Ante mortem diagnosis of paratuberculosis: A review of accuracies of ELISA, interferon-γ assay and faecal culture techniques

S.S. Nielsen, N. Toft

To cite this version:

HAL Id: hal-00532378
https://hal.science/hal-00532378
Submitted on 4 Nov 2010

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Ante mortem diagnosis of paratuberculosis: A review of accuracies of ELISA, interferon-γ assay and faecal culture techniques

S. S. Nielsen¹, N. Toft²

¹Department of Large Animal Sciences, Faculty of Life Sciences, University of Copenhagen, Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark

²Danish Meat Association, Vinkelvej 11, DK-8620 Kjellerup, Denmark

Corresponding author: Søren Saxmose Nielsen, Department of Large Animal Sciences, Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark, e-mail: ssn@life.ku.dk, fax: + 45 35 33 30 22, tel: +45 35 33 30 96.

Abstract

Infections with Mycobacterium avium subsp. paratuberculosis (MAP) can be latent for years without affecting the animal, but the animal may become infectious or clinical at some point. Diagnosis of paratuberculosis can be a challenge primarily in latent stages of the infection, and different diagnosis interpretations are usually required by the variety of decision makers. The objective of this paper was to provide a critical review of reported accuracies of ELISA tests, interferon-γ assays (IFN-γ) and faecal culture (FC) techniques used for diagnosis of three defined target conditions: MAP infected, MAP infectious and MAP affected animals.

For each animal species, target condition and diagnostic test-type, sensitivities (Se) and specificities (Sp) were summarised based on a systematic, critical review of information in literature databases. The diagnostic test information often varied substantially for tests of the same type and make, particularly ELISA, which was the most frequently reported test-type. Comparison of the various tests accuracies was generally not possible, but stratification of test-evaluations by target condition improved the interpretation of the test accuracies. Infectious and affected animals can often be detected, but Se for infected animals is generally low. A main conclusion of the review was that the quality of design, implementation and reporting of evaluations of tests for paratuberculosis is generally poor. Particularly, there is a need for better correspondence between the study population and target population, i.e. the subjects chosen for test evaluation should reflect the distribution of animals in the population where the test is intended to be used.
1. Introduction

Paratuberculosis is a chronic infection, which has been of particular concern in ruminants. The infection is caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP). The major effects of the infection on the animal level can be reduced milk yield (Benedictus et al., 1987; Kudahl et al., 2004), premature culling and reduced slaughter value (Benedictus et al., 1987), and losses due to continued spread of the infection (Kudahl et al., 2007). Not all infected animals will experience losses, which may be because of culling for other reasons than paratuberculosis or because they can resist the infection developing into the debilitating stages. The occurrence of the latter is still poorly understood (Mortensen et al., 2004).

Prevalences of the infection vary world-wide (Kennedy and Benedictus, 2001), but most notably the apparent prevalences vary by the test and test strategies used in the prevalence studies conducted. Control of the infection can be obtained via timely detection and culling of infectious animals and reduction of transmission from these animals. Eradication will usually require the detection and isolation of infected animals, as these potentially can become infectious at some point in time. “Isolation” in this regard means that infected animals and their excretions should not be allowed contact with susceptible animals.

Eradication is defined as: “The purposeful reduction of specific disease prevalence to the point of continued absence of transmission within a specified area by means of a time-limited campaign” (Yekutiel, 1980). This is to emphasize that complete eradication would require eradication of a microbial agent globally. Control is described as “any effort directed toward reducing the frequency of existing disease to levels biologically and/or economically justifiable or otherwise of little consequence” (Martin et al., 1987).

The objective of this report was to conduct a critical review of reported diagnostic test evaluations of ELISA, FC and interferon-γ tests (IFN-γ) used for ante-mortem diagnosis of conditions caused by infection with MAP. To facilitate the comparison across test studies, three target conditions: affected, infectious and infected with MAP were defined and diagnostic sensitivity (Se, probability of correct test positive classification) and diagnostic specificity (Sp, probability of correct test negative classification) reported with respect to these conditions.

2. Stages of infection /Conditions detected

2.1. Pathogenesis
Diagnosis and thereby control of the infection are hampered by a long incubation period. It is generally
assumed that infections with MAP occur in young animals, and that some age-resistance occurs. Cattle are
thought to be most susceptible from 0 to 4 months of age (Taylor, 1953), although infections have been
established in adults feed high dosages of MAP (Doyle, 1953). Similar conditions can be speculated to occur
for other animal species. Clinical disease has been observed to most frequently occur among cattle 2 to 5
years of age, although cattle from very young to very old (0 to 13 years of age) have been affected (Doyle
and Spears, 1951).

MAP is an intracellular pathogen. Subsequent to infection, MAP is initially controlled by a
predominating T helper 1 (Th1) response. Th1-cells are, among other features, characterised by their
production of interferon-γ and some IgG2. Later in the course of infection, a predominant Th2 response may
occur, and control of the infection is thought to be lost (Stabel, 2000). During the Th1 response, Map is shed
in small numbers, which may be sufficient to elicit a positive result in faecal culture tests (FC). There is
a correlation between occurrence of IgG and bacterial shedding of MAP (Nielsen and Toft, 2006a), but the
time-wise relations between the two events is not fully described. Experimental infections suggest that
bacterial shedding decreases 10 to 14 months after inoculation, to increase again later, with sero-conversion
occurring around 10 months post-inoculation (Lepper et al., 1989). Waters et al. (2003) demonstrated both
cellular and humoral immune responses approximately 100 to 150 days after initial infections with MAP, and
Eda et al. (2006) also demonstrated IgG in calves less than 1 year after inoculation. In naturally infected
animals, sero-conversion has been shown to occur in 95-98% of cows shedding MAP (Nielsen and Ersbøll,
2006). The age at which sero-conversion occurred was from 2.2 to 11.7 years. It is speculated that great
variation of the time to occurrence of bacterial shedding and occurrence of antibody responses are caused
by variation in infective doses occurring with natural MAP infections. Studies using fixed dosages and known
times of infection have also resulted in great variation in the time to occurrence of FC-positivity and ELISA-
positivity (Lepper et al., 1989), and the temporal variation in pathogenetic events may be further enhanced if
the size and number of dosages varies. In test-evaluations, it is therefore needed to consider the stage of
infection. Generally, age can be an indicator of the stage of infection, in that young animals will rarely be
expected to shed detectable amounts of bacteria and have IgG1, whereas older animals are more likely to
have bacterial shedding, antibodies and clinical disease. The age-distribution among study objects can
therefore be of great significance in the evaluation of a diagnostic test. As an example, the probability of
detecting infected cows 2 years of age using an antibody ELISA has been estimated to 0.06, whereas the
same probability was 0.50 for cows 5 years of age (Nielsen and Toft, 2006a). The distribution of animals 
among different stages of the infection can also be of great importance. Different antigens are used in various immunological tests. The antigens are usually derived from MAP or 
*Mycobacterium avium* subsp. *avium* (MAA). In the paratuberculosis literature, MAP strain 18 is MAA serovar 
2 (Chiodini, 1993). Different preservations of the antigens are being used based on their biological 
characteristics (Koets et al., 2001) and availability. Generally, antigens should be immunogenic to be used 
for diagnostic tests. Cho and Collins (2006) showed that proteins derived from culture filtrates rather than 
cellular extracts were more likely to be antigenic. The use of antigens from MAA may be as useful as 
antigens from MAP (Nielsen et al., 2001), but there have been made no extensive comparisons of antigens 
on field data indicating which antigens are the better. Irrespective of which of the two bacterial species the 
antigen is derived from, both immunogenecity and cross-reactivity should be considered. So far, the 
superiority of one specific antigen has not been demonstrated, where both immunogenecity and cross-
reactivity have been evaluated. It is likely that this may be specific to specific geographic areas, due to 
potentially different distribution of bacteria potentially causing cross-reactions, such as MAA, from one area 
to the other. In herds with high sero-prevalences estimated by two commercial ELISAs (Parachek, 
CSL/Biocor, Omaha, Nebraska, USA and HerdChek, IDEXX Laboratories Inc., Westbrook, Maine) the 
prevalence of environmental mycobacteria were higher compared to herds with a low sero-prevalence, 
suggesting that a high prevalence of environmental mycobacteria may result in many false-positive ELISA 
results (Roussel et al., 2007). It has also been demonstrated that experimental infections with environmental 
mycobacteria such as *M. intracellulare*, *M. scrofalceum* and *M. terrae* can result in significant serological 
reactions (Osterstock et al., 2007).

2.2. Target conditions

The target condition detected by any diagnostic test is essential in the evaluation of the diagnostic accuracy 
of the test. Some authors refer to this a “case definition” (e.g. Collins et al., 1991). The choice of target 
condition for the diagnosis “paratuberculosis” should vary depending on the purpose of testing, i.e. the 
effects that are of primary interest to decision makers. A schematic presentation of the pathogenesis and the 
effects is given in Fig. 1. From this, a number of conditions can be identified. The target condition chosen by 
the evaluators in a given study ideally depends on the decision makers, whom can subsequently make 
decisions based on the test results. In this report, three target conditions, affected, infectious and infected
with MAP, have been defined to classify the test evaluation studies included in the review. These target conditions are considered pivotal.

2.2.1. Animals affected by MAP.
Animals affected by MAP are usually classified based on clinical signs such as diarrhoea (persistent or intermittent), chronic weight loss or reduced milk production. MAP infection should be present, which could be documented via gross pathology, histopathology or cultivation of MAP from tissues or faeces. The animal does not have control of the infection and is affected to a degree, where parameters like milk production and general performance is decreasing due to the infection. This definition can be of use for a decision maker whom wishes to make decisions based on the performance of the cow. Reduction in milk yield is often not recorded and reported systematically in the studies, and this aspect of the infection is therefore rarely included.

2.2.2. Animals infectious with MAP
Infectious animals are defined as those that shed MAP at the time of testing with the test under evaluation and thereby they are a risk for transmission of MAP to susceptible herd-mates. The condition “Infectious” also includes animals which are “Affected”. In principle, the infectious group may also contain non-infected animals, which may be passive shedders of MAP. These animals are hypothesised to ingest MAP from heavy environmental contamination (Sweeney et al., 1992a) without being infected. The “infectious” status of these animals would therefore be considered to be transient. In many study reports, only shedding in faeces is included. The shedding is defined based on one or more tests evaluating the presence of MAP in faeces. Animals which are transmitting the agent in milk and in uteri are thereby not included specifically. Given that animals are infectious via milk and in uteri without shedding bacteria in faeces, they will therefore bias the accuracy estimated in the studies reported.

2.2.3. Animals infected with MAP.
Infected animals carry MAP intracellularly but substantial replication need not take place because the infection can be latent. The condition “Infection” also includes animals which are “Infectious” and “Affected”. The definition of “infection with MAP” is any condition where entrance and persistence of MAP have lasted
long enough to give an immune response at any time during their life; i.e. there is no time-specific cut-off for
this event to occur. It is assumed that once a cow has an established infection, the infection persists for life.

2.2.4. Animals not affected, infectious or infected with MAP

The diagnostic Se of a test reflects the ability to detect the target condition given it is present and the
diagnostic Sp of a test reflects the ability to test negative given the target condition is not present. In many
studies, the target conditions mentioned above are mixed. Studies defining one target condition for the
evaluation of Se and another for Sp can therefore be subject to peculiar interpretations. The interpretation of
“false-positives” can therefore vary.

The Sp of the condition “non-affected” has to our knowledge not been systematically assessed in any
studies.

The Sp of the condition “non-infectious” refers to an animal, which does not shed MAP on the time of testing,
but some of these non-infectious animals can be infected. Therefore, false-positive test-results include both
infected as well as non-infected (i.e. MAP free) animals. The decision maker will usually have little use of this
information, unless documentation exists to show near perfect Sp of the test for non-infected animals. This
would mean that all false-positive animals are indeed infected, but not infectious at the time of testing.

The Sp of the condition “non-infected” is the situation where an animal is free of the infection. False-positive
reactions are due to cross-reactions to other mycobacteria, laboratory errors, vaccination reactions and the
like for immunological tests (Houe et al., 2004). False-positive test results could also be a consequence of
the “pass-through” phenomenon, where ingested bacteria are shed 1 to 7 days post-ingestion, potentially
without being infected (Sweeney et al., 1992a).

2.2.5. Utility of test-results related to different conditions

The utility of the test results related to the three conditions defined above could be as follows:

Affected animals: The value of these animals is low because their production is reduced, they loose weight
and the value at slaughter is probably reduced. There is a risk they will die from the infection;

Infectious: These animals are currently infectious and are a risk to susceptible animals. They should be
managed so that transmission to susceptible animals is avoided. These animals constitute both a
short-term and a long-term economic burden, as they are likely to have a reduced milk yield or will
experience it in the near future. The long term loss will be due to their transmission of MAP to herd-mates;

Infection: Infected animals constitute a risk of becoming infectious and thereby transmit the infection to susceptible animals. A population containing infected animals cannot be declared free of infection and proper identification of infected animals is important herd-certification schemes, when trying to establish ‘MAP-free’ herds/populations or keeping MAP out of certified herds. In economic analyses, these animals may be of interest in long-term planning only, as their effect on the population will often be seen only after a number of years.

Theoretically, the conditions may be easily defined. In practice, this may not be the case. Other conditions can be defined, depending of their use, e.g. assuming that the cell-mediated immune reactions are studied, it is necessary to establish a condition “occurrence of cell-mediated immune reactions”.

2.2.6. Effect of age on the condition detected.
Considering the chronic nature of the infection, with many disease stages, it could be speculated that older animals are of greater risk of having a given condition than younger animals. However, this does not necessarily imply that it is easier to detect the condition in older animals. Higher diagnostic Se for detection of infection has been demonstrated for older cows relative to younger cows (Nielsen and Toft, 2006a), but age does not seem to influence detection of infectious and affected cattle (Nielsen and Toft, 2006a; 2006b), although this has not been formally tested.

3. The ideal test evaluation
Very few if any perfect test evaluations have yet been performed. The lack of a 100% accurate reference test and a variable incubation time seem to be primary obstacles in doing so. The effect of choice of reference test on accuracy estimates was demonstrated by McKenna et al. (2005), where accuracy estimates of ELISA based on use of tissue culture as reference test was different from accuracy estimates based on faecal culture as reference test. “Infection” may be established if a thorough microbiological examination of the animal is performed at slaughter, but it is insufficient to sample tissues only from ileum and the ileocecal lymph nodes, because this sampling procedure will fail in detecting of many infected animals (Whitlock et al., 1996). The authors state that up to 100 sites per animal are required sampled to establish the true infection status of the animal. A study containing an adequate number of animals would therefore become quite
expensive, since the prior infection status should not be known when carrying out the study, in order to ascertain a distribution of infection stages that is representative of the distribution in the target population.

Greiner and Gardner (2000) provide an extensive check list, which should be used as a starting point for any epidemiological validation of diagnostic tests. From their recommendations and the discussion of target conditions in the previous sections, we suggest that any test evaluation must at least include the following components:

a) Data must be from an observational study and the study population should be representative of the target population in which the test is to be used, i.e. the variations in incubation period, exposure dosages and age-distribution of the target population should be reflected in the study population.

b) The target condition should reflect the intended purpose of the test. The same target condition should be used for evaluation of both the Se and Sp. The interpretation of the test-information should be made in concordance with the target condition, and the interpretation of false-positives needs to be clear. As an example, the target condition “Infected” can be studied by classical test-evaluation methods, where animals are classified into truly infected, not infected based on microbiological examination of multiple tissues per animal. An alternative is to use latent class analyses (e.g. Nielsen et al., 2002), where tests that are biologically unrelated are studied, e.g. FC for detection of MAP and antibody ELISA for detection of humoral immune reactions.

c) A calculated sample size which reflects the purpose of the test evaluation, i.e. the choice of precision in the estimates should depend on the specific purpose of the test evaluation.

4. Frequent mistakes in test evaluations performed

A number of mistakes are usually made in evaluation of diagnostic tests. Some of these can be avoided by thorough planning of the study, whereas others are difficult to avoid while still making the study practically and economically feasible. Frequent mistakes include:

1) Selection of animals for the evaluation is performed by use of the test being evaluated, or a test measuring the same response. Examples include animals selected from herds classified free of infection based on negative humoral immune responses, where the test evaluated detects the same responses.
2) Variable case definitions used to classify animals across a study. All study objects should be subject to the same classification procedures. An example is the study by Sockett et al. (1992b), in which not all infected animals were subject to the same set of classification procedures, since only sero-positive and FC-negative animals were subject to the defining histopathological procedures. This would result in an over-estimation of the Se, because animals in early-stages of infection could be missed by the reference test. However, the study could be included for inference on infectious animals, given that FC was used to define the reference status, because all animals had been subject to the FC.

3) Another frequent, but to some extent unavoidable mistake is the use of a study population from a region historically known to be free of MAP to estimate Sp and sample the non-free target population to estimate Se. These two populations should be geographically comparable, because the environmental flora can be expected to give rise to different cross-reactive responses in serological tests.

5. Review of test-evaluations in literature

A review of test-evaluations was carried out by searching the available databases by November 14, 2006. These included: Agricola (1970 to Sept. 2006); Agris (1975 to Sept. 2006); Biological Abstracts /RPM (R) (1989 to 2003); BIOSIS Previews (2004- Nov. 11, 2006); Biological Abstracts (1990 to 2000), CAB Abstracts (1973 to Sept. 2006) and Medline through Pubmed (1970 to Nov. 14, 2006). The search terms were: paratuberculosis, Johne’s or Johnes combined with diagnosis, diagnostic accuracy, Se, Sp, validation or diagnostic performance.

This search generated 2137 hits including duplicate records. After exclusion of duplicate records and non-peer-reviewed publications, 312 publications remained. These were further reduced to 102 publications by exclusion of studies, where the abstract indicated that a test-evaluation was not an objective of the publication or if the language was not English. In cases of uncertainty on whether a test-evaluation had been performed, the studies were included for further assessment of the full paper. The remaining 102 publications were further evaluated for study objective and data quality.

5.1. Data extraction

Data were extracted from the publications in a standard form including the following items:
a) Animal species;

b) Type of test, i.e. serum antibody ELISA, milk antibody ELISA, FC and IFN-γ;

c) Test system, i.e. name of test for commercial tests and in-house for non-commercial tests;

d) Antigen used in serological tests;

e) Study design: Observational (case-control; cohort; cross-sectional) or experimental; longitudinal or instantaneous;

f) Data origin: field data or serum-bank or similar;

g) Conditions detected: “infectious” if animals were shedding bacteria; “infected” if animals were deemed infected by a reference method stated; “affected” if cows were clinical, had reduced production or were in some other way deemed to be affected by the infection (“shedding” alone was not considered affected); “from free region” if this had been justified (herds that had once been FC negative were not from “a free region”); “non-infectious” were from non-shedding animals);

h) Sample size, i.e. the number of animals assessed among animals with condition (for Se) and without condition (for Sp);

i) The number of test-positives among animals with condition and the number of test-negatives among animals without the condition;

j) Uncertainty measures, e.g. 95% confidence interval (95% CI) for classical methods or 95% credibility posterior intervals for Bayesian methods;

k) Cut-off used for discriminating between positive and negative animals in tests that are not dichotomous;

l) Age distribution of the study population(s);

m) Arguments for inclusion and exclusion of the data from the study;

n) Authors and year of publication;

o) Year of study;

p) Geographical origin of data.

5.2. Assessment of study quality

Many studies included some sort of bias, and it was not possible to avoid these biases generally. Therefore, it was decided to include studies if the following criteria were fulfilled:

a) One of the objectives of the study should be related to test evaluation;
b) A unique target condition should be defined for animals with the condition and animals without the condition. The target condition could differ between the estimates for Se and the estimates for Sp. If a condition could not be determined or varied across study objects, the study was excluded;

c) Animals were classified by the same criteria within the study;

d) The study population was not selected or defined by use of the test (or a related test) under evaluation, e.g. selection of herds that had previously been shown to be negative by serological methods could not be used for evaluation of a serological test;

e) Random inclusion of study objects was used, e.g. studies where study objects were included due to being suspect were excluded from the current study. Given that the selection of study objects was not described sufficiently, these studies were also excluded;

f) Infections were natural, i.e. results from experimental infections were excluded, but results of natural infections were included if these were given separately in a given publication;

g) The test-result of a given test under evaluation was based on a single test, i.e. repeated test-results could be used for defining a given condition but could not be used for the test evaluated;

Some test-results were confirmed by re-testing, but the original data were used. The argument for doing so is that by sufficient re-testing, the required result can usually be obtained if enough re-tests are carried out, and the same set of procedures should be applied to all samples.

5.3 Data analysis

The recorded data from each of the studies were tabulated in separate tables for different animal species. The target condition detected, age and test-characteristics were included in the tables, and the range of estimates was extracted.

6. Test evaluation summaries

The 102 publications evaluated in detail contained a total of 153 test-evaluations. A test-evaluation is here defined as the evaluation of one test, but some publications contained more than one test. In some publications, more than one test or more than one condition was evaluated. In Table 1, the distribution of 153 test-evaluations is given by target condition and animal species. A total of 68 studies were excluded, as they did not fulfil the inclusion criteria. The remaining studies included 58 studies from cattle, 15 studies from
goats, 11 studies from sheep and 1 study from deer. Four of these studies were conducted using latent class analyses: 1 on cattle, 2 on goats and 1 on sheep. The distribution among tests used was as follows: 64 serum antibody ELISA (SELISA) studies, 6 milk antibody ELISA (MELISA) studies, 4 IFN-γ studies and 8 studies on FC. An overview of these is given in the subsequent sections, with division into animal species, test-type and condition detected. The test-evaluations did not cover all types of tests FC, SELISA, MELISA and IFN-γ for all animal species, and studies on those not mentioned have not been reported or did not fulfil the inclusion criteria set.

Although data should have been extracted as specified section 5.1, some information was consistently not available from most publications. Therefore, only parameters shown in Tables 2 and 3 are reported further, as these were almost consistently reported. Age of the study groups was also included, although this information was only reported in some cases. Unless otherwise stated, it is assumed that adult, mature animals comprised the target population.

6.1. Cattle – Faecal Culture

Usually, the Sp of FC is considered to be almost 100%, if the isolates obtained at culture are confirmed to be MAP by molecular methods such as confirmatory IS900 PCR. However, due to the potential pass-through phenomenon (Sweeney et al., 1992a), it is possible that testing of non-infected animals on contaminated premises can lead to false-positive reactions. A latent-class approach to evaluation of the Sp of FC for detection of non-infected animals would therefore seem appropriate. Nielsen et al. (2002) estimated the Sp to be 98% in a population, where the non-infected animals were subject to contamination from infected herd-mates.

The Se from the test-evaluations in the present study are given in Table 2, for the three different target conditions. The Se of FC to detect affected animals have been estimated in one study only, resulting in an estimated of 0.70, which is similar to the Se of 0.74 estimated for detection of infectious animals. The Ses of FC for detection of infected animals were in the range 0.23 to 0.29, except for the study by Billman-Jacobe et al. (1992). However, in that study the infection status was determined at necropsy of cull cattle, many of which where clinically ill or gave strong positive reactions in serological tests. Therefore, their estimate must be considered biased to a degree which suggests it should be excluded from the review or that the true target condition of the study differ from their reported target condition.
6.2. Cattle - Serum antibody ELISA

The serum antibody ELISA for cattle is the test most frequently evaluated. The test evaluations include studies on a number of commercial ELISAs and in-house ELISAs, a variety of antigen preparations and different age-groups of animals. In Table 3, the Se and Sp is given for each study. The test used is classified into groups based on the producer and the antigen used. The antigen preparations are in most cases not comparable. Therefore, the names in the groups indicate only the source of the antigen. It is assumed that the antigen in the Parachek® (Prionics AG, Schlieren-Zurich, Switzerland) test is from the MAP VRI 316 strain, but this has not been confirmed. The producer did not respond to requests of the test specification, but the test should apparently be based on the test described originally by Milner et al. (1990). IDEXX Laboratories, Inc. (Westbrook, Maine, USA) being the provider of the IDEXX HerdChek Mycobacterium paratuberculosis test, cannot share the specification of their antigen preparation either, and it is unknown whether the both the HerdChek and the Parachek test have remained the same over the years. Other test-names were categorised as “Various” because they often do not have a name. For antigen preparations, the variety of preparations gives reason to the same “Various” group. The test names and antigen preparations are therefore not suitable for further subdivision of the data. In most studies, the age-groups studied are incompletely characterised. For those studies in which they were given, it is clear that the studies are hardly comparable, but it is assumed that the majority of animals in each study are a random collection of parturient animals, except if stated otherwise.

6.3. Cattle - Milk antibody ELISA

The constitution of milk antibody ELISAs has been less variable than the serum antibody ELISA. Only two different antigens have been used: a lipo-arabinomannan preparation (LAM) and a commercially available antigen from Allied Monitor (Fayette, Missouri, USA). The Se and Sp still vary, which may be primarily due to the choice of cut-off used in the different studies. A summary is given in Table 4. The test has not been evaluated for affected animals and very few studies have been conducted on infected animals. Therefore, the most studies have been conducted for diagnosis of infectious animals, and the Se varies from 0.29 to 0.61, with Sp in the range of 0.83 to 1.00.

6.4. Cattle – interferon-γ
The IFN-γ has only been evaluated in two studies in cattle (Paolicchi et al., 2003; Huda et al., 2004), and on a limited data material. The results are summarised in Table 5, and the Se for detection of infectious animals varied from 0.13 to 0.85. Unfortunately, the results may not be representative of the populations in general, but so far the studies are the only information available. In Paolicchi et al. (2003), the data material is based on only one herd, in which clinical paratuberculosis had occurred. In Huda et al. (2004), the negative reference herds were chosen among herds with low prevalence of antibody positive cows. It is uncertain whether both publications should be excluded based on these weaknesses. An advantage of the study by Huda et al. (2004) is the separation into three age-groups (Table 5). Neither of the studies assessed the Se of the test for detection of infected animals.

6.5. Goats – Faecal Culture

Kostoulas et al. (2006) have conducted the only study reporting on the accuracy of FC in goats. The condition studied was infected animals, and Se and Sp were assessed using latent class methods on data derived from animals > 1 year of age. The Se was estimated to 0.08 (95% credibility posterior interval (95% CPI): 0.02; 0.17) and the Sp was estimated to 0.98 (95% CPI: 0.95; 1.0), based on data from 368 goats.

6.6. Goats - Serum antibody ELISA

As in cattle, the most widely assessed test in goats has been the serum ELISA for detection of antibodies. However, contrary to cattle, the most widely condition detected has been affected and infected animals, whereas infectious animals has rarely been the condition detected (Table 6). Nonetheless, it appears that ELISA is more accurate in detection of all conditions than in cattle, with Se ranging from 0.82 to 1.0 for affected animals and 0.63 to 0.84 for infected animals. The corresponding Sp generally range from 0.92 to 1.0, although in one study with a very high Se of 0.91 the corresponding Sp was 0.79 only (Dimareli-Malli et al., 2004).

6.7. Sheep – Faecal Culture

FC has only been evaluated for the condition infected in sheep, and in one study only. The Se was estimated to 0.16 (95% CPI: 0.02; 0.48) and the Sp to 0.97 (95% CPI: 0.95-0.99) based on data from 368 animals evaluated in a latent class analysis (Kostoulas et al., 2006).
6.8. Sheep - Serum antibody ELISA

Ten serum antibody ELISAs have been evaluated for use in sheep (Table 7). In the evaluation of 5 tests, the Se were evaluated for detection of affected sheep and in the evaluation of 5 tests, infected animals were the target condition. All studies were based on detection of non-infected animals for estimation of Sp. The most widely used test was Parachek, and in 7 of the 10 studies, the antigen used was apparently MAP VRI316.

For affected animals, the Se varied from 0.36 to 0.85, and for infected animals, the Se were 0.16 to 0.44. Sp ranged from 0.95 to 0.99 (Table 7).

6.9. Deer – Serum antibody ELISA.

In clinically affected deer, two non-commercial serum antibody ELISAs have been evaluated (Griffin et al., 2005). The ELISA based on PPDj antigen from CIDC (Lelystad, The Netherlands) had a Se of 0.83 (95% CI: 0.74; 0.89) (among 102 clinically affected deer) and a Sp of 1.00 (exact 95% CI: 0.9906; 0.9998) among animals from herds with no MAP infection history or clinical signs associated with MAP. An ELISA based on PPA3 antigen from Allied Monitor (Fayette, Missouri, USA) had a Se of 0.85 (95% CI: 0.77; 0.91) with a Sp of 0.98 (95% CI: 0.96; 0.99).

6.10. Overall summary of sensitivity and specificity

An overall summary of Se and Sp for each animal species, test and target condition is shown in Table 8, based on information in Sections 6.1 to 6.9. Both FC and ELISA generally have medium to high Se for detection of affected and infectious adult cattle, with Sp of 1.0 by definition. The range of Se of ELISA for detection of infectious cattle is huge, which is partly a reflection of the number of test-evaluations included in this group. The Se of FC for detection of infected cattle are in the range 0.23 to 0.29, which may be slightly better than Se of ELISA (range 0.07 to 0.39). However, a given Se of ELISA needs to be evaluated in combination with Sp, and it cannot be determined which of the tests that are the better. The variation in test-accuracy within test-evaluations for IFN-γ is huge, but basically only two studies have been performed, with significant differences in the results. It cannot be determined which of the results that are the more reliable.

In deer, only one study including two ELISAs used for clinically affected animals have been reported, with promising results. However, the lack of studies on infected and infectious deer prompts for studies on these conditions.
ELISA used for infected, infectious and affected goats indicates that this test can have utility because of generally high Se, irrespective of target condition. The Se of FC for detection of infected goats is, however, not promising with a Se of 0.08 and a Sp of 0.98, in the one study reported. Further studies are needed to draw reasonable conclusions. However, the Se of FC is comparable to that of sheep, indicating that the estimate may be valid for the particular test. The FC method evaluated for both sheep and goats originate from the same study, and other FC methods may prove to be superior. IFN-γ and milk ELISA have not been evaluated in deer, goats and sheep. Results of test-evaluations for serum ELISA for sheep show considerable variation for both affected and infected animals, but the Se for serum ELISAs used for infected animals appear to be lower than serum ELISAs used for infected animals, as would be expected.

7. Discussion

This report summarises Se and Sp estimates for detection of animals infected by, infectious with or affected by MAP, obtained through a critical review of literature. The report is the first to divide animals into the three target conditions (infected, infectious and affected), but these can be very useful in the processes where decision makers have to make decisions related to different conditions. The division also appears to provide estimates which are more homogenous for a given test than had the estimates been reported for one group only. As an example, the Sp of FC were 0.96, 0.98 and 0.97, for cattle, goats and sheep respectively. Serum antibody ELISA used to detect infected cattle had Se in the range from 0.07 to 0.22, whereas serum antibody ELISAs for detection of infected goats were in the range 0.63 to 0.84. These ranges are much narrower than could be expected. The narrow range is of course partly a function of the low number of test-evaluations per group, but it still appears to be narrower irrespective hereof.

Division of the target conditions into affected, infectious and infected animals to some extent reduced the variation of Se within a group, as would be expected. For a decision maker, estimates of Se and Sp that are specific for a given target condition, should be preferred, because the decision maker can then report the probability of having the given condition based on the test result. If no distinction between target conditions is made, a Se and Sp is for an average of infected, infectious and affected animals will have to be assumed. In Collins et al. (2006), such an average must have been assumed, as we were unable to identify in literature the Se and Sp estimates reported as “assumptions for test Se and Sp”. As an example, they reported a Se of 0.60 ± 0.05 for FC in cattle, given the best test is used. This figure is twice the size of the Se reported in literature for detection of infected animals. The Sp of FC was reported to 0.999 ± 0.001, which would be
applicable only for non-infected animals in non-infected herds, not a randomly selected animal. The
estimates from literature suggest the Sp to be 0.98 for cattle, which is supported by the estimates of 0.97
and 0.98 for sheep and goats, respectively. The target conditions and basis for the accuracies reported by
Collins et al. (2006) as “consensus recommendations” were not given. Differences in target condition and the
choice of “best test” could be the reason for differences between their figures and the estimates reported in
literature.

Combining the estimates within groups of animals and conditions into one estimate with associated
uncertainty estimates would have been preferred. With the current approach, it is also problematic that there
is a uniform weighting of test-evaluations irrespective of sample size and year of publication. The latter is
due to potential improvement of tests with time. A formal meta-analysis was not performed, primarily due to
the differences in test protocols used, particularly differences in antigen formulations and chosen cut-offs,
making it hard to justify comparisons across studies.

The numbers given in the present report may be useful not only to decision makers, but also as input
parameters in simulation studies. There are still a number of test-target condition combinations, which have
not been evaluated, particularly in deer. There is therefore still a need for further well-planned diagnostic
test-evaluations. Also, the quality of test-evaluations was inadequate for across-study comparisons also
because very few studies report the target and the study population. Improvements in planning, conducting
and reporting test-evaluations are generally required, not only for test evaluations related to MAP infection..

It is recommended to follow the guidelines given by Greiner and Gardner (2000). The ideal test-evaluation
for a set of diagnostic paratuberculosis data have still to be published for several reasons. The complicated
nature of the long, slowly developing MAP infections and the lack of good reference tests will most likely
introduce selection bias in traditional test-evaluations. Therefore, while it is relatively simple to include
covariate effects such as animal or laboratory effects using traditional methods, the resulting accuracy
estimates will most likely be biased. As an alternative, multivariable test-evaluations without selection bias
can be performed use of some latent class methods such as Bayesian analyses. However, such methods
are not always as computationally simple to perform, and the interpretation of target conditions is not always
straight forward when these models are used. For the target condition “infection”, latent-class models are
probably the best alternative, whereas for conditions “infectious” and “affected”, traditional methods may be
better used, because these conditions generally are easier detected. An ‘ideal’ test-evaluation could include
a longitudinal study design over the entire lifetime of the animals studied. The animals would have to be
tested regularly with an agent detecting test (e.g. FC), a test detecting cell-mediated immune responses (e.g. IFN-γ) and antibody reactions (e.g. serum ELISA), ultimately ending with a post-mortem histopathological evaluation of up to 100 tissues per animal to determine the infection status of the animal. Such a study would be extremely expensive and perhaps not even provide the necessary information. However, longitudinal studies and/or more frequent use of latent class methodology could provide better test-evaluations than is currently seen. Latent class models with inclusion of covariates have been used to demonstrate the improved accuracy obtained when several tests are used, i.e. reduced milk production combined with FC and milk antibody ELISA (Wang et al., 2006). Such an approach would reduce the need for division into different conditions as suggested here. The approach needs to be developed further to include repeated test data over time. At the current stage, division into detection of different conditions will however be beneficial, particularly for decision makers.

9. Conclusion

The Se and Sp of diagnostic tests for various stages of MAP infections varied significantly, but formal comparison of the different tests cannot be justified. The main reasons are variations in study design, test components and target conditions. Stratification of target conditions into those relevant for decision makers can decrease the variation of each test in each animal species, thereby improving the interpretation. However, the accuracies of the various test types appear to vary from species to species for different target definitions, therefore interpretation of diagnostic test information should be made by species, target condition and test. There is still a profound lack of reliable test evaluations, and future assessments should be conducted more stringently to allow appropriate interpretation and comparison across populations.

Acknowledgements

This study was co-funded by the European Commission within the Sixth Framework Programme, as part of the project ParaTBTools (contract no. 023106 (FOOD)).

References

dphlei-absorbed serum enzyme-linked immunosorbent assay for diagnosis of bovine paratuberculosis in


cattle. Vet. Rec. 121, 142-146.

production data to improve the ability of enzyme-linked immunosorbent assay test results to predict

Wood, P.R., 1992. A comparison of the interferon gamma assay with the absorbed ELISA for the


Cho D., Collins, M.T., 2006. Comparison of proteosomes and antigenities of secreted and cellular proteins


Clarke, C.J., Patterson, I.A., Armstrong, K.E., Low, J.C., 1996. Comparison of the absorbed ELISA and agar
gel immunodiffusion test with clinicopathological findings in ovine clinical paratuberculosis. Vet. Rec.
139, 618-621.


692.

Assoc. 229, 1912-1919.


Fig. 1. Schematic presentation of various stages of infection and their effects. This presentation may represent the typical picture, but deviations are likely to occur.
Table 1. Overview of 102 publications on evaluation of diagnostic test for paratuberculosis in animals, divided into three target conditions and animal species

<table>
<thead>
<tr>
<th>Condition</th>
<th>Cattle</th>
<th>Goats</th>
<th>Sheep</th>
<th>Llamas &amp; alpacas</th>
<th>Deer</th>
<th>Water buffaloes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Infectious</td>
<td>36</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Infected</td>
<td>18</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>Excluded</td>
<td>45</td>
<td>9</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>68</td>
</tr>
<tr>
<td>Total</td>
<td>103</td>
<td>24</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>153</td>
</tr>
</tbody>
</table>
Table 2. Sensitivity of faecal culture methods used for diagnosis of affected, infectious and infected cattle.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Test medium</th>
<th>No. with condition</th>
<th>No. test-Positive</th>
<th>Sensitivity (95% C.I.)</th>
<th>Age</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected</td>
<td>HEYM</td>
<td>56</td>
<td>39</td>
<td>0.70 (0.56, 0.81)</td>
<td>?</td>
<td>Egan et al., 1999</td>
</tr>
<tr>
<td>Infectious</td>
<td>HEYM</td>
<td>111</td>
<td>82</td>
<td>0.74 (0.65, 0.82)</td>
<td>&gt; 2 yrs</td>
<td>Sockett et al., 1992b</td>
</tr>
<tr>
<td>Infected</td>
<td>HEYM</td>
<td>177</td>
<td>86</td>
<td>0.49 (0.41, 0.56)</td>
<td>?</td>
<td>Billman-Jacobe et al., 1992</td>
</tr>
<tr>
<td>Infected</td>
<td>TREK</td>
<td>160</td>
<td>36</td>
<td>0.23 (0.16, 0.30)</td>
<td>?</td>
<td>McKenna et al., 2005</td>
</tr>
<tr>
<td>Infected</td>
<td>HEYM</td>
<td>321</td>
<td>79</td>
<td>0.25 (0.20, 0.30)</td>
<td>All cattle in herd, &gt;0 yr</td>
<td>Whitlock et al., 2000</td>
</tr>
<tr>
<td>Infected</td>
<td>HEYM</td>
<td>232</td>
<td>67</td>
<td>0.29 (0.23, 0.35)</td>
<td>All parturient cows</td>
<td>Whitlock et al., 2000</td>
</tr>
</tbody>
</table>

HEYM = Herrold’s Egg Yolk Medium; TREK = TREK ESP culture system, TREK Diagnostics, Cleveland, Ohio, USA
<table>
<thead>
<tr>
<th>Condition for</th>
<th>Test Antigen</th>
<th>Sample size*</th>
<th>Test outcome*</th>
<th>Se§</th>
<th>Sp§</th>
<th>Age</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A NA</td>
<td>Various</td>
<td>40</td>
<td>20</td>
<td>0.50</td>
<td></td>
<td>≥12 mo.</td>
<td>Bech-Nielsen et al., 1992</td>
</tr>
<tr>
<td>A NA</td>
<td>HerdChek IDEXX</td>
<td>62</td>
<td>54</td>
<td>0.87</td>
<td></td>
<td></td>
<td>Sweeney et al., 1995</td>
</tr>
<tr>
<td>A NA</td>
<td>Parachek VRI316</td>
<td>56</td>
<td>43</td>
<td>0.77</td>
<td></td>
<td></td>
<td>Egan et al., 1999</td>
</tr>
<tr>
<td>I I</td>
<td>Various</td>
<td>134</td>
<td>62</td>
<td>64</td>
<td>0.48</td>
<td>1.00</td>
<td>≥12 mo.</td>
</tr>
<tr>
<td>I I</td>
<td>HerdChek IDEXX</td>
<td>373</td>
<td>2383</td>
<td></td>
<td></td>
<td></td>
<td>Berghaus et al., 2006</td>
</tr>
<tr>
<td>I I</td>
<td>HerdChek IDEXX</td>
<td>72</td>
<td>617</td>
<td>53</td>
<td>0.74</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>HerdChek IDEXX</td>
<td>44</td>
<td>607</td>
<td>11</td>
<td>0.25</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>HerdChek IDEXX</td>
<td>174</td>
<td>62</td>
<td></td>
<td>0.36</td>
<td></td>
<td>Sweeney et al., 1995</td>
</tr>
<tr>
<td>I I</td>
<td>HerdChek IDEXX</td>
<td>41</td>
<td>263</td>
<td>13</td>
<td>0.32</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various</td>
<td>102</td>
<td>65</td>
<td>61</td>
<td>0.60</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various PPA3</td>
<td>67</td>
<td>513</td>
<td>42</td>
<td>0.63</td>
<td>0.90</td>
<td>≥2 yr</td>
</tr>
<tr>
<td>I I</td>
<td>Various PPA3</td>
<td>8</td>
<td>16</td>
<td>5</td>
<td>0.63</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various PPA3</td>
<td>177</td>
<td>196</td>
<td>71</td>
<td>0.40</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various</td>
<td>60</td>
<td>44</td>
<td>37</td>
<td>0.62</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various</td>
<td>60</td>
<td>44</td>
<td>50</td>
<td>0.83</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various</td>
<td>36</td>
<td>156</td>
<td>34</td>
<td>0.94</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various</td>
<td>14</td>
<td>210</td>
<td>10</td>
<td>0.71</td>
<td>0.83</td>
<td>&gt;6 mo</td>
</tr>
<tr>
<td>I I</td>
<td>Parachek VRI316</td>
<td>170</td>
<td>1751</td>
<td>40</td>
<td>0.24</td>
<td>0.98</td>
<td>≥2nd lact.</td>
</tr>
<tr>
<td>I I</td>
<td>Parachek VRI316</td>
<td>177</td>
<td>196</td>
<td>61</td>
<td>0.34</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>I I</td>
<td>Various</td>
<td>60</td>
<td>304</td>
<td>22</td>
<td>0.37</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>I E</td>
<td>HerdChek IDEXX</td>
<td>415</td>
<td>359</td>
<td>127</td>
<td>0.31</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>I E</td>
<td>HerdChek IDEXX</td>
<td>198</td>
<td>346</td>
<td>48</td>
<td>0.24</td>
<td>1.00</td>
<td>&gt;2 yr</td>
</tr>
<tr>
<td>I E</td>
<td>Various</td>
<td>15</td>
<td>100</td>
<td>6</td>
<td>0.40</td>
<td>0.91</td>
<td>2 to 15 yr</td>
</tr>
<tr>
<td>I E</td>
<td>Various PPA3</td>
<td>64</td>
<td>68</td>
<td>51</td>
<td>0.80</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>I E</td>
<td>IDEXX Scand Various</td>
<td>198</td>
<td>346</td>
<td>66</td>
<td>322</td>
<td>0.33</td>
<td>0.93</td>
</tr>
<tr>
<td>I E</td>
<td>IDEXX Scand Various</td>
<td>198</td>
<td>346</td>
<td>51</td>
<td>335</td>
<td>0.26</td>
<td>0.97</td>
</tr>
<tr>
<td>I E</td>
<td>Pourquier Various</td>
<td>415</td>
<td>359</td>
<td>116</td>
<td>359</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>I E</td>
<td>SERELISA Various</td>
<td>301</td>
<td>359</td>
<td>134</td>
<td>304</td>
<td>0.45</td>
<td>0.85</td>
</tr>
<tr>
<td>I E</td>
<td>Various</td>
<td>156</td>
<td>200</td>
<td>126</td>
<td>0.81</td>
<td>1.00</td>
<td>&gt;15 mo</td>
</tr>
<tr>
<td>I E</td>
<td>Parachek VRI316</td>
<td>126</td>
<td>196</td>
<td>71</td>
<td>0.56</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>I E</td>
<td>Parachek VRI316</td>
<td>415</td>
<td>359</td>
<td>118</td>
<td>358</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>I E</td>
<td>Parachek VRI316</td>
<td>64</td>
<td>68</td>
<td>51</td>
<td>0.80</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>I NA</td>
<td>Various</td>
<td>106</td>
<td>341</td>
<td>50</td>
<td>340</td>
<td>0.47</td>
<td>1.00</td>
</tr>
<tr>
<td>I NA</td>
<td>Parachek VRI316</td>
<td>106</td>
<td>341</td>
<td>33</td>
<td>334</td>
<td>0.31</td>
<td>0.98</td>
</tr>
<tr>
<td>E E</td>
<td>HerdChek IDEXX</td>
<td>160</td>
<td>834</td>
<td>14</td>
<td>814</td>
<td>0.09</td>
<td>0.98</td>
</tr>
<tr>
<td>E E</td>
<td>Various</td>
<td>22</td>
<td>378</td>
<td>4</td>
<td>363</td>
<td>0.18</td>
<td>0.96</td>
</tr>
<tr>
<td>E E</td>
<td>SVANOVA Various</td>
<td>160</td>
<td>834</td>
<td>27</td>
<td>757</td>
<td>0.17</td>
<td>0.91</td>
</tr>
<tr>
<td>E E</td>
<td>Parachek VRI316</td>
<td>160</td>
<td>834</td>
<td>11</td>
<td>801</td>
<td>0.07</td>
<td>0.96</td>
</tr>
<tr>
<td>E NA</td>
<td>Parachek VRI316</td>
<td>1188</td>
<td>265</td>
<td></td>
<td></td>
<td>0.22</td>
<td>&gt;2 yrs</td>
</tr>
<tr>
<td>NA E</td>
<td>Parachek VRI316</td>
<td>15566</td>
<td>15467</td>
<td></td>
<td></td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>NA E</td>
<td>Parachek VRI316</td>
<td>5588</td>
<td>5579</td>
<td></td>
<td>1.00</td>
<td></td>
<td>Jubb and Galvin, 2004</td>
</tr>
<tr>
<td>NA E</td>
<td>Parachek VRI316</td>
<td>1049</td>
<td>1028</td>
<td></td>
<td>0.98</td>
<td></td>
<td>Pitt et al., 2002</td>
</tr>
</tbody>
</table>

§ Sensitivity; Sp=Specificity; A=Affected; I=Infectious; E=Infected; NA=Not available.
* Sample sizes for animals with condition (C+) and without condition (C-)
† No. of test-positive (T+) or test-negative (T-) given the condition (C)
Table 4. Sensitivity and specificity estimates from studies of milk antibody ELISAs for detection of infectious (I) and infected (E) cattle

<table>
<thead>
<tr>
<th>Condition for</th>
<th>Test</th>
<th>Antigen</th>
<th>Sample size</th>
<th>Test outcome</th>
<th>Se</th>
<th>Sp</th>
<th>Age</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I</td>
<td>Antel$^b$</td>
<td>Alli$^d$d</td>
<td>72  617</td>
<td>44</td>
<td>584</td>
<td>0.61</td>
<td>0.95</td>
</tr>
<tr>
<td>I</td>
<td>I</td>
<td>Various</td>
<td>Alli$^d$</td>
<td>67  513</td>
<td>36</td>
<td>487</td>
<td>0.54</td>
<td>0.95</td>
</tr>
<tr>
<td>I</td>
<td>I</td>
<td>Antel$^b$</td>
<td>Alli$^d$</td>
<td>170 1751</td>
<td>36</td>
<td>1724</td>
<td>0.21</td>
<td>0.98</td>
</tr>
<tr>
<td>I</td>
<td>I</td>
<td>Various</td>
<td>LAM$^a$</td>
<td>102 65</td>
<td>61</td>
<td>54</td>
<td>0.60</td>
<td>0.83</td>
</tr>
<tr>
<td>I</td>
<td>E</td>
<td>Antel$^b$</td>
<td>Alli$^d$</td>
<td>364 352</td>
<td>105</td>
<td>351</td>
<td>0.29</td>
<td>1.00</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Various</td>
<td>Alli$^d$</td>
<td>2662 comb.</td>
<td>0.39</td>
<td>0.96</td>
<td></td>
<td>Nielsen et al., 2002</td>
</tr>
</tbody>
</table>

$^a$) Se= Sensitivity; Sp=Specificity; I=Infectious; E=Infected.

$^b$) Sample sizes for animals with condition (C+) and without condition (C-)

$^c$) No. of test-positive (T+) or test-negative (T-) given the condition (C)

$^d$) Antel Biosystems Inc., Lansing, Michigan, USA
Table 5. Sensitivity and specificity estimates from studies of interferon-$\gamma$ tests for detection of infectious (I) and infected (E) cattle

<table>
<thead>
<tr>
<th>Condition for Sample size</th>
<th>Test</th>
<th>Anti-</th>
<th>Test outcome</th>
<th>Se</th>
<th>Sp</th>
<th>Age</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I</td>
<td>Various</td>
<td>PPDa</td>
<td>8</td>
<td>16</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>I</td>
<td>E</td>
<td>Bovigam</td>
<td>PPDj</td>
<td>8</td>
<td>53</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>I</td>
<td>E</td>
<td>Bovigam</td>
<td>PPDj</td>
<td>13</td>
<td>65</td>
<td>11</td>
<td>50</td>
</tr>
<tr>
<td>I</td>
<td>E</td>
<td>Bovigam</td>
<td>PPDj</td>
<td>28</td>
<td>65</td>
<td>21</td>
<td>14</td>
</tr>
</tbody>
</table>

*Se= Sensitivity; Sp=Specificity; I=Infectious; E=Infected.

^a) Sample sizes for animals with condition (C+) and without condition (C-)

^b) No. of test-positive (T+) or test-negative (T-) given the condition (C)
Table 6. Sensitivity and specificity estimates from studies of serum antibody ELISAs for detection of affected (A), infectious (I) and infected (E) goats

<table>
<thead>
<tr>
<th>Condition for</th>
<th>Test</th>
<th>Antigen</th>
<th>Sample size</th>
<th>Test outcome</th>
<th>Se</th>
<th>Sp</th>
<th>Age</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>HerdChek IDEXX</td>
<td>44 62 36</td>
<td>59</td>
<td>0.82 0.95 ?</td>
<td>Dimareli-Malli et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>Various Various</td>
<td>44 62 40</td>
<td>49</td>
<td>0.91 0.79 ?</td>
<td>Dimareli-Malli et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>Various Various</td>
<td>44 62 38</td>
<td>57</td>
<td>0.86 0.92 ?</td>
<td>Dimareli-Malli et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>Various Various</td>
<td>44 62 40</td>
<td>57</td>
<td>0.91 0.92 ?</td>
<td>Dimareli-Malli et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>Parachek VRI316</td>
<td>15 11 13</td>
<td>11</td>
<td>0.87 1.00 ?</td>
<td>Milner et al., 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>I</td>
<td>Various Various</td>
<td>16 63 14</td>
<td>59</td>
<td>0.88 0.94 ?</td>
<td>Molina et al., 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>I</td>
<td>Various PPA3</td>
<td>17 63 15</td>
<td>60</td>
<td>0.88 0.95 ?</td>
<td>Molina et al., 1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>Various PPA3</td>
<td>35 61 35</td>
<td>57</td>
<td>1.00 0.93 ?</td>
<td>Molina Cabellero et al., 1993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>E</td>
<td>HerdChek IDEXX</td>
<td>35 123 19</td>
<td>123</td>
<td>0.54 1.00 ?</td>
<td>Burnside &amp; Rowley, 1994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>NA</td>
<td>Various Various</td>
<td>36 33</td>
<td></td>
<td>0.92 ?</td>
<td>Tripathi et al., 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Pourquier Various</td>
<td>36 945 28</td>
<td>945</td>
<td>0.78 1.00 &gt; 1 yr</td>
<td>Gumber et al., 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>HerdChek IDEXX</td>
<td>368 combined</td>
<td></td>
<td>0.63 0.95 &gt; 1 yr</td>
<td>Kostoulas et al., 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Parachek VRI316</td>
<td>19 1000 16</td>
<td>997</td>
<td>0.84 1.00 ?</td>
<td>Whittington et al., 2003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>HerdChek IDEXX</td>
<td>47 1000 39</td>
<td>995</td>
<td>0.83 1.00 ?</td>
<td>Whittington et al., 2003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

§) Se= Sensitivity; Sp=Specificity; A=Affected; I=Infectious; E=Infected; NA=Not available.

#) Sample sizes for animals with condition (C+) and without condition (C-)

") No. of test-positive (T+) or test-negative (T-) given the condition (C)
Table 7. Sensitivity and specificity estimates from studies of serum antibody ELISAs for detection of affected (A), infectious (I) and infected (E) sheep

<table>
<thead>
<tr>
<th>Condition for</th>
<th>Test</th>
<th>Antigen</th>
<th>Sample size</th>
<th>Test outcome</th>
<th>Se</th>
<th>Sp</th>
<th>Age</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C+</td>
<td>C-</td>
<td>T+</td>
<td>C+</td>
<td>T-</td>
<td>C-</td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>Parachek</td>
<td>VRI316</td>
<td>32</td>
<td>43</td>
<td>20</td>
<td>42</td>
<td>0.63</td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>Various</td>
<td>Various</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0.83</td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>Parachek</td>
<td>VRI316</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0.83</td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>Parachek</td>
<td>VRI316</td>
<td>59</td>
<td>253</td>
<td>50</td>
<td>252</td>
<td>0.85</td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>Various</td>
<td>Various</td>
<td>59</td>
<td>252</td>
<td>21</td>
<td>247</td>
<td>0.36</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Parachek</td>
<td>VRI316</td>
<td>120</td>
<td>1137</td>
<td>53</td>
<td>1125</td>
<td>0.44</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Various</td>
<td>Various</td>
<td>368 combined</td>
<td>0.37</td>
<td>0.97</td>
<td>&gt; 1 yr</td>
<td>Kostoulas et al., 2006</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Parachek</td>
<td>VRI316</td>
<td>2465 combined</td>
<td>0.16</td>
<td>0.98</td>
<td>&gt; 1 yr</td>
<td>Robbe-Austerman et al., 2006</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Various</td>
<td>VRI316</td>
<td>224</td>
<td>1748</td>
<td>93</td>
<td>1661</td>
<td>0.42</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>Various</td>
<td>VRI316</td>
<td>224</td>
<td>1748</td>
<td>49</td>
<td>1731</td>
<td>0.22</td>
</tr>
</tbody>
</table>

*) Se= Sensitivity; Sp= Specificity; A=Affected; I=Infectious; E=Infected; NA=Not available.

b) Sample sizes for animals with condition (C+) and without condition (C-)

** No. of test-positive (T+) or test-negative (T-) given the condition (C)
| Table 8 | Summary (range) of reported sensitivities and specificities of faecal culture (FC), serum antibody ELISA (SELISA), milk antibody ELISA (MELISA) and interferon-γ tests for diagnosis of three stages of infection with *Mycobacterium avium* subsp. *paratuberculosis* in cattle, deer, goats and sheep.

<table>
<thead>
<tr>
<th></th>
<th>Sensitivities (range)</th>
<th>Specificities (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FC SELISA MELISA IFN-γ</td>
<td>FC SELISA MELISA IFN-γ</td>
</tr>
<tr>
<td><strong>Cattle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affected</td>
<td>0.70^(#1) 0.50; 0.87^(#3)</td>
<td>None None 1.0^(§) None None</td>
</tr>
<tr>
<td>Infectious</td>
<td>0.74^(#1) 0.24; 0.94^(#30) 0.21; 0.61^(#5)</td>
<td>0.13; 0.85^(#4) 1.0^(§) 0.40; 1.0^(#15) 0.83; 0.99^(#2) 0.88^(#3)</td>
</tr>
<tr>
<td>Infected</td>
<td>0.23;0.29^(#4) 0.07; 0.22^(#5) 0.39^(#1)</td>
<td>None 0.98^(#1) 0.85; 1.0^(#19) 0.96; 1.0^(#2) 0.94; 0.95^(#3)</td>
</tr>
<tr>
<td><strong>Deer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affected</td>
<td>None 0.83; 0.85^(#4)</td>
<td>None None 1.0^(§) 0.98; 1.0^(#2) None None</td>
</tr>
<tr>
<td>Infectious</td>
<td>None None None None</td>
<td>None None None None</td>
</tr>
<tr>
<td>Infected</td>
<td>None None None None</td>
<td>None None None None</td>
</tr>
<tr>
<td><strong>Goats</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affected</td>
<td>None 0.82; 1.0^(#9)</td>
<td>None None 1.0^(§) 0.79; 1.0^(#3) None None</td>
</tr>
<tr>
<td>Infectious</td>
<td>None 0.54; 0.92^(#2)</td>
<td>None None 1.0^(§) 0.94; 0.95^(#2) None None</td>
</tr>
<tr>
<td>Infected</td>
<td>0.08^(#1) 0.63; 0.84^(#4)</td>
<td>None None 0.98^(#1) 0.93; 1.0^(#3) None None</td>
</tr>
<tr>
<td><strong>Sheep</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affected</td>
<td>None 0.36; 0.85^(#30)</td>
<td>None None 1.0^(§) None None None</td>
</tr>
<tr>
<td>Infectious</td>
<td>None None None None</td>
<td>None None None None</td>
</tr>
<tr>
<td>Infected</td>
<td>0.16^(#1) 0.16; 0.44^(#5)</td>
<td>None None 0.97^(#1) 0.95; 1.0^(#10) None None</td>
</tr>
</tbody>
</table>

^(#) Number given is the number of test-evaluations included in the summary. One study (Billman-Jacobe et al., 1991) was excluded from group FC, Infected, Cattle because of selection procedure of animals.

^(§) Specificity is 1.0 by definition