Development and validation of a clinical respiratory disease scoring system for guiding treatment decisions in veal calves using a Bayesian framework

J. Berman, * D. Francoz, A. Abdallah, S. Dufour, and S. Buczinski

Département des sciences cliniques, Faculté de médecine vétérinaire, Université de Montréal, Saint-Hyacinthe, QC J2S 2M2, Canada
Department of Animal Medicine, Faculty of Veterinary Medicine, Zagazig University, Zagazig City 44511, Egypt
Département de Pathologie et Microbiologie, Faculté de médecine vétérinaire, Université de Montréal, St-Hyacinthe, QC J2S 2M2, Canada

ABSTRACT

Active infectious bovine respiratory disease (BRD) is an infection of the airways that needs to be diagnosed correctly so that appropriate treatment can be initiated. The simplest and most practical test to detect active BRD in dairy calves raised for veal is the detection and interpretation of clinical signs by producers or technicians. However, the clinical scoring system currently available for veal calves lacks sensitivity and specificity, contributing to economic losses and high use of antimicrobials. An accurate and reliable batch-level test to detect active BRD is essential to tailor antimicrobial use and reduce economic losses in veal calves. The objective of this study was therefore to develop and validate a new veal calf respiratory clinical scoring system (VcCRS), including reliable clinical signs (cough, ear droop or head tilt) and increased rectal temperature to detect active BRD in batches of veal calves housed individually, and to describe the accuracy of the scoring system for identifying batches of veal calves to treat. During 2017 to 2018, clinical examination, thoracic ultrasonography (TUS) and a haptoglobin concentration (Hap) were prospectively performed on 800 veal calves housed individually in Québec, Canada. Deep nasopharyngeal swabs were performed on 250 veal calves. A Bayesian latent class model accounting for imperfect accuracy of TUS and Hap was used to obtain weights for the clinical signs and develop the VcCRS. The VcCRS was then validated externally in 3 separate data sets. Finally, the applicability of the VcCRS at batch level was determined. We found that calves with 2 of the following findings—cough, unilateral or bilateral ear droop or head tilt, or increased rectal temperature ≥39.7°C—were considered positive and had a 31% chance of having active BRD. Without at least 2 of these 2 findings, a calf had a 100% chance of not having active BRD. At the batch level, we found that a batch with ≥3 positive calves among 10 calves sampled 2 wk after arrival at the fattening unit had a 94% chance of having an active BRD prevalence ≥10%. A batch with <3 positive calves had a 95% chance of not having an active BRD prevalence ≥10%. In this study, we developed a simple individual and batch-level score that is reliable across examiners and performs effectively in the detection of active BRD in veal calves. The implementation of this VcCRS in the veal calf industry would promote the elaboration of a protocol tailoring antimicrobial use.

Key words: cattle, bronchopneumonia, diagnostic, latent class model

INTRODUCTION

Infectious bovine respiratory disease (BRD) is a disease of the respiratory tract caused by viruses and bacteria (Woolums, 2015). Most of the time, the infection starts in the upper respiratory tract before descending to the lower respiratory tract and causing inflammation and lesions to the lung parenchyma (defined in this article as active BRD; Panchier and Confer, 2010; Zeineldin et al., 2019). Active BRD needs to be treated with anti-inflammatories, antimicrobials, or both treatment approaches (Woolums, 2015; Buczinski and Pardon, 2020). Once the inflammation and infection are resolved, it is common to observe lung scar tissue (defined here as inactive BRD; Ollivett et al., 2015). Because dairy calves intended for veal production are commonly commingled during collection, transport, and housing in a fattening unit, active BRD is highly prevalent in this production system (50% of calves with lung lesions at slaughter; Leruste et al., 2012) and could be responsible for an important part of mortality (16–50% of deaths are caused by BRD; Lava et al., 2016a; Winder et al., 2016), lower carcass weight, and lower carcass quality (van der Mei and van den Ingh, 1987; Pardon et al., 2013). To prevent and control active BRD in veal calves, antimicrobials are commonly used, representing up to 73% of antimicrobials used during the production cycle (Lava et al., 2016b). These antimicrobials can be given as individual
treatment, but the vast majority (>95%) are given as metaphylactic therapy—that is, the simultaneous antimicrobial therapy of clinically healthy calves and of calves that have clinical signs of active BRD in a shared compartment (Pardon et al., 2012a). Reducing individual treatment and, above all, optimal targeting of metaphylactic group treatments are key factors to substantially reduce antimicrobial use in the veal industry. Accurate and reliable individual and group-level diagnosis of active BRD is therefore essential to tailor antimicrobial use in veal production.

In Canada, the province of Québec is a major player in veal calf production, producing around 80% of Canadian veal calves (Producteurs de Bovins du Québec, 2020). Considering the relatively high number of calves in a fattening unit in Québec (mean = 470 calves; Producteurs de Bovins du Québec, 2020), the simplest and most practical test to detect active BRD in a veal fattening unit is the detection and interpretation of clinical signs by non-veterinarians (e.g., producers or technicians). However, due to the variety of infectious agents involved in BRD, clinical signs vary in intensity and duration, which can make clinical diagnosis difficult (McGuirk and Peek, 2014). Additionally, clinical diagnosis in veal calves has been reported to be variable among different examiners (Buczinski et al., 2016; Berman et al., 2021). A clinical scoring system assigns values to each predictor, which are used to determine a total score, thus making it possible to assess disease more objectively than with unstructured clinical evaluation alone (Hayes et al., 2010; Love et al., 2014). A simple, objective, and reliable clinical respiratory scoring chart (CRSC) would therefore be a useful tool to improve and standardize active BRD identification in veal calves.

Currently, 3 CRSC to detect BRD in pre-weaning dairy-breed calves have been published. The earliest used a grading system of 0 to 3 for the following findings and clinical signs: elevated rectal temperature, nasal discharge, cough, ocular discharge, and ear droop or head tilt (WiCRSC; McGuirk, 2008). Although this clinical scoring system used weights and decision rules, it did not use quantitative methods to assign these weights. Moreover, WiCRSC was reported to be unreliable between examiners with minimal training (Buczinski et al., 2016) and had only moderate diagnostic accuracy [screening sensitivity (Se) and specificity (Sp) of 46–62% and 74–91%, respectively; Buczinski et al., 2015b; Love et al., 2016]. Another CRSC was developed to circumvent these issues in pre-weaned dairy calves (CaCRSC; Love et al., 2014). This was later adapted and evaluated, accounting for the imperfect reference standard definition in pre-weaning dairy calves in Québec (QeCaCRSC; Buczinski et al., 2018b). The CaCRSC and QeCaCRSC used the same clinical signs as WiCRSC but dichotomized them (presence vs. absence). They also included an additional sign: breathing quality. Moreover, the weights assigned were determined quantitatively. However, the performances of CaCRSC and QeCaCRSC were reported to be similar to that of WiCRSC, with Se and Sp of 47% and 87%, respectively, for CaCRSC (Love et al., 2016), and Se varying from 67% to 83% and Sp of 69% to 83% for QeCaCRSC (Buczinski et al., 2018b). Different reasons could explain these moderate diagnostic performances. First, the accuracy measures of WiCRSC, CaCRSC, and QeCaCRSC were validated using different imperfect reference tests for comparison, which could increase interpretation variability as a function of both the lesions and the agents involved. Second, those scores included clinical signs that do not have high inter-rater reliability (Buczinski et al., 2016; Berman et al., 2021), which could increase variability across examiners beyond clinical variability and, thus, across studies.

We recently reported the reliability of the respiratory clinical signs (i.e., nasal discharge, cough, ocular discharge, ear droop or head tilt, and respiration) commonly used to detect respiratory disease in veal calves by different types of persons involved in veal calves’ health monitoring (producers, technicians, and veterinarians; Berman et al., 2021). We showed that induced cough (presence or absence) and ear droop or head tilt (absence or presence of slight unilateral ear droop, bilateral ear droop, or head tilt) were the most reliable clinical signs compared with all other clinical signs that were assessed (Cohen’s kappa ≥ 0.6; Berman et al., 2021). In presence of active BRD, inflammation of the trachea causes spontaneous or easily induced cough. Increased coughing frequency when using an algorithm for continuous monitoring of calves’ sounds was 99.2% specific and 50.3% sensitive for disease detection in dairy calves (Vandermeulen et al., 2016). Ear droop or head tilt occurred in case of pain or depression, or if otitis was conjointly present. In the latter case, pneumonia is commonly associated with otitis media (Francoz et al., 2004; Gosselin et al., 2012). Elaborating a new score by adding those clinical signs to the objective measure of rectal temperature to investigate fever caused by the infectious process would reduce variability due to examiners and, therefore, would likely increase assessment reliability.

The objective of the current study was therefore to develop and validate a new clinical scoring system (VeCRS) and its corresponding chart, including reliable clinical signs to detect active BRD in veal calves housed individually, and to describe its accuracy for identifying batches of veal calves to treat metaphylactically.
MATERIALS AND METHODS

We proceeded in 3 steps: (1) developing a new diagnostic score (VcCRS), (2) validating the developed score at the individual level using external validation, and (3) modeling the application of the score at the batch level. The TRIPOD guidelines (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) were used to facilitate reporting on the design, conduct, and results of the current study (Supplemental File S1; https://dataverse.harvard.edu/dataverse/clinical-scoring-system; Berman, 2022a; Moons et al., 2015). The study protocol was approved by the Comité d’éthique de l’utilisation des animaux de l’Université de Montréal (17_rech-1898; Saint-Hyacinthe, QC, Canada).

Step 1: Developing a New Clinical Score (VcCRS): Attributing Weights to Each Clinical Sign Using a Quantitative Method and Choosing a Cut-Off to Detect Active BRD (800 Veal Calves)

Study Population. During a prospective cross-sectional study performed from October 1, 2017, to December 20, 2018, a total of 800 veal calves housed individually were randomly recruited from 80 batches (10 calves per batch) of 51 commercial veal calf fattening units in Québec (Délimax, Veaux Lourds Ltée, Saint-Hyacinthe, Québec, Canada, and Les Aliments Prolacto Inc., Villeroy, Québec, Canada). All fattening units performed all-in/all-out management. Because the number of calves present in every unit was known before the visit and each box on the farms was numbered, 10 boxes’ numbers were randomly chosen before the visit (RANDOM function in Excel; Microsoft Corp.). Whenever a box was empty due to early mortality, the next random number was selected.

The calves were mostly obtained from multiple local dairy farms after commingling through auction markets (Buczinski et al., 2021). Data were collected within 2 wk after arrival at the fattening unit. We intervened before the main peak incidence of active BRD in veal calves, generally occurring around 3 wk after arrival (Miller et al., 1980; Pardon et al., 2012b). At that time, calves were about 3 to 4 wk old and were housed in individual duckboard pens. Multiple batches of calves (i.e., groups of calves that arrived at the farm to be fattened together) were considered per farm, and the selected batches were distributed across all 4 seasons. Data collection was performed in the morning (between 0900 and 1100 h), between meals.

Data Collection. Sex (male or female), breed (Holstein, Jersey, Red Holstein, Ayrshire), and treatment received before sampling (antimicrobial, anti-inflammatory, or both) were recorded. On the same day, each calf underwent a physical examination, in which the main clinical signs hypothesized to be of value for the scoring system were measured. For that purpose, we selected only the most reliable clinical signs (cough, ear droop or head tilt), as determined previously in Berman et al. (2021), and increased rectal temperature as predictors. Moreover, for each calf, thoracic ultrasonography (TUS) and a haptoglobin concentration (Hap) measurement were conducted successively, as described in the next section. These latter 2 tests were conducted to estimate the true disease status (active or non-active BRD). Finally, a sample of 250 veal calves from the last 25 sampled batches underwent a deep nasopharyngeal swab to describe the bacteria present in the studied population.

Clinical Signs. A complete physical examination was performed on each calf by the same experienced operators (A.A., J.B., or S.B.). The presence of abnormalities upon the physical examination (e.g., diarrhea, navel infection, arthritis, mass, skin or eye disorders) was recorded. The presence of respiratory clinical signs (i.e., nasal discharge, ocular discharge, induced or spontaneous cough, abnormal respiration, and ear droop or head tilt) was recorded based on the 2-level reliable combination reported previously by Berman et al. (2021) and summarized in Supplemental File S2 (https://dataverse.harvard.edu/dataverse/clinical-scoring-system; Berman, 2022b). Increased rectal temperature was considered when rectal temperature was ≥39.7°C, as reported in feedlots (Timsit et al., 2011a,b). Because the physical examination was performed first, the operators were blinded to results of both reference test (TUS and Hap).

Deep Nasopharyngeal Swab. A deep nasopharyngeal swab was taken as previously reported from the right or left nostril (Godinho et al., 2007) and placed in transport medium (BBBL Port-A-Cul Tubes, Becton, Dickinson and Company) for routine bacterial and mycoplasma cultures. All laboratory analyses were performed in a diagnostic laboratory accredited by the American Association of Veterinary Laboratory Diagnosticians. Details are available in Supplemental File S2.

Reference Tests. In the absence of a gold standard to measure active BRD, the true status of active BRD was considered as a latent variable and was investigated via the following 2 reference tests.

Thoracic Ultrasonography. Bilateral TUS was performed immediately after the physical examination by the same experienced operators (A.A., J.B., or S.B.), using a 7.5-MHz linear probe (Imago, ECM),
was expressed by the following formula:

\[
Pr_{\text{BRD}} = Pr_{\text{BRD+}} + \text{prBRD} \times \text{Sp}_{\text{TUS}} \times \text{Tu}_{\text{S}}.
\]

Active BRD detected by TUS was considered to be present if the maximal depth of consolidation observed was ≥3 cm (TUS positive; Berman et al., 2019). The operators were blinded to Hap results but not to the physical examinations.

**Haptoglobin Concentration.** Blood samples from jugular veins were collected for each calf (tube without anticoagulants) immediately after both physical examination and TUS. Within 2 h after collection, samples were centrifuged at 1,500 \times g for 15 min at approximately 20°C, and the serum was stored in aliquots at −20°C until analysis by the University of Guelph Animal Health Laboratory (Guelph, ON, Canada). Before analysis, serum samples were thawed at room temperature. Serum Hap content was measured in duplicate by the hemoglobin binding capacity method (Skinner et al., 1991), using an automated analyzer (Cobas 6000 c 501, Roche Diagnostics). The lower limit of quantification was 0.3 μmol/L (0.03 g/L); the inter-assay coefficient of variation (CV) was 6.5%; and the intra-assay CV was 0.9%. In the absence of an established cut-off inveal calves, active BRD according to Hap concentration was considered to be present if the concentration of Hap was ≥2.5 μmol/L (0.25 g/L), based on the distribution of our data and as reported by Timsit et al. (2009) in feedlots (Hap positive; details in Supplemental File S2).

**VcCRS Development.** Calf was the unit of interest. Analyses were performed using SAS version 9.4 (SAS Institute Inc.) and OpenBUGS version 3.2.3, rev. 1012 (MRC Biostatistics Unit).

**Assignment of Weight to Each Clinical Sign.** Score weights were attributed individually (univariable) for each selected clinical sign (cough, ear droop or head tilt) or increased rectal temperature (3 univariate models), to avoid underestimating the influence of individual signs due to their interaction with other signs. Briefly, we quantified the association between 1 selected clinical sign or increased rectal temperature and the real probability of active BRD using Bayesian univariable mixed logistic regression models where the dependent variable was the real status of active BRD (a latent variable), following a previously reported framework (McInturff et al., 2004; Buczinski et al., 2015). The probability that a calf had active BRD \( Pr_{\text{BRD}} \) was estimated using the observed TUS result for that calf and the Se and Sp of TUS. Briefly, the probability of the ith calf for testing TUS-positive \( Pr_{\text{TUS}} \) was described as a function of the probability of the calf being truly affected by the target condition (i.e., active BRD, \( Pr_{\text{BRD}} \)) as

\[
Pr_{\text{TUS}_{ijk}} = Pr_{\text{BRD}_{ijk}} \times Se_{\text{TUS}} + (1 - Pr_{\text{BRD}_{ijk}}) \times (1 - Sp_{\text{TUS}}).
\]

For the ith calf, the latent class variable (active BRD; \( Y_{ijk} \)) was assumed to be a Bernoulli event defined by the probability of having active BRD \( Pr_{\text{BRD}} \):

\[
Y_{ijk} \sim \text{Bernoulli} \left( Pr_{\text{BRD}_{ijk}} \right).
\]

**Choice of Priors**

In this Bayesian framework, we had to describe the prior available knowledge (as prior distributions) on the intercept \( \beta_0 \), the clinical sign’s coefficient \( w \), the mean and variance of the batch \( \mu_{\text{Batch}_{jk}} \) and farm \( \mu_{\text{Farm}_{k}} \) random intercepts, and finally, TUS Se and Sp.

Priors on \( w \) were determined considering a diffuse range of odds ratios between 0.01 and 100 (back-transformed for use in a logit scale as a range of coefficients between −2 and 4.6), corresponding to the normal distribution \( w \sim \text{Normal} (1.4, 1.7) \) and intercept \( \varepsilon_j \sim \text{Normal} (1.4, 1.7) \):

\[
\mu_{\text{Farm}} \sim \text{Normal} (0; \tau) \text{ and } \tau \sim \text{Gamma} (0.1, 0.1);\]

\[
\mu_{\text{Batch}} \sim \text{Normal} (0; \tau) \text{ and } \tau \sim \text{Gamma} (0.1, 0.1).
\]
**a: DEVELOPMENT of the new score (VcCRSC)**

Selected Predictors
- Cough (yes vs. no)
- Ear droop/Head tilt (none vs. slight uni or bi droop, or head tilt)
- Rectal temperature (≤ 39.7°C vs. ≥ 39.7°C)

Latent class model - Regression

For $j$ clinical sign in the $i^{th}$ calf of the $k^{th}$ herd:

$$\text{Logit} \; \Pr_{BRD}^{+\text{ijk}} = \beta_0 + w_j \cdot \text{predictor}_{ij} + \mu_{\text{Farm}} + \mu_{\text{Batch}}$$

**Priors**

<table>
<thead>
<tr>
<th>$w$</th>
<th>Normal (1.4, 1.7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>Normal (1.4, 1.7)</td>
</tr>
<tr>
<td>$\mu_{\text{Farm}}$</td>
<td>Normal (0; $\tau$) $\sim$ Gamma (0.1,0.1)</td>
</tr>
<tr>
<td>$\mu_{\text{Batch}}$</td>
<td>Normal (0; $\tau$) $\sim$ Gamma (0.1,0.1)</td>
</tr>
</tbody>
</table>

$$\Pr_{TUS}^{+\text{ijk}} = \Pr_{BRD}^{+\text{ijk}} \times \text{Se}_{TUS} + (1 - \Pr_{BRD}^{+\text{ijk}}) \times (1 - \text{Sp}_{TUS})$$

**b: DETERMINATION OF AN OPTIMAL CUT-OFF and VALIDATION**

**Development**
- Development population 800 veal calves
- Population 1 209 veal calves
- Population 2 513 veal calves
- Population 1 + 2 722 veal calves

**External validation**

**Figure 1.** Diagram of the latent class models (LCM): LCM regression and LCM-TUS performances (a) and LCM validation (b) used for determining and validating the clinical scoring system chart (VcCRSC) to detect active bovine respiratory disease (BRD) in veal calves. Priors of each model are shown. $\Pr_{BRD}^{+\text{ijk}}$ = probability that calf $i$ has active bovine respiratory disease; $w_j$ = weight coefficient for predictor $j$ ($j$ = batch or herd; $k$ = farm); $\mu_{\text{Farm}}$ accounted for data structure with calves clustered within a specific farm; $\mu_{\text{Batch}}$ accounted for data structure with calves clustered within specific batches; $\text{Se}_{TUS}$ = sensitivity of thoracic ultrasonography; $\text{Sp}_{TUS}$ = specificity of thoracic ultrasonography; $\text{Se}_{\text{Hap}}$ = sensitivity of haptoglobin; $\text{Sp}_{\text{Hap}}$ = specificity of haptoglobin; $\text{Se}_{\text{VcCRSC}}$ = sensitivity of VcCRSC; $\text{Sp}_{\text{VcCRSC}}$ = specificity of VcCRSC; prevalence = prevalence of active BRD.
The choice of Gamma (0.1, 0.1) for precision specification was considered as reasonably non-informative (Gelman, 2006).

In LCM-TUS performances, we assumed that TUS and Hap were conditionally independent because Hap and TUS assess different biological processes (Berman et al., 2019). Another assumption of the model was that the TUS accuracy was the same for all calves included in the study (Buczinski et al., 2018b). Beta distributions ($\beta$) were used as informative priors for all parameters of interest to optimize precision (Se and Sp of TUS, Se and Sp of Hap, and active BRD prevalence; Dunson, 2001; Gustafson, 2005). The choice of priors is detailed in Supplemental File S2 and shown in Figure 1a.

**Sample Size Calculation.** The sample size of 800 calves was obtained a priori to optimize the accurate estimation of active BRD status by physical examination, TUS, and Hap (i.e., with a 95% confidence interval varying by ±2.5–12.5% of Se and Sp for each test) and considering an expected active BRD prevalence of 20%.

**From the Logistic Analysis to the VcCRS.** The coefficients obtained from the regression analysis were used to define the specific sign score weight ($W$). The scores’ weights were obtained after rounding logistic regression coefficient parameters multiplied by 10 (Moons et al., 2002; Toll et al., 2008). At the end of the 3 different univariable analyses for the 3 selected predictors, a specific score with the different scores weight was built as follows:

$$\text{Specific score} = W_{\text{Cough}} \times \text{Cough (presence or absence)} + W_{\text{Ear droop or head tilt}} \times \text{Ear droop or head tilt (slight uni- or bilateral ear droop or head tilt)} + W_{\text{Temperature}} \times \text{Temperature (≥39.7°C)}.$$ 

**Assessment of Model Sensitivity to Priors.** The probability that a calf had active BRD ($Pr_{\text{BRD}}$) was estimated with the posterior distribution of Se and Sp of Hap instead of the Se and Sp of TUS (Table 3; sensitivity analysis scenario).

**Software.** The model was based on a total of 20,000 iterations using a 5,000-iteration burn-in. Three different chains with different initial values were run for each model. Rapid mixing and stationary distribution were sought as signs of good convergence to the posterior distribution. The convergence of the models was checked using visual trace plots and Gelman-Rubin statistic plots. Autocorrelation was detected using autocorrelation plots, and thinning was performed when required. The posterior distributions of each parameter were reported as medians and the corresponding 95% Bayesian credible interval (BCI).

**Determination of Optimal Cut-Off.** To investigate the optimal cut-off for VeCRS to detect active BRD, the posterior distributions of Se and Sp of VeCRS were computed using another Bayesian LCM (LCM validation; Figure 1b; details in Supplemental File S2) to compare the 3 tests, VeCRS, TUS, and Hap, in the development population across a range of possible cut-off values (0, 6, 9, 10, 15, 16, 19, and 25). The priors were the same as used in LCM-TUS performances for Se and Sp of TUS, Se and Sp of Hap, and prevalence of active BRD. Non-informative priors were used for Se and Sp of VeCRS corresponding to a $\beta$ (1.0; 1.0) distribution. Priors are presented in Figure 1b.

A misclassification cost-term (MCT) analysis was conducted as described by Greiner (1996). Briefly, MCT analysis is a powerful tool to illustrate the robustness of the optimal cut-off, because it considers not only the Se and Sp of VeCRS but also the prevalence of active BRD (Prev) and the cost ratio between false negatives and false positives. The MCT can be plotted for different cost ratios of false negatives to false positives (r), making it possible to develop a cut-off that accounts for different cost ratios associated with test misclassification. The exact plausible ranges for the relative costs of r are presently unknown for veal calves. We used wide ranges that were obtained in a previous study in which 4 different experts were asked to determine this value in feedlot calves (Buczinski et al., 2015a). In the absence of specific studies on veal calves reporting this ratio, we assumed plausible ratios of $r = 1:1$, $r = 3:1$, and $r = 8:1$, indicating that the cost of a false negative case is generally higher than that of a false positive case (Buczinski et al., 2018b). The MCT was calculated for each specific cut-off using the following formula:

$$\text{MCT} = (1 - \text{Prev}) \times (1 - \text{Sp}_{\text{VeCRS}}) + r \times \text{Prev} \times (1 - \text{Se}_{\text{VeCRS}}).$$

The minimum value of MCT was considered as the value that minimizes the misclassification costs (Greiner, 1996; Buczinski et al., 2018b). To determine the variability of the optimal cut-off with the prevalence, the MCT analysis was performed for different batches’ active BRD prevalence scenarios: low (Prev = 5%), medium (Prev = 10%), and high prevalence (Prev = 25%).
Step 2: External Validation of the Developed Score and Estimation of Its Performance at the Individual Level Using Different Populations (209, 313, and 722 Veal Calves)

The VcCRS was validated using 2 separate data sets from previous studies (population 1, Berman et al., 2019; population 2, Morin et al., 2020) that measured the same predictors (cough, ear droop or head tilt, and increased rectal temperature) but were sampled previously by other investigators (temporal validation; Moons et al., 2015). The target condition (active BRD) was measured identically by a Bayesian latent class model in population 1 (TUS and Hap; Berman et al., 2019) but only with TUS in population 2 (Morin et al., 2020). Details of the data sets are shown in Table 1. The Se, Sp, positive predictive value (PPV), negative predictive value (NPV) (i.e., the probability of a calf with or without active BRD receiving a positive or negative test result, respectively), positive likelihood ratio \(LR_+\) (i.e., percentage of calves with active BRD having a positive test result (Se), divided by the percentage of calves without active BRD but deemed positive based on the test result (i.e., false positive fraction \(= 1 – Sp\))], and negative likelihood ratio \(LR_-\) (i.e., percentage of calves with active BRD classified as negative based on the test result (i.e., false negative fraction \(= 1 – Se\)), divided by the percentage of calves without active BRD with a negative test result (Sp)] were estimated by LCM analysis (LCM validation; Figure 1b, details in Supplemental File S2) to compare 2 or 3 imperfect tests [TUS, Hap (only population 1), and VcCRS] in 1 or 2 populations. Thus, 3 LCM were used: (1) a LCM where TUS and Hap were compared with the VcCRS in population 1; (2) a LCM where TUS was compared with VcCRS in population 2; and (3) a LCM where TUS was compared with VcCRS in 2 distinct populations (1 and 2). The model’s priors are shown in Figure 1b.

Step 3: Simulation: Comparing and Choosing Interpretation for Batch-Level Treatment Decision (800 Veal Calves from 80 Batches)

We use the term “batch” or “herd” as a general term for any cluster or aggregate ≥100 veal (Producteurs de Bovins du Québec, 2020). To improve the applicability of the VcCRS in veal production, we estimated its batch-level PPV [HPPV, i.e., the probability of a positive batch (i.e., with ≥active BRD prevalence threshold) receiving a positive batch-level test result] and batch-level NPV [HNPV, i.e., the probability of a negative batch (i.e., with <active BRD prevalence threshold) receiving a negative batch-level test result]. The active BRD prevalence threshold can vary according to several factors, including the calves (price paid in the auction market, quality, origin), the anticipated meat marketing context, specific farm risk factors (ventilation, caretakers), the season, and the agent involved. We therefore decided to report HPPV and HNPV for a range of relevant active BRD prevalence thresholds in veal calves (i.e., 5, 10, 15, 20, 25, and 30%) according to personal communication with the veal industry in Québec (F. Beaulac, Delimax, Saint-Hyacinthe, Québec, Canada, and R. Jean, Clinique Vétérinaire de Saint Valier, Saint-Valier, Québec, Canada) and a study on a

<table>
<thead>
<tr>
<th>Item</th>
<th>Development population</th>
<th>Population 1⁠²</th>
<th>Population 2⁠³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calves (no.)</td>
<td>800</td>
<td>209</td>
<td>513</td>
</tr>
<tr>
<td>Age</td>
<td>3–4 wk</td>
<td>3–4 wk</td>
<td>3–4 wk</td>
</tr>
<tr>
<td>Male, no. (%)</td>
<td>756 (94.5)</td>
<td>194 (92.8)</td>
<td>Not available</td>
</tr>
<tr>
<td>Holsteins</td>
<td>768</td>
<td>188</td>
<td>Not available</td>
</tr>
<tr>
<td>Batches (no.)</td>
<td>80</td>
<td>51</td>
<td>4</td>
</tr>
<tr>
<td>Farms</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Sampling dates</td>
<td>October 2017–December 2018</td>
<td>July and October 2015</td>
<td>February–May 2019</td>
</tr>
<tr>
<td>Predictors</td>
<td>Nasal discharge</td>
<td>Nasal discharge</td>
<td>Nasal discharge</td>
</tr>
<tr>
<td></td>
<td>Ocular discharge</td>
<td>Ocular discharge</td>
<td>Ocular discharge</td>
</tr>
<tr>
<td></td>
<td>Ear droop or head tilt</td>
<td>Ear droop or head tilt</td>
<td>Ear droop or head tilt</td>
</tr>
<tr>
<td></td>
<td>Abnormal respiration</td>
<td>Abnormal respiration</td>
<td>Cough</td>
</tr>
<tr>
<td>Cough</td>
<td>Thoracic ultrasonography</td>
<td>Thoracic ultrasonography</td>
<td>Thoracic ultrasonography</td>
</tr>
<tr>
<td>Imperfect reference tests</td>
<td>Haptoglobin dosage</td>
<td>Haptoglobin dosage</td>
<td>Haptoglobin dosage</td>
</tr>
</tbody>
</table>

¹Predictors (cough, increased temperature, and ear droop or head tilt) were measured identically as in the score development population but sampled previously by other investigators (temporal validation). The target condition (active bovine respiratory disease) was estimated by the same reference tests (thoracic ultrasonography and serum haptoglobin concentration) as in the developing population in population 1, but only by thoracic ultrasonography in population 2.

²Population from Berman et al., 2019.

³Population from Morin et al., 2020.
feedlot (Baptiste and Kyvsgaard, 2017). Both HPPV+ and HNPV− were estimated by the following formulas (Christensen and Gardner, 2000):

\[
\text{HPPV} = \frac{\text{HTP} \times \text{HSe}}{\text{HTP} \times \text{HSe} + (1 - \text{HTP}) \times (1 - \text{HSp})};
\]

\[
\text{HPNV} = \frac{(1 - \text{HTP}) \times \text{HSp}}{(1 - \text{HTP}) \times \text{HSp} + \text{HTP} \times (1 - \text{HSe})}.
\]

Batch-level predictive values depend on the true proportion of positive batches [\text{HTP}; i.e., the pre-test probability that a batch is positive (≥active BRD prevalence threshold)]. Because the true batch prevalence of active BRD is unknown in veal calves, HPPV and HPNV were estimated for a range of HTP: 5, 10, 15, and 20%. Batch-level predictive values also depend on the batch-level Se of VcCRS [\text{HSe}; i.e., the probability that a positive batch (i.e., with ≥active BRD prevalence threshold) yields a positive batch-level test result] and the batch-level Sp of VcCRS [\text{HSp}; i.e., the probability that a negative batch (i.e., with <active BRD prevalence threshold) yields a negative batch-level test result], estimated by the following formulas (Christensen and Gardner, 2000):

\[
\text{HSe} = \sum_{k=0}^{\text{n} - 1} \binom{\text{n}}{k} \text{AP}^{k-1} (1 - \text{AP})^{n-(k-1)};
\]

\[
\text{HSp} = \sum_{k=0}^{\text{n} - 1} \binom{\text{n}}{k} \text{Sp}^{n-(k-1)} (1 - \text{Sp})^{k-1},
\]

where AP corresponds to the apparent prevalence, defined as AP = prevalence threshold × Se + (1 − prevalence threshold) × (1 − Sp). The HSe and HSp of VcCRS are dependent on the Se and Sp of the individual test, the number of animals tested (n), the prevalence threshold in infected batches, and the batch cut-off value (k; e.g., 1, 2, or 3 positive test results) used to classify the batch as positive.

We first determined the minimal number of veal calves (n) to sample (10, 20, or 30) in a batch (≥100 calves) that optimized HPPV and HPNV, considering the most probable value of HTP at 0.05 based on the batch prevalence of lung consolidation on dairy farms in Québec (Buczinski et al., 2018a). We hypothesized an active BRD prevalence threshold of 15%, which represents the threshold where the number of veal calves needing treatment is minimal according to Baptiste and Kyvsgaard (2017).

After the minimal number of veal calves was fixed (10, 20, or 30), we determined the batch cut-off value (k) that optimized HPPV and HNPV for the different scenarios of prevalence thresholds and HTP.

Additional details on the materials and methods, including the statistical models used and the data sets, are available as Supplemental Files S2 and S3 (https://dataverse.harvard.edu/dataverse клиническая оценка систем; Berman, 2022b,c).

RESULTS

Step 1: Development of the VcCRS

Study Population. Descriptive data of the 800 veal calves recruited in this study are shown in Table 2. Most of the veal calves were male Holsteins. A total of 115 calves (14%) were treated with anti-inflammatories, antimicrobials, or both approaches before collecting data. Calves’ treatments before sampling had no influence on the results (assessed by comparing proportion of TUS- and Hap-positive and proportion of bacteria isolation in treated and untreated calves; data not shown). A total of 46 calves (5%) had another disease (diarrhea, navel infection, arthritis, ringworm, mandibular mass, eye disorders) detected at physical examination. Deep nasopharyngeal results showed that most calves (n = 135; 54%) were carrying Mycoplasma bovis. Pasteurella multocida and Mannheimia haemolytica were isolated in 54 (21.4%) and 3 (1.2%) samples.

Descriptive Statistics. Data collection was possible on each of the 800 recruited calves.

Clinical Signs (Predictors). The proportions of clinical signs in the 800 veal calves were as follows: nasal discharge (n = 77; 9.6%); cough (n = 147; 18.4%); ocular discharge (n = 142; 17.8%); ear droop or head tilt (n = 35; 4.4%); abnormal respiration (n = 32; 4.0%). A total of 39 calves had increased rectal temperature (n = 39; 4.9%). The frequencies of calves with 1 of the 3 predictors included in the VcCRS development (i.e., ear droop or head tilt, cough, and increased rectal temperature), TUS positive, and Hap positive are detailed in Figure 2.

Reference Standard Tests. Thoracic ultrasonography was performed successfully on each calf. A total of 110 calves (13.8%) had lung consolidations ≥3 cm and were deemed positive on TUS. Five samples were not analyzed to determine Hap concentration (insufficient quantity or missing identification). Consequently, 5 calves had missing Hap data. Among these 5 calves, 2 were TUS positive. A total of 133 calves (16.7%) had Hap ≥2.5 μmol/L (0.25 g/L) and were deemed positive on Hap. A total of 28 calves were TUS and Hap positive, 105 calves were TUS negative but Hap positive, 80 calves were TUS positive but Hap negative, and 582 calves were TUS and Hap negative.
The posterior distribution $Se$ and $Sp$ of TUS were estimated at 76% (95% BCI: 42, 96%) and 90% (95% BCI: 87, 95%), respectively. However, the convergence of LCM regression was possible when the lower limit of $Se$ of TUS was truncated to 60%, compatible with what we know about $Se$ of TUS in the literature (Rabeling et al., 1998; Ollivett et al., 2015; Berman et al., 2019). The posterior distributions of the weight coefficients for cough, ear droop or head tilt, and increased rectal temperature are detailed in Table 3. The posterior distribution of $Se$ and $Sp$ of Hap were estimated at 62% (95% BCI: 34, 87%) and 87% (95% BCI: 85, 91%), respectively. The weight coefficients’ posterior distributions were similar to those of the main model (overlaps of 95% BCI; Table 3, sensitivity analysis scenario).

Finally, we obtained the following VcCRS:

Specific score = 10 × Cough + 9 × Ear + 6 × Fever,

where Cough takes a value of 1 if induced or spontaneous cough is present and 0 otherwise, Ear takes a value of 1 if a slight uni- or bilateral droop or head tilt is present and 0 otherwise, and, finally, Fever takes a value of 1 if rectal temperature $\geq 39.7^\circ C$ and 0 otherwise.

**Optimal Cut-Off.** The MCT values for all the possible cut-off values of the VcCRS and for $r = 1:1$, $r = 3:1$, and $r = 8:1$ are shown in Figure 3. Whatever the value of $r$, the MCT values were minimized for a cut-off of 15 in our study population, with a prevalence of 5%. The cut-off of 15 was also optimal for $r = 1:1$ and $r = 3:1$ for a prevalence of 10%. For a prevalence of 25%, even if the cut-off of 15 remained acceptable for $r = 1:1$ but should be considered lower at 6 for $r = 3:1$ and $r = 8:1$ (Figure 3). We noticed from a practical perspective that, when using a cut-off of 15, only 2 predictors are sufficient to optimize the detection of active BRD. Thus, a calf with 2 findings among the 3 studied predictors (increased rectal temperature, ear droop or head tilt, and cough) could be considered as VcCRS positive, minimizing MCT in most scenarios.

### Step 2: External Validation of the VcCRS

The accuracy parameters ($Se$, $Sp$, PPV, NPV, LR+, and LR−) of VcCRS in the developing population and in both external populations are shown in Table 4. The performances remained similar in both population 1 and population 2. When we merged both external populations data sets (population 1 + population 2), the $Se$ and $Sp$ of VcCRS were 31% (95% BCI: 14, 70%) and 100% (95% BCI: 99, 100%). These latter values were further defined as the performances of VcCRS at the individual level.

### Final VcCRS.

The posterior distribution $Se$ and $Sp$ of TUS were estimated at 76% (95% BCI: 42, 96%) and 90% (95% BCI: 87, 95%), respectively. However, the convergence of LCM regression was possible when the lower limit of $Se$ of TUS was truncated to 60%, compatible with what we know about $Se$ of TUS in the literature (Rabeling et al., 1998; Ollivett et al., 2015; Berman et al., 2019). The posterior distributions of the weight coefficients for cough, ear droop or head tilt, and increased rectal temperature are detailed in Table

### Step 3: Application at Batch Level

The determination of the minimal number of calves to sample ($n$) is shown in Figure 4. For an HTP of 0.05 and an active BRD prevalence threshold of 15%, we found no difference in HPPV and HNPV between sampling 30, 20, or 10 calves for a $k$ of 3. Because it
is practical to sample as few calves as possible in a fattening unit, the optimal number \( n \) was therefore considered at 10 random calves to screen in the batch to be assessed.

The different values of herd predictive values for the different ranges of HTP, prevalence threshold, and \( k \) are presented in Figure 5. We noticed that whatever the prevalence threshold and HTP, the optimal \( k \) was 3—except for a prevalence threshold of 5%, where the optimal \( k \) was 4, meaning that 3 or more positive calves out of 10 assessed was a practical way to determine batch-level positivity (Graphical Abstract).

---

**Table 3.** Posterior median distributions and 95% Bayesian credible interval regression coefficients and their weights for increased temperature, ear droop or head tilt, and cough using the sensitivity (Se) and specificity (Sp) of thoracic ultrasonography (TUS; main model) or the Se and Sp of haptoglobin concentration (Hap; sensitivity analysis scenario)\(^1\)

<table>
<thead>
<tr>
<th>Item</th>
<th>Priors</th>
<th>Posterior distributions</th>
<th>Main model</th>
<th>Sensitivity analysis scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Se of TUS</td>
<td>β (8.0; 2.2)</td>
<td>0.76 (0.42; 0.96) (</td>
<td>(0.60; 0.96))^2</td>
<td>β (5.7; 2.5)</td>
</tr>
<tr>
<td>Sp of TUS</td>
<td>β (99.7; 6.2)</td>
<td>0.90 (0.87; 0.95)</td>
<td>β (316.6, 30.1)</td>
<td></td>
</tr>
<tr>
<td>Regression parameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_{\text{Temperature}} )</td>
<td>N (1.4, 1.7)</td>
<td>0.59 (−0.55; 1.7)</td>
<td>Ear droop or head tilt</td>
<td>Cough Weight 6</td>
</tr>
<tr>
<td>( \beta_{\text{Ear droop or head tilt}} )</td>
<td>N (1.4, 1.7)</td>
<td>0.90 (−0.25; 1.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_{\text{Cough}} )</td>
<td>N (1.4, 1.7)</td>
<td>0.96 (0.14; 1.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>N (1.4, 1.7)</td>
<td>−2.2 (−2.9; −1.7)</td>
<td>−2.2 (−2.9; −1.7)</td>
<td>−2.3 (−3.0; −1.8) —</td>
</tr>
</tbody>
</table>

\(^1\)β = beta distribution; \( N = \) normal distribution.

\(^2\)The distribution of Se was truncated at 0.60 to promote convergence of Bayesian latent class model.

---

**Figure 2.** Upset plot showing the total set size and overlaps between the presence of the 3 predictors included in the developed score (ear droop or head tilt, cough, and increased rectal temperature), thoracic ultrasonography (TUS) results, and haptoglobin results in 795 veal calves. The number of positive tests is indicated on the y-axis. The shaded circles connected by solid lines in the lower panel show the intersecting positive test data sets.
DISCUSSION

The objective of this study was to develop and validate a new CRSC to detect active BRD in veal calves housed individually, and to describe its accuracy for identifying individuals and groups of veal calves to treat. We found that calves with 2 of the following findings—cough, unilateral or bilateral ear droop or head tilt, or rectal temperature ≥39.7°C—were considered as positive, with an individual Se and Sp of 31% (95% BCI: 14, 70%) and 100% (95% BCI: 99; 100%), respectively, for active BRD detection. At the batch level, we found that ≥3 positive calves among 10 calves sampled 2 wk after arrival at the fattening unit allowed us to detect batches with active BRD prevalence ≥10% with positive and negative batch-level predictive values ≥94% and ≥95%, respectively.

Interestingly, our score differs from the previously developed CRSC in pre-weaning dairy-breed calves by the choice of clinical signs included (McGuirk, 2008; Love et al., 2014; Buczinski et al., 2018b). Previously, all reported respiratory clinical signs (i.e., nasal discharge, ocular discharge, cough, ear droop or head tilt, abnormal respiration) and increased rectal temperature were grouped for the development of a CRSC (McGuirk, 2008; Love et al., 2014; Buczinski et al., 2018b). In contrast, we first selected the clinical signs to include in the CRSC according to their reliability when used by different persons involved in veal calves’ health monitoring (producers, technicians, and veterinarians; Berman...
et al., 2021). This a priori selection has the advantages of removing most operator variability from the CRSC used and being applied in a population-based approach, where the first-line diagnosis is generally not performed by a veterinarian.

Additionally, this selection of robust clinical signs limited the number of predictors to assess [only 3 predictors vs. 5 or 6 in WiCRSC and (Vc)CaCRSC, respectively], which could have affected the Se of Vc-\text{CRS} (Se of 31%). Moreover, this selection implied that the 3 predictors included in VcCRS are less specific of lower respiratory tract disorders than other clinical signs such as dyspnea. However, comparison of Se and Sp of VcCRS and VcCRS + dyspnea did not show any difference on Se and Sp (data shown in Supplemental File S2). Therefore, this selection simplifies the use of the VcCRS in the context of veal calves in a large fattening unit. Even simpler, we showed that the presence of only 2 findings among cough, ear droop or head tilt, and increased rectal temperature is necessary to consider a calf positive. After this initial assessment, induced cough or increased rectal temperature are tested a second time, which limits the handling of calves and the potential biosecurity issues (e.g., transmission of disease). Interestingly, we showed that randomly sampling 10 calves from a batch of calves (e.g., batches varying from 100 to 800 calves) was sufficient to apply our score at the batch level, limiting time consumption. The VcCRS is therefore an easy, simple, and quick score to use on the individual and group levels in the veal calf industry.

Estimating VcCRS performances at both the individual and the group level contrasts with previously reported CRSCs. Using VcCRS at the group level had the advantage of providing information on whether a group of calves needs to be treated with antimicrobials or not, and would help to rationalize the metaphylactic treatment approach (Baptiste and Kyvsgaard, 2017). We showed that 3 calves with positive individual VcCRS scores among 10 randomly sampled in a group allowed us to detect batches with active BRD prevalence $\geq 10\%$, with batch-level predictive values $>0.90$. Concretely, if we consider the prevalence of risk as defined at 21% in Leruste et al. (2012), being VcCRS-positive at the batch level (i.e., 3 individual VcCRS-positive calves among 10 sampled) would imply that the batch has a 99% chance of having active BRD prevalence $\geq 21\%$; otherwise, the batch has a 95% chance of not having active BRD prevalence $\geq 21\%$. However, it is difficult to define a universally accepted active BRD prevalence threshold in veal calves, as such a threshold may depend on various factors, including the calves’ price paid in the auction market, quality, origin, the anticipated meat marketing context, specific farm risk factors (ventilation, caretakers), the season, or the agent involved. That is why, in this study, we reported the VcCRS performance at the batch level for a range of active BRD prevalence scenarios (Figure 5). With this approach, one could refer to this figure to adjust the intervention in a specific group under specific circumstances. At a predefined active BRD prevalence threshold, our score could be used in the future to guide accurate metaphy-

### Table 4. Posterior median and 95% credible intervals based on 4 Bayesian latent class models for the prevalence of active bovine respiratory disease, sensitivity (Se) and specificity (Sp), positive (PPV) and negative (NPV) predictive values, positive (LR+) and negative (LR−) likelihood ratios of the clinical respiratory score chart for veal calves (VcCRS), and Se and Sp of thoracic ultrasonography (TUS) and concentration of haptoglobin (Hap) used to diagnose active bovine respiratory disease

<table>
<thead>
<tr>
<th>Item</th>
<th>Development population</th>
<th>Population 1</th>
<th>Population 2</th>
<th>Populations 1 and 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se_VcCRSC</td>
<td>0.30 (0.12; 0.72)</td>
<td>0.51 (0.22; 0.87)</td>
<td>0.33 (0.07; 0.93)</td>
<td>0.31 (0.14; 0.70)</td>
</tr>
<tr>
<td>Sp_VcCRSC</td>
<td>0.99 (0.98; 1.00)</td>
<td>1.0 (0.97; 1.00)</td>
<td>0.99 (0.98; 1.0)</td>
<td>1.0 (0.99; 1.0)</td>
</tr>
<tr>
<td>PPV_VcCRSC</td>
<td>0.53 (0.25; 0.85)</td>
<td>0.86 (0.39; 0.99)</td>
<td>0.58 (0.16; 0.96)</td>
<td>Pop 1: 0.89 (0.64; 0.99)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pop 2: 0.65 (0.26; 0.97)</td>
</tr>
<tr>
<td>NPV_VcCRSC</td>
<td>0.97 (0.92; 0.99)</td>
<td>0.97 (0.93; 1.0)</td>
<td>0.98 (0.94; 1.0)</td>
<td>Pop 1: 0.94 (0.87; 0.99)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pop 2: 0.98 (0.95; 1.0)</td>
</tr>
<tr>
<td>LR+_VcCRSC</td>
<td>27.1 (7.8, 110)</td>
<td>104.8 (14.2; 3.026)</td>
<td>59.1 (7.6; 813)</td>
<td>74.3 (18.3; 853)</td>
</tr>
<tr>
<td>LR_–VcCRSC</td>
<td>0.70 (0.28; 0.89)</td>
<td>0.49 (0.14; 0.79)</td>
<td>0.68 (0.07; 0.94)</td>
<td>0.69 (0.30; 0.87)</td>
</tr>
<tr>
<td>Se_TUS</td>
<td>0.80 (0.54; 0.96)</td>
<td>0.88 (0.68; 0.98)</td>
<td>0.79 (0.50; 0.96)</td>
<td>0.86 (0.64; 0.97)</td>
</tr>
<tr>
<td>Sp_TUS</td>
<td>0.90 (0.87; 0.93)</td>
<td>0.92 (0.89; 0.96)</td>
<td>0.95 (0.93; 0.98)</td>
<td>0.95 (0.93; 0.98)</td>
</tr>
<tr>
<td>Se_Hap</td>
<td>0.64 (0.40; 0.85)</td>
<td>0.72 (0.55; 0.86)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sp_Hap</td>
<td>0.87 (0.84; 0.90)</td>
<td>0.95 (0.92; 0.98)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prevalence</td>
<td>0.04 (0.02; 0.10)</td>
<td>0.08 (0.04; 0.15)</td>
<td>0.03 (0.006; 0.07)</td>
<td>Pop 1: 0.09 (0.04; 0.16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pop 2: 0.03 (0.007; 0.06)</td>
</tr>
</tbody>
</table>

1Population from Berman et al., 2019.
2Population from Morin et al., 2020.
3For this analysis we used a latent class model with 2 independent populations (Pop 1 and Pop 2) but assumed similar sensitivity and specificity across populations.
lactic group treatments and, ultimately, reduce antimicrobial use (Baptiste and Kyvsgaard, 2017).

At the individual level, the Se of VcCRS was 31% and, therefore, lower than the Se of previous CRSC (range: 46–83%; Buczinski et al., 2015b, 2018b; Love et al., 2016). However, the individual Sp of VcCRS was almost perfect and higher than in previous studies (range: 74–91%; Buczinski et al., 2015b, 2018b; Love et al., 2016). The almost perfect Sp of VcCRS had the advantage of limiting false-positive calves and, therefore, useless individual antimicrobial treatment and economic loss (Theurer et al., 2015). The selection of reliable clinical signs and the better definition of respiratory disease as active BRD could explain the Sp superiority of VcCRS versus previous clinical scores in individual pre-weaning dairy-breed calves (Buczinski et al., 2015b, 2018b; Love et al., 2016). Although the individual performances seem low, with Se of 31%, the goal of this study was to implement a diagnostic tool to target treatment not at the individual level but at the group level (metaphylactic). The high Sp of VcCRS permits us to have a high HPPV at the group level while avoiding false-positive batches and, therefore, unnecessary treatment of the entire batch.

Our study differs from others that have developed CRSC in that its design follows a robust statistical approach, as recommended by medical guidelines (Moons et al., 2015). First, we externally validated our score, in contrast with previous CRSC in calves (McGuirk, 2008; Love et al., 2014). This validation increases the robustness, credibility, applicability, and generalization of VcCRS across veal calf settings. Second, WiCRSC and CaCRSC used the combination of multiple tests (a composite reference test) to elaborate their scores and, therefore, used an inaccurate definition of respiratory disease to define cases (McGuirk, 2008; Love et al., 2014). In this study, we used Bayesian LCM, as recommended by the World Organization for Animal Health, in the absence of a gold standard to limit classification bias and ensure better accuracy estimation (Cheung et al., 2021). Spectrum bias was also limited by selecting prospectively random calves from commercial veal facilities that represented wide spectrums of active BRD severity (ranging from healthy calves to calves with mild and moderate active BRD to calves with severe active BRD; Buczinski and O’Connor, 2016). In contrast, the case-control designs used in McGuirk (2008) and Love et al. (2014) may have included cases at the ends of

Figure 4. Batch-level positive (solid line) and negative (dashed line) predictive values for a bovine respiratory disease prevalence threshold of 15%, considering the true proportion of positive batches of 5%, for different batch cut-off values (k) and numbers of sampled calves (n). The error bars represent the dispersion around the median value (i.e., the 95% Bayesian credible intervals).
Figure 5. Batch-level positive (HPPV) and negative (HNPV) predictive values (PV) for a minimal number of positive calves to consider batch as positive (k), true proportion of positive batches (HTP), and active bovine respiratory disease (BRD) prevalence when the clinical respiratory chart score is applied on 10 veal calves in a batch. The dashed line represents a PV of 0.9. The HPPV and HNPV with 95% Bayesian credible intervals are presented.
the spectrum of disease (either healthy calves or calves with severe active BRD), resulting in an overestimation of their scores’ coefficients.

This study has some limitations, however. First, the posterior median active BRD prevalence was low (5% vs. 20% expected for sample size calculation). This low prevalence is likely because we sampled calves before the main peak of active BRD incidence to promote early detection and limit the proportion of treated calves. Consequently, sensitivity 95% BCI estimates were wide. This uncertainty prevented us from adequately estimating the regression coefficients and forced us to truncate the lower limit of Se of the 95% BCI of TUS to promote convergence of our statistical model. However, this low active BRD prevalence implied large variation in the individual CRSC Se (median value of 31%, but 95% BCI from 14 to 70%) but not in accurately estimated Sp (median of 100%, with 95% BCI from 99 to 100%), which is reported to have the most influence on active BRD diagnosis (Theurer et al., 2015). Second, we chose by convenience a temporal external validation; that is, we used participant data collected by the same investigators, using the same predictors and target condition definitions and measurements, but sampled from an earlier period (Moons et al., 2015). Transportability from one setting to another could have been increased by using a database collected by other investigators in another country (broad validation), in other types of patients (dairy calves), or considering the probable influence of infectious agents (high prevalence of _M. bovis_ in this study). However, our study represents the first step in the development of a VcCRS, and other external validations could be performed in the future to tailor its use in other specific contexts.

CONCLUSIONS

We developed a simple batch-level score that is reliable across examiners and performs effectively in the detection of active BRD in veal calves. According to the requirements of the industry, the active BRD prevalence threshold could be defined. The positive VcCRS group could then be treated metaphylactically or investigated. The implementation of this VcCRS chart in the veal calf industry would promote the elaboration of a protocol to tailor antimicrobial use. A prospective cohort study comparing antimicrobial use and production outcomes of groups applying the VcCRS chart (exposed) and groups without (non-exposed) would be the next step to validate this CRSC and justify its large-scale implementation in veal calves.

ACKNOWLEDGMENTS

Funding was received from the Producteurs Bovins du Québec (Longueuil, Québec, Canada), Programme Innov’Action Agro-alimentaire (Project IA117743: Amélioration des veaux lourds souffrant de pneumonies) du Ministère de l’Agriculture, des Pêcheries et de l’Alimentation du Québec (MAPAQ, Saint-Hyacinthe, Québec, Canada), Délimax Veaux Lourds Ltee (Saint-Hyacinthe, Québec, Canada), Aliments Prolacto Inc. (Villeroy, Québec, Canada), and FQRNT (Fonds Québécois de Recherche en Nature et Technologie; Montréal, Québec, Canada). The authors thank Luke Palder from Proofreading Services (Raleigh, NC) for the English language proofreading. The authors have not stated any conflicts of interest.

REFERENCES


Buczinski, S., and A. M. O’Connor. 2016. Specific challenges in conducting and reporting studies on the diagnostic accuracy of ultra-


ORCIDS

J. Berman † https://orcid.org/0000-0001-7947-760X
D. Francoz † https://orcid.org/0000-0002-1634-0507
S. Dufour † https://orcid.org/0000-0001-6418-0424
S. Buczinski † https://orcid.org/0000-0002-8460-4885