



HAL
open science

Using latent class analysis to estimate the test characteristics of the interferon- γ test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions

Tracy Clegg, Anthony Duignan, Clare Whelan, Eamonn Gormley, Margaret Good, John Clarke, Nils Toft, Simon J. More

► To cite this version:

Tracy Clegg, Anthony Duignan, Clare Whelan, Eamonn Gormley, Margaret Good, et al.. Using latent class analysis to estimate the test characteristics of the interferon- γ test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions. *Veterinary Microbiology*, 2011, 151 (1-2), pp.68. 10.1016/j.vetmic.2011.02.027 . hal-00701909

HAL Id: hal-00701909

<https://hal.science/hal-00701909v1>

Submitted on 28 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Title: Using latent class analysis to estimate the test characteristics of the interferon- γ test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions

Authors: Tracy Clegg, Anthony Duignan, Clare Whelan, Eamonn Gormley, Margaret Good, John Clarke, Nils Toft, Simon J. More

PII: S0378-1135(11)00108-8
DOI: doi:10.1016/j.vetmic.2011.02.027
Reference: VETMIC 5194

To appear in: *VETMIC*



Please cite this article as: Clegg, T., Duignan, A., Whelan, C., Gormley, E., Good, M., Clarke, J., Toft, N., More, S.J., Using latent class analysis to estimate the test characteristics of the interferon- γ test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions, *Veterinary Microbiology* (2010), doi:10.1016/j.vetmic.2011.02.027

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Using latent class analysis to estimate the test characteristics of the interferon- γ test, the single intradermal comparative tuberculin test and a multiplex immunoassay under Irish conditions

Tracy Clegg¹, Anthony Duignan², Clare Whelan³, Eamonn Gormley⁴, Margaret Good², John Clarke³, Nils Toft⁵, Simon J. More¹

- 1 Centre for Veterinary Epidemiology and Risk Analysis, Veterinary Sciences Centre, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Belfield, Dublin 4, Ireland
- 2 Department of Agriculture and Food, Kildare Street, Dublin 2, Ireland
- 3 Enfer Scientific, Unit T, M7 Business Park, Newhall, Naas, Kildare, Ireland
- 4 Veterinary Sciences Centre, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Belfield, Dublin 4, Ireland
- 5 Department of Large Animal Sciences, The Royal Veterinary and Agricultural University, Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark

Corresponding author:

Tracy Clegg

Centre for Veterinary Epidemiology and Risk Analysis, Veterinary Sciences Centre, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Belfield, Dublin 4, Ireland.

Tel: +353 1 7166142; Fax: +353 1 7166147. E-mail address: tracy.clegg@ucd.ie

Abstract

Considerable effort has been devoted to improving the existing diagnostic tests for bovine tuberculosis (single intradermal comparative tuberculin test [SICTT] and γ -interferon assay [γ -IFN]) and to develop new tests. Previously, the diagnostic characteristics (sensitivity, specificity) have been estimated in populations with defined infection status. However, these approaches can be problematic as there may be few herds in Ireland where freedom from infection is guaranteed. We used latent class models to estimate the diagnostic characteristics of existing (SICTT and γ -IFN) and new (multiplex immunoassay [Enferplex-TB]) diagnostic tests under Irish field conditions where true disease status was unknown. The study population consisted of herds recruited in areas with no known TB problems (2,197 animals) and herds experiencing a confirmed TB breakdown (2,740 animals). A Bayesian model was developed, allowing for dependence between SICTT and γ -IFN, while assuming independence from the Enferplex-TB test. Different test interpretations were used for the analysis: SICTT (standard and severe interpretation), γ -IFN (a single interpretation), and a range of interpretations for the Enferplex-TB (level-1 [high sensitivity interpretation] to level-5 [high specificity interpretation]). The sensitivity and specificity (95% posterior credibility intervals; 95% PCI) of SICTT[standard] relative to Enferplex-TB[level-1] and γ -IFN were 52.9-60.8% and 99.2-99.8%, respectively. Equivalent estimates for γ -IFN relative to Enferplex-TB[level-1] and SICTT were 63.1-70.1% and 86.8-89.4%, respectively. Sensitivity of Enferplex-TB[level-1] (95% PCI: 64.8-71.9%) was superior to the SICTT[standard], and

specificity of the Enferplex-TB[level-5] was superior to γ -IFN (95% PCI: 99.6-100.0%). These results provide robust measures of sensitivity and specificity under field conditions in Ireland and suggest that the Enferplex-TB test has the potential to improve on current diagnostics for TB infection in cattle.

Keywords: Bovine tuberculosis; Ireland; latent class analysis; γ -interferon assay; single intradermal comparative tuberculin test; multiplex immunoassay.

1. Introduction

Diagnostic tests play a key role in both surveillance (to detect new cases) and control (to clear infection following case detection) activities as part of national programmes to control and eradicate bovine tuberculosis (TB; caused by infection with *Mycobacterium bovis*). Although limitations in the performance of the diagnostic tests have been identified as an important constraint to effective control and eradication (de la Rua-Domenech et al. 2006), these were not sufficient to prevent successful eradication of infection from a number of countries, including Australia (More, 2009). A range of diagnostic tests for TB are available, as reviewed by Monaghan et al. (1994) and de la Rua-Domenech et al. (2006), and considerable efforts have been made, over many years, towards the development of new, and the improvement of existing, tests. There are a number of challenges to developing improved TB diagnostics (de la Rua-Domenech et al., 2006), relating both to sensitivity and specificity (often resulting from cross-reaction with non-pathogenic mycobacteria and *Mycobacterium avium* subsp. *paratuberculosis*,

Álvarez et al., 2009). Notable advances in TB diagnostics include the development of the γ -interferon assay (γ -IFN) in Australia in the early 1990s (Wood et al., 1990) and the identification of specific antigens (including ESAT-6 and CFP-10) for use in cellular-based assays (Pollock et al., 2006). Although antibody-based diagnostics have generally proved disappointing (de la Rúa-Domenech et al., 2006), Whelan et al. (2008) reported promising results with a novel multiplex immunoassay that simultaneously measures antibody recognition of up to 25 antigens in a single well of a 96-well plate array format.

Detailed information on the operating characteristics of key diagnostic tests for TB has recently been reviewed (de la Rúa-Domenech et al., 2006). Under Irish conditions, the single intradermal comparative tuberculin test (SICTT) has an estimated sensitivity of 90.9% (89.6-91.2) (Costello et al., 1997) and specificity of 99.8-99.9% (O'Reilly, 1992). Field experience indicates that the actual percentage of false positive reactors to the SICTT on a national basis is only a fraction of 1% (O'Reilly and Mac Clancy, 1978). In a study conducted by Monaghan et al. (1997), the sensitivity and specificity of the γ -IFN under Irish conditions were estimated at 56.2-87.7% and 88.1-96.6%, respectively, depending on the cut-off selected. In a later Irish study, Gormley et al. (2006) estimated a sensitivity of 88%, based on the assessment of cattle diagnosed with TB at post-mortem, and a specificity of around 95%, based on a study of 21 herds that had been free from TB for the previous five years and were assumed to be non-infected. Recently, preliminary work has been reported by Whelan et al. (2008) on development and validation of a new multiplex immunoassay (Enferplex TB). The results indicated that this test had a sensitivity of 93.1% and specificity of 98.4% (Whelan et al 2008). The multiplex assay

allows for the detection of the host antibody response to multiple single antigens that are specific to *M. bovis*, and the data on sensitivity suggest it may assist in the detection of early infection in animals.

Classically, operating characteristics of diagnostic tests have been estimated in populations with defined infection status (sensitivity estimates in infected populations and specificity estimates in non-infected populations). However, many of the epidemiological issues associated with the validation of diagnostic tests (Greiner and Gardner, 2000) are particularly problematic with TB diagnostics, including the choice of a reference test (gold standard) and reference population, and the methods used to sample from this population. In countries where infection is endemic, such as Ireland, it is not possible to accurately estimate test specificity as there are likely to be few herds where freedom from infection with *M. bovis* can be guaranteed. Recently, a number of studies have described the use of latent class models, using either frequentist (maximum likelihood estimation, MLE) or predominantly Bayesian modelling strategies, to estimate the operating characteristics of two or more diagnostic tests where true disease status is not known (Hui and Walter 1980; Branscum et al., 2005; Toft et al., 2007). Using these methods, none of the tests are considered a reference test (or ‘gold standard’), and therefore the disease status of the study subjects is designated ‘latent’ (existing but not presently evident or realised). The assumptions underpinning the original latent class analyses are (Hui and Walter 1980): two or more tests are applied in two or more populations with differing prevalence; the test characteristics are constant across all populations; and the tests are conditionally independent. The latter assumption can be

relaxed, either by Bayesian methods through incorporating prior information (Branscum et al. 2005) or by using at least three candidate tests, with one of them being conditionally independent of the others, for the test evaluation (Toft et al. 2007).

The objective of this study was to estimate the test characteristics (sensitivity, specificity) of three diagnostic tests (the SICTT, γ -IFN and Enferplex TB) for bovine tuberculosis under Irish field conditions using latent class analysis.

2. Materials and methods

2.1 The diagnostic tests

2.1.1 Single intradermal comparative tuberculin test

The SICTT was conducted using standard methods, as described previously (Monaghan et al., 1994; de la Rúa-Domenech et al., 2006). Briefly, the test is conducted by separately injecting avian and bovine purified protein derivative (PPD) tuberculin intradermally into defined sites on the neck of cattle. The test is read 72 hours later, by comparing the relative millimetre increase in skin fold thickness (an *in-vivo* cell mediated response to each tuberculin) at each injection site. Two interpretations of the skin test were used in this study: ‘standard interpretation’ (SICTT[standard]) if the bovine reaction was both positive (≥ 4 mm) and exceeded the avian reaction by more than 4 mm, and ‘severe

interpretation' (SICTT[severe]) if the bovine reaction was either positive ($\geq 4\text{mm}$) or inconclusive ($> 2\text{mm}$) and exceeded the avian reaction.

2.1.2 The γ -interferon assay

The γ -IFN (Bovigam® γ -IFN; Prionics AG, Schlieren-Zurich, Switzerland) was conducted using standard methods, as recommended by the manufacturer. Briefly, heparinised whole blood samples were incubated for 16-24 h in the presence of test antigens (typically bovine PPD tuberculin and avian PPD tuberculin, Lelystadt) and a nil-antigen (negative) control. The level of γ -IFN in the plasma was then measured using a sandwich enzyme-linked immunosorbent assay (ELISA) (Gormley et al., 2006). The interpretation level was chosen on the basis of results from an Irish field study by Monaghan et al. (1997) where the assay was carried out on known reactor cattle (sensitivity) and clear herds (specificity). The interpretation level used in this study was the one that maximised the sensitivity of the test with respect to the lesion and *M. bovis* culture status of the animals. This interpretation is currently used in the Irish bovine TB control program. As with the SICTT, this assay acts as a measure of the cell mediated immune (CMI) response of the host to *M. bovis* infection. In Ireland, the γ -IFN is mainly used in heavily infected herds, predominantly to identify animals that have shown a negative response to the SICTT test but which may be infected with *M. bovis* (Gormley et al., 2006).

2.1.3 The multiplex immunoassay

The multiple immunoassay (Enferplex TB; Enfer Scientific Ltd, Naas, Co. Kildare, Ireland) was conducted using methods described previously (Whelan et al., 2008). Briefly, a multiplex chemiluminescent immunoassay was developed to simultaneously detect antibody recognition of up to 25 antigens in a single well in a 96-well plate array format. In this assay, the chemiluminescent signal is captured with a digital imaging system and analyzed with bespoke software that tracks each serum for its pattern of antibody recognition for *M. bovis* antigens. Five different cut-off levels were used in the current study (Enferplex TB[level 1 to 5]), the 5 interpretation levels of this test were in order of increasing specificity (from 1 to 5).

2.2 Study herds and animals

2.2.1 Animals from herds where the infection risk is very low ('low risk areas')

Methods to identify areas with low infection risk (so-called 'Greenfield sites') have been described previously (More and Collins, 2007). Briefly, using ArcView 9.1 (ESRI, Redlands, CA, USA), Ireland was divided into equal sized 1 km hexagons. The TB status of and animal numbers within each hexagon (TB history, farm size etc) was determined based on the status of all underlying farms. Hexagons were sequentially removed according to the following criteria: those with an historic TB problem (a herd was permitted a single TB breakdown with no more than 2 reactors during the 5 years between 2002 and 2006, any herds failing to meet this criteria were flagged and all

hexagons touching flagged herds were removed); those not touching farms stocked with a minimum of 17 animals and maximum of 80 animals (2nd and 3rd quartiles of the population, respectively); those within 1 km of areas treated under the Government's reactive badger culling strategy; those touching any previously excluded hexagon; and those overlaying common grazing land. The internal hexagon boundaries were then dissolved creating 'Greenfield' areas which were checked to ensure it was on or close to agricultural land. Using 2007 data, we identified 189 farms with cattle herds, in total 4,831 cattle, within 1,250m of these 'Greenfield' areas.

A full-herd SICTT is conducted annually on all cattle herds in Ireland, as part of the TB eradication programme (More and Good, 2006). For this study, the SICTT was conducted, and all samples collected, by a DAFF veterinary practitioner (veterinary inspector, VI; whole-time temporary veterinary inspector, WTVI) from all animals in each study herd greater than 6 months of age at day one (the day of tuberculin injection) of the annual herd test during 2008. The VI/WTVI collected several samples from each eligible animal, including serum and blood in lithium heparin and recorded the herd and animal identification, and the date of sampling.

2.2.2 Animals from herds where the infection risk is high ('high risk areas')

For a defined period during 2008, and as feasible, samples were collected at the direction of a VI from herds with evidence of active infection (those with more than one standard SICTT test positive at the initial breakdown test). The SICTT test had been conducted by

either a VI/WTVI or a private veterinary practitioner (PVP). In each of these herds, samples (as above) were collected, where feasible, from all reactors and in-contact cohort animals greater than 6 months of age at day 1 or no less than 10 days following day one of the breakdown test. The latter time interval was chosen in order to minimise any potential anamnestic response to the injection of tuberculin. This was evaluated in a study by Gormley et al. (2004) and Doherty et al. (1995) who showed that there was no group anamnestic response after 10 or 7 days, respectively, post-injection when immune responses were measured in the γ -IFN assay. Similarly for the Enferplex TB test there was no detectable anamnestic effect in response to the tuberculin injection when blood samples were taken 10 days after injection (unpublished results). The VI recorded the herd and animal identification, and the date of sampling.

2.3 Calculating test characteristics

A latent class analysis was adopted using a Bayesian model implemented in WinBUGS (Speigelhalter et al., 2004), which incorporated the three tests: the γ -IFN, the SICTT (standard and severe interpretations separately) and the Enferplex TB test (separately at each cut off level). The SICTT and γ -IFN were not considered independent; therefore, a model allowing for the covariance between the 2 tests, while assuming independence from Enferplex TB was used as the primary model. This model is essentially the same as the one used in Toft et al. (2007) and we refer to this study for further description of the model. In order to ensure that the model was identifiable, i.e. all parameters could be estimated based on data alone, a three test and two population model was used, with the

populations being defined as the animals from low risk and high risk areas, respectively. It was assumed that the prevalence in the samples from those two populations would differ and that the test characteristics could be considered constant across the populations. We shall address the latter assumption as well as the conditional independence issue further in the discussion. The model was run for each combination of SICTT interpretation: severe/standard and for each combination of the Enferplex TB test (5 levels), therefore in total 10 (2 by 5) different analyses were carried out.

In all analyses, we used uninformative priors in the shape of a uniform distribution on the interval between 0 and 1, modelled using the Beta (1,1) distribution. The priors for the covariances were modelled using uniform distributions with ranges given by the limits described in Toft et al. (2007). The first 10,000 Monte Carlo (MC) samples were used as a burn-in to allow convergence, and the next 100,000 samples were used for posterior inference. Posterior inference was estimated by calculating posterior means along with 95% posterior credibility intervals. Convergence after the initial burn-in was assessed by visual inspection of the time-series plots of selected variables as well as Gelman-Rubin diagnostic plots using three sample chains with different initial values. Two models with and without covariance were compared using the deviance information criterion (DIC). Using the limits for the conditional covariance described by Toft et al. (2007), the posterior mean covariance of Se/Sp ($\gamma_{Se/Sp}$) was expressed as a percentage of the maximum possible (positive) covariance. The Bayesian posterior probabilities (POPR) were used to compare the test parameters in a manner more comparable to traditional frequentist statistical methods. The POPR were used to decide in favour or against

several hypotheses such as $Se_{\gamma\text{-IFN}} < Se_{\text{SICTT}}$. The POPR was calculated as the proportion of MC samples for which the hypothesis was true (Toft et al., 2007).

3. Results

3.1 Descriptive information

3.1.1 Study enrolment

a. Animals from herds in low risk areas

Sampling was conducted in 153 of the 189 herds of the larger Greenfield sites suitable for inclusion in this study. Following exclusion of herds that did not match qualifying criteria for analysis (e.g., test already completed, no stock present, etc), the SICTT and Enferplex TB results were available for 2,285 animals in 140 herds. γ -IFN results were available for 2,197 animals from 136 herds; 88 animals (including all sampled animals from 4 herds) were not γ -IFN tested. The average herd size of the 140 study herds was 23 animals, each with an average of 16 animals tested. The primary enterprise type in these herds was either dairy (18 herds), suckler (92), beef (29) or other (1).

b. Animals from herds in high risk areas

Samples were collected from 126 herds (2,740 animals). SICTT, Enferplex TB and γ -IFN results were available for all of these animals. These 126 herds held on average 91 animals, with an average 22 animals tested. The primary enterprise type in these herds was either dairy (36 herds), suckler (73), beef (16) and other (1). In 9 of these herds, samples were collected from in-contact cohort animals, but not from the positive SICTT animals.

3.2 Test characteristics

3.2.1 Bayesian models

The number of positive animals for varying test combinations of Enferplex TB, γ -IFN and SICTT[standard, severe], by area and differing cut-off interpretations for Enferplex TB and SICTT, is presented in Table 1. The estimated operating characteristics of each test are presented in Table 2 and Table 3 based on a Bayesian estimation and after separately comparing Enferplex TB[levels 1 to 5] to γ -IFN and SICTT[standard, severe], along with the proportion of maximum covariance explained by the Se and Sp covariance terms.

<Table 1 here>

<Table 2 here>

<Table 3 here>

<Figure 1 here>

The importance of the covariance between γ -IFN and SICTT[standard, severe] was tested by comparing the DIC of a model with and without the covariance term. In all of the models presented in Table 2 and Table 3 the DIC was higher in the model that did not include the covariance between the tests, indicating that the covariance term is necessary. Furthermore in all analyses the 95% PCI for the covariance parameters did not include 0, suggesting a non-negligible effect of the covariance parameter.

The SICTT (both standard and severe) and the Enferplex TB levels 2 to 5 had the highest specificities while the Enferplex TB[level-1], SICTT(severe) and γ -IFN had the highest sensitivities (Figure 1). Based on the Bayesian POPR, the estimated sensitivity of Enferplex TB[level-1] was the same as the sensitivity for γ -IFN (POPR = 0.22) and both of these test sensitivities were higher than the SICTT[standard] sensitivity (POPR<0.01). Similarly, Enferplex TB[level-5] had the highest level of specificity (99.8%, 95% PCI: 99.6; 100.0) and based on the Bayesian POPR, the specificity of Enferplex TB[level-5] was higher than both γ -IFN (POPR <0.01) and SICTT[standard] (POPR = 0.05).

The estimated posterior mean prevalence for the low risk area varied from 0.1% to 0.4% and for the high risk area the prevalence varied from 40.4% to 51.7%.

There was little change in the sensitivity and specificity estimates after excluding 9 herds from the high risk area where a standard reactor was not sampled and a further 2 herds from the low risk area that were thought to be infected. In these calculations, the revised

sensitivities of Enferplex TB[level-1], γ -IFN and SICTT[standard] were 68.6%, 66.9% and 58.1%, respectively. The specificities were 91.9%, 87.9 and 99.7%, respectively.

4. Discussion

Latent class analysis (LCA) offers considerable promise for assisting the validation of diagnostic tests and this is the first report of this analysis applied to bovine TB. Unlike classical validation methods (which form the basis of all previous validation studies), LCA can be conducted in populations where true infection status is unknown. This study presents sensitivity and specificity estimates using LCA for three diagnostic tests (SICTT, γ -IFN, Enferplex TB), based on the application of these tests in two distinct populations of cattle selected for low and high risk of *M. bovis* infection. From the analysis of the results presented here, the Enferplex TB appears to be equivalent to or better than the SICTT and γ -IFN, and as a result, may have the potential to improve on the current diagnostics for TB infection in cattle.

The latent class models are underpinned by a number of critical assumptions. Tests are generally assumed to be conditionally independent (given that an animal is infected [or not], the probability of a positive [or negative] outcome for test 1 is the same regardless of the outcome for test 2). In the current study, the biological principles of Enferplex TB are different to those of SICTT and γ -IFN. Although the biological principles underpinning SICTT and γ -IFN are similar (each is based on a cell-mediated reaction), the models were constructed, when appropriate, to account for the covariance between

tests. These models also assume that test accuracy (sensitivity, specificity) remains constant over different populations. Johnson et al (2009) states that, in general terms, high prevalence areas might have a greater burden of clinically affected animals and more animals shedding greater numbers of the pathogen of interest. Therefore, a diagnostic test could have increased sensitivity when applied in a high prevalence population. Conversely, there might be problems with constant specificity across populations for other reasons, both related to the distribution of the disease and/or potential geographic differences in cross-reactions to other mycobacteria. Johnson et al. (2009) recommends that a biological rationale for the assumption of constant test accuracy in the two populations should be given along with an analysis of the test accuracy carried out separately in the two populations. Firstly to address the biological rationale we need to look at the prevalence of TB infection within Ireland. In Ireland the TB eradication programme was first established in 1954 and all herds are tested annually for TB. Any test positive animals and others considered infected are then culled. As a result of this programme, advanced/clinical cases of TB in cattle are rarely encountered. Since the mid 1960s, the TB animal incidence has remained relatively stable, at approximately 0.4% annually (Good, 2006; More and Good, 2006). This incidence estimate is a measure of reactors to the SICTT and not of clinical disease. In addition, cattle-to-cattle transmission in Ireland is relatively uncommon (Griffin and Dolan, 1995; Costello et. al. 1998). Therefore, although we have created two populations of cattle with differing infection prevalence, it is likely that the clinical stage of disease and associated test sensitivities in both populations will be similar. Secondly, examining test estimates in the two populations separately, the estimated test accuracies can be interpreted as estimates of the

average sensitivity/specificity across populations (Branscum et al., 2005). Toft et al. (2005) and Johnson et al. (2009) showed that the lack of constant sensitivity across populations resulted in estimates that were biased towards the population with the highest disease prevalence. Furthermore, the prevalence in the low risk area is so low that the sensitivity estimates are virtually unsupported by data from that population, hence the population estimate is similar to the one applicable in the high risk population (Johnson et al. 2009).

The assumption of differing prevalence between populations is central to latent class models, noting that it is preferable to have as large a difference as possible (Toft et al., 2005). In the current study, there was a very large difference in the animal prevalence estimates in the low risk (0.1-0.4%) and high risk (40.4-51.7%) populations.

In this study, the specificity of SICTT[standard] was similar, but the sensitivity was lower than previously reported. Estimates of sensitivity for SICTT[standard] varied between 56.8% (95% PCI: 52.9-60.8, when compared to γ -IFN and Enferplex TB[level-1]) and 67.5% (62.9-72.0, when compared to Enferplex TB[level-5]) (Table 2). As expected, the sensitivity of SICTT[severe] was higher than SICTT[standard], varying between 65.9% (62.1-69.8, when compared to Enferplex TB[level-1]) and 77.3% (73.1-81.4, compared to Enferplex TB[level-5]). Costello et al. (1997) previously reported that the sensitivity of SICTT[severe] under Irish conditions was 90.9%. In a recent review, de la Rua-Domenech et al. (2006) suggested a median sensitivity of 80% for SICTT[standard] and 93.5% for SICTT[severe], based on a meta-analysis of several studies with a range of

between 52% and 100%. Although there is a discrepancy in these results, it is important to note that the current study reports estimates of the relative performance of each test. It is likely that the classical validation methods are prone to overestimation of test sensitivity, given the bias towards late stage infection when culture is used as the ‘gold standard’. There may also be biases in the current study due to the assumption of constant sensitivity and specificity across the two populations. In the current study, the estimated specificity of SICTT[standard] was 99.5% (99.2-99.8, when compared to either Enferplex TB[level-1] or Enferplex TB[level-5]). The estimated specificity of SICTT[severe] was slightly lower, varying between 99.2% (98.7-99.6, when compared to Enferplex TB[level-5]) and 99.3% (98.9-99.7, when compared to Enferplex TB[level-1]). These estimates are similar to previous reports for SICTT[standard]: O’Reilly reported 99.8-99.9% and de la Rua-Domenech et al. (2006) a median value of 99.5% (range:78.8-100%).

In this study, the sensitivity and specificity estimates for γ -IFN were generally comparable to previously published reports. The estimate of relative sensitivity varied between 66.6% (95% PCI: 63.1-70.1, when compared to Enferplex TB[level-1]) and 76.2% (72.8-79.3, when compared to Enferplex TB[level-5]), and of specificity between 88.1% (86.8-89.4, when compared to Enferplex TB[level-1]) and 87.9% (86.4-89.3, when compared to Enferplex TB[level-5]). These estimates are similar to those reported by Monaghan et al. (1997). However, the sensitivity estimates are lower than those reported by both Gormley et al. (2006) and de la Rua-Domenech et al. (2006). As before, this difference may reflect inherent difficulties associated with the use of culture as a ‘gold

standard' when classical validation methods are used and the contributory effects of the assumptions of constant test accuracy using the LCA. The current results highlight the value of, but also the recognised problems associated with the γ -IFN test, which puts restrictions on its widespread usage. The γ -IFN assay has higher test sensitivity but lower specificity than SICTT[standard]. As a consequence, and particularly due to the potential for disclosing false-positive results, the γ -IFN test is generally applied to known infected herds in Ireland, to identify infected animals that are missed by the SICTT, and where residual infection is suspected (Gormley et al., 2006).

The results obtained with the Enferplex TB test are encouraging in the context of development of a novel diagnostic test. Using LCA, the operating characteristics of Enferplex TB were found to be equivalent or better than both the SICTT and γ -IFN, depending on the cut-off points used. The relative test sensitivity varied between 37.1% (95% PCI: 33.7-40.6) for Enferplex[level-5] (the cut-off for maximum specificity) and 68.4% (64.8-71.9) for Enferplex[level-1] (maximum sensitivity), noting that the latter estimate is higher than the estimated sensitivity of SICTT[standard] and within the estimated sensitivity range for γ -IFN and SICTT[severe]. The γ -IFN and SICTT[severe] are currently the most-sensitive diagnostic tests for routine field use. These relative sensitivity estimates are substantially less than previously reported for Enferplex TB (93.1%; Whelan et al., 2008). However, this is unsurprising as these earlier estimates were based on data generated from animals likely to have advanced infection. The estimated relative specificity varied between 92.1% (91.0-93.2) for Enferplex[level-1] and 99.8% (99.6-100.0) for Enferplex[level-5]. The latter estimate is higher than the

estimated specificity of SICTT[standard], currently the most-specific test used for routine annual surveillance of herds in Ireland.

The Enferplex TB potentially offers a number of advantages as a herd-based screening test. The existing assays are logistically complex, either because cattle need to be examined on two occasions (with the SICTT) or because samples need to be rapidly transported to a specialist laboratory (γ -IFN). Further, a serology based assay would remove some current concerns relating to the SICTT, including difficulties associated with test standardisation (including objective test interpretation) and the potential for false-positive results due to fraudulent interference. Blood tests do not require a mandatory inter-test interval and thus retests may be performed at short intervals. This would be particularly beneficial in the detection of animals with early-stage infection. With any TB screening test, specificity is of particular importance, noting that herd status is determined by the presence of one (or more) test-positive animal (Martin et al., 1992). The sensitivity and specificity of the SICTT is a function not just of the test itself but also of the potency of the tuberculin and the environment in which it is used. Therefore, to be a realistic alternative to the SICTT, the Enferplex TB (or an alternative) must have comparable specificity while maintaining at least an equivalent level of herd sensitivity. The results of this trial indicate that, in the Irish environment, the specificity of Enferplex TB[level-4] is very close to that of SICTT[standard] specificity, albeit with lower sensitivity. In common with the SICTT, the interpretation of Enferplex TB can be modified to take account of the level of TB infection within a herd, allowing for an increase in sensitivity (at the expense of specificity) when required. Thus, Enferplex

TB[level-3, level-2 or level-1] may be applied as the situation demands. If the Enferplex TB was found to be useful as a herd screening test, the SICTT and/or γ -IFN could also be used in parallel in an infected herd to detect infected animals that might not have responded to the Enferplex TB test. Based on current estimates, Enferplex TB[level-5], with high specificity and reasonable levels of sensitivity, may also be of assistance in countries that currently rely on post-mortem surveillance to detect TB infection. Although the sensitivity of post-mortem surveillance is comparatively poor (Corner et al., 1990; Collins, 1997; Frankena et al 2007), it can be increased by provision of motivational incentives (Kaneene et al., 2006). Post-mortem surveillance is labour-intensive and relatively expensive. A blood sample taken immediately before slaughter, with traceback of positive animals to the herd of origin, would be an alternative possibility. Alternatively, test results could be obtained from animals destined for slaughter, limiting post-mortem examination to Enferplex TB[level-5] positive animals only. Additional studies will be carried out to validate the Enferplex TB under a variety of scenarios.

Test validation has proved particularly problematic for bovine TB. Classical validation methods rely on test application in known infected and non-infected populations for sensitivity and specificity estimates, respectively (Greiner and Gardner, 2000). In sensitivity studies, microbiological culture is routinely used as the reference test ('gold standard'), and test sensitivity estimates have been based on the percentage of culture positive animals that test positive to the test in question. However, a positive culture result is more likely in animals with advanced, compared to early stage, infection.

Sensitivity also varies (and generally increases) with the stage of infection (Greiner and Gardner, 2000). Therefore, sensitivity estimates based on culture positive results will invariably lead to an overestimate of test sensitivity in the field. In contrast, specificity studies rely on test application in non-infected populations. In countries where TB is problematic, however, it is often difficult to substantiate freedom from infection (More et al., 2009), in part because residual infection is a key feature of TB in cattle (Olea-Popelka et al., 2008). In Ireland, for example, there are no herds where freedom from infection can be guaranteed. Extrapolation of specificity data is frequently unsatisfactory, noting that specificity may vary between populations (Greiner and Gardner, 2000) due to differences in the prevalence or distribution of cross-reacting organisms. As a consequence of these constraints, published data on the operating characteristics of diagnostic tests for bovine TB generally need to be interpreted with care.

The sensitivity and specificity estimates are robust to potential concerns with herd misclassification. In 9 herds in the high risk areas, samples were collected from cohort – but not reactor – animals. During the course of the study, infection was confirmed in 2 herds located in the low risk areas. In each case, concern was raised about the potential impact of herd misclassification on model outputs. For this reason, the models were run with and without these 11 herds, with very little difference to model estimates of test sensitivity and specificity.

It was not possible to comment on the ability of the tests at herd-level since only part of the herd was sampled in most cases. Further studies are currently underway to examine

the comparative utility of SICTT[standard] and Enferplex TB as herd-based screening tests under Irish conditions.

5. Conclusion

The use of latent class analysis (LCA) has provided robust estimates of the relative test characteristics of two currently used diagnostic tests (γ -IFN, SICTT) and a novel multiplex immunoassay (Enferplex TB) for TB under field conditions in Ireland. The new Enferplex TB test had displayed equivalent or better relative sensitivity/specificity estimates compared to the currently used tests. These results suggest that the Enferplex TB has the potential to improve on the current diagnostics for TB infection in cattle, the extent of that potential will be assessed in further studies.

6. Acknowledgements

The authors wish to acknowledge the assistance of all who contributed to this paper and, in particular: Tara Fitzsimons and Mairead Doyle of UCD who conducted all of the γ -IFN tests; Eduard Shuralev and the testing team from Enfer Scientific who performed the Enferplex and data analysis and also Hang Fai Kwok, Shane A. Olwill and James A. Johnston from Fusion Antibodies Limited who were also involved in the Enferplex data analysis.

7. Conflict of Interest Statement

TC, AD, EG, MG, NT and SJM have no financial or personal relationships with other people or organisations that could inappropriately influence (bias) this work. Examples of potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding. CW and JC are Enfer employees, however, neither Enfer nor its employees played any role that could have inappropriately influenced (biased) this work.

8. References

Álvarez, J., de Juan, L., Bezos, J., Romero, B., Sáez, J.L., Marqués, S., Domínguez, C., Mínguez, O., Fernández-Mardomingo, B., Mateos, A., Domínguez, 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.

Branscum, A.J., Gardner, I.A., Johnson, W.O., 2005. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev. Vet. Med.* 68, 145-163.

Collins, J.D., 1997. Meat plant surveillance and its role in the eradication of tuberculosis in cattle. In: *Selected Papers 1996: Tuberculosis Investigation Unit, University College Dublin*. pp. 55-59

Corner, L.A., Melville, L., McCubbins, K., Small, K.J., McCormick, B.S., Wood, P.R.,

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

de la Rúa-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, γ -interferon assay and other ancillary diagnostic techniques. *Res. Vet. Sci.* 81, 190-210.

Doherty, M.L., Monaghan, M.L., Bassett, H.F., Quinn, P.J. 1995. Effect of a recent injection of purified protein derivative on diagnostic tests for tuberculosis in cattle infected with *Mycobacterium bovis*. *Res. Vet. Sci.* 58, 217-221.

Frankena, K., White, P.W., O'Keeffe, J., Costello, E., Martin, S.W., van Grevenhof, I., More, S.J., 2007. Quantification of the relative efficiency of factory surveillance in the disclosure of tuberculosis lesions in attested Irish cattle. *Vet. Rec.* 161, 679-684.

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. Immunopathol.* 102, 413-20.

Gormley, E., Doyle, M.B., Fitzsimons, T., McGill, K., Collins, J.D., 2006. Diagnosis of *Mycobacterium bovis* infection in cattle by use of the gamma-interferon (Bovigam®) assay. *Vet. Microbiol.* 112, 171-179.

Greiner, M., Gardner, I.A., 2000. Epidemiological issues with the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45, 3-22.

Hui, S.L., Walter, S.D., 1980. Estimating the error rates of diagnostic tests. *Biometrics* 36, 167-171.

Johnson, W.O., Gardner, I.A., Metoyer, C.N., Branscum, A.J., 2009. On the interpretation of test sensitivity in the two-test two-population problem: Assumptions matter. *Prev. Vet. Med.* 91, 116-121.

Kaneene, J.B., Miller, R., Meyer, R.M., 2006. Abattoir surveillance: The US experience. *Vet. Microbiol.* 112, 273-282.

Martin, S.W., Shourkri, M., Thorburn, M.A., 1992. Evaluating the health status of herds based on tests applied to individuals. *Prev. Vet. Med.* 14, 33-43.

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

Monaghan, M., Quinn, P.J., Kelly, A.P., McGill, K., McMurray, C., O'Crowley, K., Bassett, H.F., Costello, E., Quigley, F.C., Rothel, J.S., Wood, P.R., Collins, J.D., 1997. A pilot trial to evaluate the γ -interferon assay for the detection of *Mycobacterium bovis* infected cattle under Irish conditions. *Irish Vet. J.* 50, 229-232.

More, S.J., 2009. What is needed to eradicate bovine tuberculosis successfully: An Irish perspective. *Vet. J.* 180, 275-278.

More, S.J., Cameron, A.R., Greiner, M., Clifton-Hadley, R.S., Correia Rodeia, S., Bakker, D., Salman, M.D., Sharp, J.M., De Massis, F., Aranaz, A., Boniotti, M.B., Gaffuri, A., Have, P., Verloo, D., Woodford, M., Weirup, M., 2009. Defining output-based standards to achieve and maintain tuberculosis freedom in farmed deer, with reference to member states of the European Union. *Prev. Vet. Med.* 90, 254-267.

More, S.J., Collins, D.M. (eds), 2007. Biennial report 2006-07: The Centre for Veterinary Epidemiology and Risk Analysis, the TB Diagnostics and Immunology Research Centre, the badger vaccine project. University College Dublin. pp.6-7.

More, S.J., Good, M., 2006. The tuberculosis eradication programme in Ireland: A review of scientific and policy advances since 1988. *Vet. Microbiol.* 112, 239-251.

Olea-Popelka, F.J., Costello, E., White, P., McGrath, G., Collins, J. D., O'Keeffe, J., Kelton, D.F., Berke, O., More, S., Martin, S.W., 2008. Risk factors for disclosure of

additional tuberculous cattle in attested-clear herds that had an animal with a confirmed lesion of tuberculosis at slaughter during 2003 in Ireland. *Prev. Vet. Med.* 85, 81-91.

O'Reilly, L.M., Mac Clancy, B.N., 1978. Estimation of the sensitivity, specificity and predictive value of the intradermal tuberculin test. *Irish Vet. J.* 32, 127-128

O'Reilly, L.M., 1992. Specificity and sensitivity of tuberculin tests: A review. In: *Proceedings of the International Conference of Tuberculosis in Africa and the Middle East*, April 28-30, 1992 Cairo, Egypt.

Palmer, M.V., Waters, W.R., 2006. Advances in bovine tuberculosis diagnosis and pathogenesis: What policy makers need to know. *Vet. Microbiol.* 112, 181-190.

Pollock, J.M., R. M., Girvin, R. M., Lightbody, K. A., Neill, S. D., Clements, R. A., Buddle, B. M., Andersen, P., 2006. Assessment of defined antigens for the diagnosis of bovine tuberculosis in skin test-reactor cattle. *Vet Rec.* 146, 659-665.

Spiegelhalter, D.J., Thomas, A., Best, N.G., 2004. *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit.

Toft, N., Højsgaard, 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.

Toft, N., Åkerstedt, J., Tharaldsen, J., Hopp, P., 2007. Evaluation of three serological tests for diagnosis of Maedi-Visna virus infection using latent class analysis. *Vet. Microbiol.* 120, 77-86.

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.

Wood, P.R., Corner, L.A.L., Plackett, P., 1990. Development of a simple, rapid in vitro cellular assay for bovine tuberculosis based on the production of γ interferon. *Res. Vet. Sci.* 49, 46-49.

Tables:

Table 1. The number of animals positive to Enferplex TB, γ -IFN and SICTT[standard and severe], by area and differing cut-offs for Enferplex TB and SICTT.

Table 2. Bayesian estimates of the operating characteristics of Enferplex TB (at different cut-off levels), γ -IFN and SICTT[standard], using no prior information and estimating the covariance between γ -IFN and SICTT[standard]

Table 3. Bayesian estimates of the operating characteristics of Enferplex TB (at different cut-off levels), γ -IFN and SICTT[severe], using no prior information and estimating the covariance between γ -IFN and SICTT[severe]

Figure 1: A plot of the bayesian estimates of sensitivity versus 1-specificity of Enferplex TB (at different cut-off levels), γ -IFN and SICTT[standard and severe], using no prior information and estimating the covariance between γ -IFN and SICTT[severe]

Accepted Manuscript

Table 1. The number of animals positive to Enferplex TB, γ -IFN and SICTT[standard and severe], by area and differing cut-offs for Enferplex TB and SICTT.

Enferplex TB	γ -IFN	SICTT	SICTT[Standard]		SICTT[Severe]	
			Low risk area	High risk area	Low risk area	High risk area
Enferplex TB[level-1]						
P	P	P	2	462	3	514
P	P	N	16	175	15	123
P	N	P	0	74	1	103
P	N	N	156	302	155	273
N	P	P	9	137	12	190
N	P	N	192	315	189	262
N	N	P	1	80	3	135
N	N	N	1821	1195	1819	1140
Enferplex TB[level-2]						
P	P	P	0	403	0	440
P	P	N	3	124	3	87
P	N	P	0	54	0	75
P	N	N	18	162	18	141
N	P	P	11	196	15	264
N	P	N	205	366	201	298
N	N	P	1	100	4	163
N	N	N	1959	1335	1956	1272
Enferplex TB[level-3]						
P	P	P	0	376	0	412
P	P	N	2	97	2	61
P	N	P	0	46	0	64
P	N	N	15	116	15	98
N	P	P	11	223	15	292
N	P	N	206	393	202	324
N	N	P	1	108	4	174

N	N	N	1962	1381	1959	1315
Enferplex TB[level-4]						
P	P	P	0	313	0	336
P	P	N	1	64	1	41
P	N	P	0	35	0	46
P	N	N	7	61	7	50
N	P	P	11	286	15	368
N	P	N	207	426	203	344
N	N	P	1	119	4	192
N	N	N	1970	1436	1967	1363
Enferplex TB[level-5]						
P	P	P	0	286	0	305
P	P	N	0	52	0	33
P	N	P	0	29	0	37
P	N	N	3	46	3	38
N	P	P	11	313	15	399
N	P	N	208	438	204	352
N	N	P	1	125	4	201
N	N	N	1974	1451	1971	1375
<i>Total</i>			<i>2197</i>	<i>2740</i>	<i>2197</i>	<i>2740</i>

Table 2. Bayesian estimates of the operating characteristics of Enferplex TB (at different cut-off levels), γ -IFN and SICTT[standard], using no prior information and estimating the covariance between γ -IFN and SICTT[standard]

	Parameter	Mean	Posterior credibility interval	
			2.5%	97.5%
Sensitivity	Enferplex TB[level-1]	0.684	0.648	0.719
	γ -IFN	0.666	0.631	0.701
	SICTT[standard]	0.568	0.529	0.608
Specificity	Enferplex TB[level-1]	0.921	0.910	0.932
	γ -IFN	0.881	0.868	0.894
	SICTT[standard]	0.995	0.992	0.998
	% maximum cov Se	0.393	0.311	0.471
	% maximum cov Sp	0.712	0.343	0.973
Sensitivity	Enferplex TB[level-2]	0.573	0.536	0.610
	γ -IFN	0.681	0.650	0.713
	SICTT[standard]	0.587	0.549	0.624
Specificity	Enferplex TB[level-2]	0.990	0.986	0.994
	γ -IFN	0.880	0.866	0.893
	SICTT[standard]	0.995	0.991	0.998
	% maximum cov Se	0.362	0.283	0.439
	% maximum cov Sp	0.757	0.423	0.974
Sensitivity	Enferplex TB[level-3]	0.522	0.485	0.559
	γ -IFN	0.710	0.678	0.742
	SICTT[standard]	0.626	0.585	0.666
Specificity	Enferplex TB[level-3]	0.992	0.988	0.995
	γ -IFN	0.877	0.863	0.890
	SICTT[standard]	0.995	0.991	0.998
	% maximum cov Se	0.298	0.212	0.383
	% maximum cov Sp	0.756	0.420	0.976

Sensitivity	Enferplex TB[level-4]	0.417	0.382	0.454
	γ -IFN	0.751	0.718	0.783
	SICTT[standard]	0.667	0.623	0.712
Specificity	Enferplex TB[level-4]	0.996	0.993	0.998
	γ -IFN	0.878	0.863	0.892
	SICTT[standard]	0.995	0.992	0.998
	% maximum cov Se	0.181	0.078	0.281
	% maximum cov Sp	0.775	0.424	0.981
Sensitivity	Enferplex TB[level-5]	0.371	0.337	0.406
	γ -IFN	0.762	0.728	0.793
	SICTT[standard]	0.675	0.629	0.720
Specificity	Enferplex TB[level-5]	0.998	0.996	1.000
	γ -IFN	0.879	0.864	0.893
	SICTT[standard]	0.995	0.992	0.998
	% maximum cov Se	0.143	0.036	0.249
	% maximum cov Sp	0.781	0.434	0.982

Table 3. Bayesian estimates of the operating characteristics of Enferplex TB (at different cut-off levels), γ -IFN and SICTT[severe], using no prior information and estimating the covariance between γ -IFN and SICTT[severe]

	Parameter	mean	Posterior credibility interval	
			2.5%	97.5%
Sensitivity	Enferplex TB[level-1]	0.642	0.610	0.674
	γ -IFN	0.641	0.608	0.675
	SICTT[severe]	0.659	0.621	0.698
Specificity	Enferplex TB[level-1]	0.922	0.910	0.933
	γ -IFN	0.887	0.874	0.900
	SICTT[severe]	0.993	0.989	0.997
	% maximum cov Se	0.330	0.253	0.407
	% maximum cov Sp	0.617	0.281	0.904
Sensitivity	Enferplex TB[level-2]	0.533	0.500	0.566
	γ -IFN	0.657	0.627	0.687
	SICTT[severe]	0.679	0.644	0.714
Specificity	Enferplex TB[level-2]	0.990	0.986	0.994
	γ -IFN	0.886	0.873	0.898
	SICTT[severe]	0.991	0.987	0.995
	% maximum cov Se	0.299	0.223	0.378
	% maximum cov Sp	0.637	0.371	0.867
Sensitivity	Enferplex TB[level-3]	0.488	0.455	0.522
	γ -IFN	0.682	0.651	0.713
	SICTT[severe]	0.729	0.691	0.766
Specificity	Enferplex TB[level-3]	0.992	0.988	0.995
	γ -IFN	0.881	0.867	0.894
	SICTT[severe]	0.991	0.987	0.995
	% maximum cov Se	0.266	0.175	0.363

	% maximum cov Sp	0.634	0.365	0.866
Sensitivity	Enferplex TB[level-4]	0.385	0.353	0.417
	γ -IFN	0.716	0.684	0.747
	SICTT[severe]	0.766	0.726	0.807
Specificity	Enferplex TB[level-4]	0.996	0.993	0.998
	γ -IFN	0.882	0.869	0.896
	SICTT[severe]	0.992	0.987	0.996
	% maximum cov Se	0.151	0.046	0.261
	% maximum cov Sp	0.656	0.376	0.889
Sensitivity	Enferplex TB[level-5]	0.341	0.311	0.372
	γ -IFN	0.723	0.691	0.754
	SICTT[severe]	0.773	0.731	0.814
Specificity	Enferplex TB[level-5]	0.998	0.996	1.000
	γ -IFN	0.883	0.869	0.897
	SICTT[severe]	0.992	0.987	0.996
	% maximum cov Se	0.122	0.017	0.235
	% maximum cov Sp	0.662	0.382	0.894

Figure 1: A plot of the bayesian estimates of sensitivity versus 1-specificity of Enferplex TB (at different cut-off levels), γ -IFN and SICTT[standard and severe], using no prior information and estimating the covariance between γ -IFN and SICTT[severe]

Accepted Manuscript

1 **Figure 1:**

2

3

4

