Development of a multivariable prediction model to identify dairy calves too young to be transported to auction markets in Canada using simple physical examination and body weight

S. Buczinski,1*, G. Fecteau,1 M. Cichocki,2 S. Ferraro,1 J. Arsenault,3 Y. Chorfi,4 M. Costa,4 J. Dubuc,1 D. Francoz,1 M. Rousseau,1 and M. Villetaz-Robichaud1

1Département des Sciences Cliniques, Faculté de Médecine Vétérinaire, Université de Montréal, St-Hyacinthe, QC, J2S 2M2 Canada
2École Nationale Vétérinaire de Toulouse, Toulouse, 31076 France
3Département de Pathologie et Microbiologie, Faculté de Médecine Vétérinaire, Université de Montréal, St-Hyacinthe, QC, J2S 2M2 Canada
4Département de Biomédecine, Faculté de Médecine Vétérinaire, Université de Montréal, St-Hyacinthe, QC, J2S 2M2 Canada

ABSTRACT

Calves born on Eastern Canadian dairy farms that are not kept in the herds are traditionally sold through auction markets and are raised for meat purposes such as veal calves. Since February 2020, a new Canadian federal regulation has forbidden calves <9 d old to be sold through auction markets. However, in the absence of a real-time birth registry consultation system, it would be of interest to look for predictors that could be associated with age to allow identification of calves too young to be transported. In the current retrospective cross-sectional study, 1,178 calves with a declared birth date (411 calves aged <9 d old; 34.9%) were assessed in 2 large Québec auction sites. Easy-to-record covariates [body weight (BW), breed phenotype, and presence of an umbilical cord remnant] as well as other clinical signs (umbilical swelling, enlargement, umbilical pain, wet umbilicus, skin tent, sunken eyes, ocular and nasal secretion, and hide cleanliness) were assessed. Two logistic regression models using age as a dichotomous dependent variable (<9 d old vs ≥9 d old) were built. The first model (model 1) considered all covariates, which were selected after univariable analyses and a backward stepwise selection process, whereas a more pragmatic model (model 2) only included the 3 easy-to-record variables (i.e., BW, breed, umbilical cord). Both models had similar accuracy to detect calves <9 d old (sensitivity of 38.4 and 37.5%, and specificity of 85.7 and 84.6% for model 1 and 2, respectively). Model 2 was subsequently more specifically studied as it employs a faster and easier assessment. Decision thresholds were tested for their robustness based on misclassification cost term (MCT) analysis with various prevalence of calves <9 d old and various costs of false-negative: false-positive ratio. Despite statistical significance, model accuracy, even if refined with MCT analysis, was limited at the individual level, showing the limits of using physical signs and BW or their combination as a reliable proxy of age. The sensitivity of these models to find calves <9 d old was not to be used for monitoring compliance with the Canadian federal regulation. The relatively high model specificity may help to use this model as a rule-in test (i.e., targeting positive calves for further investigation) rather than a rule-out test (due to its low sensitivity).

Key words: surplus calves, regulation, fitness for transport, diagnosis

INTRODUCTION

In dairy cattle production, some calves in the herd act as replacement heifers, whereas male and extra female calves (also called surplus calves) are traditionally sold at an early age. In Eastern Canada, most these surplus calves are transported at a young age to auction markets to be sold for veal production (Renaud et al., 2018; Marquou et al., 2019; Wilson et al., 2020b). Until recently, there were no transportation restrictions in Canada regarding minimum age of calves. Since February 2020, a new regulation put in place by the Government of Canada (2019) for improving welfare of dairy calves forbids transportation to the auction market of calves less than 9 d old. It also limits the length of travel time (≤12 h) and restricts the maximal time delay between 2 milk feeds during the journey (≤12 h). Previous research has shown that both calf characteristics and transportation conditions may influence calf welfare and health (Jongman and Butler, 2014; Deters and Hansen, 2020). For instance, despite exhibiting less stress response relative to transportation (as assessed by lower corticosteroid response), young calves have a greater risk of mortality after transportation than older calves, as reviewed by Eicher (2001).
This mortality risk is negatively correlated with age (Knowles, 1995; Eicher, 2001). However, it is difficult to distinguish the effect of transportation versus the effect of age on mortality because mortality is also higher in nontransported calves compared with older calves. The behavior of 3-d-old calves during transportation is also different from that of older calves but it remains unclear how this may affect their ability to cope with transportation stress (Jongman and Butler, 2014). The latter study also reported that 3-d-old calves laid down more often during transportation and had a longer recovery period compared with 5- and 10-d-old calves. In that context, younger calves were less interested in drinking water during recovery compared with older animals.

As previously reported, it is difficult at auction markets to assess compliance with this new transport regulation in the absence of reliable ways to provide real-time information about the exact birth date (Buczinski et al., 2020). Due to this specific limitation, it appears important to investigate indirect indicators of calf age that could be used when there is suspicion that a calf <9 d old is present at the auction site. Development of such tools would be helpful to improve application of the new regulation with direct feedback to the dairy producers or drovers. This would also discourage the transportation of young calves with a higher risk of being negatively affected by the journey and commingling. A predictive model using serum gamma-glutamyl transferase (GGT) activity and BW detected calves <9 d old with a sensitivity and specificity of 70.4 and 77.3%, respectively (Buczinski et al., 2020). However, beside its prohibitive cost, the measurement of serum GGT activity cannot be performed on site, meaning that an extra step would be necessary before intervening, making it impractical for rapid decision making. Several other individual indicators of calves’ age have been investigated, such as the presence of umbilical cord remnant (Hides and Hannah, 2005) or low BW (Buczinski et al., 2020). As recently highlighted by a Canadian consensus group on health and welfare challenges in the marketing of male dairy calves, improving age traceability or at least identifying measures strongly associated with age through the marketing system would be important for both the dairy and veal industries (Wilson et al., 2020). This would help to monitor and reinforce the current compliance with the new Canadian federal regulation forbidding transportation to auction markets of calves <9 d old.

Therefore, the objective of the study was to develop a predictive model with information that can be easily collected at the auction markets that could be used to determine the probability of a calf to be <9 d old. Our hypothesis was that clinical information that can be collected rapidly on site could be used as a screening test to detect calves with a high risk of being <9 d old, thus being useful to monitor the compliance with the new Canadian federal regulation.

**MATERIALS AND METHODS**

This study was reported according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist for prediction model development (Collins et al., 2015). Data collection for this retrospective cross-sectional study was performed in compliance with the animal care and ethics committee of the Université de Montréal (Protocol number 20-Rech-2058).

**Calf and Health Variables Available for Age Prediction**

The data set used for the present study was a subset of a larger study in which clinical information of calves sold in auction markets were collected to study dairy farm risk factors associated with calf price and quality during summer 2019 and winter 2020 (Ferraro et al., 2020). This scoring system was adapted from previously reported scoring charts with the aim of being applicable in auction settings (Graham et al., 2018; Marquou et al., 2019).

Briefly, 3,656 calves were examined in 2 large auction markets in Québec (sites where 400–700 calves/d per site are sold), which represent 74% of the annual provincial auction market calves’ sales (Buczinski et al., 2021). Each auction market was visited twice during the summer of 2019 and winter of 2020 (total of 4 visits per market site). The calves were examined using a standardized scoring system that included presence of umbilical cord remnant (present, absent), wet external umbilical structures (present, absent), and umbilical swelling of more than 2 fingers wide (present, absent). Evidence of umbilical pain was noted when abdominal contraction, defense reaction, or both were present at palpation (present, absent). Hydration status was assessed using skin tent after pinch and twist perpendicular to the neck with return time ≥2 s as an indication of dehydration as previously validated (Kells et al., 2020). Eye position in the orbit was also noted as a trichotomous variable (normal, mildly sunken, and severely sunken). Hide cleanliness was assessed using the presence or absence of >25% of the caudal thighs or upper leg under the tail covered by either liquid (indicator or possible recent diarrhea) or crusted fecal material (i.e., coming from either dried previous diarrhea...
or normal crusted feces). Ocular secretions were noted as absent, serous (epiphora), or purulent (crusted or purulent). Nasal secretions were assessed for the presence or absence of purulent secretion. Ear position was assessed and defined as positive in case of bilateral ear droop. Evaluation of this scoring chart reliability was performed between the different operators performing the tests in a pilot study on a local dairy farm and in auction market setting after training with 1 senior clinician with extensive experience on calf health (SB). A total of 19 calves were screened by 8 different persons during 2 training sessions although the auction data collection was mostly performed by 4 raters. With the exception of ear drop (0.31), the percentage of agreement varied between 0.87 and 1 (median = 0.95). The associated Gwet agreement coefficient type 1 (AC1), which is not affected by missing agreement pairs (e.g., operators from different training sessions did not screen the same calves) and is more robust to kappa paradox, was then used (Wongpakaran et al., 2013; Gwet, 2014). Except for ear drop (AC1 = 0.1) the AC1 varied from 0.76 to 1, which was considered good. The chart was considered robust enough to be applied in a larger population.

The provincial unique identification number of all calves in the study were collected after the auction days and sent to the Quebec agency in charge of national birth registry (Attestra; https://attestra.com/) to obtain information on recorded birth declarations. For a particular calf, the birth date could either be marked as declared by the producer, estimated by the producer, or declared by the auction market (i.e., where the birth date is estimated based on calf weight at the day of auction). Only records of dates declared by the producers were kept for further analysis because it was anticipated that records based on calving dates were the most accurate estimation of the true birth date. The calf breed phenotype was also noted and categorized as Holstein, beef crossbred, or another dairy breed. The calf’s official sale weight was recorded and the centered BW was defined as the specific calf BW minus mean calf weight. We chose to report the centered weight (vs. crude weight) as a predictor because of practical relevance. The estimation of the regression slope could therefore be interpreted as the effect of +1 kg versus the “average” calf weight.

We focused on calves less than 4 wk of age for model building because we anticipated that older calves would have lower risk of being suspected to be <9 d old and would not be the target population. This 4 wk of age limit was based on the definition of the neonatal period (Roy, 1983) and on the fact that it is half of the typical recommended preweaning period length in dairy calves in North America (Kertz et al., 2017). The data set kept for further analysis only contained calves <28 d of age.

**Statistical Analysis**

Descriptive statistics and statistical analyses were performed using the R version 4.0.3 open access software (R Development Core Team, 2020). The general study framework is depicted in Figure 1.

**Model Building Strategy**

The dependent variable was a priori defined as the age of the calf used as a dichotomous variable (<9 d old vs. ≥9 d old) modeling the probability of a calf being <9 d old. The explanatory variables offered to at least 1 model were the presence of clinical dehydration (skin tent and sunken eyes), umbilical cord remnant, umbilical swelling, wet umbilicus and umbilical pain at palpation, hide cleanliness, ocular and nasal secretions as well as centered BW, sex, and breed phenotype. The correlation between all reported explanatory variables that were mostly categorical were assessed using the Goodman Kruskal’s τ (tau) statistics. This asymmetric test was based on the fraction of variability in a categorical variable A that can be explained by another categorical variable B. Forward [τ (A, B)] and backward [τ (B, A)] calculations were therefore performed (Mirkin, 1992). We defined a priori a Goodman Kruskal’s τ >0.6 as indicator of possible collinearity between 2 different variables.

Modeling approach then used 2 different analysis steps. For the first model, univariable analyses were performed with Chi-squared statistics against the dependent variable for categorical independent variables (step 1). Centered weight, which was normally distributed (Shapiro-Wilk test), was directly tested against the dependent variable using logistic regression Wald test. Variables with \( P \)-values <0.2 were then selected to be included in the multivariable analysis (step 2) using a backward stepwise approach until all remaining variables had \( P \)-values <0.05 to obtain the final model (model 1). A second model, which was considered easier to implement in a large group of animals, was based on predictors that can be assessed quickly and objectively which were the breed phenotype, the presence of umbilical cord, and calves’ centered weight. With this approach, the 3 main predictors of interest were directly put as independent variables in the multivariable logistic regression model (model 2).

The logistic models’ fits were then assessed using Nagelkerke R-squared statistics (Harrell Jr., 2015). The
2 models were also compared using Akaike information criterion (AIC) as a compromise between increase of the maximum likelihood of the model and model complexity. Discrimination and calibration of the models were assessed as the 2 complementary model characteristics in terms of prediction, as recommended (Fenlon et al., 2018). Model calibration was also assessed using calibration plot. These plots represent the observed probability of being <9 d old versus the mean predicted probability obtained after classifying population prediction probabilities in deciles (Fenlon et al., 2018). Discrimination of the models was determined using the concordance test statistic (which is equivalent to the area under the receiver operating characteristic curve), model sensitivity defined as the percentage of calves <9 d old correctly predicted by the model, and specificity defined as the percentage of calves ≥9 d old adequately predicted by the model. The probability threshold level used from the logistic regression model was by default assumed as ≥0.5. If the predicted probability was ≥0.5, the result was compatible with a calf <9 d old whereas if the predicted probability was <0.5, the calf was initially considered ≥9 d old.

The statistical significance level was set at \( P < 0.05 \). We simulated a prediction probability for model 2 to provide a real-world interpretation of the data, which was then modeled using various scenarios of covariates characteristics. Briefly, a new data set with 1,000 fictive calves covering various covariates profiles was created. Mean predicted probabilities and bootstrapped con-

![Flowchart](image)

**Figure 1.** Flowchart indicating the data selection process and the modeling strategies used to predict the probability that a calf is <9 d old.
Confidence intervals were obtained using 10,000 different iterations model estimation with finalfit R package (Harrison et al., 2019).

The initial prediction from logistic regression assumes by default that a prediction higher than 0.5 is used as a decision threshold for detecting calves <9 d old. Robustness of the decision threshold of model 2 (i.e., changing from the default 0.5 decision level with various other possible threshold from 0 to 1) was tested across various scenario of false-negative versus false-positive costs using a misclassification cost term analysis (MCT; Greiner, 1996). The MCT was determined using the following formula:

\[
\text{MCT} = (1 - p) \times (1 - \text{Sp}) + r \times p \times (1 - \text{Se}),
\]

where \( p \) is the prevalence of calves <9 d old and \( \text{Se} \) and \( \text{Sp} \) are, respectively, the sensitivity and specificity of the model for a specific probability threshold. The variable \( r \) is the false-negative to false-positive cost ratio. In the current study, it was difficult to precisely know this ratio because it depends on the objective of this type of monitoring, which may depend on political context. Therefore, we hypothesized 3 different possible conditions where (1) false-negative cases (calves <9 d old not correctly classified by the model) were considered as having higher costs than false-positive cases (calves ≥9 d old not adequately classified) with \( r \) of 5:1 and 2:1; (2) neutral scenario where false-negative and false-positive cases have the same costs; and (3) finally a situation where false-positive cases have higher costs than false-negative cases with \( r \) = 1:2 and 1:5. Scenario 1 can be considered relevant in conditions where the animal welfare perspective has higher importance and therefore assuming a lower cost for false-positive cases, whereas scenario 2 is neutral, and scenario 3 is trying to avoid false-positive results with a calf falsely considered as too young (avoiding unwarranted penalties for farmers and transporters). This analysis was performed using the original prevalence of young calves in the data set (35%). The same scenarios were also repeated for 3 different simulated prevalence ranges of the reported outcome.

![Figure 2. Distribution of age of calves that were assessed in Québec auction markets. The minimal age limit for calves to be transported to an auction market based on the federal regulation (9 d) is indicated in blue dashed line. A dotted line indicates the upper limit of inclusion of older calves (28 d) in the prediction model. Density represents the relative likelihood (probability) of the age distribution in the observed population.](image-url)
The prevalence of young calves was lower than observed in the initial data set were also drawn until reaching prevalence of 5 and 20%. In this situation all calves ≥9 d old were kept and the data set was filled with a random draw of calves <9 d old with no replacement using sample_n() function from the tidyverse library (Wickham and Grolemund, 2016).

RESULTS

The general flowchart of the study is shown in Figure 1. A total of 3,656 calves were examined in the current study. From these animals, the declared birth date could be obtained for 1,303 calves (35.7%) using Attestra registry. For these calves, the median age was 11 d [interquartile range (IQR): 8–18 d]. The distribution of calves’ age is presented in Figure 2. A total of 1,178 calves were finally available for model building after removing calves ≥28 d old. Of these 1,178 calves, 411 (34.9%) were <9 d old. The main descriptive characteristics of these calves are indicated in Table 1. The mean (SD) BW was 50 (5.8) kg. Distribution of centered BW is shown in Figure 3. Univariable analyses revealed that calves with umbilical cord remnant, wet umbilical area, and umbilical swelling had higher odds of being <9 d old with associated odds ratio of 3.12 (95% CI: 2.43–4.00), 1.95 (1.20–3.16), and 1.62 (1.21–2.19), respectively.

The main results of both models—model 1 with a first univariable selection of covariables and subsequent multivariable analysis, and the simpler model 2 with only the 3 predictors of interest (breed, weight, and umbilical cord remnant)—are depicted in Table 2. The models had similar accuracy with a low sensitivity and moderate specificity of 38.4% (95% CI: 33.7–43.3) and 85.7% (83.0–88.1) for model 1, and 37.5% (32.8–42.4) and 84.6% (81.9–87.1) for model 2, respectively, using a probability threshold of 0.5. The discrimination and Nagelkerke R-squared were similar for both models (Table 2). The calibration plots are reported for both models in Figure 4. For practical reasons, model 2 was selected for further investigations despite its higher AIC due to its simplicity of recording and similar test accuracy and calibration characteristics. Specific predicted probabilities depending on various calves’ breed, umbilical cord status, and BW obtained from bootstrapped models (i.e., model 2) are indicated in Figure 5.

The MCT analysis to determine the effect of prevalence and relative cost of false-negative and false-positive cases is presented in Figure 6. As a general trend, the decision threshold increased when the prevalence of the target condition (calf <9 d old) decreased and when the false-negative: false-positive (FN:FP) cost ratio (r) decreased. Two different possible thresholds emerged in addition to the 0.5 decision probability. Using 0.25
(< vs. ≥) decision threshold was useful in the 2 high-prevalence scenario (35% and 50%) and when FN:FP cost is high. A threshold of 60% (< vs. ≥) is useful when the prevalence is low or when FN:FP cost is low but its added value vs the 50% (< vs. ≥) decision probability was limited.

**DISCUSSION**

In this study, we evaluated practical and simple predictors that could be used in the absence of a reliable way to determine birth date of young dairy calves that are sold at auction markets. We found that various physical predictors and BW are associated with the probability of a calf being too young to be transported in auction market under the Canadian federal regulation (<9 d old). However, the current study demonstrated that the ability to predict age <9 d old is limited when using various combinations of physical signs or using a simpler approach including centered BW and the presence of umbilical cord remnant. The sensitivities of the models were low, whereas the specificities were moderate (>80%). High specificity can be useful in clinical decision making because a positive test can be used as a rule-in condition. When adapted to the current situation, a calf that is positive (predicted probability with regression models ≥ threshold) based on our model would still require a more specific approach using various pre-test suspicion levels. Because the posttest probability of being abnormal depends on pretest probability using test likelihood ratio (Timsit et al., 2018), even a model with a relatively low false-positive rate (i.e., defined as a calf with a predicted probability of ≥0.5); LR− = likelihood ratio of a negative test (i.e., defined as a calf with a predicted probability of <0.5).

**Table 2. Multivariable logistic regression models aiming to identify calves <9 d old when sold through the Québec auction markets**

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Coefficient estimate (SE)</th>
<th>Odds ratio1 estimate (95% CI)</th>
<th>Coefficient estimate (SE)</th>
<th>Odds ratio1 estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercepts</td>
<td>−1.271 (0.131)</td>
<td>—</td>
<td>−1.216 (0.102)</td>
<td></td>
</tr>
<tr>
<td>Holstein Referent</td>
<td>Referent</td>
<td></td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Crossbred</td>
<td>0.380 (0.153)</td>
<td>0.348 (0.148)</td>
<td>1.416 (1.06–1.893)</td>
<td></td>
</tr>
<tr>
<td>Dairy</td>
<td>−0.205 (0.355)</td>
<td>−0.107 (0.348)</td>
<td>0.899 (0.454–1.777)</td>
<td></td>
</tr>
<tr>
<td>Umbilical cord</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent Referent</td>
<td>Referent</td>
<td></td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>1.079 (0.134)</td>
<td>2.942 (2.262–3.825)</td>
<td>1.044 (0.131)</td>
<td>2.841 (2.197–3.672)</td>
</tr>
<tr>
<td>Umbilical swelling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent Referent</td>
<td>Referent</td>
<td></td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>0.761 (0.167)</td>
<td>2.140 (1.543–2.969)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye discharge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No discharge Referent</td>
<td>—</td>
<td></td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Serous</td>
<td>−0.121 (0.145)</td>
<td>0.886 (0.667–1.177)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purrulent/crusted</td>
<td>−1.662 (0.791)</td>
<td>0.190 (0.04–0.894)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hide cleanliness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean Referent</td>
<td>—</td>
<td></td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Diarrhea</td>
<td>0.387 (0.371)</td>
<td>1.473 (0.712–3.047)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crusted</td>
<td>−0.435 (0.178)</td>
<td>0.647 (0.457–0.916)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centered weight²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per 1-kg increase</td>
<td>−0.081 (0.011)</td>
<td>0.922 (0.903–0.942)</td>
<td>−0.076 (0.011)</td>
<td>0.927 (0.907–0.947)</td>
</tr>
<tr>
<td>R² Nagelkerke</td>
<td>0.193</td>
<td>0.159</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>0.733</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>1.366</td>
<td>1.389</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 411 &lt;9 d old</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 767 ≥9 d old³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1The odds ratio is comparing the odds of calves <9 d old versus calves 9 to 28 d old.

2Centered weight was obtained by subtracting the mean weight (50 kg) from the individual calf weight. AUC = area under the receiver operator characteristic curve; AIC = Akaike information criterion.

3The decision threshold was a predicted probability of ≥0.5 compatible with a calf younger than 9 d old and <0.5 to detect a calf of at least 9 d old. Se = sensitivity; Sp = specificity; LR+ = likelihood ratio of a positive test (i.e., defined as a calf with a predicted probability of ≥0.5); LR− = likelihood ratio of a negative test (i.e., defined as a calf with a predicted probability of <0.5).
with a 50% pretest probability of being <9 d old and a positive test would only have a posttest probability of 71% of being <9 d old. The accuracy of the model for detecting young calves was even more limited especially for sensitivity than our previous study using serum GGT activity and BW where a sensitivity and specificity of 70.4 and 77.3% were respectively obtained (Buczinski et al., 2020).

Marketing male dairy calves is an important topic for the dairy industry because it is associated with multiple health and welfare challenges (Bolton and von Keyserlingk, 2021). Regulations recently established in Canada imposing a minimum age for a calf to be fit for transportation to auction markets is designed to improve the health and welfare of these animals by avoiding transportation of young calves. Application of this new federal regulation is challenging in a country such as Canada because the systems for identifying cattle and control of birth registry are different from one province to another. Currently, age verification is not mandatory in Canada and therefore cannot be used as a simple way to verify the new regulation on young calf transportation. This explains why simple predictors that can be monitored at the auction market remain of great interest. A highly accurate classification of age was not expected, but rather a screening system based on clinical signs (e.g., presence of umbilical cord) historically used as indicators of age (Hides and Hannah, 2005).

Interestingly, the current study identified that birth date information in the available registry was missing for most calves not kept in the herd. Even after consultation of the provincial Attestra registry, only 1,303 out of 3,656 calves (35.6%) had declared birth dates. Ideally, calves without the specific age requirement to be commingled in auction markets should not be transported. Assessment of their age and fitness for transport should therefore be primarily done by dairy producers and then confirmed by the transporter, who should ultimately decide whether the calf is fit for the intended journey. This is a key challenge for drivers and auction markets who need to be able to judge rapidly on the age criterion. The prevalence of umbilical cord presence was almost 2 times greater in <9-d-old calves (59.9%) versus ≥9-d-old calves (32.3%). However, despite this higher prevalence, presence of umbilical cord cannot be used alone as an accurate indicator of age. Even after adding information on calf breed and weight, the accuracy of the prediction was limited. However, the relatively high specificity of the model would potentially be helpful to rule in cases detected as positive by the model to further investigation.

In the absence of an affordable and accurate way to determine age fitness for transportation to a commingling site, one can control for the imperfect sensitivity and specificity of the prediction test to determine the true prevalence of calves <9 d old in a certain population. Therefore, an indirect way of applying the results of this study is a possible group-level test to determine the prevalence of calves <9 d old using the following formula (Dohoo et al., 2009): True prevalence = (apparent prevalence + Sp − 1) ÷ (Se + Sp − 1), which does not allow an individual prediction but rather a group-level assessment of the percentage of calves with age <9 d old. We have to note that the study period ended just at the beginning of the new regulation. A 2-yr period was allowed for producers and transporters to comply with this new regulation. Therefore, the prevalence of <9-d-old calves should not be extrapolated as representative of practices put in place after
the new regulation, but rather as a population assessed before the regulation took place to have a relatively large population of calves <9 d old, which was helpful for multivariable modeling.

The 35.6% of calves that had been selected based on birth registry had similar breed characteristics (70.8% Holstein, 25.3% crossbreed, and 3.9% of other dairy breed) when compared with the initial population of 3,656 calves (70.9% Holstein, 24.6% crossbreed, and 4.5% of other dairy breed) and similar proportion of calves with the presence of umbilical cord (42.0 vs. 41.9% in the initial calf population). Interestingly, the crossbred calves had higher probabilities of being <9 d old in the current study versus Holstein (referent) breed. It is difficult to interpret this finding due to the observational nature of this study. Dairy herds that commonly use beef breed semen for AI with or without sexed semen may have different characteristics than more conventional dairy herds (Holden and Butler, 2018). The perceived better price for beef cross calves in Québec auction markets may also be an incentive to ship these calves at a younger age (Buczinski et al., 2021), as their weight and muscle development are generally higher than that of Holstein calves at the same age. However, this has to be confirmed in future work.

This study comes with specific limitations. We opted for selecting the simplest model despite the fact that the complete model had lower AIC and a slightly better Nagelkerke R-squared. The choice of the best model in the presence of multiple possible competing models is never straightforward. Various techniques can be used to address model robustness and applicability, but a simpler model with easy-to-record information has a better chance of being implemented in practice. We tried to optimize the decision threshold based on the MCT analysis which allows the robustness of the decision model to be determined when using a dichotomous approach (based on an adapted probability decision threshold). As expected, the optimal decision threshold depends both on the FN:FP costs ratio and prevalence of calves <9 d old. The decision threshold decreased when the prevalence rose or when the FN:FP ratio increased.

Figure 5. Predicted probabilities of calves to be <9 d old using umbilical cord presence, centered BW, and breed phenotype (model 2). The corresponding plot was derived from 10,000 bootstrapped estimations of a fictive set of 1,000 calves with different simulated weight (with centered weight ± 20 kg), breed, and umbilical cord (Cord) association profiles. The line corresponds to the mean bootstrap estimate, and the colored area corresponds to the observed spread of 95% bootstrapped estimates. Predicted thresholds of 0.25, 0.50, and 0.60 are indicated as horizontal dotted lines.
CONCLUSIONS

In this study, we were able to find simple predictors that are associated with the risk of calves for being unfit for transportation based on Canadian federal regulation (i.e., <9 d old). However, these predictors were not accurate enough to be used at the calf level. Real-time monitoring of high-quality birth registry at the farm level could ultimately be used to decide whether transportation to auction markets is legal or not.

ACKNOWLEDGMENTS

This project was a specific addendum to a larger project supported by Programme de développement sectorial du Québec-Canada, the Producteurs de bovins du Québec (Longueuil, QC, Canada), and MITACS (Ottawa, ON, Canada). The authors also thank Réseau Encan Québec (QC, Canada) for their willingness to participate in the study, Jean-Philippe Pelletier, Nicolas Barbeau-Grégoire (St-Hyacinthe, QC, Canada) and

---

**Figure 6.** Misclassification cost term (MCT) analysis to determine the effect of false-negative: false-positive (FN:FP) cost ratio with different prevalence of calves <9 d old on the decision threshold obtained from the pragmatic prediction model (model 2). Five different FN:FP ratio models are indicated (1:1, 1:2, 1:5, 2:1, 5:1) in different colored lines. Different prevalence of calves <9 d old are presented using the original dataset (prevalence = 35%; panel A), then using a prevalence of 50% (panel B) obtained from downsampling the calves ≥9 d old and prevalence of 5% (panel C) and 20% (panel D) using internal resampling. Specific optimal decision thresholds at 0.25, 0.50, and 0.60 are indicated as vertical lines. The optimal threshold for each curve is obtained when the curves reach its minimum. A robust threshold is a threshold that is not heavily influenced by the prevalence and FN:FP cost ratio.
REFERENCES


ORCIDS

S. Buczinski https://orcid.org/0000-0002-8460-4885
J. Arsenault https://orcid.org/0000-0001-8382-7336
Y. Chorfi https://orcid.org/0000-0001-7893-9649
J. Dubuc https://orcid.org/0000-0003-1013-8372
D. Francoz https://orcid.org/0000-0002-1634-0507
M. Rousseau https://orcid.org/0000-0002-2702-434X
M. Villetazz-Robichaud https://orcid.org/0000-0001-9685-1827