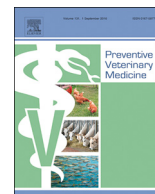Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Validation of a clinical scoring system for bovine respiratory disease complex diagnosis in preweaned dairy calves using a Bayesian framework

S. Buczinski*, G. Fecteau, J. Dubuc, D. Francoz

Faculté de médecine vétérinaire, Université de Montréal, Saint-Hyacinthe, J2S 2M2, Québec, Canada

ABSTRACT

Bovine respiratory disease complex is a major cause of illness in dairy calves. The diagnosis of active infection of the lower respiratory tract is challenging on daily basis in the absence of accurate clinical signs. Clinical scoring systems such as the Californian scoring system, are appealing but were developed without considering the imperfection of reference standard tests used for case definition. This study used a Bayesian latent class model to update Californian prediction rules. The results of clinical examination and ultrasound findings of 608 preweaned dairy calves were used. A model accounting for imperfect accuracy of thoracic ultrasound examination was used to obtain updated weights for the clinical signs included in the Californian scoring system. There were 20 points (95% Bayesian credible intervals: 11–29) for abnormal breathing pattern, 16 points (95% BCI: 4–29) for ear drop/head tilt, 16 points (95% BCI: 9–25) for cough, 10 points (95% BCI: 3–18) for the presence of nasal discharge, 7 points (95% BCI: −1 to 8) for rectal temperature ≥39.2 °C, and −1 points (95% BCI: −9 to 8) for the presence of ocular discharge. The optimal cut-offs were determined using the misclassification cost-term term (MCT) approach with different possible scenarios of expected prevalence and different plausible ratio of false negative costs/false positive costs. The predicted probabilities of active infection of the lower respiratory tract were also obtained using posterior densities of the main logistic regression model. Depending on the context, cut-off varying from 9 to 16 can minimized the MCT. The optimal cut-off decreased when expected prevalence of disease and false negative/false positive ratio increased.

1. Introduction

The diagnosis of bovine respiratory disease complex (BRD) has been of interest for many researchers over the past 30 years. The BRD complex is one of the 2 most frequent diseases that can affect young-stock cattle including dairy, veal, preweaned beef, and feedlot calves (Assie et al., 2004; McGuirk, 2008; Pardon et al., 2012; Woolums et al., 2013). In dairy calves specifically, BRD is commonly enzootic in a herd with periodic outbreaks (Ames, 1997).

Various systemic clinical signs (fever, depression, anorexia) and respiratory signs (nasal discharge, tachypnea, dyspnea, cough, etc…) characterize the classical BRD presentation (McGuirk, 2008). Due to the variety of infectious agents involved in BRD, the clinical signs vary in terms of intensity and duration. The term of subclinical BRD can also be used for defining calves not detected as sick by clinical examination but presenting lung lesions (e.g. detected at slaughter or by ultrasound examination). The primary surveillance of BRD is generally performed by the farmers or their workers who observe calves daily (especially during the feeding period). However, this BRD detection strategy, based on human observation is most often considered as an art rather than a science (Portillo, 2014). Pen riders detecting BRD in feedlot calves are good examples of this concept. It is difficult to "teach" how to be a good pen rider (Portillo, 2014). For these reasons, using a simple and practical systematic approach is of interest to develop a common BRD definition and a structured examination process. Two of them have gained in popularity for dairy calves in North America (McGuirk, 2008; Love et al., 2014). The first scoring system is based on 5 clinical signs assessment (rectal temperature, cough, nasal and ocular discharges, and ear position) which are categorized in 4 (0–3) different ordinal levels (McGuirk, 2008). The accuracy of this clinical scoring system was recently assessed using a Bayesian latent class framework (Buczinski et al., 2015b). The score sensitivity (Se) for detection of active BRD was 62.4% (95% Bayesian credible intervals (BCI): 47.9–75.8%) and specificity (Sp) was 74.1% (95% BCI: 64.9–82.8%). Recently, a second clinical scoring system was developed considering previous clinical signs included in Wisconsin score and trying to assess weight of each clinical signs based on logistic regression coefficients (Aly et al., 2014; Love et al., 2014). One of the advantages of this new clinical scoring

system is that each clinical sign is assessed using a dichotomous way (normal vs. abnormal). The weight attributed for each selected clinical item was determined using a conditional logistic regression analysis. Two different BRD case definitions were used in this study. The calves with a positive nasopharyngeal PCR for viral pathogens (bovine respiratory syncytial virus, herpesvirus type 1, or bovine viral diarrhea virus) were defined as cases. The calves with a nasopharyngeal swab culture positive for an aerobic respiratory pathogen (*Pasteurella multocida, Mannheimia haemolytica, Histophilus somni or Bibersteinia trehalosi)* or *Mycoplasma* spp. were defined as cases if their Wisconsin score was ≥ 5. The rounded regression coefficients obtained from logistic regression models were used as prediction rule's weight for each retained clinical sign. Three different scoring systems were reported but the third presented model (BRD3 score) using a threshold ≥ 5 to define a case was selected for further use by the authors (Aly et al., 2014; Love et al., 2014, 2016). One of the limitations of the methodology presented for obtaining clinical signs "weights" was that the calves' BRD definition was determined partly using Wisconsin score results (80% of BRD cases and 71% of control cases were classified according to their Wisconsin score result). This implies that the same clinical signs for which the authors tried to define accuracy were also included in case definition. This "testimation" problem could potentially artificially inflate prediction weight due to collinearity between the test to assign BRD status (Wisconsin score) and the clinical signs assessed (same clinical signs than the Wisconsin score) as previously described (Steyerberg, 2008). Attribution of weight for prediction rules and their validation is an important but challenging task in medical science (Toll et al., 2008; Collins et al., 2015). This challenge is even more important in the absence of a gold standard to define the true disease status of each patient (Magder and Hughes, 1997; McInturff et al., 2004). Bayesian latent class models have received a lot of attention in the recent years to study diagnostic tests in the absence of gold standard (Branscum et al., 2005; van Smeden et al., 2014). The use of Bayesian methodology is also flexible and allows incorporation of classification error in the reported outcome from logistic regression modeling (McInturff et al., 2004). It would therefore be of interest to assess the accuracy of the Californian scoring system considering the imperfection of thoracic ultrasonography.

Our hypothesis was that the weights of clinical signs used in Californian score would differ from the originally published score when using Bayesian latent class modeling. The main objective of the study was therefore to update the Californian prediction rules system using a Bayesian framework to attribute score weights using thoracic ultrasonography as an imperfect reference standard to determine BRD status. A second objective of the study was to determine the optimal strategy for using this updated test in situations with different clinical settings (expected prevalence of disease and relative cost of false negative/false positive cases).

2. Materials and methods

2.1. Animals selection and thoracic examination diagnostic test

The database that was used for this study validation was described in a previous cross-sectional study which aimed to assess the prevalence of lung lesions using thoracic ultrasonography in pre-weaned dairy calves from the Province of Québec, Canada (Buczinski et al., 2018). The sample size calculation for the number of herds to include in that previous study was based on an expected prevalence of 10% of lung consolidation with a 7.5% precision estimate. In participating herd, 6–12 preweaned female calves randomly selected from all preweaned calves were included in summer 2015 and winter 2016. A total of 608 calves from 39 herds were then recruited and the clinical signs used in the Californian scoring system were collected on these calves during farms visits. A systematic bilateral thoracic ultrasonography (TUS) was also performed in all enrolled calves by one experienced veterinarian (SB) and a recently graduated veterinarian who received a specific training on TUS. The agreement for finding lung consolidation using this training technique was previously reported with κ values from 0.6 to 1.0 (moderate to perfect) in preweaned dairy calves (Buczinski et al., 2013). The lung field caudal to the heart (right and left) and the cranial right lung field were screened as previously reported (Buczinski et al., 2016; Ollivett and Buczinski, 2016). The TUS examination was positive if consolidation (depth ≥ 1 cm) was diagnosed. The TUS was negative if consolidation was < 1 cm (Buczinski et al., 2015b).

2.2. Clinical signs data collection

The clinical signs data were collected as described elsewhere (Aly et al., 2014). These clinical signs were assessed by 2 different operators (not the same than those performing TUS examination) who received the same background information on clinical scoring using the reported visual Californian scoring chart (Aly et al., 2014). Briefly, the presence of nasal (NAS) or ocular (EYE) discharge were recorded as well as the presence of ear drop/head tilt (EAR), the presence of spontaneous cough during the examination (COUGH), abnormal breathing (BREATH) or increased rectal temperature (≥ 39.2 °C; TEMP using a digital thermometer (#8076, Formedica, Montréal, QC, CAN)). These dichotomous variables were recorded for every calf as well as its TUS tests results. The clinical signs were also assessed using Wisconsin scoring rules (McGuirk, 2008).

2.3. Statistical analyses

The analyses were performed using different software (SAS, v9.4, Cary, NC and OpenBUGS, version 3.2.3 rev 1012, MRC, UK).

The general assumption for the model was that there was no covariance between ultrasonography test results and clinical signs data included in the Californian scoring system. Because clinical signs and
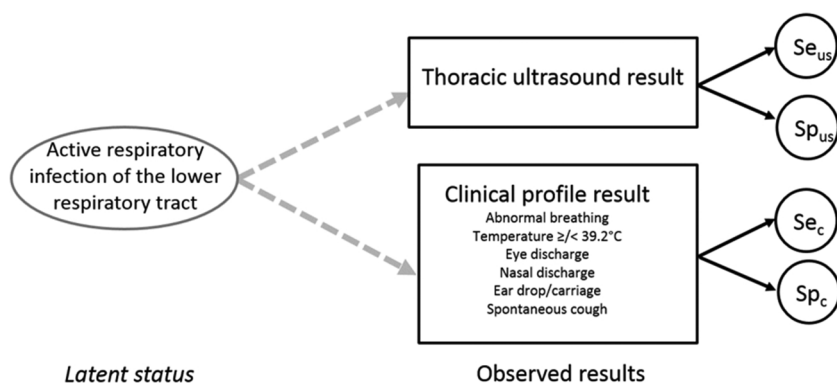

Fig. 1. Diagram representation of the latent class model used for determining the clinical score accuracy for the diagnosis of active infection of the lower respiratory tract in dairy calves. The rectangles correspond to observed variables (ultrasound result and clinical score results) and the oval represents the latent variable (active infection of the lower respiratory tract). The circles represent the sensitivity and specificity of ultrasound (Se_{us} and Sp_{us}) and clinical score (Se_c and Sp_c).

ultrasonography assess different biological processes we thought that these assumptions were most likely true. Another assumption of the model was that the accuracy of ultrasonography would be the same for all calves included in the study.

2.3.1. Prediction rules of score using logistic regression model considering the imperfect nature of ultrasound examination

The latent variable was an active infection of the lower respiratory tract that could be either bacterial, viral or of mixed etiology. This latent variable is applicable because it requires an immediate therapeutic action for either limiting bacterial growth (in cases of bacterial pneumonia) or bacterial complication in case of initial viral infection. Healthy animals, animals sick with other diseases than lower respiratory tract BRD or previously affected by BRD (with no further associated signs of inflammation/infection) were therefore not included in that definition. The general framework is illustrated in Fig. 1.

The probability of the ith calf for being TUS positive (P_{us}^+) was described as a function of the probability of the calf being BRD positive (PBRD+), TUS sensitivity (Se_{us}) and specificity (Sp_{us}).

$$P_{us}^+{}_i = PBRD^+{}_i \times Se_{us} + (1 - PBRD^+{}_i) \times (1 - Sp_{us}) \quad (1)$$

with Se_{us} and Sp_{us} priors assumed to follow beta distributions which are specified in the next section.

For the ith calf, the latent BRD status (Y_i) was assumed a Bernoulli event defined by the probability of being BRD affected ($PBRD^+$ defined as the probability of active infection of the lower respiratory tract (ie. latent variable)).

$$Y_i \sim \text{Bernoulli} (PBRD^+) \quad (2)$$

A general mixed logistic regression model was built using all clinical signs from the Californian scoring chart:

$$\text{Logit } PBRD^+ = \beta 7 + \beta 1 \times BREATH + \beta 2 \times TEMP + \beta 3 \times EYE + \beta 4 \times NAS + \beta 5 \times EAR + \beta 6 \times COUGH + \varepsilon$$

The herd was considered as a random effect (Robert et al., 2012). The specific herd term (ε) accounted for data structure with calves clustered within specific farms:

(4) $\varepsilon \sim \text{norm} (0;\tau)$ and $\tau \sim \text{dgamma} (1,1)$

The choice of gamma ($\gamma = 1$, $\gamma = 1$) for precision specification was considered as reasonably non-informative (Gelman, 2006) with $\gamma \leq 1$ avoiding $\gamma \to 0$ (eg 0.01 or 0.001). Relaxing to gamma (0.001, 0.001) did not significantly change the posterior density estimates (supplemental files).

Main model implementation (**MODEL 1**) was performed using OpenBUGS. The priors used for modeling TUS accuracy were based on previously reported posterior results of TUS accuracy ($Se_{us} = 79.4\%$ (95%BCI, 66.4–90.0%); $Sp_{us} = 93.9\%$ (95%BCI, 88.0–97.6%)) obtained in preweaned dairy calves populations different from the present study (Buczinski et al., 2015b). The corresponding beta distribution chosen for TUS were beta (27.02, 7.92) for Se_{us} (best guess 79%, 5th percentile 65%) and beta (80.58, 6.08) for Sp_{us} (best guess 94%, 5th percentile 88%). Since it is not clinically intuitive to define informed priors on logistic regression parameters from Eq. (4), we used the method of conditional means priors elicitation of different clinical profiles by two different experts on calf health (McInturff et al., 2004).

The priors of $PBRD^+$ depending on the different clinical profiles (x_i:x_1, ..., x_7) that were assessed and $PBRD^+$ probabilities (y_i:p_1, ..., p_7) used were used for determining the vector [β] as described by McInturff et al. (2004) using the inverse relation:

$$\log\left(\frac{y_i}{1-y_i}\right) = \beta_7 + \sum_{i=1}^{6} \beta_i X_i + \varepsilon = [\beta X] \quad (5)$$

$$[\beta] = [X]^{-1} \times \text{logit}(y) \quad (6)$$

Briefly, two experts on dairy calves' health (GF, DF) were independently solicited to elicit probability of active infection of the lower respiratory tract based on 7 different clinical profiles (Table 1). For example, for the first clinical profile, the experts were asked to consider the example of a calf with no abnormal clinical signs to be affected by active infection of the lower respiratory tract. The second clinical profile was a hypothetical calf with increased rectal temperature, abnormal breathing pattern and nasal discharge. The experts were asked for their best guess for $PBRD^+$ of 7 hypothetical different calves as well as their 5th or 95th percentile for this prior distribution. The experts were not aware of the score distribution relative to ultrasonographic results. This blinding aimed avoiding bias on $PBRD^+$ assignment.

The model was based on a total of 50,000 iterations using a 5000 burn-in. Three different chains with different inits values were performed for each model. Rapid mixing and stationary distribution were searched as signs of good modeling ability. The convergence of the model was checked using visual trace-plots and Gelman-Rubin statistic plot. Autocorrelation was detected using autocorrelation plots and thinning was performed when required. The distribution of BRD probabilities and 95% BCI of posterior densities from different clinical profiles were then obtained as practical ways for interpreting the study findings and uncertainty for the different possible clinical presentations. These clinical profiles were obtained from all possible permutations of clinical signs (6 dichotomous clinical signs with 2^6 (64) possible permutations obtained with Interactive Matrix Language (IML) of SAS software)

2.3.2. Sensitivity analysis

Sensitivity analysis was performed using non-informative priors for all probability densities (**MODEL 2**). In this model, all probabilities profiles (p1,..,p7) were assumed following a uniform non-informative probability density from 0 to 1 (beta (1,1)). A third model (**MODEL 3**) was built using non-informative priors for ultrasonography accuracy which was also assumed to follow the same flat distribution. The deviance information criteria (DIC) and effective number of models' parameters (pD) were noted as indicators of model fit accounting for overfitting risk (Spiegelhalter et al., 2002). A DIC difference of 5 or more was considered as indicative of a better fit although this is a debatable issue (Adrion and Mansmann, 2012).

2.3.3. From logistic regression parameters to prediction rules

The coefficients obtained from the main model (MODEL 1) regression analysis were used for using the score weight of every parameter included in the main model. The score's weights were obtained after rounding logistic regression coefficient parameters multiplied by 10 (Moons et al., 2002; Toll et al., 2008). All possible obtained scores from the calves were then used as possible cut-offs (defining positive score if $\geq$ to cut-off values, negative if $<$ to cut-off) to obtain cross-classification (2 by 2 tables) with lung consolidation results. The sensitivity

Table 1

Conditional mean prior distributions based on two experts' opinion for active infection of the lower respiratory tract in dairy calves based on different clinical signs combinations.

i	Breathing	T	Eye discharge	Nasal discharge	Ear drop/head tilt	Cough	Expert 1 Guess	Expert 1 Range	Expert 2 Guess	Expert 2 Range	Priors Mean	Priors ext	Beta distribution
1	Normal	< 39.2 °C	Absent	Absent	Absent	No cough	0.01	0–0.1	0.03	0.01–0.07	0.02	0.1	Beta (1.839, 42.112)
2	Rapid or abnormal breathing	≥39.2 °C	Absent	Present	Absent	No cough	0.5	0.3–0.75	0.9	0.7–0.98	0.7	0.3	Beta (3.641, 2.132)
3	Rapid or abnormal breathing	< 39.2 °C	Absent	Absent	Present	Spontaneous cough	0.9	0.75–1.0	0.85	0.7–0.98	0.875	0.7	Beta (18.56, 3.509)
4	Rapid or abnormal breathing	< 39.2 °C	Absent	Absent	Absent	Spontaneous cough	0.5	0.25–0.75	0.6	0.3–0.9	0.55	0.25	Beta (4.052, 3.497)
5	Rapid or abnormal breathing	≥39.2 °C	Absent	Absent	Absent	No cough	0.5	0.25–0.75	0.5	0.2–0.85	0.5	0.2	Beta (3.262, 2.262)
6	Normal	≥39.2 °C	Absent	Present	Absent	No cough	0.25	0.1–0.75	0.8	0.5–0.9	0.525	0.1	Beta (1.5, 1.453)
7	Normal	≥39.2 °C	Present	Absent	Absent	No cough	0.5	0.25–0.75	0.4	0.1–0.6	0.45	0.75	Beta (3.497, 4.052)

The 2 experts were asked to give their best probability guess for the presence of an active infection of the lower respiratory tract in a calf i with different clinical profiles as well as the values they were 90% probability range (5th and 95th percentile).

The mean guess and larger probability range value were used for generating beta distributions.

and specificity for every cut-off was obtained using a latent class model using 1 population 2 independent tests modeling. Non-informative priors (beta(1,1)) were used for score Se_c/Sp_c and prevalence of BRD. The prior of TUS were the same as in model 1. Posterior distributions were obtained in OpenBUGS using the same method that previously described (see Section 2.3.1). Median estimate and 95% BCI of score sensitivity and specificity for each cut-off were then compiled.

2.3.4. Determination of the optimal cut-off chosen to detect BRD in dairy calves

The accuracy of the scores using across all possible cut-offs was explored using the misclassification cost-term (MCT) approach considering the differential cost of false negative versus false positive cases (Greiner, 1996; Dufour et al., 2017). The MCT was calculated for each specific cut-off using the equation:

$$MCT = (1 - PBRD^+) * (1 - Sp_c) + r*PBRD^+ * (1 - Se_c) \quad (7)$$

This term depends on the prevalence of BRD ($PBRD^+$), the false-negative to false-positive cost ratio (r), and the sensitivity and specificity of the score at a specific cut-off (Se_c, Sp_c). The minimum value of MCT can be considered as the value which minimizes the costs. The plausible ranges for the relative costs of false negative to false positive cases ratio are presently unknown for dairy calves. We used wide ranges that were obtained on a previous study where four different experts were asked to determine this value in feedlot calves (Buczinski et al., 2015a). In the absence of specific studies on dairy calves reporting this ratio, we assumed plausible ranges variations (1:1, 3:1; 8:1, and 20:1) indicating that the cost of false negative case is generally higher than a false positive case. We thought that this general rule applies in dairy calves, however, the last ratio (20:1) was perceived as a low probability event.

3. Results

Data from a total of 608 preweaned calves were included for this study (no missing data). Two hundred twenty (36.2%) calves showed evidence of lung consolidation (≥1 cm consolidation). The repartition of abnormal clinical signs conditional to ultrasound findings are presented in Fig. 2.

The prediction according to different clinical profiles provided by the 2 clinical experts are summarized in Table 1. Convergence and good mixing were obtained rapidly for all the models. The regression parameters obtained from the different modeling strategies are summarized in Table 2. The expected predicted probabilities for the 7 different clinical profiles are summarized in Fig. 3. Briefly, the different priors had a relatively low impact on posterior densities of all tested models (Table 2, Supplemental file). The BRD probabilities and their 95% BCI are presented in Fig. 4 for all possible 64 clinical signs permutations using the posterior distributions of Model 1.

The regression coefficients were obtained (using rounded value of 10*β (Moons et al., 2002)) and the final weight attributed to every clinical sign is presented in Table 3 as well as comparison of the relative weight of each clinical sign when compared with the original BRD3 Californian score proposed by Love et al. (2014) and Aly et al. (2014). The Se_c and Sp_c were then obtained for all possible scores obtained in the database (Table 4) and associated accuracy was recorded in Fig. 4. The MCT analysis is presented in Fig. 5 using a BRD prevalence of 5, 20 and 40% respectively. In a high prevalence scenario (40% prevalence as in a clinical outbreak) or for r = 1:20, a very low score cut-off would be the optimal choice (eg treating all calves with one abnormal sign). For a low prevalence scenario and for r of 1:8 or less, a cut-off of 13 or 15 would minimize MCT. In average BRD prevalence situation (20%), a cut-off between 9 to 16 would minimize MCT depending on r values.
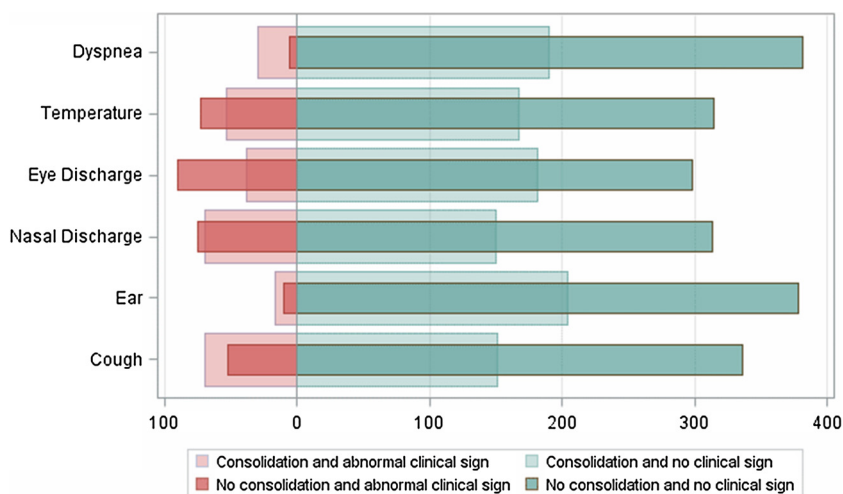
Fig. 2. Repartition of clinical signs observed in 608 pre-weaned dairy calves in association with their and consolidation status.

For each clinical sign included in the scoring system, the calves with no abnormal clinical sign are indicated in green and the calves with abnormal clinical sign are indicated in red. The relative number of calves with lung consolidation conditional on their clinical sign status (normal or abnormal) is graphically represented as the relative length of dark (no consolidation) vs pale (consolidation) colored bars. Lung consolidation was defined as a calf with at least one site of visible lung tissue ≥ 1 cm of depth when performing thoracic ultrasonography. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Using detection strategy that can be used daily is of key importance to adequately detect and manage BRD in dairy calves. A simple clinical scoring rule appears as a promising tool for the farmers and their workers. However, it is crucial to establish an unbiased estimate of the accuracy of this scoring rule. This study considered the absence of gold standard for BRD definition in calves to assess clinical score accuracy. Rather than basing our approach on an hypothetical gold standard we used a reference standard test with imperfect accuracy (thoracic ultrasound). With our latent class modeling approach, we establish that the optimal cut-off to define a positive case varied from 9 to 16 in low to average BRD prevalence situations. However, in high prevalence situation ($\geq 40\%$) treating all calves with one abnormal clinical sign would be the most cost-effective strategy. Relaxing accuracy results from thoracic ultrasound with non-informative priors had a limited impact on our predicting regression coefficient scores, which reinforces the external validity of the present findings.

Our study results can be interpreted and used in two different ways. First, it can be intuitive to look for the predicted probability of suffering from active lower respiratory tract infection for a calf with a specific clinical profile using the posterior probabilities from the main logistic regression model based on the 64 different clinical signs permutations. These predicted probabilities can be directly used by either veterinarians or farmers for assessing the risk of being sick based on our decision tree. Secondly, the clinical prediction rule using updated weights for each clinical item can also be used with some flexibility. Depending on the context (low vs. high expected prevalence of BRD, producer efficiency to detect calves), the cut-offs used for defining an affected animal could be modified considering the relative cost of misclassified calves. This scoring rule could also serve as a screening test before applying other more specific tests such as thoracic auscultation or determination of serum acute phase protein concentrations in calves with abnormal clinical signs profiles. Used for this purpose, the cut-off could be adjusted to be more sensitive since the false positive cases would be correctly classified by the second test.

The absence of a gold standard test is a recurrent issue in BRD research (White and Renter, 2009; Abdallah et al., 2016; Timsit et al., 2016; White et al., 2016). Due to this limitation, composite reference standard definition is frequently used. Association of multiple positive test (ex: increased rectal temperature, specific clinical signs, abnormal auscultation or ultrasound findings) for assessing new test accuracy or when reporting impact of different interventions on BRD are good examples. However, this type of composite reference standard can lead to biased estimates due to spectrum bias risk (Schiller et al., 2016). One can easily admit that a calf with clinical signs of BRD but without abnormal ancillary test may still be truly affected by BRD. This is particularly true when the ancillary test lacks sensitivity. In this situation, composite reference standard would classify this calves with discordant result i) as either a negative case or ii) as a calf not enrolled in the study (depending on negative case definition). These 2 types of considerations would i) decrease the apparent specificity of the new test or ii) increase the new test accuracy excluding cases with discordant results (which obviously may be the typical clinical cases where the clinician want to use another test to determine the "true" patient's status).

The Californian scoring system is relatively easy to implement in a dairy farm since the clinical signs evaluated are dichotomised (normal vs. abnormal). The different weights we found for these specific clinical signs slightly differs from what was originally reported (Aly et al., 2014; Love et al., 2014). This is not surprising since external validation of prediction rules most often justified some changes when compared with initial findings (Steyerberg, 2008; Collins et al., 2015). When looking more specifically on the main differences between the relative weights found in the current study vs the initial report of Californian score (Love et al., 2014) we can note that the weight attributed to the dyspnea and cough was higher in our study (representing 28.6% and 22.8% of the total points, respectively) than initially reported (11.8% for both signs). In contrast, eye discharge's weight was lower in the current study (1.4%) than in Love et al. (2014) study where it represents 11.8% of the score. The rectal temperature weight was similar in both scores (10% vs 11.8% in Californian score). In the present study, rather than presenting new clinical signs or new categorization for the clinical signs (for example: temperature) that were assessed, we focused our attention on updating prediction rules. This strategy has been reported as a better approach to optimize prediction rule rather than creating a different rule, which can in term limit the application in the field (Toll et al., 2008). As an example, in 2008 more than 60 different prediction rules (with different covariates) existed to estimate the risk breast cancer in human medicine (Toll et al., 2008). This would obviously become a limitation for the physician who cannot intuitively know which one is the best. Therefore, it could be discouraging to use any prediction rule.

The present score's accuracy varied, as expected, depending on the

Table 2

Median and 95% Bayesian credible intervals (BCI) for the Se and Sp of thoracic ultrasonography, regression parameters and calves' probability of active infectious respiratory process of the lower respiratory tract.

		MODEL 1		MODEL 2		MODEL 3	
		Prior	Posterior median (95% BCI)	Prior	Posterior median (95% BCI)	Prior	Posterior median (95% BCI)
	SeTUS	beta (27.02, 7.92)	82.7% (71.0–91.5)	beta (27.02, 7.92)	83.1% (71.4–91.8)	beta (1, 1)	93.0% (73.7–99.7)
	SpTUS	beta (80.58, 6.08)	94.9% (91.5–97.6)	beta (80.58, 6.08)	94.9% (91.3–97.6)	beta (1,1)	96.4% (91.8–99.7)
Regression parameter	Intercept	-	-2.84 (−3.82 to −2.15)	-	-2.77 (−4.07 to −1.93)	-	-2.66 (−3.62 to −2.02)
	b(dyspnea)[a]	-	1.99 (1.11–2.93)	-	2.21 (1.13–3.49)	-	1.98 (1.18–2.89)
	b(temperature)	-	0.70 (−0.08 to 0.78)	-	0.24 (−0.63 to 1.20)	-	0.50 (−0.18 to 1.32)
	b(eye)	-	−0.08 (−0.93 to 0.78)	-	−0.41 (−1.43 to −0.49)	-	−0.18 (−0.92 to 0.57)
	b(nasal)	-	0.99 (0.26–1.76)	-	1.04 (0.27–1.1.93)	-	0.88 (0.28–1.59)
	b(ear)	-	1.55 (0.39–2.87)	-	1.88 (0.29–4.12)	-	1.44 (0.45–2.59)
	b(cough)	-	1.60 (0.85–2.51)	-	1.73 (0.93–2.77)	-	1.42 (0.76–2.22)
	pBRD1[b]	beta (1.84, 42.11)	5.5% (2.2–10.5)	beta(1,1)	5.9% (1.7–12.7)	beta (1.84, 42.11)	6.5% (2.6–11.8)
	pBRD2	beta (3.64, 2.13)	69.5% (46.2–86.8)	beta(1,1)	66.4% (36.0–88.9)	beta (3.64, 2.13)	66.6% (45.2–84.2)
	pBRD3	beta (18.56, 3.51)	90.7% (77.3–97.3)	beta(1,1)	95.4% (76.1–99.7)	beta (18.56, 3.51)	89.7% (76.9–96.6)
	pBRD4	beta (4.05, 3.50)	67.5% (46.8–83.5)	beta(1,1)	75.8% (49.4–92.2)	beta (4.05, 3.50)	67.3% (48.1–82.6)
	pBRD5	beta (3.26, 2.26)	45.8% (25.7–68.1)	beta(1,1)	41.3% (17.1–70.8)	beta (3.26, 2.26)	45.2% (26.4–65.7)
	pBRD6	beta (1.50, 1.45)	23.8% (11.1–43.8)	beta(1,1)	18.0% (6.7–37.3)	beta (1.50, 1.45)	21.6% (10.6–39.0)
	pBRD7	beta (3.50, 4.05)	9.8% (3.5–22.7)	beta(1,1)	4.9% (1.0–14.7)	beta (3.50, 4.05)	8.7% (3.2–19.7)
	DIC		521.7		519.0		517.8
	pD		28.0		29.1		28.9

Model 1, 2 and 3 used the herd as a random effect ε ($\varepsilon \sim$ dnorm (0;τ) and τ ~ dgamma (1,1)).

SeTUS: sensitivity of thoracic ultrasound examination using consolidation depth ≥1 cm as a positive test; SpTUS: specificity of thoracic ultrasound examination.

DIC: Deviance information criteria; pD: effective number of parameter.

Model 1 (M1) was obtained using informative prior information from thoracic ultrasonography accuracy and clinical profiles BRD probability (see Table 2).

Model 2 (M2) is obtained using informative prior information from thoracic ultrasonography accuracy and non-informative priors for clinical profiles BRD probability.

Model 3 (M3) is obtained using non-informative prior for thoracic ultrasonography accuracy and clinical profiles BRD probability.

[a] b(dyspnea): regression parameter obtained from multivariable mixed logistic regression model for calves with abnormal breathing.

[b] pBRD1: probability of active infectious respiratory process of the lower respiratory tract (latent variable) for a calf with the clinical profile 1 (see Table 1 for the definition of the 7 clinical profiles used for expert elicitations).

Median and 95% credible intervals of predicted probabilities and ultrasound accuracy for bovine respiratory disease complex diagnosis in dairy calves
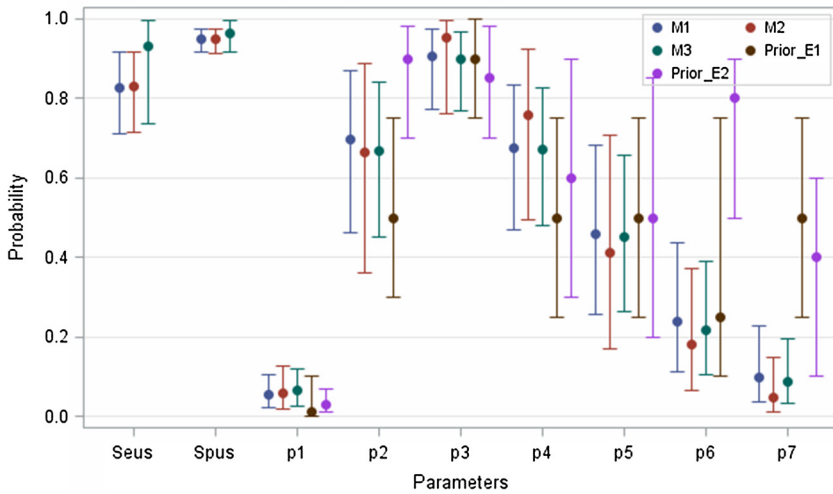

Fig. 3. Median and 95% credible intervals of predicted probabilities and ultrasound accuracy for the diagnosis of active infection of the lower respiratory tract in dairy calves compared with experts' conditional means probabilities.

M1: Model 1 is obtained using informative prior information from thoracic ultrasonography accuracy and clinical profiles BRD probability (see Table 2).

M2: Model 2 is obtained using informative prior information from thoracic ultrasonography accuracy and non-informative priors for clinical profiles BRD probability (see Table 2).

M3: Model 3 is obtained using non-informative prior for thoracic ultrasonography accuracy and clinical profiles BRD probability (see Table 2).

p1, …p7 are corresponding to 7 different calves' clinical profiles (see Table 1).

Prior_E1, Prior_E2 are the conditional means priors obtained from expert 1 and 2 consultation (see Table 1).
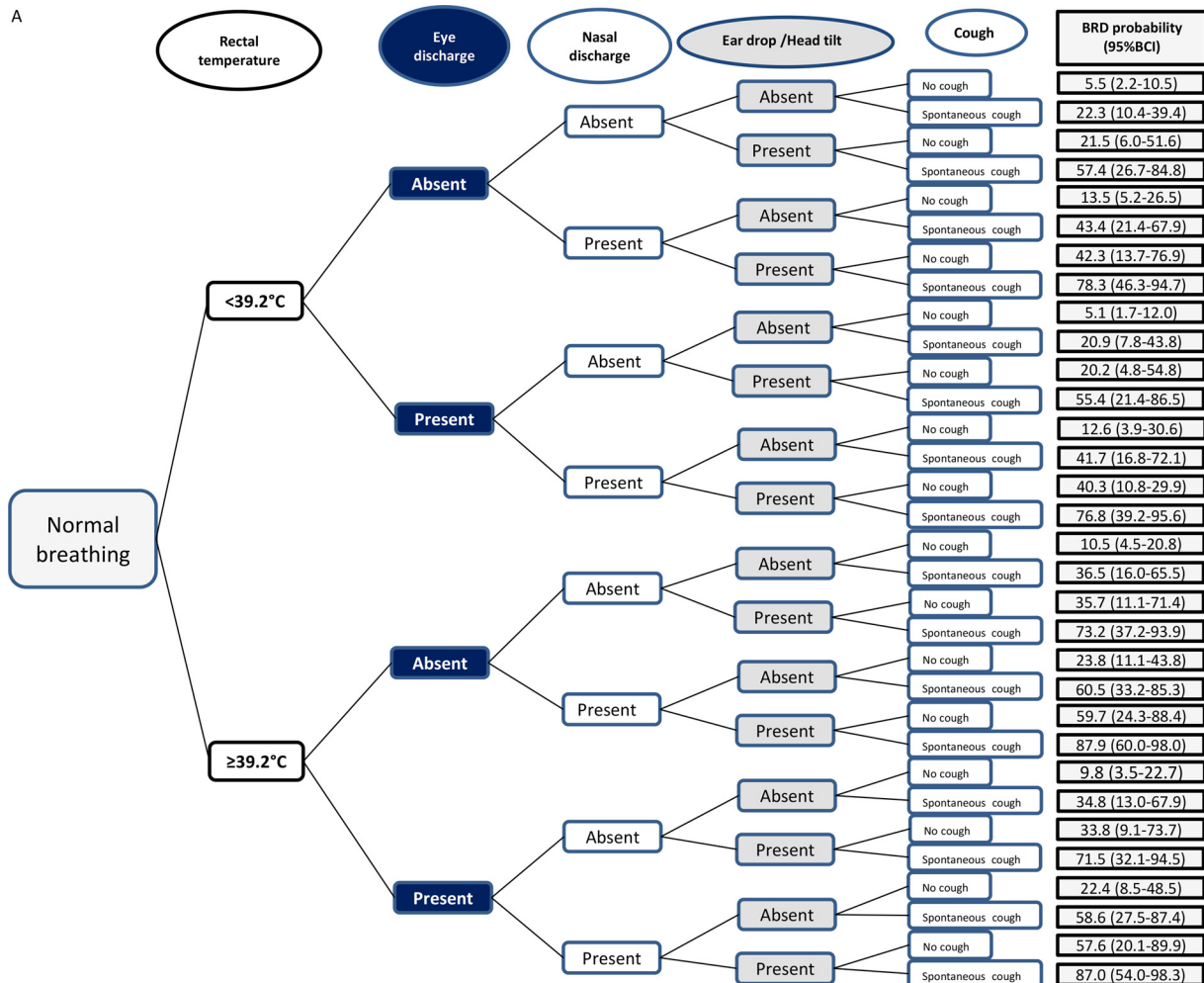

Fig. 4. Practical application of clinical assessment of dairy calves using 6 different clinical signs. The decision tree is based on 64 possible different clinical signs profiles to predict the probability (95% credible intervals) of an active infection of the lower respiratory tract based on posterior densities from model 1 (Table 2). The calves with normal vs rapid or abnormal breathing pattern are presented in Fig. 4A and 4B respectively. The clinical signs are dichotomous as reported by Love et al. (2014).
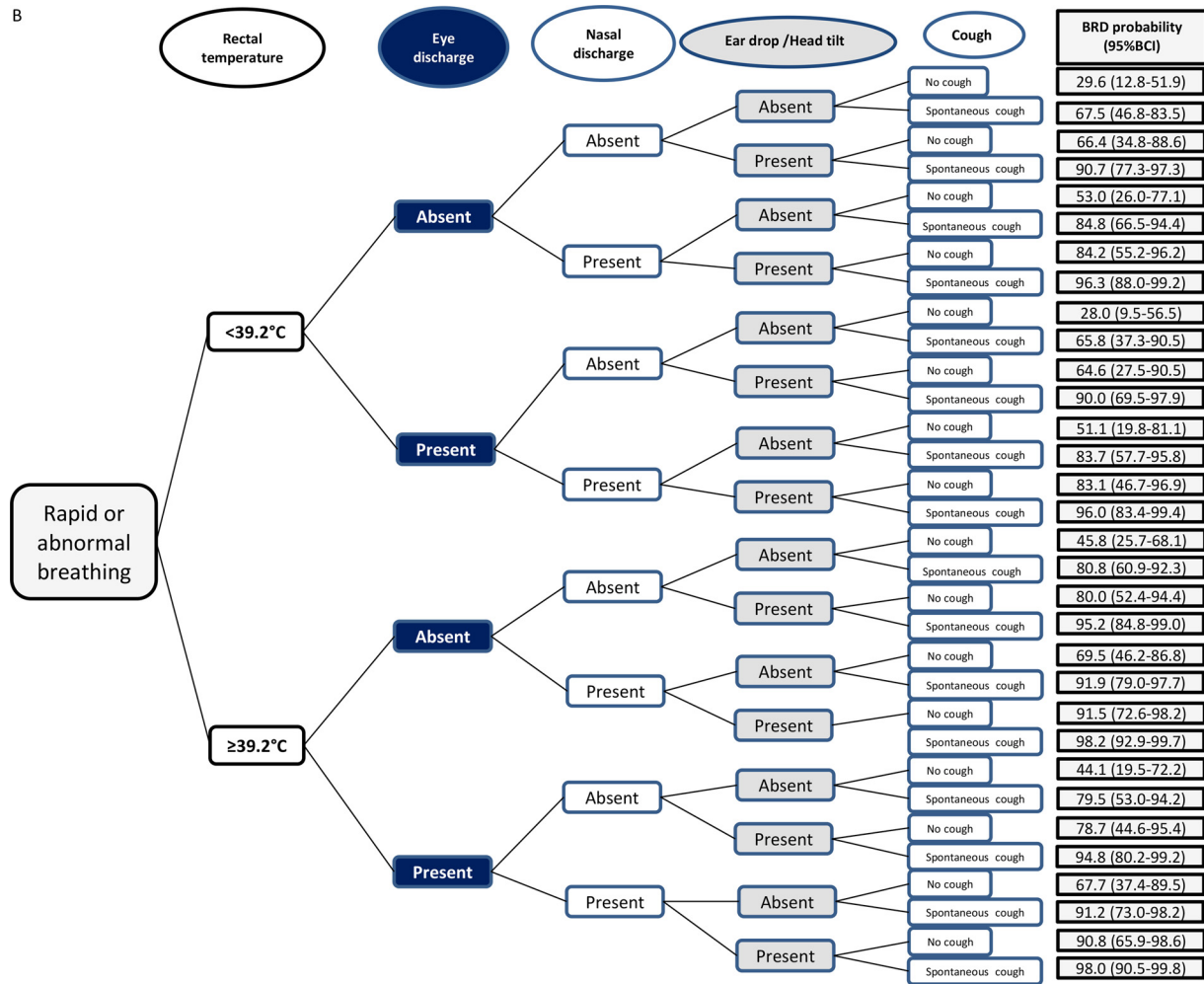
Fig. 4. (*continued*)

selected cut-off. The MCT analyses revealed that cut-offs from 9 to 16 were optimal decision thresholds in the different scenarios with Se varying from 83.0 to 66.9% and Sp from 69.1 to 82.7% respectively.

These ranges can be compared with reported for the Californian score with screening sensitivity of 46.8% (95% confidence interval; 39.5–54.3%) and specificity of 87.6% (82.6–91.1%) using a cutoff score

Table 3
Proposed clinical scoring system for detection of active infection of the lower respiratory tract in dairy calves and comparison with original Californian scoring chart.

	MODEL 1		Californian scoring system
	Modified scoring rule (relative weight%)	95% credibility intervals for modified scoring rule	Scoring rule from the Californian score (Love et al., 2014; Aly et al., 2014) (relative weight (%))
Presence of dyspnea	20 (28.6)	11–29	2 (11.8)
Temperature ≥39.2 °C	7 (10.0)	−1 to 8	2 (11.8)
Eye discharge	−1 (1.4)	−9 to 8	2 (11.8)
Nasal discharge	10 (14.3)	3–18	4 (23.5)
Ear drop/Head tilt	16 (22.8)	4–29	5 (29.3)
Spontaneous cough	16 (22.8)	9–25	2 (11.8)
Maximal score (points)	70		17

[a]The number of points for every abnormal clinical sign was obtained using the rounded 10*b (b:logistic regression coefficients) value according to Moons et al. (2002).
[b]The relative weight of each clinical sign was obtained dividing the number of point for this specific clinical sign by the total number of points (17 possible points in the Californian score and up to 70 points for the updated prediction rule).

Table 4

Sensitivity and specificity of different scores cut-offs for the diagnosis of active infection of the lower respiratory tract in dairy calves using updated predicting rule.

Cutoff to define a positive score	Sensitivity (95% BCI)	Specificity (95% BCI)
≥0	97.3 (92.0–98.2)	11.7 (8.8–15.1)
≥6	84.3 (72.2–98.2)	57.7 (52.5–63.8)
≥7	82.5 (70.1–97.8)	60.3 (55.1–66.6)
≥9	83.0 (69.5–98.2)	69.1 (63.7–75.9)
≥10	81.6 (68.2–97.8)	71.0 (65.7–77.5)
≥15	71.1 (56.6–92.3)	81.2 (76.4–87.5)
≥16	66.9 (52.3–92.3)	82.7 (78.0–88.4)
≥17	50.8 (37.4–80.3)	90.3 (86.5–95.0)
≥20	49.3 (35.6–81.8)	93.5 (89.9–98.1)
≥22	43.1 (30.8–73.4)	93.3 (90.0–97.4)
≥23	42.2 (30.0–76.4)	94.2 (90.9–98.0)
≥25	36.7 (25.3–64.5)	95.8 (92.8–99.0)
≥26	35.3 (24.0–64.6)	97.0 (94.3–99.6)
≥27	26.5 (16.8–50.5)	98.5 (96.5–99.9)
≥30	26.7 (16.7–51.9)	99.0 (97.2–99.9)
≥32	26.5 (16.9–56.5)	99.1 (97.5–100)
≥33	24.3 (15.2–46.0)	99.2 (97.6–100)
≥35	21.0 (12.7–42.1)	99.5 (98.3–100)
≥36	18.4 (10.7–40.5)	99.5 (98.2–100)
≥37	13.3 (7.1–27.7)	99.4 (98.1–100)
≥42	11.4 (5.7–24.3)	99.3 (98.0–100)
≥43	10.3 (5.0–24.7)	99.5 (98.2–100)
≥45	6.2 (2.5–16.0)	99.7 (98.7–100)
≥46	5.9 (2.3–14.8)	99.8 (98.9–100)
≥42	3.2 (0.9–9.5)	99.7 (98.9–100)
≥52	3.0 (0.8–8.94)	99.9 (99.2–100)

The computations of sensitivity and specificity were obtained from latent class model (1 population, 2 independent tests) using informative prior for ultrasound accuracy (Se ~ beta (27.02, 7.92); Sp ~ beta(80.58, 6.08)) and non informative prior for score accuracy and respiratory disease prevalence (beta (1,1)). The median estimates were used for generating the misclassification cost-term analyses.

of 5 or more. In this study, case definition was a composite reference standard with abnormal thoracic ultrasound or auscultation findings which may either include misclassified animals or calves with non-active pneumonia lesions. This risk of bias was avoided using a latent class approach.

One limitation of the present study is due to the imperfect test that was used as a proxy of BRD. Thoracic ultrasound is a fast and reliable test to detect BRD induced lung lesions (Ollivett et al., 2015). Moreover, it can be performed even by relatively novice ultrasound operator (Buczinski et al., 2013). Thoracic ultrasound findings correlates with necropsy findings (which can be considered as a gold standard) in chronic (Rabeling et al., 1998), subclinical BRD cases (Ollivett et al., 2015) and experimentally induced *Mannheimia haemolytica* infection (Ollivett et al., 2013). However, lung consolidation can be absent in early BRD cases or viral pneumonia (false negative case) and false positive cases may occur when lung consolidations are the consequences of a previous BRD episode but now inactive (ie. animal that previously had BRD but who will now not benefit from treatment). Other lung lesions associated with ultrasonographic lung consolidation (atelectasia or pulmonary tumor) may also occur (Lichtenstein et al., 2009). The sensitivity analysis showed that the informative prior on ultrasound accuracy had little impact on posterior distributions (model 2).

We used conditional mean priors based on experts' opinion of expected BRD probability associated with different clinical profiles. These priors had limited impacts on posterior probabilities as shown by sensitivity analysis. Interestingly, there was some discrepancies between experts and for some clinical profiles (posterior densities very different from priors). These discrepancies may reflect several mechanisms inherent to dichotomous scoring system and clinical decision-making. Three of the four profiles where either the expert did not agree or where posterior densities changed from priors densities included increased rectal temperature (≥39.2 °C). Although very simple to understand, categorization of a dynamic continuous biologic process has some limitations. For example, a 39.2 °C calf with slight nasal discharge and a 40.5 °C calf with bilateral purulent nasal discharge would have received the same clinical score. However, most of clinicians would admit that the probability of BRD in these two calves are quite different. The clinical thinking and reasoning is impacted when estimating the risk (best guess) of BRD probability by the spectrum of clinical signs included in a scoring system. This process would indeed influence the predicted probability. This may also explain why some posterior densities where significantly different from prior densities because LCM models considers the information lost in the categorization process.

Our model assumes that the herds have a specific BRD probability (using herd random intercept) but it was assumed a constant slope for each predictor of the model for a specific herd. However, since we did not have any bacterial or viral samples from calves in specific farms we could not investigate the impact of the specific etiological agent on the score accuracy (impact on the model slopes). In Québec dairy farms with preweaned calves' enzootic pneumonia, bacteria were most frequently isolated from deep nasopharyngeal swabs. Viruses were uncommonly present except for bovine coronavirus at least in one recent study (Francoz et al., 2015). The exact role of coronavirus in BRD complex remains uncertain. We tried to avoid bias in the present study using a random sampling of farms considered representative of the Québec dairy industry. From 6 (where only 6 calves were present) to 12 calves (where 12 or more preweaned calves were present) were selected as a requirement of our initial study (Buczinski et al., 2018). No specific internal validation of the proposed clinical scoring method was performed which can also be a study limitation.

In conclusion, this is the first study upgrading currently predicted rules for BRD diagnosis in calves considering the absence of gold standard for this disease. Using a Bayesian framework and including expert conditional means priors, it was possible to optimize an existing score which can be further used in practice. The different predicted probability for calves with various clinical signs have been reported and can be helpful for decision-making when using this score in practice. Dairy practitioners could either use the updated clinical score selecting the optimal cut-off depending on the clinical context of application (high vs. low expected prevalence of respiratory problems) or directly use the results of the 64 clinical profiles probability results. As with any diagnostic or prediction study, the findings of this study need to be confirmed, updated, and refined by future studies, particularly using different populations of calves. The applications of these clinical signs at the group level would also be helpful to decide when a group of calves is at high (vs. low) risk of being BRD affected.

Conflict of interest

None of the authors had any conflict of interest related to this study.
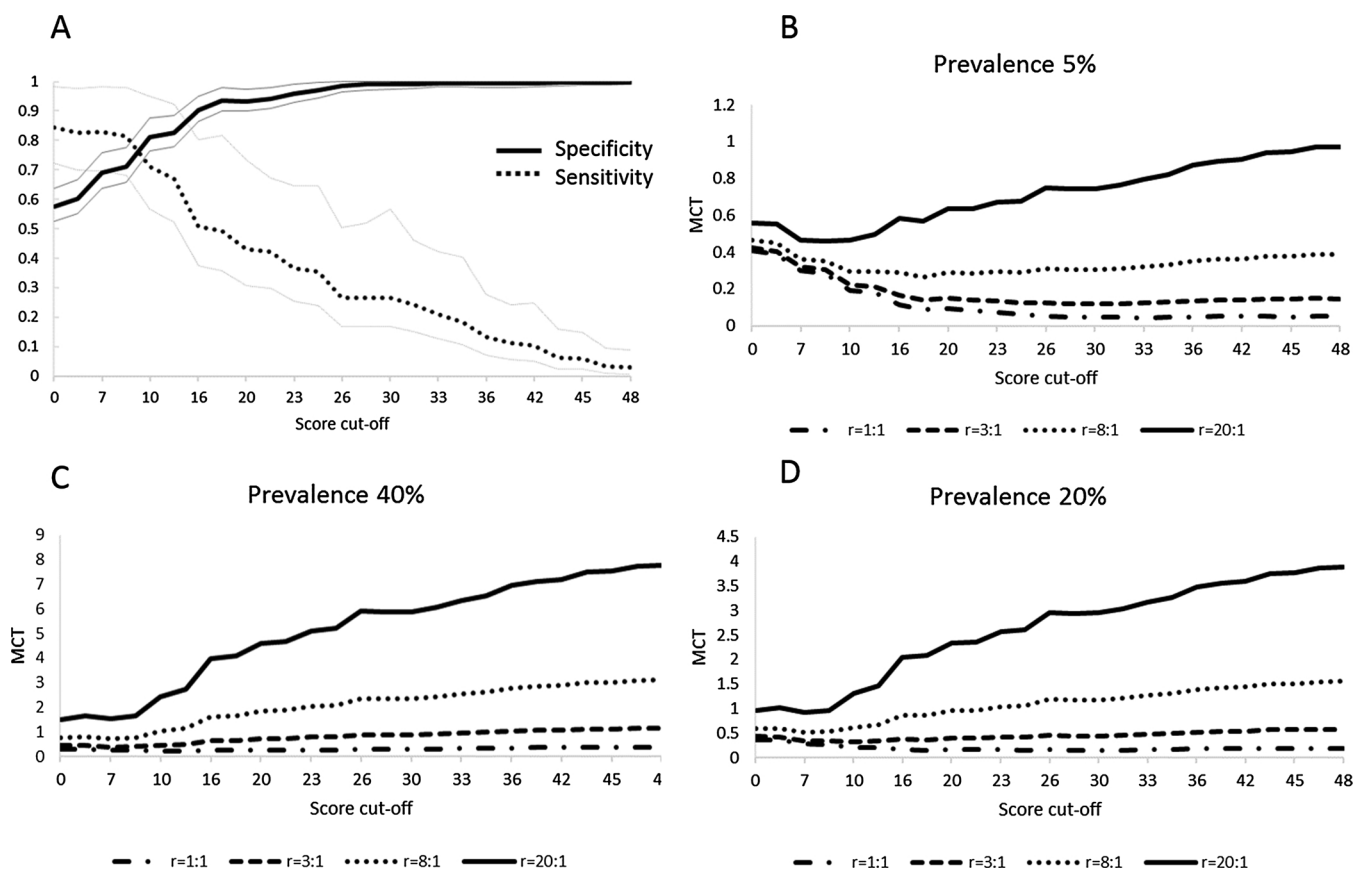
Fig. 5. Sensitivity and specificity of clinical score for the diagnosis of active infection of the lower respiratory tract in dairy calves and misclassification cost-term of the cut-offs used depending on expected respiratory disease prevalence and relative value of false negative to false positive cases.
The 95% credible intervals are represented for sensitivity and specificity.
MCT: misclassification cost term, Se: sensitivity, Sp: specificity, r: ratio of the cost of false negative case/false positive case.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.prevetmed.2018.05.004.

References

Abdallah, A., Hewson, J., Francoz, D., Selim, H., Buczinski, S., 2016. Systematic review of the diagnostic accuracy of Haptoglobin, Serum Amyloid A, and Fibrinogen versus clinical reference standards for the diagnosis of bovine respiratory disease. J. Vet. Intern. Med. 30, 1356–1368.

Adrion, C., Mansmann, U., 2012. Bayesian model selection techniques as decision support for shaping a statistical analysis plan of a clinical trial: an example from a vertigo phase III study with longitudinal count data as primary endpoint. BMC Med. Res. Method 12, 137.

Aly, S.S., Love, W.J., Williams, D.R., Lehenbauer, T.W., Van Eenennaam, A., Drake, C., Kass, P.H., Farver, T.B., 2014. Agreement between bovine respiratory disease scoring systems for pre-weaned dairy calves. Anim. Health Res. Rev. 15, 148–150.

Ames, T.R., 1997. Dairy calf pneumonia. The disease and its impact. Vet. Clin. North Am. Food Anim. Pract. 13, 379–391.

Assie, S., Seegers, H., Beaudeau, F., 2004. Incidence of respiratory disorders during housing in non-weaned Charolais calves in cow-calf farms of Pays de la Loire

(Western France). Prev. Vet. Med. 63, 271–282.

Branscum, A.J., Gardner, I.A., Johnson, W.O., 2005. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev. Vet. Med. 68, 145–163.

Buczinski, S., Borris, M.E., Dubuc, J., 2017. Herd-level prevalence of the ultrasonographic lung lesions associated with bovine respiratory disease and related environmental risk factors. J. Dairy Sci. 101, 2423–2432.

Buczinski, S., Forté, G., Bélanger, A.M., 2013. Short communication: ultrasonographic assessment of the thorax as a fast technique to assess pulmonary lesions in dairy calves with bovine respiratory disease. J. Dairy Sci. 96, 4523–4528.

Buczinski, S., Ménard, J., Timsit, E., 2016. Incremental value (Bayesian Framework) of thoracic ultrasonography over thoracic auscultation for diagnosis of bronchopneumonia in preweaned dairy calves. J. Vet. Intern. Med. 30, 1396–1401.

Buczinski, S., Rademacher, R.D., Tripp, H.M., Edmonds, M., Johnson, E.G., Dufour, S., 2015a. Assessment of L-lactatemia as a predictor of respiratory disease recognition and severity in feedlot steers. Prev. Vet. Med. 118, 306–318.

Buczinski, S., Ollivett, T.L., Dendukuri, N., 2015b. Bayesian estimation of the accuracy of the calf respiratory scoring chart and ultrasonography for the diagnosis of bovine respiratory disease in pre-weaned dairy calves. Prev. Vet. Med. 119, 227–231.

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Eur. J. Clin. Invest. 45, 204–214.

Dufour, S., Durocher, J., Dubuc, J., Dendukuri, N., Hassan, S., Buczinski, S., 2017. Bayesian estimation of sensitivity and specificity of a milk pregnancy-associated glycoprotein-based ELISA and of transrectal ultrasonographic exam for diagnosis of pregnancy at 28–45 days following breeding in dairy cows. Prev. Vet. Med. 140, 122–133.

Francoz, D., Buczinski, S., Bélanger, A.M., Forté, G., Labrecque, O., Tremblay, D., Wellemans, V., Dubuc, J., 2015. Respiratory pathogens in Quebec dairy calves and their relationship with clinical status, lung consolidation, and average daily gain. J. Vet. Intern. Med. 29, 381–387.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 1, 515–534.

Greiner, M., 1996. Two-graph receiver operating characteristic (TG-ROC): update version supports optimisation of cut-off values that minimise overall misclassification costs. J. Immunol. Methods 191, 93–94.

Lichtenstein, D., Meziere, G., Seitz, J., 2009. The dynamic air bronchogram. A lung ultrasound sign of alveolar consolidation ruling out atelectasis. Chest 135, 1421–1425.

Love, W.J., Lehenbauer, T.W., Kass, P.H., Van Eenennaam, A.L., Aly, S.S., 2014.

Development of a novel clinical scoring system for on-farm diagnosis of bovine respiratory disease in pre-weaned dairy calves. PeerJ 2, e238.

Love, W.J., Lehenbauer, T.W., Van Eenennaam, A.L., Drake, C.M., Kass, P.H., Farver, T.B., Aly, S.S., 2016. Sensitivity and specificity of on-farm scoring systems and nasal culture to detect bovine respiratory disease complex in preweaned dairy calves. J. Vet. Diagn. Invest. 28, 119–128.

Magder, L.S., Hughes, J.P., 1997. Logistic regression when the outcome is measured with uncertainty. Am. J. Epidemiol. 146, 195–203.

McGuirk, S.M., 2008. Disease management of dairy calves and heifers. Vet. Clin. North Am. Food Anim. Pract. 24, 139–153.

McInturff, P., Johnson, W.O., Cowling, D., Gardner, I.A., 2004. Modelling risk when binary outcomes are subject to error. Stat. Med. 23, 1095–1109.

Moons, K.G., Harrell, F.E., Steyerberg, E.W., 2002. Should scoring rules be based on odds ratios or regression coefficients? J. Clin. Epidemiol. 55, 1054–1055.

Ollivett, T., Hewson, J., Schubotz, R., Caswell, J., 2013. Ultrasonographic progression of lung consolidation after experimental infection with Mannheimia haemolytica in Holstein calves. J. Vet. Intern. Med. 673.

Ollivett, T.L., Buczinski, S., 2016. On-farm use of ultrasonography for bovine respiratory disease. Vet. Clin. North Am. Food Anim. Pract. 32, 19–35.

Ollivett, T.L., Caswell, J.L., Nydam, D.V., Duffield, T., Leslie, K.E., Hewson, J., Kelton, D., 2015. Thoracic ultrasonography and bronchoalveolar lavage fluid analysis in Holstein calves with subclinical lung lesions. J. Vet. Intern. Med. 29, 1728–1734.

Pardon, B., De Bleecker, K., Hostens, M., Callens, J., Dewulf, J., Deprez, P., 2012. Longitudinal study on morbidity and mortality in white veal calves in Belgium. BMC Vet. Res. 8, 26.

Portillo, T., 2014. Pen riding and evaluation of cattle in pens to identify compromised individuals. Proceedings of the 47th Annual Conference of the American Association of Bovine Practitioners. pp. 5–8.

Rabeling, B., Rehage, J., Döpfer, D., Scholz, H., 1998. Ultrasonographic findings in calves with respiratory disease. Vet. Rec. 143, 468–471.

Robert, C., Christensen, R., Johnson, W., Branscum, A., Hanson, T., 2012. Bayesian Ideas and Data Analysis. Taylor & Francis Group.

Schiller, I., van Smeden, M., Hadgu, A., Libman, M., Reitsma, J.B., Dendukuri, N., 2016. Bias due to composite reference standards in diagnostic accuracy studies. Stat. Med. 35, 1454–1470.

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. Ser. B 64, 583–639.

Steyerberg, E., 2008. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer Science & Business Media.

Timsit, E., Dendukuri, N., Schiller, I., Buczinski, S., 2016. 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. Prev. Vet. Med. 135, 67–73.

Toll, D.B., Janssen, K.J., Vergouwe, Y., Moons, K.G., 2008. Validation, updating and impact of clinical prediction rules: a review. J. Clin. Epidemiol. 61, 1085–1094.

van Smeden, M., Naaktgeboren, C.A., Reitsma, J.B., Moons, K.G., de Groot, J.A., 2014. Latent class models in diagnostic studies when there is no reference standard—a systematic review. Am. J. Epidemiol. 179, 423–431.

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.

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. Diag. Invest. 21, 446–453.

Woolums, A.R., Berghaus, R.D., Smith, D.R., White, B.J., Engelken, T.J., Irsik, M.B., Matlick, D.K., Jones, A.L., Ellis, R.W., Smith, I.J., Mason, G.L., Waggoner, E.R., 2013. Producer survey of herd-level risk factors for nursing beef calf respiratory disease. J. Am. Vet. Med. Assoc. 243, 538–547.