Validation and inter-device reliability of a behavior monitoring collar to measure rumination, feeding activity, and idle time of lactating dairy cows

J. V. R. Lovatti,1 K. A. Dijkinga,2 J. F. Aires,3 L. F. C. Garrido,2 J. H. C. Costa,1 and R. R. Daros2*

Abstract: Inter-device precision and accuracy are not investigated for precision livestock farming (PLF) technologies, but are fundamental for the use of data in populational metrics and to compare cows’ data. This study aimed to validate a behavior monitoring collar (BMC; CowMed, Santa Maria, RS, Brazil) and its inter-device reliability. First, comparing observations with the BMC and second the inter-device precision and accuracy for rumination, feeding activity, and idle time of lactating dairy cows. Holstein cows (n = 23) were housed in a voluntary milk system free-stall barn and fitted with 2 devices within the same cow. Observations were made over 2 periods of one day (0700 to 1100h; 1400 to 1700h); the 7 h per cow were summarized for each behavior to assess the agreement of observed behavior and BMC data. To assess the inter-device reliability 26d of BMC data was summarized by day per cow for both devices. Pearson correlation (r), coefficient of determination (R²), Lin’s concordance correlation coefficient (ρc), linear regression, and Bland-Altman plots (BAP) were calculated for each period of observation. For the validation, we found high correlations for feeding activity, very high for idle time, but low correlations for rumination. The BAP was deemed acceptable and without bias; BAP mean differences ± SD were 0.83 ± 4.01, −0.48 ± 4.15, and 7.17 ± 3.94 min/h for rumination, feeding activity, and idle time, respectively. The slope of the linear regression did not differ from 1 for all behaviors but idle. For inter-device comparison, we found moderate correlations for feeding activity, idle time and a low correlation for rumination. The BAP was deemed acceptable and without bias; BAP mean differences were −0.36 ± 2.84, 0.45 ± 3.51, and −0.06 ± 2.81 min/h for rumination, feeding activity, and idle time, respectively. All slopes of the linear regressions differed from 1 but feeding time. Thus, the inter-device did not meet the accuracy criteria. In summary, this study validated the precision of the BMC for recording feeding activity of lactating dairy cows.

Monitoring animal behavior visually is subjective and requires substantial amount of time (Eerdekens et al., 2021). Precision livestock farming technologies are a noninvasive, objective measurement of animal’s behavior using algorithms to process raw data (Costa et al., 2021) and able to continuously detect real-time behavioral changes (Borchers et al., 2016). Technologies are deemed valid when they achieve satisfactory precision and accuracy compared with a gold standard (Royston and Altman, 2013).

Precision of PLF devices for monitoring cows’ behavior have been assessed by Pearson correlation coefficient and Lin’s concordance correlation coefficient (Bikker et al., 2014; Borchers et al., 2016), or coefficient of determination, while very few studies have reported accuracy results (Grinter et al., 2019). Accuracy of PLF devices has been examined by using the slope of the regression line and Bland-Altman Plots. The Bland-Altman Plots are useful to evaluate the bias between the mean differences and to estimate an agreement between 2 methods (Giavarina, 2015). The evaluation of the accuracy is essential, once it represents how closely the measures (i.e., automated recorded behaviors) are to the true values (i.e., observations) (Tedeschi, 2006). Thus, accuracy enables the development of benchmarking, allowing the comparison of the behavior recorded by the PLF device under research or farm conditions.

Despite the popularity of PLF devices, there have been few to no studies investigating inter-device reliability. Inter-device reliability is relevant and should be minimal when comparing data between and within-subject (Santos-Lozano et al., 2012). The use of data for populational measurements to make comparisons between subjects are an opportunity for PLF, but it requires inter-device reliability. In sport-tracking devices, inter-device reliability of accelerometers was found to be highly variable (Nicolleta et al., 2018). Thus, we suggest that the inter-device variability may exist between devices, and it varies depending on the behavior measured. The aim of this study is to validate the device and its inter-device precision and accuracy of a behavior monitoring collar (BMC) for lactating dairy cows for ruminating, feeding activity, and idling time. To our knowledge there were no other studies validating inter-device precision and accuracy of a PLF device in commercial settings.

This study was approved by the animal use ethics committee of the Pontifícia Universidade Católica do Paraná (CEUA-PUCPR #02090) and conducted at the Fazenda Experimental Gralha Azul of PUCPR (Fazenda Rio Grande, Paraná, Brazil).

Animals were housed in a free-stall barn divided into 2 pens approximately 85 m²/pen with a 17 m² feed alley, stocked with approximately 31 cows/pen. Stall stocking density was < 100%; stalls were fitted with mattress covered by 2–5 cm of sawdust and cleaned daily. The barn was equipped with a voluntary milking...
system. Cows were fed a partial mixed ration, plus a commercial pellet (approximately 4 kg/day). The mixed ration was formulated following the National Research Council recommendations (NRC, 2001) using RLM 3.3 Software. The diet was set to meet the requirements of lactating dairy cows producing at least 36 kg of milk/d. Cows were fed twice a day at approximately 08:00 and 16:00 h. Also, cows had ad libitum access to fresh water. Sample size was determined following Friedman (1982). Seventeen cows were the minimum number to detect an assumed effect size of 0.70 (correlation coefficient as a measure of effect size) for a correlation as described by Friedman (1982); power of 0.90, and a type I error probability of 0.05 (2-sided). From a herd of 62 dairy cows, 24 Holstein cows (mean ± SD; DIM: 208.78 ± 127.69; parity: 1.3 ± 0.6; and milk yield: 34.88 ± 8.66 kg/d) were selected using the DIM and lactations (primiparous and multiparous) as criteria and were divided into 2 randomly selected groups within pens.

The cows were fitted with a commercially available BMC (CowMed, Santa Maria, RS, Brazil) one week before observation started as recommended by the manufacturer as the adaptation period. The BMC is composed by a device (11.5 × 7 x 3.3 cm; 140g) + nylon band (120g) + counterweight (240g). The BMC’s life expectancy of batteries is up to 5 years. The BMC data was wirelessly transmitted hourly to a base station connected to the internet placed inside the barn. The barn base station was able to store data for up to 24 h. All BMC devices were synchronized to a local hour (GMT-03). Each enrolled cow had 2 devices within the same collar positioned longitudinally in the middle of the left side of the collar near the animal’s ear. The BMC uses a preprocessing data mechanism where the data is recorded by minute but encoded in 1 h bouts – i.e., the data cloud received the data in minutes per hour for each behavior (ruminating, feeding activity and idle time).

The observations were made into 2 periods (0700 to 1100 h, and 1400 to 1700 h) within a 24 h time frame to attempt to record a range of behaviors from diurnal variation (DeVries et al., 2003). To match the BMC data recording scheme the observers were trained to scan sample the focus cows every minute with the aid of smartphones synchronized to the same local hour (GMT-03). Five smartphones were synchronised to the same local hour (GMT-03). Each enrolled cow had 2 devices within the same collar positioned longitudinally in the middle of the left side of the collar near the animal’s ear. The BMC uses a preprocessing data mechanism where the data is recorded by minute but encoded in 1 h bouts – i.e., the data cloud received the data in minutes per hour for each behavior (ruminating, feeding activity and idle time).

The slope of the regression and Bland-Altman plots (Bland and Altman, 1986) were used to assess the accuracy for each behavior. Bland-Altman statistical results were used to obtain the mean differences of the plots. The BMC was considered accurate if the slope from the linear regressions did not differ significantly from 1, and if the 95% interval of the agreement included zero for mean bias from the Bland-Altman plots. All statistical analyses were performed in R, version 4.1.3 (https://r-project.org).

Descriptive analyses for data observed and BMCs are presented in Table 1. For the validation comparisons, the Pearson correlation coefficient was \( r = 0.50, 0.87, \) and 0.93 \( (P = 0.03) \) for rumination time, feeding activity time, and idle time, respectively. The coefficient of determination was \( R^2 = 0.25, 0.75, \) and 0.87 \( (P = 0.03) \) for rumination time, feeding activity time, and idle time, respectively. Also, Lin’s concordance correlation coefficient was \( \rho_c = 0.48, 0.86, \) and 0.63 for rumination time, feeding activity time, and idle time, respectively. Slopes of linear regressions for observations vs. BMC did not differ significantly from 1 but idle behavior. The slope of regression used to assess accuracy for observations compared with BMC was found to be 1.03 (95% CI:
0.92 – 1.14; P < 0.001) for rumination time; 0.97 (0.88 – 1.06; P < 0.001) for feeding activity time; and 1.47 (1.36 – 1.59; P < 0.001) for idle time. The Bland-Altman plot was used to assess the bias between the mean difference of observed and BMC, and the agreement interval, for rumination (Figure 1A), feeding activity (Figure 1B), and idle (Figure 1C). The BMC was found to have most cows within the 95% confidence interval agreement of the Bland-Altman plot, where cows out of the 95% interval of agreement were 2 for feeding activity. Also, all the Bland-Altman plots included zero within the confidence interval for observations compared with BCM. Mean differences were used to determine whether one measure was over or underestimating another. The results of the mean difference between observations and BMC were rumination time: 0.83 ± 4.01; feed activity time: −0.48 ± 4.15; and idle time: 7.17 ± 3.94 min/h.

For inter-device measures summarized by total time recorded per day for each of the 23 cows over 22 d, the Pearson correlation coefficient [median (Q1, Q3)] was (r) = 0.48 (0.29, 0.63), 0.53 (0.40, 0.72), and 0.53 (0.33, 0.79) for rumination, feeding time, and idle time, respectively (Figure 2). The coefficient of determination was (R^2) = 0.45, 0.60, and 0.56 (P < 0.01) for rumination time, feeding activity time, and idle time, respectively. The concordance correlation coefficient [median (Q1, Q3)] for inter-device was (ρc) = 0.38 (0.16, 0.60), 0.50 (0.13, 0.64), and 0.40 (0.24, 0.75) for rumination time, feeding time, and idle time, respectively. The Bland-Altman plot was used to assess the bias between the mean difference of BMCs, and the agreement interval, for rumination (Figure 1D), feeding activity (Figure 1E), and idle (Figure 1F). For inter-device comparison, all slopes differed from 1, but feeding time. The slope was found to be 0.96 (95% CI: 0.90 – 0.99; P = 0.02) for rumination time; 0.97 (0.89 – 1.03; P = 0.26) for feeding time; and 0.95 (0.93 – 0.99; P = 0.01) for idle time. The BMCs were found to have most cows within the 95% interval of agreement of the Bland-Altman plot. Also, all the Bland-Altman plots included zero within the interval of agreement for inter-device comparisons. Additionally, the results of the mean differences between BMCs were rumination time: −0.36 ± 2.84; feeding activity time: 0.45 ± 3.51; and idle time: −0.06 ± 2.81 min/h.

The behavior monitoring collar used in this study showed a high correlation between the BMC compared with a trained observer for feeding activity and idle behavior, but low correlations for rumination. Studies that validated other similar commercial monitoring behavior devices found comparable results to this BMC. Bikker et al. (2014), studying free-stall housed dairy cows had very high correlations for feeding and idle time. Borchers et al. (2016) when validating PLF devices, using free-stall housed dairy cows, had very high results for feeding behavior. Grinter et al. (2019) validated a very similar device under similar conditions to those of this study, showed very high results for ruminating, feeding, and resting behaviors. Overall, we deemed the BMC assessed in this study precise to measure feeding activity when compared with observations, but more refinements are needed to precisely monitor ruminating and idle time.

Accuracy has been assessed in validation studies by analyzing the slope of the regression line (Chizzotti et al., 2015; Grinter et al., 2019) and the Bland-Altman plots (Renaud et al., 2022; Cantor et al., 2022) to assess the agreement between 2 measures. Previous research suggested that accuracy is not only important for helping farmers to monitor dairy herds in real-time, but also allows data
to be compared across farms (Grinter et al., 2019). Although all Bland-Altman plots satisfied accuracy requirements, the slope of the regression line for idle time showed that the BMC overestimates idle behavior when compared with observations. We may visually assess in the Bland-Altman plot (Figure 1C) where the 0 was found to be close to the lower limit of agreement, meaning a tendency to overestimating the idle time, even though all the cows were encompassed within the 95% limit of agreement. The overestimation of idle behavior may be attributed to open-set recognition, where different activities are misclassified into known activities on algorithms trained for a limited set of behaviors (Mao et al., 2023). The BMC may account for walking, standing, lying and other activities not identified as rumination or feeding activity within the idle behavior, thus resulting in a difference between the observations and BMC. However, Bland-Altman Plots define the intervals of agreements, and it does not state whether those limits of agreement are acceptable or not (Giavarina, 2015). Thus, it is essential to take into consideration the biological aspect of the variables investigated. Future research should investigate and clarify the factors that affect PLF devices’ accuracy.

There is a lack of discussion regarding inter-device reliability in PLF devices, and to our knowledge, this is the first study investigating the inter-device precision and accuracy of a BMC for lactating dairy cows. Inter-device reliability is directly correlated to the accuracy of the data recorded. Low reliability may lead to inaccurate measures, impacting the detection of abnormal cow behavior. Pearson correlation and coefficient of determination did not meet the criteria of precision for the inter-device comparison of the BMC in this study. Furthermore, the BMC did not meet all the accuracy criteria, but there were no bias observed when evaluating the data obtained from inter-device comparison. However, an increase in variation as the time increased was observed for feeding activity time (Figure 1E). In a recent study by Benaissa et al. (2023) the integration of data collected from accelerometers and Ultra-Wideband location devices yielded improved outcomes for feeding and ruminating time when compared with the utilization of accelerometer data alone. The context-aware modeling, such as location, enables accurate categorization of behaviors, suggesting a prospective future approach for enhancing the accuracy of the BMC investigated in our current study.

Studies investigating triaxial accelerometer inter-device reliability and factors affecting data collection, demonstrated high reliability between devices exposed to different applications [veterinary use (Martin et al., 2017); human health and activity (Takaes et al., 2014; Dontje et al., 2015; Nickerson et al., 2020)]. Overall, triaxial accelerometer devices proved to have high reliability between devices in different applications, thus it may be applied for monitoring cows’ behavior such as ruminating, feeding activity and idle time. Nevertheless, we deemed the BMC to have low accuracy when comparing both devices measuring ruminating, feeding activity, and idle time. Despite of low accuracy, on field applications are based on individual machine learning, thus the BMC is applicable for monitoring animal behavior individual data and detect temporal variability, once the algorithm evaluates each cow’s average daily rate of acceleration and creates a behavior index. However, between animals, in which accuracy is demanded, future investigations are needed to improve the reliability of the BMC.

![Figure 1](image.png)

**Figure 1.** Bland-Altman plots illustrating agreement between the differences in observations (OBS) and the behavior-monitoring collar (BMC) measures for rumination (A), feeding activity (B), and idle (C) time; and the agreement between the differences in both BMCs (BMC1 – BMC2) for rumination (D), feeding activity (E), idle (F) time. The solid line indicates the mean difference between the measures and the dotted lines represent the standard deviation from the mean difference. The x-axis represents the range of the mean values between the measures. The y-axis represents the difference between the measures.
There are some limitations to be considered when evaluating the data obtained in this study. The objective of this study was to validate the BMC and its inter-device reliability of lactating dairy cows. Inter-activity similarity, which is a case where different animal behaviors have similar characteristics or movement patterns (Mao et al., 2021), such as panting and licking, may result in interference of behavior detection by the BMC. Thus, the algorithm may have classified other behaviors as one of the behaviors of interest in this study. Furthermore, this is an independent validation of the algorithm, and likely the ethogram employed in this investigation may exhibit discrepancies in comparison to the ethogram that served as the basis for developing the BMC algorithm. Future research should investigate factors affecting the validity of PLF devices analyzing larger data sets to understand the magnitude of the variability.

This study evaluated the precision, accuracy, and inter-device reliability of a commercially available BMC. Feeding activity was found to be highly correlated with observations deeming the device useful to measure feeding activity autonomously. However, while the BMC allows the collection of constant and consistent data on an individual basis it still lacks accuracy.

**References**


Figure 2. Boxplots of the Pearson correlations for inter-device comparisons (BMC1 vs. BMC2) for rumination, feeding activity, and idle behavior of 23 Holstein dairy cows within 24 h timeframe over 22 d. The the median (50th percentile) is represented by the heavy dark line within each box; lower and upper lines of the box represent the 25th and 75th percentiles, respectively. Whiskers extend to the smallest and largest non-outlier statistical values. Each dot represents and individual cow.
Lovatti et al. | Inter-device reliability of a behavior monitoring collar for lactating dairy cows


**NOTES**

We thank the Fazenda Experimental Gralha Azul team from the Pontificia Universidade Católica do Paraná and Murilo V. F. Barros for their help during data collection. We are grateful to CowMed’s team for their support during this study execution. This project was a result of a partnership among CowMed (CowMed, Santa Maria, RS, Brazil), the Pontificia Universidade Católica do Paraná, and the Dairy Science Program at the University of Kentucky. This research was partially funded by a United States Department of Agriculture NIFA Hatch Grant Project KY007100 at the University of Kentucky and partially supported by the Agriculture and Food Research Initiative Grant number 2020-07265 from the USDA National Institute of Food and Agriculture, Inter-Disciplinary Engagement in Animal Systems (IDEAS), A1261. KAD received scientific scholarship (PUCPR-PAIT 210029) from the Brazilian National Council for Scientific and Technological Development (CNPq) during the study period. The authors have not stated any conflicts of interest.

JDS Communications TBC; TBC