Performance of various interpretations of clinical scoring systems for diagnosis of respiratory disease in dairy calves in a temperate climate using Bayesian latent class analysis

John D. Donlon,1,2* Conor G. McAloon2,1* and John F. Mee3
1School of Veterinary Medicine, University College Dublin, Belfield, Dublin 4, D04 W6F6, Ireland
2Animal and Bioscience Research Department, Teagasc, Animal & Grassland Research and Innovation Centre, Grange, Dunsany, Co. Meath, C15 PW93
3Animal and Bioscience Research Department, Teagasc, Moorepark Research Centre, Fermoy, Co.Cork, P61 P302, Ireland

ABSTRACT

Bovine respiratory disease (BRD) presents a challenge to farmers all over the globe not only because it can have significant impacts on welfare and productivity, but also because diagnosis can prove challenging. Several clinical scoring systems have been developed to aid farmers in making consistent early diagnosis, 2 examples being the Wisconsin (WCS) and the California (CALIF) systems. Neither of these systems were developed in or for use in a temperate environment. As environment may lead to changes in BRD presentation, the weightings and cut offs designed for one environmental presentation of BRD may not be appropriate when used in a temperate climate. Additionally, the interpretation of the scores recorded varies between studies; this may also influence conclusions. Hence, the objective of this work was to investigate the sensitivity (Se) and specificity (Sp) of these tests in a temperate climate and investigate the influence of varying the interpretation on the performance of the WCS. In this prospective study, 98 commercial spring calving dairy farms were recruited (40 randomly, 58 targeted) and visited. Thoracic ultrasound and WCS was performed on 20 randomly sampled calves between 4 and 6 weeks of age on each farm. On a subset of 32 farms, the CALIF score was also undertaken. The data were then used in a hierarchical Bayesian latent class model to estimate the Se and Sp of 5 different interpretations of the Wisconsin clinical score and one interpretation of the California clinical score. In total, 1,936 calves were examined. The Se of the Wisconsin score varied from 0.336 to 0.577 depending on the interpretation used and the Sp varied from 0.943 to 0.977. The Se of the California score was 0.529 (95% Bayesian credible interval (BCI); 0.403, 0.651) and the Sp was 0.903 (95% bci; 0.883, 0.922). In conclusion, the performance of the clinical scores in a temperate environment were similar to previously published work from more extreme climates, however the performance varied widely depending on the score interpretation. Authors should justify their usage of a particular clinical score interpretation to improve clarity in publications.

Key words: Bovine respiratory disease, Bayesian latent class analysis, California clinical score, Wisconsin clinical score

INTRODUCTION

Rapid diagnosis and treatment is essential for reducing the deleterious impact of BRD on calf welfare and productivity (Lorenz et al., 2011). In a clinical setting, diagnosis of BRD is generally on the basis of a clinical examination by a veterinarian. Clinical respiratory scoring (CRS) has been developed to standardise some aspects of the clinical exam, facilitating producer detection and standardising the detection of cases and non-cases in a research setting (McGuirk and Peek, 2014; Lago et al., 2006; van Leenen et al., 2020; Rhodes et al., 2021). Examples of some clinical scoring systems which can be used to diagnose BRD are the Wisconsin clinical score (WCS) (McGuirk and Peek, 2014) and the California clinical score (CALIF) (Love et al., 2014). Both can be carried out relatively quickly, use clinical signs which are readily observable and only require a thermometer making them cheap to implement. These characteristics have led to widespread implementation of both WCS and CALIF in research. Initially McGuirk (2008) used unpublished broncho-alveolar cytology and bacterial culture to define the WCS case thresholds, making it difficult to assess their validity. Other researchers have subsequently conducted research to validate the sensitivity (Se) and
specificity (Sp) of both the WCS and CALIF (Buczinski et al., 2015; Decaris et al., 2022).

The estimates of Se and Sp for the WCS were shown to vary widely (Se; 0.48–0.78, Sp; 0.74–0.99) (Buczinski et al., 2015; Berman et al., 2019; Decaris et al., 2022). Buczinski et al. (2016) also showed that the inter-rater reliability of WCS can be poor. The CALIF system attempted to reduce inter-rater variability by making each component of score dichotomous rather than having several levels (Love et al., 2016). Preliminary efforts have been made to validate and refine CRS, however, it is clear that their performance requires further investigation.

The implementation of the WCS has not been consistent across publications, with some authors reinterpreting the scores or altering the case threshold (Lago et al., 2006; Medrano-Galarza et al., 2018; Binversie et al., 2020; Calderón-Amor and Gallo, 2020). Alteration of the interpretations or threshold of WCS is likely to lead to changes in the operating characteristics of the test. For example Decaris et al. (2022) examined the performance of the WCS using 3 different cut-offs (≥4, ≥ 5, 2 scores ≥ 2) and found that the median Se varied from 0.75 – 0.97 while the median Sp varied from 0.53 – 0.82. Changes in test characteristics may be useful where the test is being used for different purposes, e.g., screening v diagnosis/confirmation. However, it is often the case that publications give no reason for the alteration in threshold. In addition, when an inconsistent methodology is used, comparison of results between publications is difficult and thus synthesis of higher evidence, i.e., meta-analysis, is hampered. This is an area that requires more research so that readers, authors and reviewers are well informed about test performance when alternative interpretations are used.

In addition, the methodological and interpretative limitations of existing studies linking CRS to BRD, the environment in which these studies were conducted may influence outcomes. Both the CALIF and the WCS were developed in continental climates: these climates may expose calves to extremes of ambient temperature and relative humidity at particular times of the year. For example, in California, the minimum temperature detected in calf housing by Louie et al. (2018) was 11.5°C with a maximum temperature of 41.9°C. In contrast, in Wisconsin, Lago et al. (2006) found a mean temperature within calf housing of 3.9°C (minimum −6.7°C, maximum: 12.2°C). It is possible that these exposures may affect the presentation of BRD in various climates, and therefore affect the diagnostic accuracy of clinical scoring systems in more temperate climates. Thus, it is important to verify that the weightings used when developing the CALIF and WCS scoring systems are appropriate for use in different climate/production systems and produce similar Se and Sp to those seen in the USA.

Apart from sensitivity and specificity, there are several other metrics that can be useful in understanding the results of a test in a given population: positive predictive value (PPV), negative predictive value (NPV) and accuracy. In the context of this work PPV represents the likelihood that a calf diagnosed with BRD according to a CRS is a true case of BRD, NPV represents the likelihood that a calf considered negative according to a CRS is truly BRD free. Accuracy reflects the likelihood that calves are correctly categorized as either BRD positive or negative. These metrics are a function of both the Se, Sp of CRS as well as the prevalence of BRD in the population which they are applied. Understanding PPV, NPV and accuracy will allow us to better understand if CRS adequately categorizes cases of BRD in a population such as the one examined in this work.

Given the potential limitations of the existing research on the performance of CRS in diagnosing BRD, we aimed to investigate the effects of altering threshold and interpretation on CRS accuracy using a hierarchical Bayesian latent class model. We hypothesize that the performance of WSC and CALIF may differ in calves from pasture-based dairy herds in a temperate climatic region when compared with the regions where they were developed. Second, we hypothesize that the alteration of CRS thresholds would result in altered Se and Sp of those tests.

**MATERIALS AND METHODS**

**Reporting Guidelines**

This work was undertaken using the STARD BLCM framework (Kostoulas et al., 2017). This was a prospective cross-sectional study; the data collection was planned for estimation of the sensitivity and specificity of the WCS and CALIF.

**Ethics**

This study was approved by University College Dublin, Animal Research Ethics Committee and the Health Products Regulatory Authority (V016/2020Q1).

**Herd selection**

The herds used in this work were recruited primarily for a study of housing environment and BRD. The sample size was based on a power calculation to determine the relationship between BRD prevalence and bacterial air count. Two herd types were recruited for this work. One cohort was randomly recruited to be representative of typical Irish dairy herds (n = 40). A problem herd cohort was identified as herds with a prior history of BRD in...
dairy calves, these may have been herds identified by private veterinary practitioners (PVP) or the farmers themselves. The randomly recruited Irish herds were contacted through the national animal breeding organization (ICBF) as previously described by Donlon et al. (2023). Problem herds were recruited through 2 means. The first was a letter distributed to PVPs who had submitted calves for post mortem which were subsequently diagnosed with BRD or related samples (i.e., nasal swabs) to any regional veterinary laboratory in Ireland. It asked the PVPs to refer any farmers they felt had a significant issue with BRD to the first author (n = 42). The second way that problem farms were recruited was by ICBF: sending a letter to (n = 100) dairy farmers detailing the study and asking any farmers with a significant BRD problem to contact the first author (n = 16). No limits were put on the farm location or calf vaccination status. However, if the farmer had less than 70 lactating cows they were excluded because they would not have had a sufficient number of calves within or close to the required age range at the time of the visit. All farms (n = 98) enrolled on the farm location or calf vaccination status. However, if the farmer had less than 70 lactating cows they were excluded because they would not have had a sufficient number of calves within or close to the required age range at the time of the visit. All farms (n = 98) enrolled (40 normal and 58 problem farms) were visited once over a 3 year period in the spring of 2020 (n = 29), 2021 (n = 37) and 2022 (n = 32).

Calf enrolment

At each visit, 20 calves, of either sex, were recruited for examination. Calves were selected by age: calves between 4 and 6 weeks of age were sampled. If the farm did not have 20 calves between 4 to 6 weeks of age, then the calves greater than 6 weeks of age that were not yet weaned were sampled preferentially to calves less than 4 weeks. Four to 6 weeks old has previously been identified as the age at which the peak prevalence of BRD is observed (Lago et al., 2006; Rhodes et al., 2021). The researchers sampled all age-appropriate calves in the same pen as opposed to moving between pens. In pens where there were more than 20 age-appropriate calves, 20 calves were selected by entering a pen and reading tag numbers until a calf in the specified age range was found, the calf was then sampled and marked to avoid repeated sampling. Exclusion criteria such as breed or other signs of illness were not used. All calves enrolled had their weight estimated using a weigh band (Volac International Limited).

Clinical assessment

A clinical assessment was undertaken on each calf (n = 1,936). In 2020 and 2021, the WCS was undertaken (n = 1,308) (Table 1). In 2022, in addition to the WCS, respiratory effort was included (n = 628) so that the CALIF score could also be calculated. Respiratory effort was scores as 0 if the calf appeared normal and 2 if tachypnea or dyspnea were noted. The first author visited all farms assisted by 25 operators (3 operators/farm) who conducted the clinical scoring. The clinical assessments were supervised at all times by the first author and the operators were all provided with a reference WCS chart. Logistically it was not possible to have a single operator carry out all of the clinical scoring. Inter-rater reliability was not investigated as part of this work.

Thoracic ultrasound scoring (TUS)

All TUS was undertaken by the first author. The TUS was carried out immediately after the clinical assessment was performed; it was not possible to blind the TUS operator to the clinical assessment results. A detailed description of the TUS technique used can be found in Donlon et al. (2023). Thoracic ultrasound scoring was chosen as the second diagnostic test because of its ability to be conducted rapidly calf-side with readily available diagnostic equipment. In summary, calves were examined from the 10th intercostal space to the 2nd on the left hand side and from the 10th intercostal space to the 1st on the right hand side. A score between 0 and 5 was given to the degree of consolidation as described by Ollivett and Buczinski (2016). Score 0 was assigned to a calf with no lung consolidation with or without isolated comet tail artifacts. Score 1 was assigned to a calf with diffuse comet tail artifacts. Score 2 was assigned to a calf with isolated patches of ultrasonographic lung consolidation. Score 3 was assigned where a calf had consolidation of a full lung lobe. Score 4 was assigned where a calf had consolidation of 2 full lung lobes. Score 5 was assigned where a calf had consolidation of 3 or more lung lobes. The case definition of BRD as diagnosed by TUS was a score of ≥ 3. This definition was chosen because it has previously been shown to be associated with reduced growth rate in Irish dairy calves (Rhodes et al., 2021). Because the author had previously shown that the prevalence of BRD in Irish dairy calves was low (4%, 95% Bayesian confidence interval (1%, 8%)) (Donlon et al., 2023) a cut off which would produce a higher TUS Sp was desirable for this work.

Data management

Data management was as described in Donlon et al. (2023). All data were recorded using a file maker pro application (Claris) which produced a CSV file containing the calf tag number, sex, age, each component clinical score, weight and thoracic ultrasound result. All record-ings were imported into R version 4.0.0 (https://www.r-project.org/). All data manipulation, visual presentation and statistical analysis was conducted using ´tidyverse´
Calves \( (n = 25) \) with missing CRS or TUS data were removed before modeling.

**Statistical analysis**

The statistical analysis was split into 2 parts: an assessment of the Se and Sp of various interpretations of the WCS using all of the data collected, and an estimation of Se and Sp of the CALIF respiratory score using the data collected in Spring 2022.

**Various interpretations of the Wisconsin score.** Five different published interpretations of the Wisconsin score were chosen for Se and Sp estimation. Hereunder is a list of the interpretations from each publication:

1. McGuirk and Peek (2014): If a sum of all component scores is \( \geq 5 \), or 2 or more of the component scores is \( \geq 2 \), the calf is considered BRD positive
2. Lago et al. (2006): If a sum of all component scores is \( \geq 6 \) the calf is considered BRD positive
3. Calderón-Amor and Gallo (2020): If the cough score is \( \geq 2 \) and nasal discharge or ocular discharge score is \( \geq 1 \) the calf is considered BRD positive.
4. Medrano-Galarza et al. (2018): If the sum of all component scores is \( \geq 5 \) the calf is considered BRD positive.
5. Binversie et al. (2020): If any of the 2 component scores are \( \geq 2 \) the calf is considered BRD positive.

These publications do not represent an exhaustive list of WCS variations. McGuirk and Peek (2014) was chosen because they are the originators of the WCS. Lago et al. (2006), Calderon-Amor and Gallo (2020) and Medrano-Galarza et al. (2018) were all chosen because they had previously been cited by the author in work comparing the prevalence of BRD (Donlon et al., 2023). In that publication it was noted that they all used different interpretations of the WCS making comparison of results difficult. Binversie et al. (2020) was also included as it was the interpretation used by Decaris et al. (2022) in an investigation of WCS performance in a tropical climate.

For each of these interpretations a data frame was created that categorized the calves using a combination of the dichotomous tests results from TUS and WCS. Using these results, a hierarchical Bayesian latent class model (HBLCM) was constructed. We assumed that the Se and Sp were constant across all herds, that the within herd prevalence of BRD varied across the herds with a proportion being disease-free (no cases of BRD diagnosed), and that the 2 tests being used (TUS and WCS) were conditionally independent of each other.

The likelihood of each of the 4 possible outcomes from the test results of each interpretation was specified using a multinomial distribution:

\[
Y_k \sim \text{multinomial}(n_k, (P_{11k}, P_{10k}, P_{01k}, P_{00k}))
\]

For the \( k \)-th herd, the \( n_k \) observations gave the data vector for the joint test results \( P_{11k}, P_{10k}, P_{01k}, P_{00k} \). Where the \( P_{11k} \) is the number of calves from the \( k \)-th herd that test positive on TUS and each WCS interpretation, and \( P_{10k} \) is the number that test positive on TUS and negative on each WCS interpretation, and so on.

The multinomial cell probabilities are given by:

\[
P_{11k} = \pi_k \times Se_{TUS} \times Se_{WCS} + (1 - \pi_k) \times (1 - Sp_{TUS}) \times (1 - Sp_{WCS})
\]

\[
P_{10k} = \pi_k \times Se_{TUS} \times (1 - Se_{WCS}) + (1 - \pi_k) \times (1 - Sp_{TUS}) \times Sp_{WCS}
\]

\[
P_{01k} = \pi_k \times (1 - Se_{TUS}) \times Se_{WCS} + (1 - \pi_k) \times Sp_{TUS} \times (1 - Sp_{WCS})
\]

\[
P_{11k} = \pi_k \times (1 - Se_{TUS}) \times (1 - Se_{WCS}) + (1 - \pi_k) \times Sp_{TUS} \times Sp_{WCS}
\]

### Table 1. Distribution of Wisconsin clinical score components in the calves examined as part of this work. Each cell contains the number of calves that were attributed a given score, the corresponding percentage and the criteria used to define that score

<table>
<thead>
<tr>
<th>Score</th>
<th>Cough</th>
<th>Nose</th>
<th>Eye</th>
<th>Ear</th>
<th>Temperature score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1617  (83.52%) (no cough)</td>
<td>1646  (85.02%) (normal serous discharge)</td>
<td>1354  (69.94%) (normal eye)</td>
<td>1904  (98.35%) (normal ear)</td>
<td>148  (7.64%) (&gt;38.5°C)</td>
</tr>
<tr>
<td>1</td>
<td>150   (7.75%) (induced single cough)</td>
<td>221   (11.42%) (small amount of unilateral, cloudy discharge)</td>
<td>548    (28.31%) (mild ocular discharge)</td>
<td>19    (0.98%) (ear flicking)</td>
<td>461  (23.81%) (38.9–39.4°C)</td>
</tr>
<tr>
<td>2</td>
<td>128   (6.61%) (induced repeated coughs or occasional spontaneous)</td>
<td>61    (3.15%) (bilateral, cloudy, or excessive mucus)</td>
<td>31     (1.60%) (moderate bilateral ocular discharge)</td>
<td>6     (0.31%) (slight unilateral ear drop)</td>
<td>1103  (56.94%) (38.9–39.4°C)</td>
</tr>
<tr>
<td>3</td>
<td>39    (2.01%) (repeated spontaneous coughing)</td>
<td>3     (0.15%) (copious, bilateral mucopurulent nasal discharge)</td>
<td>2      (0.10%) (heavy ocular discharge)</td>
<td>2     (0.10%) (severe head tilt, or bilateral ear droop)</td>
<td>222  (11.47%) (&gt;39.4°C)</td>
</tr>
<tr>
<td>NA</td>
<td>2     (0.10%)</td>
<td>5     (0.26%)</td>
<td>1     (0.05%)</td>
<td>5     (0.26%)</td>
<td>2     (0.10%)</td>
</tr>
</tbody>
</table>

Donlon et al.: Performance of various…
Where $\pi_k$ is the prevalence in the $k$-th herd and $\text{Se}_\text{TUS}$, $\text{Se}_\text{WCS}$, $\text{Sp}_\text{TUS}$, $\text{Sp}_\text{WCS}$ are the sensitivities and specificities of thoracic ultrasound and each interpretation of the Wisconsin clinical score.

The method of prevalence modeling used has previously been described in Donlon et al. (2023). Briefly, the within-herd prevalence ($\pi_k$) was modeled as a product of herd-level prevalence ($h_k$) and the ‘conditional’ within-herd prevalence ($\psi_k$), where $\mu$ represented the proportion of herds that were BRD-free. The conditional within-herd prevalence was the prevalence of BRD in herds where BRD was present and was modeled using an intercept-only random effect logistic regression where $\alpha$ was the intercept, and $\epsilon_k$ was the farm-level random effect.

$$\pi_j \sim h_k \times \psi_k$$
$$\text{logit}(\psi_k) < -\alpha + \epsilon_k$$
$$h_k \sim \text{Bernoulli} (\mu)$$
$$\epsilon_k \sim \text{normal} (0, 1/\tau)$$
$$\tau \sim \text{gamma} (10,10)$$
$$\mu \sim \beta (1, 1)$$

The Se and Sp of each test were modeled using $\beta$ distributions:

$$\text{Se}_\text{WCS} \sim \beta (\alpha \text{Se}_\text{WCS}, \beta \text{Se}_\text{WCS})$$
$$\text{Se}_\text{TUS} \sim \beta (\alpha \text{Se}_\text{TUS}, \beta \text{Se}_\text{TUS})$$
$$\text{Sp}_\text{WCS} \sim \text{beta} (\alpha \text{Sp}_\text{WCS}, \beta \text{Sp}_\text{WCS})$$
$$\text{Sp}_\text{TUS} \sim \text{beta} (\alpha \text{Sp}_\text{TUS}, \beta \text{Sp}_\text{TUS})$$

Posterior inferences for each parameter ($\text{Se}_\text{WCS}$, $\text{Se}_\text{TUS}$, $\text{Sp}_\text{WCS}$, $\text{Sp}_\text{TUS}$, $\pi_k$, $\epsilon_k$ and $h_k$) were obtained using JAGS called from R statistical software using the ‘rjags’ package (41). Markov chains ran for 15,000 iterations after a burn in period for 5,000 iterations. Convergence of the Markov chains was assessed by visual assessment of Markov chain and autocorrelation plots and by running multiple ($n = 2$) chains from dispersed starting values (e.g., 0.05 and 0.95 for variables bounded between 0 and 1).

**Model priors - test characteristics.** In a previous publication we undertook a literature review to estimate the sensitivity and specificity of both TUS and WCS (Donlon et al. 2023). In brief, PubMed and CAB Direct were searched on the 30th of November 2022. Publica-
tions were included if they used calves that were tested on farm with a clearly defined testing methodology and case definitions for both TUS and WCS. The publications must also have used a statistical framework that accounted for the lack of a gold standard test. Estimates from 4 publications were included for TUS and 3 publications for WCS.

For the model priors in this study, the estimates were taken from that publication as informative priors. A consistent prior was used for WCS across all of the interpretations as Se and Sp estimates were not available for each individual interpretation.

Model priors – prevalence. Within herd prevalence ($\alpha$) was modeled as a normal distribution (−2.3, 0.4); to reflect a mean prevalence of 0.1 and a 95% confidence that the prevalence <0.5 this was calculated using Excel (Microsoft). A gamma distribution of 10, 10 was used as the prior for tau. The gamma distribution equates to the variance of the logit of normal distribution and considering the priors used, this allowed for the within-herd prevalence to vary from 0 to 100%.

Sensitivity analysis. As part of the sensitivity analysis, several measures were undertaken: the final model was checked by changing the prior information of WCS Se and Sp to non-informative $\beta$ (1,1). The analysis was also conducted assuming conditional dependence between TUS and WCS.

PPV, NPV and accuracy estimates. The Se and Sp estimates for each interpretation were then used to produce positive predictive value, negative predictive value (Table 7) and accuracy (Table 8) estimates for each of the interpretations at different prevalence levels. In the context of this work, PPV is the proportion of positive CRS results that is truly BRD cases (Monaghan et al., 2021). Negative predictive value is the proportion of negative CRS results that is truly BRD free (Monaghan et al., 2021). Accuracy represents the proportion of calves that were correctly categorized as either BRD-free or a case from the entire population examined (Monaghan et al., 2021).

Estimate of the Se and Sp of CALIF. For the second part of this work, a subset of recordings ($n = 628$) was used which included scores of respiratory effort. The data were used to categorise calves as positive or negative for BRD according to the CALIF case definition of Love et al. (2014). The data were handled in the same manner as previously described for each interpretation of the WCS.

Model priors. The model priors previously described were used for both TUS and BRD prevalence. A prior estimate of the sensitivity and specificity of CALIF was taken from the posterior produced by Decaris et al. (2022).

Sensitivity analysis. Two methods were used to conduct a sensitivity analysis. First, a model was produced where the results of TUS, CALIF and WCS were modeled as 3 tests in a HBLCM. The WCS and CALIF were considered co-dependent in this model as the majority of the data used to calculate them were the same. This model was ran using informed and uninformative priors. Initially autocorrelation was observed in some of the sample parameters so the Markov chains were thinned by 10 for inference. The structure of this model can be found in the appendix. Second CALIF and TUS only were modeled using the method previously described for the overall data set; a sensitivity analysis was conducted whereby uninformative priors and codependency with TUS was also modeled. For comparison, these models were also rerun using the McGuirk and Peek (2014) interpretation of WCS individually on the same subset of calves.

RESULTS

Farm descriptions

All farms recruited (98) had spring-calving herds ranging from 70 to 900 dairy cows with a median of 147. The majority of farms were located in the province of Munster in the south of Ireland (67.3%, $n = 66$).

All farmers housed preweaning calves in groups indoors; 87.2% ($n = 86$) used natural ventilation and 12.2% ($n = 12$) used mechanical ventilation. Straw was used as a bedding on 92% ($n = 90$) of farms with the other farmers using woodchip (5.1%, $n = 5$), sawdust (2.0%, $n = 2$) or rushes (1.0%, $n = 1$). The majority of farmers fed calves using teat feeders (59.2%, $n = 58$), but automatic feeders (36.7%, $n = 36$) and buckets (4.1% $n = 4$) were also used. Calves were fed milk replacer (53.1%, $n = 52$) or whole milk (23.5%, $n = 23$) on the majority of farms with

<table>
<thead>
<tr>
<th>TUS Score</th>
<th>No.</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1025</td>
<td>52.92%</td>
</tr>
<tr>
<td>1</td>
<td>525</td>
<td>27.1%</td>
</tr>
<tr>
<td>2</td>
<td>176</td>
<td>9.09%</td>
</tr>
<tr>
<td>3</td>
<td>161</td>
<td>8.32%</td>
</tr>
<tr>
<td>4</td>
<td>33</td>
<td>1.70%</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.31%</td>
</tr>
<tr>
<td>NA</td>
<td>10</td>
<td>0.52%</td>
</tr>
</tbody>
</table>
Table 4. Estimates of sensitivity and specificity of the Wisconsin clinical score interpretations from the hierarchical Bayesian latent class model including a 95% Bayesian confidence interval. Each set of rows for a given interpretation contains the results of three different model variations, the top row of each interpretation contains the estimates using and independent model, the second row contains the results where the model was considered co-dependent and the final row contains the results of the model where uninformed priors were used for the Wisconsin clinical score. The Table also contains the β distributions used for both the Wisconsin clinical score and thoracic ultrasound in each model as well as the corresponding posterior estimates of sensitivity and specificity for thoracic ultrasound produced by each model.

<table>
<thead>
<tr>
<th>Interpretation</th>
<th>Test dependency</th>
<th>WCS- Prior Beta</th>
<th>WCS-Posterior</th>
<th>TUS- Prior Beta</th>
<th>TUS- Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough ≥ 2 and nasal/ocular discharge ≥ 1 (Calderon-Amor and Gallo (2020))</td>
<td>Independent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.336 (0.251, 0.424)</td>
<td>0.968 (0.959, 0.977)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>Co-dependent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.358 (0.25, 0.471)</td>
<td>0.967 (0.955, 0.977)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>(1, 1)</td>
<td>(1, 1)</td>
<td>0.31 (0.223, 0.396)</td>
<td>0.98 (0.97, 0.99)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td>≥6 (Lago et al. (2006))</td>
<td>Independent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.361 (0.269, 0.458)</td>
<td>0.977 (0.969, 0.984)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>Co-dependent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.399 (0.293, 0.513)</td>
<td>0.977 (0.968, 0.985)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td>≥5 or more scores ≥ 2 (McGuirk and Peek (2014))</td>
<td>Independent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.361 (0.269, 0.459)</td>
<td>0.952 (0.938, 0.965)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>Co-dependent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.577 (0.488, 0.667)</td>
<td>0.943 (0.927, 0.959)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>(1, 1)</td>
<td>(1, 1)</td>
<td>0.647 (0.557, 0.734)</td>
<td>0.945 (0.928, 0.961)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td>≥5 (Medrano-Galarza et al. (2019))</td>
<td>Independent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.567 (0.468, 0.668)</td>
<td>0.956 (0.945, 0.97)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>Co-dependent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.625 (0.52, 0.722)</td>
<td>0.958 (0.944, 0.971)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>(1, 1)</td>
<td>(1, 1)</td>
<td>0.551 (0.451, 0.655)</td>
<td>0.976 (0.96, 0.989)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td>2 or more scores ≥ 2 (Binversie et al. (2022))</td>
<td>Independent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.512 (0.426, 0.601)</td>
<td>0.95 (0.936, 0.962)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>Co-dependent</td>
<td>(9.8, 5.6) (138.8, 18.5)</td>
<td>0.575 (0.481, 0.667)</td>
<td>0.95 (0.935, 0.963)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
<tr>
<td></td>
<td>(1, 1)</td>
<td>(1, 1)</td>
<td>0.511 (0.423, 0.597)</td>
<td>0.95 (0.936, 0.963)</td>
<td>(27.6, 10.7) (89.3, 8.4)</td>
</tr>
</tbody>
</table>
the remaining calves fed a combination of milk replacer, whole saleable milk and waste milk (23.5%, n = 23).

The majority of farmers did not vaccinate their calves against BRD pathogens (61.2%, n = 60), 34 (34.7%) farmers were using BRD vaccines and 4 (4.1%) had unclear vaccination status.

The median volume of colostrum fed in the first feed was 3L (min. 1L, max. 4L). A bottle and teat was the most common method of colostrum feeding (59.2%, n = 58). Fifty-nine percent of farmers fed colostrum only from the calf’s dam (n = 58), 19.6% of farmers (n = 29) fed pooled colostrum from more than one cow and 11.2% of farmers had unclear colostrum management strategies (n = 11).

Calf descriptions

In total, 1,936 calves were examined using TUS and WCS and a subset of 628 were also examined using the CALIF. The most common breed examined was Holstein Friesian (55.3%, n = 1,070), followed by Jersey cross dairy (17.8%, n = 344), beef cross dairy (13.4%, n = 260) and other dairy breeds (11.7%, n = 227); 35 (1.8%) calves did not have a breed recorded. Seventy-nine percent of the calves examined were female (1,535), 19.8% were male (n = 383) and 18 (0.009%) calves did not have their sex recorded. The median age of calves examined was 34 d, (range 10 - 72 d). The mean estimated weight of the calves examined was 58.8 kg, (Standard deviation: 10.05 kg).

Scores - WCS, CALIF and TUS

The descriptions and frequency of each of the clinical signs can be found in Table 1 and 2. Overall, the mean rectal temperature was 39.0°C, (Standard deviation: 0.45°C). Figure 1 is an upset plot that represents the distribution of the intersection in case definition between
the various clinical scores. The median score of both the WCS and of the CALIF was 2 (range 0–10). The median TUS was 0 (range 0 – 5); the frequency of each score is listed in Table 3.

The apparent calf-level prevalence of BRD in this population was 10.3% according to TUS. However it was 11.2% according to the McGuirk and Peek (2014) interpretation, 9.6% according to the Binversie et al. (2020) interpretation, 9.2% according to the Medrano-Galarza et al. (2018) interpretation, 4.9% according to the Calderon-Amor and Gallo (2020) interpretation, and 4.1% according to the Lago et al. (2006) interpretation of the WCS.

Various interpretations of the Wisconsin score

Bayesian latent class analyses. Convergence was assessed visually using autocorrelation trace plots and found to be adequate for all models. The potential scale reduction factor (PSRF) values for Se and Sp of all WCS interpretations and TUS were <1.0015 indicating adequate convergence. The effective sample size ranged from 2,928 to 13,407.

The results of each of the Bayesian latent class models for part one can be found in Table 4. Across the 5 WCS interpretations the median Se ranged from 0.310 to 0.647 and the median Sp ranged from 0.943 to 0.980. The WCS interpretation proposed by McGuirk and Peek (2014) had the highest Se (Median: 0.647, 95% BCI (0.557, 0.734). The interpretation proposed by Calderon-Amor and Gallo (2020) had the highest Sp (Median: 0.98, 95% BCI (0.97, 0.99). The Se and Sp estimates for TUS varied depending on the interpretations. The median TUS Se estimates varied from 0.541 to 0.790 and Sp estimates from 0.950 to 0.977.

Sensitivity analysis. Overall, the models for each of the interpretations were robust. Uninformative priors only led to a max 0.026 change in the Se and a change of 0.025 in Sp of WCS interpretations while maintaining similar 95% Bayesian confidence intervals (BCI). Modeling with a co-dependency to TUS resulted in an increase in Se in all cases with the maximum increase of 0.07. In the case of Sp, co-dependency resulted in a maximum change of 0.025.

Estimate of the Se and Sp of CALIF

Convergence of the 3 test (WCS, CALIF, TUS) model was assessed using autocorrelation plots. Some autocorrelation was detected so the model was thinned by 10. The PSRF values for the Se and Sp of all tests were <1.0017 and the effective sample size for each variable ranged from 1,931 to 20,000. The estimates of Se and Sp for WCS, CALIF and TUS can be found in Table 5. The WCS had higher Se than the CALIF in models using informative and non-informative priors. The CALIF had higher Sp than WCS in the model using non-informative priors.

Table 6 contains a comparison of the Se and Sp from the models using TUS and either WCS or CALIF in separate HBLCM models. The Se of CALIF improved (0.529 vs 0.594) while the Sp remained stable (0.903 vs 0.921). However, when WCS was used instead of CALIF in this 2 tests model (WCS, TUS), it had a higher Se compared with the CALIF model (0.668 vs 0.594) and similar Sp (0.935 vs 0.921).

Estimates of PPV, NPV and accuracy

Tables 7 and 8 detail the PPV, NPV and accuracy for various prevalence levels of BRD in a given population
using the Se and Sp estimated from the primary models. Lago et al. (2006) had the highest performing positive predictive values across all of the prevalence levels of BRD. McGuirk and Peek (2014) had the highest negative predictive values across all BRD prevalence levels. Lago et al. (2006) had the highest test accuracy at 5% BRD prevalence, but as prevalence increased, Medrano-Galarza et al. (2018) out performed Lago et al. (2006).

**DISCUSSION**

We hypothesized that the performance of WSC and CALIF may differ in calves from pasture-based dairy herds in a temperate climatic region when compared with the regions where they were developed. Second we hypothesized that the alteration of CRS thresholds would result in altered Se and Sp of those tests. Our objectives were therefore to investigate the diagnostic performance of WSC and CALIF in Irish calves with a second objective of using the data to investigate out alteration of the WCS might affect its performance.

This is the largest study (n = 1936) conducted to characterize the Se and Sp of calf clinical scoring systems. Previous publications that characterize the Se and Sp of a clinical scoring system include Decaris et al. (2022) (n = 428), Bermann et al. (2022) (n = 800), Bermann et al. (2019) (n = 510), Buczinski et al. (2018) (n = 608), Love et al. (2016) (n = 536), and Buczinski et al. (2015) (n = 191). The WCS system is the most commonly used calf clinical scoring system internationally (Johnson et al., 2017; Cramer and Ollivett, 2019; van Leenen et al., 2020). Many WCS interpretations have been described in the literature (Lago et al., 2006; McGuirk and Peek, 2014; Calderón-Amor and Gallo, 2020). Decaris et al. (2022) chose to model 3 interpretations of the WCS; in this work we modeled those interpretations and identified other interpretations which had not been evaluated. The results generated here indicate that alterations of the WCS can lead to substantial variation in Se and Sp of the test. Due to the variation in test performance found in this work, the authors would suggest that altering WCS interpretation should be avoided. Using a variation of WCS may lead to inaccurate effect size estimates in work investigating BRD risk factors or may lead to inaccurate prevalence estimates. Test performance variation may also be compounded by differences between operators performing tests between studies and differences in BRD presentation due to the etiological agents present in a given population. Thus alteration of WCS interpretation is adding even more unnecessary uncertainty to the results those publications. It is clear from this work that publications that use alterations of WCS cannot be compared with an assumption of similar test performance. In future, publications that alter the interpretation of the WCS should be required to outline the justification for the alteration.

**Table 5.** Sensitivity and specificity estimates (median and 95% BCI) for the California clinical score, Wisconsin clinical score and thoracic ultrasound from two hierarchical Bayesian latent class models using all three tests. The results of the first model using informed priors for both clinical scores can be found in the top row for each diagnostic test, the results of a second model using uninformed priors can be found in the row below.

<table>
<thead>
<tr>
<th>Diagnostic Test</th>
<th>Prior β</th>
<th>Posterior β</th>
<th>Se</th>
<th>Sp</th>
<th>Se</th>
<th>Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALIF</td>
<td>(22.6, 13.5)</td>
<td>(176.0, 48.0)</td>
<td>0.529 (0.403, 0.651)</td>
<td>0.903 (0.883, 0.922)</td>
<td>0.516 (0.356, 0.705)</td>
<td>0.986 (0.957, 1)</td>
</tr>
<tr>
<td>WCS</td>
<td>(9.8, 5.6)</td>
<td>(138.8, 18.5)</td>
<td>0.733 (0.609, 0.85)</td>
<td>0.871 (0.845, 0.895)</td>
<td>0.679 (0.523, 0.832)</td>
<td>0.948 (0.912, 0.974)</td>
</tr>
<tr>
<td>TUS</td>
<td>(27.6, 10.7)</td>
<td>(89.3, 8.4)</td>
<td>0.638 (0.533, 0.742)</td>
<td>0.948 (0.93, 0.967)</td>
<td>0.632 (0.53, 0.734)</td>
<td>0.948 (0.927, 0.967)</td>
</tr>
</tbody>
</table>

**Table 6.** Posterior sensitivity and specificity estimates for clinical scores and thoracic ultrasound from four hierarchical Bayesian latent class models built only using the subset of data for which California clinical scores were built. Two models used only California clinical score results (one with informed priors and one with uninformed priors) and thoracic ultrasound results. Two models (one with informed priors and one with uninformed priors) only used the Wisconsin clinical score results and thoracic ultrasound.

<table>
<thead>
<tr>
<th>Prior β</th>
<th>Se</th>
<th>Sp</th>
<th>TUS- Se</th>
<th>TUS- Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALIF</td>
<td>0.594 (0.466, 0.719)</td>
<td>0.921 (0.901, 0.939)</td>
<td>0.743 (0.631, 0.848)</td>
<td>0.945 (0.922, 0.965)</td>
</tr>
<tr>
<td>WCS</td>
<td>0.587 (0.417, 0.775)</td>
<td>0.988 (0.973, 0.999)</td>
<td>0.673 (0.561, 0.787)</td>
<td>0.948 (0.926, 0.968)</td>
</tr>
<tr>
<td></td>
<td>0.664 (0.532, 0.784)</td>
<td>0.935 (0.914, 0.954)</td>
<td>0.706 (0.593, 0.813)</td>
<td>0.954 (0.934, 0.972)</td>
</tr>
</tbody>
</table>
Overall, the performance of all the models examining the various interpretations of WCS appeared to be relatively robust to changes; modeling with uninformative priors only led to small decreases in the median Se and a stable Sp. When a co-dependency was modeled between TUS and WCS, there was a general trend for a small increase in the Se of the WCS interpretations. Current evidence suggests that of all the clinical signs assessed as part of the WCS, only cough is significantly associated with lung consolidation (Lowie et al., 2022). For that reason the tests were considered independent in the primary models. It is noteworthy that the Se and Sp of TUS appeared to vary depending on the interpretation of WCS. This may suggest that the latent class being identified in each of the models varied slightly.

In the second part of this work when the 3 tests were modeled together, the model appeared less robust to uninformative priors. The 95% BCI for Se and Sp widened for both WCS and CALIF making the authors less confident in the results. When the WCS and CALIF were modeled separately on the subset of calves used in part 2, the Se of CALIF increased. However, when modeled individually on the same data set the Se estimates of WCS remained similar with wider confidence intervals suggesting that the model using 3 tests together was stronger with more reliable results.

The Se and Sp of the WCS vary in the literature. Decaris et al. (2022) estimated that when 2 component scores had a score of ≥2, the Se of WCS was 0.75 (0.608, 0.882) and the Sp was 0.822 (0.770, 0.873). Using the same interpretation we found a lower Se 0.512 (0.4266, 0.601) and a higher Sp 0.950 (0.936, 0.962). The 95% BCI did not overlap for either estimate of Se or Sp. There may be multiple reasons for this; model structure, differences in BRD prevalence or a difference in the TUS cut off. Buczinski et al., (2015) used a score of ≥ 5 to define a BRD positive and estimated the Se and Sp of WCS to be 0.624 (0.479–0.758) and 0.741 (0.649–0.828), respectively. Berman et al. (2019) using the same interpretation found Se and Sp values of 0.69 (0.40, 0.97) and 0.95 (0.92, 0.97), respectively. Both of these are similar to the 0.567 (0.468, 0.668) of the present study, though the latter had narrower confidence intervals attributed to the larger population. The Sp of the present study was similar to that of Berman et al. (2019) but higher than that of Buczinski et al. (2015). This may be due to the TUS score of ≥ 3 being used as a positive result in the present study, which is a higher cut off than the ≥1cm cut off used by Buczinski et al. (2015) and similar to the ≥3cm used by Berman et al. (2019).

There appears to be large differences in the Se of the various interpretations of the WCS. The poorest performing interpretation in regard to Se was in the study by Calderón-Amor and Gallo (2020) which only used the

### Table 7

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.356 (0.452, 0.452)</td>
<td>0.348 (0.415, 0.406)</td>
</tr>
<tr>
<td>0.10</td>
<td>0.538 (0.636, 0.636)</td>
<td>0.529 (0.616, 0.616)</td>
</tr>
<tr>
<td>0.15</td>
<td>0.649 (0.735, 0.735)</td>
<td>0.641 (0.724, 0.724)</td>
</tr>
<tr>
<td>0.20</td>
<td>0.740 (0.827, 0.827)</td>
<td>0.731 (0.814, 0.814)</td>
</tr>
<tr>
<td>0.25</td>
<td>0.831 (0.912, 0.912)</td>
<td>0.824 (0.905, 0.905)</td>
</tr>
<tr>
<td>0.30</td>
<td>0.921 (0.994, 0.994)</td>
<td>0.914 (0.986, 0.986)</td>
</tr>
</tbody>
</table>
results of cough and nasal or ocular discharge to define positive results. The best performing interpretation with regard to Se was McGuirk and Peek (2014) which defined a case as a score ≥5 or 2 or more component scores ≥2. When the highest and lowest median Se was compared, there was a difference in Se of 0.241 (0.336 vs 0.577). Three of the interpretations had relatively similar results in regard to Se (McGuirk and Peek, 2014; Medrano-Galarza et al., 2018; Binversie et al., 2020) and of those McGuirk and Peek (2014) had the highest sensitivity but lowest specificity. When the positive predictive value of these tests were calculated using various prevalence levels, it was found that the cut off used by Lago et al. (2006) had the highest PPV at a low prevalence (0.05). This could be attributed to the higher specificity of the test due to the high cut off used to define a BRD positive calf (≥6). The 3 interpretations that were most closely related (Lago et al., 2006; Medrano-Galarza et al., 2018; Binversie et al., 2020) had similar Se and Sp. However, the interpretation that used ≥5 as its cut off marginally but consistently outperformed the other 2 interpretations in PPV and accuracy (Tables 7 and 8). It is clear that altering the cut off chosen can have a major effect on the test characteristics and may serve to increase sensitivity or specificity. This may be useful if the test is to be used in a screening capacity or to be used in a low BRD prevalence population.

This is only the second time that the California scoring system has been evaluated using a Bayesian technique to assess Se and Sp (Decaris et al., 2022). It has not been assessed before in a temperate climate, such as Ireland. The California scoring system is simpler to implement than other scoring systems (Love et al., 2016) and so may present an opportunity for use on Irish farms if it was found to be accurate. As the environment in which the CALIF weightings were developed is quite different to that of Irish dairy farms in Spring (arid desert vs temperate climate), there may have been differences in the presentation of BRD between the 2 settings. Changes in climate may affect the presentation of certain clinical signs, such as rectal temperature. In a hot arid desert environment the rectal temperature of calves is likely to be higher (Decaris et al., 2022), than those found in calves in a temperate environment. Thus the weighting given to this clinical sign may not be appropriate for an Irish setting.

As seen in Table 5 and 6, the CALIF score performed worse than the WCS score, having lower Se and Sp. This is in line with other publications which have compared the 2 scores. Love et al. (2016) found a screening Se of 0.468 and a Sp of 0.874, which was similar to the Se and Sp found in the present study (0.529, 0.903). However Love et al. (2016) did not account for the lack of a gold standard ante mortem BRD diagnostic test. Decaris et al. (2022) estimated that the Se of CALIF was 0.634 and the Sp was 0.788; a higher Se estimate than was found in the present study and a lower Sp estimate. Overall the Se and Sp estimates for CALIF were similar to those already published in other climates, but lower than that of the WCS and in the authors’ opinions, too low to be implemented on typical Irish dairy farms that have a relatively low prevalence of BRD (Donlon et al., 2023).

The inclusion of respiratory effort in CALIF is one of its major differentiating factors from WCS. In this work, the prevalence of increased respiratory effort was relatively low. There may be 2 reasons for this. First, the calves in the present study were all kept in group pens. The process of handling and examining calves inside of the pen was likely to have excited them, making it difficult to distinguish excitement from genuine labored breathing. Second, the calves were under a different type of thermal stress to that seen in California, where heat may exacerbate breathing patterns in those affected by BRD to a greater extent than that seen in Ireland. Increased respiratory effort is still a valuable clinical sign of BRD: of the the 7 calves that were given a score of 2 for respiratory effort, 3 were positive on TUS. Thus, even if breathing difficulty is an infrequently noted sign, it is a strong indicator of BRD and should be included as a highly relevant clinical sign. In terms of practical implementation in group penned calves breathing can also be assessed from a distance and therefore could be assessed more easily than other signs such as ocular or nasal discharge. However as highlighted by Berman et al. (2021) it is a clinical sign with one of the lowest inter-
rater agreement and thus its implementation as a screening tool would require training to improve consistency.

Inter-rater variability was not explored in this work, however it has been established as an issue for scoring systems (Buczinski et al., 2016; Berman et al., 2021). Having a single operator would have led to more consistency in scoring. However, the use of a single operator would have reduced the applicability of the results to a more general population, a point acknowledged by Decaris et al. (2022). Berman et al. (2021) has previously found that producers were less consistent in scoring BRD clinical signs than veterinarians apart from ear droop/head tilt. In the present study, the scorers were all veterinarians or veterinary students and so the results are not truly reflective of the Se and Sp that might be achieved by farmers. However, the results obtained are a useful as a benchmark of the Se and Sp achievable in a research setting where the scorers are closely supervised. Previous publications investigating the Se and Sp of CRS have used a small number of scorers (1 – 4 scorers) (Decaris et al., 2022; Buczinski et al., 2018, Berman et al., 2019). The estimates from those works might only reflect the performance of those scores and not CRS in general when applied in a research setting. However, the estimates generated here are similar to those already published with relatively narrow confidence intervals, suggesting that these results are externally valid.

Others have used numerous tests (TUS, auscultation, haptoglobin) in a population to allow for the model to be identifiable. Haptoglobin has been used to determine active vs non-active forms of pneumonia (Berman et al., 2019; Decaris et al., 2022). Auscultation is a common method of BRD diagnosis used by veterinarians (Pardon et al., 2019). We chose not to use auscultation in this work because it has been shown to have poor diagnostic performance and large variability between raters (Buczinski et al., 2013, Pardon et al., 2019). Haptoglobin and auscultation were not used in this work; instead to increase the degrees of freedom we chose to visit numerous farms and used a hierarchical model, making these estimates more generalisable.

Another limitation of this work was an absence of information relating to the infectious agents present on the farms visited as part of this study. This may have provided additional useful information and allowed us to investigate if particular etiological agents caused particular presentations that might affect the overall clinical score. In future work reweighting or investigating clinical scores the influence of infectious agents should be investigated.

As this was a cross sectional study, it was not possible to analyze the temporal association between clinical signs and lung lesions. A potential limitation of this study design is that the clinical signs and lesions used to diagnose BRD via TUS and CRS may appear at distinct times during the disease process. This may lead to misclassification of calves as demonstrated by Cuevas-Gómez et al. (2021) who found that the majority of dairy beef calves that developed clinical BRD had lung lesions detectable by TUS a median of 10.5 d before they were considered CRS cases. A longitudinal study design would allow us to better understand what proportion of calves with lung lesions at the time of first examination go onto become CRS cases and vice versa. As previously mentioned TUS lung lesions which represent damage from a resolved BRD event may cause CRS to appear to be less sensitive however a longitudinal study where screening of calves for lung lesions before inclusion would help to avoid such a possibility.

The data that were collected in this study could be used for a reweighting of the clinical signs, similar to the work of Buczinski et al. (2018), to produce a clinical score that would most accurately characterize Irish dairy calves’ BRD status. Such a reweighted scoring system would be ideal for dissemination to Irish dairy farmers to help improve health monitoring and accurate and timely diagnosis of BRD. However, this task was outside of the scope of this study.

CONCLUSIONS

The interpretations of a clinical scoring system needs to be well justified in the context of the objectives of a study and in the context of the pre-existing knowledge of BRD prevalence in that population. The Wisconsin score appears to be more sensitive and specific than the California score in Irish dairy calves, however, future work reweighting the score for Irish calves may yield a test with similar attributes that is easier to implement.

Conflict of Interest statement

John Denis Donlon was paid an honorarium by MSD and Boehringer Ingelheim for teaching workshops on BRD and thoracic ultrasound.

ACKNOWLEDGMENTS

This research was funded by Dairy Research Ireland.

REFERENCES


