Methodology and preliminary results of a systematic literature review of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Corresponding author: Sara Downs

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

Postal address:

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

Abstract

A systematic review was conducted to identify studies with data for statistical meta-analyses of sensitivity (Se) and specificity (Sp) of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis (bTB) in cattle. Members of a working group (WG) developed and tested search criteria and developed a standardised two-stage review process, to identify primary studies with numerator and denominator data for test performance and an agreed range of covariate data. No limits were applied to year, language, region or type of test in initial searches of electronic databases. In stage 1, titles and available abstracts were reviewed. References that complied with stage 1 selection criteria were reviewed in entirety and agreed data were extracted from references that complied with stage 2 selection criteria. At stage 1, 9,782 references were reviewed and 261 (2.6%) passed through to stage 2 where 215 English language references were each randomly allocated to two of 18 WG reviewers and 46 references in other languages were allocated to native speakers. Agreement regarding eligibility between reviewers of the same reference at stage 2 was moderate (Kappa statistic=0.51) and a resolution procedure was conducted. Only 119 references (published 1934-2009) were identified with eligible performance estimates for one or more of 14 different diagnostic test types; despite a comprehensive search strategy and the global impact of bTB. Searches of electronic databases for diagnostic test performance data were found to be nonspecific with regard to identifying references with diagnostic test Se or Sp data. Guidelines for the content of abstracts to research papers reporting diagnostic test performance are presented. The results of meta-analyses of the sensitivity and specificity of the tests, and of an evaluation of the methodological quality of the source references, are presented in accompanying papers (Nuñez-Garcia et al., 2017; Downs et al., 2017).

Keywords: bovine tuberculosis; diagnostic tests; performance; sensitivity; specificity; systematic review

Introduction

Virtually all mammals show susceptibility to infection with *Mycobacterium bovis* and bovine tuberculosis (bTB) is endemic in cattle in many parts of the world (O'Reilly and Daborn, 1995, FAO, 2012, OIE, 2012). Field surveillance for bTB in cattle relies on ante-mortem diagnostic tests that can detect subclinical infection (Monaghan et al., 1994, de la Rua-Domenech et al., 2006, OIE, 2012, EEC, 1964). Tuberculin skin tests which include the Single Intradermal tuberculin (SIT) test, the Single Intradermal Comparative Cervical Tuberculin (SICCT) test and the Caudal Fold (CF) test based on purified protein derivative (PPD) from mycobacterial cultures are standalone ante-mortem tests for diagnosing bTB in cattle and designation of herds as officially bTB free (OFT). Post-mortem meat inspection for lesions characteristic of bTB at routine slaughter is mandatory for every bovine entering the human food chain in most countries. In the EU and elsewhere, post-mortem, culture and/or molecular methods are also used as confirmatory tests where *M. bovis* infection is suspected after a positive ante-mortem test result (de la Rua-Domenech et al., 2006).

Tests vary in their ability to detect different components or phases of the immunological response in an infected animal and their sensitivity (Se), the probability of a positive test if an animal is truly infected, varies accordingly. Specificity (Sp), the probability of a negative test result if an animal is uninfected, also varies with test type. The accuracy (Se and Sp) of diagnostic tests has important implications for the design of bTB control and eradication strategies in cattle (Salman, 2003, pp. 59). However, a range of estimates for Se and Sp for bTB tests of the same type is reported in the veterinary literature. Test performance is influenced not only by specific test characteristics such as test antigen formulation, positive cut-off point but

also pathogenesis stage of infection in the host, differences in test conduct operators, prevalence of cross-reacting environmental mycobacteria and other factors (de la Rua-Domenech et al., 2006, Bezos, 2014, Clegg et al., 2015).

A systematic literature review attempts to bring together all work in a subject area and to provide an impartial, objective and accurate assessment of evidence. The methodology has been used extensively by Cochrane and others, in conjunction with statistical meta-analyses, to generate summary estimates for the efficacy of healthcare interventions (Bero et al., 1998, Sargeant et al., 2006, Liberati et al., 2009, Sargeant et al., 2010). There has been less work attempting to summarise estimates of the performance of diagnostic tests in domestic animals, although the methodology has been used to obtain pooled estimates of the performance of tests of pulmonary tuberculosis in humans (Steingart et al., 2007, Leeflang et al., 2008) and bTB in deer (EFSA, 2008). Our aim was to develop robust search methodology to identify primary research sources for estimates of Se and Sp of ante- and post-mortem diagnostic tests for bTB in cattle. These data were to be used in statistical meta-analyses to provide summary estimates of performance of different test types (Nu□ez-Garcia et al., 2017).

Materials and methods

Working Group (WG) and overview

A standardised process of review was discussed, developed and agreed at WG meetings. A two-stage review of relevant literature was conducted. Stage 1 was conducted through reading the abstract of a reference or title if the abstract was

unavailable. Stage 2 was a more detailed review of entire references that had passed through stage 1 and it led to the extraction of numerator and denominator data for the calculation of Se and/or Sp and other data that the WG had identified as influential on test performance. The methodology was adapted from an approach taken previously in a review of the performance of diagnostic tests for bTB in deer (EFSA, 2008) and from guidelines for systematic review and meta-analyses (Irwig et al., 1994, Stroup et al., 2000, Bossuyt et al., 2004, Westwood et al., 2005).

The WG included 22 scientists. The expertise of the group can be broadly summarised as follows: 12 epidemiologists (10 of whom were also veterinarians), four immunologists specialising in the development of diagnostic tests, one pathologist (also a veterinarian), two bacteriologists, one bioinformatician, one livestock geneticist and one biologist who had specialised in the development of bTB databases for over 15 years.

Two linked, bespoke databases were developed for stages 1 and 2 (VLA 2010, Appendix 1). Communication between WG members was facilitated by a series of workshops, email and a web-based Sharepoint portal.

Stage 1 review

a. Overview

The stage 1 review was a review of titles and abstracts (where available) to identify references likely to be relevant for detailed (stage 2) consideration. Inclusion and

exclusion criteria were discussed and agreed at the first WG meeting (see a. below). A search strategy was developed that included testing different electronic search strings (see b. below).

b. Inclusion and exclusion criteria at stage 1

Inclusion criteria:

- The reference related to primary research (therefore excluding review papers)
- The reference included either report(s) of Se and/or Sp of a diagnostic test for bTB, or provided data enabling these statistics to be calculated
- The diagnostic test performance was measured on bovines

Exclusion criteria:

The Se estimates were from studies where cattle had been experimentally infected with *M. bovis*

c. Sources for references and search strategy

Sources for 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 (not in the public domain) that may not have undergone peer review and reports identified from research institutions and laboratories (grey literature). A list of research institutions, Government Departments, agencies and laboratories known to members of the working group and from searches on the World Wide Web was compiled and a member of the WG contacted

each by email asking for sight of reports and other references that reported the performance of diagnostic tests for bTB. VLA librarians attempted to obtain other grey literature that potentially contained performance data identified during the searches of electronic databases e.g. abstracts published in conference proceedings.

3. References known to WG members including an electronic database at the Animal and Plant Health Agency (APHA, previously the Veterinary Laboratories Agency, VLA) that listed references to bTB research
4. Bibliographies in the reports and references already retrieved

Different algorithms of search strings with variations of search terms for bovine tuberculosis and different names for diagnostic tests for bTB were tested in the electronic databases. No limits were placed in terms of diagnostic test type, year of publication or region. The different search strings were compared in terms of the i) their ability to maximise the proportion of a list of 65 references known to contain performance data they identified (sensitivity) and ii) the total number of references identified in the electronic databases.

d. Standardisation of the stage 1 review

Guidelines were developed to standardise the review of abstracts before the stage 1 review itself was conducted (VLA 2010, Appendix 2). To facilitate the development of the guidelines, five studies were conducted to assess repeatability between reviewers. In the first four of these studies, 100 abstracts were allocated to three or four reviewers. Agreement between reviewers was measured using a 2-way Anova model and intra-class correlation coefficients were calculated (Landis and Koch,

1977) and the guidelines were clarified after each study to attempt to improve repeatability. In the fifth study, 500 references were reviewed by the two WG members who were to conduct the stage 1 review of all references.

Stage 2 review

a. Overview

Entire references were obtained for those that passed through the stage 1 review. References written in English were randomly allocated to two WG members whilst references written in Spanish or German were only allocated to native speakers in the WG. Volunteer scientists at the APHA who were native speakers read references in languages other than English, Spanish or German. Relevant data were obtained from the volunteers during structured interviews with WG members.

Each reviewer was asked to evaluate each reference paper or report against agreed inclusion and exclusion criteria (see b. below). Reason(s) for rejection, if relevant, were recorded. Otherwise, data were entered into the stage 2 database according to a guide developed for data-entry (VLA 2010, Appendix 3). Data entered included information about the study population, test characteristics, characteristics of the reference standard and the numerator and denominator data required to estimate Se and/or Sp (VLA 2010, Appendix 1).

b. Inclusion and exclusion criteria at stage 2

Inclusion criteria for Se estimates:

- Se could be calculated

- The bovine population had been naturally exposed to bTB
- Each study animal had been individually examined using one of the following positive reference tests: post mortem examination (PM) (meat inspection or detailed laboratory inspection), culture, microscopic inspection (histology or histopathology), SICCT test

Exclusion criteria for Se estimates:

- 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*) and there were no results from acceptable reference tests on each animal

Inclusion criteria for Sp estimates:

- Sp could be calculated
- There was good evidence that the bovine population was free from infection with, and exposure to, *M. bovis*, including herds with Officially Tuberculosis Free (OTF) status, herds from OTF area or OTF country, herds from a non-endemic bTB area where the authors stated that the area had been free of bTB for several years, or herds that in authors' opinion were tuberculosis-free and had been free for several years

Exclusion criteria for Sp estimates:

- Any other evidence of lack of exposure to bTB, including groups of animals with negative tuberculin skin tests of all individual animals in an area where bTB was endemic or had existed in the recent past, negative culture in some animals from the herd in an area where bTB was endemic or had existed in the

recent past, negative PM in some animals from the herd in an area where bTB was endemic or had existed in the recent past, or herds of apparently healthy animals with no other evidence suggesting bTB freedom (VLA 2010, Appendix 5).

c. Resolution of differences after stage 2 review

The data from each reference that potentially contained eligible data were entered into the stage 2 database, by the reviewers using instructions for data-entry including screen shots of the database (VLA 2010, Appendix 3). Two WG members who had not conducted any stage 2 reviews compared data from references reviewed by two reviewers. Agreement between reviewers was measured at different levels of detail:

Level 1: Whether the reference was eligible based on Se and Sp inclusion and exclusion criteria

If both reviewers agreed that the reference was eligible, agreement at levels 2 and 3 was also measured:

Level 2: The number of Se and/or Sp estimates that could be extracted by test type.

Level 3: The Se and/or Sp numerator and denominator data for each test-type and of population.

Each reference that had been reviewed by two reviewers was then randomly assigned one lead reviewer from the pair. The lead reviewer led the resolution of inconsistencies during a WG meeting dedicated to this procedure, and through telephone and email correspondence; and returned a database with revised corrected data to APHA. Remaining disagreements between data in one or more descriptive

covariates were resolved by an APHA WG member by cross-checking related fields and otherwise accepting the data entered by the lead reviewer as the final response.

d. Latent class analysis (LCA)

The following criteria were applied to the primary studies identified in stage 2 to identify those where sensitivity and specificity could be estimated using LCA (Toft et al., 2005); which could include studies where a gold standard or reference test was included but nevertheless met the following criteria::

a) Observed frequencies (counts) are reported for all possible combinations of positive or negative results for each (sub) population.
b) Animals were sampled from (sub) populations, i.e. no pooled serum panels were used.
c) Animals tested have not been pre-selected according to any diagnostic tests for for bTB; i.e. if any (sub) populations had been used for the calculation of test performance they were defined epidemiologically (e.g. observed or stipulated risk factors, epidemiologically distinct production units).
d) It could be assumed that each animal tested in each of the (sub) populations had the same unknown, but non-zero, probability of being infected with *M. bovis* and the unknown infection prevalences varied among the (sub) populations.
e) The tests were applied in parallel; i.e. all test were applied to all animals and the decision to conduct a second or third test was independent on any other test results.
f) The simplifying assumption was justified that the sensitivity and specificity of each tests was the same in all (sub) populations within a study;

g) It was justified to consider that the diagnostic tests used on all animals in the study were conditionally independent; i.e. the application of or result from one test was not associated with application or result from another test in the study and false positive and false negative errors were uncorrelated (Gardner et al., 2010).

e. Assessment of study quality

The WG decided that a review of the methodological quality of studies of diagnostic tests should be also conducted. As a result of this discussion, the QUADAS instrument developed by Whiting and others (Whiting et al., 2003) for assessing quality of studies of diagnostic test performance was modified for a veterinary context (VLA 2010, Appendix 4) and was incorporated into the stage 2 database. The assessment of the methodological quality of references is reported elsewhere (Downs et al., 2017).

Statistical analyses

Summary distributions of data were examined using Microsoft Excel 2011. Calculation of the intra-class correlation coefficient using a 2-way Anova model was conducted using Stata release 12.1 (StataCorp) and interpreted according to Landis and Koch, 1977, where values<0 indicate no agreement, values 0-0.20 as slight, 0.21-40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial and 0.81-1.0 as almost perfect.

Results

Stage 1 review

a. Search strings

The number of references identified in electronic databases using the search strings tested ranged from 5,354 to 11,772. The search string eventually used to select references to be reviewed at stage 1 (see below) identified 62/65 of the test references. Of the three test references of the 65 not identified by the search, two were publications from the grey literature which were not indexed on any of the databases searched through Web of Knowledge and Dialog and could only be retrieved from the Defra website (Vordermeier and Ewer, 2006, Veterinary Laboratories Agency, 2006). The other reference was a Norwegian paper published in 1949 (Holth, 1949) which described a diagnostic test no longer used that was classed as ineligible for inclusion in the systematic review by the WG (SE3238 Annex 1 Ineligible Test).

The final search string run on 1st December 2008 (with no date or language limitations) was:

(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)

*indicates truncation or stemming. This technique was used to broaden the search to include various word endings to the root of the word.

The number of references identified by applying the above search string to the Web of Knowledge and associated electronic databases was 8,089. Adding additional references identified using the same search string through searching Dialog increased this total to 9,756. Hand searching, referrals from the WG and external research institutions resulted in a total on the stage 1 database of 9,782. Duplicates were removed at intervals as the database was compiled.

b. Review of abstracts and selection of references at stage 1

b i) Results of study to standardise stage 1 review

Agreement between reviewers ranged between poor and almost perfect with no discernible trend in improvement in level of agreement after each study despite revision of guidelines for reviewing abstracts (Table 1). In the fifth study reviewer A selected twice the number of abstracts as reviewer C. However, agreement was classified as moderate (Landis and Koch, 1977) because all but one (14/15) abstracts selected by reviewer C were also selected by reviewer A. The data suggested that the total number of abstracts classed as complying with Stage 1 inclusion and exclusion criteria increased with the number of reviewers reviewing abstracts, but there was a threshold over which the number of reviewers did not increase the number of abstracts identified as complying with inclusion and exclusion criteria e.g. there was no

Table 1

Agreement between reviewers that abstracts passed stage 1 eligibility criteria

Study	Abstracts reviewed [a]	Abstracts that passed stage 1 Reviewer					
		A	B	C	D	Intraclass correlaton coefficient	Agreement classification [b]
1	98	5	6	7	[d]	0.59[c]	Moderate
2	99	3	2	3	3	0.29[c]	Poor
3	100	9	7	6	10	0.84	Almost perfect
4	100	5	3	4	2	0.46	Moderate
5	500	34	[d]	15	[d]	0.55	Moderate

Footnote to Table 1:
[a] Randomly selected from electronic searches
[b] (Landis and Koch, 1977)
[c] Up to 2 abstracts of the 100 allocated were omitted by one reviewer in study
[d] Indicates that the reviewer did not participate in this study

Table 2

Cumulative totals of abstracts that were passed in development of stage 1 selection criteria

Study	Abstracts[a] that passed stage 1 by reviewer			
	A	A+B	A+B+C	A+B+C+D
1	5	8	10	[b]
2	3	5	7	7
3	9	9	10	10
4	5	7	8	8
5	34	[b]	35	[b]

Footnote to Table 2
[a] Randomly selected from electronic searches. A, B, C and D are the four different reviewers that tested stage 1 selection criteria
[b] No review conducted

increase in the number of abstracts identified by three compared to four reviewers (Table 2). As a result of these studies it was agreed by the WG group that:

i) All abstracts to references identified through electronic searches were to be reviewed by two WG members

ii) Any references identified by either reviewer through review of the abstract as possibly containing information on test performance be passed for detailed review at stage 2.

b ii) Review of stage 1 database of references

The procedure followed is shown in Figure 1. Of the 9,782 references identified from all sources, over 90% had an abstract available. Of the references identified by the electronic searches approximately 80% were from journals or books, and 20% were grey literature (10% reports and 10% other). Over 97% of references reviewed by both reviewers were initially excluded by one or other reviewer, as unlikely to comply with our selection criteria, and only 3% were classified by one or both reviewers as likely to contain eligible data. The reasons for exclusion were re-examined and further review of the abstracts (and titles if no abstract was available) conducted to attempt to ensure that references likely to have eligible test performance data had not been excluded in error.

During the stage 1 review, 87% (8541/9782) of references were excluded by both reviewers as being the wrong subject material; 7% (717/9782) of references were initially excluded because an abstract was not available and there was insufficient information in the title with which to determine eligibility and 3% (257/9782) of references were excluded on the basis that it was unlikely that the reference would

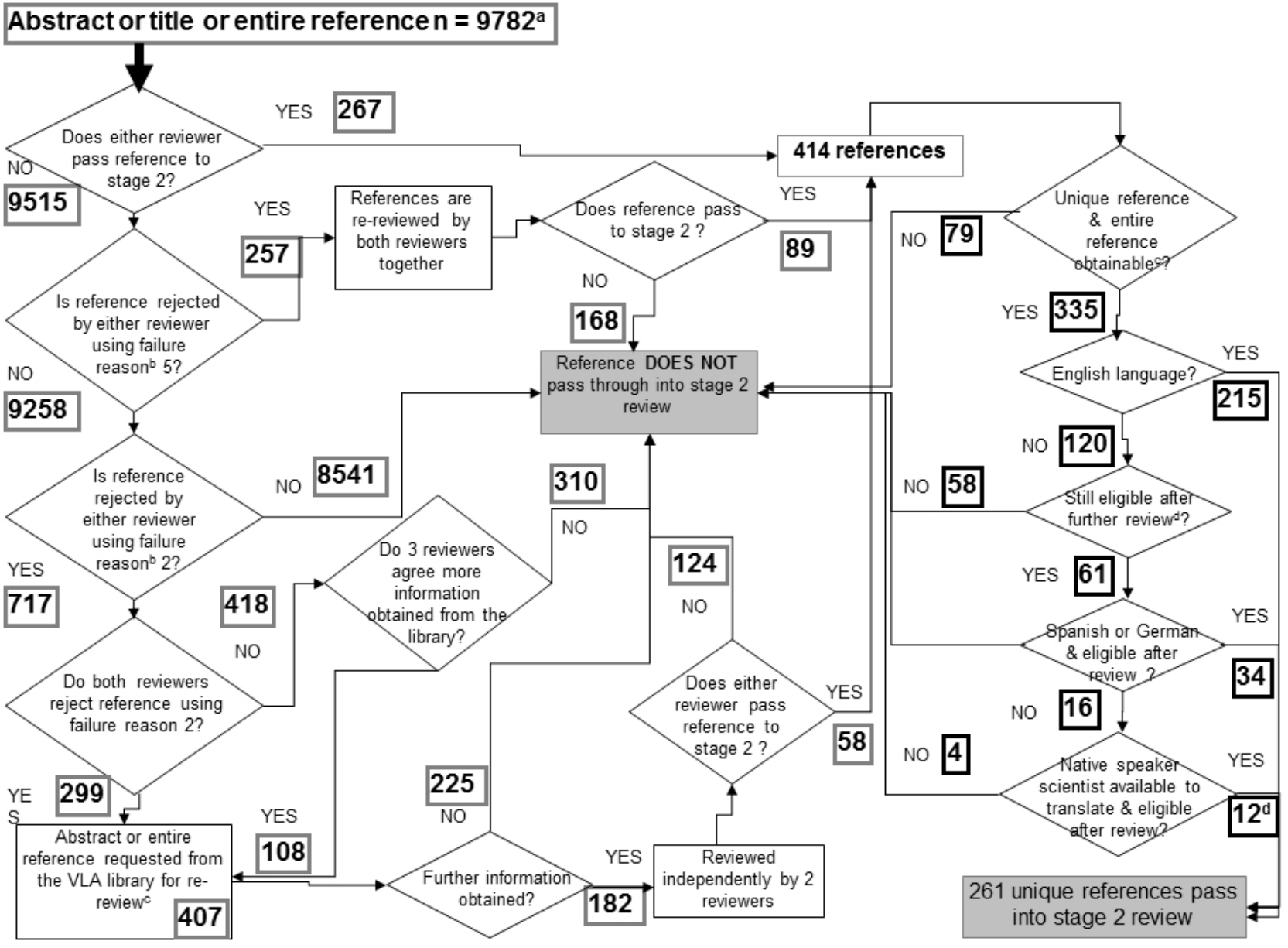

Abstract or title or entire reference n = 9782[a]
Does either reviewer pass reference to stage 2?
YES 267
NO 9515
414 references
YES 257
References are re-reviewed by both reviewers together
Does reference pass to stage 2 ?
YES 89
NO 168
Is reference rejected by either reviewer using failure reason[b] 5?
NO 9258
Reference DOES NOT pass through into stage 2 review
Is reference rejected by either reviewer using failure reason[b] 2?
NO 8541
YES 717
Do both reviewers reject reference using failure reason 2?
418 NO
Do 3 reviewers agree more information obtained from the library?
310 NO
124 NO
Does either reviewer pass reference to stage 2 ?
YES 58
YES 299
Abstract or entire reference requested from the VLA library for re-review[c] 407
YES 108
225 NO
Further information obtained?
YES 182
Reviewed independently by 2 reviewers
Unique reference & entire reference obtainable?
NO 79
YES 335
English language?
YES 215
NO 120
Still eligible after further review[d]?
NO 58
YES 61
Spanish or German & eligible after review ?
YES 34
NO 16
Native speaker scientist available to translate & eligible after review?
NO 4
YES 12[d]
261 unique references pass into stage 2 review

Fig. 1. Decision tree showing how references with eligible data were selected during the systematic review

Footnote to Fig. 1.
[a] Identified through searches of electronic bibliographies and other sources
[b] Failure reasons 1: Does not contain performance information for diagnostic test for bTB in bovines (wrong subject material) 2. Insufficient information in record 3: Surveillance or prevalence report 4: Review not primary study 5: Not possible to calculate either Se of Sp
[c] By APHA library using reasonable means, cost and effort
[d] By three WG members in the light of stage 2 exclusion criteria
[e]French, Chinese, Polish, Dutch and Italian references reviewed by APHA scientific staff

contain enough data to calculate test performance (e.g. prevalence or surveillance study or review). Reasonable attempts were made to obtain entire references where there was no abstract and only a title available and both reviewers considered that there was insufficient information to judge eligibility (407 references).

Of the references that appeared to have eligible data after stage 1 (Figure 1), 37% (153/414) were excluded for the following reasons: around 11% because they were duplicates, 12% based on further consideration of exclusion criteria by other WG members, 9% because the reference could not be obtained through reasonable means, and 4% because a native speaker to review the reference could not be found.

Stage 2 review

a. Results of review

There were 261 references that passed into stage 2, of which 5% (12/261) were from unpublished sources (grey literature). The remaining 249 references were published in peer reviewed scientific press. Of references that passed through to stage 2, 215 English language references were each randomly allocated to two of the 18 reviewers within the WG. Sixteen and 18 references respectively were reviewed by two Spanish native speakers and one German native speaker in the WG. A further 12 references (six Chinese, two Italian, two Dutch, one French and one Polish) were reviewed by native speakers among APHA veterinarians and scientists. Sixty-one % (160/261) of references were initially classed as having eligible Se and/or Sp data by at least one WG reviewer. Of the references reviewed by two WG reviewers, 71% (173/241) were initially classed by both reviewers as containing eligible Se and/or Sp data (Table 3).

Table 3

Results of stage 2 review of references by pairs of reviewers

Stage 2 phase	Number of references[a]
Reviewed by two [b] WG members	241
Both WG reviewers considered that a reference should be included or excluded based on eligibility criteria	173 (71%[c])
One WG reviewer classed reference as having eligible data but other did not	68 (28%)
Sensitivity and/or specificity data considered eligible following resolution procedure	119 (49%)

Footnote to Table 3:
[a]15 foreign language references were reviewed by one reviewer. Of these three references were classified by the reviewer as containing eligible test performance data
[b] References reviewed by two of 18 reviewers randomly allocated to English language references and otherwise assigned to native speakers.
[c] Kappa statistic=0.512 (moderate agreement)
During the review process a reference reviewed by two reviewers might be included by one and excluded by other. However after the resolution process a reference would either be included or excluded.

Following the resolution process 45% (119/261) of references were identified as having eligible Se and/or Sp data and 5.9% (7/119) were from grey literature. Of the 119 references, 100 were written in English, 9 in Spanish, four in German and the remainder in Chinese, Dutch, French, Italian, Portuguese or Polish. Twenty-five % (30/119) of the references contained eligible data for Se and Sp, 61% (73/119) for Se only and 6% (16/261) for Sp only. Some references contained more than one performance estimate for the same test, for different types of tests or test modifications. Over 10% of references measuring Se and Sp of the IFN-γ blood test using Bovine PPD - Avian PPD diagnostic antigen had more than 7 and 19 estimates respectively.

b. References with eligible estimates of Se and/or Sp

Table 4 shows the number of references with eligible estimates for Se and/or Sp by test type. Individual references could contain one or more estimates for one or more test types. The WG identified 14 different diagnostic test types in total. Lists of the references with eligible performance data for each test type and plots showing the estimates for sensitivity and specificity by reference and test type are in the supplementary material.

The tuberculin skin tests and the IFN-γ and ELISA blood tests included modifications of the basic test format. The Se and Sp of the SICCT test were measured at different cut-off values for a positive response (standard and severe interpretation). An even wider range of cut-off values and algorithms was used to define a positive response in the references reporting the performance of blood tests e.g IFN-γ and ELISA blood

Table 4 Distribution of references [a] with eligible estimates of Sensitivity and Specificity by year of publication

Test Name	Sensitivity		Specificity		Percentage of references by year of publication									
	References	Estimates	References	Estimates	1931-69		1970-79		1980-89		1990-99		2000-10	
	n	n	n	n	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
Single intradermal skin test	8	18	4	10	50.0	25.0	25.0	0.0	0.0	0.0	12.5	0.0	12.5	75.0
SICCT test	15	46	8	18	13.3	0.0	20.0	12.5	13.3	0.0	20.0	37.5	33.3	50.0
Caudal fold skin test	17	75	2	3	11.8	0.0	23.5	50.0	11.8	0.0	35.3	50.0	17.7	0.0
IFN-γ blood tests	27	172	19	145	0.0	0.0	0.0	0.0	0.0	0.0	40.7	26.3	59.3	73.7
ELISA	23	62	12	27	0.0	0.0	0.0	0.0	26.1	25.0	47.8	50.0	26.1	25.0
Rapid test	0	0	2	3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Latex bead agglutination assay	2	3	1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
Multiplex immunoassay	1	5	1	4	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
Glutaraldehyde	1	1	0	0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Fluorescence polarization assay	2	2	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Necropsy [b]	8	14	1	3	0.0	0.0	12.5	0.0	25.0	10.0	0.0	0.0	62.5	0.0
Microscopic examination	13	21	1	1	15.4	100.0	15.4	0.0	15.4	0.0	7.7	0.0	46.2	0.0
Culture	8	16	1	1	0.0	0.0	25.0	0.0	37.5	0.0	0.0	0.0	37.5	100.0
PCR	12	25	4	5	0.0	0.0	0.0	0.0	0.0	0.0	25.0	50.0	75.0	50.0

Footnote to Table 4:
[a] All references are listed in the online supplement.
single intradermal comparative cervical tuberculin skin (SICCT) test
[b] Includes meat inspection and detailed/laboratory based examination
[c] Includes histopathology

tests. Additionally the blood tests varied by diagnostic antigen composition. Around 6%, 43%, 6% and 4% of IFN-γ test performance estimates were based on the Bovine PPD, Bovine PPD-Avian PPD, ESAT6/CFP10 and MPB70 diagnostic antigens respectively, with the remainder a heterogeneous collection. All the skin tests in the review were based on response i.e. change in skin thickness to a intradermal injection of Bovine PPD (the SIT test and CF test) or Bovine PPD minus Avian PPD (the SICCT test).

The choice of the positive reference standard and the level of evidence that a population was free from *M. bovis* infection (the negative reference standard) varied (Figures 2 and 3). Sp was measured in Officially Tuberculosis Free populations in 20, 31.8, 21.4, 33.3 and 100% of percent of references respectively reporting performance of the SIT test, IFN-γ blood test, ELISA blood test, Rapid blood test and LLBA . Although all estimates of Sp were from bTB-free populations, many studies used negative test results from the SICCT test and post-mortem tests as additional criteria for freedom (results not shown). A positive SICCT test result was used as a cattle selection criterion in 33.3, 26.1, 33.3, 21.5, 25.0 and 16.7% of studies measuring Se of the IFN-γ blood test, ELISA blood test, necropsy, microscopic examination, culture of *M. bovis* and PCR respectively and therefore could be considered as part of the reference standard in these studies.

Many factors, known or suspected by the WG as influencing test performance were not reported or not reported in a consistent manner. Ninety percent of the studies did not indicate cattle breed or production class. Sixty percent of references about the IFN-γ blood test did not indicate whether a tuberculin skin test had been conducted

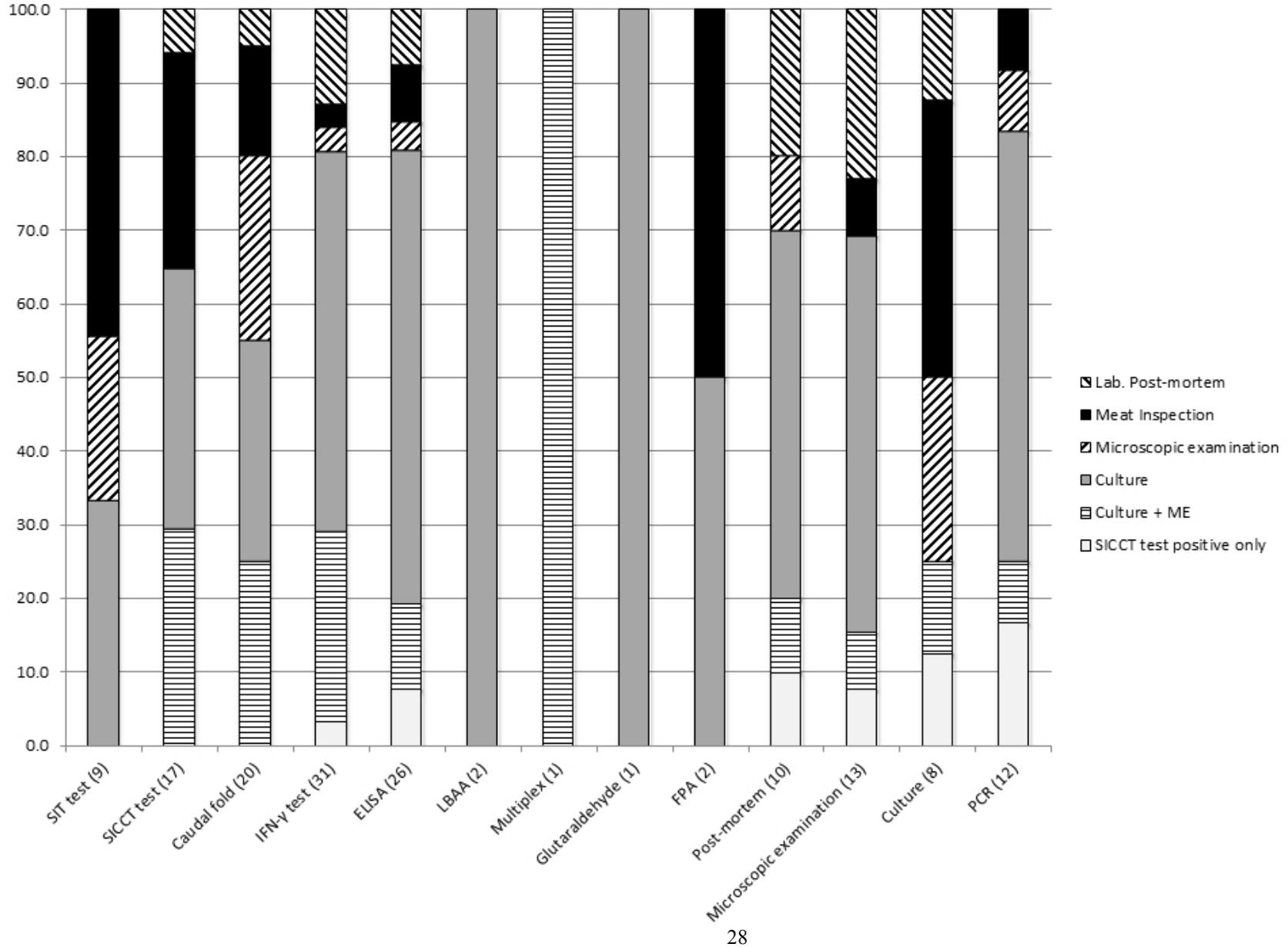

100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
SIT test (9)
SICCT test (17)
Caudal fold (20)
IFN-γ test (31)
ELISA (26)
LBAA (2)
Multiplex (1)
Glutaraldehyde (1)
FPA (2)
Post-mortem (10)
Microscopic examination (13)
Culture (8)
PCR (12)
Lab. Post-mortem
Meat Inspection
Microscopic examination
Culture
Culture + ME
SICCT test positive only

Fig. 2. Distribution of positive reference standards across references with sensitivity estimates

Footnote to Fig. 2.
The number in brackets indicates the denominator, which is the number of reference standards used in references that report Se of the test. This is sometimes larger than the number of references with at least one Se estimates of test because some references reported Se against more than one reference standard.
Lab. Post-mortem = detailed or laboratory based necropsy as opposed to meat Inspection
ME=Microscopic examination such as histopathology
SICCT test= Single Intradermal Comparative Cervical Tuberculin test where the change in skin fold thickness in response to Bovine PDD- response to Avian PPD > 4mm for a positive test

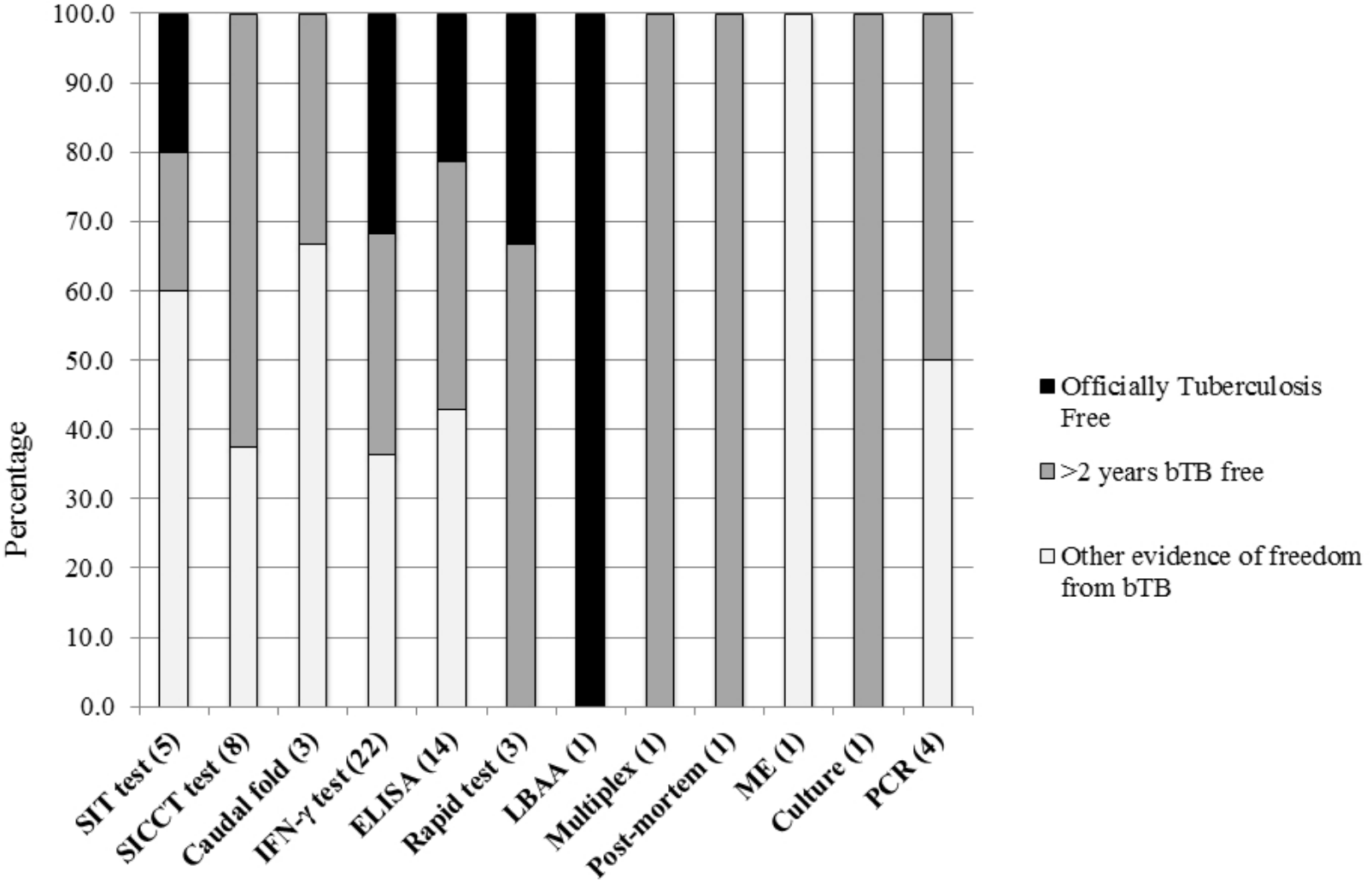

100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
Percentage
SIT test (5)
SICCT test (8)
Caudal fold (3)
IFN-γ test (22)
ELISA (14)
Rapid test (3)
LBAA (1)
Multiplex (1)
Post-mortem (1)
ME (1)
Culture (1)
PCR (4)
Officially Tuberculosis Free
>2 years bTB free
Other evidence of freedom from bTB

Fig. 3. Distribution of levels of evidence that the study population was bovine tuberculosis free across references with specificity estimates

Footnote to Fig. 3.
The number in brackets indicates the denominator, which is the number of reference standards used in references that report Sp of the test. This is sometimes larger than the number of references with at least one Sp estimates of test because some references reported Sp for more than one population and for different levels of evidence that the populations were bovine tuberculosis (bTB) free
LBAA=Latex bead agglutination assay
ME=Microscopic examination such as histopathology
SICCT test= Single Intradermal Comparative Cervical Tuberculin test where the change in skin fold thickness in response to Bovine PDD- response to Avian PPD > 4mm for a positive test

beforehand. It was not possible to compare bTB prevalence in the Se studies because the scales on which change was measured varied and could not be standardised in a meaningful way.

Study populations where Se was estimated were in general smaller than those estimating Sp. The median number of cattle designated truly positive was 46 (Inter Quartile Range (IQR) 22 to 212) in the Se studies compared with a median of 200 cattle (IQR 25 to 904) designated as truly negative in the Sp studies. Overall, crude Se estimates extracted from the references ranged from 0 to 100% (median 79.3, IQR 57.5 to 91.3) and for Sp ranged from 24.1 to 100% (median 97.0, IQR 90.9 to 100). Summary estimates of Se and Sp by test type calculated through statistical meta-analysis are reported in the accompanying paper (Nu□ez-Garcia et al., 2017).

c. Studies with suitable data for Latent Class Analysis

There were 66 references identified by at least one reviewer during the stage 2 review as reporting a study with results of two or more diagnostic tests for cattle sampled from one or more populations. Detailed review by a WG member determined that only six references met all LCA inclusion criteria. The remaining 60 failed inclusion criterion a) and three also failed inclusion criterion c). Bayesian latent class models were fitted to the results of the six studies that met inclusion criteria (Claxton et al., 1979, Wood et al., 1991, Wood et al., 1992, Gaborick et al., 1996, Liebana et al., 2008, Mumtaz et al., 2008). The inclusion of test performance estimates from the LCA, were explored in a meta-analysis reported in Nu□ez-Garcia et al 2017).

Discussion

In a systematic review of scientific literature to identify primary studies from which the Se and Sp of diagnostic tests for bTB in cattle could be calculated, 14 different types of contemporary tests were distinguished (Table 4). Only 1% of references identified in the initial search of electronic databases contained eligible performance data despite attempts to improve the accuracy of electronic searches and reviewer agreement. There were relatively few estimates for most test types including the performance of the tuberculin skin tests, which are official tests for measuring bTB incidence worldwide. Reference standards against which performance was estimated varied between studies (Figures 2 and 3), including those of the same test type, and the information provided about other factors known to influence test performance was inconsistent. There was no common scale over which performance of blood tests measuring immunological parameters could be compared. Furthermore, cut-off values to define a positive response to tests often varied between studies.

We set out to design a comprehensive search for published and unpublished studies reporting test performance. Publication bias leading to over-representation of reports of statistically significant findings is well known (Saveleva and Selinski, 2008). With diagnostic tests, there may be a publication bias towards reporting studies with better estimates of test performance (Deeks et al., 2005). In an attempt to reduce publication bias, at stage 1, both published and unpublished literature was obtained and references were not excluded on the basis of test type, date of research and region. We tried to improve the specificity of our search string for the electronic databases, but could not improve it substantially without omitting references that contained eligible data and over 9,000 abstracts were reviewed. Others have also reported difficulty in designing

searches in electronic databases for studies reporting diagnostic test performance that are specific without compromising sensitivity (Leeflang et al., 2008).

Fewer than 300 references from the 9,782 reviewed at stage 1 passed through to stage 2. Stage 1 reviewers excluded 85% of references on the basis of 'wrong subject material' reflecting both inadequate reporting in abstracts and poor specificity of search terms. Inadequate reporting in abstracts of diagnostic accuracy studies has been documented elsewhere (Korevaar et al., 2015). Checklists for the information that should be reported both in reference papers and abstracts have been developed, the most well-known being STARD (Standards for Reporting Diagnostic Accuracy) (Bossuyt et al., 2004; Vandenbroucke et al., 2007; Bossuyt et al., 2015). Our study also suggests that the efficacy of electronic searches for diagnostic test performance would be improved by requiring abstracts to comply with minimum reporting criteria. One problem is that the words 'Se' and 'Sp' are used in a variety of contexts and their use is not limited to their technical meaning within the diagnostic test context. Inclusion of the words "denominator" and "numerator" with the associated raw data in abstracts and coupled with Se and Sp as index terms could be explored as an avenue to improve the accuracy of searches (Table 5).

Levels of agreement between reviewers as low as 20% have been reported in systematic reviews, signifying that relying on one reviewer to determine eligibility may lead to substantial misclassification (Chalmers, 1991). At stage 2 in the current project, over 80% of references were randomly allocated to two reviewers who reviewed entire references independently using a standardised procedure before comparing assessments. Agreement regarding whether a reference should be included

Table 5

Proposal for minimum content required of abstracts reporting results from studies measuring the Sensitivity and Specificity of performance of diagnostic tests in cattle

1	Usual name of index diagnostic test/s and active diagnostic reagent
2	Usual name of test to be used as a reference standard (if relevant)
3	Cattle breed and/or production type
4	Estimate of background disease prevalence as a percentage of cattle population (for SE study) , or evidence of freedom from disease (Sp study)
5	Sampling method of cattle population (census, random, purposive/ haphazard/convenience)
6	Numerator – being the number of cattle that test positive (Se study) or the number of animals that test negative (Sp study)
7	Denominator - being the number of cattle positive or negative based on the reference standard (if relevant)
8	Definition for cut-off value/s for a positive response

or excluded was moderate (Landis and Koch, 1977) and the resolution procedure led to almost 40% of references that were initially classified by at least one reviewer as having performance data, being excluded. Discussion within the WG about the possible paucity of studies with eligible data may have predisposed reviewers to initially retain studies if compliance with eligibility criteria was ambiguous, thereby delaying exclusion of such studies until after discussion with fellow reviewers.

Dynamic biological processes such as disease pathogenesis (affecting Se) and interference by colonisation by environmental mycobacteria (affecting Sp) are difficult to measure and control for when evaluating test performancefrom. Other factors such as time interval between tests, diagnostic reagent, animal breed, etc. may be more easily measured. We tried to capture as much information as possible about factors that could explain heterogeneity in estimates of test performance and could be controlled for in a multivariable statistical analysis. Many factors known or suspected by the WG as influencing test performance were not reported or not reported consistently however, limiting exploration or adjustment for influential co-factors (Nu□ez-Garcia et al., 2017). Selection criteria for studies from which Sp could be calculated included a requirement that the Sp was estimated in a population sample where there was good evidence that the population was bTB free, since the parameter relies on comparing the number of cattle that were negative to the diagnostic test under evaluation to the number of truly negative cattle. This is likely to have improved the internal validity of our Sp estimates. However, it also means that test Sp may be over-estimated (and less externally valid) for populations infected with bTB since the estimates will not take account of the higher risk of false positives in an infected population. Ideally test performance is evaluated in a population as similar as

possible to the population in which the test is eventually to be used (Berkvens et al,.2006). Furthermore, the majority of references for the IFN-γ and ELISA blood tests reported more than one performance estimate, each relating to a different cut-off value for a positive response. The algorithms used to define a positive test result varied between studies and were often based on different scales, hindering comparative assessment of performance and control in combined analyses.

Differences in choice of reference standard and associated accuracy may bias between test comparisons (Lijmer et al., 1999). Culture and microscopic examination were the most common reference standards for Se estimates. Both are likely to have high specificity because the endpoints are isolation of the bacteria or confirmation through histopathology (Liebana et al., 2008). Both however are conditional on the performance of post-mortem examination which is likely to negatively bias Se. Meat inspection is likely to have lower Se than laboratory based necroscopy because it relies on existence of macroscopic lesions that can be observed through visual inspection, palpation and incision of organs and may also have less than perfect Sp because of the misclassification of the cause of the lesion (Corner et al., 1990, Liebana et al., 2008). The least common reference standard in eligible papers was the SICCT test (Figure 2). The SICCT test (at standard interpretation) is reported to have very high specificity, higher than the other tuberculin tests, because attempts to exclude immunological responses caused by exposure to environmental mycobacteria by measuring the response to *M. avium* as well as *M. bovis* (de la Rua-Domenech et al., 2006, Goodchild et al., 2015). However, the SICCT test is likely to classify a different spectrum of animals as infected than post-mortem tests because it measures

immunological changes early in pathogenesis whereas most other reference standards detect later stage pathology (Pollock et al., 2005).

Despite the comprehensive search strategy, only 119 references were identified with eligible estimates for one of 14 different diagnostic test types between 1934 and 2009 (see supplementary material). The number of references varied widely, from one each for the Multiplex Immunoassay and Glutaraldehyde tests, 15 for the SICCT test and 27 for the IFN-γ blood test. This is relatively few given that the methodological quality of the studies may vary (see Downs et al., 2017) and other factors may affect the accuracy of reported estimates. Changes in the molecular evolution of *M. bovis* have been recorded and subtle changes in test format over time may have affected performance and encouraged selection for particular strains of *M. bovis* that can evade detection (Smith et al., 2006). The move from using PPD (tuberculin) based on *M. tuberculosis* to PPD produced from *M. bovis* strain AN5 in the 1970s (Lesslie et al., 1975) and changes in the companies that supply tuberculin may have led to differences in test performance (Downs et al., 2013, Tameni et al., 1998).

Since the conduct of this systematic review and early 2016, Se and Sp have been estimated for the SICCT test in three studies: Karolemeas et al., 2012 (Se), EFSA, 2012 (Se and Sp), Goodchild et al., 2015 (Sp). The Se and Sp of the SIT test was assessed by Casal et al., 2014 (Se) and EFSA, 2012 (Se and Sp) (see Nu□ez-Garcia et al., 2017, for further discussion). There have been no new studies measuring the performance of the CF test, although a meta-analysis has been published (Farnham et al., 2012).

In conclusion, in a comprehensive and systematic search of the literature spanning almost 80 years there were comparatively few references providing estimates of the performance of diagnostic tests for bTB in cattle despite the global importance of the disease. The systematic literature review also showed that there is substantial scope for improvement in search methodology and technology for the detection of veterinary studies of the performance of diagnostic tests and that standardised reporting of diagnostic test performance is urgently needed.

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

Authors' contributions

All authors contributed to the design of the systematic review, and the identification of variables for which data should be extracted prior to the conduct of the review. PU with input from JP and SD designed and created the bespoke databases stages 1 and 2 for data entry. Protocol documents were developed by JP, PU and SD with input from the rest of the WG. EP, JP and SD designed the search string for electronic databases and tested it. EP obtained references through APHA library and SD contacted research institutions for unpublished work. SD, JP, TG, JB, and RCH developed the methodology and guidelines for stage 1 review of abstracts. SD and JP reviewed the abstracts at stage 1. DA, JB, AC, RdlR, AG, JG, JNG, S.R, MS, MV, EW, MW AW, JW, SM, SD and RCH conducted stage 2 reviews of references. JP and PU compared agreement between authors at stage 2 and designed and conducted the resolution

procedure to obtain agreed estimates of performance. SD cleaned the covariate data. JNG plotted the unadjusted estimates of Se and Sp by reference and country. SD drafted the first version of the manuscript circulated to the WG. All authors contributed to the discussion of results, read, commented on and approved the final manuscript. SD was project leader.

Acknowledgements

We would like to thank authors of references and research institutes who provided unpublished data or articles in press including Animal Health Board New Zealand, Animal and Plant Health Inspection Service United States Department of Agriculture, Centre for Veterinary Epidemiology and Risk Analysis University College Dublin, Department of Agriculture and Rural Development Northern Ireland, Prionics AG Switzerland, Veterinary Division Scottish Government, Veterinary Research Institute Czech Republic, VISAVET Health Surveillance Centre UCM Spain and APHA Tuberculosis Science. We thank APHA veterinarians and scientists Guanghui Wu, Zhensheng Pan, Anna La Rocca, Monika Klita, Guda van der Burgt, and Sandrine Lesellier who reviewed references in their native language, Robin Sayers for calculating the intra-class correlation coefficients and Mary O'Mara for organising the references. We thank Dirk Pfeiffer for helpful comments on an earlier draft of the paper.

Additional Material

The supplementary material lists all the references with eligible estimates of Se and/or Sp identified in the systematic review and displays the distribution of estimates used in the meta-analysis of test performance (Nu□ez-Garcia et al, 2017).

References

Bero, L.A., Grilli, R., Grimshaw, J.M., Harvey, E., Oxman, A.D., Thomson, M.A., 1998. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. Brit. Med. J. 317, 465-468.

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

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

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Lijmer, J.G., Moher, D., Rennie, D., de Vet, H.C., Kressel,H.Y., Rifai, N., Golub, R.M., Altman, D.G., Hooft, L., Korevaar, D.A., Cohen, J.F., for the STARD Group, 2015. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Brit. Med. J. 351. DOI: 10.1136/bmj.h5527.

Casal, C., Diez-Guerrier, A., Alvarez, J., Rodriguez-Campos, S., Mateos, A., Linscott, R., Martel, E., Lawerence, J.C., Whelan, C., Clarke, J., O'Brien, A., Dominguez, L., Aranaz, A., 2014. Strategic use of serology for the diagnosis of bovine tuberculosis after intradermal skin testing. Vet. Microbiol. 170, 342-351.

Chalmers, T.C., 1991. Problems induced by meta-analyses. Stat. Med. 10, 971-980.

Claxton, P.D., Eamens, G.J., Mylrea, P.J., 1979. Laboratory diagnosis of bovine tuberculosis. Aust. Vet. J. 55, 514-520.

Clegg, T.A., Duignan, A., More S.J., 2015. The relative effectiveness of testers during field surveillance in unrestricted low-risk herds in Ireland. Prev. Vet. Med. 119, 85-89.

Corner, L.A., Melville, L., McCubbin, K., Small, K.J., McCormick, B.S., Wood, P.R., Rothel, J.S., 1990. Effeciency of inspection procedures for the detection of

tuberculosis lesions in cattle. Aust. Vet. J. 67, 389-392.

de la Rua-Domenech, R., 2006. Human *Mycobacterium bovis* infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis. Tuberculosis. 86, 77-109.

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.

Deeks, J.J., Macaskill, P., Irwig, L., 2005. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J. Clin. Epidemiol. 58, 882-893.

Downs, S.H., Clifton-Hadley, R.S., Upton, P.U., Milne, I., Ely, E., Gopal, R., Goodchild, A.V., Sayers, A.R., 2013. Tuberculin manufacturing source and breakdown incidence rate of bovine tuberculosis in British cattle, 2005-2009. Vet. Rec. 26, 98. DOI: 10.1136/vr.100679.

Downs, S.H., More, S.J., Goodchild, A.V., Whelan, A.O., Abernethy, D.A., Broughan, J.M., Cameron, A., Cook, A.J., de la Rua-Domenech, R., Greiner, M., Gunn, J., Nu□ez-Garcia, J., Rhodes, S., Rolfe, S., Sharp, M., Upton, P.A., Vordermeier, H. M., Watson, E., Welsh, M., Woolliams, J.A., Clifton-Hadley, R.S., Parry, J.E., 2017. Evaluation of the methodological quality of studies of the performance of diagnostic tests for bovine tuberculosis in cattle, using QUADAS.Prev. Vet. Med doi:10.1016/j.prevetmed.2017.03.006

EEC (European Economic Community), 1964. Council Directive 64/432/EEC, of 26 June 1964 on animal health problems affecting intra-Community trade in bovine animals and swine. Official Journal 121, 1977-2012 (as amended).

EFSA (European Food Safety Agency), 2008. Scientific Opinion of the Scientific Panel on Animal Health and Welfare on request from the European Commission on "Tuberculosis testing in deer". The EFSA Journal. 645, 1-34. http://www.efsa.europa.eu/sites/default/files/scientific_output/files/main_documents/645.pdf (accessed 4 August 2017).

EFSA (European Food Safety Agency), 2012. EFSA Panel on Animal Health and Welfare (AHAW); Scientific Opinion on the use of a gamma interferon test for the diagnosis of bovine tuberculosis. The EFSA Journal. 10, 2975. http://www.efsa.europa.eu/en/efsajournal/pub/2975.htm (accessed 4 August 2017).

Farnham, M.W., Norby, B., Goldsmith, T.J., Wells, S.J., 2012. Meta-analysis of field studies on bovine tuberculosis skin tests in United States cattle herds. Prev. Vet. Med. 102, 234-242.

FAO (Food and Agriculture Organisation), 2012. Bovine tuberculosis at the animal-human-ecosystem interface. EMPRES Transboundary Animal Diseases Bulletin. 40, 2-11.

http://www.fao.org/3/a-i2811e/i2811e01.pdf (accessed 20 March 2016).

Gaborick, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996. Evaluation of a five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. J. Am. Vet. Med. Assoc. 209, 962-966.

Gardner I.A., Stryhn H., Lind P., Collins M.T., 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Prev Vet Med. 45, 107-122.

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

Holth, H., 1949. Tuberculosis Anatoxin as Diagnostic for Tuberculosis in Cattle. Nordisk Veterinaermed. 1, 601-612.

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

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

Korevaar, D.A., Cohen, J.F., Hooft, L., Bossuyt, P.M., 2015. Literature survey of high-impact journals revealed reporting weaknesses in abstracts of diagnostic accuracy studies. J. Clin. Epidemiol. 68, 708-715.

Landis, J.R., Koch, G.G., 1977. Measurement of observer agreement for categorical data. Biometrics. 33, 159-174.

Leeflang, M.M., Deeks, J.J., Gatsonis, C., Bossuyt, P.M., 2008. Systematic reviews of diagnostic test accuracy. Ann. Intern. Med. 149, 889-897.

Liebana, E., Johnson, L., Gough, J., Durr, P., Jahans, K., Clifton-Hadley, R., Spencer, Y., Hewinson, R.G., Downs, S.H., 2008. Pathology of Naturally Occurring Bovine Tuberculosis in England and Wales. Vet. J. 176, 354-360.

Lesslie, I.W., Hebert, C.N., Barnett, D.N., 1975. Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 South-eastern England. Vet. Rec. 96, 335-338.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 6, e1000097.

Liebana, E., Johnson, L., Gough, J., Durr, P., Jahans, K., Clifton-Hadley, R., Spencer, Y., Hewinson, R.G., Downs, S.H., 2008. Pathology of Naturally Occurring Bovine Tuberculosis in England and Wales. Vet. J. 176, 354-360.

Lijmer, J.G., Mol, B.W., Heisterkamp, S., Bonsel, G.J., Prins, M.H., van der Meulen, J.H., Bossuyt, P.M., 1999. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 282, 1061-1066.

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

Mumtaz, N., Chaudhry, Z.I., Mahmood, N., Shakoori, A.R., 2008. Reliability of PCR for detection of Bovine tuberculosis in Pakistan. Pakistan J. Zool. 40, 347-351.

Nu□ez-Garcia, J., Downs, S.H., Parry, J.E., Abernethy, D.A., Broughan, J.M., Cameron, A.R., Cook, A.J., de la Rua-Domenech, R., Goodchild, A.V., Gunn, J., More, S.J., Rhodes, S., Rolfe, S., Sharp, M., Upton, P.A., Vordermeier, H.M., Watson, E., Welsh, M., Whelan, A.O., Woolliams, J.A., Clifton-Hadley, R.S., Greiner, M., 2017. Meta-analysis of the sensitivity and specificity of a range of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland, Prev. Vet. Med. doi: 10.1016/j.prevetmed 2017.02.017.

OIE, 2012. Terrestrial Manual (Chapter 2.4.7 – Bovine Tuberculosis). World Organisation for Animal Health. Paris. http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.04.06_BOVINE_TB.pdf (accessed 4 August 2017).

O'Reilly, L.M., Daborn, C. J., 1995. The Epidemiology of *Mycobacterium bovis* Infections in Animals and Man - a Review. Tubercle Lung Dis. 76, 1-46.

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. Immun. Immunop.108, 37-43.

Salman, M.D. 2003. Animal disease surveillance and survey systems: Methods and Applications. Iowa State Press.

Sargeant, J.M, Rajic A., Read S., Ohlsson A., 2006. The process of systematic review and its application in agri-food public-health. Prev. Vet. Med. 75, 141-151.

Sargeant JM, O'Connor AM, Gardner IA, Dickson JS, Torrence ME, Dohoo IR, Lefebvre SL, Morley PS, Ramirez A, Snedeker K., 2010. The REFLECT statement: reporting guidelines for randomized controlled trials in livestock and food safety: explanation and elaboration. J Food Prot. 73, 579-603.

Saveleva, E., Selinski, S., 2008. Meta-analyses with binary outcomes: how many studies need to be omitted to detect a publication bias? J. Toxicol. Env. Heal. A. 71, 845-850.

Smith, N.H., Gordon, S.V., de la Rua-Domenech, R., Clifton-Hadley, R.S., Hewinson,

R.G., 2006. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis.* Nat. Rev. Microbiol. 4, 670-681.

Steingart, K.R., Henry, M., Laal, S., Hopewell, P.C., Ramsay, A., Menzies, D., Cunningham, J., Weldingh, K., Pai, M., 2007. A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. Thorax. 62, 911-918.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B., 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 283, 2008-2012.

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

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

Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gotzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J., Egger, M., 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 4, e297.

VLA (Veterinary Laboratories Agency) 2010. Meta-analysis of diagnostic tests and modeling to identify appropriate testing strategies to reduce *M. bovis* infection GB herds. SE3238 Final Report and appendices: Appendix 1 Specification meta-analysis database V3.0, Appendix 2 Stage 1 reference eligibility, Appendix 3 stage 2 data-entry, Appendix 4 VETQUADAS, Appendix 5 Stage 2 reference eligibility, Appendix 6 electronic search string, Annex 1 Ineligible Tests, Annex 2 Eligible Tests 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#Description (accessed 4 August 2017).

Vordermeier, M., Ewer, K., 2006. Specificity trial of the Bovigam IFN-Gamma test in GB cattle. Veterinary Laboratories Agency.

Westwood, M.E., Whiting, P.F., Kleijnen, J., 2005. How does study quality affect the results of a diagnostic meta-analysis? BMC Medical Research Methodology 5, 1-16.

Whiting, P., Rutjes A.W., Dinnes J., Reitsma J., Bossuyt P.M., Kleijnen J., 2003. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med. Res. Methodol. 3, 1-13.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., McCormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991. Field comparison of the interferon-gamma assay and the intradermal tuberculin test for the diagnosis of bovine tuberculosis. Aust. Vet. J. 68, 286-290.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., Mccormick, B.S., Francis, B., Melville, L., Small, K., Dewitte, K., Tolson, J., Ryan, T.J., Delisle, G.W., Cox, J.C., Jones, S.L., 1992. A Field-Evaluation of Serological and Cellular Diagnostic-Tests for Bovine Tuberculosis. Vet. Microbiol. 31, 71-79.

Online supplement

References identified in the systematic review with eligible estimates for sensitivity and/or specificity of diagnostic tests for bovine tuberculosis

Single Intradermal skin test (SIT)

Davidson, R.M., 1965. A Comparison of the Efficacy of Canadian and Australian Tuberculins. Part I. New Zeal. Vet. J. 13, 109-115.

Kubin, G., 1951. Experiments in Cattle With Graded Dilutions of Tuberculin for the Determination of the Actual Reaction. Wien. Tierarztl. Monat. 38, 277-310.

Lilenbaum, W., Ferreira, R., Marassi, C.D., Ristow, P., Oelemann, W.M.R., Fonseca, L.D., 2007. Interference of Tuberculosis on the Performance of Elisas Used in the Diagnosis of Paratuberculosis in Cattle. Braz. J. Microbiol. 38, 472-477.

Llamazares, O.R.G., Martin, C.B.G., Nistal, D.A., Redondo, V., Rodriguez, L.D., Ferri, E.F.R., 1999. Field Evaluation of the Single Intradermal Cervical Tuberculin Test and the Interferon-Gamma Assay for Detection and Eradication of Bovine Tuberculosis in Spain. Vet. Microbiol. 70, 55-66.

O'Reilly, L.M., MacClancy, B.N., 1975. A Comparison of the Accuracy of a Human and a Bovine Tuberculin PPD for Testing Cattle With a Comparative-Cervical Test. Irish Vet. J. 29, 63-70.

Pollock, J.M., McNair, J., Bassett, H., Cassidy, J.P., Costello, E., Aggerbeck, H., Rosenkrands, I., Andersen, P., 2003. Specific delayed-type hypersensitivity responses to ESAT-6 identify-tuberculosis-infected cattle. J. Clin. Microbiol. 41, 1856-1860.

Ragassa, A., Ameni, G., 2001. Sensitivity and Specificity of a Single Intradermal Tuberculin Test at the Cervical and Caudal Folds in Zabu Cattle in Ethiopia. Indian J. Anim. Sci. 71, 325-327.

Roswurm, J.D., De Kantor, I.N., Marchevsky, N., Spinelli, R., Spath, Y.E., 1979b. Tuberculin Test Sensitivity in Cattle Naturally Infected With *Mycobacterium-bovis* in Argentina 1. Simple Tuberculin Tests. B. Ofic. Sanit. Panam. 86, 420-427.

Waveren, G.M., 1953. Tuberculosis Testing in Cows With PPD. Tuberculin. Tijdschr. Diergeneesk. 78, 735-760.

Zeller, H., 1936. Diagnosis of Bovine Tuberculosis. Berl. Tierarztl. 52, 789-791.

SICCT test

Ameni, G., Miorner, H., Roger, F., Tibbo, M., 2000. Comparison between comparative tuberculin and gamma-interferon tests for the diagnosis of bovine tuberculosis in Ethiopia. Trop. Anim. Health Pro. 32, 267-276.

Aranaz, A., De Juan, L., Bezos, J., Alvarez, J., Romero, B., Lozano, F., Paramio, J.L., Lopez-Sanchez, J., Mateos, A., Dominguez, L., 2006. Assessment of Diagnostic Tools for Eradication of Bovine Tuberculosis in Cattle Co-Infected With *Mycobacterium bovis* and *M-Avium subsp. paratuberculosis*. Vet. Res. 37, 593-606.

Berggren, S.A., 1978. Evaluation of Tuberculin Test in Malawi. Nord. Vet. Med. 30, 364-370.

Costello, E., Egan, J.W.A., Quigley, F.C., O'Reilly, P.F., 1997a. Performance of the single intradermal comparative tuberculin test in identifying cattle with tuberculous lesions in Irish herds. Vet. Rec. 141, 222-224.

DARD, 1996. Possible use of the gamma interferon blood test in the control of bovine tuberculosis in Northern Ireland. Department of Agriculture and Rural Development. Unpublished.

DARD, 2006. A report on the DARD IFN-g initiatives up to May 2006 from the IFN-g working group. Department of Agriculture and Rural Development. Unpublished.

Lauzi, S., Pasotto, D., Amadori, M., Archetti, I.L., Poli, G., Bonizzi, L., 2000. Evaluation of the Specificity of the Gamma-Interferon Test in Italian Bovine Tuberculosis-Free Herds. Vet. J. 160, 17-24.

Lesslie, I.W., Hebert, C.N., Barnett, D.N., 1975a. Comparison of the specificity of human and bovine tuberculin PPD for testing cattle. 2 South-eastern England. Vet. Rec. 96, 335-338.

Lesslie, W., Hebert, C.N., 1975b. Comparison of the Specificity of Human and Bovine Tuberculin PPD for Testing Cattle. 3. National Trial in Great Britain. Vet. Rec. 96, 338-341.

O'Reilly, L.M., 1986. Field trials to determine a suitable injection dose of bovine PPD tuberculin for the diagnosis of bovine tuberculosis in naturally infected cattle. Dev. Biol. Stand. 58.

Quirin, R., Rasolofo, V., Andriambololona, R., Ramboasolo, A., Rasolonavalona, T., Raharisolo, C., Rakotoaritahina, H., Chanteau, S., Boisier, P., 2001. Validity of Intradermal Tuberculin Testing for the Screening of Bovine Tuberculosis in Madagascar. Onderstepoort J. Vet. 68, 231-238.

Roswurm, J.D., de Kantor, I.N., Marchevsky, N., Spinelli, R., Spath, Y.E., 1979a. Tuberculin Test Sensitivity in Cattle Naturally Infected With *Mycobacterium-bovis* in Argentina 2. Comparative Tuberculin Tests. B. Ofic. Sanit. Panam. 86, 428-434.
Rushford, B.H., 1964. Investigations Into the Problem of Non-Specific Reactors to the Single Caudal Fold Tuberculin Test in Victorian Dairy Cattle. Part I. Aust. Vet. J. 40, 406-411.

Saegerman, C., Delville, J., Waele, L., Gilson, D., 1995. Serological and cutaneous testing of bovine tuberculosis with the A60 antigen complex from *Mycobacterium bovis*, strain Calmette-Guérin. Prev. Vet. Med. 23, 239-248.

Silva, E., 2001. Evaluation of an enzyme-linked immunosorbent assay in the diagnosis of bovine tuberculosis. Vet. Microbiol. 78, 111-117.

Silva, E., Gonzalez, N., Regalado, L., Alfonso, M.L., Barcenas, Y.J., Menendez, J., 1997. Especificidad del elisa para el diagnostico serologico de la tuberculosis bovina en areas libres. Rev. Salud Anim. 19, 9-13.

Thom, M.L., Hope, J.C., Mcaulay, M., Villarreal-Ramos, B., Coffey, T.J., Stephens, S., Vordermeier, H.M., Howard, C.J., 2006. The Effect of Tuberculin Testing on the Development of Cell-Mediated Immune Responses During *Mycobacterium bovis* Infection. Vet. Immunol. Immunop. 114, 25-36.

Waveren, G.M., 1953. Tuberculosis Testing in Cows With PPD. Tuberculin. Tijdschr. Diergeneesk. 78, 735-760.

WHS, 1996. TB Breakdown - Second Interim Report. Report of whole herd slaughter in Scotland. Unpublished.

Zorawski, C., Karpinski, T., Wisniewski, Z., 1981. Preparation of PPD Tuberculin From *Mycobacterium bovis* Strain An5. Med. Weter. 37, 215-218.

Caudal fold test

Davidson, R.M., 1965. A Comparison of the Efficacy of Canadian and Australian Tuberculins. Part I. New Zeal. Vet. J. 13, 109-115.

Dorset, M., 1934. A Comparison of Koch's Old Tuberculin With a New Synthetic-Medium Tuberculin. J. Am. Vet. Med. Assoc. 84, 439-456.

Francis, J., Seiler, R.J., Wilkie, I.W., O'Boyle, D., Lumsden, M.J., Frost, A.J., 1978. The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. Vet. Rec. 103, 420-435.

Gaborick, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996. Evaluation of a five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. J. Am. Vet. Med. Assoc. 209, 962.

Lepper, A.W.D., Carpenter, M.T., Williams, O.J., Scanlan, W.A., Mcewan, D.R., Andrews, L.G., Thomas, T.R., Corner, L.A., 1979. Comparison of the Efficiency of 2

Doses of Bovine PPD Tuberculin in Single Caudal Fold Tests on Australian Cattle. Aust. Vet. J. 55, 251-256.

Lepper, A.W.D., Mckay, D.J., Lord, V., 1979b. Comparison of the efficiency of 4 doses of bovine PPD tuberculin in cattle and guinea pigs: Recommendations for the most suitable dose for use in the single caudal fold test in cattle. Aust. Vet. J. 55, 507-513.

Lepper, A.W.D., Newtontabrett, D.A., Corner, L.A., Carpenter, M.T., Scanlan, W.A., Williams, O.J., Helwig, D.M., 1977. Use of Bovine PPD Tuberculin in Single Caudal Fold Test to Detect Tuberculosis in Beef-Cattle. Aust. Vet. J. 53, 208-213.

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

Plackett, P., Ripper, J., Corner, L.A., Small, K., Witte, K., Melville, L., Hides, S., Wood, P.R., 1989. An ELISA for the detection of anergic tuberculous cattle. Aust. Vet. J. 66, 15-19.

Ragassa, A., Ameni, G., 2001. Sensitivity and Specificity of a Single Intradermal Tuberculin Test at the Cervical and Caudal Folds in Zabu Cattle in Ethiopia. Indian J. Anim. Sci. 71, 325-327.

Ramirez Casillas, C., Valero Elizondo, G., Arriaga Diaz, C., 1995. A comparison between the ELISA test and tuberculin testing in the diagnosis of bovine tuberculosis. Tec. Pec. Mex. 33, 148-158.

Ryan, T.J., de Lisle, G.W., Wood, P.R., 1991. The performance of the skin and gamma interferon tests for the diagnosis of tuberculosis infection in cattle in New Zealand. Massey University Continuing Education, 143-150.

Small, K.J., Thomson, D., 1986. The Efficiency of Bovine PPD Tuberculin in the Single Caudal Fold Test to Detect Tuberculosis in Water Buffalo. Buffalo Bull. 5, 62-64.

Surujballi, O.P., Romanowska, A., Sugden, E.A., Turcotte, C., Jolley, M.E., 2002. A Fluorescence Polarization Assay for the Detection of Antibodies to *Mycobacterium bovis* in Cattle Sera. Vet. Microbiol. 87, 149-157.

Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Nabors, R.S., Payeur, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995. Comparison of the Sensitivity of the Caudal Fold Skin-Test and a Commercial Gamma-Interferon Assay for Diagnosis of Bovine Tuberculosis. Am. J. Vet. Res. 56, 416-419.

Whipple, D.L., Bolin, C.A., Miller, J.M., 1996. Distribution of lesions in cattle infected with *Mycobacterium bovis*. J. Vet. Diagn. Invest. 8, 351-354.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., Mccormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991. Field Comparison of the Interferon-Gamma Assay and the Intradermal Tuberculin Test for the Diagnosis of Bovine Tuberculosis. Aust. Vet. J. 68, 286-290.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., Mccormick, B.S., Francis, B., Melville, L., Small, K., Dewitte, K., Tolson, J., Ryan, T.J., Delisle, G.W., Cox, J.C., Jones, S.L., 1992. A Field-Evaluation of Serological and Cellular Diagnostic-Tests for Bovine Tuberculosis. Vet. Microbiol. 31, 71-79.

IFN-γ blood test

Aagaard, C., Govaerts, M., Meikle, V., Vallecillo, A.J., Gutierrez-Pabello, J.A., Suarez-Guemes, F., Mcnair, J., Cataldi, A., Espitia, C., Andersen, P., Pollock, J.M., 2006. Optimizing Antigen Cocktails for Detection of *Mycobacterium bovis* in Herds With Different Prevalences of Bovine Tuberculosis: ESAT6-CFP10 Mixture Shows Optimal Sensitivity and Specificity. J. Clin. Microbiol. 44, 4326-4335.

Alvarez, J., de Juan, L., Bezos, J., Romero, B., Saez, J.L., 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. Vet. Microbiol. 135, 389-393.

Amadori, M., Lyashchenko, K.P., Gennaro, M.L., Pollock, J.M., Zerbini, I., 2002. Use of Recombinant Proteins in Antibody Tests for Bovine Tuberculosis. Vet. Microbiol. 85, 379-389.

Ameni, G., Miorner, H., Roger, F., Tibbo, M., 2000. Comparison between comparative tuberculin and gamma-interferon tests for the diagnosis of bovine tuberculosis in Ethiopia. Tropical Animal Health and Production 32, 267-276.

Aranaz, A., De Juan, L., Bezos, J., Alvarez, J., Romero, B., Lozano, F., Paramio, J.L., Lopez-Sanchez, J., Mateos, A., Dominguez, L., 2006. Assessment of Diagnostic Tools for Eradication of Bovine Tuberculosis in Cattle Co-Infected With *Mycobacterium bovis and M-Avium subsp. paratuberculosis.* Veterinary Research 37, 593-606.

Archetti, I.L., Amadori, M., Bardotti, M., Pasotto, D., Lauzi, S., Bonizzi, L., Zanella, A., Poli, G., 1999. The γ-IFN test for diagnosis of bovine tuberculosis: evaluation of specificity under field conditions. SISVet Annual Meeting selected abstracts. I, 170.

Archetti, I.L., Amadori, M., Scaccaglia, P., Vezzoli, F., Luini, M., Fabbi, M., Belloli, A., Sala, L., 1996. Use of the Gamma-Interferon Test in Diagnosing Bovine Tuberculosis in the Lombardia Region. Selezione Vet. 4, 329-338.

Buddle, B.M., de Lisle, G.W., Ryan, T.J., 2005. Optimisation and definition of the criteria of using a *Mycobacterium bovis*-specific ancillary serial test for the diagnosis of bovine tuberculosis in cattle. Wallaceville Animal Research Centre & New

Zealand Food Safety Authority. 1-17. Final Report to Animal Health Board. Unpublished.

Buddle, B.M., Mccarthy, A.R., Ryan, T.J., Pollock, J.M., Vordermeier, H.M., Hewinson, R.G., Andersen, P., De Lisle, G.W., 2003. Use of Mycobacterial Peptides and Recombinant Proteins for the Diagnosis of Bovine Tuberculosis in Skin Test-Positive Cattle. Vet. Rec. 153, 615-620.

Buddle, B.M., Ryan, T.J., Pollock, J.M., Andersen, P., De Lisle, G.W., 2001. Use of Esat-6 in the Interferon-Gamma Test for Diagnosis of Bovine Tuberculosis Following Skin Testing. Vet. Microbiol. 80, 37-46.

Cagiola, A., Feliziani, F., Severi, G., Pasquali, P., Rutili, D., 2004. Analysis of Possible Factors Affecting the Specificity of the Gamma Interferon Test in Tuberculosis-Free Cattle Herds. Clin. Diag. Lab. Immun. 11, 952-956.

Cockle, P.J., Gordon, S.V., Hewinson, R.G., Vordermeier, H.A., 2006. Field Evaluation of a Novel Differential Diagnostic Reagent for Detection of *Mycobacterium bovis* in Cattle. Clin. Vaccine Immunol. 13, 1119-1124.

DARD, 1996. Possible use of the gamma interferon blood test in the control of bovine tuberculosis in Northern Ireland. Department of Agriculture and Rural Development. Unpublished.

DARD, 2006. A report on the DARD IFN-g initiatives up to May 2006 from the IFN-γ working group. Department of Agriculture and Rural Development. Unpublished.

de Lisle, G.R., T., 2000. Specificity of the gamma interferon test in calves. Wallaceville Animal Research Centre & Ruakura Research Centre. Unpublished.

Gormley, E., Doyle, M.B., Mcgill, K., Costello, E., Good, M., Collins, J.D., 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. Vet. Immunol. Immunop. 102, 413-420.

Lauzi, S., Pasotto, D., Amadori, M., Archetti, I.L., Poli, G., Bonizzi, L., 2000. Evaluation of the Specificity of the Gamma-Interferon Test in Italian Bovine Tuberculosis-Free Herds. Vet. J. 160, 17-24.

Llamazares, O.R.G., Martin, C.B.G., Nistal, D.A., Redondo, V., Rodriguez, L.D., Ferri, E.F.R., 1999. Field Evaluation of the Single Intradermal Cervical Tuberculin Test and the Interferon-Gamma Assay for Detection and Eradication of Bovine Tuberculosis in Spain. Vet. Microbiol. 70, 55-66.

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

Monaghan, M., Quinn, P.J., Kelly, A.P., Mcgill, K., Mcmurray, C., Ocrowley, K., Bassett, H.F., Costello, E., Quigley, F., Rothel, J.S., Wood, P.R., Collins, J.D., 1997. A Pilot Trial to Evaluate the Gamma-Interferon Assay for the Detection of *Mycobacterium bovis* Infected Cattle Under Irish Conditions. Irish Vet. J. 50, 229-232.

O'Nuallain, E.M.O., Davis, W.C., Costello, E., Pollock, J.M., Monaghan, M.L., 1997. Detection of *Mycobacterium bovis* Infection in Cattle Using an Immunoassay for Bovine Soluble Interleukin-2 Receptor-Alpha (Sil-2r-Alpha) Produced by Peripheral Blood T-Lymphocytes Following Incubation With Tuberculin PPD. Vet. Immunol. Immunop. 56, 65-76.

Pollock, J.M., Girvin, R.M., Lightbody, K.A., Clements, R.A., Neill, S.D., Buddle, B.M., Andersen, P., 2000. Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. Vet. Rec. 146, 659-665.

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.

Ritelli, M., Amadori, M., Dondo, A., Begni, B., Zoppi, S., Archetti, I.L., 2004. Combined Use of the G-Interferon and Interleukin-2 Receptor Assays for Diagnosis of Bovine Tuberculosis. Online J. Vet. Res. 8, 16-21.

Ryan, T.J., Buddle, B.M., de Lisle, G.W., 2000. An evaluation of the gamma interferon test for detecting bovine tuberculosis in cattle 8 to 28 days after tuberculin skin testing. Res. Vet. Sci. 69, 57-61.

Ryan, T.J., de Lisle, G.W., Wood, P.R., 1991. The performance of the skin and gamma interferon tests for the diagnosis of tuberculosis infection in cattle in New Zealand. Massey University Continuing Education, 143-150. Unpublished.

Scacchia, M., Lelli, R., Petrini, A., Prencipe, V., Calistri, P., Giovannini, A., 2000. Use of Innovative Methods in the Eradication of Bovine Tuberculosis. J. Vet. Med. B. 47, 321-327.

Sidders, B., Pirson, C., Hogarth, P.J., Hewinson, R.G., Stoker, N.G., Vordermeier, H.M., Ewer, K., 2008. Screening of Highly Expressed Mycobacterial Genes Identifies Rv3615c as a Useful Differential Diagnostic Antigen for the *Mycobacterium Tuberculosis* Complex. Infect. Immun. 76, 3932-3939.

Vordermeier, H.M., Whelan, A., Cockle, P.J., Farrant, L., Palmer, N., Hewinson, R.G., 2001. Use of synthetic peptides derived from the antigens ESAT-6 and CFP-10 for differential diagnosis of bovine tuberculosis in cattle. Clinical Diagnostic Laboratory Immunology 8, 571-578.

Vordermeier, M., 2009. Improved diagnostics for cattle. SE3005 Final Report. Veterinary Laboratories Agency.

Vordermeier, M., Ewer, K., 2006. Specificity Trial of the BOVIGAM â IFN-Gamma Test in GB Cattle. Veterinary Laboratories Agency, 1-29. Unpublished.

Whipple, D.L., Bolin, C.A., Davis, A.J., Jarnagin, J.L., Johnson, D.C., Nabors, R.S., Payeur, J.B., Saari, D.A., Wilson, A.J., Wolf, M.M., 1995. Comparison of the Sensitivity of the Caudal Fold Skin-Test and a Commercial Gamma-Interferon Assay for Diagnosis of Bovine Tuberculosis. Am. J. Vet. Res. 56, 416-419.

Whipple, D.L., Bolin, C.A., Miller, J.M., 1996. Distribution of lesions in cattle infected with *Mycobacterium bovis*. J. Vet. Diagn. Invest. 8, 351-354.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., Mccormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991. Field Comparison of the Interferon-Gamma Assay and the Intradermal Tuberculin Test for the Diagnosis of Bovine Tuberculosis. Aust. Vet. J. 68, 286-290.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., Mccormick, B.S., Francis, B., Melville, L., Small, K., Dewitte, K., Tolson, J., Ryan, T.J., Delisle, G.W., Cox, J.C., Jones, S.L., 1992. A Field-Evaluation of Serological and Cellular Diagnostic-Tests for Bovine Tuberculosis. Vet. Microbiol. 31, 71-79.

ELISA

Auer, L.A., 1987. Assessment of an enzyme linked immunosorbent assay for the detection of cattle infected with *Mycobacterium bovis*. Aust. Vet. J. 64, 172-176.

Cavirani, S., Flammini, C.F., Allegri, G., Ansuini, A., Arrigoni, N., Cavalca, M., Magnani, R., 1993. Serological Diagnosis of Tuberculosis in Cattle. Elisa Reactions in Animals Naturally Infected With *M. bovis* and *M. Paratuberculosis*. Annali Della Facolta Di Medicina Veterinaria, Universita Di Parma 13, 211-222.

Costello, E., O'Reilly, P.F., Yearsley, D.K., OGrady, D.P., OReilly, L.M., Collins, J.D., Monaghan, M.L., Bassett, H.F., 1997b. A study of an enzyme-linked immunosorbent assay for the diagnosis of tuberculosis in cattle. Irish Vet. J. 50, 35-38.

Eid, G.E., Mousa, I.M.I., Selim, S.A.K., 2001. Comparison Between Tuberculin Test and Enzyme Linked Immunosorbant Assay for Diagnosis of Tuberculosis in Cattle and Buffaloes. Vet. Med. J. Giza. 49, 355-369.

Gaborick, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996. Evaluation of a five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. J. Am. Vet. Med. Assoc. 209, 962-966.

Hammam, H., Refai, M., Bisping, W., Kirpal, G., 1989. Studies on the Efficiency of Absorbed Bovine PPD in Tuberculin and Serological Tests. Zbl. Vet. Med. B. 36, 175-179.

Lilenbaum, W., Ribeiro, E.R., Souza, G.N., Moreira, E.C., Fonseca, L.S., Ferreira, M.A.S., Schettini, J., 1999. Evaluation of an ELISA-PPD for the diagnosis of bovine tuberculosis in field trials in Brazil. Res. Vet. Sci. 66, 191-195.

McNair, J., Corbett, D.M., Girvin, R.M., Mackie, D.P., Pollock, J.M., 2001. Characterization of the early antibody response in bovine tuberculosis: MPB83 is an early target with diagnostic potential. Scand. J. Immunol. 53, 365-371.

O'Nuallain, E.M.O., Davis, W.C., Costello, E., Pollock, J.M., Monaghan, M.L., 1997. Detection of *Mycobacterium bovis* Infection in Cattle Using an Immunoassay for Bovine Soluble Interleukin-2 Receptor-Alpha (Sil-2r-Alpha) Produced by Peripheral Blood T-Lymphocytes Following Incubation With Tuberculin PPD. Vet. Immunol. Immunop. 56, 65-76.

Plackett, P., Ripper, J., Corner, L.A., Small, K., Witte, K., Melville, L., Hides, S., Wood, P.R., 1989. An ELISA for the detection of anergic tuberculous cattle. Aust. Vet. J. 66, 15-19.

Ramirez Casillas, C., Valero Elizondo, G., Arriaga Diaz, C., 1995. A comparison between the ELISA test and tuberculin testing in the diagnosis of bovine tuberculosis. Tec. Pecu. Mex. 33, 148-158.

Ritacco, V., de Kantor, I.N., Barrera, L., Nader, A., Bernardelli, A., Torrea, G., Errico, F., Fliess, E., 1987. Assessment of the sensitivity and specificity of enzyme-linked immunosorbent assay (ELISA) for the detection of mycobacterial antibodies in bovine tuberculosis. J. Vet. Med. B. 34, 119-125.

Ritacco, V., Lopez, B., Barrera, L., Nader, A., Fliess, E., Dekantor, I.N., 1990. Further Evaluation of an Indirect Enzyme-Linked-Immunosorbent-Assay for the Diagnosis of Bovine Tuberculosis. J. Vet. Med. B. 37, 19-27.

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.

Saegerman, C., Delville, J., Waele, L., Gilson, D., 1995. Serological and cutaneous testing of bovine tuberculosis with the A60 antigen complex from *Mycobacterium bovis*, strain Calmette-Guérin. Prev. Vet. Med. 23, 239-248.

Silva, E., 2001. Evaluation of an enzyme-linked immunosorbent assay in the diagnosis of bovine tuberculosis. Vet. Microbiol. 78, 111-117.

Silva, E., Gonzalez, N., Regalado, L., Alfonso, M.L., Barcenas, Y.J., Menendez, J., 1997. Especificidad del elisa para el diagnostico serologico de la tuberculosis bovina en areas libres. Rev. Salud. Anim. 19, 9-13.

Silva, E., Merlini, J.C., de Kantor, I.N., Bernardelli, A., Robinson, A., Nibot, C., Gonzalez, N., Regalado, L., Alfonso, M.J., Martinez, N., 1999. Evaluation of the ELISA technique (DAVIH-TBB system) for the diagnosis of bovine tuberculosis. Rev. Med. Vet. Argent. 80, 361-363.

Sugden, E.A., Stilwell, K., Rohonczy, E.B., Martineau, P., 1997. Competitive and indirect enzyme-linked immunsorbent assays for *Mycobacterium bovis* infections based on MPB70 and lipoarabinomannan antigens. Can. J. Vet. Res. 61, 8-14.

Thoen, C.O., Malstrom, C., Himes, E.M., Mills, K., 1981. Use of enzyme-linked immunosorbent assay for detecting mycobacterial antigens in tissues of *Mycobacterium bovis*-infected cattle. Am. J. Vet. Res. 42, 1814-1815.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., Mccormick, B.S., Francis, B., Melville, L., Small, K., Dewitte, K., Tolson, J., Ryan, T.J., Delisle, G.W., Cox, J.C., Jones, S.L., 1992. A Field-Evaluation of Serological and Cellular Diagnostic-Tests for Bovine Tuberculosis. Vet. Microbiol. 31, 71-79.

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

Latex bead agglutination assay

Koo, H.C., Park, Y.H., Ahn, J., Waters, W.R., Hamilton, M.J., Barrington, G., Mosaad, A.A., Palmer, M.V., Shin, S., Davis, W.C., 2004. New Latex Bead Agglutination Assay for Differential Diagnosis of Cattle Infected With *Mycobacterium bovis* and *Myobacterium Avium subsp paratuberculosis*. Clin. Diagn. Lab. Immun. 11, 1070-1074.

Koo, H.C., Park, Y.H., Ahn, J., Waters, W.R., Palmer, M.V., Hamilton, M.J., Barrington, G., Mosaad, A.A., Park, K.T., Jung, W.K., Hwang, I.Y., Cho, S.N., Shin, S.J., Davis, W.C., 2005. Use of Rmpb70 Protein and Esat-6 Peptide as Antigens for Comparison of the Enzyme-Linked Immunosorbent, Immunochromatographic, and Latex Bead Agglutination Assays for Serodiagnosis of Bovine Tuberculosis. J. Clin. Microbiol. 43, 4498-4506.

Multiplex immunoassay

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.

Glutaldehyde

de Kantor, I.N., Lopez, B., Torres, P., Nader, A., Garcia, V., 1993. Preliminary Evaluation of a Simple Method for Detection of Bovine Tuberculosis: the Glutaraldehyde Test. J. Vet. Med. B. 40, 27-30.

Fluorescence polarization assay

Jolley, M.E., Nasir, M.S., Surujballi, O.P., Romanowska, A., Renteria, T.B., De La Mora, A., Lim, A., Bolin, S.R., Michel, A.L., Kostovic, M., Corrigan, E.C., 2007. Fluorescence Polarization Assay for the Detection of Antibodies to *Mycobacterium bovis* in Bovine Sera. Vet. Microbiol. 120, 113-121.

Surujballi, O.P., Romanowska, A., Sugden, E.A., Turcotte, C., Jolley, M.E., 2002. A Fluorescence Polarization Assay for the Detection of Antibodies to *Mycobacterium bovis* in Cattle Sera. Vet. Microbiol. 87, 149-157.

Serological Lateral Flow

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.

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.

Necroscopy

Asseged, B., Woldesenbet, Z., Yimer, E., Lemma, E., 2004. Evaluation of abattoir inspection for the diagnosis of Mycobacterium bovis infection in cattle at Addis Ababa abattoir. Trop. Anim. Health Pro. 36, 537-546.

de Arrunda Pinto, P.S., Vragas Viloria, M.I., de Faria, E.J.E., de Almeida, L.P., 2004. Performance Evaluation of the Anatomopathological and Histopathological Analysis in the Post Mortem Inspection of Bovine Suspects or Reagents to the Tuberculin Test. Rev. Bras. Cienc. Vet. 11, 27-31.

de Kantor, I.N., Nader, A., Bernardelli, A., Giron, D.O., Man, E., 1987. Tuberculous Infection in Cattle Not Detected by Slaughterhouse Inspection. J. Vet. Med. B. 34, 202-205.

Liebana, E., Johnson, L., Gough, J., Durr, P., Jahans, K., Clifton-Hadley, R., Spencer, Y., Hewinson, R.G., Downs, S.H., 2008. Pathology of Naturally Occurring Bovine Tuberculosis in England and Wales. Vet. J. 176, 354-360.

O'Reilly, L.M., Costello, E., 1988. Bovine Tuberculosis. Irish Vet. News. 10, 17-25.

Selim, S.A., Amin, M.M., El-Sherif, M.T., El-Saifi, A.A., 1978. Indirect Hem Agglutination Test as a Means of Diagnosis of Tuberculosis in Egyptian Buffaloes. B. Anim. Hlth. Prod. Afr. 26, 315-320.

Teklu, A., Asseged, B., Yimer, E., Gebeyehu, M., Woldesenbet, Z., 2004. Tuberculous lesions not detected by routine abattoir inspection: the experience of the Hossana municipal abattoir, southern Ethiopia. Rev. Sci. Tech. OIE. 23, 957-964.

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

Microscopic examination

Busch, G., 1940. Fluorescence-Microscopic Demonstration of Tubercle Bacilli in Cattle. Dtsch. Tietarztl. Wschr. 48, 587-589.

Claxton, P.D., Eamens, G.J., Mylrea, P.J., 1979. Laboratory diagnosis of bovine tuberculosis. Aust. Vet. J. 55, 514-520.

Cousins, D.V., Francis, B.R., Gow, B.L., 1989. Advantages of a New Agar Medium in the Primary Isolation of *Mycobacterium-bovis*. Vet. Microbiol. 20, 89-95.

de Arrunda Pinto, P.S., Vragas Viloria, M.I., de Faria, E.J.E., de Almeida, L.P., 2004. Performance Evaluation of the Anatomopathological and Histopathological Analysis in the Post Mortem Inspection of Bovine Suspects or Reagents to the Tuberculin Test. Rev. Bras. Cienc. Vet. 11, 27-31.

de Kantor, I.N., Jimenez, J.L., 1977. Mycobacteria Isolated From Lymph Nodes of Cattle in Argentina. Rev. Med. Vet. Argent. 58, 381-384.

Gutierrez Cancela, M.M., Garcia Marian, J.F., 1993. Comparison of Ziehl-Nielsen Staining and Immunohistochemistry for the Detection of *Mycobacterium bovis* in Bovine and Caprine Tuberculous Lesions. J. Comp. Pathol. 109, 361-370.

Latini, O., Canal, A.M., Ferrara, M.E., Sequeira, M.D., Sequeira, G., Bagnaroli, R., Torres, P., 1997. Reliability of the Prevalence of *Mycobacterium Bovis* in Cattle by Rejection in Abattoirs. Arch. Med. Vet. 29, 197-204.

Liebana, E., Johnson, L., Gough, J., Durr, P., Jahans, K., Clifton-Hadley, R., Spencer, Y., Hewinson, R.G., Downs, S.H., 2008. Pathology of Naturally Occurring Bovine Tuberculosis in England and Wales. Vet. J. 176, 354-360.

Neill, S.D., Hanna, J., O'Brien, J.J., 1986. Radiometric detection of mycobacteria in abattoir specimens taken from tuberculin reactive cattle. Vet. Rec. 118, 304-306.

Taylor, M.J., Hughes, M.S., Skuce, R.A., Neill, S.D., 2001. Detection of Mycobacterium bovis in bovine tuberculous specimens utilising real-time fluorescence and FRET probe rapid-cycle PCR. Irish Vet. J. 54, 105-106.

Teklu, A., Asseged, B., Yimer, E., Gebeyehu, M., Woldesenbet, Z., 2004. Tuberculous lesions not detected by routine abattoir inspection: the experience of the Hossana municipal abattoir, southern Ethiopia. Rev. Sci. Tech. OIE. 23, 957-964.

Varello, K., Pezzolato, M., Mascarino, D., Ingravalle, F., Caramelli, M., Bozzetta, E., 2008. Comparison of Histologic Techniques for the Diagnosis of Bovine Tuberculosis in the Framework of Eradication Programs. J. Vet. Diagn. Invest. 20, 164-169.

Watrelot-Virieux, D., Drevon-Gaillot, E., Toussaint, Y., Belli, P., 2006. Comparison of Three Diagnostic Detection Methods for Tuberculosis in French Cattle. J. Vet. Med. B. 53, 321-325.

Culture

Claxton, P.D., Eamens, G.J., Mylrea, P.J., 1979. Laboratory diagnosis of bovine tuberculosis. Aust. Vet. J. 55, 514-520.

Cousins, D.V., Francis, B.R., Gow, B.L., 1989. Advantages of a New Agar Medium in the Primary Isolation of *Mycobacterium-bovis*. Vet. Microbiol. 20, 89-95.

de Kantor, I.N., Jimenez, J.L., 1977. Mycobacteria Isolated From Lymph Nodes of Cattle in Argentina. Rev. Med. Vet. Argent. 58, 381-384.

Neill, S.D., Hanna, J., O'Brien, J.J., 1986. Radiometric detection of mycobacteria in abattoir specimens taken from tuberculin reactive cattle. Vet. Rec. 118, 304-306.

Perez, A., Reniero, A., Forteis, A., Meregalli, S., Lopez, B., Ritacco, V., 2002. Detection of *Mycobacterium bovis* in Milk by Bacteriological Methods and Polymerase Chain Reaction. Rev. Argent. Microbiol. 34, 45-51.

Yearsley, D., O'Rourke, J., Griffin, M., Nearney, M., Egan, J., 2000. Detection of *Mycobacterium tuberculosis* complex organisms in tissue samples from bovine lymph nodes: comparison of two rapid methods and Lowenstein-Jensen medium with the bactec 460 TB system. Irish Vet. J. 53, 29-31.

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

PCR

Bean, D.C., Hills, A., Ryan, T., Aitken, J., 2007. Evaluation of the Bd Probetec Et System for Direct Detection of *Mycobacterium bovis* in Veterinary Specimens. J. Clin. Microbiol. 45, 3434-3435.

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

Estrada-Chavez, C., Diaz Otero, F., Arriaga Diaz, C., Villegas-Sepulveda, N., Perez Gonzalez, R., Gonzalez Salazar, D., 2004. Agreement Between PCR and

Conventional Methods for Diagnosis of Bovine Tuberculosis. Vet. Mexico. 35, 225-236.

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

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

Mishra, A., Singhal, A., Chauhan, D.S., Katoch, V.M., Srivastava, K., Thakral, S.S., Bharadwaj, S.S., Sreenivas, V., Prasad, H.K., 2005. Direct Detection and Identification of *Mycobacterium tuberculosis* and *Mycobacterium bovis* in Bovine Samples by a Novel Nested Pcr Assay: Correlation With Conventional Techniques. J. Clin. Microbiol. 43, 5670-5678.

Parra, A., Garcia, N., Garcia, A., Lacombe, A., Moreno, F., Freire, F., Moran, J., De Mendoza, J.H., 2008. Development of a Molecular Diagnostic Test Applied to Experimental Abattoir Surveillance on Bovine Tuberculosis. Vet. Microbiol. 127, 315-324.

Perez, A., Reniero, A., Forteis, A., Meregalli, S., Lopez, B., Ritacco, V., 2002. Detection of *Mycobacterium bovis* in Milk by Bacteriological Methods and Polymerase Chain Reaction. Rev. Argent. Microbiol. 34, 45-51.

Romero, R.E., Garzon, D.L., Mejia, G.A., Monroy, W., Patarroyo, M.E., Murillo, L.A., 1999. Identification of *Mycobacterium bovis* in bovine clinical samples by PCR species-specific primers. Can. J. Vet. Res. 63, 101-106.

Taylor, M.J., Hughes, M.S., Skuce, R.A., Neill, S.D., 2001. Detection of *Mycobacterium bovis* in bovine tuberculous specimens utilising real-time fluorescence and FRET probe rapid-cycle PCR. Irish Vet. J. 54, 105-106.

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

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

SICCT test with IFN-γ blood test

Aranaz, A., De Juan, L., Bezos, J., Alvarez, J., Romero, B., Lozano, F., Paramio, J.L., Lopez-Sanchez, J., Mateos, A., Dominguez, L., 2006. Assessment of Diagnostic

Tools for Eradication of Bovine Tuberculosis in Cattle Co-Infected With *Mycobacterium bovis* and *M-Avium subsp. paratuberculosis*. Vet. Res. 37, 593-606.

SIT with IFN-γ blood test

Llamazares, O.R.G., Martin, C.B.G., Nistal, D.A., Redondo, V., Rodriguez, L.D., Ferri, E.F.R., 1999. Field Evaluation of the Single Intradermal Cervical Tuberculin Test and the Interferon-Gamma Assay for Detection and Eradication of Bovine Tuberculosis in Spain. Vet. Microbiol. 70, 55-66.

Caudal fold with ELISA

Gaborick, C.M., Salman, M.D., Ellis, R.P., Triantis, J., 1996. Evaluation of a five-antigen ELISA for diagnosis of tuberculosis in cattle and Cervidae. J. Am. Vet. Med. Assoc. 209, 962-966.

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

Plackett, P., Ripper, J., Corner, L.A., Small, K., Witte, K., Melville, L., Hides, S., Wood, P.R., 1989. An ELISA for the detection of anergic tuberculous cattle. Aust. Vet. J. 66, 15-19.

Caudal fold with IFN-γ blood test

Ryan, T.J., de Lisle, G.W., Wood, P.R., 1991. The performance of the skin and gamma interferon tests for the diagnosis of tuberculosis infection in cattle in New Zealand. Massey University Continuing Education, 143-150. Unpublished.

Whipple, D.L., Bolin, C.A., Miller, J.M., 1996. Distribution of lesions in cattle infected with *Mycobacterium bovis*. J. Vet. Diagn. Invest. 8, 351-354.

Wood, P.R., Corner, L.A., Rothel, J.S., Baldock, C., Jones, S.L., Cousins, D.B., Mccormick, B.S., Francis, B.R., Creeper, J., Tweddle, N.E., 1991. Field Comparison of the Interferon-Gamma Assay and the Intradermal Tuberculin Test for the Diagnosis of Bovine Tuberculosis. Aust. Vet. J. 68, 286-290.

Wood, P.R., Corner, L.A., Rothel, J.S., Ripper, J.L., Fifis, T., Mccormick, B.S., Francis, B., Melville, L., Small, K., Dewitte, K., Tolson, J., Ryan, T.J., Delisle, G.W., Cox, J.C., Jones, S.L., 1992. A Field-Evaluation of Serological and Cellular Diagnostic-Tests for Bovine Tuberculosis. Vet. Microbiol. 31, 71-79.

Caudal fold with SICCT test

Norby, B., Bartlett, P.C., Fitzgerald, S.D., Granger, L.M., Bruning-Fann, C.S., Whipple, D.L., Payeur, J.B., 2004. The sensitivity of gross necropsy, caudal fold and comparative cervical tests for the diagnosis of bovine tuberculosis. J. Vet. Diagn. Invest. 16, 126-131.

Culture with microscopic examination

Liebana, E., Johnson, L., Gough, J., Durr, P., Jahans, K., Clifton-Hadley, R., Spencer, Y., Hewinson, R.G., Downs, S.H., 2008. Pathology of Naturally Occurring Bovine Tuberculosis in England and Wales. Vet. J. 176, 354-360.

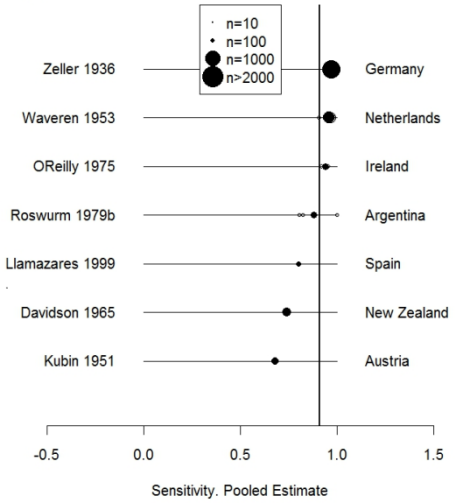
Multiple IFN-γ blood tests

Buddle, B.M., de Lisle, G.W., Ryan, T.J., 2005. Optimisation and definition of the criteria of using a *Mycobacterium bovis*-specific ancillary serial test for the diagnosis of bovine tuberculosis in cattle. Wallaceville Animal Research Centre & New Zealand Food Safety Authority. 1-17. Final Report to Animal Health Board. Unpublished.

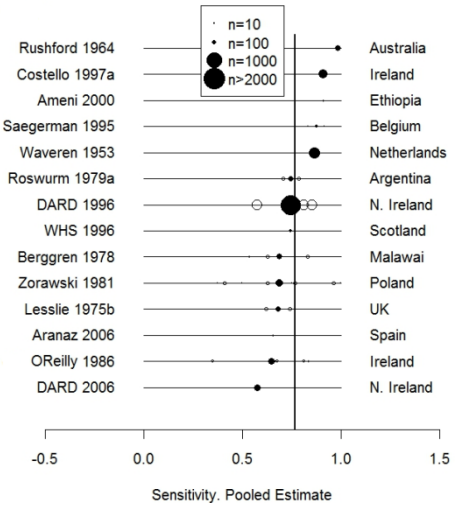
Plots of crude pooled estimates[1] of diagnostic test performance by reference and country that formed data in statistical meta-analyses of test performance[2], that were extracted from references identified in the systematic review.

1. Sensitivity estimates

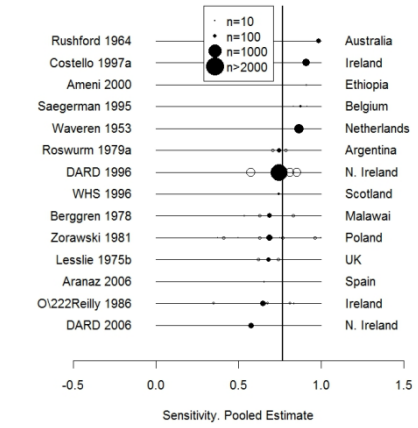
Single Intradermal skin Test



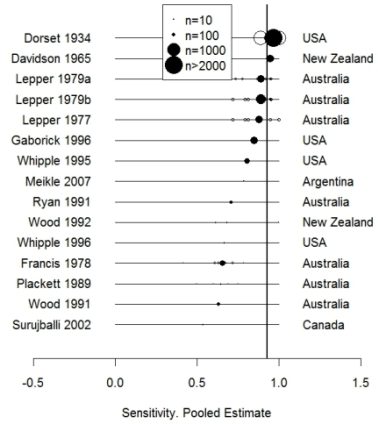
SICCT test – Standard Interpretation



SICCT test – Severe Interpretation



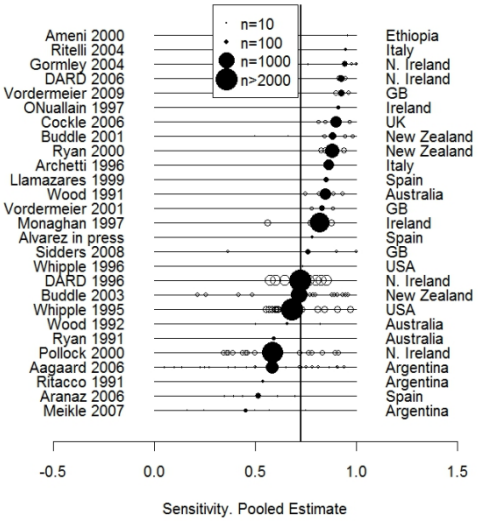
Caudal Fold skin test



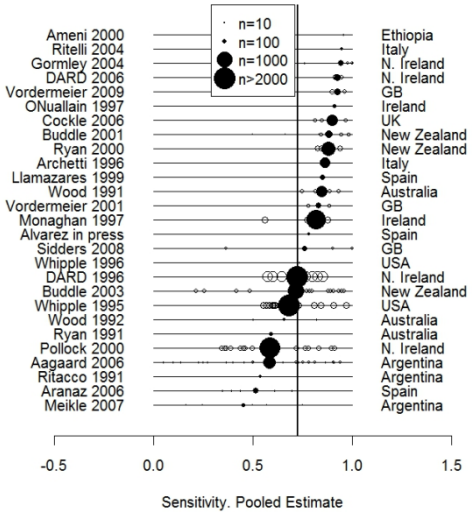
[1] Plots created using R software: https://www.r-project.org/. Size of bubbles is indicative of population size.

[2] Nu□ez-Garcia, J., Downs, S.H., Parry, J.E., Abernethy, D.A., Broughan, J.M., Cameron, A.R., Cook, A.J., de la Rua-Domenech, R., Goodchild, A.V., Gunn, J., More, S.J., Rhodes, S., Rolfe, S., Sharp, M., Upton, P.A., Vordermeier, H.M., Watson, E., Welsh, M., Whelan, A.O., Woolliams, J.A., Clifton-Hadley, R.S., Greiner, M., 2017. Meta-analysis of the sensitivity and specificity of a range of ante-mortem and post-mortem diagnostic tests for bovine tuberculosis in the UK and Ireland, Prev. Vet. Med. doi: 10.1016/j.prevetmed 2017.02.017.

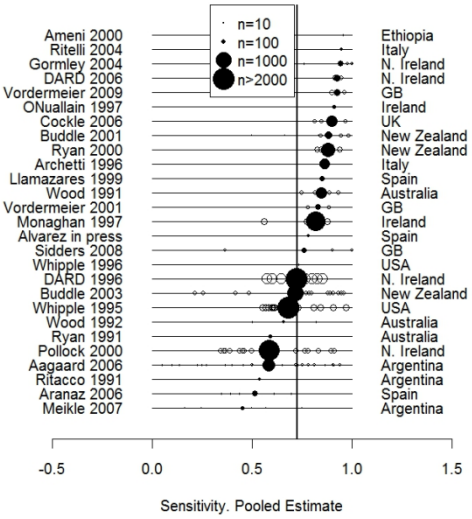
IFN-γ blood test – Bovine PPD



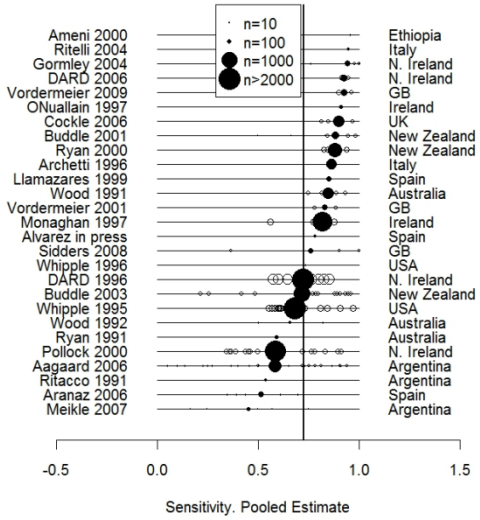
IFN-γ blood test – Bovine-Avian PPD



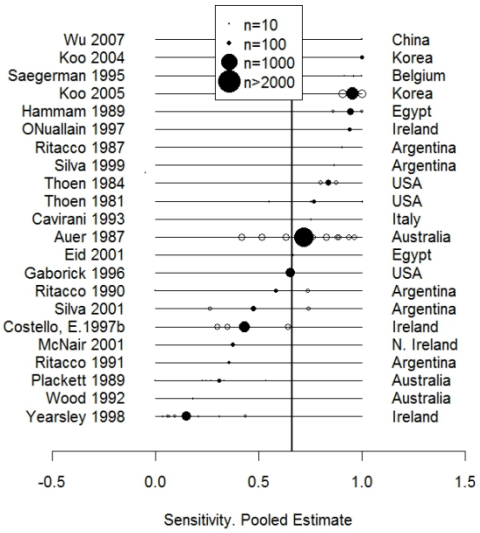
IFN-γ blood test – ESAT6/CFP10



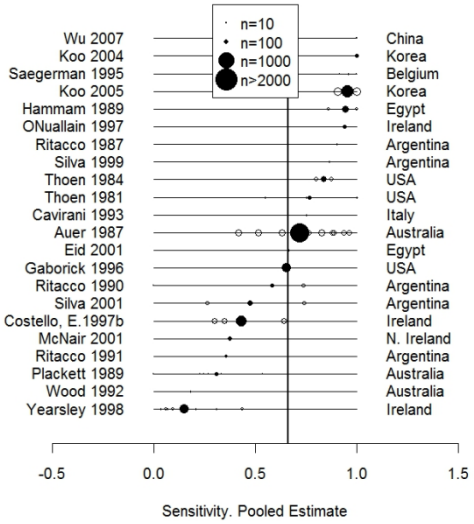
IFN-γ blood test – MPB70



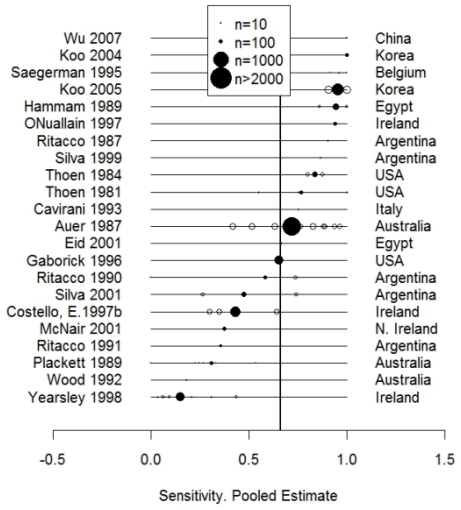
ELISA – Bovine PPD



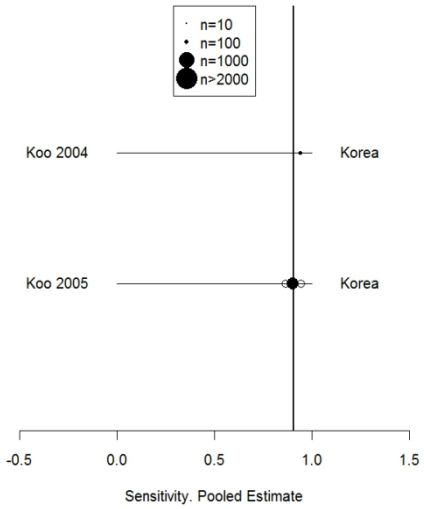
ELISA – Bovine PPD -Avian PPD



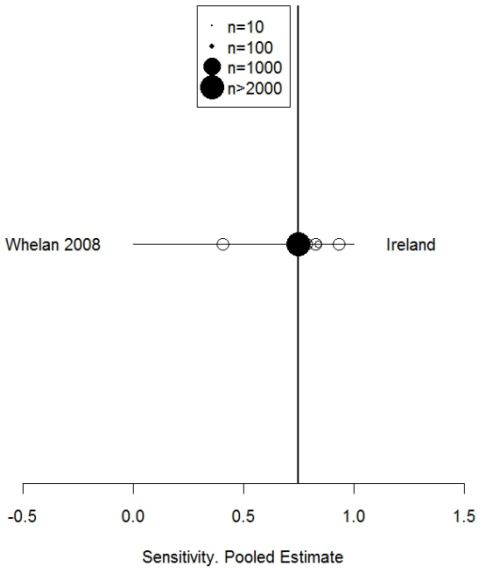
ELISA – MPB70



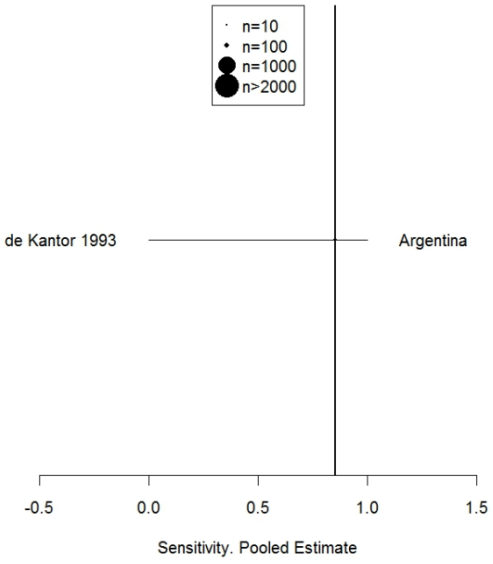
Latex Bead Agglutination Assay



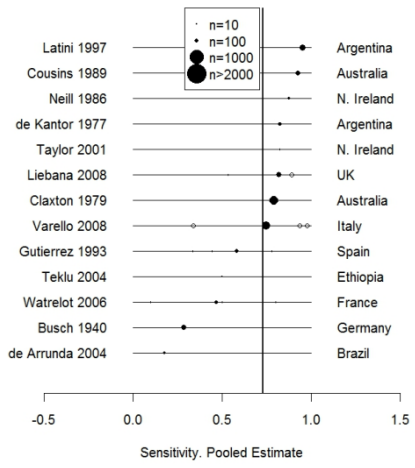
Multiplex Immunoassy



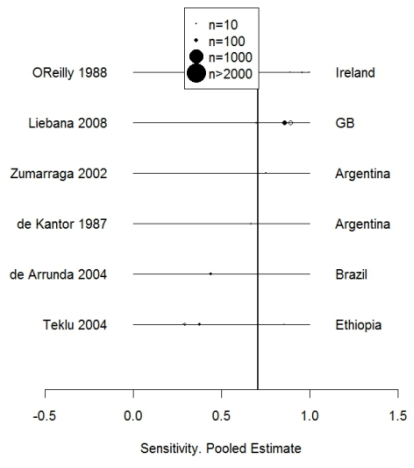
Glutaraldehyde



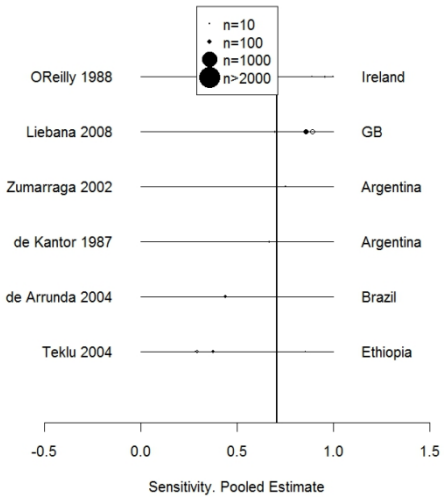
Microscopic Examination



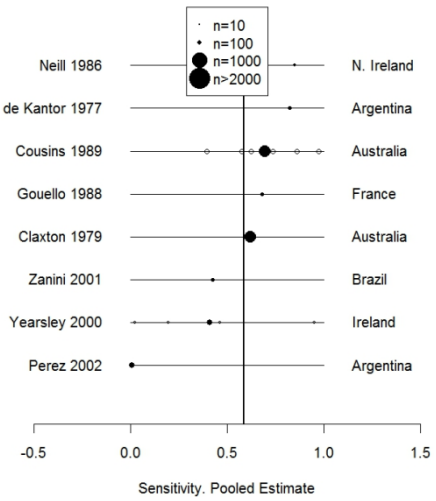
Post-mortem (Detailed/Laboratory)



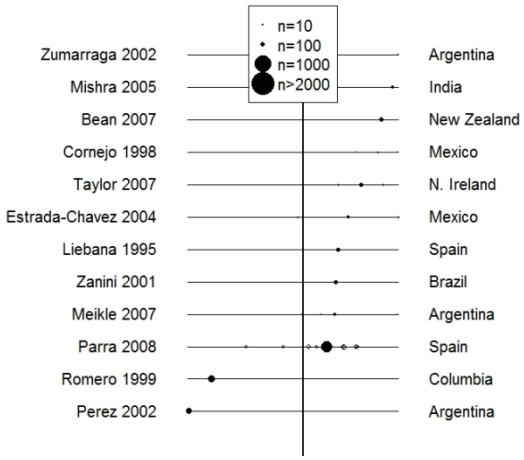
Post-mortem (Meat Inspection)



Culture

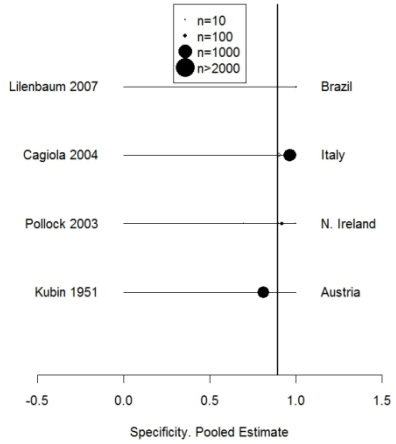


PCR

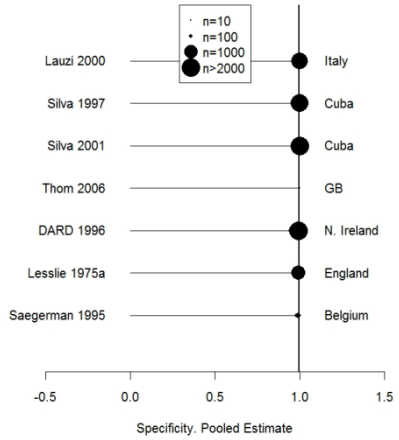


2. Specificity estimates

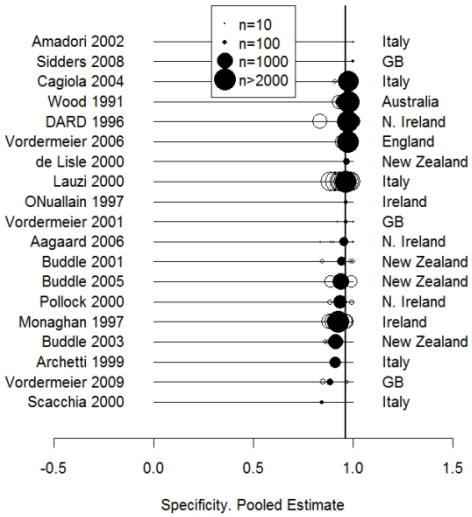
Single Intradermal Skin test



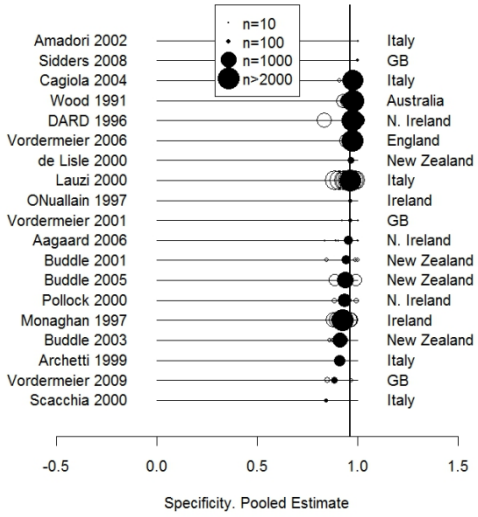
SICCT test – Standard Interpretation



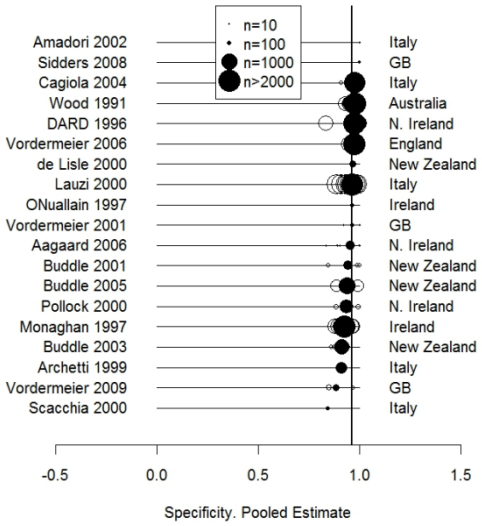
IFN-γ blood test – Bovine PPD



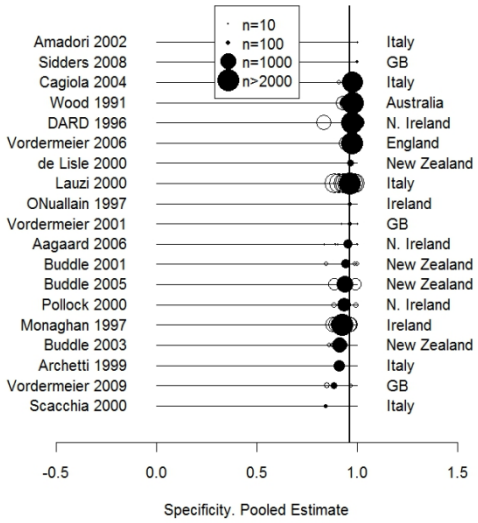
IFN-γ blood test – Bovine PPD - Avian PPD



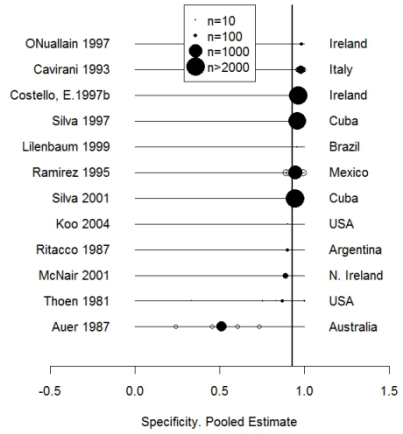
IFN-γ blood test – ESAT6/CFP10



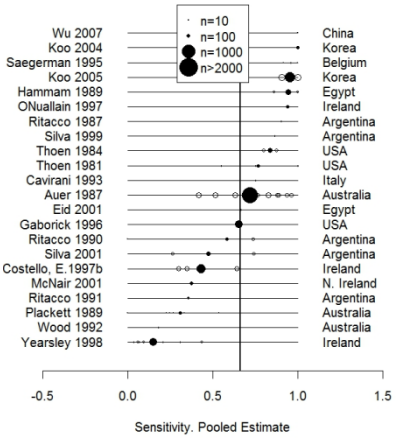
IFN-γ blood test – MPB70



ELISA – Bovine PPD

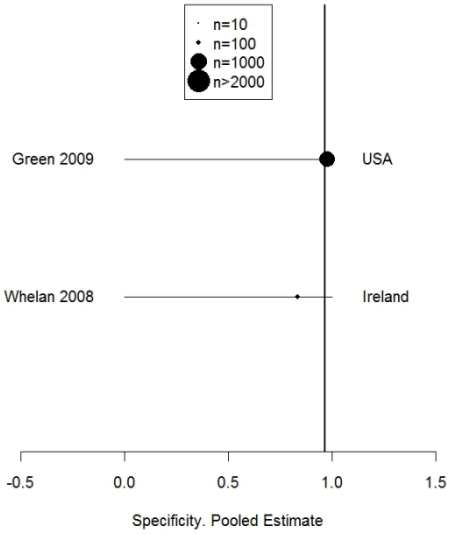


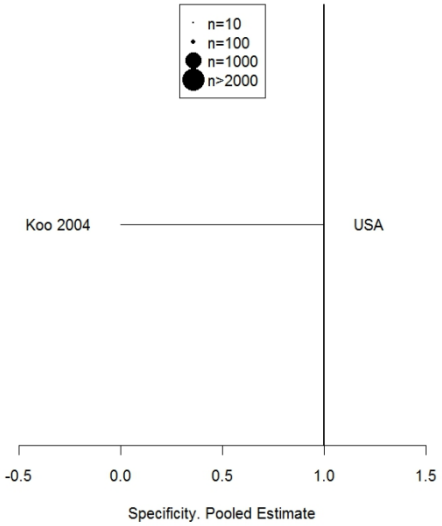
ELISA – Bovine-Avian PPD



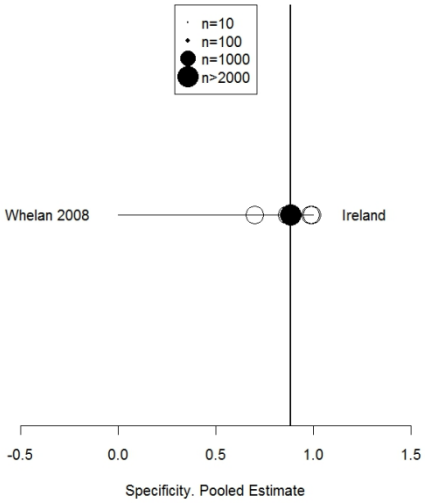
Rapid test

Latex Bead Agglutination Assay

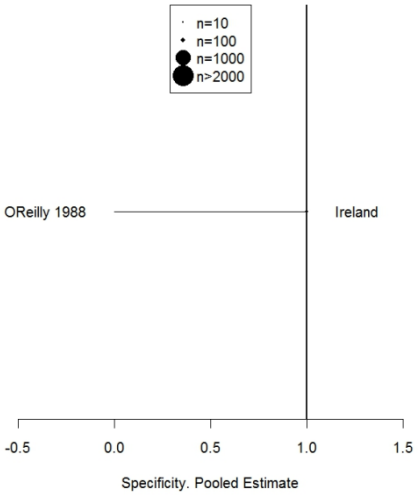


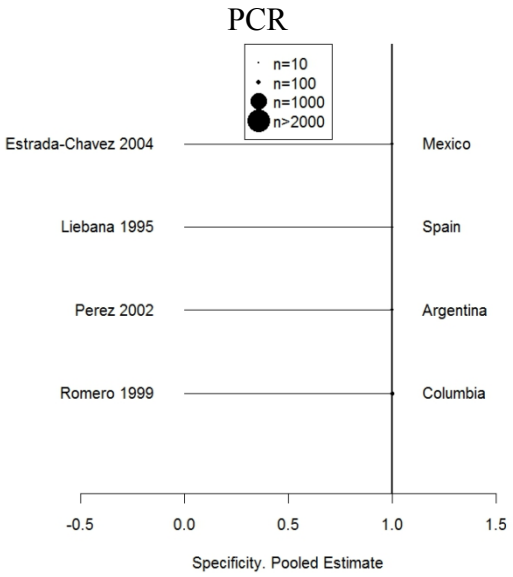


Multiplex Immunoassay



Post-mortem




PCR
n=10
n=100
n=1000
n>2000
Estrada-Chavez 2004
Mexico
Liebana 1995
Spain
Perez 2002
Argentina
Romero 1999
Columbia
-0.5
0.0
0.5
1.0
1.5
Specificity. Pooled Estimate