Journal of Veterinary Internal Medicine
Open Access

J Vet Intern Med 2016;30:1356–1368

Systematic Review of the Diagnostic Accuracy of Haptoglobin, Serum Amyloid A, and Fibrinogen versus Clinical Reference Standards for the Diagnosis of Bovine Respiratory Disease

A. Abdallah, J. Hewson, D. Francoz, H. Selim, and S. Buczinski

Background: Bovine respiratory disease (BRD) is a worldwide animal health concern especially in feedlot, dairy, and veal calves. One of the greatest challenges is the absence of a gold standard for achieving an accurate *antemortem* diagnosis. Various blood markers, including the acute-phase proteins (AAP), have been proposed as potential valuable tools for BRD diagnosis.

Objectives: To perform a systematic review of the literature to assess the accuracy of selected APP (haptoglobin [Hp], serum amyloid A [SAA], and fibrinogen [Fb]) as diagnostic tools for cattle with naturally occurring BRD when compared with clinical reference standards of diagnosis.

Methods: This review was performed with eligible studies selected from CAB Abstract and MEDLINE from 1946 to 2015, as well as the "gray literature." Methodological quality of included studies was assessed using the QUADAS-2 tool developed for diagnostic accuracy studies. The accuracy parameters sensitivity (Se) and specificity (Sp) were obtained from the articles or through contact with the authors when not directly reported.

Results: A total of 314 studies were identified, from them, 23 met inclusion criteria as diagnostic studies for naturally occurring BRD. Quality of studies showed high risk of bias for case selection (70% of articles) and unclear risk of bias for index test (65%), reference standard (74%), and flow and timing (61%). There were high concerns regarding applicability for case selection (61% of studies) and reference standards used for defining BRD (48%). The concerns regarding index test application were low (83% of the studies). Only 4–8 studies could be included in the meta-analysis for each APP. No pooled estimates or pooled accuracy measurements were performed due to the low number of studies and multiple differences between studies, including reference standard definitions.

Conclusions and Clinical Importance: On the basis of these findings, it is not possible to make conclusions regarding the accuracy of APP for BRD diagnosis. The reporting of accuracy of APP for BRD detection is inconsistent among studies. Recommendations to improve capability for future meta-analyses in this area include reporting studies on diagnostic tests following the Standard for the Reporting of Diagnostic Accuracy Studies (STARD), as well as trying to standardize BRD definition across future studies.

Key words: Bias; Gold standard; Quality assessment; Sensitivity; Specificity.

Despite advances in veterinary medicine, animal husbandry, and animal welfare, bovine respiratory disease (BRD) continues to be the most economically significant disease in feedlots[1,2] and one of the most important causes of morbidity and mortality in dairy calves[3] and veal calves.[4] Relapses, mortality, propagation of infectious agents, and retarded growth or impaired production performance can be observed as consequences of BRD in addition to the associated costs for processing sick calves (monitoring and treatments).[5–7] Effective control of BRD has proven to be difficult in the North American dairy and beef industries, at least in part due to the complexity of disease pathogenesis and the ubiquity of BRD-associated pathogens.[2]

One of the greatest challenges for feedlot personnel or dairy farmers is an early and accurate diagnosis of BRD. In previous studies,[8–10] it has been reported that 37–68% of steers that never receive a diagnosis of BRD during the finishing period have lung lesions at

Abbreviations:

APP	acute-phase proteins
BRD	bovine respiratory disease
BRDneg	BRD-negative cases
BRDpos	BRD-positive cases
CRP	C-reactive protein
Fb	fibrinogen
Hp	haptoglobin
P_{BRD}	apparent BRD prevalence based on reference standard
QUADAS2	quality assessment of diagnostic test accuracy studies-2
SAA	serum amyloid A
Se	sensitivity
Sp	specificity
sROC	summary receiver-operating characteristic
STARD	standard for the reporting of diagnostic accuracy studies

From the Faculté de Médecine Vétérinaire, Département des Sciences Cliniques, Université de Montréal, Saint-Hyacinthe, QC, Canada (Abdallah, Francoz, Buczinski); Faculty of Veterinary Medicine, Zagazig University, Zagazig, Egypt (Abdallah, Selim); and the Department of Clinical Studies, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada (Hewson).

The main results of this study were presented at the 2015 ACVIM Forum, Indianapolis, IN, June 4, 2015, and at the Annual Congress of the Association des Médecins Vétérinaires Praticiens du Québec (AMVPQ), Saint-Sauveur, QC, Canada, September 2015.

Corresponding author: S. Buczinski, Clinique Ambulatoire Bovine, Faculté de Médecine Vétérinaire, C.P. 5000, St-Hyacinthe, QC, Canada J2S7C6; e-mail: s.buczinski@umontreal.ca.

Submitted December 22, 2015; Revised April 4, 2016; Accepted April 28, 2016.

DOI: 10.1111/jvim.13975

slaughter. Pen-rider accuracy has been reported to be only moderate with a sensitivity (Se: proportion of accurately diagnosed BRD cases) of 61.8% and specificity (Sp: proportion of accurately diagnosed non-BRD cases) of 62.8% as determined using a latent class analysis with no gold standard.[11] In dairy herds, clinical diagnosis also lacks accuracy,[12,13] and the same is true in veal calves where clinical signs, such as abnormal breathing, nasal discharge, and cough, were not accurate to predict moderate to severe lung lesions at slaughter.[4]

To improve diagnostic accuracy, several authors have focused on ancillary tests using various blood biomarkers.[14–16] Among the potential biomarkers, the acute-phase proteins (APP) change in concentration after infection, inflammation, surgical trauma, or stress[17] and can either increase (positive APP) or decrease (negative APP) as a consequence of inflammatory stimuli. Haptoglobin (Hp), serum amyloid A (SAA), and fibrinogen (Fb) are among the most commonly reported APP.[18] The C-reactive protein (CRP) has also been mentioned in various species (eg, human or dog) as an important APP but has received limited interest in cattle.[18] The serum increase of APP can occur as soon as 4 hours after the insult for SAA or CRP or later (24–48 hours) for Hp or Fb.[18]

The diagnostic and prognostic potential of measuring APP in cattle with BRD has been suggested[a] but is variably supported by the literature. As an example, Hp has been considered useful for identifying beef calves with BRD needing treatment and for monitoring treatment efficacy.[14,19] In contrast, other studies found that Hp had limited capacity as a diagnostic clinical tool for BRD in feedlot cattle.[8,20] Such discrepancy might indicate: (1) that not all pathogens involved in BRD increase serum Hp concentrations to the same extent, which might mitigate its accuracy in naturally occurring BRD (where the exact etiology is rarely known); (2) differences might occur due to sampling heterogeneity; or (3) differences might result from variability in study design and case definition. Diagnostic test accuracy refers to the ability of a test to distinguish between patients with disease (Se) and those without disease (Sp).[21] As for therapeutic and preventive measures, evidence-based veterinary medicine using systematic reviews can be used for assessing published literature on test accuracy to assess the strengths and weaknesses of the evidence available.[22] To date, the systematic assessment of diagnostic test accuracy has been uncommonly used in veterinary medicine and animal science.[23,24]

The aim of this study was, therefore, to perform a systematic review/meta-analysis concerning selected APP (Hp, SAA, and Fb) as diagnostic tools for cattle with naturally occurring BRD when compared with the reference standard used to diagnose BRD in each study.

Materials and Methods

Search Strategy

The literature search strategy used was based on the recommendations for Systematic Reviews of Diagnostic Test Accuracy from the Cochrane Diagnostic Test Accuracy Working Group.[21] Relevant publications were identified through an initial search of Ovid (CAB Abstract and MEDLINE) from 1946 to February 2015. We used, with the help of a librarian, a combination of controlled vocabulary keywords and Medical Subject Headings (MeSH) focusing on BRD and APP (Appendix S1). A "gray" literature search was also performed through Google Scholar to identify research abstracts from meeting proceedings or unpublished studies. Furthermore, the reference lists of review articles were consulted to look for any references not included in the systematic review list. The titles and abstracts of all studies identified in the initial literature search were subsequently screened for eligibility by 2 authors (AA and SB).

Eligibility Criteria

In the first round of screening, titles and abstracts were screened for inclusion based on potential relevance. Experimental BRD research articles, articles presenting case series (ie, only BRDpos cases), review articles, nonrelevant articles (ie, not reporting cattle respiratory disease and topics not related with the study objective), and articles not published in English were excluded. Prognostic studies were also excluded from this systematic review. A prognostic study was a study where the APP was measured and the BRD status was determined prospectively (time to event approach, in general >24 hours after APP measurement). Since the delay between the index test (APP) and the reference standard (BRDpos or BRDneg case assessment) not standardized at a fixed time (time to event studies with censored data), those studies were excluded. By extension, we excluded studies where the reference standard and the index test were performed with >1 day between them. Following preliminary screening, full-text copies of potentially relevant articles were obtained and further reviewed. Eligibility criteria for inclusion were independently assessed by 2 reviewers (AA and SB) using the same strategy as performed at the title and abstract level. A consensus between both reviewers was required to exclude an article at this stage of screening.

Data Extraction

Study information and available raw data were extracted from all the selected studies by the same person (AA) and compiled in a spreadsheet.[b] A second author (SB) independently extracted the same information and then a face-to-face meeting was done to double check the data. Study characteristics that were systematically recorded included APP studied, the type of patient (dairy or beef), age range (defined as calves if <1 year, adults if ≥1 year), and criteria used for defining the clinical reference standard (ie, definitions of BRDpos and BRDneg cases in the study) to be used as the gold standard against which to assess the APP accuracy. Since study design can impact assessment of diagnostic test accuracy,[25] the design of each study was recorded as a 1-gate design (if BRDpos and BRDneg cases were selected from the same population following the same diagnostic testing in both groups) or as a 2-gate design (if BRDpos and BRDneg cases were diagnosed using different inclusion criteria, eg, the BRDneg definition was based on healthy animals that received no diagnostic testing or cases with a disease other than BRD that received alternate testing). Since disease prevalence[26] can also impact the Se/Sp of a diagnostic test, the BRD prevalence (P_{BRD}) of each study was also recorded using the formula:

$$P_{BRD} = \frac{\text{nBRDPos}}{(\text{nBRDPos} + \text{nBRDNeg})}. \quad (1)$$

The tests used for APP quantification, the cutoff used to define BRDpos, and the method used to determine this cutoff (ie, whether it was predetermined based on previously published

Table 1. Examples of QUADAS 2 categories for the systematic review on the accuracy of acute-phase proteins (APP) for the diagnosis of bovine respiratory disease complex (BRD).

Risk of Bias	Signaling Questions	Example of Low-Risk Categories	Examples of High-Risk Categories
Domain 1: patient assessment	Was a consecutive or random sample of patients enrolled?	1-gate design	2-gate design
	Was a case-control design avoided?	Yes	No
	Did the study avoid inappropriate exclusions?	Study including all the prespecified selected patients	Study excluding cases based on index test or reference standard results
Domain 2: index test(s)	Were the index test results interpreted without knowledge of the results of the reference standard?	The reference standard test result was not known when reading the APP results	The APP results were interpreted knowing the final BRD status
	If a threshold was used, was it prespecified?	Statement that a particular cutoff was used before looking at the data	The cutoff chosen was obtained from the study results
Domain 3: reference standard	Is the reference standard likely to correctly classify the target condition?	Composite reference standard based on clinical examination and ancillary tests	Chronic respiratory conditions in a referral hospital
	Were the reference standard results interpreted without knowledge of the results of the index test(s)?	The BRD status of each patient is determined independently of APP measurement	The reference standard incorporates APP results
Domain 4: flow and timing	Was there an appropriate interval between the index test(s) and reference standard?	BRD status assessment and APP determination were determined at the same moment	Exclusion of prognostic studies[a]
	Did all patients receive a reference standard?	All patients were tested for BRD status determination	Some patients were not tested (eg, healthy without any specific definition)
	Did patients receive the same reference standard?	All patients had the same tests for BRD status determination	The tests for defining BRDpos and BRDneg were not the same
	Were all patients included in the analysis?	The number of patients with test results match the number of included cases	For some unspecified reason, there is a mismatch between included patients and patients with tests results

Applicability Concerns		Examples of Low Applicability Concerns	Examples of High Applicability Concerns
Domain 1: patient assessment	Is there concern that the included patients do not match the review question?	The patients are first-line cases with no selection of more advanced BRD stages	The patients are chronic cases with obvious BRD signs
Domain 2: index test(s)	Is there concern that the index test, its conduct, or interpretation differ from the review question?	The test is a commercially validated test that can be used in future studies	The test is developed for that particular study without further details
Domain 3: reference standard	Is there concern that the target condition as defined by the reference standard does not match the review question?	The BRD status is determined by the use of multiple tests results (combination of physical examination, temperature, and ancillary tests)	BRD-positive cases defined as anorectic and pyretic animals with no further testing

[a]A study was excluded if BRD status was assigned >1 day after the APP sampling date.

studies or was data driven) were noted if present, as well as details of the 2 by 2 tables [true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) cases]. In situations where either the cutoff or the 2 by 2 table were available in the report, the authors were contacted 3 times if an e-mail address was listed, in order to get the raw data or unpublished 2 by 2 tables. When raw data were obtained, the cutoff used was the cutoff that minimized misclassification by the APP (ie, provided the maximal Se + Sp value).

Methodological Quality Assessment

A quality assessment of every retained article was performed using the Quality Assessment of Diagnostic Test Accuracy 2 (QUADAS-2) tool.[27] The QUADAS-2 process categorizes the risk of potential bias and study applicability when evaluating diagnostic tests, generating ratings of low, unclear or high risk of bias (Table 1). This tool assesses internal and external validity of studies based on the design, the diagnostic (index) test used, and the definitions of positive and negative cases. The QUADAS-2 tool consists of 4 key domains to assess the risk of bias, including patient selection, the index test used (ie, the APP), the reference standard (the test considered as the gold standard [in our case the definition of BRDpos and BRDneg] against which index test results are compared), and flow and timing of patients and index test application in the study. The risk of bias for the timing was assessed as low if it was explicitly stated that the APP and BRD status were determined at the same moment. It was classified as unclear when no explicit statement indicated the exact delay, but there was no indication that this delay could not exceed 1 day. The risk of bias for study flow was assessed taking into account whether all the patients received a reference standard and if it was the same reference standard (eg, not a different test to define BRDpos versus BRDneg cases) as well as if all the patients were included in the analysis. When the information was not present,

an unclear risk was assigned. Furthermore, applicability of each retained article to the systematic review question was also assessed by the QUADAS-2 tool through 3 items that consider patient selection, the index test itself, and the reference standard test used. Both reviewers (AA and SB) scored each study independently using this 7-item tool. Any disagreements were resolved to consensus during a face-to-face discussion about each disagreement. When continued disagreement was present, a third reviewer was solicited (JH or DF) and the final decision was based on that review and final consensus.

Statistical Analysis

Descriptive statistics were used for the study characteristics. Paired forest plots of sensitivity and specificity were obtained based on 2 by 2 table results using specific freeware for systematic review.[c] The information gathered included the true-positive cases (TP: cases ≥ APP cutoff and BRDpos); false-positive cases (FP: cases ≥ APP cutoff and BRDneg); false-negative cases (FN: cases < APP cutoff and BRDpos), and true-negative cases (TN: cases < APP cutoff and BRDneg). Moses-Littenberg summary receiver-operating characteristic (sROC) curves were used by default in RevMan for a visual description of the accuracy obtained for the multiple tests in the different studies. The SROC curve is a good visual graph that helps for preliminary data analysis before exploring more complex models. This curve is obtained based on the true-positive fraction (TPF = Se = TP/(TP + FN)) and the false-positive fraction (FPF = (1–Sp) = FP/(FP + TN) of every study based on fitting the linear mixed model:

$$D_i = a + b \times S_i + e_i, \tag{2}$$

where a and b are the intercept and regression coefficient estimated from the studies' results and e_i is the random error in the study i

$$D = \text{logit}(\text{TPF}) - \text{logit}(\text{FPF}), \tag{3}$$

$$S = \text{logit}(\text{TPF}) + \text{logit}(\text{FPF}). \tag{4}$$

As the number of studies in which Se/Sp could be obtained was low, with heterogeneity in BRDpos and BRDneg definitions as well as test cutoffs chosen, it was not appropriate to calculate pooled estimates of these values.[28] For these reasons, no further statistical analysis models were attempted.

Results

Our initial search strategy retrieved 385 articles from CAB Abstracts and MEDLINE and 34 articles from Google Scholar. The flow diagram outlining the selection strategy to obtain 25 articles is summarized in Figure 1. Twenty-three studies,[29–49,d,e] were ultimately included. Two articles described studies already reported in different journals.[46,47]

Study Characteristics and BRD Definitions

Table 2 shows the characteristics of the 23 studies that were retained for analysis. Eighteen studies used calves, while 5 studies enrolled adult cattle as target animals. Seventeen studies (74%) used a 2-gate design, where healthy animals (n = 12) and animals with another diagnosis than BRD (n = 5) were included as controls. Only 6 studies (26%) used a 1-gate design. Four of 23 (17%) studies estimated all selected APP

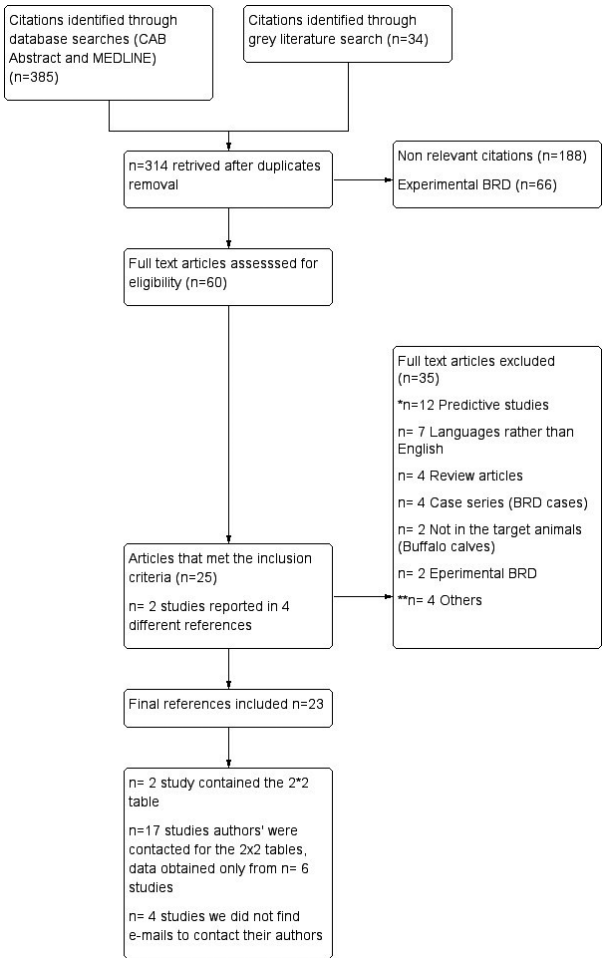

Fig 1. Flow diagram for the study selection process. Gray literature searched through Google Scholar. *A predictive study was a study where the acute-phase protein (APP) was measured and the status relative to bovine respiratory disease (BRD) was determined prospectively. Since the delay between the index test (APP) and the reference standard (BRDpos or BRDneg case assessment) was delayed and not standardized at a fixed time (time to event studies with censored data), those studies were excluded. **n = 4 others: Analytical papers (n = 2), BRD status based on serologic status (n = 1), and study assessing pathogenic factors of pneumonic pasteurellosis (n = 1).

(Hp, SAA, and Fb) for BRD diagnosis; 8 (35%) studies assessed both Hp and SAA, 2 (9%) studies assessed both Hp and Fb, 5 (22%) studies estimated Hp, and 4 (17%) studies evaluated Fb as a single diagnostic marker.

The BRDpos and BRDneg case definitions used in the different studies are summarized in Table 3. Sixteen (70%) studies used clinical examination only as the reference standard for diagnosing BRD, whereas 6 (26%) studies applied a combination of clinical and laboratory methods as their reference standard for BRD diagnosis (2 of them also included APP results in the case definition). The remaining study[46] used clinical examination in addition to ultrasonography and endoscopic examination in some cases to diagnose BRD. Two studies

Table 2. Characteristics of the studies included in the systematic review.

Study ID	APPs Measured	Type of Cattle	Study Type	Age Range
Al Qudah, 2009	Fb	NR	2-gate[a] (healthy controls[c])	2–3 m
Alsemgeest, 1994	Hp and SAA	NR	2-gate (alternative diagnosis[d] + healthy controls)	Diseased: Mixed age, control cows mean 3 y
Angen, 2009	Hp and SAA	Dairy	2-gate (healthy controls)	14 d to 4 m
Arslan, 2010	Hp and SAA	Beef	2-gate (healthy controls)	1–3 y
Coskun, 2012	Hp and SAA	NR	2-gate (healthy controls)	Mean 38 d (15–65)
Dudek, 2011	Hp and SAA	NR	2-gate (healthy controls)	NR
Fathi, 2013	Hp, SAA, and Fb	Dairy	2-gate (healthy controls)	2 w to 6 m
Fratric, 2011	Fb	Dairy	2-gate (healthy controls)	3 m
Ganheim, 2007	Hp, SAA, and Fb	Dairy	1-gate[b]	Group A 4–13 w (8 w) and group B 9–13 w (11 w)
Idoate, 2015	Hp	Beef	2-gate (healthy controls)	Calves (NR)
Lee, 2005	Fb	Dairy	1-gate	<1 m
Mohammadi, 2008	Hp and Fb	Dairy	2-gate (healthy controls)	2 w to 6 m
Nazifi, 2008	Hp	NR	2-gate (alternative diagnosis + healthy controls)	7 < 2 y. 11 2–4 y, 18 > 4 y
Nazifi, 2010	Hp and SAA	NR	2-gate (alternative diagnosis + healthy controls)	NR
Prathaban, 1990	Fb	NR	2-gate (alternative diagnosis + healthy controls)	NR
Svensson, 2006	Hp	Dairy	1-gate	2–35 d
Svensson, 2007	Hp	Dairy	1-gate	24–56 d
Timsit, 2009	Hp and Fb	Beef	1-gate	NR
Timsit, 2011	Hp	Beef	1-gate	NR
Tothova, 2010	Hp and SAA	NR	2-gate (healthy controls)	3–6 m
Tothova, 2011	Hp, SAA, and Fb	Dairy	2-gate (alternative diagnosis + healthy controls)	2 w to 6 m
Tothova, 2013	Hp, SAA, and Fb	NR	2-gate (healthy controls)	4–6 m
Wolfger, 2015	Hp and SAA	Beef	2-gate (healthy controls)	NR

Hp, haptoglobin; SAA, serum amyloid A; Fb, fibrinogen; NR, not reported; d, days; w, weeks; m, months; y, years.
[a]Two-gate: Case-control design: the composition of control group is indicated in brackets.
[b]One-gate: Cohort design.
[c]The most commonly reported 2-gate design was using clinically healthy animals as the controls.
[d]In the case of an alternative diagnosis, the nondiseased animals were animals that had a disease different from BRD.

used transtracheal and bronchoalveolar fluid culture to determine the causative agent of BRD.[31,33] Only 5 (22%) of the 23 studies clearly reported their criteria for definition of BRDneg cases.

Quality Assessment

Overall, quality assessment was rated as a high or unclear risk of bias or applicability, with only index test applicability being rated as low for risk of bias (Fig 2). The risk of bias in patient selection was considered high in 17 (74%) studies, mainly because of the 2-gate (case-control) design used in the majority of studies. The risk of bias in performance of the index test was considered unclear in 15 (65%) studies. The risk of bias for reference standard definition was unclear in the majority of studies (n = 17; 74%). The risk of bias arising from patient flow and timing of procedures was considered unclear in the majority of studies, mostly because in almost all studies it was not clear if a reference standard was performed in every patient (or whether that standard could have been performed differently in the BRDneg and BRDpos cases), if all cases were included for analysis, or if the interval between index test and reference standard determination was appropriate or not. Regarding applicability, there was high risk identified for patient selection and the reference standard, but a low risk for index test application since most studies reported commercially validated tests used under controlled conditions.

Index Test (APP) Application

The different techniques used for APP quantification are illustrated in Table 4. Cutoffs and data relative to the 2 by 2 tables were available in 8 studies (35%) involving Hp, 5 studies (22%) with SAA, and 4 studies (17%) with Fb. Results also showed that in 15 (65%) studies, it was unclear if the cutoff point was prespecified or not. We obtained the cutoff points and the 2 by 2 tables from 6 out of the 17 contacted authors (35%).

Accuracy Measurements

The accuracy of the different APP is summarized in Figures 3 and 4. The sROC was indicative of higher accuracy for Hp due to its closer fit to the upper left corner of the ROC space. For Hp, the Se varied from

Table 3. Reference standards used for classification of positive (BRDpos) and negative (BRDneg) bovine respiratory disease (BRD) cases and apparent BRD prevalence in the studies reporting diagnostic accuracy of acute-phase protein measurement for BRD diagnosis.

Study ID	BRDpos Definition	BRDneg Definition	P_{BRD}[a]
Al Qudah, 2009	Acute cases: T: 40°C, anorexia, depression with expiratory grunt, tachycardia, reluctance to move, crackles and harsh breath sounds on auscultation. Chronic cases: Normal or moderate persistent fever, rough hair coat, gaunt appearance, history of pneumonia and bronchitis for 2 weeks, increased HR and RR, copious bilateral mucopurulent nasal discharge, and chronic productive cough with loud breath sounds over the ventral part of lung	NR	55
Alsemgeest, 1994	Clinical diagnosis applied (details not provided) and confirmation by pathological diagnosis	NR	17
Angen, 2009	T > 39.5°C with nasal discharge, coughing, or unprovoked RR >40/min	T < 39.5°C, no nasal discharge, no coughing, and an unprovoked RR <40/min	38
Arslan, 2010	T ≥ 39.5°C, with RR >50/min, coughing and/or nasal discharge, and anorexia	NR	67
Coskun, 2012	(T > 39.5°C, with coughing, RR ≥40/min, nasal discharge, anorexia, depression, crackles, and harsh sounds on auscultation) and laboratory analysis	T < 39.5°C, no nasal discharge, no coughing, and RR <40/min	79
Dudek, 2011	Clinical diagnosis applied (details not provided) and confirmed with serological and microbiological examination	NR	50
Fathi, 2013	T > 39.5°C, signs of depression with abnormal lung sounds on auscultation	NR	50
Fratric, 2011	T > 39.5°C, depression, lack of involvement of other body systems, abnormal lung sounds on auscultation	NR	50
Ganheim, 2007	Coughing and other signs (details not provided)	NR	41
Idoate, 2015	Wisconsin CHSC ≥5	Wisconsin CHSC ≤4	18
Lee, 2005	Clinical diagnosis applied (details not provided)	NR	19
Mohammadi, 2008	T > 39.5°C, signs of depression, lack of involvement of other body systems, abnormal lung sounds on auscultation	NR	50
Nazifi, 2008	Clinical diagnosis applied (details not provided) and laboratory analysis (not defined)	NR	6
Nazifi, 2010	Fever, signs of acute pulmonary involvement, coughing and dyspnea with crackles and wheezes on auscultation, bacteriology, virus identification, and postmortem examination (details not provided)	NR	12
Prathaban, 1990	Clinical diagnosis applied (details not provided)	NR	6
Svensson, 2006	Coughing or sneezing for >2 days, severely or moderately increased respiratory sounds (obvious increased bronchial or vesicular breath sounds or presence of adventitious sounds synchronous with breathing), and/or nasal discharge	NR	31
Svensson, 2007	T > 39.5°C with coughing or sneezing for >2 days, severely or moderately increased respiratory sounds on auscultation, and/or nasal discharge	NR	18
Timsit, 2009[b]	Visual appraisal performed twice daily by owners. At the first detection of BRD, clinical examinations performed by a veterinarian on every animal in the pen	NR	19
Timsit, 2011	T ≥ 39.7°C and abnormal pulmonary sounds on lung auscultation or animals having T ≥ 39.7°C, depression and at least one other BRD clinical sign	NR	88
Tothova, 2010	General health status (T, food intake, and behavior) and respiratory system examination by visual inspection, breathing rate, nasal discharge, type of breathing, dry or wet spontaneous coughing and labored breathing with open mouth, increased or decreased loudness of the breathing sound, bronchial sounds, abnormal breath sounds (crackles and/or wheezes) and in some cases by ultrasonography and endoscopic examination	NR	64
Tothova, 2011	General health state (T, food intake, behavior) and respiratory system examination including recording of the clinical signs of the disease (details not provided)	NR	53

(continued)

Table 3 (Continued)

Study ID	BRDpos Definition	BRDneg Definition	P_{BRD}[a]
Tothova, 2013	General health state (T, food intake, and behavior) and respiratory system examination by visual inspection (breathing rate, nasal discharge, type of breathing, dyspnea, dry or wet spontaneous cough) and auscultation (increased or decreased loudness of breathing sounds, bronchial sounds, abnormal breathing sounds such as crackles and/or wheezes) and signs of breathing with mouth open; calves included that had clinical signs of the disease manifested for more than 2 weeks	NR	44
Wolfger, 2015[b]	Clinical definition: having ≥2 clinical signs of BRD (reluctance to move, crusted nose, nasal or ocular discharge, drooped ears or head, and gaunt appearance) and T ≥ 40°C	T < 40°C and no severe sickness observed in the treatment chute, or healthy	92[c]

BRD, bovine respiratory disease; T, temperature; HR, heart rate; RR, respiratory rate; NR, not reported or no clear definition; CHSC, calf health scoring chart; Hp, haptoglobin; Fb, fibrinogen.

[a]P_{BRD}: apparent BRD prevalence in the study.

[b]In this study, the classification presented is not the classification that was initially used. The original classification used a combination of clinical signs and Hp concentrations (Wolfger et al 2015) or different combination of Hp and Fb values and clinical signs (Timsit et al 2009).

[c]For this specific study, the BRD prevalence was calculated using the animals where APP was measured (Fig 3), versus 41% from calculated directly from the original study.

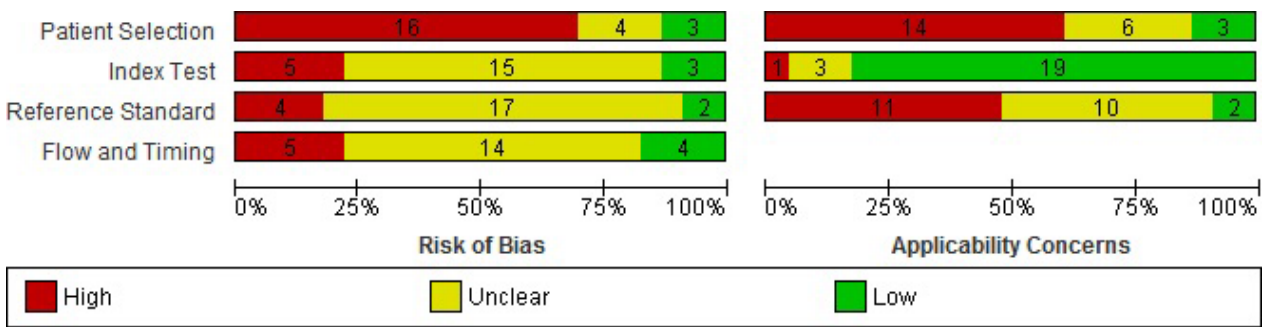

Fig. 2. Summary of the methodological quality of included studies on the basis of authors' assessments regarding the 4 domains assessing the risk of bias and the 3 domains assessing applicability concerns of the Quality of Diagnostic Accuracy Study-2 checklist for each study. In green are highlighted the number of studies with a low risk of bias or low applicability concern and in red the studies with a high risk of bias or high applicability concerns. The number of studies in yellow indicates the studies where these risks of bias or applicability concerns could not be assessed properly (unclear).

61 to 100% and Sp from 80 to 100%, whereas for SAA the Se varied from 53 to 100% and Sp ranged from 43 to 94%. Sensitivity for Fb was between 57 and 80% and Sp varied from 89 to 95%.

Discussion

We reviewed the literature to determine the diagnostic accuracy of haptoglobin, serum amyloid A, and fibrinogen, as relevant acute-phase proteins in naturally occurring BRD using an evidence-based approach. When performing a systematic review, it is essential that a very defined research question is posed and that inclusion and exclusion criteria are determined a priori to directly investigate that question. In our systematic review on the sensitivity and specificity of acute-phase protein determination to diagnose BRD, our research question was confined to use of these tests in the clinical setting. This restricted our final database to only those studies involving naturally occurring BRD, which we felt would more likely yield a better representation of APP test performance under realistic conditions of test use. Test assessment during experimentally induced disease or when applied to use on only severely affected animals (ie, case series reports) would have shown inflated Se and Sp outcomes, and so these studies were intentionally excluded during our systematic review of the literature.

Table 4. Different methods used to quantify acute-phase proteins in studies focusing on naturally occurring bovine respiratory disease.

Study ID	Methods of Index Test Application
Al Qudah, 2009	Fb: Heat precipitation method as described by Schalm et al[50]
Alsemgeest, 1994	Hp: Hemoglobin binding assay as described by Makimura and Suzuki[51] (Eastman-Kodak Co., Chicago USA) SAA: Indirect ELISA method as described by Boosman et al[52]
Angen, 2009	Hp: Method as described by Heegaard et al[53] (monoclonal antibody-based capture ELISA as described previously Godson et al[54]), SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Arslan, 2010	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Coskun, 2012	Hp: Sandwich ELISA (Life Diagnostic Inc, West Chester, PA) SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Dudek, 2011	Methods for determination of Hp and SAA were not provided
Fathi, 2013	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland) Fb: Heat precipitation method
Fratric, 2011	Fb: Fibrinogen reagent kit (Techno clone Gmbh, Vienna, Austria)
Ganheim, 2007	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland) Fb: Automated analyzer (Konelab 30, Konelab Corporation) as described by Becker et al[55]
Idoate, 2015	Hp: Commercially available bovine Hp ELISA kit (Immunology Consultants Laboratory, Inc, Portland, OR)
Lee, 2005	Fb: Ready to use kits (Inhwa Pharma, Korea) and UV VIS spectrophotometer (Hanson, Tech, Korea)
Mohammadi, 2008	Hp: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland) Fb: Method for determination not provided
Nazifi, 2008	Hp: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Nazifi, 2010	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Prathaban, 1990	Fb: Biuret method as described by Phillips et al[56] and Bauer et al[57]
Svensson, 2006	Hp: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Svensson, 2007	Hp: Commercial kit (PHASE Range Haptoglobin Assay Kit) based on Hp-hemoglobin binding and preservation of peroxidase activity as measured by spectrophotometry using a Cobas Mira (Roche Basel)
Timsit, 2009	Methods for determination of Hp and Fb were not provided
Timsit, 2011	Hp: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Tothova, 2010	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Tothova, 2011	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Tothova, 2013	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)
Wolfger, 2015	Hp and SAA: ELISA commercially available kit (Tridelta Development Limited, Maynooth, Co, Kildare, Ireland)

Fb, fibrinogen; Hp, haptoglobin; SAA, serum amyloid A; UV VIS, ultraviolet visible.

We have found that there are a relatively low number of studies where accuracy of the APP was presented in terms of Se and Sp. By including only studies of APP assessment under field conditions, we were able to identify only 23 studies over the period of 1946–2015 that met our inclusion criteria. These 23 studies reported variable Se and Sp of each APP in diagnosing BRD, with generally improved test performance for Hp over fibrinogen and SAA in cattle with BRD although not statistically different. These 23 reports included a variable mix of which APP were studied; therefore, the actual number of studies for each APP (Fb, SAA, and Hp) upon which our systematic review conclusions could be based was even fewer. In our systematic review, we sought to determine the collective Se and Sp of each APP for diagnosing BRD based on the existing literature. However, there were a relatively small number of studies where 2 by 2 tables or data were available to assess diagnostic test accuracy, even after contacting 17 of the authors with 3 reminders. This emphasizes the need for future reporting of data in such a way as to allow calculation of Se/Sp estimates as well as BRD prevalences. The use of different cutoffs between studies was also a limitation to obtain pooled estimates of Se and Sp using a bivariate analysis.[58] In this situation, the determination of a hierarchical summary ROC (HSROC) model can be used, assuming that a cutoff effect exists (ie, when the Se increases, the Sp decreases).[59] However, because of the relatively small number of studies and the highly variable definitions of BRDpos and BRDneg cases and studies particular settings, we could not determine these accuracy parameters. Due to this, our ability to make firm conclusions or to perform meta-analysis of the data was impaired. Another limitation of true APP accuracy determination is that APP response can also be caused by various stressors, which could potentially affect its specificity for BRD detection. As an example, castration might occur for feedlot calves with a short delay before being on feed and is also associated with increased Hp concentrations.[60] Haptoglobin concentration can also increase after stressful experimental events such as transportations for 2-day[61] or 3-day starvation.[62] All these situations could potentially have a negative impact on APP specificity.

During the systematic review process, it is also important to assess the scientific merits of the papers ultimately selected for inclusion, in order to make judgment on the strength of the conclusions that are drawn from the systematic review analysis. In our study, we employed the QUADAS-2 system for assessing applicability of the studies to our research question and for assessment of any risks of bias in our included

FP Study	TP	FP	FN	TN	Cut off (mg/ml)	Study design	Sensitivity (95% CI)	Specificity (95% CI)
Tothova, 2013	16	2	12	34	2.916	Two-gate	0.57 [0.37, 0.76]	0.94 [0.81, 0.99]
Tothova, 2011	32	5	19	41	3.017	Two-gate	0.63 [0.48, 0.76]	0.89 [0.76, 0.96]
Timsit, 2009	12	5	9	86	4.2	One-gate	0.57 [0.34, 0.78]	0.95 [0.88, 0.98]
Fathi, 2013	24	3	6	27	5.0	Two-gate	0.80 [0.61, 0.92]	0.90 [0.73, 0.98]

Hp Study	TP	FP	FN	TN	Cut off (mg/ml)	Study design	Sensitivity (95% CI)	Specificity (95% CI)
Fathi, 2013	21	3	9	27	0.05	Two-gate	0.70 [0.51, 0.85]	0.90 [0.73, 0.98]
Tothova, 2011	32	4	19	42	0.142	Two-gate	0.63 [0.48, 0.76]	0.91 [0.79, 0.98]
Wolfger, 2015	114	2	0	8	0.15	Two-gate	1.00 [0.97, 1.00]	0.80 [0.44, 0.97]
Coskun, 2002	21	0	9	8	0.15	Two-gate	0.70 [0.51, 0.85]	1.00 [0.63, 1.00]
Tothova, 2010	23	1	4	14	0.2	Two-gate	0.85 [0.66, 0.96]	0.93 [0.68, 1.00]
Timsit, 2009	16	5	5	86	0.25	One-gate	0.76 [0.53, 0.92]	0.95 [0.88, 0.98]
Tothova, 2013	17	1	11	35	0.293	Two-gate	0.61 [0.41, 0.78]	0.97 [0.85, 1.00]
Idoate, 2015	13	2	1	12	0.81	Two-gate	0.93 [0.66, 1.00]	0.86 [0.57, 0.98]

SAA Study	TP	FP	FN	TN	Cut off (mg/ml)	Study design	Sensitivity (95% CI)	Specificity (95% CI)
Coskun, 2002	30	4	0	3	0.0090	Two-gate	1.00 [0.88, 1.00]	0.43 [0.10, 0.82]
Tothova, 2010	16	3	11	12	0.05	Two-gate	0.59 [0.39, 0.78]	0.80 [0.52, 0.96]
Tothova, 2013	23	2	5	34	0.054	Two-gate	0.82 [0.63, 0.94]	0.94 [0.81, 0.99]
Tothova, 2011	33	6	18	40	0.069	Two-gate	0.65 [0.50, 0.78]	0.87 [0.74, 0.95]
Wolfger, 2015	60	1	54	9	0.13	Two-gate	0.53 [0.43, 0.62]	0.90 [0.55, 1.00]

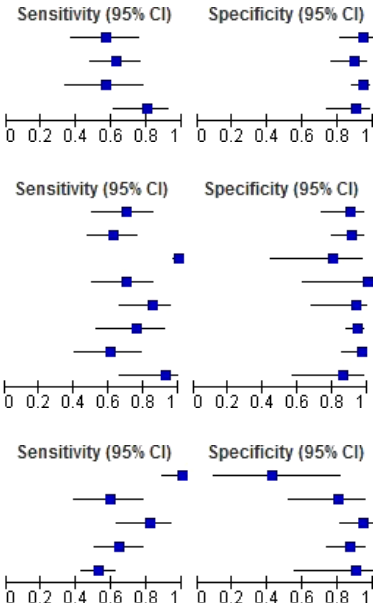

Fig 3. Accuracy table and forest plots illustrating sensitivity (Se) and specificity (Sp) for haptoglobin (Hp, n = 8 studies), serum amyloid A (SAA, n = 5 studies), and fibrinogen (Fb, n = 4 studies) as reported in studies of naturally occurring bovine respiratory disease. CI, confidence interval, TP, true-positive cases (cases with APP value ≥ cutoff and BRDpos); FP, false-positive cases (cases with APP value ≥ APP cutoff and BRDneg); FN, false-negative cases (FN: cases with APP value < APP cutoff and BRDpos); TN, (cases with APP value < APP cutoff and BRDneg). The APP cutoffs were predetermined in Wolfger et al 2015 (for Hp) and Coskun et al 2012 studies. They were data driven in other studies except for Fathi et al 2013, where it was unclear how they were determined. The raw datasets were available from Coskun et al (2012); Timsit et al (2009); Tothova et al (2011, 2013); and Wolfger et al 2015.

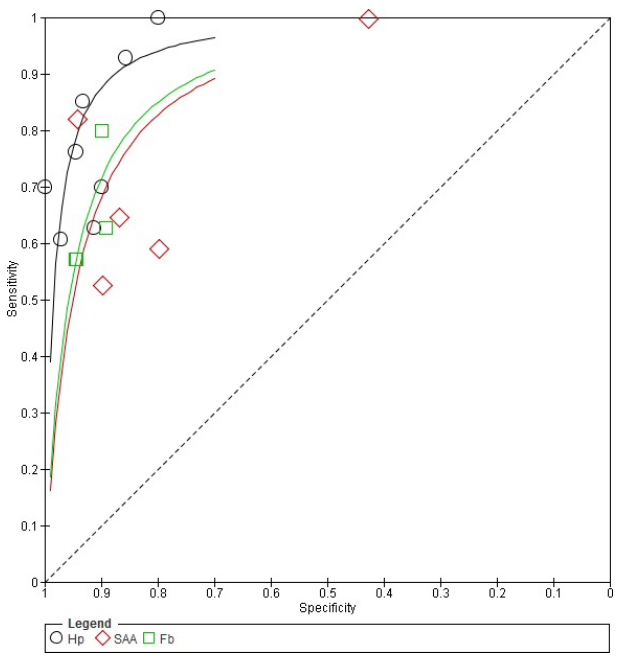

Fig 4. Summary receiver-operating characteristic (SROC) curve of acute-phase proteins accuracy for the diagnosis of naturally occurring bovine respiratory disease. The studies reporting haptoglobin (black circle), serum amyloid A (red diamond), and fibrinogen (blue square) are used to obtain the respective sROC curves based on Equation 2. A higher area under the sROC curve (ie, curve moved toward the upper left corner of the ROC space) is an indicator of higher test accuracy.

studies, in order to better understand whether our conclusions were appropriate and to identify any limitations to our review. Applicability of each study was evaluated in terms of relevance of the index test, reference standard, and population used to our research question. Areas of potential bias that were assessed and included the following: method of case classification (the reference standard used to diagnose BRD in each study), how the index test (APP measurement) was employed during the diagnostic process, methods of patient selection, and timing of measuring the APP.

In general, the majority of studies had high applicability to our research question in that the index test was used in our population of interest in a manner similar to the intended clinical application during field diagnosis of naturally occurring BRD in cattle. However, in terms of potential bias of the included studies as assessed by the QUADAS2 tool, high risk of bias was identified regarding allocation of cattle into BRDpos and BRDneg groups. The BRD case definition (ie, the reference standard) was variable between studies, which is a common feature across all types of BRD studies and arises from the absence of a consensus concerning how to diagnose BRD in dairy or beef cattle.[f] The BRDpos definition was mainly clinically derived and based on various combinations of clinical signs, including fever, depression, nasal discharge, cough, increased respiratory rate, and abnormal lung sounds on auscultation (crackles, wheezes, and harsh sounds). Of these signs, using differing combinations of signs had the potential to significantly impact BRD classification.

Using differing combinations of clinically derived signs in populations of differing true BRD prevalence would result in underestimating the number of BRD cases in situations of high disease prevalence and imperfect sensitivity, and more likely, in populations of low disease prevalence and imperfect specificity, to overestimate the number of BRD cases. The use of an established clinical scoring system can improve disease detection to some degree.[5,63] If misclassification by the reference standard is not estimated, this would bias the accuracy estimates of the index tests (APPs measured) in each study. Similarly, the exact BRDneg definition was infrequently defined, which could also be perceived as a limitation on the interpretation of apparent test specificity; in general, the use of healthy animals to represent the BRDneg comparison group would lead to overestimating the index test specificity.[25]

The differential reference bias, in which the reference standard is not applied in the same manner between BRDpos and BRDneg cases, can also have an impact on Se/Sp.[64] Only 5 studies clearly defined how BRDneg cases were determined. When applying the same reference standard to classify the cases, one could guess that this definition is the complement of the BRDpos definition. However, if not specifically stated, it is difficult for the reader to be sure of that precise definition which often resulted in us having to assign an "unclear" risk of bias in many studies for the reference standard.

The use of APPs (index test) as a part of the reference standard tests (known as incorporation bias) was only observed in 2 studies where case definition was based on both the presence of clinical signs and Hp values above a specified cutoff. Using this definition could obviously overestimate the Se of APP.

Using the QUADAS2 tool, we detected "high" or "unclear" risk of bias on the Se and Sp of the various acute-phase proteins, resulting from how patients were identified or enrolled. Of the various methodological weaknesses that could make diagnostic studies vulnerable to bias, the use of 2 gate-design (case-control) studies is certainly the most commonly known design prone to overestimate index test accuracy.[65] Using healthy controls traditionally leads to spectrum bias, due to enrollment of patients with high suspicion of the disease arising from serial testing or the presence of several positive clinical signs or paraclinical tests (ie, representing more severely affected animals so being worse in terms of the spectrum of BRD signs) compared to control animals which are healthy (ie, absence of all the clinical signs and negative to the paraclinical tests). In the studies we included, the risk of bias was assessed as "high" based on how BRDpos cases were defined, versus those classified as BRDneg healthy controls. Commonly, a 2-gate approach was used, where cases were based on different selection criteria than those used to define healthy controls. For instance, in several studies, BRDpos was based on the presence of a fever, whereas BRDneg cases were those animals not pulled for any assessment (including any determination of rectal temperature to say they do not have a fever) when pen riders visually assessed the animals each day. This approach is recognized as having high potential to overestimate the sensitivity of a test[66] because only the more severely affected animals would undergo the reference standard of rectal temperature measurement for inclusion as BRDpos cases.

Details such as randomized or consecutive sampling in the population were often not included to allow determination of whether the patient selection was unbiased. The fact that the majority of studies did not provide an inclusion flow diagram of the patients, detailing how many were eligible and alternatively excluded, also suggested potential bias. These types of study characteristics have been shown to increase the diagnostic odds ratio of a test (which is an index of global test accuracy; increasing when either Se or Sp increases) in human medical literature.[65] Using flow diagrams such as reported in a recent study[49] can be a quick and effective way to help to understand the sampling strategy as well as any case not retained in the article.

Another concern arising from assessing the risk of bias of the studies was that frequently it was not possible to ensure that interpretation of either the reference standard (BRD diagnosis) or the index test application (cases selected to undergo APP measurement) occurred independent of each other. For example, often it was unclear whether the researcher selected cases for APP testing based on a previous suspicion of BRD (observation of clinical signs) by the producer. Similarly, in some studies, it was not clear whether the test cutoff used to assign animals to BRDpos or BRDneg groups in a binary fashion was determined in advance of the study, versus being selected after data collection, which would have a clear impact on the assessment of diagnostic accuracy in such studies.[67]

As an overall observation regarding results from the QUADAS2 evaluation of our included studies, the majority of studies were assessed as having either high or unclear risk of bias, or applicability concerns were rated as high or unclear. It was very interesting to note that most of time we could not differentiate a high versus a low risk of bias within a study due to the absence of a clear mention of the methodology used by the authors; therefore, by default, an "unclear" risk or concern category was chosen in the QUADAS2 tool. This is also a common pitfall of many diagnostic studies in human medicine.[68] However, this could very readily be avoided through more explicit reporting of methodology by authors, allowing stronger conclusions to be drawn from their studies. We voluntarily included studies where the diagnostic accuracy was not the primary objective of the study. This might have been another reason why the reporting could have been heterogeneous when compared with studies in which assessment of diagnostic accuracy of an index test was the primary objective.

In conclusion, we have found that the data concerning the diagnostic accuracy of APP for BRD diagnosis are scant and partially reported. Prior studies have provided the sensitivity and specificity of various acute-phase proteins in the diagnosis of BRD under field conditions, with variable results being seen in terms of the individual APP studied, the test cutoff used to define disease, whether the

APP were used alone or in parallel, and timing of test application to case diagnosis. Through our systematic review process, we attempted to take this knowledge surrounding the use of APP to a deeper level of understanding in terms of diagnosing BRD, but could not as a result of such differing study designs and reporting. Since definition of BRDpos and BRDneg cases varied from study to study, it was difficult to explain if observed variation in APP test accuracy was due to study characteristics, sampling, the test cutoff used, or difference in the reference standard. We could, therefore, not conclude on the added value of using APP tests to rule-in or rule-out BRD. In light of the challenges experienced during this systematic review of the literature, we recommend that effort be invested into first establishing a more standardized case definition of BRD in cattle, as well as enforcing a more structured approach to reporting studies on diagnostic test accuracy, using for example, the STAndard for Reporting Diagnostic accuracy studies (STARD) statement.[69] Systematic reviews and meta-analyses provide an opportunity for drawing more meaningful interpretations from the collective data of multiple studies. Challenges encountered in this study, as in other systematic reviews, highlight the need for improved and standardized study reporting such that the scientific community is able to maximize the value and usefulness of future research data.

Footnotes

[a] Eckersall PD. Acute-phase proteins as monitoring tools in farm animals. 13th International Conference Production Diseases in Farm Animals. Leipzig, Germany, 2007:374–380.

[b] Excel, Microsoft, Redmond, WA.

[c] Review Manager (RevMan) Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

[d] Mohammadi GR, Mohri M, and Fathi A. Identify relationship among these acute-phase proteins (Hp and Fb), albumin and clinical finding in dairy calf pneumonia. XV Congress of the Mediterranean Federation for Health and Production of Ruminants 2008.

[e] Timsit E, Assié S, Bareille N. Failure to detect bovine respiratory disease in newly received young beef cattle by visual appraisal. Proceedings of the 12th Symposium of the International Society for Veterinary Epidemiology and Economics, Durban, South Africa, 2009:585.

[f] Hewson J, Buczinski S. Researching BRD-Part II: apples, oranges or pears? Defining the ingredients needed for future research in BRD. Annual convention of the ACVIM 2014, Nashville, TN.

Acknowledgments

The authors thank Drs Alparslan Coşkun, Ezzatollah Fathi, Karin Orsel, Catarina Svensson, Edouard Timsit, and Csilla Tothova for their kind help through sharing datasets or accuracy results. We also thank Mrs Huguette Mallet, Librarian of the Faculty of Veterinary Medicine, University of Montréal, for her tremendous help in searching strategy.

Conflict of Interest Declaration: The authors declare no conflict of interest.

Off-label Antimicrobial Declaration: The authors declare no off-label use of antimicrobials.

References

1. Loneragan GH, Dargatz DA, Morley PS, Smith MA. Trends in mortality ratios among cattle in US feedlots. J Am Vet Med Assoc 2001;219:1122–1127.

2. Gorden PJ, Plummer P. Control, management and prevention of bovine respiratory disease in dairy calves and cows. Vet Clin North Am Food Anim Pract 2010;26:243–259.

3. USDA APHIS. Dairy Heifer Raiser. An overview of operations that specialize in raising dairy heifers. In: APHIS, ed. Dairy 2011. USDA–APHIS–VS–CEAH–NAHMS; 1–152. https://www.aphis.usda.gov/animal_health/nahms/dairy/downloads/dairyheifer11/HeiferRaiser.pdf.

4. Leruste H, Brscic M, Heutinck LFM, et al. The relationship between clinical signs of respiratory system disorders and lung lesions at slaughter in veal calves. Prev Vet Med 2012;105:93–100.

5. McGuirk SM. Disease management of dairy calves and heifers. Vet Clin North Am Food Anim Pract 2008;24:139–153.

6. Jim K. Impact of bovine respiratory disease (BRD) from the perspective of the Canadian beef producer. Anim Health Res Rev 2009;10:109–110.

7. Guterbock WM. The impact of BRD: The current dairy experience. Anim Health Res Rev 2014;15:130–134.

8. Wittum TE, Young CR, Stanker LH, et al. Haptoglobin response to clinical respiratory tract disease in feedlot cattle. Am J Vet Res 1996;57:646–649.

9. Gardner BA, Dolezal HG, Bryant LK, et al. Health of finishing steers: Effects on performance, carcass traits, and meat tenderness. J Anim Sci 1999;77:3168–3175.

10. Thompson PN, Stone A, Schultheiss WA. Use of treatment records and lung lesion scoring to estimate the effect of respiratory disease on growth during early and late finishing periods in South African feedlot cattle. J Anim Sci 2006;84:488–498.

11. White BJ, Renter DG. Bayesian estimation of the performance of using clinical observations and harvest lung lesions for diagnosing bovine respiratory disease in post-weaned beef calves. J Vet Diagn Invest 2009;21:446–453.

12. Van Donkersgoed J, Ribble CS, Boyer LG, et al. Epidemiological study of enzootic pneumonia in dairy calves in Saskatchewan. Can J Vet Res 1993;57:247–254.

13. Sivula NJ, Ames TR, Marsh WE, et al. Descriptive epidemiology of morbidity and mortality in Minnesota dairy heifer calves. Prev Vet Med 1996;27:155–171.

14. Carter JN, Meredith GL, Montelongo M, et al. Relationship of vitamin E supplementation and antimicrobial treatment with acute-phase protein responses in cattle affected by naturally acquired respiratory tract disease. Am J Vet Res 2002;63:1111–1117.

15. Berry BA, Confer AW, Krehbiel CR, et al. Effects of dietary energy and starch concentrations for newly received feedlot calves: II. Acute-phase protein response. J Anim Sci 2004;82:845–850.

16. Buczinski S, Rademacher RD, Tripp HM, et al. Assessment of L-lactatemia as a predictor of respiratory disease recognition and severity in feedlot steers. Prev Vet Med 2015;118:306–318.

17. Gruys EM, Toussaint JM, Niewold TA, et al. Acute phase reaction and acute phase proteins. J Zhejiang Univ Sci B 2005;6:1045–1056.

18. Petersen HH, Nielsen JP, Heegard PMH. Application of acute phase protein measurements in veterinary clinical chemistry. Vet Res 2004;35:163–187.

19. Humblet MF, Coghe J, Lekeux P, et al. Acute phase proteins assessment for an early selection of treatments in growing

calves suffering from bronchopneumonia under field conditions. Res Vet Sci 2004;77:41–47.

20. Young CR, Wittum TE, Stanker LH, et al. Serum haptoglobin concentrations in a population of feedlot cattle. Am J Vet Res 1996;57:138–141.

21. Leeflang MMG, Deeks JJ, Gatsonis C, et al. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149:889–897.

22. Deeks JJ, Bossuyt PM, Gatsonis C (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2013. Available at: http://srdta.cochrane.org/.

23. Wilkins W, Rajic A, Parker S, et al. Examining heterogeneity in the diagnostic accuracy of culture and PCR for Salmonella spp. in swine: A systematic review/meta-regression approach. Zoonoses Public Health 2010;57:121–134.

24. Lamb CR, Nelson JR. Diagnostic accuracy of tests based on radiologic measurements of dogs and cats: A systematic review. Vet Radiol Ultrasound 2015;56:231–244.

25. Rutjes AWS, Reitsma JB, Vandenbroucke JP, et al. Case–control and two-gate designs in diagnostic accuracy studies. Clin Chem 2005;51:1335–1341.

26. Leeflang MMG, Rutjes AWS, Reitsma JB, et al. Variation of a test's sensitivity and specificity with disease prevalence. Can Med Assoc J 2013;185:537–544.

27. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2 Group: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;55:529–536.

28. Umscheid CA. A primer on performing systematic reviews and meta-analyses. Clin Infect Dis 2013;57:725–734.

29. Al Qudah KM. Oxidative stress in calves with acute or chronic bronchopneumonia. Rev Méd Vét 2009;160:231–236.

30. Alsemgeest SPM, Kalsbeek HC, Wensing TH, et al. Concentration of serum Amyloid-a (SAA) and haptoglobin (Hp) as parameters of inflammatory diseases in cattle. Vet Q 1994;16:21–23.

31. Angen O, Thomsen J, Larsen LE, et al. Respiratory disease in calves: Microbiological investigations on trans-tracheally aspirated bronchoalveolar fluid and acute phase protein response. Vet Microbiol 2009;137:165–171.

32. Arslan HH, Yavuz O, Nisbet C, et al. Comparative evaluation of the effects of florfenicol and tulathromycin on clinical recovery and acute phase proteins in undifferentiated natural bovine respiratory disease. Rev Méd Vét 2010;161:535–539.

33. Coskun A, Guzelbektes H, Simsek A, et al. Haptoglobin and SAA concentrations and enzyme activities in bronchoalveolar lavage fluids from calves with bronchopneumonia. Rev Méd Vét 2012;163:615–620.

34. Dudek K, Dariusz B, Symanska-Czerwinska M. The immune response in seropositive calves against *Mycoplasma bovis*. Vet Stan 2011;42:182–183.

35. Fathi E, Farahzadi R, Imani M. Approach to treatment of bronchopneumonia by evaluation of selected acute-phase proteins in calf herds. Comp Clin Pathol 2013;22:125–129.

36. Fratric N, Ilic V, Gvozdic D, et al. A hematological profile and plasma proteins of three months old calves with clinical signs of bronchopneumonia. Vet Stan 2011;42:326–331.

37. Ganheim C, Alenius S, Waller KP. Acute phase proteins as indicators of calf herd health. Vet J 2007;173:645–651.

38. Idoate J, Ley BV, Schultz L, Heller M. Acute phase proteins in naturally occurring respiratory disease of feedlot cattle. Vet Immunol Immunopathol 2015;163:221–226.

39. Lee S, Park M, Shah DH, et al. Survey of critical clinicpathological values in relation to neonatal diseases in native calves. Indian Vet J 2005;82:1049–1052.

40. Nazifi S, Rezakhani A, Koohimoghadam M, et al. Evaluation of serum haptoglobin in clinically healthy cattle and cattle with inflammatory diseases in Shiraz, a tropical area in southern Iran. Bulgarian J Vet Med 2008;11:95–101.

41. Nazifi S, Ansari-Lari M, Tabandeh MR, et al. Clinical relevance of serum sialic acids evaluation and correlation with haptoglobin and serum amyloid A in diseased cattle. Bulgarian J Vet Med 2010;13:45–54.

42. Prathaban S, Gnanaprakasam V. Study on plasma fibrinogen level of Indian crossbred cows in health and disease. Indian Vet J 1990;67:453–456.

43. Svensson C, Liberg P, Hultgren J. Evaluating the efficacy of serum haptoglobin concentration as an indicator of respiratory-tract disease in dairy calves. Vet J 2007;174:288–294.

44. Svensson C, Liberg P. The effect of group size on health and growth rate of Swedish dairy calves housed in pens with automatic milk-feeders. Prev Vet Med 2006;73:43–53.

45. Timsit E, Assié S, Quiniou R, et al. Early detection of bovine respiratory disease in young bulls using reticulo-rumen temperature boluses. Vet J 2011;190:136–142.

46. Tothova C, Nagy O, Seidel H, et al. The effect of chronic respiratory diseases on acute phase proteins and selected blood parameters of protein metabolism in calves. Berl Munch Tierarztl Wochenschr 2010;123:307–313.

47. Tothova C, Nagy O, Seidel H, et al. The influence of various inflammatory diseases on the concentrations of acute phase proteins in calves. Vet Stan 2011;42:250–256.

48. Tothova C, Oskar N, Gabriel K. The serum protein electrophoretic pattern and acute phase proteins concentrations in calves with chronic respiratory diseases. Acta Vet Beograd 2013;63:473–486.

49. Wolfger B, Schwartzkopf-Genswein KS, Barkema HW, et al. Feeding behavior as an early predictor of bovine respiratory disease in North American feedlot systems. J Anim Sci 2015;93:377–385.

50. Schalm W, Jain N, Carroll E. Veterinary Haematology. Philadelphia, PA: Lea and Febiger; 1975.

51. Makimura S, Suzuki N. Quantitative determination of bovine serum haptoglobin and its elevation in some inflammatory diseases. Jpn J Vet Sci 1982;44:15–21.

52. Boosman R, Niewold TA, Mutsaers CW, et al. Serum amyloid-A concentrations in cows given endotoxin as an acute phase stimulant. Am J Vet Res 1989;50:1690–1694.

53. Heegaard PMH, Godson DL, Toussaint MJM, et al. The acute phase response of haptoglobin and serum amyloid A (SAA) in cattle undergoing experimental infection with bovine respiratory syncytial virus. Vet Immunol Immunopathol 2000;77:151–159.

54. Godson DL, Campos M, Attah-poku SK, et al. Serum haptoglobin as an indicator of the acute phase response in bovine respiratory disease. Vet Immunol Immunopathol 1996; 51:277–292.

55. Becker U, Bartl K, Whalefeld A. A functional photometric assay for plasma fibrinogen. Thromb Res 1984;35:475–484.

56. Phillips LL, Jenkins EB, Hardaway RM. Comparison of two methods for determination of fibrinogen. Technical bulletin of the Registry of Medical Technologists 1969.

57. Bauer JD, Ackermann PG, Toro G. Clinical Laboratory Methods. St. Louis, MO: The Mosby Company; 1974.

58. Reitsma JB, Glas AS, Rutjes AWS, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982–990.

59. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001;20:2865–2884.

60. Earley B, Crowe MA. Effects of ketoprofen alone or in combination with local anesthesia during the castration of bull calves on plasma cortisol, immunological and inflammatory responses. J Anim Sci 2002;80:1044–1052.

61. Katoh N, Oikawa S, Oohashi T, et al. Decreases of apolipoprotein B-100 and A-I concentrations and induction of haptoglobin and serum amyloid A in nonfed calves. J Vet Med Sci 2002;64:51–55.

62. Murata H, Miyamoto T. Bovine haptoglobin as a possible immunomodulator in the sera of transported calves. Br Vet J 1993;149:277–283.

63. Panciera RJ, Confer AW. Pathogenesis and Pathology of Bovine Pneumonia. Vet Clin North Am Food Anim Pract 2010;26:191–214.

64. Whiting PF, Rutjes AWS, Westwooda ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol 2013;66:1093–1104.

65. Rutjes AW, Reitsma JB, Di NM, et al. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;174:469–476.

66. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. J Am Med Assoc 1999;282:1061–1066.

67. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: Mechanisms, magnitude, and solutions. Clin Chem 2008;54:729–737.

68. Korevaar DA, van Enst AW, Spijker R, et al. Reporting quality of diagnostic accuracy studies: A systematic review and meta-analysis of investigations on adherence. Evid Based Med 2014;19:47–54.

69. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Ann Intern Med 2003;138:1–12.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Search terms used to perform the literature search.