Preventive Veterinary Medicine 135 (2016) 67–73

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Diagnostic accuracy of clinical illness for bovine respiratory disease (BRD) diagnosis in beef cattle placed in feedlots: A systematic literature review and hierarchical Bayesian latent-class meta-analysis

E. Timsit[a,b], N. Dendukuri[c], I. Schiller[c], S. Buczinski[d,*]

[a] Department of Production Animal Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada
[b] Feedlot Health Management Services, Okotoks, AB, Canada
[c] Division of Clinical Epidemiology, McGill University Health Centre, Montreal, QC, Canada
[d] Department of Clinical Science, Faculté de Médecine Vétérinaire, Université de Montréal, St-Hyacinthe, QC, Canada

ARTICLE INFO

Article history:
Received 22 July 2016
Received in revised form 4 November 2016
Accepted 7 November 2016

Keywords:
Diagnostic test
Latent class model
Pen-rider
Imperfect reference
Lung lesion
Shipping fever

ABSTRACT

Diagnosis of bovine respiratory disease (BRD) in beef cattle placed in feedlots is typically based on clinical illness (CI) detected by pen-checkers. Unfortunately, the accuracy of this diagnostic approach (namely, sensitivity [Se] and specificity [Sp]) remains poorly understood, in part due to the absence of a reference test for ante-mortem diagnosis of BRD. Our objective was to pool available estimates of CI's diagnostic accuracy for BRD diagnosis in feedlot beef cattle while adjusting for the inaccuracy in the reference test. The presence of lung lesions (LU) at slaughter was used as the reference test. A systematic review of the literature was conducted to identify research articles comparing CI detected by pen-checkers during the feeding period to LU at slaughter. A hierarchical Bayesian latent-class meta-analysis was used to model test accuracy. This approach accounted for imperfections of both tests as well as the within and between study variability in the accuracy of CI. Furthermore, it also predicted the Se_{CI} and Sp_{CI} for future studies. Conditional independence between CI and LU was assumed, as these two tests are not based on similar biological principles. Seven studies were included in the meta-analysis. Estimated pooled Se_{CI} and Sp_{CI} were 0.27 (95% Bayesian credible interval: 0.12–0.65) and 0.92 (0.72–0.98), respectively, whereas estimated pooled Se_{LU} and Sp_{LU} were 0.91 (0.82–0.99) and 0.67 (0.64–0.79). Predicted Se_{CI} and Sp_{CI} for future studies were 0.27 (0.01–0.96) and 0.92 (0.14–1.00), respectively. The wide credible intervals around predicted Se_{CI} and Sp_{CI} estimates indicated considerable heterogeneity among studies, which suggests that pooled Se_{CI} and Sp_{CI} are not generalizable to individual studies. In conclusion, CI appeared to have poor Se but high Sp for BRD diagnosis in feedlots. Furthermore, considerable heterogeneity among studies highlighted an urgent need to standardize BRD diagnosis in feedlots.

1. Introduction

Bovine Respiratory Disease (BRD) is the most important health problem in the beef feedlot industry (USDA, 2013). Cattle of all ages can be affected by BRD; however, they are most likely to be affected soon after arrival at the feedlot, as they are exposed to a wide range of respiratory pathogens (due to co-mingling) concurrent with various stressors negatively affecting their immune system (Babcock et al., 2010; Taylor et al., 2010).

Diagnosis of BRD in cattle placed in feedlots is based on clinical illness (CI) detected by pen-checkers (Portillo, 2014). Unfortunately, this diagnostic approach is not always accurate (White and Renter, 2009). Cattle are prey animals and consequently will often mask signs of sickness, especially in presence of humans (Weary et al., 2009). Therefore, a proportion of cattle with BRD are never detected by pen-checkers (i.e. false negatives). Furthermore, clinical signs typically used to diagnose BRD (e.g. depression, anorexia, fever) are not always specific to this disease condition (Portillo, 2014). Cattle diagnosed with BRD may thus not be truly affected by this disease (i.e. false positives). However, in the absence of a perfect test for ante-mortem BRD diagnosis, the diagnostic accuracy

Abbreviations: BRD, bovine respiratory disease; BCI, Bayesian credible interval; CAB, Commonwealth Agricultural Bureau; CI, clinical illness; DIC, deviance information criterion; HSROC, hierarchical summary receiver operating characteristic curve; LU, lung lesion; Se, sensitivity; Sp, specificity.

* Corresponding author.
E-mail addresses: eftimsit@ucalgary.ca (E. Timsit), nandini.dendukuri@mcgill.ca (N. Dendukuri), Ian.Schiller@clinepi.mcgill.ca (I. Schiller), s.buczinski@umontreal.ca (S. Buczinski).

http://dx.doi.org/10.1016/j.prevetmed.2016.11.006
0167-5877/

of CI (namely, sensitivity [Se] and specificity [Sp]) remains largely unknown.

Comparing a diagnosis of BRD (based on CI) with the presence of lung lesions (LU) associated with BRD at slaughter would provide estimates of CI's diagnostic accuracy. Indeed, the presence of LU at post-mortem examination is considered a reference test for BRD diagnosis (Caswell et al., 2012). However, as LU can occur before feedlot placement or resolve before slaughter, presence of LU at slaughter is also an imperfect measure of true BRD status of feedlot cattle during the feeding period. Using Bayesian latent class models that took into account imperfect accuracy of CI and LU for BRD diagnosis, White and Renter (2009) estimated Se and Sp of CI at 61.8% (97.5% Bayesian credible interval [BCI]: 55.7–68.4%) and 62.8% (97.5% BCI: 60.0–65.7%), respectively, based on two different populations. This study provided very useful information. However, its internal validity as well as generalizability are questionable, as this study used a model that assumed a constant accuracy of tests across populations, which seems unlikely due to differences in terms of location (Africa *versus* North America), type of cattle (ranch derived *versus* auction market derived), definition of CI and prevalence of disease (low *versus* high). Failure to take into account potential variations of Se and Sp between populations could have significantly impacted the reported estimates of CI's accuracy (Johnson et al., 2009).

Recently, additional studies comparing CI to LU at slaughter have been published (Schneider et al., 2009; Leach et al., 2013; Rezac et al., 2014). These studies, if combined in a meta-analysis taking into account the imperfection of LU, could provide more robust estimates of CI's accuracy. The objective of this study was therefore to evaluate accuracy of CI for BRD diagnosis in feedlots, based on a literature review and meta-analysis of studies that compared CI to LU at slaughter. A hierarchical Bayesian approach to model test accuracy was used to account for imperfection of LU and the fact that CI's accuracy may vary among studies due to within-study characteristics.

2. Materials and methods

A systematic review of the literature and a diagnostic test accuracy meta-analysis were conducted in accordance with PRISMA guidelines (Preferred Reporting Items for Systematic reviews and Meta-Analyses statement; Liberati et al., 2009).

2.1. Systematic review of the literature

A systematic review of the literature was performed to identify published research articles comparing CI detected in beef cattle placed in feedlots to LU detected at slaughter. Original articles in English published up to February 24th, 2016 were searched in CAB (Commonwealth Agricultural Bureau) Abstracts and PubMed/MEDLINE databases. The following keywords were used for the search strategy: cattle (medical subject heading [MeSH] or all fields) AND lung (all fields) AND lesions (all fields) AND feedlot (all fields). The references of retrieved articles were also searched independently and checked for any additional studies not captured by the initial search.

Selection and assessment of eligibility were performed by 2 investigators (ET and SB), with disagreement resolved by discussion. Original articles were included in the meta-analysis based on the following criteria: reporting of results in sufficient detail to allow reconstruction of a contingency table comparing CI detected during the feedlot phase with LU at slaughter; differentiating BRD events that occurred pre- and post-weaning (i.e. no total BRD morbidity since birth); and defining CI based on clinical signs of BRD, with or without elevated rectal temperature. Authors of original articles were not routinely contacted if results were not reported in sufficient detail to be included in the meta-analysis.

The following data were extracted from each included study (where available): basic study characteristics (authors; year of publication); population characteristics (weight; sex; location); diagnostic methods (CI definition; lung lesion scoring system); and outcomes (contingency table comparing CI to LU at slaughter). Data extraction was performed by one author (ET) and verified by a second author (SB).

2.2. Hierarchical Bayesian latent-class meta-analysis

A hierarchical Bayesian latent-class meta-analysis approach previously described by Dendukuri et al. (2012) was used to estimate accuracy of CI (index test) using presence of LU at slaughter as an imperfect reference test. The hierarchical structure took into account both within- and between-study variability, whereas the latent-class structure acknowledged that neither CI nor LU was a perfect test for BRD diagnosis.

The meta-analysis model selected is commonly referred to as the hierarchical summary receiver operating characteristic (HSROC) model (Macaskill et al., 2010). This model assumes that the observed CI result is obtained by dichotomizing an unobserved continuous variable at an unknown cut-off θ_i in the i^{th} study. The unobserved continuous variable is assumed to follow two separate logistic distributions – one among infected animals, and the other among uninfected animals.

This model had 2 levels: (i) a within-study level for study-specific parameters, and (ii) a between-study level for parameters common to all studies. Study-specific parameters included θ_i (the cut-off value for defining a positive test in the study i), α_i (the difference in means of the two logistic distributions in study i). The parameters of the HSROC model were then transformed into the parameters of interest, namely Se_{CIi} (sensitivity of study i for CI) and Sp_{CIi} (specificity of study i for CI). In addition, because of the latent class structure of the model, an estimate for the prevalence of BRD for study i (π_i) was also obtained.

Between-study parameters included Λ (pooled difference in means), σ_α (between-study standard deviation of the difference in means), Θ (pooled mean cut-off value for defining a positive test), σ_θ (between-study standard deviation in the cut-off), β (logarithm of the ratio of the standard deviation of test results among cattle with the disease and among cattle without the disease). These parameters were transformed into pooled Se and Sp of the test under evaluation (CI). The Se and Sp of the reference test, namely Se_{LU} and Sp_{LU}, were assumed to be constant across studies. This assumption was made because: (i) all included studies adapted their lung scoring system based on one report (Bryant et al., 1999); and (ii) they all defined a positive lung based on the presence of at least one lesion, irrespective of the extent of the lesion.

The model selected assumed conditional independence between CI and LU in each study as these 2 tests are not based on similar biological principles. This means that an unmeasured confounding variable that simultaneously impacts the accuracy of both tests is unlikely. Furthermore, the previous study by White and Renter (2009) had suggested that there was no conclusive evidence of conditional dependence between these 2 tests.

For the Bayesian estimation, prior distributions for most between-study parameters were non informative (i.e., objective), but had values that covered a reasonable range (Dendukuri et al., 2012). Prior distributions differed slightly from those previously described (Dendukuri et al., 2012), as a logit rather than a probit link function was used to express sensitivities and specificities (to improve convergence). Hyperprior distributions for Λ, Θ and β were selected so that the resulting marginal distribution of pooled Se and Sp were approximately uniform over the interval (0,1): Λ

$\sim$ U (-5.2, 5.2), $\Theta \sim$ U (-2.6, 2.6) and $\beta \sim$ U (−1.3, 1.3). Both σ^2_α and σ^2_θ parameters followed a gamma distribution: Γ(shape = 4, scale = 2). Informative priors were only used for the accuracy of LU at slaughter, with Se ranging from 60 to 85% (Beta [36.28, 13.76]) and Sp ranging from 80 to 95% (Beta [67.18,9.60]). These priors were based on a previous study (White and Renter, 2009). To evaluate the effect of these informative priors, a sensitivity analysis was conducted using uninformative prior distributions (Beta [1,1]) instead of informative priors.

Posterior distributions were obtained using a Monte Carlo Markov Chain algorithm implemented with WinBUGS software (Medical Research Council and the Imperial College of Science, Technology and Medicine, London, available at http://www.mrc-bsu.cam.ac.uk/bugs; WinBUGS program for HSROC model [logit link] is provided as Supplementary materials). The first 5,000 iterations were discarded as burn-in, whereas the next 110,000 were used to obtain posterior medians of the parameters of interest and their corresponding 95% BCI. Convergence of models was assessed by running multiple chains with disparate initial values and then calculating the Gelman-Rubin statistic to compare within and between chain variability. All chains converged quickly to a common region of the parameter space, as is desirable.

Results of the meta-analysis were also reported in the form of a hierarchical summary receiver operating characteristic curve (HSROC) using R software (R Core Team (2013), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org; SROC plot code is provided as Supplementary materials). This presentation showed how Se_{CI} and Sp_{CI} varied as the threshold for positivity changed. In addition, this plot depicted the 75% and 95% prediction regions which cover possible values of Se_{CI} and Sp_{CI} in future studies. These values were obtained from the predictive distribution of Λ, Θ and β. The deviance information criterion (DIC) was used to compare models with and without informative prior distributions. A lower DIC indicated a better fit of the model, with a difference ≥ 5 points in DIC considered as an important difference (Spiegelhalter et al., 2002).

2.3. Sensitivity analysis

Two extensions to the main model were considered to study the robustness of the findings. Both extensions were only considered exploratory, as a more parsimonious model was preferred given the small number of studies available (Trikalinos et al., 2014). First, the accuracy of LU was allowed to vary across studies by using hierarchical prior distributions over Se_{LU} and Sp_{LU} (Chu et al., 2009). Second, the possibility of conditional dependence between Se_{CI} and Se_{LU} was allowed, hypothesizing that both CI and LU may detect more severe cases and also miss less severe cases. The relative conditional dependence was fixed at varying values ranging from 0-0.3 of the maximum possible conditional dependence, which corresponded to a moderate degree of conditional dependence (Wang et al., *in press*). Higher values were deemed to be highly unlikely as it would require the presence of a covariate that had a very high degree of influence on both Se_{CI} and Se_{LU} simultaneously (Wang et al., *in press*).

3. Results

3.1. Systematic review of the literature

The flow chart of the literature search is shown in Fig. 1. The search of PubMed/MEDLINE and CAB abstracts databases provided a total of 76 studies after eliminating duplicates. Of these, 63 were discarded based on title and abstract. The full text of the remaining 13 studies was examined in more detail and 6 studies did not meet the inclusion criteria; 3 did not provide enough information to cross-classify CI and LU (Young et al., 1996; Reinhardt et al., 2009; Munson et al., 2012); 2 did not differentiate CI diagnosed pre- and post-weaning (Wittum et al., 1996; Bryant et al., 1999) and 1 did not provide a definition for CI (Loneragan et al., 2002).

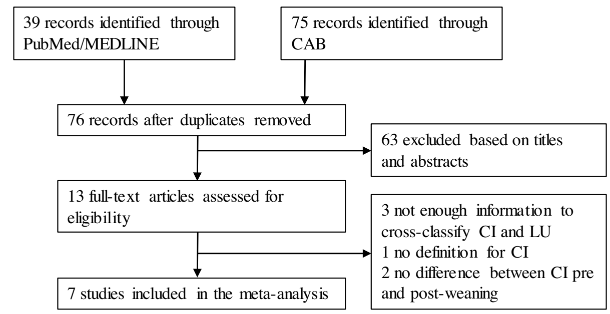

Fig. 1. Flow chart for selection of studies included in the meta-analysis. (adapted from Liberati et al., 2009)

In total, seven studies were identified for inclusion in the meta-analysis (Table 1). These studies were published between 1999 and 2014 and were mostly conducted in North America (6 of 7 studies). From 170 to 19,299 steers or heifers or both were included in these studies, with an average body weight upon arrival at the feedlots ranging from 233 to 344 kg. It is noteworthy that cattle enrolled in the most recent studies were heavier.

3.2. Hierarchical Bayesian latent-class meta-analysis

Posterior medians and 95% BCI of Se_{CI} and Sp_{CI} obtained from informative and non-informative models are shown in Table 2 and Fig. 2. Between-study parameters for both models are presented in Table 3. The model with the smallest DIC was the model with the non-informative prior distribution (171.65 *versus* 178.60). This model estimated the pooled Se_{CI} and Sp_{CI} at 0.27 (0.12–0.65) and 0.92 (0.72–0.98), respectively. Interestingly, studies published after 2006 had a lower Se_{CI} than those published earlier (Table 2).

The medians and 95% BCI for predicted Se_{CI} and Sp_{CI} in a future study were 0.27 (0.01–0.96) and 0.92 (0.14–1.00), respectively. Wide credible intervals around predicted Se_{CI} and Sp_{CI} estimates indicated considerable heterogeneity among studies (also evident in the prediction regions in Fig. 2).

The estimated Se_{LU} and Sp_{LU} for BRD diagnosis were 0.91 (0.82–0.99) and 0.67 (0.64–0.79), respectively, using the non-informative model and were 0.84 (0.80–0.89) and 0.85 (0.74–0.92) using the informative model (Table 3). Therefore, the model with non-informative prior distributions estimated Se_{LU} and Sp_{LU} as higher and lower, respectively, than our prior subjective assumptions.

3.3. Sensitivity analysis

Both sensitivity analyses did not significantly alter results of the main model. When the accuracy of LU was allowed to vary across studies, the pooled Se_{CI} and Sp_{CI} were 0.30 (0.13, 0.66) and 0.95 (0.78, 0.99), respectively. When allowing for a relative covariance of 0.3 between Se_{CI} and Se_{LU}, the pooled Se_{CI} and Sp_{CI} were 0.25 (0.11, 0.64) and 0.91 (0.69, 0.97), respectively.

4. Discussion

Accurate ante-mortem diagnosis of BRD remains a challenge for the beef feedlot industry. This study modelled the diagnostic accuracy of current methods for BRD diagnosis in feedlots,

Table 1
Studies (n = 7) included in the meta-analysis, ordered by year of publication. These studies compared BRD diagnosis based on clinical illness (CI) during the feeding period with the presence of lung lesions associated with BRD (LU) at slaughter.

Study	Location	# of feedlots	# of cattle	# of lungs	Sex (steers [S], heifers [H])	Mean weight or age at arrival	Definition for CI	Definition for LU	CI/LU			
									+/+	+/–	–/+	–/-
Gardner et al. (1999)	USA (Kansas)	1	222	204	S	291 kg ± 29	BRD signs + > 40 °C	Lesion in cranio-ventral lung lobes	49	53	38	64
Buhman et al. (2000)	USA (Texas)	1	170	146	H	249 kg ± 16 and 235 kg ± 18[b]	BRD signs[c]	Lesion in any lung lobe[f]	37	1	90	18
Thompson et al. (2006)	South Africa	2	2036	2036	S, H[a]	233 kg ± 1	BRD signs +/– >40 °C[d]	Lesion in cranio-ventral lung lobes or pleural adhesion[g]	265	196	606	969
Schneider et al. (2009)	USA (Iowa)	10	5976	1665	S (80%), H (20%)	288 kg ± 44	BRD signs[c]	Lesion in any lung lobe or pleural adhesion[h]	121	42	910	592
Leach et al. (2013)	USA (Nebraska)	1	2182	2023	–	165 d ± 17	BRD signs and temperature[e]	Lesion in any lung lobe	195	60	1395	373
Tennant et al. (2014)	USA (Texas)	1	1530	2336	S	312 kg ± 10	BRD signs +/– >40 °C[d]	Lesion in any lung lobe or pleural adhesion[i]	157	29	1344	806
Rezac et al. (2014)	USA (Texas [5], Kansas [1])	6	19299	13191	S (54%), H (46%)	344 kg ± 42	BRD signs and temperature[e]	Lesion in any lung lobe or pleural adhesion[j]	127	157	4591	8316

[a] Predominantly steers but proportions were not reported.

[b] Two groups of heifers were studied.

[c] Rectal temperature was not included in the definition of CI.

[d] Cattle with severe BRD signs were treated even if rectal temperature was below than 40 °C.

[e] Rectal temperature was included in the definition of CI but thresholds used were not reported.

[f] LU scoring: no lesion, <20% of total lung = minor lung lesions, >20% of total lung = major lung lesions.

[g] LU scoring: lung consolidation = 0, <50% on a single lobe, >50% of a single lobe; pleural adhesion = 0, <50% of the pleural surface, >50% of the pleural surface.

[h] LU scoring: 0 = normal, 1 = <1 cranio-ventral lobe and < 5% of lung volume, 2 = pleural adhesions, consolidation in > 1 cranio-ventral lobe and between 5 and 10% total lung, 3 = consolidation in > 1 cranio-ventral lobe and between 10 and 15% of total lung, 4 = > 15% of total lung, 5 = active bronchial lymph nodes.

[i] LU scoring: 0, Fib = pleural adhesion, 5con = <5% lung consolidation, 15con = >5%-<15% lung consolidation, 50con = >15%-<50% lung consolidation and/or pleural adhesion, ALLCON = >50% lung consolidation and/or pleural adhesion.

[j] LU scoring: no, <50% on a single lobe, >50% of a single lobe or pleural adhesion.

Table 2
Posterior medians (95% credible intervals) of sensitivity and specificity of clinical illness for BRD diagnosis and prevalence of BRD in feedlot cattle obtained from 2 hierarchical Bayesian latent-class meta-analysis models (with and without prior information on the accuracy of lung lesions at slaughter).

Study	Model with informative prior[a] (DIC[b] = 178.60)			Model with non informative prior (DIC = 171.65)		
	Sensitivity	Specificity	BRD prevalence	Sensitivity	Specificity	BRD prevalence
Gardner et al. (1999)	0.59 (0.45–0.75)	0.57 (0.46–0.67)	0.40 (0.24–0.52)	0.71 (0.36–0.98)	0.55 (0.46–0.64)	0.17 (0.05–0.33)
Buhman et al. (2000)	0.26 (0.19–0.34)	0.93 (0.49–1.00)	0.97 (0.89–1.00)	0.28 (0.21–0.37)	0.95 (0.72–1.00)	0.89 (0.76–0.99)
Thompson et al. (2006)	0.34 (0.30–0.44)	0.85 (0.83–0.88)	0.40 (0.28–0.48)	0.56 (0.37–0.81)	0.84 (0.82–0.87)	0.17 (0.11–0.31)
Schneider et al. (2009)	0.13 (0.10–0.15)	0.96 (0.93–0.99)	0.68 (0.59–0.75)	0.14 (0.11–0.17)	0.94 (0.92–0.98)	0.50 (0.42–0.62)
Leach et al. (2013)	0.13 (0.11–0.14)	0.87 (0.71–0.99)	0.92 (0.85–0.99)	0.12 (0.10–0.14)	0.86 (0.79–0.94)	0.78 (0.68–0.93)
Tennant et al. (2014)	0.11 (0.09–0.13)	0.99 (0.97–1.00)	0.71 (0.63–0.77)	0.13 (0.11–0.15)	0.98 (0.96–1.00)	0.54 (0.46–0.65)
Rezac et al. (2014)	0.03 (0.02–0.05)	0.98 (0.98–0.99)	0.30 (0.17–0.37)	0.09 (0.03–0.60)	0.98 (0.98–0.98)	0.04 (0.01–0.21)
Pooled	0.20 (0.10–0.46)	0.92 (0.70–0.98)	–	0.27 (0.12–0.56)	0.92 (0.72–0.98)	–

[a] Informative priors were used for the accuracy of lung lesions at slaughter with a sensitivity ranging from 60 to 85% (Beta [36.279, 13.761]) and a specificity ranging from 80 to 95% (Beta [67.181,9.597]).
[b] DIC: Deviance Information Criterion.

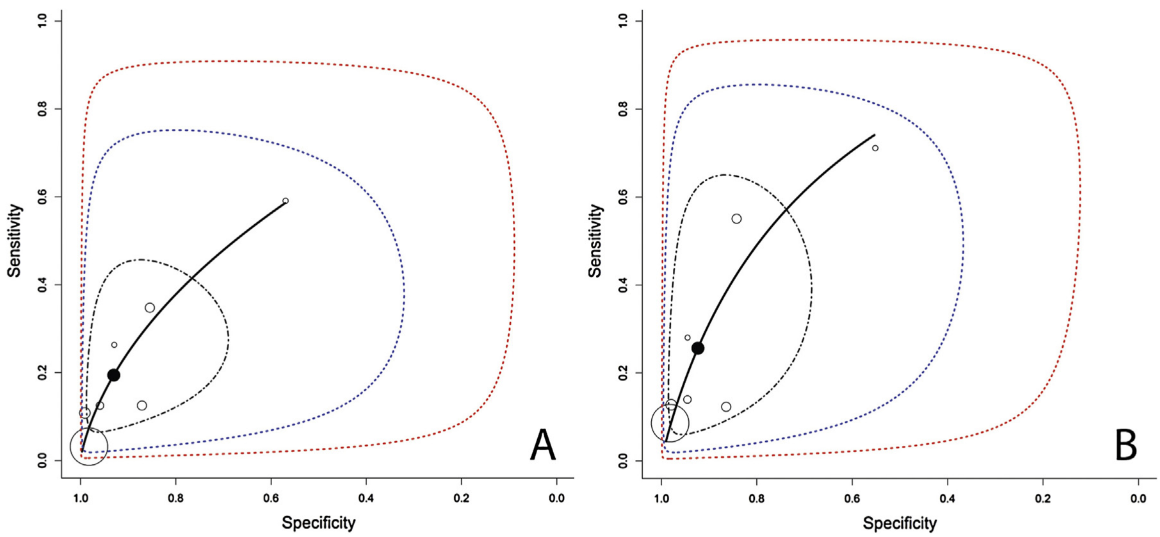

Fig. 2. Summary receiver operating characteristic curves (solid line) of clinical illness for BRD diagnosis resulting from informative priors (A) and non-informative priors (B) for diagnostic accuracy of lung lesions at slaughter. Each clear circle represents the posterior median (sensitivity [Se], specificity [Sp]) pair from an individual study (n = 7). Larger circles reflect a higher sample size (the size is not proportional to the sample size). The black circle marks the pooled Se and Sp across the 7 studies included in the meta-analysis. The black dot-dashed-curve indicates the boundary of the 95% credible region for the pooled estimates of Se and Sp across the 7 studies. The blue dotted-curve and red dotted-curve represent the 75% and 95% prediction regions, respectively, for a future study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Posterior medians (95% Bayesian credible intervals) of between-study parameters obtained from 2 hierarchical Bayesian latent-class meta-analysis models.

Parameters	Model with informative prior[a] (DIC[b] = 178.60)	Model with non informative prior (DIC = 171.65)
Θ	−1.94 (−2.56;-0.68)	−1.74 (−2.52;−0.48)
Λ	0.61 (−1.30;2.41)	1.23 (−0.72;3.54)
β	0.27 (−0.38;1.02)	0.11 (−0.74;1.01)
σ_θ	1.51 (0.92;2.77)	1.47 (0.87;2.73)
σ_α	1.27 (0.55;2.69)	1.29 (0.59;2.73)
Sensitivity LU	0.84 (0.80;0.89)	0.91 (0.82;0.99)
Specificity LU	0.85 (0.74;0.92)	0.67 (0.64;0.79)

[a] Informative priors were used for accuracy of lung lesions (LU) at slaughter with a sensitivity ranging from 60 to 85% (Beta [36.279, 13.761]) and a specificity ranging from 80 to 95% (Beta [67.181,9.597]).
[b] DIC: Deviance Information Criterion.

using LU at slaughter as an imperfect reference test. A systematic review of the literature combined with a hierarchical Bayesian latent-class meta-analysis yielded a pooled Se_{CI} that was low (0.27; 95% BCI = 0.12–0.65) but a pooled Sp_{CI} that was high (0.92; 95% BCI = 0.72–0.98). The meta-analysis also revealed considerable heterogeneity among studies, with 95% credible intervals around predicted Se_{CI} (0.27; 95% BCI = 0.01–0.96) and Sp_{CI} (0.92; 95% BCI = 0.14–1.00) much wider than those around pooled Se_{CI} and Sp_{CI} estimates. Therefore, we inferred that pooled Se_{CI} and Sp_{CI} cannot be generalized to individual studies and more research is needed to explain this heterogeneity.

The hierarchical Bayesian latent-class model used in the present study has numerous advantages (Dendukuri et al., 2012; Held et al., 2015). First, it adjusted for imperfections of the reference test i.e. lung lesion at slaughter. Indeed, with the latent class approach, the test under evaluation (i.e. CI) and the reference test (i.e. LU) were assumed to measure the same unobservable (latent) variable, namely true disease status (Held et al., 2015). Secondly, the hierarchical approach assumed that Se_{CI} and Sp_{CI} can vary across studies due to different cut-off values for defining a positive test, differences in diagnostic accuracy, or both (Dendukuri et al., 2012). To assume an identical Se and Sp of the index test in all studies included in the meta-analysis can be difficult to justify (given the variability in population and design aspect of individual studies) and can lead to biased estimates of test's accuracy (Dendukuri et al., 2012). For example, White and Renter (2009) assumed a constant accuracy of CI and LU across 2 studies (Gardner et al., 1999; Thompson et al., 2006) to model Se_{CI} and Sp_{CI}. However, we determined that Se_{CI} and Sp_{CI} actually varied significantly between these 2 studies (Se_{CI} = 0.59 [0.45–0.75] and Sp_{CI} = 0.57 [0.46–0.67] in Gardner et al., 1999 *versus* Se_{CI} = 0.34 [0.30-0.44] and Sp_{CI} = 0.85 [0.83–0.88] in Thompson et al., 2006). Furthermore, using the same model previously described by White and Renter (2009), a smaller DIC was obtained when Se_{CI} and Sp_{CI} were able to vary between studies (DIC: 49.68 *versus* 57.81; SB, personal communication) indicating a better fit with the data. This highlighted the importance of assumptions when evaluating accuracy of diagnostic tests (Johnson et al., 2009).

A first limitation of the present study was that the small number of studies identified prevented a more detailed exploration of the reasons for heterogeneity among studies using a meta-regression approach. Another limitation could be that the accuracy of LU at slaughter for BRD diagnosis was assumed to be constant across studies. However, when accuracy of LU was allowed to vary across studies (i.e. sensitivity analysis), similar pooled results for Se_{CI} and Sp_{CI} were obtained. Therefore, we chose to use a more parsimonious model that assumed fixed LU accuracy, as parsimonious models are recommended in empirical studies of meta-analyses of small numbers of studies (Trikalinos et al., 2014). It is noteworthy that informative priors were only used for LU accuracy. Indeed, we chose not to use informative priors for CI's accuracy in order to allow posterior densities to be impacted more by the data than by the priors. Furthermore, to the authors' best knowledge, there is very little information available in the scientific literature on pen checker's diagnostic accuracy for BRD diagnosis.

The Se of current methods for BRD diagnosis in feedlot was much lower than expected. Previously, White and Renter (2009) estimated the Se_{CI} for diagnosing BRD in feedlots at 61.8% (97.5% BCI: 55.7–68.4). This discrepancy can be explained by the inclusion of more recent studies, which had a lower Se_{CI}. Indeed, studies published in 2009 and later had estimated Se_{CI} ranging from 9 to 14% (non-informative model) whereas studies published in 2006 and before had estimated Se_{CI} ranging from 28 to 71%. This lower Se_{CI} could firstly be explained by a higher number of cattle to be managed by a pen-checker, and thus a reduced time for detection of illness per animal. Lee et al. (2015) recently reported that the average recommended number of cattle at high-risk of developing BRD to be managed by one pen checker increased from 2,739 to 3,464 head between 2009 and 2014 in the United States and Canada. It could also be explained by the fact that cattle enrolled in the latter studies were heavier. These heavier (and thus probably older) cattle had more chance to develop BRD before entering feedlots and may already have had LU upon arrival. Furthermore, heavier and more mature cattle are less likely to have identifiable clinical signs of sickness than younger, lightweight cattle, which make them more difficult to detect when sick (Portillo, 2014). Pen checkers also spend less time to check heavier animals, as they are at a lower risk to develop BRD during the feeding period (USDA, 2013). Irrespective of the cause, the low Se_{Ci} observed in the present study should encourage veterinarians and producers to improve the Se of BRD diagnosis. Such improvement could be obtained, for example, with health monitoring systems that can detect subtle physiological or behavioural changes associated with BRD that are not readily detectable by clinical observation (White et al., 2016; Wolfger et al., 2015).

The Sp_{CI} for BRD diagnosis was higher than previously reported (White and Renter, 2009), which is positive for the beef industry. However, it is noteworthy that even if BRD is the most common cause of illness in feedlots, its prevalence is usually low during the feeding period (apparent BRD prevalence = 21.2% in cattle < 318 kg when placed in feedlot and 8.8% in cattle > 318 kg when placed; USDA, 2013). Therefore, numerous false-positive BRD cases can still be treated with antibiotics (i.e. standard therapy for BRD). For example, with a Se_{CI} = 27%, a Sp_{CI} = 92% and a BRD prevalence = 30%, the positive predictive value of a BRD diagnosis based on CI would be only 59%. In a context of prudent and rationale use of antibiotics (OIE, 2014), this low positive predictive value should encourage veterinarians and producers to implement a confirmatory test such as thoracic auscultation (DeDonder et al., 2010; Mang et al., 2015) or lung ultrasonography (Rademacher et al., 2014) before initiating BRD therapy. Furthermore, it was recently reported that increasing Sp for BRD diagnosis in feedlots created more rapid, positive change in net economic returns than increasing Se (Theurer et al., 2015). Unfortunately, the number of feedlot veterinary consultants training feedlot employees to use lung auscultation to diagnose BRD is decreasing over the last years in North America and Canada (65.2% in 2009 *versus* 21.7% in 2015) (Lee et al., 2015).

The heterogeneity in CI's diagnostic accuracy among studies was expected, as detection of cattle with BRD based on clinical observation is highly subjective and often described as an art rather than a science (Portillo, 2014). Nevertheless, this high heterogeneity highlighted the importance of better standardizing detection and confirmation of a BRD case. For example, the number of pen checkers involved in BRD detection should be reported and an inter-observer agreement should be performed in studies using multiple pen checkers. With such heterogeneity, how can one be sure that a difference in BRD morbidity observed between 2 treatment groups is real and not due to differences in diagnostic accuracy among pen checkers? Guidelines for reporting the level of reliability and agreement among observers are available (Kottner et al., 2011). Furthermore, methods (e.g. rectal temperature measurement, lung auscultation, etc.) and threshold used for BRD confirmation should also be reported in sufficient detail. For example, 6 of the 7 studies included in the present meta-analysis indicated that rectal temperature was part of CI definition but only 3 studies reported the threshold used. Guidelines for reporting diagnostic accuracy studies are available (STARD 2015 (Bossuyt et al., 2015); STRADAS-paratuberculosis (Gardner et al., 2011)). Finally, to the authors' best knowledge, there is no consensus on the definition of a BRD case. However, such a consensus would reduce discrepancies among studies and improve external validity.

In conclusion, current methods for BRD diagnosis in feedlots based on CI appears to have poor Se but high Sp. Furthermore, there was considerable heterogeneity in CI's diagnostic accuracy among studies, which highlighted an urgent need to standardize BRD diagnosis in feedlot cattle.

Conflict of interest

None of the authors have any financial or personal relationships that could inappropriately influence or bias the content of this paper.

Acknowledgements

We gratefully acknowledge Dr. John Kastelic for editing the manuscript. The authors also acknowledge Dr D.J. Rezac for providing additional data on his study comparing CI to LU at slaughter (Rezac et al., 2014). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.prevetmed.2016.11.006.

References

Babcock, A.H., Renter, D.G., White, B.J., Dubnicka, S.R., Scott, H.M., 2010. Temporal distributions of respiratory disease events within cohorts of feedlot cattle and associations with cattle health and performance indices. Prev. Vet. Med 97, 198–219.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L., Lijmer, J.G., Moher, D., Rennie, D., de Vet, H.C., Kressel, H.Y., Rifai, N., Golub, R.M., Altman, D.G., Hooft, L., Korevaar, D.A., Cohen, J.F., Group, S., 2015. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351, h5527.

Buhman, M.J., Perino, L.J., Galyean, M.L., Wittum, T.E., Montgomery, T.H., Swingle, R.S., 2000. Association between changes in eating and drinking behaviors and respiratory tract disease in newly arrived calves at a feedlot. Am. J. Vet. Res. 61, 1163–1168.

Bryant, L.K., Perino, L.J., Griffin, D.R.D.A., Wittum, T.E., 1999. A method for recording pulmonary lesions of beef calves at slaughter, and the association of lesions with average daily gain. Bov. Pract. 33, 163–173.

Caswell, J.L., Hewson, J., Slavic, E., DeLay, J., Bateman, K., 2012. Laboratory and postmortem diagnosis of bovine respiratory disease. Vet. Clin. North Am. Food Anim. Pract. 28, 419–441.

Chu, H., Chen, S., Louis, T.A., 2009. Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. J. Am. Stat. Assoc. 104, 512–523.

DeDonder, K.D., Thomson, D.U., Loneragan, G.H., Noffsinger, T., Taylor, W., Apley, M., 2010. Lung auscultation and rectal temperature as a predictor of lung lesions and bovine respiratory disease treatment outcome in feedyard cattle. Bov. Pract. 44, 146–153.

Dendukuri, N., Schiller, I., Joseph, L., Pai, M., 2012. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics 68, 1285–1293.

Gardner, B.A., Dolezal, H.G., Bryant, L.K., Owens, F.N., Smith, R.A., 1999. Health of finishing steers: effects on performance, carcass traits, and meat tenderness. J. Anim. Sci. 77, 3168–3175.

Gardner, I.A., Nielsen, S.S., Whittington, R.J., Collins, M.T., Bakker, D., Harris, B., Sreevatsan, S., Lombard, J.E., Sweeney, R., Smith, D.R., Gavalchin, J., Eda, S., 2011. Consensus-based reporting standards for diagnostic test accuracy studies for paratuberculosis in ruminants. Prev. Vet. Med. 101, 18–34.

Held, U., Brunner, F., Steurer, J., Wertli, M.M., 2015. Bayesian meta-analysis of test accuracy in the absence of a perfect reference test applied to bone scintigraphy for the diagnosis of complex regional pain syndrome. Biom. J. 57, 1020–1037.

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

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M., Streiner, D.L., 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J. Clin. Epidemiol. 64, 96–106.

Leach, R.J., Chitko-McKown, C.G., Bennett, G.L., Jones, S.A., Kachman, S.D., Keele, J.W., Leymaster, K.A., Thallman, R.M., Kuehn, L.A., 2013. The change in differing leukocyte populations during vaccination to bovine respiratory disease and their correlations with lung scores, health records, and average daily gain. J. Anim. Sci. 91, 3564–3573.

Lee, T.L., Terrell, S.P., Bartle, S.J., Reinhardt, C.D., Apley, M.D., Rethorst, D., Thomson, D.U., 2015. Current feedlot cattle health and well-being program recommendations in the United States and Canada: the 2014 feedlot veterinary consultant survey. Bov. Pract. 49, 124–131.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Medicine 6, e1000100.

Loneragan, G.H., Morley, P.S., Wagner, J.J., Mason, G.L., Yost, G.S., Thoren, M.A., Wittum, T.E., Bray, T.M., 2002. Effects of feeding aspirin and supplemental vitamin E on plasma concentrations of 3-methylindole 3-methyleneindolenine-adduct concentrations in blood and pulmonary tissues, lung lesions, and growth performance in feedlot cattle. Am. J. Vet. Res. 63, 1641–1647.

Macaskill P., Gatsonis C., Deeks J., Harbord R., Takwoingi Y., 2010. Analysing and presenting results. In Hand-book for Diagnsotic Test Accuracy Reviews Version 1.0, J.J. Deeks, P. M. Bossuyt, and C. Gatsonis (eds), The Cochrane Collaboration (http://srdta.cochrane.org/).

Mang, A.V., Buczinski, S., Booker, C.W., Timsit, E., 2015. Evaluation of a computer-aided lung auscultation system for diagnosis of bovine respiratory disease in feedlot cattle. J. Vet. Intern. Med. 29, 1112–1116.

Munson, R.D., Thomson, D.U., Reinhardt, C.D., 2012. Effects of delayed steroid implanting on health, performance, and carcass quality in high health risk, auction market-sourced feedlot steers. J. Anim. Sci. 90, 4037–4041.

OIE, World Organisation for Animal Health, 2014, Chapter 6.9, Terrestrial Animal Health Code (2014), Responsible and prudent use of antimicrobial agents in veterinary medicine. http://www.oie.int/international-standard-setting/terrestrial-code/access-online/. (Accessed 30 November 2015).

Portillo, T.A., 2014. Pen riding and evaluation of cattle in pens to identify compromised individuals. In: American Association of Bovine Practitioners Annual Conference, Albuquerque, pp. 5–8.

Rademacher, R.D., Buczinski, S., Tripp, H.M., Edmonds, M.D., Johnson, E.G., 2014. Systematic thoracic ultrasonography in acute bovine respiratory disease of feedlot steers: impact of lung consolidation on diagnosis and prognosis in a case-control study. Bov. Pract. 48, 1–10.

Reinhardt, C.D., Busby, W.D., Corah, L.R., 2009. Relationship of various incoming cattle traits with feedlot performance and carcass traits. J. Anim. Sci. 87, 3030–3042.

Rezac, D.J., Thomson, D.U., Bartle, S.J., Osterstock, J.B., Prouty, F.L., Reinhardt, C.D., 2014. Prevalence, severity, and relationships of lung lesions, liver abnormalities, and rumen health scores measured at slaughter in beef cattle. J. Anim. Sci. 92, 2595–2602.

Schneider, M.J., Tait Jr., R.G., Busby, W.D., Reecy, J.M., 2009. An evaluation of bovine respiratory disease complex in feedlot cattle: Impact on performance and carcass traits using treatment records and lung lesion scores. J. Anim. Sci. 87, 1821–1827.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Series B 64, 583–639.

Taylor, J.D., Fulton, R.W., Lehenbauer, T.W., Step, D.L., Confer, A.W., 2010. The epidemiology of bovine respiratory disease: What is the evidence for predisposing factors? Can. Vet. J. 51, 1095–1102.

Tennant, T.C., Ives, S.E., Harper, L.B., Renter, D.G., Lawrence, T.E., 2014. Comparison of tulathromycin and tilmicosin on the prevalence and severity of bovine respiratory disease in feedlot cattle in association with feedlot performance, carcass characteristics, and economic factors. J. Anim. Sci. 92, 5203–5213.

Theurer, M.E., White, B.J., Larson, R.L., Schroeder, T.C., 2015. A stochastic model to determine the economic value of changing diagnostic test characteristics for identification of cattle for treatment of bovine respiratory disease. J. Anim. Sci. 93, 1398–1410.

Thompson, P.N., Stone, A., Schultheiss, W.A., 2006. 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. 84, 488–498.

Trikalinos, T.A., Hoaglin, D.C., Schmid, C.H., 2014. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. Stat. Med. 33, 1441–1459.

USDA, 2013. Feedlot 2011 Part IV: Health and Health Management on US feedlots with a capacity of 1000 or more Head, https://www.aphis.usda.gov/animal_health/nahms/feedlot/downloads/feedlot2011/Feed11_dr_PartIV.pdf. Accessed 28 August 2015.

Wang, Z., Dendukuri, N., Joseph, L., 2016. Understanding the effects of conditional dependence in research studies involving imperfect diagnostic tests. Stat. Med., in press.

Weary, D.M., Huzzey, J.M., von Keyserlingk, M.A., 2009. Board-invited review: Using behavior to predict and identify ill health in animals. J. Anim. Sci. 87, 770–777.

White, B.J., Renter, D.G., 2009. 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. 21, 446–453.

White, B.J., Goehl, D.R., Amrine, D.E., Booker, C., Wildman, B., Perrett, T., 2016. Bayesian evaluation of clinical diagnostic test characteristics of visual observations and remote monitoring to diagnose bovine respiratory disease in beef calves. Prev. Vet. Med. 126, 74–80.

Wittum, T.E., Woollen, N.E., Perino, L.J., Littledike, E.T., 1996. Relationships among treatment for respiratory tract disease, pulmonary lesions evident at slaughter, and rate of weight gain in feedlot cattle. J. Am. Vet. Med. Assoc. 209, 814–818.

Wolfger, B., Timsit, E., White, B.J., Orsel, K., 2015. A systematic review of bovine respiratory disease diagnosis focused on diagnostic confirmation, early detection, and prediction of unfavorable outcomes in feedlot cattle. Vet. Clin. North Am. Food Anim. Pract. 31, 351–365.

Young, C.R., Wittum, T.E., Stanker, L.H., Perino, L.J., Griffin, D.D., Littledike, E.T., 1996. Serum haptoglobin concentrations in a population of feedlot cattle. Am. J. Vet. Res. 57, 138–141.