Bovine tuberculosis:
A review of diagnostic tests for M. bovis infection in cattle

Sam A. J. Strain
James McNair
Stanley W. J. McDowell

Bacteriology Branch
Veterinary Sciences Division
Agri-Food and Biosciences Institute

May 2011
'there seems to be no valid reason for doubting the opinion, never seriously doubted before 1901, that human and bovine bacilli belong to the same family. On this view the answer to the question, can the bovine bacillus affect man, is obviously in the affirmative. The same answer must also be given to those who hold the theory that human and bovine tubercle bacilli are different in kind, since the ‘bovine kind’ are readily to be found as the causal agents of many fatal cases of human tuberculosis.'

Royal Commission on Tuberculosis interim report 1907

[in response to the erroneous views of Calmette and Guerin] ‘The inhalation of tubercle bacilli suspended in the atmosphere is a very certain method of infection in susceptible animals even when small doses of bacilli are employed’

Professor Sir John M’Fadyean. The Veterinary Journal 1911

‘..it seems probable that this goal [eradication of bovine TB in Great Britain] will be reached in another ten years.’

J. Francis ‘Tuberculosis in Animals and Man.’ 1958
Executive summary:

General Conclusions

A. The diagnosis of bovine tuberculosis, like human tuberculosis, remains extremely challenging. There is currently no single test which will fulfil all the criteria necessary to identify all infected animals. Of necessity a combination of approaches is likely to be needed to achieve an adequate level of diagnosis.

B. Traditionally the focus of test and cull programmes, and as a result the assessment of test performance, has been on the infected animal. In reality, not all animals infected with bTB will be infectious while some such as those that are immuno-compromised or in the later stages of disease may well be responsible for a disproportionally high proportion of disease transmission. While much more difficult than a simple comparison with infected / uninfected disease status, ideally the assessment of tests should also consider the ability of the test to remove what may be a relatively small proportion of more highly infectious individuals.

C. Sensitivity and specificity estimates for diagnostic tests can be dependent upon the animal population studied and the disease characteristics within the population. Caution should be exercised in sensitivity and specificity estimates made from populations of animals different from those within N. Ireland.

D. In overall terms current bTB prevalence in N. Ireland is relatively low and as a result it is critical that any general screening diagnostic test should have a high specificity, as this characteristic will have the greater impact on test positive predictive values i.e. the likelihood that a positive test indicates a truly infected animal. Tests with poorer specificities may still have important roles in specific disease circumstances for example in herds with high disease prevalence.

E. There is now substantial evidence that other factors influence not only the course of the disease but also the sensitivity and perhaps the specificity of the most commonly used diagnostic tests. Most importantly it appears that co-infection with parasites, most notably liver fluke, and the mycobacterial disease Johne’s disease influence the diagnostic sensitivity of both the comparative skin test and the interferon-gamma test.

F. While evidence exists on the performance of individual tests, there is a significant knowledge gap in relation to how tests could be best applied in varying field situations and disease circumstances in order to achieve optimal disease control and cost-benefit.

Skin Test

A. A large body of evidence now exists to support the high specificity of the single comparative intradermal skin test (SICTT) with typical specificity estimates being greater than 99.9%. However the sensitivity of the test remains moderate with estimates typically in the range 50-60% using the standard test interpretation. Nonetheless, the SICTT remains the best single test currently available for the diagnosis of bovine tuberculosis.

B. Importantly the comparative intradermal skin test is entirely dependent upon the quality and potency of the reagents used i.e. the purified protein derivatives (PPDa & PPDb). These reagents are central reagents to both the comparative intradermal skin test and
the interferon-gamma test. There is now substantial evidence of considerable variation in the quality of PPD tuberculins produced around the world with some showing very poor potency and likely to lead to a failure to detect disease as well as in some cases poor specificity. Further work to improve and standardise PPD potency and performance is indicated. Consideration should be given to using ancillary potency testing of tuberculin batches used by DARD to give additional assurance of their quality.

C. The use of more stringent skin test interpretations, for example severe interpretation, leads to increases in test sensitivity albeit with reduced specificity. It is likely that the application of more stringent interpretations will lead to a significant improvement in disease control in clearly defined situations where increased sensitivity is desirable and a somewhat reduced specificity acceptable.

Interferon-gamma Test

A. The interferon-gamma test is a generic assay for measuring the release of interferon-gamma from cells. As such it is misleading to describe it as one test as its performance is dependent on a large number of factors which can be varied, including the antigens used (tuberculins or \textit{M. tuberculosis} complex specific antigens), the way samples are stored and processed and the criteria used to define positive results.

B. The interferon-gamma test, as most commonly used, is generally accepted to have a higher sensitivity than the SCITT and the ability to detect infected animals that are SICTT negative. However the test does not have a sufficiently high specificity to allow its use as a general screening tool. The application of the IFN-g test in its current format is likely to be most useful in specific disease situations for example as a serial test to TB inconclusive animals in order to better define their disease status.

C. The use of tuberculins in the interferon-gamma test can lead to apparent false positive reactions. The use of TB complex specific antigens is likely to lead to greater test specificity albeit at the cost of reduced sensitivity. However this may be an acceptable trade-off and analysis of accumulated data comparing conventional interferon-gamma tests using PPDs and specific antigens is required.

D. It is unclear which method for calculating positive results from the IFN-g test is optimal (net OD or S/P ratio) and what the optimal test criteria for defining a positive result is. Analysis is urgently required to identify the most appropriate methods and disease circumstances for their application.

E. The on-going discovery of newer antigenic targets such as Rv3615c and antigenic combinations continues to offer opportunities to improve the performance of the interferon-gamma test and should be kept under review.

Antibody Tests

A. Antibody tests are generally regarded as having a low sensitivity, although potentially they could have high specificities, depending upon the antigenic targets used. It is possible that antibody tests could be of value in identifying animals with more advanced disease and therefore more likely to be infectious. Similarly it is possible that anergic (non-responsive) animals to cell mediated immune response
(CMI) based tests such as the ‘skin test’ and the ‘interferon-gamma test’ could have measurable antibody responses.

B. It seems unlikely antibody based tests could be used to identify all or most infected animals. However they may have significant value as ancillary tests to identify important sub-groups of infectious animals. A general principle has been proposed where approximately 20% of the host population contributes 80% of the net transmission potential of a very wide variety of diseases (the ‘20/80’ rule), i.e. a small proportion of a population having a disproportionate effect on the overall transmission of disease. It is possible that antibody tests could have a role in identifying some of this sub-population of infected animals.

C. It is uncertain what effect the intradermal skin test may have on the serological (antibody) response in infected cattle. As part of any antibody test evaluation an assessment of prior skin tests on antibody responses to bovine tuberculosis will need to be carried out to investigate if skin testing influences subsequent ancillary tests either by reducing their specificity or increasing their sensitivity.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Section Heading</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Executive Summary</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>Objectives</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>Methods</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Introduction</td>
<td>8</td>
</tr>
<tr>
<td>2.1</td>
<td>Diagnostic Test Characteristics</td>
<td>8</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Sensitivity/Specificity</td>
<td>9</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Predictive Values</td>
<td>9</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Reproducibility</td>
<td>10</td>
</tr>
<tr>
<td>2.1.4</td>
<td>Ease of Implementation</td>
<td>10</td>
</tr>
<tr>
<td>2.1.5</td>
<td>Test Confounders</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Direct Diagnostics</td>
<td>11</td>
</tr>
<tr>
<td>3.1</td>
<td>Methods Currently Available</td>
<td>11</td>
</tr>
<tr>
<td>3.2</td>
<td>Future Developments</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>Indirect Diagnostics</td>
<td>15</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary of the Cellular Immune Response to Mycobacteria</td>
<td>15</td>
</tr>
<tr>
<td>4.2</td>
<td>Cell Mediated Immunology Based Tests</td>
<td>16</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Skin Test</td>
<td>16</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Historical Context</td>
<td>17</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Critical Reagents</td>
<td>18</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Intradermal Test Formats</td>
<td>20</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Performance of the SCITT</td>
<td>20</td>
</tr>
<tr>
<td>4.2.6</td>
<td>Developments of the Intradermal Skin Test</td>
<td>22</td>
</tr>
<tr>
<td>4.3</td>
<td>Interferon gamma tests</td>
<td>24</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Historical Context</td>
<td>24</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Performance of the Interferon-gamma test</td>
<td>24</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Developments of the Interferon-gamma test</td>
<td>28</td>
</tr>
<tr>
<td>4.4</td>
<td>Antibody Tests</td>
<td>30</td>
</tr>
<tr>
<td>4.5</td>
<td>Other Disease Markers</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>References</td>
<td>34</td>
</tr>
<tr>
<td>6</td>
<td>Appendix 1</td>
<td>44</td>
</tr>
</tbody>
</table>
Objectives of this review:

A review of the published work or work nearing completion to inform an understanding of which tests for use in cattle provide the greatest sensitivity and specificity, and in which circumstances, and how this can be applied in order to improve TB control in NI per se and, also, to best inform the most practical and cost-effective deployment of test resources in what may become a more constrained economic environment.

[DARD Animal Health and Welfare Branch 2010]

Methods:
This review was written after an extensive review of the available scientific literature. On-line resources (PubMed and Web of Science), were used to find appropriate peer-reviewed literature.

PubMed (http://www.ncbi.nlm.nih.gov/pubmed) comprises more than 20 million citations for biomedical literature from MEDLINE, life science journals, and online books.

The DEFRA web-pages on bovine tuberculosis including published research projects and the Final Report of the Independent Scientific Group (ISG, 2007); ‘Bovine TB: the Scientific Evidence’ were referenced throughout.

The following relevant areas have been discussed:

- The characteristics of tests and the factors that influence their applicability and value.
- Diagnostics based upon the direct detection of infection.
- The value of the intradermal skin test including its historical context and its different formats.
- Blood based tests including the interferon-gamma test and the currently available antibody tests.
- Other technologies that are currently in development and an assessment of their current status.
2. Introduction

The genus Mycobacterium comprises over 100 species. The majority of these are found in the environment and are not normally associated with disease in humans or animals. However a small number are pathogenic. Within this pathogenic sub-group is the so-called *M. tuberculosis* complex which comprises *M. tuberculosis, M. africanum, M. canettii, M. pinnipedii, M. microti, M. bovis* and *M. bovis* subspecies *caprae*. These are generally regarded as host adapted but with the ability to spill over into other species. So in the case of *M. bovis*, the primary cause of bovine tuberculosis, it appears to be strongly cattle adapted but is able to infect humans. However it does not appear to easily transmit between people. A fuller treatment of the phylogenetic structure of the *M. tuberculosis* complex is described in the associated Literature review to this one, “A Review of Cattle – Cattle Transmission, Risk Factors and Susceptibility.”

For convenience the methods for the diagnosis of tuberculosis in cattle are examined in two broad categories namely direct and indirect tests. For the purposes of this review direct tests are those designed to directly identify the organism in the host animal. Primarily this relates to the post mortem examination of animals and the associated tests used to confirm infection. Indirect tests are those that identify infection in the live animal using indirect indicators of infection and they constitute the main focus of this review. This category of test is dominated by those based upon immunological markers, i.e. those that identify and measure immune responses in the animal to the *M. bovis* organism. They fall into 2 categories, those that are based upon the cellular immune response, primarily the intradermal skin test and the interferon-gamma test, and those based upon antibody responses. Of course there is considerable overlap in how the various tests are or could be implemented. For example most post mortem histopathology and culture is carried out on animals found infected using an immunological test.

2.1 Diagnostic Test Characteristics

The usefulness or otherwise of a diagnostic test is heavily dependent upon the expectations placed upon the test. For example a diagnostic test may be required to identify all infected animals or all infectious animals or all animals exposed to the pathogen. In most cases the tests are required to identify all infected animals. However it can be argued that only those animals infectious or likely to become infectious need be identified by current diagnostic tests in order to control the disease at the national level over time. For example an animal infected with bovine tuberculosis but unlikely to become infectious, either because it will be slaughtered at a young age or it is able to control the infection, will not, in the long term, contribute to the maintenance of the disease at the national level. This may be particularly important in the case of tuberculosis. In the case of human TB most individuals infected will not progress to active disease. Given the apparent dynamics of bovine TB in herds it appears not to be a highly infectious pathogen with only some animals apparently progressing to advancing disease and therefore presumably an infectious state. While it may be more straightforward to measure a test by its ability to identify all infected animals this may not be the optimal approach and may inadvertently lead to the rejection of certain tests with apparent poor
sensitivities but which may nonetheless be useful for the detection of important sub-groups of infected animals.

Compounding this is the possibility of intrinsic and extrinsic factors that might influence disease outcome, e.g., host/pathogen genotype, co-infection with other pathogens or nutrition. Addressing these issues is a necessary pre-requisite before anything meaningful can be said about diagnostic test sensitivity or specificity i.e., how sensitive a test is at detecting an animal infected compared to exposed is likely to vary as is the specificity of the test, i.e., how good is the test at telling if an animal is infected/infectious/exposed compared to the background population.

In the case of bovine tuberculosis it is still not clear if all animals exposed to the pathogen become infected and of those that do how many progress to becoming infectious.

2.1.1 Sensitivity/Specificity
Sensitivity (Se) can be defined as the probability that a truly infected animal will test positive. Specificity (Sp) can be defined as the probability that an uninfected animal will test negative (Banoo et al., 2008). Normally both are measured against a reference standard test sometimes referred to as a ‘Gold Standard’ i.e. a test with a high level of both sensitivity and specificity. However, where the gold standard is deficient i.e. where it fails to have perfect sensitivity and specificity, the estimation of Sp/Se of a novel test becomes increasingly challenging. This is seen in TB where the ‘Gold Standard’ is normally taken to be culture. In this case the specificity can be assumed to be very high (>99%) but the sensitivity is relatively poor. This is due to a combination of inappropriate/insufficient material submitted for culture and the inherent failure of culture to isolate the organism on occasion particularly where there are low levels of the organism. It should be noted that in all estimates of sensitivity and specificity the credible intervals for the estimates should be taken into account as well as the central figure. For example in a recent meta-analysis of bovine TB test performances the single comparative intradermal test had an estimated sensitivity of 51% with a 95% credible interval of 29-73%, (Downs, 2011) i.e. the true figure in likely to lie somewhere between these numbers, with an increasing probability that the true figure is near 51%.

2.1.2 Predictive Values
Positive predictive value is defined as the probability that an animal testing positive is truly infected. Conversely the negative predictive value is the probability that an animal testing negative is truly uninfected. Both of these measures depend upon the sensitivity and specificity of the test but also on the prevalence of the infection in a population. The larger the proportion of a population is infected the higher will be the positive predictive value and the lower the negative predictive value of the test. As the proportion of infected animals changes so too does the predictive values. An example is illustrated in appendix 1 which shows the effect of disease prevalence on the predictive values of tests with differing sensitivity and specificity characteristics. In summary they illustrate the need to have tests with high sensitivities and specificities, but most crucially where disease prevalence is low the overriding characteristic for a test to have a good predictive value is for it to have a high specificity.
2.1.3 Reproducibility
This is a measure of the percentage of times the same result is obtained when the test is repeated on the same sample. It can be used to measure factors that might be expected to have an effect on test outcome such as the effect of different operators or the effect of different batches of the same test reagents on the test outcome. It is an important measure of the robustness of a test and any validation exercise must take it into account in order to give assurance that the test does not give unacceptably high levels of variation, or if there is variation that the causes of this variation can be identified and controlled.

2.1.4 Ease of Implementation
A number of other characteristics of a test need to be considered before it can be implemented. Notably these will include the speed and complexity of the test, the equipment needed for doing the test, the operator expertise required to perform and interpret the test, and the stability of the test if it is to be used under different conditions. For example, a test shown to be excellent under trial conditions may have serious flaws when implemented if the test matrix to be used in the test must be very fresh or if the test requires a very high level of expertise held only by a few people.

2.1.5 Test Confounders
It is worth highlighting that the usefulness of the tests to identify tuberculosis infected animals is almost certainly confounded by a number of other factors such as co-infection (e.g. parasitism), masking infection (e.g. M. avium-intracellulare including M. avium paratuberculosis), nutrional stress (including production cycle stage e.g. around parturition), inadequate testing techniques and drug interactions (e.g. dexamethasone therapy). For example one early study found infected cattle failing to respond to the skin test immediately following parturition but testing positive 4-6 weeks later (Kerr, 1946a). One study in humans identified a strong association between malnutrition and helminth infection and reduced interferon-gamma responses to M. tuberculosis (Thomas et al. 2010). In cattle there is now increasing evidence that co-infection with other pathogens affects the diagnosis of bovine TB. In herds where there are cattle co-infected with Johne’s disease and bovine TB, a strongly significant reduction in the diagnostic sensitivity of the interferon-gamma test was found. For example one study found that for co-infected animals the apparent sensitivity of the interferon-gamma test in cattle bacteriologically confirmed to be co-infected with M. bovis and M. avium paratuberculosis, the cause of Johne’s disease infected, was 50% compared to 78% in M. bovis only infected cattle (Álvarez et al., 2009). Similarly work done at AFBI-VSD in collaboration with University College Dublin has found a significant effect of co-infection with Fasciola hepatica (liver fluke) and M. bovis on bovine TB diagnostics. Co-infection depressed both the interferon-gamma response and the comparative skin test response compared to M. bovis alone infected cattle (Flynn et al., 2009). In this experiment 2 out of 6 co-infected animals failed to show a comparative skin test positive response using the standard test interpretation. Interestingly other work has suggested a spatial association between liver fluke and TB in Great Britain (D. Williams pers comm). It is clear that there is an accumulating body of evidence that various extrinsic factors are likely to influence the test outcomes for bovine tuberculosis and possibly the disease transmission dynamics, all of which indicate an urgent need to identify and quantify the effects.
3 Direct Diagnostics

The diagnosis of bovine tuberculosis by clinical examination is of very limited value given that most animals infected with the bacterium do not show clinical signs of the disease and that there are no pathognomic signs of bovine tuberculosis in cattle. Instead direct diagnosis depends upon identification of the organism following post mortem examination which can be in conjunction with ante-mortem immunological tests where positive animals are slaughtered or through passive surveillance as part of routine meat inspection protocols. Broadly there are two approaches to the direct diagnosis of *Mycobacterium bovis* in cattle. They are based upon the detection of the organism either through direct culture or using molecular methods. In contrast, the direct diagnosis of TB in humans is largely based upon sputum bacteriology and microscopy with live imaging such as radiography and magnetic resonance imaging (MRI) used where available. These test approaches are not directly applicable to cattle for a variety of reasons which include the stage of disease, where human diagnosis is often directed to patients with more advanced disease, and the size and practicality of testing animals.

3.1 Methods Currently Available

Passive abattoir surveillance can be a cost effective method for surveying TB in animals (Schiller et al. 2010a). However its major drawback is a lack of sensitivity, with one estimate of 28.5%, given the normal constraints for meat inspection in abattoirs which limits the detail of post mortem examinations. This fact is important in using Post Mortem information to assess other ante-mortem tests particularly if derived from routine abattoir surveillance. For example it is well recognised that most of the conventional immunological tests, notably the intradermal skin and the interferon-gamma tests, are much more sensitive (see below). However if compared to post mortem results at face value these tests would appear to identify a large number of false positive animals. In reality most of these animals will be infected but the post mortem surveillance, typically in conjunction with bacteriology and histopathology, will have failed to identify infection (McIlroy et al., 1986). Interestingly in a study in the US estimating the sensitivity of thorough gross post mortems in known infected herds a sensitivity of 86% was estimated (Norby et al., 2004).

While typical gross pathological changes can be indicative of infection they are not definitive, and a final determination of infection status can only be reached on using confirmatory tests, most notably using bacteriology and/or molecular methods. While histopathology is frequently used as a confirmatory method to abattoir surveillance it can normally only be suggestive as lesions defined as granulomatous on histopathology can be caused by other bacteria, unless a molecular method such as in situ hybridisation is used to detect *M. bovis* DNA in tissue samples.

The polymerase chain reaction (PCR) is a technique for identifying the presence of bacteria-specific sequences of DNA. As such in principle it presents a very attractive method for identifying a range of pathogens as it is very quick, it can be easily standardised, it can detect very low quantities of DNA and it can be very specific provided the DNA target is specific to the organism being studied and the reagents used are tailored correctly to avoid false positive reactions. Its use is increasingly being applied in the field of virology, with its application now routinely used for many viral diseases.
However the biology of viruses is considerably different from bacteria with PCR techniques successfully applied to virus detection having proved unreliable when applied to mycobacteria.

A number of applications have been suggested for the use of PCR based tests. These include screening of post mortem material and analysis of environmental samples including faeces, respiratory secretions and air including exhaled air. A considerable amount of work and study has gone into developing robust DNA detection methods in both human and veterinary medicine. However to date no version of this technique is being used to detect human tuberculosis. For example in one study an optimised PCR test applied to bovine post mortem tissues resulted in a test sensitivity of 61-65% compared to conventional pathology and culture (Parra et al., 2008). Work such as this has led to conclusions such as that of a 2008 review into bovine TB which concluded that ‘the PCR technique is not yet able to perform as well as conventional bacterial culture in the detection of M. bovis in terms of sensitivity, specificity or reliability’ (Wilsmore, 2008). A more recent review drawing on preliminary results from a DEFRA funded project (SE 3231) chaired by the Chief Scientific Advisor to DEFRA concluded that ‘PCR was not a test that could be usefully used for detecting TB in badgers’ (DEFRA, 2010).

The reasons for PCR technologies not yet fulfilling their potential as diagnostic tests include problems of sampling i.e. samples with low levels of organism, samples containing inhibitors which prevent efficient PCR reactions and difficulties in extracting DNA from mycobacteria due to the organism’s robust cell wall. Additionally versions of the test which are designed to identify very low levels of DNA such as ‘nested PCR’ are very susceptible to laboratory cross contamination and require very stringent laboratory procedures to avoid false positive results.

Recently a single use closed PCR system has been developed, the Cepheid GeneXpert System (Sunnyvale California) for the diagnosis of M. tuberculosis complex infections in humans and the identification of antimicrobial multi-resistance using sputum samples (Helb et al. 2010). This test allows for considerable automation with all reagents used for bacterial disruption, nucleic acid extraction, amplification and amplicon detection inside a disposable cartridge. Results are generated within 2 hours of processing. The system has shown promising results. Amongst culture positive patients a test sensitivity was estimated of 98.2% in smear positive (microscopy) patients and 72.5% in smear negative patients (Boehme et al. 2010). The test had a sensitivity of 90.3% compared to 67.1% for microscopy in culture positive patients (Boehme et al. 2011). While the test has only been used in humans on sputum samples and no studies have been undertaken in cattle to date these results at least suggest that the possibility of this type of technology may become a useful rapid test in the future for diagnosis of bovine TB in animal samples. However financial constraints may limit their widespread usefulness as concerns have been expressed regarding the cost of the GeneXpert system (Kranzer 2011).

In summary the current molecular tools are inadequate for the direct detection of M. bovis either from ante or post mortem samples. However molecular methods such as spoligotyping are widely used to further
characterise mycobacteria isolated by routine cultural methods and to definitively identify them.

3.2 Future Developments

Recently an alternative system for the isolation and identification of mycobacteria has been assessed using a phage assay system. Phages (bacteriophages) are viruses that infect bacteria. They tend to be specific to particular bacterial species or bacterial groups. A rapid phage assay (FastPlaque) has been available for some time to identify M. tuberculosis complex organisms in sputum samples for humans based upon the D29 mycobacteriophage. Sputum samples immobilised in an agar gel are exposed to a suspension of the phage which is able to infect and kill viable mycobacteria releasing more phage viruses. The position and number of times this has happened can be measured by using an indicator organism (M. smegmatis) which grows rapidly and forms a visible confluent ‘lawn’ of growth. Any position where a mycobacterium has been killed by the phage in the sample can then be seen, as phage released from killed mycobacteria will then kill a small area of the M. smegmatis lawn leaving a visible patch or plaque of ‘no growth.’ The main advantage of this system is that it allows for a rapid test result – typically within 2 days although it also allows a semi-quantitative measure of the amount of mycobacteria in any sample tested as measured by the number of plaques formed. However, the major disadvantage of the test is that the phage is not specific to the M. tuberculosis complex and so further confirmatory tests are required to definitively identify any organisms detected.

Recently this test has been further developed to identify Mycobacterium avium subspecies paratuberculosis, the cause of Johne’s disease, using an immunomagnetic separation stage applied to the test material to improve the overall specificity of the test (Foddai et al. 2010a). Here bacterial species-specific antibodies attached to magnetic beads are used to capture any organism in the test material. These beads are then extracted using magnets and subjected to the phage assay. The usefulness of this method has now been successfully demonstrated for bulk milk tank monitoring (Foddai et al. 2010b). A similar approach is currently being developed for isolating M. bovis from animal tissues in a joint research project involving AFBI-VSD and QUB although at the time of writing it is impossible to assess whether such an approach will be a viable alternative to culture.

Passive abattoir surveillance is based upon identifying typical pathological changes in animal tissues as part of routine meat inspection. However, as discussed above, given the time constraints and need for lesions to be plainly visible to inspectors there is ample possibility of gross lesions being missed during this process. In principle this could be addressed using some of the imaging techniques now available most notably magnetic resonance imaging (MRI). For example in one study the use MRI scanning of lungs from tuberculosis infected experimental animals was shown to have a high detection rate for lesions down to less than 1mm in diameter throughout the lungs (Kraft et al., 2004). This method allows not only the identification of pathology that might otherwise be missed but also the assessment of the size and number of lesions present. However how useful this technology might be as an ancillary aid to post mortem examination remains uncertain given the cost of equipment to carry out the imaging and the expertise to interpret images generated. Also, the pathology of tuberculosis in cattle will
be different from that found in experimental animal models. Therefore the
applicability of the method would have to first be validated in infected and
uninfected cattle to assess if the methods used could differentiate TB
pathology from other lung pathologies commonly seen in cattle.
4 Indirect Diagnostics

The majority of indirect diagnostic tests available or likely to become available are based upon measuring various parameters of the immune response to the *M. bovis* organism. Broadly they fall into two categories those based upon the cellular immune response and those based upon antibody responses. In order to set these tests in context it is necessary to give a brief summary of the typical immune responses to mycobacteria.

4.1 Summary of the Cellular Immune Response to Mycobacteria

Tuberculosis is primarily a respiratory pathogen and gains entry to the animal through inhalation. Most of the immunology of bovine TB is assumed to be largely equivalent to human TB and most of the detailed immunology is derived from mice studies. The following is a basic summary of what is believed to happen during infection. However much of this has not been directly demonstrated in bovine TB infected cattle, although it is probably reasonable to assume the general information is applicable to infection in cattle. A basic knowledge of this is helpful in understanding diagnostic tests which measure the immune response and the potential limitations these tests may have.

The immune response can be thought to fall into two broad categories, the innate and acquired responses. The innate response does not require any priming to act. It is a basic system which identifies non-self targets and acts to eliminate them through phagocytosis. So within the lung of a previously unexposed animal the organism is engulfed by dendritic cells (DCs) and macrophages. The primary function of these cells is to clear foreign material by engulfing, degrading and removing it. This initial response is largely mediated by a range of receptors (Toll-Like Receptors) on the surface of the cells which recognise foreign molecules and trigger the engulfment of the bacteria. This interaction of the organism with the cell surface receptor leads to a cascade of further cellular responses and the development of an acquired immune response. The acquired response is specific to the presence of the organism and can be categorised as a Th1 or Th2 response. In general terms the Th1 response is cellular in nature with immune cells coordinating responses to the organism and the Th2 response humoral i.e. the production of antibodies from B cells. However there is a great deal of interplay between the Th1 and Th2 responses and it would be mistaken to think of them as mutually exclusive responses with one response on while the other is off. It is however broadly recognised that for an animal to control the infection there needs to be a dominant Th1 response. If the Th2 response dominates this is typically associated with disseminated active disease.

Once engulfed by DCs or Macrophages these cells send a range of signals to alert and recruit other cells to the area of infection leading ultimately to the formation of a granuloma. At this point some of the DCs will migrate to the lung associated lymph glands where sites of infection are likely to form centres of immunological memory. In DCs or macrophages the phagosome enclosing the organism is directed by a range of intracellular signals to fuse with a cell lysosome. The lysosome contains a complex mix of enzymes and chemicals used to degrade a wide variety of substances. However it is now clear that mycobacteria have evolved mechanisms that inhibit phagosome maturation inhibiting their fusion with lysosomes, and the production of enzymes by the lysosome. In reality an arms race ensues where the
organism inhibits the degradation process and the immune system responds by producing chemical messengers to improve the efficiency of the cellular processes and the recruitment of additional cells. For example the production of interferon-gamma by neighbouring cells is able to overcome the organism driven blocking of phagosome maturation. Other secreted chemical messengers play a variety of other roles. For example Tumour Necrosis Factor – α (TNF-α) secreted by activated macrophages appears to be essential in the formation of granulomas and appears to work synergistically with IFN-γ. Other cytokines released during the infection process have opposite effects. Most notably Interleukins 4 and 13 inhibit phagosome maturation. These molecules are released probably to avoid run-away reactions to foreign material, without them the immune system would end in a constant positive feedback loop leading to disease caused by the immune system itself. Probably in most cases a stalemate is reached where immune responses are unable to eliminate infection but the infection is effectively kept largely under control.

It appears that the formation of a granuloma is a mark of a protective immuno-pathological response to the organism as it is a means of curtailing the spread of the organism to other tissues. Historically this was regarded as biologically inert with little immunological activity once a granuloma was formed and the organism ‘walled off’ from surrounding tissues. However recent work with human TB suggests the granuloma is more immunologically active with what appear to be lymphoid tissue at the periphery where a continual communication between immune cells takes place, leading to ongoing stimulation of the immune system.

Therefore any immune based test is predicated on its ability to detect some aspect of the acquired immune response. It is clear that in most cases of TB the cellular immune response dominates and so the most widely used tests measure this response. It also seems clear that in most cases there is ongoing stimulation of the immune system albeit at differing magnitudes throughout the disease process whether the animal is controlling infection through the development of granulomas or not. The type and magnitude of immune responses are likely to in some way reflect the underlying disease process. For example it might be expected that a dominant antibody response may be an indicator of uncontrolled disease.

4.2 Cell Mediated Immunology Based Tests
4.2.1 Skin Test
The intradermal skin test is one of two currently approved tests for bovine TB within the European Union. It is based upon the measurement of a delayed type hypersensitivity response to intradermally injected tuberculins. It is typically performed either in the neck or caudal tail fold. In general the skin of the neck is regarded as more sensitive. As many of the antigenic targets within PPDb are shared with other bacteria, particularly mycobacteria, many countries have adopted a comparative test approach. In these cases comparisons are made between the responses to PPDb and PPDa, with reactions characterised as positive where there is a greater response to PPDb.
4.2.2 Historical Context

Eight years after isolating the bacterial cause of TB in humans, Robert Koch produced a culture filtrate of the bacterium which he initially believed could be used as a treatment for TB naming it tuberculin. Subsequently he discovered it was ineffective. However it was clear that when injected subcutaneously into TB patients it resulted in significant clinical signs such as vomiting and fever while in uninfected individuals it led to none or very mild symptoms (Snider, 1982). This was the foundation of the tuberculin skin test. Indeed this form of test was initially used as a diagnostic tool in cattle where tuberculous cattle given a subcutaneous injection of Koch’s tuberculin developed a measurable fever whereas uninfected animals did not. Although very laborious it was used in Europe around the turn of the 20th century with significant success (Monaghan et al., 1994).

The early use of tuberculin as a skin test was predominantly in the field of veterinary rather than human medicine. Pearson, an American who worked with Koch, on returning to the US tested a herd of cattle to find approximately half of them showing tuberculin skin reactivity. On post mortem all had lesions typical of TB. In this case none of the animals were showing clinical signs of disease. This gave early evidence that the test could be applied in animals before they showed any clinical signs of the disease. The application of this test to herds known to have disease resulted in significant reductions in the disease burden of cattle. However as the test started to be applied to areas apparently free of the disease, numbers of reactor animals were found that did not apparently have any other post mortem signs of tuberculosis. The interpretation of these animals was highly contentious with some claiming these animals were truly infected while others such as Hastings working in the University of Wisconsin in 1924 claiming that there were other sensitising bacteria that the cattle were exposed to leading to a cross reaction and therefore a false positive result. This was formally demonstrated by Hastings and co-workers who demonstrated that non-tuberculosis mycobacteria recovered from reactor cattle showing no visible signs of disease were able to sensitise other cattle to tuberculin (Crawford, 1926; Hagan, 1929; Snider, 1982). A more standardised tuberculin formulation was produced in 1934 by Florence Seibert which was termed purified protein derivative (PPD).

It was later discovered that in humans the responses to PPD was dose dependent. Given sufficient dose many people would have reactions to the PPD that had no history of TB or known contact with TB. These individuals would have no reactions at lower doses. Therefore it was found in humans the specificity of the test could be improved by optimising the dose of tuberculin used. A second observation in humans was the booster effect. In these cases, patients with an initial poor response to tuberculin subsequently developed strong responses on repeated testing. In most cases this was attributable to patients with latent infection who were anergic to the skin test. The stimulus of the skin test was sufficient to elicit a measurable reaction on subsequent tests.

For bovine tuberculosis diagnosis, a repeat intradermal test was developed where tuberculin was injected into the skin of an animal and repeated at the same injection site 48 hours later with the increase in skin thickness measured 24 hours later. Variations of this were developed including the ‘Stormont technique’ where PPD was injected intradermally, and repeated at
the same place 7 days later with the subsequent skin thickness measured 24 hours later (Kerr, 1949). An increase of over 5mm between days 7 and 8 were classified as positive (Kerr, 1946a, b). Interestingly in the animals tested for these experiments it was found this technique had a greater sensitivity and specificity compared to both the repeat single intradermal and the repeat comparative intradermal skin tests as measured by post mortem findings. This test continued to be used in N. Ireland until 1958 when it was replaced by the single comparative intradermal test (Monaghan et al., 1994) following two related large scale experiments comparing the repeat and single intradermal comparative tests carried out by MAFF in which no significant increase in sensitivity or specificity could be detected using the more laborious double comparative test (Francis, 1958). This test is now in essence what is now called the single intradermal comparative skin test. The only major modification was when in 1975 tuberculin derived from \textit{M. bovis} was used instead of \textit{M. tuberculosis} due to the increased measurable potency of \textit{M. bovis} tuberculin (Lesslie et al., 1975).

\subsection*{4.2.3 Critical Reagents}

Most critical to any immunological test is the antigenic targets used to measure the immune response. Pivotal for the intradermal skin test are the purified protein derivative tuberculins (PPDs). These are a complex mixture of heat treated molecular components of the organism as well as components produced by the organism during culture (British Pharmacopoeia). In summary the target organism is cultured in a liquid medium which is then heat treated to kill the organism. The material produced is filtered, concentrated by precipitation, washed and then solubilised in sterile water. The potency of the tuberculin is measured using an animal model where guinea-pigs sensitised to the target organism are intradermally injected with the tuberculin to measure skin test reactivity. This potency testing is then used to standardise the tuberculin before its application in diagnostic tests. The organisms currently used to produce the tuberculins currently in use in N. Ireland are the AN5 \textit{M. bovis} strain and the D4ER \textit{M. avium} strain. The AN5 strain was isolated in England around 1948 (Paterson, 1948). It was used for tuberculin production in 1975. Prior to this strains of \textit{M. tuberculosis} were used but it was decided that the use of \textit{M. bovis} would yield a more specific test (Lesslie et al., 1975; Smith et al., 2006). Importantly the choice of AN5 was due to its ability to grow at high levels in the lab rather than any empirical evidence that it was an optimal strain of \textit{M. bovis} to use as measured by specificity/sensitivity estimates (Inwald et al., 2003). The spoligotype pattern for AN5 has not been seen in GB over the past 15 years and not in NI or RoI in any isolates studied since 2003. However the strain shares significant commonalities to spoligotype SB0129 leading to the suggestion that it is a lab adapted variant of this spoligotype. This spoligotype is infrequently identified in the British Isles and is genetically somewhat distant from most spoligotypes currently present (R. Skuce Pers Comm). It has been suggested that this could mean that tuberculin derived from AN5 may not be the optimal strain for current use (Smith et al., 2006). However it must be noted that most antigenic targets will be retained across all the strains of \textit{M. bovis}, any differences are likely to be small and may not be significant. Nonetheless there remains uncertainty over whether there is an optimal \textit{M bovis} strain for current TB diagnostics.
Irrespective of the *M. bovis* strain used there is now considerable evidence that different sources of PPD have potentially highly significant differences in their activity. For example one study found that different PPDs had significant differences in their ability to detect both experimentally and naturally infected cattle using both the SICCT and the IFN-γ test (Schiller et al. 2010b). Based upon the protein concentrations of the PPDs used there was a 17 fold difference in the potency of the PPDs between the best and worst performing ones. Sub-optimal PPDs had reduced sensitivity compared to the best performing ones. When used in known uninfected cattle the best performing PPDs also displayed the best specificity i.e. use of poorest PPDs also led to some false positive responses. Although PPDs are produced to standard guidelines such as those of the OIE and the European Pharmacopoeia, variations in media, culture time, growth phase for bacterial harvesting, and the treatment, precipitation and filtration steps are very likely to occur which could lead to both quantitative and qualitative variations in PPDs. For example in one study most internationally and commercially available tuberculins, if used at a dose rate of 1mg/ml, would fail to meet the minimum dose of 2000 international units (Bakker et al, 2009). Similarly the in vivo potency testing of PPDs in guinea pigs is known to be variable and so a wide potency range is acceptable for skin test reagents (Council Dir 64/432). It is also unclear whether the guinea-pig animal model is the most appropriate method for assessing PPD potency. For example Schiller et al (2010b) reported one PPD which had apparently poor potency in guinea-pigs but good potency when used in cattle. It would appear that there is a necessity to more rigorously assess the potency of PPDs. This could be through the assessment of them in artificially infected cattle. This would necessitate the use of Containment level 3 facilities to house the cattle. An alternative method could be to use BCG sensitised cattle which would not require such stringent containment facilities as it is recognised that BCG vaccinated cattle raise a strong and persistent cellular immune response measurable by intradermal testing (Buddle et al., 2008). Further to this it appears that the in vitro assessment of PPD potency using the interferon-gamma test is a reasonable measure of the potency of the tuberculin as measured by the intradermal test (Schiller et al. 2010b) suggesting that the interferon-gamma test using blood samples from infected or BCG vaccinated animals could be used as a proxy measure of the potency of PPDs for intradermal skin testing.

These observations on the variability of PPDs is of critical importance in determining how useful or otherwise any test based upon them is, where they are inadequate any test using them will also be inadequate. For example the potency of the tuberculin has a significant influence on the number of test positive animals disclosed using the single comparative intradermal skin test. A recent assessment of bovine PPD at 3 different potencies (low, normal and high) revealed that the high potency tuberculin revealed 40% more reactors than the normal potency tuberculin while the low potency one revealed 50% fewer (Good et al. 2011). Consideration should be given to assessing the usefulness of higher potency tuberculins in disclosing additional infected animals.

There is some uncertainty about the extent to which cattle are exposed to *M. avium avium* (Maa) the cause of avian TB. Laboratory isolations of the organism within N. Ireland are rare. Where infection does occur it appears not to persist and the animal is able to mount an immune response which
quickly wanes over time as measured by both the skin test and the interferon-gamma test (Barry et al., 2011). However in comparison there is considerable exposure of the N. Ireland cattle population to *M. avium* subspecies *paratuberculosis* (Map) the cause of Johne’s disease. Infection with this organism leads to persistent infection and a persistent cellular immune response which is measurable by the skin test using PPDa. It could be that while PPDa contains most of the relevant immune targets for Map it may not be the optimal reagent in a population of Map infected cattle. Studies at AFBI-VSD have suggested that tuberculins based upon Map rather than Maa may be more efficient in identifying Map infected cattle as non TB infected when used in a comparative interferon-gamma test (Barry et al., 2011).

### 4.2.4 Intradermal Test Formats

Broadly there are now two test formats used across the world, a single intradermal skin test (SIT) and the comparative test as described above. The decision which test to apply is influenced by a variety of considerations which include the prevalence of disease and the exposure of cattle to other mycobacteria. The SIT can be used as the caudal fold test in the USA, Canada and New Zealand, in some cases as a first test with subsequent comparative tests used to confirm infection. Across Europe the cervical SIT test is widely used. However early studies using the SIT in the British Isles identified a large number of non-specific reactions which effectively precluded the widespread use of this test format (Lesslie et al., 1975; Monaghan et al., 1994). Sensitivity and specificity estimates for the test have been previously reviewed (de la Rua-Domenech et al., 2006), with sensitivity ranging from 63.2-100%. In general the cervical SIT appears to be more sensitive than the caudal fold test. Specificity was estimated to range from 75.5-99.0% suggesting at least for the UK and Ireland that it could not be applied in a widespread way.

### 4.2.5 Performance of the SCITT

The intradermal skin test relies upon a delayed type hypersensitivity reaction to the injected tuberculin. This response is a complex cellular reaction composed of both specific and non-specific components, with the predominant response due to previously sensitised T cells (i.e. cells previously exposed to components of mycobacteria) which recruit and orchestrate the infiltration of other cell types into the injection area thereby leading to the development of a measurable transient swelling. This swelling, by definition, takes some time to develop but it has been shown that the maximal measurable reaction is typically 72 hours following the injection (Lepper, 1977b) (Francis, 1978). Importantly there appears to be a strong association between the degree of disease progression and the skin test response. Those cattle with large skin responses to tuberculin are much more likely to have visible pathology at post mortem than those with more moderate skin responses (Clifton-Hadley, 2005).

While it is generally accepted that recently infected cattle fail to respond to the intradermal skin test (Monaghan et al., 1994) there is uncertainty how long this period lasts. Experimental work where cattle infected with moderate infective doses of *M. bovis* tested positive within 3 weeks of infection suggests it may not be very long (Thom et al., 2006). At the other extreme it has been long recognised that cattle with advanced disease may
become non-responsive to the test, so-called anergic animals (Lepper, 1977a) (Pollock and Neill, 2002).

In practice herds with disclosed reactor animals are re-tested in order to identify newly infected animals and infected animals not detected at the previous test. However it appears that the interval between repeated tests on animals may have an important effect on the sensitivity of the test. This has been described as desensitisation. Tested infected cattle can fail to react if the test is repeated shortly afterwards (Radunz, 1985). This appears to be most intense in the week immediately following the test (Doherty et al., 1995). Earlier studies suggested this desensitisation has largely subsided after 60 days (Thom et al., 2006) (Radunz, 1985). However more recent work has suggested that repeated intradermal testing at 60 day intervals does have a significant effect (Coad et al., 2009). This observation may be most important for animals testing inconclusive and undergoing repeated tests. These animals are likely to have marginal skin responses which repeated testing may depress sufficiently to prevent their subsequent disclosure. Interestingly in this and a previous study (Doherty et al., 1995) this depressing effect was not seen with the interferon-gamma test, lending evidence that this test may be the test of choice in this category of animal i.e. as a serial test.

Estimating the sensitivity of the SCITT test has proven to be very challenging. Any estimate will be based upon the animals it was estimated on which may not necessarily reflect the actual population of animals the test will be used on. For example an estimation of the test sensitivity in one country may not reflect the situation in a different one (different confounders, different genetic make up, different husbandry etc.). To accurately assess sensitivity all animals tested should be slaughtered and subjected to a thorough and accurate test to determine the actual disease status of each animal. In reality rarely have all tested animals been slaughtered and the post mortem examinations thorough enough to have a high level of certainty of the actual disease status of the animals under study. Many studies have tended to be small, selecting animals from high prevalence herds (de la Rua-Domenech et al., 2006). Estimates have ranged from 55.1% to 90.9% (Costello et al., 1997; Neill et al., 1994) using the standard skin interpretation. Sensitivity increases when the interpretation is severe but at the expense of specificity. Similarly where the non-comparative test is used (single intradermal test) studies have found this to have in general higher sensitivities but at the expense of reduced specificities (de la Rua-Domenech et al., 2006a). For example one report suggested that the sensitivity of the test could be increased from 83% to 93% if the test interpretation is changed from standard to severe (de la Rua-Domenech et al., 2006b). More recently modelling work has estimated the sensitivity (Se) and specificity (Sp) ranges for standard and severe interpretations of the SCITT to be 58.6%-67.5% (Se Standard), 65.9-77.3% (Se Severe) 99.2-99.8% (Sp Standard) and 98.9-99.7% (Sp Severe) (Clegg et al. 2011). This compares to other unpublished work which suggests that standard interpretation has a Se of 53-59% and a specificity of >99.9% and severe interpretation has a Se of 67-74% and a Sp of 99.8-100% (Skuce pers comm).

It should be noted however that while most estimates of sensitivity are made at the animal level, the herd level sensitivity will be inherently higher even
where the within herd prevalence is low (de la Rua-Domenech et al., 2006). This opens up the possibility of employing ancillary tests with greater individual animal sensitivity in herds where TB has been diagnosed using the SCITT even if the ancillary test has a reduced specificity. This could include for example the use of the single intradermal skin test.

As described above there is an inverse relationship between the specificity of a test and its sensitivity. Where the SCITT has been applied to uninfected populations (i.e. to assess the number of false positive reactions) the specificity estimates range from 94% to 100% using the standard test interpretation (Buddle et al., 2001; Francis, 1978; Lesslie et al., 1975; Neill, 1994; O’Reilly, 1975). When looking at putatively non-infected animals Goodchild & Clifton-Hadley estimated test specificity as 99.99%.

A recent meta-analysis of current diagnostic tests was published to estimate using published data TB diagnostic test performance. Of 9782 references found only 261 met the criteria of the study for including in the analysis. The analysis suggested the SCITT to have a very high specificity (>99%) and a sensitivity of approximately 51%. In comparison the single intradermal test had a sensitivity of 93% but a specificity of approximately 92% (Downs, 2011).

Given the inherent difficulties, using conventional methods, of reaching accurate estimates of the sensitivity and specificity of diagnostic tests, novel statistical approaches are increasingly being deployed most notably Bayesian methods such as latent class modelling for estimating sensitivities and specificities. These methods allow the inference from observed variables the structure of more fundamental variables which cannot directly be accurately measured. For example they can be used to take into account the inaccuracies of several test results as part of the process of making estimations of basic information about a disease where no perfect test is available. In a similar way it can be used to estimate true test characteristics where there is no gold standard available to measure a test against. While there is some controversy about the usefulness of this approach, for example it can be sensitive to prior assumptions which can be mistaken, its use is being increasingly applied to understanding diagnostic test accuracy where there are no adequate ‘gold standard’ methods available to measure data against, particularly where newer methodologies are being developed which allow for more robust estimations (Xu and Craig, 2009). In human TB it has been applied to a variety of current diagnostic tests which estimated a sensitivity/specificity of 99.9%/64.2% for the skin test and 76.3%/93.6% for the quantiferon (interferon-gamma) test (Girardi et al., 2009). Interestingly initial analysis using this approach has resulted in estimated test attributes for bovine TB broadly very similar to previous estimates using conventional methods, with estimates of test sensitivity for the SCITT using standard interpretation of 56% and test specificity of greater than 99.9% for (Skuce pers comm.).

4.2.6 Developments of the Intradermal Skin Test
Attempts have been made to improve the current SCITT. Most notable has been the use of more defined antigens. For example in one study the use of ESAT-6 resulted in an increased specificity in the skin test when compared to PPD based tests (Pollock et al., 2003). However it also led to a decrease in
sensitivity. The best results could only be achieved using high concentrations of the recombinant protein making the test very expensive. This response could in principle be augmented if an adjuvant (which potentiates immunological responses) was included. However this approach carries with it the risk of inducing sensitisation to ESAT-6 which would be counterproductive as it could lead to false positive reactions in subsequently tested animals (Whelan et al., 2003). That ESAT-6 induces reduced responses compared to PPDs is almost certainly a consequence of the large range of antigenic targets within PPDs. In response to this further work using a combination of recombinant proteins (ESAT-6, CFP-10, Rv3615c and MPB83) have suggested their use had a comparable sensitivity to PPD based tests on naturally infected cattle (Whelan et al., 2009) which could be further enhanced using overlapping peptides of the same antigens. Other variations on this approach are being developed elsewhere with initial promising results (Schiller et al. 2009). Importantly, given the inherent specificity associated with using specific antigens, if the use of defined antigens is proven to have equivalent sensitivity to the skin test in other large scale trials this approach would allow for a more standardised approach to intradermal skin testing as the reagents would be much more characterised than PPDs and their use would obviate the need for a comparative test.

The current method for reading skin test results remains crude and somewhat subjective relying upon the application of callipers to measure skin thickness. One alternative approach is to use infrared thermography which measures the local temperature of each injection site. An initial evaluation of this technology on a small group of animals suggested an improved classification of sensitised animals compared to conventional skin measurements (Johnson & Dunbar, 2008) and that the reading of results could be undertaken 24 hours following injection. However the test would need considerable further evaluation in naturally and experimentally infected animals before its use could be recommended. In principle other approaches to the intradermal test could be explored, for example using needle free transcutaneous injections and ultrasonic skin thickness measurements although at this stage these would need to be proven to work in principle before any field evaluations could be done.
4.3 Interferon gamma tests

The interferon-gamma test is an assay commonly applied as an ancillary test for the ante-mortem diagnosis of bovine TB. It measures the cellular response to mycobacterial antigens and as such broadly measures the same immune response as the intradermal tests. Briefly blood samples are incubated with mycobacterial antigens. Immune cells present in the blood respond to the antigens releasing a cascade of chemical signals, most notably interferon-gamma. A positive result is seen where there is a preferential release of IFN-g to constituents of *M. bovis* compared to other mycobacteria (Wood and Jones, 2001a). IFN-g is regarded as a pivotal cytokine released during the immune response to TB. It is produced by a wide variety of cell types but primarily by T cells which on exposure to mycobacterial antigens release IFN-g which leads to a cascade of immune responses – most notably the activation of macrophages. Therefore the assay is a measure of the responsiveness of T cells to the organism which is part of the acquired immune responses i.e. the presence of reactive T cells implies exposure to the organism.

4.3.1 Historical Context

Interferon-gamma is a pivotal chemical messenger involved in orchestrating a wide variety of immune responses and the immune cells involved in those responses. Its major role is in activating macrophages and therefore in the immune system’s attempt to clear intracellular pathogens such as mycobacteria. The interferon-gamma assay was developed in Australia in the late 1980s although its development and validation was largely after the bulk of bovine TB was eradicated in Australia (Wood et al., 1991). It is now an OIE listed test for international trade and approved under EU directive 64/432 annex B as an ancillary test for bovine TB. It is now widely used throughout the world either as a serial test to the intradermal test (usually the SIT), where its use is to enhance overall disease detection specificity or in parallel with the intradermal test where its use is to increase overall disease detection sensitivity (Pollock et al., 2005). It is currently marketed as Bovigam® (Prionics, Switzerland).

4.3.2 Performance of the Interferon-gamma Test

In summary the test measures the release of interferon-gamma (IFN-g) from sensitised blood lymphocytes exposed to mycobacterial antigens using a sandwich ELISA. The antigens currently used in the N. Ireland program are purified protein derivatives from *M. avium* (PFDa) and *M. bovis* (PPDb) and ESAT-6, a putative TB complex specific antigen.

The laboratory only defines a positive test when a sample shows a detectable level of IFN-γ above a background and where there is evidence of tuberculosis complex specific cell mediated immune responses. The cut-off (the degree to which there is a biased response to bovine TB) currently used in N. Ireland is comparatively low, i.e. a small bias triggers a positive result. This was set by DARD to maximise the sensitivity of the test (Anonymous, 2006). However lower cut-offs inevitably lead to reduced test specificity. There may be circumstances where a higher cut-off may be more appropriate in order to maximise the specificity of the test for example if it is to be used more extensively although the cut-off should be tailored to the situation the test is being used in (for example where higher sensitivity is required and a reduced specificity acceptable such as in testing SCITT inconclusive or high risk animals).
Fundamentally it has a number of advantages over the intradermal tests. It has better sensitivity (where standard SCITT interpretation is applied), it can be repeated on the same animal immediately, animals need only be handled once, and the test interpretation can be regarded as less subjective. However there are also a number of significant limitations compared to the intradermal tests. These include lower specificity, particularly in young animals, the need for rapid laboratory processing following blood sampling, substantial laboratory expertise and significant test kit costs.

A measurable interferon-gamma response can be detected as early as 3-5 weeks following infection even when the level of exposure is very modest (Dean et al., 2005). This response does not appear to be dose dependent with those calves exposed to less than 10 colony forming units and which develop pathology developing responses as quickly as calves given much higher doses.

A number of studies have attempted to quantify the diagnostic performance of the IFN-g test in cattle. The specificity estimates range from 85% (Buddle et al., 2001) to 100% (Lilenbaum et al cited in de la Rua-Domensch 2006). A test specificity of 99.6% was estimated in N. Ireland (Neill et al., 1994). However care needs to be exercised in making comparisons between studies as there have tended to be a wide variation in the criteria used to define positive animals, the numbers of animals used in the studies, the populations the animals were selected from and differences in how blood samples were handled following collection. Nonetheless the broad sweep of evidence points to the interferon-gamma test as having an overall poorer specificity than the single comparative intradermal test. In a recent meta-analysis of studies a specificity estimate for the test of 96% was reached (Downs et al., 2011).

The most significant improvement in specificity of the test is likely to come from the use of more defined antigens, most notably the early secretory antigenic target (ESAT-6) and the culture filtrate protein (CFP-10). Both of these are regarded as largely specific to the Mycobacterium tuberculosis complex that includes M. bovis, M. africanum and M. tuberculosis. They are noted in their absence from M. avium and BCG M. bovis, although they are conserved in M. kansasii (Veyrier et al. 2011). Therefore with the noteworthy exception of M. kansasii their use would be expected to significantly improve the specificity of the test (Aagaard et al., 2006; Buddle et al., 2001; Pollock et al., 2000). However disappointingly a recent study in GB suggests the improvement in specificity may not be sufficient to allow its widespread use (Vordermeier cited in Schiller et al 2010) although to the author’s knowledge this work has not yet been published. Interestingly in some recent work carried out at AFBI, cattle experimentally infected with Mycobacterium avium subsp. paratuberculosis (Map) showed, on a small number of occasions, a biased response towards PPDb, which would have categorised them as reactors using current cut-offs. This lack of specificity would probably have been overcome had more defined antigens been used (Barry et al., 2011).

Estimates of test sensitivity range from 73% (Lilenbaum) through to 100% (Monaghan et al., 1997). A test sensitivity of 84.3% was estimated in N. Ireland (Neill et al., 1994). As with specificity estimates, care needs to be exercised in the interpretation of these figures. However the overall sense of
these studies is that the test does have a greater sensitivity to detect infected cattle than the SCITT. In a meta-analysis of studies a sensitivity estimate of 67% was made (compared to 51% for the SCITT using standard interpretation) (Downs et al., 2011). This is borne out by other studies where the IFN-g test was able to identify animals with an increased risk of infection (for example animals in-contact with infected animals) which the SCITT was unable to detect (Coad et al., 2008; Gormley et al., 2006). This may, at least in part, be due to the ability of the test to detect infected animals at an earlier stage of infection (Pollock et al., 2005). These animals are much more likely to subsequently become SCITT reactors when compared to interferon – gamma test negative animals. While most research has identified the use of more defined antigens as leading to a reduction in the sensitivity of the test (Buddle et al., 2003; Pollock et al., 2000; Vordermeier, 2006), there are at least theoretical situations where the use of defined antigens might increase sensitivity (Schiller et al. 2010a). For example in one experiment it was observed that during the initial days following infection there was a biased response to PPDa against PPDb (Schiller et al. 2010b). In principle this could be overcome by using more defined antigens. Similarly it is recognised that in animals co-infected with *Mycobacterium avium subsp. paratuberculosis* (Map) the conventional use of PPDs in the test fails to identify a significant number of *M. bovis* infected cattle (Álvarez et al., 2009). In order to improve the overall sensitivity of a test using defined antigens additional antigens have been assessed for their potential usefulness as diagnostic reagents. For example while ESAT-6 and CFP-10 remain the outstanding specific reagents for this test adding additional antigens have given additional diagnostic benefit to the test. In one study adding 4 antigens which are present in the *M. bovis* genome but absent from BCG improved the sensitivity of the test from 91.4% of infected cattle to 96.6% at a test specificity of 93% (Cockle et al., 2006). More recently the antigen Rv3615c was found to identify additional *M. bovis* infected cattle that ESAT-6/CFP-10 did not suggesting that this *M. tuberculosis* complex specific antigen could be used in addition to the currently used specific antigens to increase sensitivity without compromising specificity (Sidders et al., 2008). A similar effect was seen when the outer membrane protein (OmpATb) was used. Again the use of this antigen identified infected cattle which ESAT-6 and CFP-10 did not (Schiller et al., 2009). There appears therefore to be a significant need to assess these newer cocktails of antigens in the field in order to identify if their use could allow the more widespread application of the test within the N. Ireland TB eradication program.

There is some controversy over the effect of prior intradermal skin testing on subsequent IFN-g testing in cattle. Early work suggested there might be a transient depression in the IFN-g responses of cattle following single intradermal testing (SIT) (Rothel et al., 1992) although this was only seen in 2 out of 4 infected cattle. A similar effect was seen in experimentally infected cattle tested 3 days following a SCITT (Whelan et al., 2004). This is in contrast to a similar study carried out by the same group where naturally infected cattle did not show any depression in the IFN-g responses 3 days following skin test (Coad et al., 2007). In this latter study there was a trend towards increased IFN-g responsiveness. Another study found that intradermal testing actually boosted the IFN-g responses of blood samples for several weeks (Whipple et al., 2001) and did not appear to have any detrimental effect on test result. This observation is supported by work
carried out in New Zealand with naturally infected cattle which concluded that the IFN-g test could be reliably used 8 days following the SIT reading without any loss in diagnostic sensitivity or specificity (Ryan et al., 2000). Similarly studies in both UK and Ireland failed to find any detrimental effect of repeated SCITT on the diagnostic value of the IFN-g test when applied to infected cattle (Doherty et al., 1995; Gormley et al., 2004; Thom et al., 2004).

It is well recognised that the magnitude of the IFN-g response following blood stimulation with antigens is closely related to the length of time and conditions blood is stored following collection. Several studies have attempted to evaluate the effect of time between blood sample collection and blood testing with a variety of contradictory conclusions drawn. Some suggest a delay in blood processing has no effect on test sensitivity while others that it reduces sensitivity. Similarly there appears limited agreement on the effect of blood storage on test specificity (Vordermeier et al., 2006). The original recommendation was that samples should be processed within 8 hours of sampling (Rothel et al., 1992). This has been supported by other studies which have shown marked reductions in the responsiveness of blood cells to antigens over time. For example in one Irish study only about 50% of animals that were SCITT positive and IFN-g positive using 8 hour old blood samples remained IFN-g positive using blood samples 24 hours old (Gormley et al., 2004). However in the same study no similar reduction in test sensitivity was identified using animals with confirmed tuberculosis at post mortem. When the more specific antigens ESAT-6 and CFP-10 were used in assays comparing 8 hours and 24 hours blood storage the longer storage period resulted in a reduction in the predictive value of the test which is probably a reflection of the overall lower IFN-g responses to specific antigens compared to PPDs (Whelan et al., 2004). In work carried out at the Veterinary Laboratories Agency no effect of time was seen in the sensitivity of the test to detect experimentally infected cattle but there was a modest decrease in sensitivity to detect field reactors, although with this category of animal there was an increase in specificity (Vordermeier et al., 2006). Whether storing bloods for longer before processing has a detrimental effect on the test outcome may reflect the stage of disease being detected with blood samples from more advanced disease, and therefore animals with stronger IFN-g responses, more resistant to the inevitable reduction in IFN-g production following antigenic stimulation. Therefore it would appear that there may be scope for a study to identify those disease situations where the application of a test version incorporating a 24 hour delay in sample processing could be applied for example in situations where there is suspicion of animals with more advanced disease which the SCITT is failing to detect. In addition it is noteworthy that blood collected post-mortem is a poorer matrix for the test leading to reduced test sensitivity (Rothel et al., 1992).

Clearly it is not feasible to deploy the IFN-g test as it is currently used with its current estimated specificity of approximately 96% in any widespread TB eradication program particularly where there is a relatively low animal disease prevalence as its use will invariably lead to the identification of an unacceptably high number of false positive animals. Its most useful deployment is likely to be as an ancillary parallel test to the SCITT in clearly defined circumstances where removal of a small number of ‘false positives’ is acceptable for example where there is likely to be recently infected animals (for example in a new explosive infection), in better defining inconclusive
reactor animals (Monaghan et al., 1997; Neill et al., 1994), and in pre­
movement testing of cattle from reactor herds or those deemed to have a
high risk of infection. An additional use has been as a serial test of SIT
reactors in herds or areas which are TB free (Wood and Jones, 2001b), i.e.,
where there is doubt based on epidemiological evidence of the validity of the
skin test result. The usefulness of this approach could be further refined by
increasing the specificity of the antigens used (Buddle et al., 2001). New
Zealand affords an example of the various applications that the test can be
used for (Anonymous, 2005). Here it is agreed that the test is used as a SIT
ancillary serial test in 3 distinct circumstances dependent upon the farm
risk factors identified. They are as an additional test to the Caudal Fold
Skin (CFT) test where the standard interpretation is applied, as a serial test
to the CFT where there is good reason to assume the CFT result is a false
positive (in this case positive cut­off is set high to increase specificity), and
the use of ESAT-6/CFP10 as specific antigens (where there are possible false
positive CF test results in areas assumed to be TB free). There is also the
option to use the test in parallel to the CF test in circumstances where the
CFT result is negative where there is a high risk of infection.

To date there has been no comprehensive study into the cost effectiveness of
the interferon-gamma test either used on its own or in combination with the
intradermal tests. One study concluded that for the test to be cost-effective
its use would need to demonstrate a significant reduction in the within and
between herd infection transmission rates (Anonymous, 2002). To maximise
the impact of the test on disease outcomes this study recommended that it
was inappropriate to use the test in a wholesale way in multiple reactor
herds but rather in alternative targeted approaches. As part of the approach
to this it was recommended that a statistical modelling approach be adopted to
identify the optimal situations for the application of the test as well as
attempts to reduce the overall cost of the test through for example the use of
a reduced panel of antigens. To the authors knowledge however, a thorough
study has not been completed to assess the usefulness and cost-
effectiveness or otherwise of the test in controlling the disease (Vordermeier
et al., 2005).

4.3.3 Developments of the Interferon-gamma Test
A number of constraints which limit the usefulness of the IFN-g test have
been recognised and they include sub-optimal specificity, high logistical
costs due to the assay needing rapid transport of blood sample to the
processing lab and the cost of the test kit. Further refinements of the test
are urgently needed to allow its fuller use in the control of bovine TB. For
example it is at least theoretically possible to develop an ‘in-tube’ stimulation
where blood samples are mixed and incubated with antigen at the point of
collection from the animal (Schiller et al. 2010a). Similarly there are in
principle methods which might allow the maintenance of cell viability
between sample collection and processing which could allow an extension of
the transit time for blood samples.

The cost of the test could be reduced if the patent rights to the test expire
leading to an increase in commercial competition in kit production as well as
opening up the possibility of laboratories developing their own ‘in-house’
methods. Important would be the automation of some of the assay test
steps thereby reducing the personnel costs associated with the test and
increasing test throughput. This is likely to require further test development to ensure the test is sufficiently stable to allow this (M. Welsh pers comm.).

Currently the test result is based upon a net optical density (OD) value. This is calculated by subtracting the reactivity to PPDa from PPDb. Where there is a larger response to PPDb this is regarded as a positive test. Reducing the degree of PPDb bias needed to categorise a test as positive increases the test sensitivity at the expense of specificity whereas increasing it will do the reverse. The appropriate cut off may vary depending upon a range of factors which should be modelled in order to optimise the test performance as well as identify the circumstances it is best suited for use in. In addition alternative test readouts, for example sample to positive control ratios (S/P ratios) which are widely used in immunological assays for other diseases, may be a more appropriate method for measuring test results.

It is very likely that the test will be improved by the further use of mixtures of *M. tuberculosis* complex specific antigens. The primary value of this will be to increase test specificity as in principle measurable responses to these antigens should be largely TB complex specific. ESAT-6 and CFP-10 remain the most useful specific antigens. However their use leads to a significant reduction in test sensitivity. Recent work has suggested that the addition of other antigens could significantly increase the sensitivity of the test (Schiller et al., 2009; Sidders et al., 2008). However these now need to be evaluated in the field situation. It is important that this should be carried out in N. Ireland as evaluations in other countries may not be directly applicable particularly in relation to the disease prevalence and the background exposure of cattle to other mycobacteria.

It is possible that other immune signalling molecules released following blood stimulation with antigen may be useful either instead of or as an adjunct to interferon-gamma. One promising candidate is the monocyte derived chemokine IP-10. In one study with humans it was found to have a comparable accuracy to the IFN-g test. In patients with active TB it had a reported sensitivity of 81% compared to 83% for the IFN-g test, although the test specificity was somewhat less (97% compared to 100% in this study) (Ruhwald et al. 2011). It has been suggested that this is a better target for testing latently infected people or HIV co-infected patients than IFN-g. Although, to the authors knowledge, this target has not been evaluated in cattle it is at least feasible that it or other similar immune signalling molecules could be better targets or useful as additional targets for future assays of this type.
4.4 Antibody Tests

Antibody responses to *Mycobacteria tuberculosis* complex organisms are generally regarded as muted in most animals with the predominant immune response being cellular. Classically following infection the animal mounts a cell mediated immune response which can with time wane to be replaced with an antibody response (de la Rua-Domenech et al., 2006; Pollock and Neill, 2002; Welsh et al., 2005). In some animals this shift has been associated with anergic animals which though diseased fail to respond to the normal diagnostic tests based on CMI responses (skin test and IFN-γ). Some of these anergic animals are detectable using antibody tests (Yearsley et al., 1998). However it remains unclear how many such animals exist given the intensive and regular TB testing carried out within N. Ireland. Importantly however the switch from cellular to antibody responses has been related to an advanced disease profile (Welsh et al., 2005). Indeed the antibody response may act as a marker for advancing disease or it may actually have a role in the development of disease by down-regulating cell mediated control mechanisms (Hussain et al., 2001). Therefore antibody responses may be a marker for more advanced disease and therefore of animals more likely to be infectious.

Antibody tests hold many advantages over the conventional CMI based tests. They only require animals to be handled once to obtain blood samples which are often taken for other disease monitoring purposes allowing less frequent animal handling and less time taken by sampler/testers in the field. They are usually rapid allowing a short result turnaround time, are typically easily scalable so that large numbers of samples can be processed quickly (for example much of the sample handling can be undertaken by robots) and usually the consumable costs per sample are modest. Taken together the overall cost per test is likely to be much less than that for CMI based tests.

At any single time-point most infected animals will not be displaying a measurable specific antibody response (Fifis et al., 1992; Plackett et al., 1989). There is some evidence that the SCITT performed on recently infected cattle can lead to a measurably increased response (Thom et al., 2004; Waters et al., 2006) thereby potentially increasing the sensitivity of antibody testing. However other researchers have found some evidence for non-specific antibody responses in non-infected cattle following intradermal skin testing (D. Bakker pers comm.). Therefore there remains uncertainty about the usefulness of antibody tests in cattle populations routinely subjected to skin tests. This remains a significant knowledge gap and it would appear urgent that a carefully constructed field trial is undertaken to assess the usefulness of current antibody tests and their effect on long term disease outcomes at the animal and herd levels in an animal population subjected to routine comparative intradermal skin testing (Graham et al., 2010).

A wide variety of antibody tests have been trialled. However it is difficult to make comparisons between them as many have used different combinations of antigenic targets and different platforms for measuring antibody responses as well as testing them on different populations of animals with different disease profiles. Some have been ‘in house’ methods that have not been rigorously evaluated for sensitivity, specificity or test stability. However there has been a concerted effort by several research groups to identify potential TB specific antigenic targets for antibody assays (Schiller et al. 2010a). In principle these targets are likely to be different from the prime
candidates for CMI based tests due to the differences in antigen processing and cell priming for cellular and antibody responses (Male et al., 2006). Therefore, for example, ESAT-6 and CFP-10 are poorly recognised in antibody assays. To date the most promising antigens for sero-diagnosis appear to be MPB70 and MPB83 (Lightbody et al., 1998; McNair et al., 2001; Wiker, 2009). These appear to be *M. tuberculosis* complex specific and are likely to be the core reagents in any antibody based test for bovine tuberculosis. However it is increasingly clear that to improve overall test sensitivity it will be necessary to include additional antigens which individually are not recognised by many infected animals but bring overall improved sensitivity when included in assays. (Lyashchenko et al., 2008; Waters et al., 2006; Whelan et al. 2010).

In addition to the improvements coming from optimised reagents significant improvements in test sensitivity (and potentially test specificity) are likely to be achieved using new technology platforms. Potentially these allow for increased sensitivity by being able to detect very low levels of antibody, as well as developing novel algorithms for improving test specificity. Current technologies either in use or known to be in development include:

- Multiantigen print immunoassay (MAPIA) where several antigens are printed onto a membrane and antibody responses to each are measured for each animal (Waters et al., 2006)
- Fluorescence polarisation assays (FPA) where a tracer (the target antigen or part of it) with a fluorescent molecule bound to it is added to serum. If antibody is present in the serum it binds to this and the measurable fluorescent polarisation increases due to the increased size of the combined antigen-antibody complex (Jolley et al., 2007).
- Immunochromatographic lateral flow test (e.g. ‘Stat-Pak™’) where coloured latex beads impregnated with antigen are mixed with serum. The material flows along a membrane by capillary action across a line where antigen has been impregnated onto the membrane. Where coated latex beads have bound to antibody in the serum these then bind to the membrane making a coloured line (Lyashchenko et al., 2006)
- Single antigen chemiluminescence where magnetic iron beads coated with the antigen of choice are mixed with animal serum and collected magnetically. Any bound antibody is then detected using a specific anti-bovine antibody and identified using a chemiluminescent reaction (e.g., SeraLyte-*Mbu*™ using the Total Optical Assay Device (TOAD) (Green et al., 2009))
- Multiplex chemiluminescence immunoassay (e.g. Enferplex) where individual antigens are printed in small dots on a 96 multi-well plate. Individual serum samples are added to each well and a conventional immunoassay detection method detects antibody to each spotted antigen with colour change on each spot indicating presence of specific antibody to that antigen (Whelan et al., 2008).
- Improved conventional enzyme linked immuno-sorbet assays (ELISA) where a cocktail of *M. tuberculosis* complex specific antigens are used in a conventional antibody detection system using optimised reagents (e.g. IDEXX *M. bovis* assay (J. Lawrence pers comm.)).
The performance of these newer tests has only started to be assessed in recent years. However preliminary results have led to some cautious optimism on their potential value. For example using MAPIA researchers detected a response in 7 out of 9 animals experimentally infected by an aerosol method (Waters et al., 2006). However in another group only 2 out of eight animals infected by the intratonsillar route were detectable. In both groups however the responses were boosted by the comparative skin test such that 8/9 and 5/8 of the animals in each respective group were detected 3 days following the skin test. Broadly a similar response was seen in the same animals when the lateral flow ‘Stat-Pak’ was used. Here 60% of infected animals were detectable 7 weeks after infection and 96% 18 weeks after infection although this was only achieved 2 weeks after a comparative intradermal skin test.

In one study using a fluorescence polarisation assay the specificity of the assay was estimated as 99.9% using samples from presumed *M. bovis* free herds (Jolley et al., 2007). The sensitivity of the test was estimated to be 61.3% when compared to post mortem PCR. In this study this was probably an over estimate of sensitivity given the poor sensitivity of PCR on post mortem tissues. However this study was based upon detecting responses to MPB70. Other studies have suggested more infected animals will be positive to MPB83 (Whelan et al., 2008) than to MPB70 therefore there may be opportunities to further refine this assay by optimising the antigenic target used.

Using a multiplex assay incorporating 13 antigens one study concluded the test had a sensitivity of 93.1% and a specificity of 98.6% calculated from known infected or uninfected animals and based upon a positive response detected to 2 or more antigens (Whelan et al., 2008). Interestingly a recent latent class analysis of this test (Enferplex) suggested that the true test sensitivity varied between 37.1% (using a high positive cut-off) and 68.4% (using a low positive cut-off), and that the true specificity varied between 92.1% and 99.8% for the respective cut-offs (Clegg et al. 2011). However it remains unclear what the comparative skin test history of the animals used in the studies was and so it is uncertain what effect this might have on test performance in the field, as it does with many of the other antibody assay trials.

In general it appears that tests for bovine tuberculosis based upon antibody detection have a low sensitivity. However this may be misleading in that although the tests are not likely to identify all or indeed most infected animals they may be useful in identifying animals with a high risk of infecting other animals. For example there is increasing evidence for a suggested general principle where approximately 20% of any host population for a disease contributes 80% of the net transmission potential for a very wide variety of diseases (the ‘20/80’ rule), i.e. a small proportion of any infected population has a disproportionate effect on the overall transmission of disease (Woolhouse et al., 1997). This principle is applicable to cattle populations (Woolhouse et al., 2005) and may well be in action with bovine tuberculosis where the majority of infected animals are found on a minority of farms, suggesting that in most farms where TB is found it is not readily transmitted but in a small number it is. This implies that a small sub-population of infected animals is infectious. While it seems unlikely antibody based tests could be used to identify all or most infected animals
they may have value in identifying important sub-groups of highly infectious animals which are otherwise poorly detected using standard tests. There is at least a theoretical possibility that these tests could have a disproportionately positive effect if they are proven to be able to identify diseased animals that other methods fail to detect. There remain important knowledge gaps as to the usefulness of the currently available antibody tests and the field circumstances they could most appropriately be applied to. Therefore there appears an urgent need to undertake a field trial to assess the applicability of a number of the current leading candidates to identify if they could bring additional value to the current N. Ireland TB eradication program.

4.5 Other Disease Markers
A number of studies have been carried out to attempt to identify if there are other biomarkers that could be used to identify infection in cattle. Amongst these are volatile organic compounds (VOCs) which are present in blood serum or urine. It has been reported that the balance of VOCs can change depending upon the disease status of the animal (Fend et al., 2005). A recently completed DEFRA funded research project (SE3221) suggested that changes in VOCs in serum and urine from infected cattle could be identified. However while the technologies, ‘eNoses,’ were able to distinguish animals at the group level, i.e. infected compared to uninfected groups it was not able to distinguish animals at the individual animal level. Interestingly the study suggested a possible link between serum ammonia and infection in badgers although this finding needs further validation in future studies. It is worth considering that while the present technology does not appear to have a good sensitivity at the individual animal level there is at least potential for it to be useful in identifying sub-groups of infected animals, for example it could be that animals with more advanced disease are more likely to have measurable VOC changes. However given current knowledge this is at best conjecture.

A number of other studies have been conducted to attempt to identify other biomarkers particularly in blood samples that might indicate infection. For example ‘proteonomic fingerprints’ which screen for a wide variety of serum proteins using mass spectrometry have suggested that certain profiles could have a sensitivity of 94% and a specificity of 95% in human TB patients (McNerney and Daley 2011). However the sophisticated analytical technology required for these approaches means that they are not currently easily applicable until new test platforms are developed and validated for their use. Other approaches are currently under study such as DNA transcriptional technologies which identify patterns of genes either expressed or turned off in response to infection in order to develop a signature of gene activity which would indicate infection. While some preliminary work has been promising using this approach (J. McNair pers comm) it remains largely experimental, requiring further research and validation.
5 References


Coad, M., Clifford, D., Rhodes, S.G., Hewinson, R.G., Vordermeier, H.M., Whelan, A.O., 2009, Repeat tuberculin skin testing leads to desensitisation in naturally infected tuberculous cattle which is associated with elevated interleukin-10 and decreased interleukin-1 beta responses. Veterinary research 41, 1-12.


Crawford, A.B., 1926, Tuberculin sensitization in Guinea-Pigs caused by various acid-fast organisms. Journal of the American Veterinary Medical Association 69, 10.

de la Rua-Domenech, R., Goodchild, A.T., Vordermeier, H.M., Hewinson, R.G., Christiansen, K.H., Clifton-Hadley, R.S., 2006a, Ante mortem diagnosis of


Francis, J., 1958, Tuberculosis in Animals and Man.


assay for rapid detection of Mycobacterium bovis infection in multiple species of free-ranging wildlife. Veterinary Microbiology 132, 283-292.


their Relevance for Disease Control and Eradication. Transboundary and Emerging Diseases 57, 205-220.


Appendix 1. The effect of disease prevalence on the predictive value of tests with varying sensitivity and specificity characteristics

In this hypothetical case the positive and negative predictive values of a test with a sensitivity of 95% and a specificity of 99% has been plotted over a range of population disease prevalences. Most importantly the positive predictive value of the test is very high when there is a high prevalence of infection within the population but decreases markedly as the proportion of infected animals decreases. So for example with a disease prevalence of 1% the positive predictive value of a test with these characteristics is less than 50% i.e. more than half of all animals testing positive will not be infected. The predictive values will change depending upon the estimated sensitivity and specificity. For example if the sensitivity of the test is reduced to 60% (approximately that estimated for the comparative tuberculin skin test) the positive predictive value at low herd prevalences is further reduced but most dramatic is the effect on the negative predictive value at high herd prevalences (figure 2). The most dramatic effect on positive predictive values is when the specificity of the test is reduced. So if this falls to 90% then approximately half of test positives are likely to be incorrect when the disease prevalence is 10% (figure 3). This illustrates the need to have tests with high sensitivities and specificities, but most crucial at low disease prevalences is the specificity of the test.

![Figure 1. Effect of Changes in the Proportion of Infected Animals on the Predictive Values of a test with 95% sensitivity and 99% specificity.](image-url)
Figure 2. Effect of Changes in the Proportion of Infected Animals on the Predictive Values of a test with 60% sensitivity and 99% specificity.

Figure 3. Effect of Changes in the Proportion of Infected Animals on the Predictive Values of a test with 95% sensitivity and 90% specificity.