



J. Dairy Sci. TBC

<https://doi.org/10.3168/jds.2023-23554>

© TBC, The Authors. Published by Elsevier Inc. and FASS Inc. on behalf of the American Dairy Science Association®.
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Comparison and inter-observer reliability between a Visual Analogue Scale and the Wisconsin Calf Health Scoring Chart for detection of respiratory disease in dairy calves

Henrik H. Møller,^{1*} Mogens A. Krogh,² Mette B. Petersen,¹ Liza R. Nielsen,³ and Nynne Capion¹

¹Department of Veterinary Clinical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Taastrup, Denmark

²Department of Animal and Veterinary Sciences, Aarhus University, Foulum, Denmark

³Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg, Denmark

ABSTRACT

Respiratory disease is an ongoing challenge for calves in the dairy sector with a relatively high prevalence and impact on welfare and economics. Applying scoring protocols for detecting respiratory disease requires that they are easily implemented, consistent between observers and fast to use in daily management. This study was conducted in one Danish dairy farm from September 2020 through January 2021. The study included 126 heifer calves enrolled in the age of 17 to 24 d. All calves were observed every second day for a period of 46 d. At each visit all calves were scored with a new Visual Analog Scale (VAS) and the Wisconsin Calf Health Scoring Chart (WCHSC). We calculated agreement between the 2 scoring systems based on conditional probability to score higher or lower than a cut-off in the VAS compared with a specified cut-off in WCHSC used as reference test. A generalized mixed effects regression model was performed to develop a model that could estimate the prevalence of respiratory disease and the overall agreement between the 2 scoring systems. The overall agreement between the VAS and WCHSC was 89.6%. The second part of the study assessed inter-observer reliability between 2 experienced observers and between an experienced observer and veterinary students. The inter-observer reliability was calculated by intra-class correlation coefficient and was 0.58 between experienced observers and was 0.34 between an experienced observer and veterinary students indicating a moderate to poor reliability between the observers. It was possible to use VAS as an alternative clinical scoring method, which primarily focuses on the general condition of the individual calf rather than specific categories of clinical signs. Our study set up

lacked a comparison to other diagnostic tools i.e., thoracic ultrasound to confirm the findings which should be considered in future studies when exploring VAS as a screening tool for detection of respiratory disease in dairy calves.

Keywords: VAS, Wisconsin Calf Health Scoring Chart, Respiratory disease, Inter-observer reliability

INTRODUCTION

Respiratory disease is an ongoing challenge in dairy calves worldwide and affects more than 18% of pre-weaned dairy replacement heifers with estimated economic losses ranging between 250 and 280 USD per heifer (Cummings et al., 2022). Respiratory disease accounts for up to 61% of antimicrobial treatments in calves in the US (Zhang et al., 2022) which makes it a target disease syndrome with a potential to reduce antimicrobial use. Detection and quantification of respiratory disease using standardised methods are needed so only calves in need of it, are given treatment.

According to McGuirk and Peek (2014) detection of respiratory disease is often based on pattern recognition performed by farm staff with different prior experiences. Poor standardization of the diagnostic methods can lead to imprecise estimates of disease occurrence on farm level and risk of treatment failures at the individual level. Failure to correctly diagnose respiratory disease can impair animal welfare and lead to unjustified use of antimicrobials.

The Wisconsin Calf Health Scoring Chart (WCHSC) (McGuirk & Peek, 2014) is used in research of respiratory disease in calves (Buczinski et al., 2014; Mahendran et al., 2017; Lowie et al., 2022). The WCHSC consists of clinical observations and grades coughing, nasal discharge, eye and ear constitution and rectal temperature, which is then aggregated on the individual calf. The WCHSC cut-off at ≥ 5 is developed based on paired results with broncho-alveolar fluid cytology and validated by culture in the work by McGuirk (2008,

Received March 31, 2023.

Accepted August 15, 2023.

* Correspondence: Henrik Hjul Møller, Agrovej 8, DK-2630 Taastrup, Denmark, +45 61119625, henrik.hjul.moller@sund.ku.dk

unpublished data), but no diagnostic performance was provided in that study. Recent studies calculated sensitivity (Se) of 30% and specificity (Sp) of 99% (Berman et al., 2022), and a Se of 78% and a Sp of 82% (Decaris et al., 2022) by Bayesian techniques based on clinical examination, thoracic ultrasonography (TUS) and serum haptoglobin concentrations in veal calf respiratory scoring systems.

McGuick & Peek (2014) estimated the additional labor associated with implementing the use of the WCHSC in full scale in the range of 3–4 h a day for each 100 calves. This level of time consumption can be a major barrier for implementation in commercial dairy farms. Therefore, there is a need to look for and evaluate other protocols that focus on other clinical aspects that can more easily be observed and aggregated simultaneously.

Several other relevant behavioral and clinical signs exist which are not directly incorporated in the WCHSC. Calves are less active and consume less milk at presence of respiratory disease (Swartz et al., 2017), and behavioral changes such as abnormal posture when standing or lying, individual isolation, lethargy, reduced grooming and feeding time, fewer visits to automated feeding stations and fewer social interactions are found to appear earlier than more noticeable signs of respiratory disease like nasal- and eye discharge (Cramer et al., 2016; Hixson et al., 2018; Duthie et al., 2021). Love et al. (2016) addressed the possibility to include screening protocols to detect calves that need subsequent examination to make a precise diagnosis of respiratory disease with other tools. This could potentially lead to a more widespread on-farm practical tool in the management and handling of respiratory disease cases.

‘Visual Analogue Scale (VAS) score’ is a method developed to assess the severity of pain using several behavioral signs, including alertness. It is defined as a continuous scale between 0 and 100 mm, where 0 is no pain and 100 mm is severe pain (Martin et al., 2020). VAS has been reported as a valuable tool to assess pain in experimentally *M. haemolytica*-infected calves (Martin et al., 2022).

A requirement for a feasible health scoring chart is good inter-rater reliability, and for practical implementation, preferably with limited prior expertise or training. Buczinski et al. (2016) reported a slight to fair agreement with kappa ranging from 0.06 to 0.35 at different cut-offs between relatively unexperienced observers using the WCHSC for detection of respiratory disease. Berman et al. (2020) reported WCHSC agreement by kappa ranging between 0.2 and 0.8 based on different professional backgrounds of the observers and by agreement between the individual clinical signs that appear in cases of respiratory disease. Inter-rater reliability of the VAS score has been described

for other conditions in other animal species including elbow osteoarthritis in dogs with an intra-class correlation coefficient (ICC) > 0.9 (Aulakh et al., 2020), and in qualitative behavior assessment of housed sheep (Diaz-Lundahl et al., 2019) which reported a Kendall coefficient of 0.45 – 0.91.

We hypothesized that it was possible to use VAS as an alternative system to the WCHSC for diagnosing signs of respiratory disease in dairy calves. The first objective was to determine the level of agreement of the VAS compared with the WCHSC as reference test to detect signs of respiratory disease. The second objective was to estimate inter-observer reliability between 2 trained observers in calf health; and between an experienced observer in calf health and a group of veterinary students using the VAS and the WCHSC to detect signs of respiratory disease.

MATERIALS AND METHODS

The study protocol was reviewed and accepted by the Local Ethical Committee at Department of Veterinary Clinical Sciences, University of Copenhagen (project no. 2020–023). A written consent from the farm owner to conduct the study was obtained before the start of the study. The veterinary students were informed that the data collected would be included in a research project.

Study population

The study was performed in a commercial Danish dairy farm with approximately 650 dairy cows. The breeds were Danish Holstein (90%) and Red Danish Holstein (10%). Only heifer calves were enrolled in the study because bull calves were moved from the farm for veal production around 2 weeks of age. The calves were housed in individual outdoor hutches after birth and subsequently at 35–40 d of age moved either to outdoor group hutches or indoor group pens with pen sizes of 6 to 14 calves. Within a week after moving to group housing all calves were dehorned and received a *Trichostrongylus axei* (ringworm) vaccination. The calves were not vaccinated against respiratory disease-associated pathogens.

Study design and data collection

The study was carried out from September 2020 through January 2021 as an observational study following a cohort of dairy calves. Visits to the farm took place every second day leading to 76 visits in total during the study period. The first author of this paper carried out 71 of the visits and the last author 5. The

Table 1: Overview of the number of calves, observers, and observations, used for the analysis in this study

Study	Study period	No. calves	No. observers	No. observations
WCHSC-VAS agreement	Sep. 2020 – Jan. 2021	126	2	3011
Inter-observer reliability – experienced observers	Nov. 2020	53	2	106
Inter-observer reliability, veterinary students	Nov. 2020	47	19	224

calves were enrolled between 17 and 24 d of age and observed by the researchers for development of signs of respiratory disease every second day for 46 consecutive days (6 weeks). Every calf was scored 24 times in total using both the VAS and the WCHSC. Farm visits were done between 9 a.m. and 4 p.m. between feedings to limit disturbances to the daily routines in the farm. All calves were first scored using the VAS system and then all calves were scored using the WCHSC (see below for detailed description).

New calves were enrolled once weekly during the study until 6 weeks before the end of the study period. The 2 observers were veterinarians with substantial training in clinical research. Data for inter-observer reliability evaluation between first and last author of this paper was collected on the 13th of November 2020.

Data from 4 d in November 2020 (17th, 19th, 23rd and 25th) were used for inter-observer reliability between the first author and 19 Danish veterinary students. The veterinary students were introduced to the VAS and the WCHSC by 10-min verbal instructions in the calf rearing area in the farm and all veterinary students got a handout template based on the work done by McGuirk & Peek (2014). Table 1 provides an overview of the number of enrolled calves, observers, and observations.

Clinical scoring of signs of respiratory disease

VAS score procedure. The VAS score given to each calf at each farm visit was a score that corresponded to the general clinical appearance in the moment the

calf was observed. The score was given by observing the calf from a short distance (~1–2 m) without touching or being in physical contact with the calf and with no other interaction and intervention than the presence of the observer.

The VAS score was noted in the range between 0 and 10 cm with the possibility to use 0.1 increments and was given within approximately 10 s of observation. Based on the clinical assessment the VAS score is decided with support from the 4 anchor points described in Table 2. If the calf lacks some of the described clinical findings but have others within the scale anchor points then the calf will remain in the particular part of the VAS scale but might be placed in the lower or higher end of the scale depending on the severity of the clinical findings. The description of the VAS anchor points along the scale used in this study can be seen in Table 2 and is defined as points anchored at 0, 2.5, 5, 7.5 and 10 cm at the continuous scale.

Wisconsin Calf Health Scoring Chart procedure. The WCHSC was used at each visit according to the method developed and described by McGuirk & Peek (2014). The clinical scores were given in the following order: Nasal discharge score, eye score, ear score and rectal temperature. Coughing was assessed as either spontaneous or induced by manual laryngeal/pharyngeal provocation during the measurement of rectal temperature. The fecal score was not noted in this study. Each clinical score was noted as an ordinal number between 0 and 3. A calf that had a total WCHSC score of 3 could e.g., be a calf with unilateral cloudy nasal discharge, a single induced cough and ear

Table 2: Definitions of VAS score anchor points along the numeric axis between 0 and 10 cm and grouped into the following 4 intervals on the numeric scale from 0 to 10 cm

VAS score interval	Definition
0.0 ; 2.4	The calf is interested in its surroundings, is very bright and alert, has clear eyes, the ears are pointing in an upright position and the head is held above spine level if the calf is standing
2.5 ; 4.9	The calf is less interested in the surroundings, slightly depressed, less bright, alert and responsive, has less clear eyes, the calf uses longer time to get up, the hair layer is not intact, and/or drooping ears
5.0 ; 7.4	A depressed calf with obvious clinical signs of disease including nasal- or eye discharge, coughing, and drooping ears and/or abdominal respiration in combination with increased frequency and head position below spine level
7.5 ; 10.0	A significant depressed or apathetic calf, unwilling to stand up, or lying in lateral recumbency and appearing very ill or terminally ill often assessed upon the signs obtained in the lower score intervals

flick or head shake. Scores ≥ 5 could be a calf with bilateral, cloudy, or excessive mucus discharge, moderate bilateral amount of ocular discharge and single induced cough even if the calf were jumping around and seemed healthy. Using the WCHSC protocol, it is recommended that a total score ≥ 5 or that have 2 or more clinical parameters with a score of 2 or 3 are considered to have respiratory disease (McGuirk & Peek, 2014).

Statistical analysis

Data management and analyses were conducted in the statistical platform R version 4.2.3 (R Core Team, 2022), and RStudio version 2023.03.0+386 (RStudio Team, 2022). All variables were considered statistically significant at a level of $P < 0.05$.

The WCHSC score was divided into 2 groups to describe high (≥ 5) vs. low (< 5) probability that the calf had respiratory disease as described by McGuirk and Peek (2014). The level of agreement was defined as the estimated probability of the VAS correctly classifying the low vs. high probability of respiratory disease based on the WCHSC. Since observations of the VAS and WCHSC were repeated on the individual calf, this was adjusted for in a generalized mixed effects regression model, with calf as a random effect. Initially, an empty model with the dichotomized WCHSC scores as outcome variable and the random intercept was created to provide an unadjusted prevalence estimate of the calves having high probability of respiratory disease based on the WCHSC. Subsequently, the generalized mixed model analysis was repeated multiple times with a dichotomized VAS score (dVAS) as explanatory variable for each of the possible cut-off of the VAS score. Based on these models the conditional probabilities $p(\text{WCHSC} \geq 5 \mid \text{dVAS} = 1)$ and $p(\text{WCHSC} < 5 \mid \text{dVAS} = 0)$ were predicted for each cut-off of the VAS score. Based upon the prevalence estimate and the 2 conditional probabilities the agreement could be estimated solving 2 equations with 2 unknowns. The level of agreement was then estimated for all possible cut-off values of VAS score and the cut-off with the best possible agreement was identified.

The autocorrelation functions of the residuals were used to evaluate if the calf random effect was sufficiently accounting for correlation in the repeated measurements. The parameter estimates of the calf random effects were also evaluated between the different models to assess model stability. Model control was done by assessing simulated scaled residuals from the DHARMA-package (Hartig, 2022). The generalized mixed effects regression model was analyzed using the glmer-function from the lme4-package (Bates et al., 2015) and the pre-

diction of the conditional probabilities and prevalence was done with the emmeans-package (Lenth, 2022).

To compare the difference in mean values of VAS scored by both the experienced observers and by the author and veterinary students, a 2-sample paired t -test was performed in R.

The ICC was used to estimate the inter-observer reliability in use of VAS to clinical scoring of the calves. The ICC was analyzed and interpreted based on the work and descriptions by Koo & Li (2016). ICC estimates and their 95% confident intervals were calculated in the irr-package (Gamer et al., 2019) and the psych-package (Revelle, 2022) by a 2-way random effects model based on a specific rating by different observers and absolute agreement, because we wanted to assess the agreement in VAS scores from specific calves scored by a fixed set of observers. The calculations were based on a single unit level meaning that the VAS score was compared equally between all observers participating in the reliability study.

RESULTS

Comparison of VAS and WCHSC

A total of 126 calves were enrolled in the study. This resulted in a total of 3011 calf observations by 24 observations per calf throughout the study period. Details of the results can be seen in Table 3. Based on the WCHSC there were 380 observations with a high probability of respiratory disease and 2631 observations with a low probability of respiratory disease. In the low probability group, the VAS score had a mean of 0.9 (median of 0.7) and ranged between 0.1 and 6.4 whereas the mean VAS score was 1.8 (median of 1.3) and ranged between 0.2 and 7.5 in the high probability group. The means of the VAS score were significantly different in the high and low probability groups and calculated by a 2-sided paired t -test ($P < 0.0001$). The prevalence of high probability of respiratory disease across all observations, but adjusted for the repeated observations on the same calves in the generalized mixed effects regression model, was 0.125 (95% CI: 0.106; 0.146). The highest accuracy between a dichotomized VAS score and the high vs. low probability of respiratory disease based on the dichotomized WCHSC scores was 0.896 at a VAS score cut-off of ≥ 2.8 . At this cut-off the conditional probability of $p(\text{high VAS} \mid \text{WCHSC} \geq 5)$ and $p(\text{low VAS} \mid \text{WCHSC} < 5)$ were 0.267 and 0.984.

Inter-observer reliability in using the VAS score

Two experienced observers, A and B, observed the same 53 calves on the same day. Details can be seen in

Table 3. Observer A had a mean of VAS score of 1.1 which was equal to Observer B's mean VAS score of 1.1. The mean difference in VAS score between Observer A and Observer B was non-significant ($t_{paired} = -0.54$, $P = 0.59$). The ICC between Observer A and B was 0.58 (95% CI: 0.37; 0.73) which indicated poor to moderate reliability between observers according to Liljequist et al. (2019).

Observer A and 19 veterinary students observed the same 47 calves on the same day, which gave a total of 224 calf observations. Observer A had VAS scores ranging between 0.2 and 2.7 with a mean of 0.8 (median = 0.7). The veterinary student group had VAS scores ranging from 0.0 to 6.8 with a mean of 1.8 (median = 1.7). The veterinary students had significant higher VAS scores than Observer A ($t_{paired} = 13.04$, $P < 0.0001$). The ICC between observer A and the veterinary students was 0.34 (95% CI: 0.26; 0.44) which indicated a poor reliability between the observers (Liljequist et al., 2019).

DISCUSSION

We investigated the use of VAS to evaluate signs of respiratory disease in dairy calves by comparing it to the WCHSC as reference test. The WCHSC is a well-established clinical scoring system that has been used for quantifying clinical signs of respiratory disease in calves.

The first part of the study focused on comparing 2 methods that based the assessment of the health or disease status of a calf by 2 different approaches. We found a good overall agreement between the 2 methods, which were estimated to agree approximately 90% of the time, when dichotomized into high and low disease probability, at a cut-off of ≥ 2.8 in VAS and ≥ 5 in WCHSC, respectively.

The WCHSC is based on specific clinical signs that are considered obvious to perceive when detecting respiratory disease in calves, but it takes longer time to perform, requires handling and to some extent restraining of the calves. The VAS system requires observation of the calves from a distance but does not assess clinical signs the same way, as WCHSC does. VAS assesses the disease status of the calf by behavioral changes and changed appearance and focus less on specific clinical signs. If VAS was considered as a clinical screening tool in this study, it might have been appropriate to pay more attention to the specific clinical signs and assess whether it was respiratory disease or other organ system diseases involved as the cause of a higher VAS score. A concept of doing follow-up diagnostics to confirm the clinical findings as serial testing could be further analyzed before and after the data collection took place. This could have improved the quality of the VAS

Table 3: Distributions of VAS scores for each of two outcomes of the reference test WCHSC scores dichotomized into low/high probability of respiratory disease being present in the calf at cut-off of ≥ 5 . Distributions of VAS scores are described for three different parts of the study: 1) Comparison study of VAS and WCHSC including all calves observed primarily by the first author, 2) Inter-observer reliability between two experienced observers, and 3) Inter-observer reliability between observer A and a group of veterinary students. Distributions are given by number of observations, minimum, 25% quartile, median, 75% quartile, maximum and mean distribution of VAS scores by each category of dichotomized WCHSC scores

Parameter	Category	No. observations	Min	25% quartile	Median	Mean	75% quartile	Max
Comparison study	WCHSC score < 5.0	2631	0.1	0.5	0.7	0.9	1.0	6.4
	WCHSC score ≥ 5.0	380	0.2	0.8	1.3	1.8	2.6	7.5
Agreement experienced observers	First author	53	0.3	0.3	0.8	1.1	2.8	5.3
	Last author	53	0.2	0.3	1.0	1.1	2.7	2.8
Agreement veterinary students	Observer A	224	0.2	0.5	0.7	0.8	0.9	2.7
	Veterinary students	224	0.0	0.0	1.7	1.8	2.5	6.8

system by using i.e., the Fagans nomogram (Fagan, 1975) but such an analysis was not performed in this study. When focusing on specific clinical signs in the calf, the observer might not emphasize on behavioral changes because the primary focus is on the specific clinical signs of disease. A major limitation in our study was that the same single observer both scored the VAS and the WCHSC with only a short time between, which could be an observer bias in the data. Martin et al. (2020) stated that the VAS score was a mark on a 100 mm line and the distance from the mark to the ends of the line were measured manually with a ruler which indicated the clinical score at the specific point on the line. In our study we just noted the specific number on the 10 cm scale without using a physical drawn line and noted the score with 0.1 increments. It is not clear whether this difference in how to perform the score introduced any unintended biases to our results.

It is not clear whether the clinical signs obtained in the WCHSC or in the VAS truly reflect lesions in the lungs which are an important parameter of respiratory disease in calves. Leruste et al. (2012) found that a high proportion of lungs with moderate or severe lesions detected by post mortem examination, could not be accurately detected by a protocol assessing clinical signs of respiratory disease alone (AUC = 0.77). Therefore, it could be a possibility that the signs we conclude on mostly reflect upper respiratory disease signs.

In this study, neither the WCHSC nor the VAS scoring system can be considered a perfect test, and therefore an appropriate gold standard test, for respiratory disease detection in calves. For a more precise diagnosis, the clinical signs could have been followed up by TUS which could have added value as an alternative reference test for VAS instead of using the WCHSC even though TUS neither is a perfect gold standard test. Reported performances of TUS ranging from 82% to 99% for Se and from 81% to 95% for Sp (Amrine et al., 2013). Cuevas-Gómez et al. (2021) found that 61% of calves had lung lesions up to 10.5 d before a clinical respiratory disease diagnosis which further emphasizes that the clinical diagnosis is lacking diagnostic quality without other diagnostics to support the diagnosis. Cramer et al. (2019) concluded that only 23% of clinical cases of respiratory disease showed signs of depression compared with controls and the authors claimed that many cases of respiratory disease will go undetected unless an additional screening tool is used. This is an argument for why clinical scoring primarily may act as a screening tool in the recommendation for on-farm usage. Love et al. (2014) developed an alternative clinical scoring protocol to detect respiratory disease based on the WCHSC and included nasal swab qPCR viral pathogen detection as a follow-up diagnosis of the clinical

diagnosis; even though that pathogen detection from nasal swabs is also of limited value because pathogens can act as commensals in some calves but cause disease in others (Angen et al., 2009; Klompaker et al., 2021).

Despite the 2 scoring systems use different approaches in identifying diseased calves, they overall agree about the disease decision in the calves when the VAS cut-off was 2.8. On the other hand, when looking at the conditional probability of $p(\text{high VAS} | \text{WCHSC} \geq 5)$ (0.276) and $p(\text{low VAS} | \text{WCHSC} < 5)$ (0.984), we found that the agreement between the 2 methods were highest when WCHSC was below the specified cut-off. There was more disagreement when the WCHSC was high, and this also emphasizes the difference in nature of the 2 different scoring systems. The WCHSC protocol is one among several protocols that all categorize specific clinical signs into a general clinical score that are guiding decisions in treatment of respiratory disease (Love et al., 2014; McGuirk & Peek, 2014; Berman et al., 2022). VAS might be a tool to use for detection of respiratory disease cases that hypothetically can be initially treated with NSAIDs. Only few studies have evaluated this concept and found varying effect (Carvallo Chaigneau et al., 2021; Mahendran et al., 2020), but it is a very interesting way of trying to reduce the amount of antimicrobials used for treatment of respiratory disease in calves.

In the second part of the study, we found a poor to moderate reliability between experienced observers in calf health and between an experienced observer and veterinary students (ICC = 0.58 and 0.34, respectively) in using the VAS for detecting signs of respiratory disease. Berman et al. (2020) reported that clinical signs were not detected equally between veterinarians, technicians, and producers which we also found in our reliability assessment. Few studies have evaluated the inter-rater reliability of using VAS in calves. A behavioral study of calves with focus on specific temperament patterns, found ICC's in VAS between observers ranging between 0.12 and 0.76 (Vogt et al., 2017). A study reported ICC's ranging between 0.52 to 0.70 when analyzing reliabilities of raters by using TUS in calves with and without respiratory disease (Buczinski et al., 2018). Overall, the ICC's in this and other studies indicate that it can be difficult to achieve general high levels of reliabilities when assessing animal health and disease status or behavioral expressions by continuous measuring scales. Stakeholders in calf health e.g., veterinarians, calf caretakers, farm managers, technicians, producers and consultants, shall be aware of how differences in reliability in clinical scoring can have an impact on decision-making outcomes based on observations of calves by different people. Also, it is important that the people that are involved in clinical

scoring screenings receive a thorough introduction and education within the field to improve the reliability and the quality in the scoring.

More studies are needed to validate clinical scoring as a valuable tool to detect respiratory disease, including a more precise definition of the VAS so the reliability between observers can be optimized. Continuous scales might already be of high quality when doing intra-observer reliability and must be further analyzed in future studies.

CONCLUSIONS

We used a VAS to evaluate signs of respiratory disease in dairy calves and compared it to the WCHSC. We showed that it might be possible to implement another scoring protocol to detect clinical signs of respiratory disease in calves because of a relatively high level of agreement between the 2 scoring systems. As also found in previous studies, there was poor to moderate inter-observer reliability in the VAS of calves with signs of respiratory disease between experienced observers and less experienced observers in calf health. This calls for better training and calibration of users of the different disease scoring systems. Further studies to define outcome-based cut-offs in using VAS for the desired purposes are needed. We also recommend making a follow-up diagnosis with other respiratory disease diagnostic tools to verify the diagnosis after using a clinical scoring protocol, like VAS, to screen for respiratory disease in dairy calves.

ACKNOWLEDGMENTS

The Danish Veterinary and Food Administration and The University of Copenhagen funded this work. A huge thank you to the farmer that was willing to let us conduct the study in their dairy farm. Also, a great thank you to the participating veterinary students that agreed to participate in the study and contribute to veterinary clinical research.

Conflict of Interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

- Amrine, D. E., B. J. White, R. Larson, D. E. Anderson, D. A. Mosier, and N. Cernicchiaro. 2013. Precision and accuracy of clinical illness scores, compared with pulmonary consolidation scores, in Holstein calves with experimentally induced *Mycoplasma bovis* pneumonia. *Am. J. Vet. Res.* 74:310–315. <https://doi.org/10.2460/ajvr.74.2.310>.
- Angen, Ø., J. Thomsen, L. E. Larsen, J. Larsen, B. Kokotovic, P. M. H. Heegaard, and J. M. D. Enemark. 2009. Respiratory disease in calves: Microbiological investigations on trans-tracheally aspirated bronchoalveolar fluid and acute phase protein response. *Vet. Microbiol.* 137:165–171. <https://doi.org/10.1016/j.vetmic.2008.12.024>.
- Aulakh, K. S., K. R. Dongaonkar, K. Barnes, A. J. Gines, J. T. Bordenon, D. Hulse, H. K. Aulakh, and C. Liu. 2020. Influence of orthopedic examination on lameness scores and interobserver and intraobserver agreement in dogs with naturally occurring elbow osteoarthritis. *Vet. Surg.* 49:455–462. <https://doi.org/10.1111/vsu.13390>.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting Linear Mixed-Effects Models Using *lme4*. *J. Stat. Softw.* 67:1406–5823 <https://doi.org/10.18637/jss.v067.i01>
- Berman, J., D. Francoz, A. Abdallah, S. Dufour, and S. Buczinski. 2022. Development and validation of a clinical respiratory disease scoring system for guiding treatment decisions in veal calves using a Bayesian framework. *J. Dairy Sci.* 105:9917–9933. <https://doi.org/10.3168/jds.2021-21695>.
- Berman, J., I. Masseau, G. Fecteau, S. Buczinski, and D. Francoz. 2020. Comparison between thoracic ultrasonography and thoracic radiography for the detection of thoracic lesions in dairy calves using a two-stage Bayesian method. *Prev. Vet. Med.* 184:105153. <https://doi.org/10.1016/j.prevetmed.2020.105153>.
- Buczinski, S., C. Buathier, A. M. Bélanger, H. Michaux, N. Tison, and E. Timsit. 2018. Inter-rater agreement and reliability of thoracic ultrasonographic findings in feedlot calves, with or without naturally occurring bronchorespiratory disease. *J. Vet. Intern. Med.* 32:1787–1792. <https://doi.org/10.1111/jvim.15257>.
- Buczinski, S., C. Faure, S. Jolivet, and A. Abdallah. 2016. Evaluation of inter-observer agreement when using a clinical respiratory scoring system in pre-weaned dairy calves. *N. Z. Vet. J.* 64:243–247. <https://doi.org/10.1080/00480169.2016.1153439>.
- Buczinski, S., G. Forté, D. Francoz, and A. M. Bélanger. 2014. Comparison of thoracic auscultation, clinical score, and ultrasonography as indicators of bovine respiratory disease in preweaned dairy calves. *J. Vet. Intern. Med.* 28:234–242. <https://doi.org/10.1111/jvim.12251>.
- Carvalho Chaigneau, F. R., P. Walsh, M. Lebedev, V. Mutua, H. McEligot, H. Bang, and L. J. Gershwin. 2021. A randomized controlled trial comparing non-steroidal anti-inflammatory and fusion protein inhibitors singly and in combination on the histopathology of bovine respiratory syncytial virus infection. *PLoS One* 16:e0252455. <https://doi.org/10.1371/journal.pone.0252455>.
- Cramer, M. C., T. L. Ollivett, and A. L. Stanton. 2016. Associations of behavior-based measurements and clinical disease in preweaned, group-housed dairy calves. *J. Dairy Sci.* 99:7434–7443. <https://doi.org/10.3168/jds.2015-10207>.
- Cramer, M. C., K. L. Proudfoot, and T. L. Ollivett. 2019. Short communication: Behavioral attitude scores associated with bovine respiratory disease identified using calf lung ultrasound and clinical respiratory scoring. *J. Dairy Sci.* 102:6540–6544. <https://doi.org/10.3168/jds.2018-15550>.
- Cuevas-Gómez, I., M. McGee, J. M. Sánchez, E. O’Riordan, N. Byrne, T. McDanel, and B. Earley. 2021. Association between clinical respiratory signs, lung lesions detected by thoracic ultrasonography and growth performance in pre-weaned dairy calves. *Ir. Vet. J.* 74:7. <https://doi.org/10.1186/s13620-021-00187-1>.
- Cummings, D. B., N. F. Meyer, and D. L. Step. 2022. Bovine Respiratory Disease Considerations in Young Dairy Calves. *Vet. Clin. North Am. Food Anim. Pract.* 38:93–105. <https://doi.org/10.1016/j.cvfa.2021.11.007>.
- Decaris, N., S. Buczinski, D. I. C. Tárdon, L. Camargo, N. R. Schlemmer, S. C. F. Hagen, A. R. Woolums, and V. Gomes. 2022. Diagnostic accuracy of Wisconsin and California scoring systems to detect bovine respiratory disease in preweaning dairy calves under subtropical environmental conditions. *J. Dairy Sci.* 105:7750–7763. <https://doi.org/10.3168/jds.2021-21491>.
- Diaz-Lundahl, S., S. Hellestveit, S. M. Stubbsjøen, J. Phythian, C. Oppermann, R. Moe, and K. Muri. 2019. Intra- and Inter-Observer Reliability of Qualitative Behaviour Assessments of Housed

- Sheep in Norway. *Animals (Basel)* 9:569. <https://doi.org/10.3390/ani9080569>.
- Duthie, C. A., J. M. Bowen, D. J. Bell, G. A. Miller, C. Mason, and M. J. Haskell. 2021. Feeding behaviour and activity as early indicators of disease in pre-weaned dairy calves. *Animal* 15:100150. <https://doi.org/10.1016/j.animal.2020.100150>.
- Fagan, J. 1975. Nomogram for Bayes's Theorem. *N. Engl. J. Med.* 293:257. <https://doi.org/10.1056/NEJM197507312930513> <https://www.nejm.org/doi/full/10.1056/NEJM197507312930513>.
- Gamer, M., J. Lemon, I. Fellows, and P. Singh. 2019. Package “irr”: Various Coefficients of Interrater Reliability and Agreement (0.84.1). Accessed Mar. 16, 2023 <https://CRAN.R-project.org/package=irr>
- Hartig, F. 2022. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.6. Accessed Mar. 17, 2023 <https://CRAN.R-project.org/package=DHARMa>
- Hixson, C. L., P. D. Krawczel, J. M. Caldwell, and E. K. Miller-Cushon. 2018. Behavioral changes in group-housed dairy calves infected with *Mannheimia haemolytica*. *J. Dairy Sci.* 101:10351–10360. <https://doi.org/10.3168/jds.2018-14832>.
- Klomp maker, A. F., M. Brydensholt, A. M. Michelsen, M. J. Denwood, C. T. Kirkeby, L. E. Larsen, N. B. Goecke, N. D. Otten, and L. R. Nielsen. 2021. Estimating Clinically Relevant Cut-Off Values for a High-Throughput Quantitative Real-Time PCR Detecting Bacterial Respiratory Pathogens in Cattle. *Front. Vet. Sci.* 8:674771. <https://doi.org/10.3389/fvets.2021.674771>.
- Lenth, R. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means, R package version 1.8.3. Accessed Jan. 13, 2023. <https://CRAN.R-project.org/package=emmeans>
- Leruste, H., M. Brscic, L. F. M. Heutinck, E. K. Visser, M. Wolthuis-Fillerup, E. A. M. Bokkers, N. Stockhofe-Zurwieden, G. Cozzi, F. Gottardo, B. J. Lensink, and C. G. van Reenen. 2012. The relationship between clinical signs of respiratory system disorders and lung lesions at slaughter in veal calves. *Prev. Vet. Med.* 105:93–100. <https://doi.org/10.1016/j.prevetmed.2012.01.015>.
- Liljequist, D., B. Elfving, and K. S. Roaldsen. 2019. Intraclass correlation – A discussion and demonstration of basic features. *PLoS One* 14:e0219854. <https://doi.org/10.1371/journal.pone.0219854>.
- Love, W. J., T. W. Lehenbauer, P. H. Kass, A. L. van Eenennaam, and S. S. Aly. 2014. Development of a novel clinical scoring system for on-farm diagnosis of bovine respiratory disease in pre-weaned dairy calves. *PeerJ* 2:e238. <https://doi.org/10.7717/peerj.238>.
- Love, W. J., T. W. Lehenbauer, A. L. van Eenennaam, C. M. Drake, P. H. Kass, T. B. Farver, and S. S. Aly. 2016. Sensitivity and specificity of on-farm scoring systems and nasal culture to detect bovine respiratory disease complex in preweaned dairy calves. *J. Vet. Diagn. Invest.* 28:119–128. <https://doi.org/10.1177/1040638715626204>.
- Lowie, T., K. van Leenen, S. Jourquin, M. L. Pas, J. Bokma, and B. Pardon. 2022. Differences in the association of cough and other clinical signs with ultrasonographic lung consolidation in dairy, veal, and beef calves. *J. Dairy Sci.* 105:6111–6124. <https://doi.org/10.3168/jds.2021-21570>.
- Mahendran, S. A. 2020. Use of fever detection in combination with thoracic ultrasonography to identify respiratory disease, and compare treatments of antimicrobials and NSAID: A randomised study in dairy calves. *Vet. Rec. Open* 7:e000415. <https://doi.org/10.1136/vetreco-2020-000415>.
- Mahendran, S. A., R. Booth, L. Beekhuis, A. Manning, T. Blackmore, A. Vanhoudt, and N. Bell. 2017. Assessing the effects of weekly preweaning health scores on dairy calf mortality and productivity parameters: Cohort study. *Vet. Rec.* 181:196–196. <https://doi.org/10.1136/vr.104197>.
- Martin, M., M. Kleinhenz, K. Kleinhenz, E. Reppert, D. Blasi, T. Parks, A. Baysinger, J. Hutcheson, and J. Coetzee. 2020. Comparison of the effect of tildipirosin administered alone or in combination with transdermal flunixin on the performance, health, activity, and well-being of transported feedlot calves on arrival at the feedlot. *Transl. Anim. Sci.* 4:452–459. <https://doi.org/10.1093/tas/txaa005>.
- Martin, M. S., M. D. Kleinhenz, B. J. White, B. T. Johnson, S. R. Montgomery, A. K. Curtis, M. M. Weeder, D. A. Blasi, K. M. Almes, R. G. Amachawadi, H. M. Salih, M. D. Miesner, A. K. Baysinger, J. S. Nickell, and J. F. Coetzee. 2022. Assessment of pain associated with bovine respiratory disease and its mitigation with flunixin meglumine in cattle with induced bacterial respiratory disease. *J. Anim. Sci.* 100:1–13. <https://doi.org/10.1093/jas/skab373>.
- McGuirk, S. M. 2008. Disease Management of Dairy Calves and Heifers. *Vet. Clin. North Am. Food Anim. Pract.* 24:139–153. <https://doi.org/10.1016/j.cvfa.2007.10.003>.
- McGuirk, S. M., and S. F. Peek. 2014. Timely diagnosis of dairy calf respiratory disease using a standardized scoring system. *Anim. Health Res. Rev.* 15:145–147. <https://doi.org/10.1017/S1466252314000267>.
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Accessed Jan. 12, 2023. <https://www.R-project.org/>
- Revelle, W. 2022. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, version = 2.2.9. Accessed Mar. 17, 2023. <https://CRAN.R-project.org/package=psych>
- Rstudio Team. 2022. RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. Accessed Jan. 12, 2023 <http://www.posit.co/>
- Swartz, T. H., A. N. Findlay, and C. S. Petersson-Wolfe. 2017. Short communication: Automated detection of behavioral changes from respiratory disease in pre-weaned calves. *J. Dairy Sci.* 100:9273–9278. <https://doi.org/10.3168/jds.2016-12280>.
- Vogt, A., E. L. Aditia, I. Schlechter, S. Schütze, K. Geburt, M. Gauly, and U. König von Borstel. 2017. Inter- and intra-observer reliability of different methods for recording temperament in beef and dairy calves. *Appl. Anim. Behav. Sci.* 195:15–23. <https://doi.org/10.1016/j.applanim.2017.06.008>.
- Zhang, X., X. Yi, H. Zhuang, Z. Deng, and C. Ma. 2022. Invited Review: Antimicrobial Use and Antimicrobial Resistance in Pathogens Associated with Diarrhea and Respiratory disease in Dairy Calves. *Animals (Basel)* 12:771. <https://doi.org/10.3390/ani12060771>.

ORCID

- Henrik H. Møller  <https://orcid.org/0009-0007-6259-7614>
 Mogens A. Krogh  <https://orcid.org/0000-0003-0731-6676>
 Mette B. Petersen  <https://orcid.org/0000-0001-7195-2103>
 Liza R. Nielsen  <https://orcid.org/0000-0003-2046-2387>
 Nynne Capion  <https://orcid.org/0000-0001-8278-9321>