Animal Health Research Reviews

cambridge.org/ahr

Systematic Review

Cite this article: Buczinski S, Arsenault J, Kostoulas P, Corbière F, Fecteau G, Dendukuri N (2019). Accuracy of paratuberculosis diagnostic tests in small ruminants: protocol for a systematic review and meta-analysis. *Animal Health Research Reviews* **20**, 98–102. https://doi.org/10.1017/S1466252319000082

Received: 13 May 2019
Revised: 26 June 2019
Accepted: 8 July 2019
First published online: 19 November 2019

Author for correspondence:
S. Buczinski, Faculté de médecine vétérinaire, Université de Montréal, Québec Canada.
E-mail: s.buczinski@umontreal.ca

© Cambridge University Press 2019

CAMBRIDGE UNIVERSITY PRESS

Accuracy of paratuberculosis diagnostic tests in small ruminants: protocol for a systematic review and meta-analysis

S. Buczinski[1], J. Arsenault[1], P. Kostoulas[2], F. Corbière[3], G. Fecteau[1] and N. Dendukuri[4]

[1]Faculté de médecine vétérinaire, Université de Montréal, Québec Canada; [2]Faculty of Veterinary Medicine, University of Thessaly, Volos, Greece; [3]UMR INRA ENVT 1225 IHAP, Ecole Nationale Vétérinaire de Toulouse, Toulouse, France and [4]Department of Medicine, McGill University, Technology Assessment Unit, McGill University Health Centre, Montréal, Québec, Canada

Abstract

Paratuberculosis is a worldwide infectious disease caused by *Mycobacterium avium* ssp. *paratuberculosis* (MAP). Various ruminant species can be affected by the disease, and the diagnosis of the disease is challenging in the absence of a gold standard test. The aim of this systematic review protocol is to determine the accuracy of the direct and indirect diagnostic tests for MAP infection with a special focus on sheep and goats.

Background information

Paratuberculosis is a chronic enteric disease of ruminants with a worldwide distribution. It is caused by *Mycobacterium avium* subspecies *paratuberculosis* (MAP) (Windsor, 2015). The disease is mostly characterized in small ruminants by a slow progression. Weight loss associated with malabsorption and, in some cases, end-stage diarrhea are the most common clinical signs. Premature culling and death can also be observed as a consequence of the disease. The detection of infected animals is challenging because of the relatively poor performance of diagnostic tests in the early to mid-stage infection (Windsor, 2015). More recently, it has been suggested that the resilience of some animals can be another challenge of disease (de Silva *et al*., 2018). Although resilient animals can present a different immune response compared to others (which may facilitate the diagnosis of infection), they are most challenging for disease control, as they can be infected but can clear the infection and probably should be kept in the herd.

The infection can either be diagnosed by tests that directly target MAP detection (e.g. culture or polymerase chain reaction (PCR) detection of MAP in feces or biopsy samples) or indirect tests targeting immune response to MAP (e.g. enzyme-linked immunosorbent assay (ELISA) test in serum or milk samples or interferon-γ (IFNγ) test on blood samples). There are important differences in the diagnosis of MAP infection between sheep versus goats. The MAP strains isolated from sheep are generally 'S' (sheep) type, whereas goats can be infected by either 'S' or 'C' (cattle) strains. One of the main clinical implications of the presence of the two strains is that 'S' strain is more difficult to culture than 'C' strain which can lead to false-negative results (Fawzy *et al*., 2018). Moreover, the difference in the immune response of the two species may at least partially explain the difference in the accuracy of indirect tests such as ELISA (Bauman *et al*., 2016*b*). For these reasons, it is important to consider sheep and goats separately in the approaches used to estimate test accuracy. The presence of MAP in feces can also be due to passive shedding or 'pass-through' phenomenon (Whittington and Sergeant, 2001) that can lead to a false-positive test result. However, this phenomenon does not seem to be as prevalent in sheep and goats compared to cattle (Marquetoux *et al*., 2018).

The indirect tests have the advantage of being relatively inexpensive and rapid but with the most important limitations of becoming positive in later stages of infection (except for resilient animals) and therefore at a higher risk of false-negative results. On the other side, they can also lead to false-positive results in cases of atypical mycobacterial exposure (Gilardoni *et al*., 2012).

In that specific context, it is of interest to determine the accuracy (sensitivity and specificity) of available tests for the diagnosis of small ruminant MAP infection using an evidence-based approach. Diagnostic tests can either be used at the individual level or at the farm (or group) level with the possibility of pooling samples of different individuals (e.g. feces or milk) or using environmental samples. We will focus on tests used at the individual level in this study.

The diagnostic test accuracy assessment is challenging when compiling the results from different studies due to multiple sources of variations. Differences obtained for sensitivity and

specificity of the tests in different studies may be associated with random biological variation but also due to the bivariate nature of diagnostic test studies (sensitivity and specificity are determined in two different populations). For example, in a two-gate design (Rutjes *et al*., 2005) where specificity is assessed in healthy animals and sensitivity is assessed in advanced stages of the disease, the accuracy of the test may be overestimated (Leeflang *et al*., 2013). Other covariates studied or types of biases based on study design may also influence test accuracy estimation. Moreover, not adequately accounting for the absence of a gold standard test could also lead to misinterpretation of some test results (Branscum *et al*., 2005; Rutjes *et al*., 2007). As an example, when evaluating a new test (index test) using an imperfect reference standard test, discordant results can be falsely considered as an error of the index test if not accounting for possible misclassification of the reference test. Misclassification due to the imperfect reference standard test therefore needs to be considered. The Bayesian latent class modeling approach has gained popularity in veterinary medicine due to the flexibility to account for misclassification of the pseudo-gold standard or reference test. The latent class models can easily handle imperfect reference standard tests and are commonly implemented for the diagnosis of different infectious diseases of ruminants including paratuberculosis (Gardner *et al*., 2000; Kostoulas *et al*., 2006; Bauman *et al*., 2016*a*).

Objectives

The objective of this study was to perform a systematic review of the literature on the accuracy of different diagnostic tests used at the individual level for the diagnosis of MAP infection in sheep and goats. Both species will be separated throughout the manuscript because of inherent differences in MAP strains and immune response. The different targeted index tests were identified as ELISA (in the serum or milk samples), Agar Gelose ImmunoDiffusion (AGID) tests in the serum, IFNγ in the blood and direct fecal or tissue PCR in sheep and goats, respectively, using direct fecal or tissue culture or histopathology and acid-fast staining as imperfect reference standard tests (Fig. 1). The protocol for this systematic review and meta-analysis is reported following the Cochrane Diagnostic Test Accuracy taskforce recommendations (https://methods.cochrane.org/sdt/handbook-dta-reviews).

Methods

Eligibility criteria

Studies reporting at least one reference standard test and one of the different diagnostic tests or ≥2 index tests in sheep or goat will be eligible to be included in the systematic review. The study design will either be cross-sectional or longitudinal where at least one index test is compared with one or more reference standard tests at single or multiple time points. Studies comparing two or more index tests will also be included allowing for possible network-based latent class model analysis (Menten and Lesaffre, 2015). We will include either naturally occurring MAP infection studies or studies with the experimental oral challenge of MAP. When the same animals had a serial assessment of diagnostic tests, the different test results within time will be recorded separately in the spreadsheet. One randomly selected dataset will be used for meta-analytic methods. One-gate (where a population or a random sample of a population is sampled) or two-gate studies (where two different target populations are sampled, e.g. healthy animals versus sick animals) will be eligible to be included. The study design will be recorded as a potential covariate influencing test accuracy (Rutjes *et al*., 2005). Experimental versus naturally occurring infection will also be recorded as a potential covariate of interest.

Target conditions being diagnosed

The target condition is an active infection by MAP in either sheep or goat (latent status that cannot be directly assessed). Different stages of infection exist due to slow progression of the disease from infection to clinical manifestations of the disease and eventual death (Windsor, 2015). The stage of infection will be noted based on specific study characteristics (e.g. study focusing on sick animals (with clinical signs associated with paratuberculosis or other health problems versus subclinically infected animals (i.e. animal with no clinical anomalies) versus random sample of a population).

Reference standard tests

The reference standard used to diagnose the target condition is generally based on direct detection of MAP by either fecal or tissue culture using specific media for mycobacteria. Histopathology and acid-fast staining can also be used for the diagnosis. Both tests can be considered with a high although not perfect specificity (low risk of false positive associated with pass through phenomenon or other mycobacteria) but false-negative results commonly occur due to either absent/low/transient shedding in the early stages of infection (Gilardoni *et al*., 2012).

Index tests

There are multiple index tests that can be used for the diagnosis of MAP-infected animals. The main index tests we propose to include are tests that can be rapidly done in either blood, milk, or fecal samples. The ELISA and AGID tests for serum samples and ELISA on milk samples as well as IFNγ test on blood and fecal PCR are targeted index tests for that systematic review. Different manufacturers may provide these tests and different thresholds can be used to define positivity in different studies as well as within the same study. This information will be recorded as well as the results obtained using different thresholds within the same study.

Clinical pathway to use the index tests

The index tests are generally used as screening tests because of the rapidity and easiness to be processed and their lower cost than direct MAP identification methods. Milk samples are convenient samples in milking goats and ewes, whereas blood sample are more difficult to obtain. The tests can either be used at the individual level or population (herd/flock) level. Most of the time, they are used at the population level as a proxy to estimate the prevalence of the disease in herds or geographical areas (Liapi *et al*., 2011). We will focus on the use of these tests at the individual level.

Criteria for considering studies for this review

Participants

The unit of interest will be the individual animal. Animals will either be domesticated sheep or goats, either male or female.

https://doi.org/10.1017/S1466252319000082 Published online by Cambridge University Press

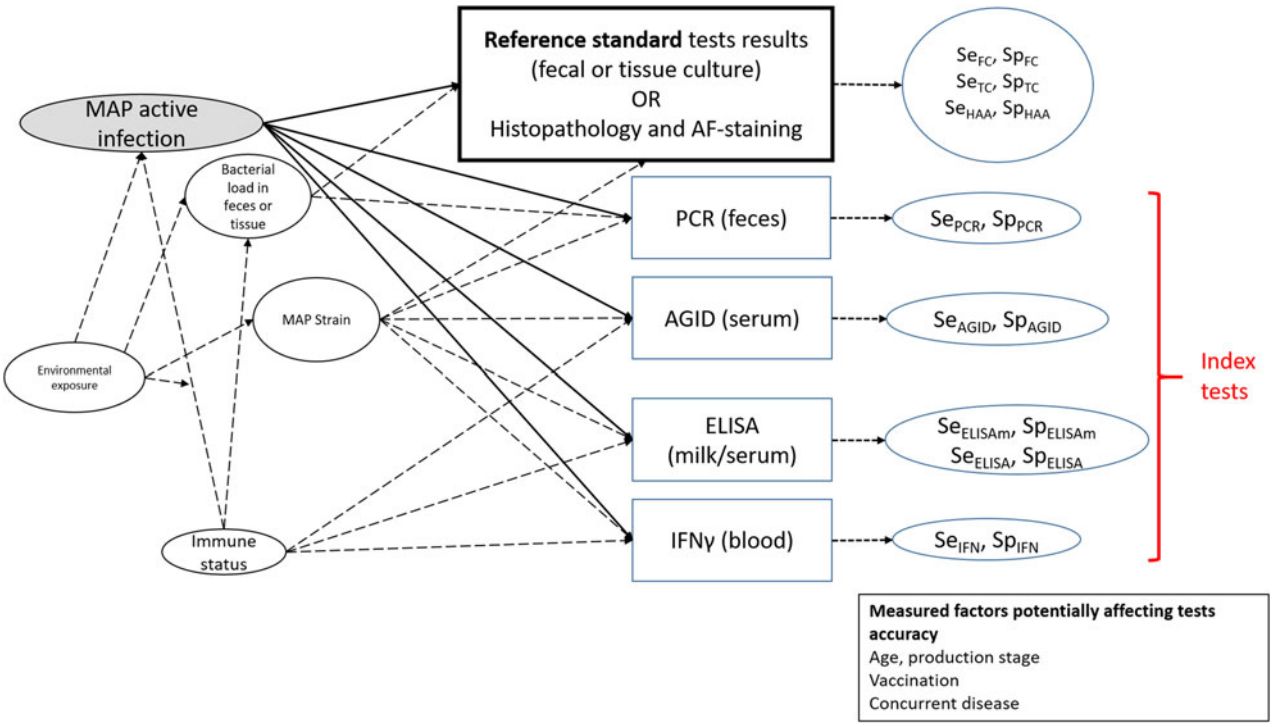


Fig. 1. Model specifying prior beliefs of relations between latent and observed variables for diagnostic test accuracy in small ruminant paratuberculosis. The latent variable in the grey oval is MAP active infection status of the animal. Other latent variables that may impact test accuracy (indicated as dotted lines) are environmental exposure, bacterial load in the feces, or tissue as well as the MAP strain ('C' versus 'S' strain). The immune status of the animal may also interfere with indirect and direct tests. The variables of interest are the index test accuracy (which is determined by the bivariate Se (sensitivity) and Sp (specificity)). The measured covariates that may impact test accuracy are detailed in a specific box. FC, fecal culture; TC, tissue culture; HAA, histopathology and acid-fast staining; Em, ELISA on milk; E, ELISA on serum; AGID, Agar Gelose ImmunoDiffusion; A, AGID on serum; PCR, polymerase chain reaction; IFN, interferon γ.

Studies where the distinction between sheep and goat results is not possible will be excluded due to the difference explained above (see background information). Studies on wild ruminants will also be excluded. No geographic limitations will restrict the search but this information will be noted.

Search strategy for identification of studies

The identification of potential studies will be performed using two different electronic databases: PubMed and CAB Abstracts. The grey literature will also be assessed using Google-Scholar search and Pro-Quest Dissertation and Thesis Global database. The distinction between peer-reviewed articles, congress abstracts, and non-peer-reviewed articles will be noted, but studies will not be excluded based on this characteristic. The exact search strategy is indicated in the Appendix. We will use a previous search strategy used for the systematic review of a diagnostic test which has been shown to be highly sensitive (98%) and with 74% specificity (Relevo, 2012). For disease identification, the same search strategy that previously reported for the systematic review in cattle paratuberculosis will be used (Rangel *et al.*, 2015). Finally, species filters will be used as previously reported in the small ruminant meta-analysis (Saccareau *et al.*, 2017).

Data collection and analysis

Selection of studies

The studies will be selected following a two-step screening process. At each step, the first 20 potential references will be assessed simultaneously by two different persons in the same room in order of consistency for the full screening method (with at least ten retained papers by the corresponding author). At the first step, the titles and abstracts of the studies will be used to exclude studies irrelevant to the systematic review due to imperfect specificity of the research. Two different reviewers will perform the screening of articles in parallel. The agreement between the reviewers will be calculated using Cohen's κ statistic. Any disagreement will be solved by either consensus after face-to-face discussion or a third resolver rater. For the second step, all eligible studies will then be entirely read by two reviewers. Studies will ultimately be retained if direct or indirect information allowing the extraction of animal cross-classification using at least one of the previous index tests and a reference standard test is presented in the study results or supplemental files. A specific restriction to English or French written full-text documents will be applied. At each step of the review, the specific reason for exclusion will be recorded. Relevant studies where full text is not available will also be excluded due to limited information at the abstract level.

Data extraction and management

The data extraction will be performed using a standardized spreadsheet (Excel for Windows, Microsoft, Redmond, WA). General information including authors, journal, date of publication, country where the study was done as well as study design (one-gate versus two-gate study) will be noted. Information concerning the species and breeds of interest, age and physiological status of the animals tested, and type of production (meat, milk, mixed, or fiber) will be noted. The specific information on the

https://doi.org/10.1017/S1466252319000082 Published online by Cambridge University Press

type of sample use and diagnostic tests used will be recorded including manufacturer and any specific threshold, target gene, or antigen information. Because vaccination against paratuberculosis is part of the strategy to control the disease, the vaccinal status of the animals will also be noted as a potential covariate that may affect test accuracy.

Assessment of methodological quality

The methodological quality will be assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2 tool) validated for systematic reviews of diagnostic test accuracy studies (Whiting and Davenport, 2018). The QUADAS-2 tool assesses the risk of bias and applicability concerns of selected studies. Two different operators will appraise the studies' quality after a simultaneous pilot assessment study of five retained articles. Any discrepancy between the two authors that cannot be solved by face-to-face discussion will be submitted to a third author.

Statistical analysis and data synthesis

The statistical analyses will be performed using a hierarchical model of receiving operating characteristics (HSROC) in a Bayesian framework (Schiller and Dendukuri, 2015). The reference standard tests will be either fecal culture, tissue culture, or histopathology combined to acid-fast staining. Since it is difficult to know *a priori* the exact number of studies where two or more index tests were assessed without the use of any reference test, we would potentially use a network-based latent class model analysis as recently investigated for dealing with these practical issues (Menten and Lesaffre, 2015). As these modeling alternatives are relatively new and not extensively validated, we propose to use this approach as a secondary statistical analysis method.

Sensitivity analysis

Sensitivity analysis will be performed using conditional dependence between index test results and reference standard test results. The sensitivity analysis of the different models will also be performed using different prior information for the reference standard tests. A further step of sensitivity analysis would be to compare the two complementary statistical approaches described in the previous paragraph.

Investigations of heterogeneity

Assessment heterogeneity will be performed running the main model analyses after excluding studies at high risk of bias.

Comparison of different tests for the diagnosis of small ruminant paratuberculosis

The implication of these review findings will then be applied using a specific summary of findings tables and graphics (Buczinski *et al.*, 2018; Whiting and Davenport, 2018). In a fictive population of 1000 sheep or goats for the test, accuracy and associated uncertainty will be used for classifying animals as true-positive, true-negative, false-positive, and false-negative cases based on different relevant prevalence scenarios. Construction of different scenarios at the herd level to detect infected herds will also be used with a stochastic modeling approach.

Conclusions

At the end of this study, we hope to have extensive knowledge of individual diagnostic tests that can be used for the diagnosis of MAP infection in small ruminants at the individual and herd/flock level. We would also point out the strengths and weaknesses of available evidence based on our critical appraisal process.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1466252319000082

Author contributions. SB is the main author and guarantor of the study. He provides the first draft of the protocol. All authors have been significantly involved in the preparation, reviewing, and suggestions concerning the current systematic review protocol.

Financial support. This review received no specific source of funding. None of the authors had any specific conflict of interest regarding the current study.

References

Bauman CA, Jones-Bitton A, Jansen J, Kelton D and Menzies P (2016*a*) Evaluation of fecal culture and fecal RT-PCR to detect *Mycobacterium avium* ssp. *paratuberculosis* fecal shedding in dairy goats and dairy sheep using latent class Bayesian modeling. *BMC Veterinary Research* **12**, 212.

Bauman CA, Jones-Bitton A, Menzies P, Toft N, Jansen J and Kelton D (2016*b*) Prevalence of paratuberculosis in the dairy goat and dairy sheep industries in Ontario, Canada. *Canadian Veterinary Journal* **57**, 169–175.

Branscum AJ, Gardner IA and Johnson WO (2005) Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Preventive Veterinary Medicine* **68**, 145–163.

Buczinski S, Gicquel E, Fecteau G, Takwoingi Y, Chigerwe M and Vandeweerd JM (2018) Systematic review and meta-analysis of diagnostic accuracy of serum refractometry and brix refractometry for the diagnosis of inadequate transfer of passive immunity in calves. *Journal of Veterinary Internal Medicine* **32**, 474–483.

de Silva K, Plain K, Purdie A, Begg D and Whittington R (2018) Defining resilience to mycobacterial disease: characteristics of survivors of ovine paratuberculosis. *Veterinary Immunology and Immunopathology* **195**, 56–64.

Fawzy A, Zschöck M, Ewers C and Eisenberg T (2018) Genotyping methods and molecular epidemiology of *Mycobacterium avium* subsp. *paratuberculosis* (MAP). *International Journal of Veterinary Science and Medicine* **6**, 258–264.

Gardner IA, Stryhn H, Lind P and Collins MT (2000) Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Preventive Veterinary Medicine* **45**, 107–122.

Gilardoni LR, Paolicchi FA and Mundo SL (2012) Bovine paratuberculosis: a review of the advantages and disadvantages of different diagnostic tests. *Revista Argentina Microbiologica* **44**, 201–215.

Kostoulas P, Leontides L, Enoe C, Billinis C, Florou M and Sofia M (2006) Bayesian estimation of sensitivity and specificity of serum ELISA and faecal culture for diagnosis of paratuberculosis in Greek dairy sheep and goats. *Preventive Veterinary Medicine* **76**, 56–73.

Leeflang MM, Rutjes AW, Reitsma JB, Hooft L and Bossuyt PM (2013) Variation of a test's sensitivity and specificity with disease prevalence. *Canadian Medical Association Journal* **185**, E537–E544.

Liapi M, Leontides L, Kostoulas P, Botsaris G, Iacovou Y, Rees C, Georgiou K, Smith G and Naseby D (2011) Bayesian estimation of the true prevalence of *Mycobacterium avium* subsp. *paratuberculosis* infection in Cypriot dairy sheep and goat flocks. *Small Ruminant Research* **95**, 174–178.

Marquetoux N, Mitchell R, Ridler A, Heuer C and Wilson P (2018) A synthesis of the patho-physiology of *Mycobacterium avium* subspecies *paratuberculosis* infection in sheep to inform mathematical modelling of ovine paratuberculosis. *Veterinary Research* **49**, 27.

Menten J and Lesaffre E (2015) A general framework for comparative Bayesian meta-analysis of diagnostic studies. *BMC Medical Research Methodology* **15**, 70.

https://doi.org/10.1017/S1466252319000082 Published online by Cambridge University Press

Rangel SJ, Pare J, Dore E, Arango JC, Cote G, Buczinski S, Labrecque O, Fairbrother JH, Roy JP, Wellemans V and Fecteau G (2015) A systematic review of risk factors associated with the introduction of *Mycobacterium avium* spp. *paratuberculosis* (MAP) into dairy herds. *Canadian Veterinary Journal* **56**, 169–177.

Relevo R (2012) Effective search strategies for systematic reviews of medical tests. In Chang SM, Matchar DB, Smetana GW and Umscheid CA (eds), *Methods Guide for Medical Test Reviews*. Rockville, MD, pp. 28–32.

Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS and Bossuyt PM (2005) Case-control and two-gate designs in diagnostic accuracy studies. *Clinical Chemistry* **51**, 1335–1341.

Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS and Bossuyt PM (2007) Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment (Winchester, England)* **11**, iii, ix–51.

Saccareau M, Sallé G, Robert-Granié C, Duchemin T, Jacquiet P, Blanchard A, Cabaret J and Moreno CR (2017) Meta-analysis of the parasitic phase traits of *Haemonchus contortus* infection in sheep. *Parasites & Vectors* **10**, 201.

Schiller I and Dendukuri N (2015) HSROC: an R package for Bayesian meta-analysis of diagnostic test accuracy.

Whiting P and Davenport C (2018) Understanding test accuracy research: a test consequence graphic. *Diagnostic and Prognostic Research* **2**, 2.

Whittington R and Sergeant E (2001) Progress towards understanding the spread, detection and control of *Mycobacterium avium* subsp *paratuberculosis* in animal populations. *Australian Veterinary Journal* **79**, 267–278.

Windsor PA (2015) Paratuberculosis in sheep and goats. *Veterinary Microbiology* **181**, 161–169.

https://doi.org/10.1017/S1466252319000082 Published online by Cambridge University Press