
IFN-γ Bovine	IFN-γ –B	Gamma - interferon test with bovine diagnostic antigen
IFN-γ CFP10 ESAT6	IFN-γ –CE	Gamma - interferon test with CFP10 and ESAT6 diagnostic antigens
IFN-γ MPB70	IFN-γ –MPB	Gamma - interferon test with MPB70 diagnostic antigen
IFN-γ BACE	IFN-γ –BACE	Gamma - interferon test with bovine PPD and avian PPD diagnostic antigens and CFP10 and ESAT6 diagnostic antigens
IFN-γ OTHER	IFN-γ -OTHER	Gamma - interferon test with OTHER diagnostic antigens – PLEASE SPECIFY
ELISA Bovine-Avian	Elisa-BA	Enzyme-linked immunosorbant assay with bovine PPD and avian PPD diagnostic antigens
ELISA Bovine	Elisa-B	Enzyme-linked immunosorbant assay with bovine diagnostic antigen
ELISA MPB70	Elisa-MPB	Enzyme-linked immunosorbant assay with MPB70 diagnostic antigen
Latex Bead Agglutination assay	Latex	Latex bead agglutination assay (LBAA)
Multiplex immunoassay	Multiplex	Multiplex chemiluminescent immunoassay developed by Enfer Scientific
Serological Rapid	Rapid	Rapid immunochromatographic assay (rapid test)

Confidentiality and unpublished data

Specific issues relating to confidentiality of the data provided will be discussed between the owners and EFSA.

Closing date for the data submissions is 26 April 2012.

The data should be transmitted to EFSA in electronic format using the MS Excel® table TBtest.xls.

Use worksheet RawData to enter experimental data for individual animals included in a study, one row per animal.

Use worksheet TestSummaries to report the total number of animals testing positive and negative in test comparison studies, complete only the area in blue.

A data dictionary for each column can be found in the first worksheet and the controlled terminologies and definitions are provided in Terms.

Data submissions should be sent to this e-mail address: ahaw@efsa.europa.eu

For inquires regarding the reporting format please contact us at: ahaw@efsa.europa.eu

F. DATA RECEIVED – PUBLIC DATA CALL

Data was received from both public institutions and private commercial organizations. A total of 54 data sets were received, a summary of the data sets provided is presented in table 8. The data was evaluated in agreement with the inclusion and exclusion criteria defined in the public data call. Furthermore, during the data analysis it was concluded that, because of the differences in test protocol, it was not possible to combine datasets from different countries. To be able to estimate all model parameters only datasets in which more than one population and at least three test results were available could be used. Data from animal populations selected based on post mortem results (infected /not infected) was excluded to avoid bias.

Table 8: Received datasets

Data provider	Workbook (a)	Study identifier (a)	Year	Country	Population (b)	Study strategy	Dataset type	Test name	Skin test	Timing	Post mortem
Asturias	data Asturias - EFSA	Asturias	2011	ES	OUT	Census	Summary	IFN-γ-BA	SIT cervical severe	BEFORE	Culture or macroscopic lesions
Galicia	Data Galicia - EFSA	Galicia	2001	ES	OUT	Census	Summary	IFN-γ-BA	SIT cervical severe	BEFORE	Culture
Castilla y León	dataCyLEFSA	CyL	2009	ES	OUT	Census	Summary	IFN-γ-BA	SIT cervical severe	BEFORE	No post-mortem results reported
Castilla y León	dataCyLEFSA	CyL	2010	ES	OUT	Census	Summary	IFN-γ-BA	SIT cervical severe	BEFORE	No post-mortem results reported
Castilla y León	dataCyLEFSA	CyL	2011	ES	OUT	Census	Summary	IFN-γ-BA	SIT cervical severe	BEFORE	Culture (37)
EURL	Population1	Example 1	2010	ES	OUT	Selective sampling	Animal level	ELISA-MPB70	SICCT severe	BEFORE	Culture or macroscopic lesions
EURL	Population1	Example 1	2010	ES	OUT	Selective sampling	Animal level	ELISA-MPB70	SIT cervical severe	BEFORE	Culture or macroscopic lesions
EURL	Population1	Example 1	2010	ES	OUT	Selective sampling	Animal level	IFN-γ-BA	SICCT severe	BEFORE	Culture or macroscopic lesions
EURL	Population1	Example 1	2010	ES	OUT	Selective sampling	Animal level	IFN-γ-BA	SIT cervical severe	BEFORE	Culture or macroscopic lesions
EURL	Population2	Example 2	2010	ES	OUT	Census	Animal level	IFN-γ-BA	SIT cervical severe	BEFORE	Culture or macroscopic lesions (62)
EURL	Population3	Example 3	2010	ES	OUT	Convenient sampling	Animal level	ELISA-MPB70	SICCT severe	BEFORE	Culture or macroscopic lesions (7)
EURL	Population3	Example 3	2010	ES	OUT	Convenient sampling	Animal level	ELISA-MPB70	SIT cervical severe	BEFORE	Culture or macroscopic lesions (7)
EURL	Population3	Example 3	2010	ES	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT severe	BEFORE	Culture or macroscopic lesions (7)
EURL	Population3	Example 3	2010	ES	OUT	Convenient sampling	Animal level	IFN-γ-BA	SIT cervical severe	BEFORE	Culture or macroscopic lesions (7)

EURL	Population4	Example 4	2010	ES	OTFH	Selective sampling	Animal level	ELISA MPB70	SIT cervical severe	BEFORE	No post-mortem results reported
EURL	Population4	Example 4	2010	ES	OTFH	Selective sampling	Animal level	IFN-γ-BA	SIT cervical severe	BEFORE	No post-mortem results reported
EURL	Population5	Example 7	2008	ES	OTFH	Objective sampling	Animal level	IFN-γ-BA	SIT cervical severe	BEFORE	No post-mortem results reported
EURL	Population5	Example 7	2008	ES	OTFH	Objective sampling	Animal level	Multiplex	SIT cervical severe	BEFORE	No post-mortem results reported
EURL	Population5	Example 7	2008	ES	OTFH	Objective sampling	Animal level	Multiplex	SIT cervical severe	BEFORE	No post-mortem results reported
Prionics	120326ax1_CH_B2G_LS_BA	FS2010_CH	2011	CH	OTFC	Objective sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	No post-mortem results reported
Prionics	120326ax1_FR_B2G_LS_BA	FS2010_FR	2011	FR	OTFC	Objective sampling	Animal level	IFN-γ-BA	SICCT severe	BEFORE	No post-mortem results reported
Prionics	120326ax1_IRL_B1G_LS_BA	FS2010_IRL	2011	IE	OTHER	Suspect sampling	Animal level	IFN-γ-BA	SICCT severe	AFTER	No post-mortem results reported
Prionics	120326ax1_UK_B2G_LS_BA	FS2010_UK	2011	GB	OUT	Selective sampling	Animal level	IFN-γ-BA	SICCT severe	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial1	high risk	2008	IE	OUT	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial3	high risk	2009	GB	OUT	Selective sampling	Summary	Multiplex	SIT caudal standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial3	high risk	2009	GB	OUT	Selective sampling	Summary	Multiplex	SIT caudal standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial3	high risk	2009	GB	OUT	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial3	high risk	2009	GB	OUT	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial2fileb	high_risk_area	2008	IE	OUT	Selective sampling	Summary	Multiplex	SICCT severe	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial2fileb	high_risk_area	2008	IE	OUT	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial2filea	high_risk_area	2008	IE	OUT	Selective sampling	Summary	Multiplex	SICCT-SEVERE	AFTER	No post-mortem results reported
Enferplex	120326ax1E	high_risk_ar	2008	IE	OUT	Selective	Summary	Multiplex	SICCT-	AFTER	No post-mortem results

	nferTrial2filea	ea				sampling			STANDARD		reported
Enferplex	120326ax1EnferTrial1	low risk	2008	IE	NOOUT	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial3	low risk	2009	GB	NOOUT	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial3	low risk	2009	GB	NOOUT	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial4	low_risk area	2010	CH	OTFC	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Enferplex	120326ax1EnferTrial2filea	low_risk_area	2008	IE	NOOUT	Selective sampling	Summary	Multiplex	SICCT severe	BEFORE	No post-mortem results reported
Enferplex	120326ax1EnferTrial2filea	low_risk_area	2008	IE	NOOUT	Selective sampling	Summary	Multiplex	SICCT standard	BEFORE	No post-mortem results reported
Enferplex	120326ax1EnferTrial2filea	low_risk_area	2008	IE	NOOUT	Selective sampling	Summary	Multiplex	SICCT severe	BEFORE	No post-mortem results reported
Enferplex	120326ax1EnferTrial2fileb	low_risk_area	2008	IE	NOOUT	Selective sampling	Summary	Multiplex	SICCT standard	BEFORE	No post-mortem results reported
Enferplex	120326ax1EnferTrial4	low_risk_area	2010	CH	OTFC	Selective sampling	Summary	Multiplex	SICCT standard	AFTER	No post-mortem results reported
Madrid	Raw data CAM 2010-2011.xlsx	Madrid2010	2010	ES	OUT	Census	Animal level	IFN-γ-BA	SIT cervical severe	BEFORE	Culture
Madrid	Raw data CAM 2010-2011.xlsx	Madrid2011	2011	ES	OUT	Census	Animal level	IFN-γ-BA	SIT cervical severe	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	ESATData	Prob_Net_ESAT6	2005	GB	OUT	Convenient sampling	Animal level	IFN-γ other	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	ESATData	Prob_Net_ESAT6	2006	GB	OUT	Convenient sampling	Animal level	IFN-γ-other	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	ESATData	Prob_Net_ESAT6	2007	GB	OUT	Convenient sampling	Animal level	IFN-γ-other	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	ESATData	Prob_Net_ESAT6	2008	GB	OUT	Convenient sampling	Animal level	IFN-γ- other	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	ESATData	Prob_Net_ESAT6	2009	GB	OUT	Convenient sampling	Animal level	IFN-γ-other	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-	ESATData	Prob_Net_E	2010	GB	OUT	Convenient	Animal	IFN-γ-	SICCT	BEFORE	Culture or macroscopic

NI-UK		SAT6				sampling	level	other	standard		lesions
AFBI-NI-UK	NIData	Prob_Net_IFNg	2004	UK (NI)	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	NIData	Prob_Net_IFNg	2005	UK (NI)	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	NIData	Prob_Net_IFNg	2006	UK (NI)	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	NIData	Prob_Net_IFNg	2007	UK (NI)	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI-NI-UK	NIData	Prob_Net_IFNg	2008	UK (NI)	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI	NIData	Prob_Net_IFNg	2009	UK (NI)	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	Culture or macroscopic lesions
AFBI	NIData	Prob_Net_IFNg	2010	UK (NI)	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	BEFORE	Culture or macroscopic lesions
IZSLER	120326ax1-1-1v2	120608	2010	IT	OTFR	Suspect sampling	Animal level	IFN-γ-BA	SIT cervical severe	AFTER	No post-mortem results reported
IZSLER	120326ax1-2-1v2	5143	2012	IT	OTFR	Suspect sampling	Animal level	IFN-γ-BA	SIT cervical standard	AFTER	No post-mortem results reported
AHVLA	EFSAdata18-4-12GJ		2002-2005	GB	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	AFTER	Culture or macroscopic lesions
AHVLA	EFSAdata18-4-12GJ		2002-2011	GB	OUT	Convenient sampling	Animal level	IFN-γ-BA	SICCT standard	AFTER	Culture or macroscopic lesions
AHVLA	EFSAdata18-4-12GJ		2002-2011	GB	OUT	Convenient sampling	Animal level	IFN-γ-CE	SICCT standard	AFTER	Culture or macroscopic lesions
AHVLA	EFSAdata18-4-12GJ		2003-2011	GB	OUT	Convenient sampling	Animal level	IFN-γ-CE	SICCT standard	AFTER	Culture or macroscopic lesions
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	Austria Negatives	2010	AT	OTFC	Convenient sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Macroscopic lesions
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	England AHVLA Negatives	2010	GB	OTFH	Selective sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Macroscopic lesions
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	England AHVLA TB	2010	GB	OUT	Convenient sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Culture
IDEXX	IDEXX EFSA TB ELISA Data	Ireland UCD Negatives	2010	IE	OTFH	Selective sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Macroscopic lesions

	25April2012										
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	Ireland UCD TB Set 1	2010	IE	OUT	Selective sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Macroscopic lesions
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	Ireland UCD TB Set 2	2010	IE	OUT	Selective sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Macroscopic lesions
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	Ireland UCD TB Set 3	2010	IE	OUT	Selective sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Macroscopic lesions
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	Ireland UCD TB Set 4	2010	IE	OUT	Selective sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Macroscopic lesions
IDEXX	IDEXX EFSA TB ELISA Data 25April2012	Wales AHVLA Wales TB	2010	GB	OUT	Convenient sampling	Animal level	ELISA-MPB70	SICCT severe	AFTER	Culture
UCD-IE	120326ax1_ Ireland(GormleyE)_TC _1st August2012 .xlsx	high_risk_area	2008	IE	OUT	Selective sampling	Summary	IFN-γ-BA	SICCT severe	AFTER	Macroscopic lesions
UCD-IE	120326ax1_ Ireland(GormleyE)_TC _1st August2012 .xlsx	high_risk_area	2008	IE	OUT	Selective sampling	Summary	IFN-γ-BA	SICCT standard	AFTER	Macroscopic lesions
UCD-IE	120326ax1_ Ireland(GormleyE)_TC _1st August2012 .xlsx	low_risk_area	2008	IE	NOOUT	Selective sampling	Summary	IFN-γ-BA	SICCT severe	BEFORE	Macroscopic lesions
UCD -IE	120326ax1_ Ireland(Gor	low_risk_area	2008	IE	NOOUT	Selective sampling	Summary	IFN-γ-BA	SICCT standard	BEFORE	Macroscopic lesions

	mleyE)_TC_1st August2012.xlsx										
Prionics	120326ax1_CH_B2G_PCEC_0.1	FS2010_CH	2010	CH	OTFC	Objective sampling	Animal level	IFN-γ -CE	SICCT standard	BEFORE	No post-mortem results reported
Prionics	120326ax1_CH_B2G_PC-HP_0.1	FS2010_CH	2010	CH	OTFC	Objective sampling	Animal level	IFN-γ-PC-HP	SICCT standard	BEFORE	No post-mortem results reported
Prionics	120326ax1_FR_B2G_PCEC_0.1	FS2010_FR	2011	FR	OTFC	Objective sampling	Animal level	IFN-γ -CE	SICCT severe	BEFORE	No post-mortem results reported
Prionics	120326ax1_FR_B2G_PC-HP_0.1	FS2010_FR	2011	FR	OTFC	Objective sampling	Animal level	IFN-γ PC-HP	SICCT severe	BEFORE	No post-mortem results reported
Prionics	120326ax1_IRL_B2G_PC-HP_0.1	FS2010_IRL	2011	IE	OTHER	Suspect sampling	Animal level	IFN-γ- PC-HP	SICCT severe	AFTER	No post-mortem results reported
Prionics	120326ax1_UK_B2G_PCEC_0.1	FS2010_UK	2011	GB	OUT	Selective sampling	Animal level	IFN-γ -CE	SICCT severe	AFTER	No post-mortem results reported
Prionics	120326ax1_UK_B2G_PC-HP_0.1	FS2010_UK	2011	GB	OUT	Selective sampling	Animal level	IFN-γ- PC-HP	SICCT severe	AFTER	No post-mortem results reported
Prionics	120326ax1_IRL_B2G_PC-HP_0.05	FS2010_IRL	2011	IE	OTHER	Suspect sampling	Animal level	IFN-γ- PC-HP	SICCT severe	AFTER	No post-mortem results reported

(a) Work book : Study identifier that represents all animals or results included in the same study
(b) Source population status regarding bovine TB infection: OTFC, officially TB-free country; OTFR, officially TB-free region or province; OTFH, officially TB-free herd; OUT, outbreaks or reactors in the past 2 years in the same herds; NOOUT; not officially free but no outbreaks; OTHER, other.
(c) Timing : When the blood sample was taken in relation to skin test, before or after inoculation of tuberculin

G. MODEL DETAILS

Materials and methods

Latent class models have been used previously in the context of test diagnosis when gold standard tests are absent (Toft et al., 2005). The latent class model approach can be used for a binary latent variable, Y, the categories of which are called latent classes and indicate the disease status: Y takes the value 1 if the disease or infection is present and 0 otherwise. The outcomes of p diagnostic tests in the j-th subpopulation are expressed using manifest binary variables, X_{ij}, assuming the value 1 if the i-th diagnostic test is positive and 0 otherwise, $i = 1, \ldots, p$ ($p = 3$ in our case), $j = 1, \ldots, J$, where J denotes the number of populations.

The model was used to fit four different data sets (Spain SIT, Spain SICCT, Northern Ireland and Ireland). In order to solve identifiability issues only datasets in which more than one population and at least three test results were collected could be used.

Only the dataset from Northern Ireland was complete regarding the results of the post-mortem test. The post-mortem test results were missing for samples that were negative for the other two tests (IFN-γ-BA and skin test, but also for samples in which at least IFN-γ-BA was positive. Three different scenarios were considered to simulate the missing data.

- Scenario I: post-mortem test will be 98 % negative when IFN-γ and skin test are negative. The proportions observed in the available data are maintained for the cases when both tests are positive or IFN-γ is positive (Table 9). The WG members considered this to be the most likely scenario.

- Scenario II : The post-mortem result is in agreement with the IFN-γ result.

- Scenario III : The post-mortem result is in agreement with the skin test result.

Table 9: Test results observed in Spain and Ireland

	T101	T100	T111	T110
Spain 2010	70	162	55	21
Spain 2011	69	232	32	18
Ireland OUT	88	326	326	21
Ireland NOOUT	0	9	5	5

T101, positive results to IFN-γ and post mortem, negative to skin test; T100, positive results to IFN-γ, negative to skin test and post mortem; T111, positive results to all; T110, positive results to IFN-γ and skin test, negative to post mortem.

The model parameters include the prevalence in the populations considered, η_j, the sensitivities and specificities of the various tests in the different populations, denoted by Se_{ij} and Sp_{ij}, respectively, $i = 1,2,3$.

In this particular case, the j-th population counts (O_j) of the different patterns of test results (in a total of eight possible patterns) follow a multinomial distribution:

$$O_j \mid Se_{ij}, Sp_{ij}, \eta_j \sim Multinomial(\pi_j = P(\pi_j = x_j), n_j),$$

where n_j is the sample size of j-th population, , $i = 1,2,3$, and π_j is a vector of probabilities of observing the individual pattern $x_j = (x_{1j}, x_{2j}, x_{3j})^t$ of test results in population j.

Then the probability of an outcome pattern $\mathbf{x_j}$, considering that test $\mathbf{X_1}$ (IFN-γ-BA) and $\mathbf{X_2}$ (e.g., SICCT standard) are correlated in the infected and non-infected classes and $\mathbf{X_3}$ (post-mortem) is assumed to be conditional independent from the other two tests, can be written in the following way:

$$P(\mathbf{x_j}) = \eta_j \cdot \left[Se_{1j}^{x_{1j}} \cdot Se_{2j}^{x_{2j}} \cdot (1 - Se_{1j})^{1-x_{1j}} \cdot (1 - Se_{2j})^{1-x_{2j}} + (-1)^{x_{1j}-x_{2j}} \cdot cov\, se_{12|Y=1}\right] \cdot$$
$$\left[Se_{3j}^{x_{3j}} \cdot (1 - Se_{3j})^{1-x_{3j}}\right] + (1 - \eta_j) \cdot \left[Sp_{3j}^{1-x_{3j}} \cdot (1 - Sp_{3j})^{x_{3j}}\right] \cdot$$
$$\left[Sp_{1j}^{1-x_{1j}} \cdot Sp_{2j}^{1-x_{2j}} \cdot (1 - Sp_{1j})^{x_{1j}} \cdot (1 - Sp_{2j})^{x_{2j}} + (-1)^{x_{1j}-x_{2j}} \cdot cov\, sp_{12|Y=0}\right]$$

where $cov\, se_{12|Y=1} = cov(X_1, X_2|Y = 1)$ and $cov\, sp_{12|Y=0} = cov(X_1, X_2|Y = 0)$ are the measures of dependencies between the two tests for both classes (infected and non-infected). The model considers a different prevalence for each population and equal sensitivities ($Se_{ij} \equiv Se_i, \forall\, j = 1,2$ **and** $i = 1,2,3$, i.e., that the test sensitivity is not different across the populations) and specificities ($Sp_{ij} \equiv Sp_i, \forall\, j = 1,2$ **and** $i = 1,2,3$, i.e., that the test specificity is not different across the populations) of each test across populations.

The Bayesian paradigm was used to fit the latent class model, considering non-informative prior for the parameters ($Se_i, Sp_i,$ **where** $i = 1,2,3$). The prior distributions for the prevalence, sensitivities and specificities were considered to be β distributed, with parameters $\alpha = 1$ and $b = 1$. The priors used for the covariance between the tests were uniformly distributed between the limits of the parameters, which are derived from the sensitivities and specificities of the two tests and they are:

$$\max\left[(1 - Se_1) \cdot (1 - Se_2) - Se_1 \cdot Se_2,\ -Se_1 \cdot Se_2\right] \le cov\, se_{12|Y=1} \le \min\left[Se_2 \cdot (1 - Se_1), Se_1 \cdot (1 - Se_2)\right]$$
$$\max\left[(1 - Sp_1) \cdot (1 - Sp_2) - Sp_1 \cdot Sp_2,\ -Sp_1 \cdot Sp_2\right] \le cov\, sp_{12|Y=0} \le \min\left[Sp_2 \cdot (1 - Sp_1), Sp_1 \cdot (1 - Sp_2)\right]$$

Once the parameters of the model are estimated then the kappa (κ) measure of agreement can be estimated as follows:

$$\kappa_{Se} = \frac{cov\, se_{12|Y=1}}{Se_1 \cdot (1 - Se_2) + Se_2 \cdot (1 - Se_1)} \text{ and } \kappa_{Sp} = \frac{cov\, sp_{12|Y=0}}{Sp_1 \cdot (1 - Sp_2) + Sp_2 \cdot (1 - Sp_1)}.$$

The latent class model was fitted using the statistical software R (R Development Core Team, 2009), the package R2WinBUGS was used to fit the Bayesian model, assuming conditional dependency between the test under evaluation and the comparator. The convergence was assessed using the CODA package (Plummer et al., 2006).

Results

The estimates of sensitivity and specificity of IFN-γ-BA of the Bayesian latent class analysis are summarised in Table 4. Sensitivity, specificity, covariance and agreement measure (kappa) for all countries (Scenario I) are shown in Figures 3–10.

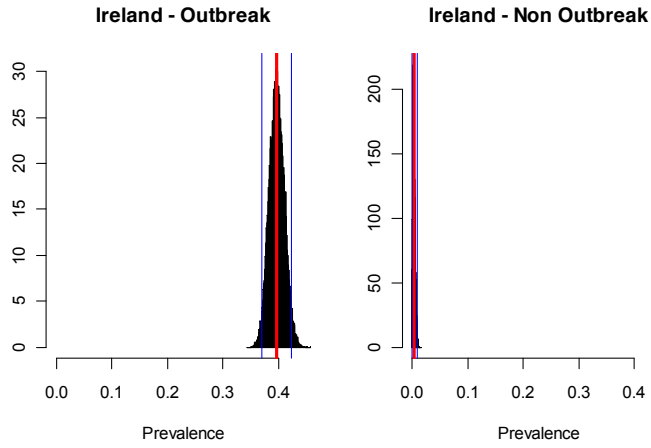

Figure 3: Prevalence estimates from the latent class model for Ireland, Scenario I.

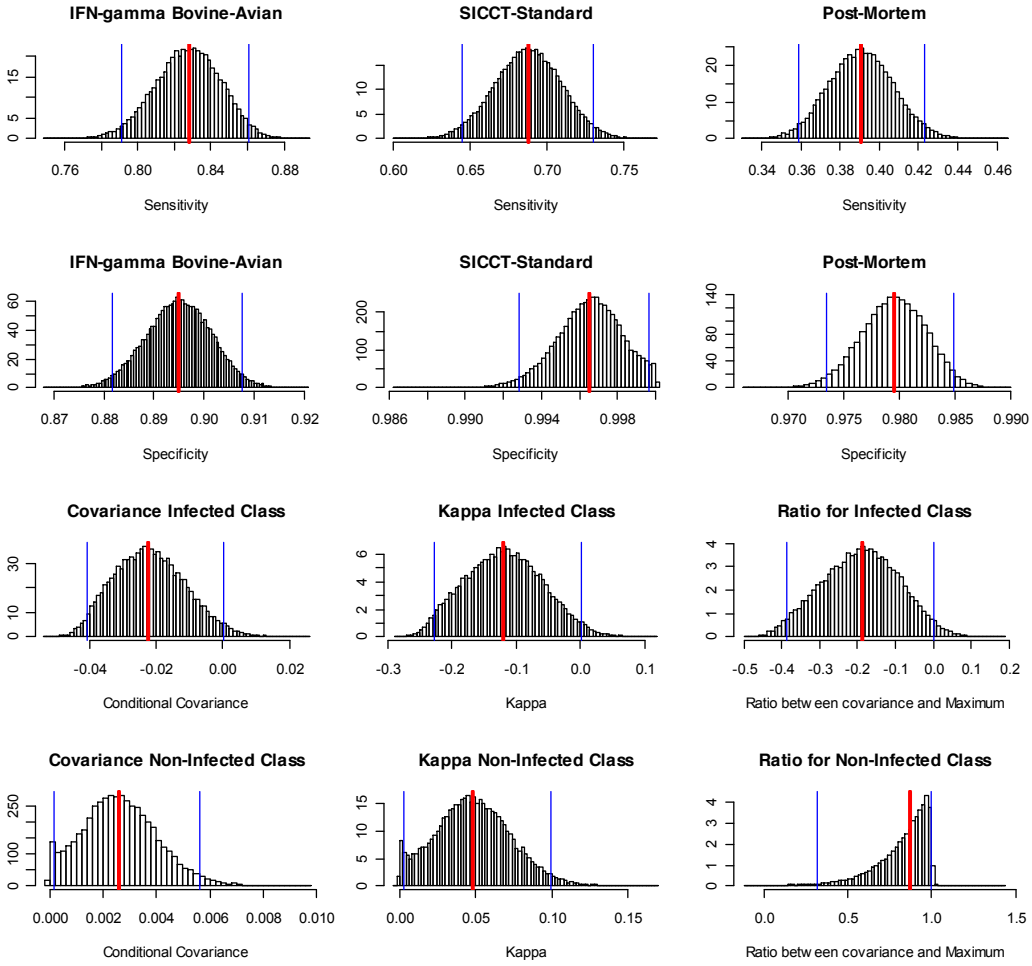

Figure 4: Sensitivities, specificities, covariance and agreement measure (kappa) from latent class model for Ireland, Scenario I.

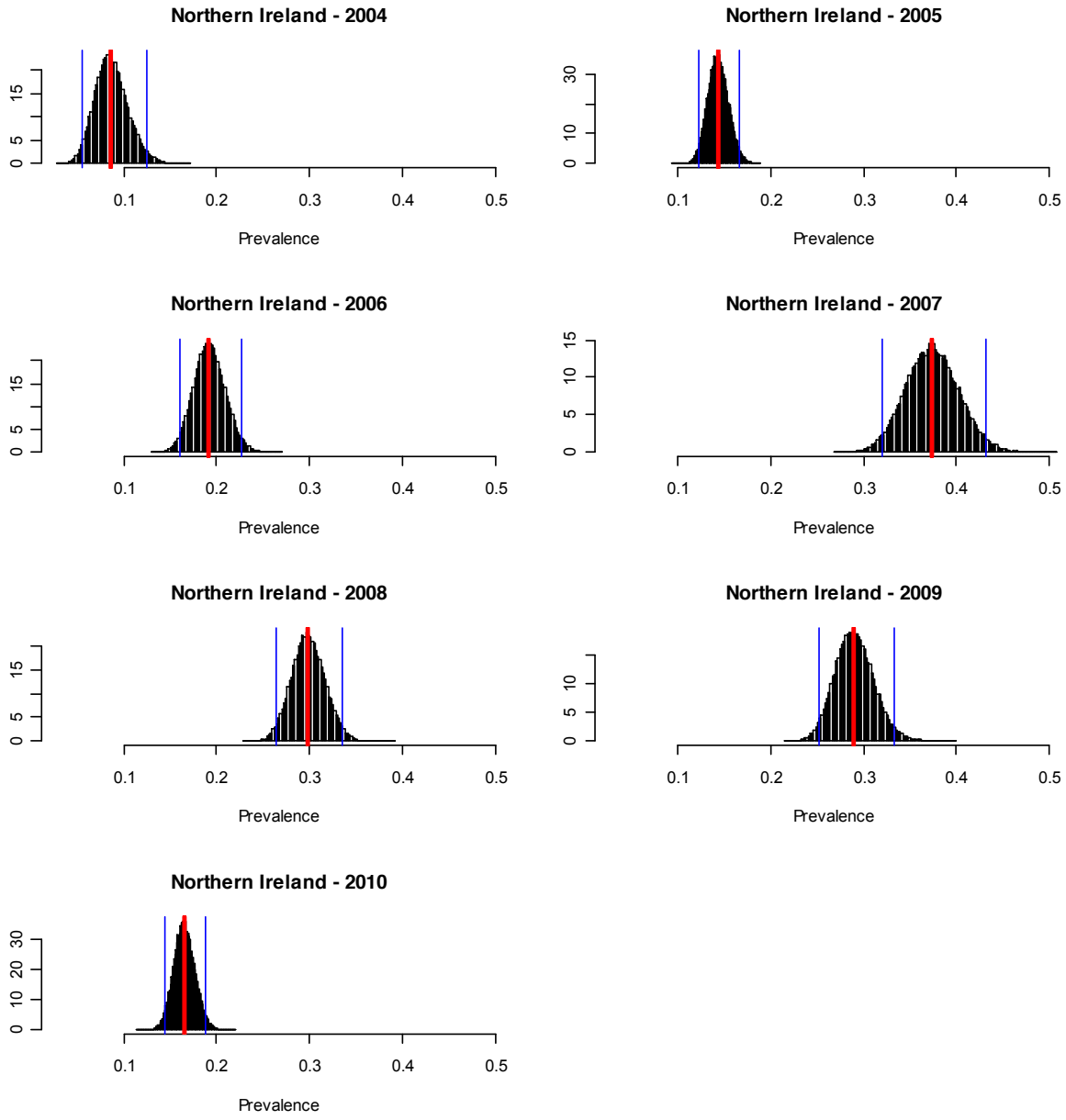

Figure 5: Prevalence estimates from the latent class model for Northern Ireland

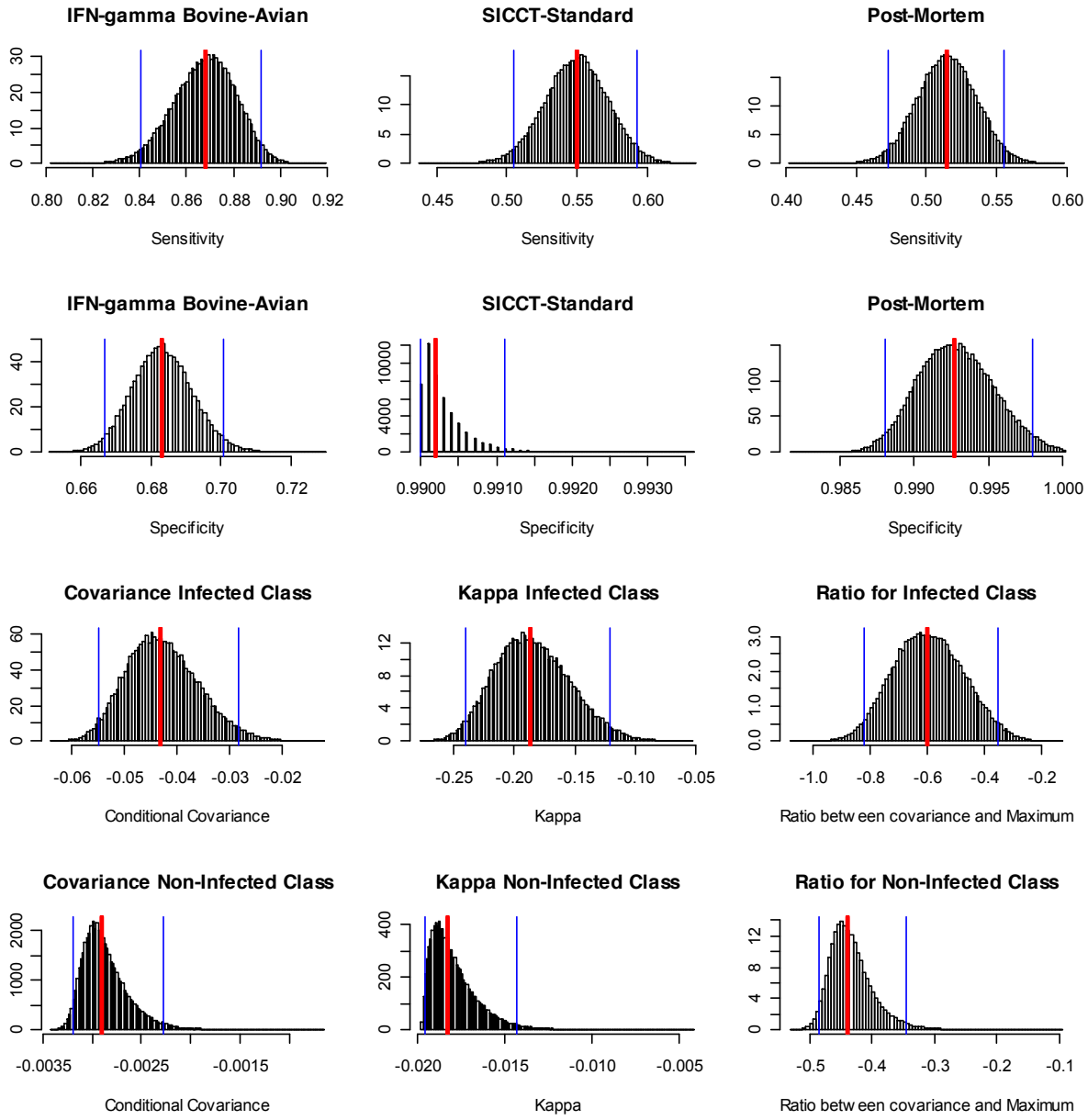

Figure 6: Sensitivities, specificities, covariance and agreement measure (kappa) from the latent class model for Northern Ireland.

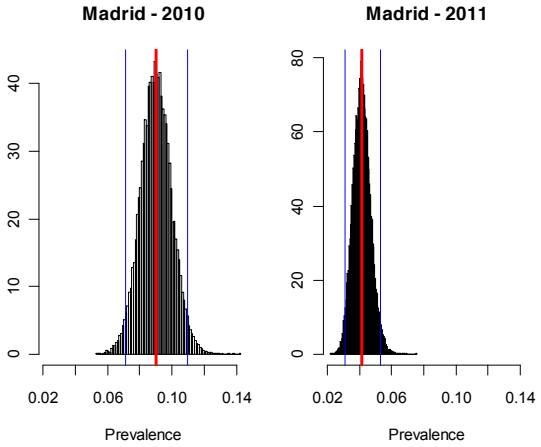

Figure 7: Prevalence estimates from the latent class model for Spain – SICCT, Scenario I.

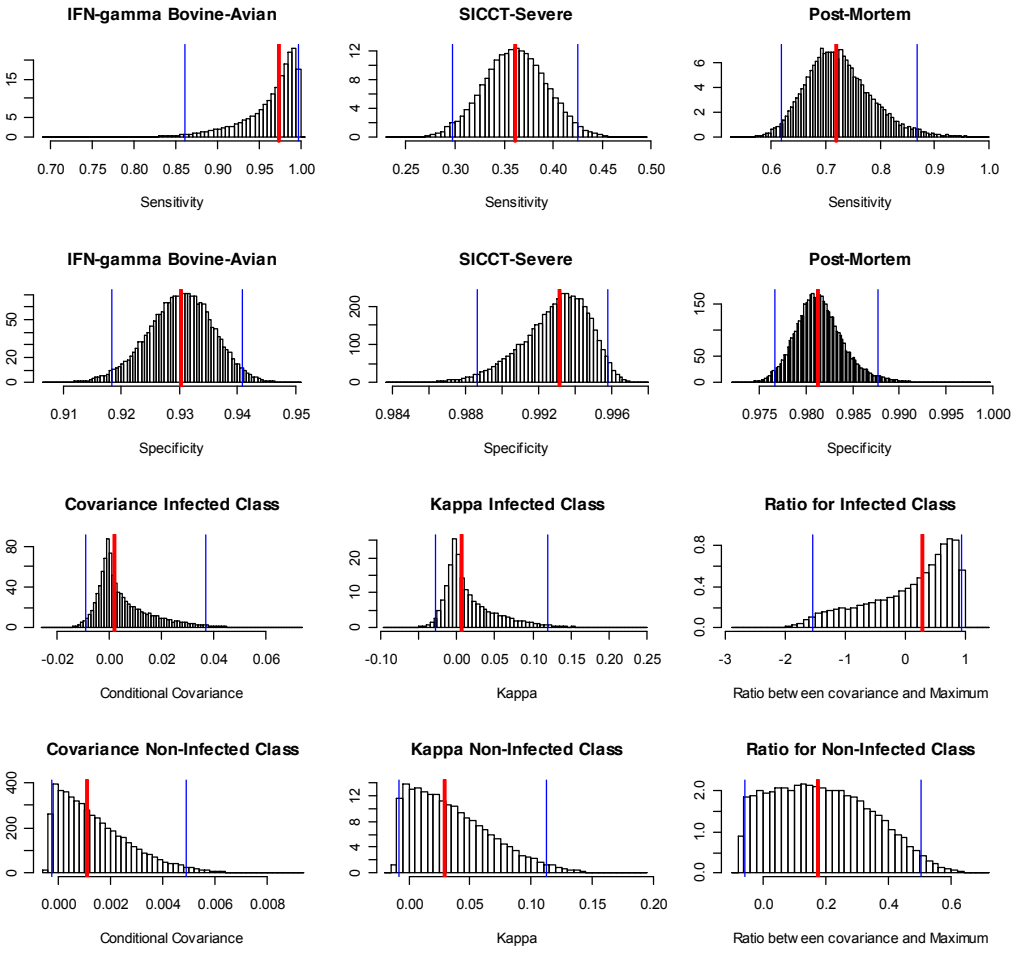

Figure 8: Sensitivities, specificities, covariance and agreement measure (kappa) from the latent class model for Spain, Scenario I.

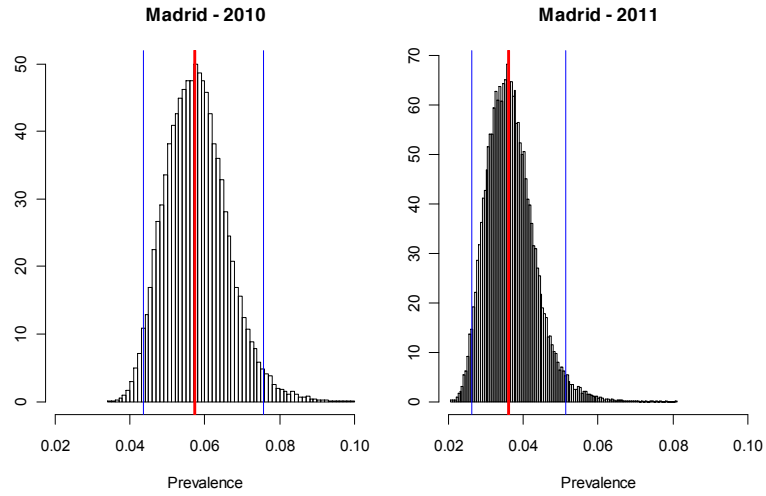

Figure 9: Prevalence estimates from the latent class model for Spain – SIT, Scenario I.

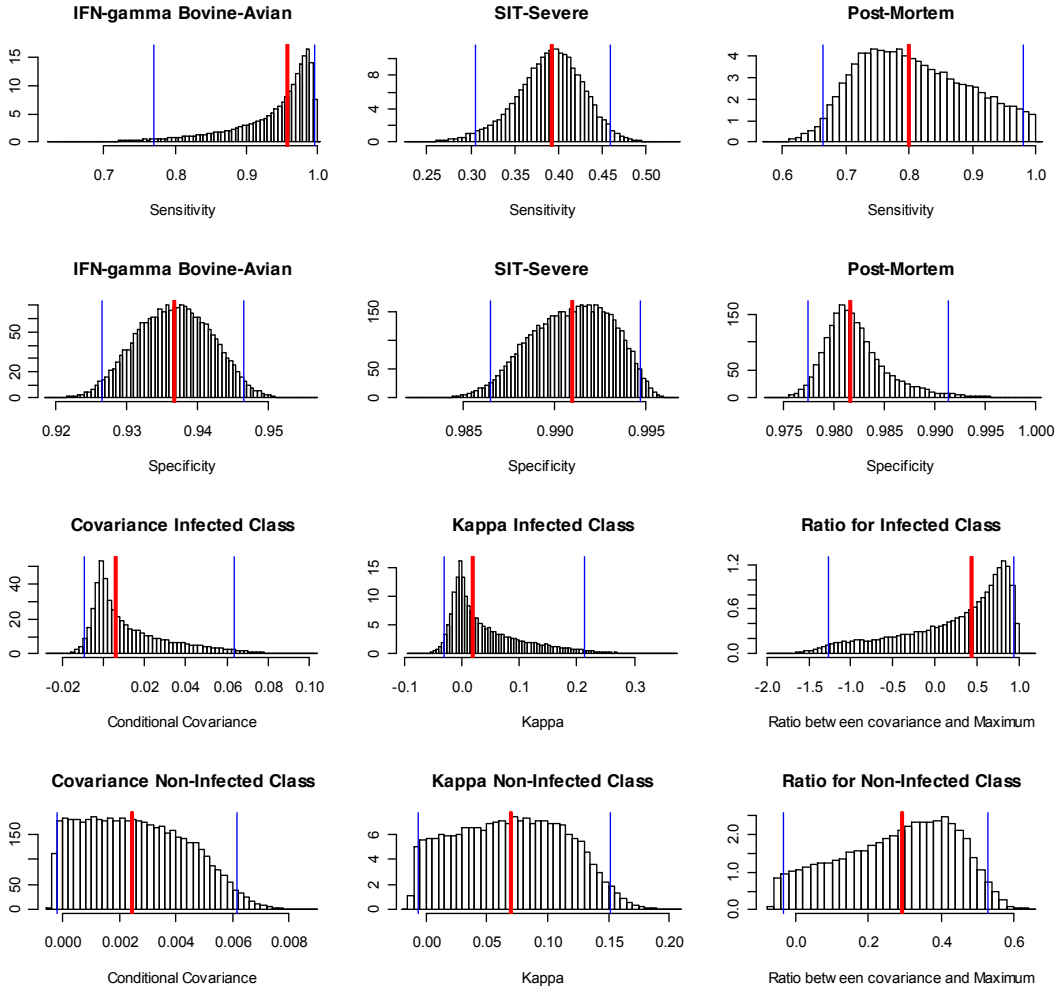

Figure 10: Sensitivities, specificities, covariance and agreement measure (kappa) from the latent class model for Spain – SIT, Scenario I.

ABBREVIATIONS

AHVLA	Animal Health and Veterinary Laboratory Agency
EFSA	European Food Safety Authority
ELISA	Enzyme-linked immunosorbent assay
ELISA-B	Enzyme-linked immunosorbent assay with bovine diagnostic antigen
ELISA-BA	Enzyme-linked immunosorbent assay with bovine PPD and avian PPD diagnostic antigens
ELISA-MPB	Enzyme-linked immunosorbent assay with MPB70 diagnostic antigen
EU	European Union
GLP	Good laboratory practice
GS	Gold standard
IFN-γ	Gamma interferon test
IFN-γ-B	Gamma interferon test with bovine diagnostic antigen
IFN-γ-BACE	Gamma interferon test with bovine PPD and avian PPD diagnostic antigens and CFP10 and ESAT6 diagnostic antigens
IFN-γ-CE	Gamma interferon test with CFP10 and ESAT6 diagnostic antigens
IFN-γ-MPB	Gamma interferon test with MPB70 diagnostic antigen
IFN-γ-BA	Gamma interferon test where the reaction to avian PPD is subtracted from the reaction to bovine PPD
LBBA	Latex bead agglutination assay
LCA	Latent class analysis
Multiplex	Multiplex chemiluminescent immunoassay
NIL_{OD}	OD obtained after stimulation with PBS
NOOUT	Not officially free but no outbreaks
NPV	Negative predictive value
OD	Optical density
OIE	World Organisation for Animal Health
OTFC	Officially tuberculosis-free country
OTFH	Officially tuberculosis-free herd

OTFR	Officially tuberculosis-free region or province
OUT	Outbreaks or reactors in the past 2 years in the same herds
PPD	Purified protein derivative
PPD-based IFN	Gamma interferon test based on purified protein derivative
PPDa	Tuberculin PPD from *M. avium*
$PPDa_{OD}$	OD observed after stimulation with avian PPD
PPDb	Tuberculin PPD from *M. bovis*
$PPDb_{OD}$	OD observed after stimulation with bovine PPD
Rapid	Rapid immunochromatographic assay (rapid test)
SIT	Single intradermal tuberculin test
Skin test	Tuberculin skin test
TB	Tuberculosis
TBOF	Officially tuberculosis free
T0R	Term of reference
WG	*Ad hoc* working group

GLOSSARY

Accuracy: In the diagnostic test context, accuracy is defined as the overall (proportion of) agreement between a (new) test result and the true disease status (gold standard) in a sample of individuals.

Bias: The extent to which a prevalence estimate produced by the surveillance system deviates from the true prevalence value. Bias is reduced as representativeness is increased

Confidence interval: An interval estimate statistically derived from a sample. The interval estimation is designed to include (capture) an unknown (true) population parameter with a certain level of confidence

Credible interval: the Bayesian equivalent of the confidence interval. A 95 % confidence interval states that the estimated probability that the process used to generate the interval includes the correct value of the parameter is 95 %.

Diagnostic sensitivity: the ability of a diagnostic test to correctly identify a single infected animal.

Diagnostic specificity: the ability of a diagnostic test to correctly identify an uninfected animal.

Meta-analysis: A statistical analysis that combines the results of several studies that have addressed the same research question. As combination may increase statistical power of the estimation, results may be a more accurate reflection of the unknown property than those derived from a single study under one set of conditions.

Systematic literature review: Conducting a literature review using prior criteria for searching/selection the literature with scientific tools to assess the findings from the published studies.