J. Dairy Sci. 104:4467–4477
https://doi.org/10.3168/jds.2020-19603

Validation of milk mid-infrared spectroscopy for predicting the metabolic status of lactating dairy cows in Australia

P. N. Ho,[1,*] T. D. W. Luke,[1,2] and J. E. Pryce[1,2]
[1]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia
[2]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

ABSTRACT

Increased concentrations of some serum biomarkers are known to be associated with impaired health of dairy cows. Therefore, being able to predict these biomarkers, especially in the early stage of lactation, would enable preventive management decision. Some health biomarkers may also be used as phenotypes for genetic improvement for improved animal health. In this study, we validated the accuracy and robustness of models for predicting serum concentrations of β-hydroxybutyrate (BHB), fatty acids, and urea nitrogen, using milk mid-infrared (MIR) spectroscopy. The data included 3,262 blood samples of 3,027 lactating Holstein-Friesian cows from 19 dairy herds in Southeastern Australia, collected in the period from July 2017 to April 2020. The models were developed using partial least squares regression and were validated using 10-fold random cross-validation, herd-year by herd-year external validation, and year by year validation. The coefficients of determination (R^2) for prediction of serum BHB, fatty acids, and urea obtained through random cross-validation were 0.60, 0.42, and 0.87, respectively. For the herd-year by herd-year external validation, the prediction accuracies held up comparatively well, with R^2 values of 0.49, 0.33, and 0.67 for of serum BHB, fatty acids, and urea, respectively. When the models were developed using data from a single year to predict data collected in future years, the R^2 remained comparable, however, the root mean squared errors increased substantially (4–10 times larger than compared with that of herd-year by herd-year external validation) which could be due to machine differences in spectral response, the change in spectral response of individual machines over time, or other differences associated with farm management between seasons. In conclusion, the mid-infrared equations for predicting serum BHB, fatty acids, and urea have been validated. The prediction equations could be used to help farmers detect cows with metabolic disorders in early lactation in addition to generating novel phenotypes for genetic improvement purposes.
Key words: mid-infrared spectroscopy, metabolic disorder, prediction accuracy

INTRODUCTION

Early postpartum dairy cows often enter a period of negative energy balance (Ingvartsen and Andersen, 2000), which predisposes them to metabolic and microbial diseases such as ketosis, milk fever, displaced abomasum, retained placenta, metritis, and mastitis (Collard et al., 2000; Esposito et al., 2014; Pryce et al., 2016). In fact, between 30 to 50% of cows are affected by some form of diseases in early lactation with 75% of these occurring in the first month after calving (LeBlanc, 2010). Such problems not only impair farm profitability, they also directly increase veterinary and reproductive costs (Hogeveen et al., 2011; Shalloo et al., 2014), environmental losses (Bell et al., 2013), and affect animal welfare outcomes (Oltenacu and Broom, 2010). Given the high incidence and the costs of these disorders, there has been growing interest in predicting the metabolic status of dairy cows in the early stages of lactation (see the recent review of Pralle and White, 2020). This information could either be used to help farmers make informed interventions to prevent the development of these diseases, or to generate novel phenotypes for genetic improvement purposes, most likely through genomic selection.

In dairy cows, widely used indicators of energy status include blood BHB and fatty acid concentrations, whereas blood urea nitrogen concentrations indicate the efficiency of protein utilization (Melendez et al., 2003; Wathes et al., 2007; Urdl et al., 2015). Unfortunately, taking blood samples from cows on a regular basis is not only laborious and costly, but also potentially stressful to the animals. Therefore, noninvasive alternative methods are of great interest. Milk mid-infrared (**MIR**) spectroscopy is a fast and cost-effective method that has been routinely used by milk recording organi-

Received September 8, 2020.
Accepted November 13, 2020.
*Corresponding author: phuong.ho@agriculture.vic.gov.au

zations worldwide to quantify the composition of milk samples (De Marchi et al., 2014). Furthermore, many studies have reported promising accuracies from using milk MIR to predict serum metabolic profiles, with the R^2 obtained through random cross-validation ranging between 0.21 and 0.92 (Belay et al., 2017; Grelet et al., 2018; Pralle et al., 2018; Benedet et al., 2019; Luke et al., 2019b). Although the initial results are promising, the models should be properly validated, preferably through external validation, before they can be implemented for farmers to use as management tools (Pralle and White, 2020). This is because random cross validation is often overly optimistic when compared with the more stringent external validation (i.e., using data from a different herd; Wang and Bovenhuis, 2019).

Luke et al. (2019b) investigated the prediction accuracy of models for predicting 7 metabolites measured in the serum of dairy cows using milk MIR spectroscopy using cross-validation and 2 external validation approaches. The results were promising, showing similar R^2 values for cross-validation and external validation. However, the data were limited to just one year of sampling. Before recommending the use of these predictions for any purpose, more extensive testing is warranted. Thus, the objective of this study was to examine how robust prediction accuracies of these equations are using data from multiple herds and years.

MATERIALS AND METHODS

Animal Data

Between July 2017 to April 2020, 3,262 blood samples were taken from 3,027 lactating Holstein-Friesian cows of 19 dairy herds in Southeastern Australia, according to the protocol described in Luke et al. (2019b). The cows were between first and eighth parity. At the time of sampling, all cows were between 0 and 50 d in milk, with a mean of 21.4 ± 18.7 d. The farms included in the present study operated a seasonal calving system, where the intent is for most cows to calve in a period of time to match the pasture availability. Additionally, although details of specific dietary regimens are rarely reported for commercial farms in Australia, many farms implement a feeding system reliant on grazed pasture plus other forages with some supplementary cereals often also fed. In this data set, all cows were milked twice daily in accordance with the standard commercial practices of herd-testing organizations in Australia. All procedures undertaken in this study were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (NHMRC, 2013). Approval to proceed was granted by the Agricultural Research and Extension Animal Ethics Committee of the Department of Jobs, Precincts and Resources Animal Ethics Committee (Attwood, Victoria, Australia), and the Tasmanian Department of Primary Industries, Parks, Water and Environment (Animal Biosecurity and Welfare Branch, New Town, Tasmania, Australia).

The sera were analyzed for concentrations of BHB, fatty acids [formerly referred as nonesterified fatty acids (**NEFA**)], and urea by Regional Laboratory Services (Benalla, Victoria, Australia), using the following assays: enzymatic kinetic assays for BHB (McMurray et al., 1984) and urea (Wilcox et al., 1966) and enzymatic end-point assay for fatty acids (proprietary formulation, Randox Laboratories, Crumlin, UK). Milk samples were collected either immediately before or after blood sampling or at the same time (i.e., blood samples taken in the parlor) and sent to Hico Pty Ltd. (Maffra, Victoria, Australia) or TasHerd Pty Ltd. (Hadspen, Tasmania, Australia) to analyze for fat, protein, and lactose percentages, and SCC using a NexGen Series FTS Combi machine (Bentley Instruments, Chaska, MN), from where the MIR spectra were obtained for this study. A recorded spectrum includes 899 data points, with each point representing the absorption of infrared light through the milk sample at a particular wavenumber in the 649 to 3,999 cm^{-1} region. It should be noticed that one machine from Hico and 2 from TasHerd were used for analysis of milk samples. Additionally, all milk samples that were collected in 2017 plus samples from the 2 Tasmanian herds collected in 2018 were analyzed by TasHerd, whereas the remaining milk samples were analyzed by Hico.

Data Preprocessing

Before the model development, several mathematical treatments were applied to the metabolites and the spectra. For BHB and fatty acids, because their distributions were skewed with an over-representation of the lower values, which might impair the accuracy of predicting high values in partial least square (**PLS**) regression (Grelet et al., 2016), a logarithmic (10) and a square root transformation were applied to their original records, respectively (Figure 1). For the spectra, noisy regions (1,615–1,652 cm^{-1} and 649–925 cm^{-1}) characterized by a low signal-to-noise ratio, which is the consequence of a high water absorption and noninformative region (2,998–3,998 cm^{-1}), were first removed (Hewavitharana and van Brakel, 1997). Second, to discard the spectra that are potentially outliers, a standardized Mahalanobis distance [i.e., global H distance (Shenk and Westerhaus, 1995)] between each spectrum and the population average was calculated. Then, the spectra with a global distance greater than 3 (n = 38)

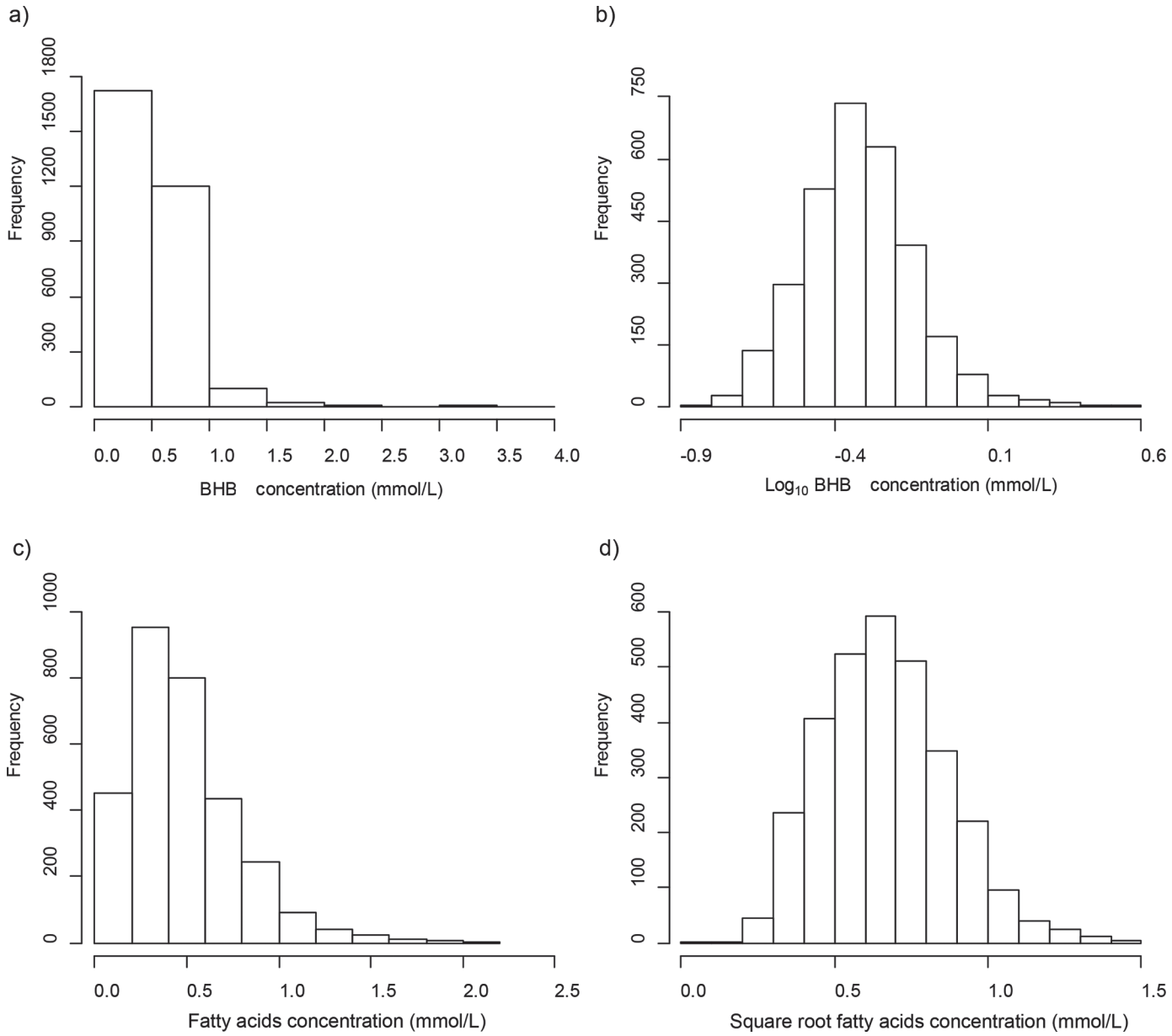

Figure 1. (a) Frequency distribution of untransformed serum BHB concentrations, (b) serum BHB concentrations following $\log_{10}$ transformation, (c) untransformed serum fatty acids, and (d) serum fatty acid concentrations following square root transformation.

were considered outliers and eliminated. Last, extended multiplicative correction (Kohler et al., 2009) and a first-order Saviztky–Golay derivative (Savitzky and Golay, 1964) were applied to the reduced spectra. A final spectrum used for model development consisted of 536 wavenumbers.

Model Development and Evaluation of Performance

Luke et al. (2019b) observed that using literature-based thresholds, to define subclinical cases of disease, the prevalence of some diseases was low, especially for hyperketonemia when using a cut-off value of $\geq$1.2 mmol/L of BHB. Additionally, the epidemiology of these diseases in the Australian pasture-based context is the focus of other research conducted by our group. Accordingly, in this study we only focused on the prediction equations for continuous traits that is predicting serum metabolite concentrations.

The prediction models for blood metabolic profiles were developed using PLS regression and implemented with the pls R package of Mevik and Wehrens (2007).

Table 1. Summary statistics (mean ± SD) of the serum metabolites (mmol/L)

Season	N[1]	BHB	Fatty acids	Urea
Spring 2017	1,041	0.53 ± 0.23	0.50 ± 0.33	5.67 ± 2.29
Autumn 2018	105	0.56 ± 0.26	0.29 ± 0.26	3.74 ± 0.78
Spring 2018	1,003	0.47 ± 0.23	0.50 ± 0.32	5.27 ± 1.47
Spring 2019	759	0.62 ± 0.47	0.47 ± 0.28	5.69 ± 2.63
Autumn 2020	354	0.58 ± 0.24	0.60 ± 0.23	4.89 ± 1.21
All data	3,262	0.54 ± 0.31	0.49 ± 0.31	5.40 ± 2.06

[1]N = number of records.

Model performance was evaluated through 10-fold random cross-validation, herd-year by herd-year external validation, and year by year validation.

1) In the 10-fold random cross-validation, the data set was first randomly split into 10 parts and then one part was reserved for validation, whereas the remaining data were used for model training. This process was repeated 10 times until each part of the data had been validated once.
2) In the herd-year by herd-year external validation, the data of a given herd-year was excluded and used as a validation of the model trained with the data of the other herd-years. The process was repeated until every herd-year had been validated once (i.e., 24 times, as there were 24 herd-years in this study). The number of herds in 2017, 2018, 2019, and 2020 were 4, 10, 8, and 2, respectively. To make sure that there was no carryover effect of cows occurring in the same herd from one year to the next, we also examined models developed using records of cows from the herds that were completely independent of the herd being validated and this produced comparable prediction accuracies to our assumption of unique herd-year.
3) The process of year by year validation was similar to herd-year by herd-year validation, but here data from 1 yr was used to validate models created using data collected in the other 2 yr.

All analyses in the present study were performed using R statistical software version 3.6.1 (R Development Core Team, 2020).

The optimal number of PLS components was determined based on first local minimum value in root mean squared error (**RMSE**) of prediction. The measures of prediction accuracy included mean bias of estimation, slope between true and predicted values, coefficient of determination (R^2) and RMSE, and ratio of performance to interquartile distance (**RPIQ**). According to Bellon-Maurel et al. (2010), RPIQ is more preferred to the ratio of performance to deviation on the skewed data.

RESULTS AND DISCUSSION

Summary statistics of the concentrations of the serum metabolites in this study are presented in Table 1. The means and standard deviations (mmol/L) of BHB, fatty acids, and urea in the pooled data set were 0.54 ± 0.31, 0.49 ± 0.31, and 5.40 ± 2.06, respectively. These results are very consistent with those reported by Luke et al. (2019b), with the means of BHB and fatty acids being identical, whereas that of urea being slightly lower in this study (5.40 vs. 5.54 mmol/L). Because the study of Luke et al. (2019b) comprised samples that were a subset of the current study, similarities in means were expected. The current data set includes data collected from more herds in different geographical locations and calving in different seasons, which we expected to capture the diverse conditions of dairy farms in Southeastern Australia. There was not much difference between calving seasons (i.e., Autumn vs. Spring) and years of collection, with the exception of Spring 2019 that had higher average concentrations of each metabolite, as well as being more variable.

When used to predict subclinical diseases, MIR models often assume arbitrary thresholds that have been derived using detailed epidemiological studies and therefore applying these thresholds requires caution and the farming system or country need to be taken into consideration. In the present study, using cut-off values of ≥1.2 mmol/L for BHB (Iwersen et al., 2009) we found only 2.9% of cows that had hyperketonemia, whereas 20.1% of cows had elevated fatty acids, defined using a threshold of ≥0.7 mmol/L (Ospina et al., 2010). However, in a study of 208 cows from 22 farms in Australia with blood samples collected earlier in lactation (i.e., between 2 and 21 DIM), Brunner et al. (2018) reported a higher concentration of BHB of 0.7 ± 0.7 and accordingly a higher prevalence of hyperketonemia incidence of 9.6%. In our study, cows were between 0 to 50 DIM with only 15% being in the first week of

lactation, which may be why there was a lower prevalence of hyperketonemia. McArt et al. (2012) showed that the peak incidence of hyperketonemia in US dairy cows occurred at d 5 postpartum. Unfortunately, in the seasonal calving systems typical in Australia and New Zealand, farmers are often reluctant to perform milk testing in very early lactation, because it is generally a busy time of year with a lot of cows calving within a few weeks. However, there are several studies that illustrate the benefits of milk samples collected in early lactation for decision making purposes [e.g., Grelet et al. (2018); Ho et al. (2019); Pralle and White (2020)]. It is conceivable that having more information in early lactation could facilitate practice change of farmers to initiate herd-testing in early lactation to access data for decision making. For urea, using the cut-off value of $\geq$6.78 mmol/L, which has been reported to reduce conception rate in dairy cows (Butler et al., 1996), we found 24.1% of cows exceeded this threshold. However, it should be noted that the concentration threshold values used here are solely based on the research undertaken in other countries, which are not necessarily true for Australian farming conditions and thus an epidemiological investigation to derive the appropriate thresholds is essential to warrant these results (Raboisson et al., 2017).

In terms of the model performance, the prediction accuracy obtained through 10-fold random cross-validation (Table 2) are comparable to those reported by Luke et al. (2019b) and the other studies in the literature (Grelet et al., 2018; Pralle et al., 2018; Bonfatti et al., 2019). More interestingly, the accuracy held up comparatively well in the external herd-year by herd-year validation. When compared with random cross validation results, the R^2 values derived from external validation dropped from 0.60 to 0.49, 0.42 to 0.33, and 0.87 to 0.67 for BHB, fatty acids, and urea, respectively. This was an important finding as some previous studies have reported a significant drop in prediction accuracy on external validation compared with random cross-validation. McParland et al. (2012), for example, indicated that the model for predicting energy balance developed using data of cows fed TMR from the Scotland's Rural College research farm did not work when applied to the data of grazing cows from the Teagasc Animal and Grassland Research and Innovation Center in Moorepark, Ireland, with the correlation coefficient dropping from 0.7 to 0.1. Furthermore, Wang and Bovenhuis (2019) reported a massive drop in the prediction accuracy of methane emissions using MIR from 0.49 to 0.01 on random cross-validation and herd-by-herd external validation, respectively. Therefore, it was very encouraging that random cross-validation results compared reasonably well to external validation in our study.

Table 2. Prediction accuracy of the models (mean ± SD) obtained through 10-fold random cross-validation and external herd-year by herd-year validation[1]

Metabolite	LV	Bias	Slope	R^2_{cv}	$RMSE_{cv}$	$RPIQ_{cv}$	Bias	Slope	R^2_{val}	$RMSE_{val}$	$RPIQ_{val}$
BHB	16	0.22 ± 0.04	0.54 ± 0.08	0.60 ± 0.05	0.20 ± 0.05	1.35 ± 0.29	0.27 ± 0.11	0.43 ± 0.12	0.48 ± 0.14	0.19 ± 0.12	1.20 ± 0.31
Fatty acids	19	0.26 ± 0.02	0.41 ± 0.03	0.42 ± 0.04	0.23 ± 0.02	1.56 ± 0.19	0.35 ± 0.29	0.38 ± 0.15	0.35 ± 0.17	0.33 ± 0.23	1.21 ± 0.47
Urea	18	0.62 ± 0.13	0.88 ± 0.02	0.87 ± 0.03	0.75 ± 0.08	3.84 ± 0.49	1.33 ± 1.06	0.75 ± 0.14	0.69 ± 0.17	0.72 ± 0.27	2.48 ± 1.04

[1]LV = number of latent variables in the model; R^2_{cv} = coefficient of determination of cross-validation; $RMSE_{cv}$ = root mean square error of cross-validation; $RPIQ_{cv}$ = ratio of performance to interquartile distance of cross-validation; R^2_{val} = coefficient of determination of external validation; $RMSE_{val}$ = root mean square error of external validation; $RPIQ_{val}$ = ratio of performance to interquartile distance of external validation.

It could be argued that the reliability of the prediction accuracy of an MIR equation is trait dependent, and it relies on the covariance structures between traits of interest and milk composition, which can be affected by breed and feed, and others (Eskildsen et al., 2014). We also attempted to investigate the possible reasons behind variation in prediction accuracy between herd-years. It was found that R^2 correlated with the standard deviation of the metabolic profiles in the validating set (i.e., the values of Pearson correlation were 0.12, 0.46, and 0.44 for BHB, fatty acids, and urea, respectively). No difference could be observed between the absorbance levels in the spectra of training and validating set. To the best of our knowledge, this is the first study to examine the robustness of MIR prediction models of blood metabolic profile data using a comparatively large number of cows and a diverse range of farms. In our study, the number of cows with serum metabolite phenotypes was 3,027, whereas this was 826 in Belay et al. (2017), 241 in Grelet et al. (2018), 1,013 in Pralle et al. (2018), and 124 in Benedet et al. (2019), respectively. Although the values of coefficient of determination looked promising, only the model for predicting urea could be claimed to be excellent (RPIQ $>$ 2) whereas the models for predicting BHB and fatty acids were nonreliable (RPIQ $<$ 1.4), according to the thresholds recommended by Chang et al. (2001).

Implementation of prediction equations requires careful consideration, including how the information will be used. Understanding the epidemiology associated with (sub)clinical disease is an important part of the process. In our study we used MIR predictions of serum metabolites, whereas others have considered the same or similar metabolites in milk. For example, in Canada, Lactanet provides MIR-predicted milk BHB for screening types of ketosis at the herd-level (i.e., Type I ketosis occurs around 2 weeks after calving, due to cows not eating enough to fulfil energy requirements, whereas Type 2 ketosis occurs in the first week after calving due to cows being over-conditioned before calving, which is often known as fat cow syndrome; Schwarz, 2019). In the Walloon region of Belgium, a package with several MIR predictions related to milk processing, such as milk technological properties and quality, is being evaluated with the intention that results are included in the milk recording reports provided to farmers (P. Delhez, Gembloux Agro-Bio Tech, University of Liège; personal communication).

Apart from management perspectives, these prediction models can also be used to generate phenotypes on a large-scale for genetic improvement of metabolic health, a trait that is difficult and expensive to measure (Pryce et al., 2016; Gengler et al., 2018; Bresolin and Dórea, 2020). Using the data of 1,393 Holstein-Friesian cows, Luke et al. (2019a) indicated that despite being lowly heritable, with genomic heritability estimates of 0.09, 0.18, and 0.18, for BHB, fatty acids, and urea, respectively, genomic selection of these traits is possible with accuracies ranging between 0.31 and 0.51. Although increasing the reference population through performing more blood sampling would be expected to improve the prediction accuracy, it is expensive, time consuming, and invasive. In this context, the use of MIR predictions as indicator traits may be a cost-effective alternative to generate many more records than can be measured, provided that milk samples can be collected in the early weeks of lactation. For example, the work of van den Berg et al. (2021) shows that for urea, MIR predictions of serum urea are highly genetically correlated with actual serum urea measurements, and therefore, either can be used for genomic prediction.

We also attempted to test the robustness of models through years (i.e., training the models using data of a single year to predict data collected in future years). The results from Table 3 imply that although the R^2 remained consistent, RMSE increased substantially compared with those reported in Table 2 (i.e., 2.12 vs. 0.21, 1.59 vs. 0.33, and 3.43 vs. 0.80 mmol/L, on average for BHB, fatty acids, and urea, respectively). Accordingly, the values of bias and slope also got worse following the direction of RMSE. The increase in RMSE is also shown in Figure 2a, 2c, and 2e, where the predicted values significantly deviated from the observed ones. Several factors could be argued to contribute to those deviations, such as machine differences in spectral response, the change in spectral response within the same machine over time, or other differences associated with management between seasons where data were collected.

In this study, the data were obtained in multiple years with milk samples collected in spring 2017, autumn 2018, and the 2 Tasmanian herds in spring 2018 were analyzed by TasHerd Pty Ltd. (Hadspen, Tasmania, Australia), whereas the milk samples collected in spring 2018 (excluding the 2 Tasmanian herds), spring 2019, and autumn 2020 were analyzed by Hico Pty Ltd. (Maffra, Victoria, Australia). As milk samples were analyzed by different machines, some differences in spectral response might be expected. In this context, analysis of identical milk samples is often recommended to standardize each machine and to overcome instrument-to-instrument variations (Grelet et al., 2017). Unfortunately, this was not possible in the present study because reference samples were not available. Alternatively, we examined the method of Bonfatti et al. (2017), which was developed to work on retrospective data, but considerable deviations persisted. Interestingly, the deviations disappeared when a randomly

Table 3. Prediction accuracy of the models trained using the data of a single year and validated on the data of the future years[1]

		BHB					Fatty acids					Urea				
Training set	Validation set	Bias	Slope	R^2	RMSE	RPIQ	Bias	Slope	R^2	RMSE	RPIQ	Bias	Slope	R^2	RMSE	RPIQ
Spring 2017	Autumn 2018	0.04	0.06	0.41	0.53	0.54	0.32	0.4	0.21	0.59	0.50	−5.02	0.89	0.76	5.71	0.30
Spring 2017	Spring 2018	0.08	0.25	0.49	0.39	0.54	1.01	0.68	0.32	0.86	0.44	−0.71	0.75	0.60	2.56	0.94
Spring 2017	Spring 2019	0.03	0.03	0.32	0.62	0.40	1.38	1.01	0.41	1.57	0.27	−5.20	0.86	0.66	5.83	0.22
Spring 2017	Autumn 2020	0.04	0.06	0.48	0.55	0.50	0.38	0.48	0.30	0.29	1.31	−4.59	0.82	0.67	5.71	0.22
Autumn 2018	Spring 2018	0.04	0.12	0.26	0.42	0.42	0.74	1.25	0.25	1.12	0.37	0.08	0.66	0.43	2.05	0.83
Autumn 2018	Spring 2019	1.44	8.42	0.11	9.05	0.10	5.14	−2.17	0.24	3.93	0.09	−4.02	0.77	0.39	5.04	0.28
Autumn 2018	Autumn 2020	3.44	9.72	0.13	8.97	0.11	7.88	−2.03	0.14	6.33	0.05	1.84	0.56	0.19	1.54	0.88
Spring 2018	Spring 2019	0.25	0.21	0.34	0.34	0.86	0.28	−0.20	0.30	0.44	0.73	2.33	0.68	0.60	0.95	1.64
Spring 2018	Autumn 2020	0.14	0.19	0.44	0.31	0.88	0.33	−0.20	0.20	0.49	0.56	7.19	0.58	0.56	4.48	0.28
Spring 2019	Autumn 2020	0.23	0.61	0.57	0.16	1.67	0.26	0.38	0.30	0.22	1.28	1.48	0.66	0.64	0.76	1.68

[1]RMSE = root mean square error; RPIQ = ratio of performance to interquartile distance.

selected herd-year in the validation season was incorporated into the training set (Table 4). For example, when training a model using the data of spring 2017 to predict the data of spring 2018, records from one herd in spring 2018 were randomly selected and incorporated into the training set. This result might be explained by the fact that the model has learned some prior information on the new machine (via the herd-year we provided) and incorporated that into the algorithm during the training process. Figure 2 presents the plots between the observed and predicted values of a selected herd-year without (Figure 2a, 2c, 2e) and with (Figure 2b, 2d, 2f) an inclusion a random herd-year from the validation season into the training set. These results probably imply that the deviations in predictions obtained here were mainly arisen from machine differences, which can be resolved by standardization (Grelet et al., 2015), rather than by feeding or management. Indeed, when we trained and validated the models using the data collected in different years, but analyzed by the same machine, no deviations could be observed (Table 5). These results also agree with the principal component analysis (data not shown), where a shift was observed if data were from different machines, but not within machine through time.

CONCLUSIONS

Using a data set obtained from a comparatively large number of cows and multiple calving seasons and herds, this study has validated the MIR prediction equations developed to predict blood BHB, fatty acids, and urea as indicators of the metabolic status of dairy cows. The results also emphasize the importance of spectral standardization among machines for application of these prediction equations to practice.

ACKNOWLEDGMENTS

This study was undertaken as part of the DairyBio program, which is jointly funded by Dairy Australia (Melbourne, Australia), the Gardiner Foundation (Melbourne, Australia), and Agriculture Victoria (Melbourne, Australia). The authors thank Di Mapleson, Brigid Ribaux, and the staff at Ellinbank Dairy Research Centre (Ellinbank, Australia) for their technical expertise and assistance, in addition to Erika Oakes and Michelle Axford of Datagene (Bundoora, Australia) for their work coordinating this study, and the farmers who participated in this project. Hico Pty Ltd. (Maffra, Victoria, Australia) and TasHerd Pty Ltd. (Hadspen, Tasmania, Australia) are gratefully acknowledged for collecting milk samples and providing MIR spectral

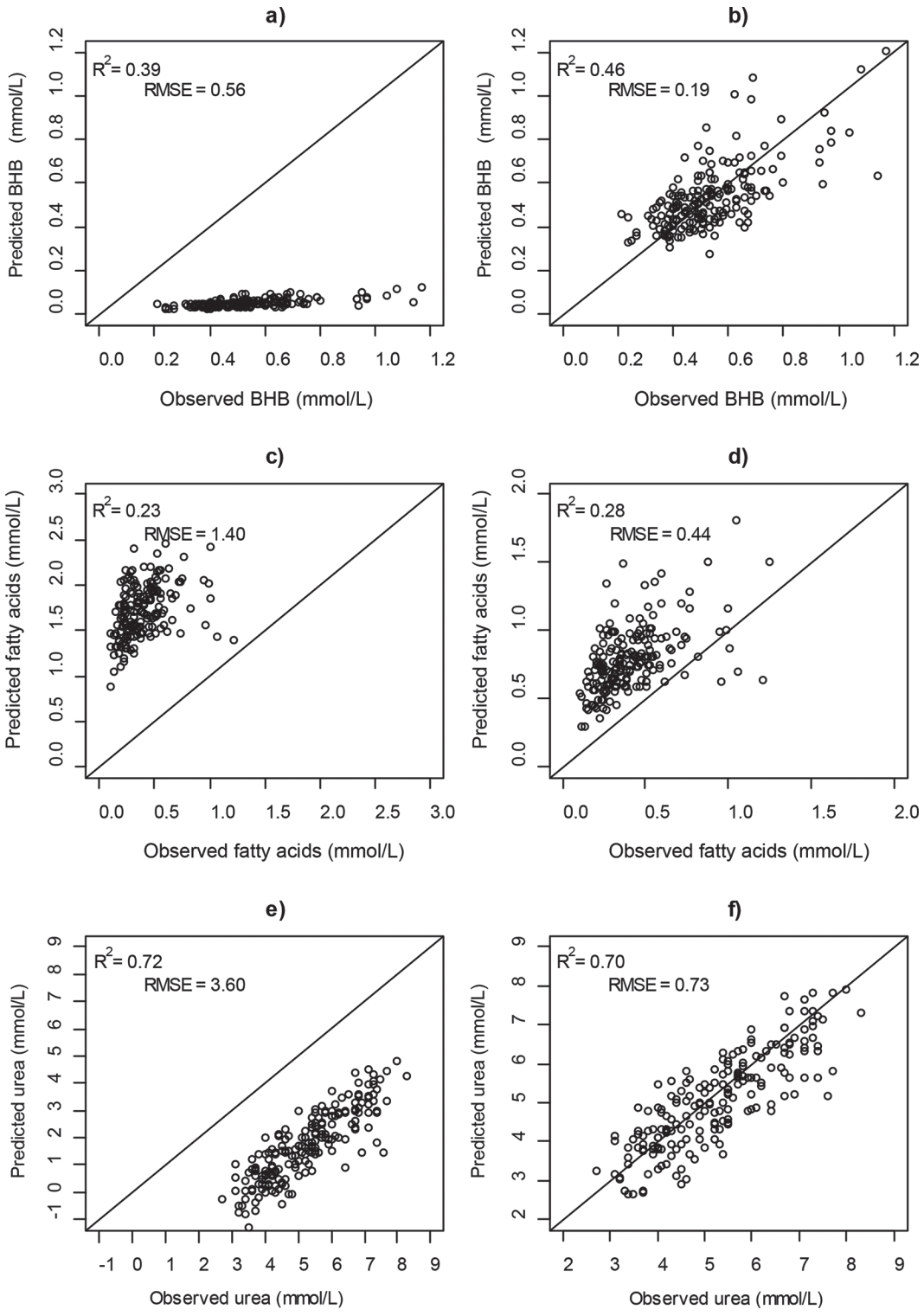

Figure 2. Plots of observed and predicted values of BHB, fatty acids, and urea for a sample herd when training the model without (a, c, e) and with (b, d, f), respectively, prior information from a randomly selected herd in the validation set. RMSE = root mean square error.

Table 4. Prediction accuracy of the models trained using the data of a single year and validated on the data of the future years, with prior information provided by a randomly selected herd from the validation set[1]

		BHB					Fatty acids					Urea				
Training set	Validation set	Bias	Slope	R^2	RMSE	RPIQ	Bias	Slope	R^2	RMSE	RPIQ	Bias	Slope	R^2	RMSE	RPIQ
Spring 2017	Autumn 2018	0.32	0.39	0.47	0.14	1.58	0.14	0.38	0.36	0.18	1.44	1.11	0.68	0.55	0.50	2.01
Spring 2017	Spring 2018	0.20	0.51	0.50	0.22	0.91	0.57	0.50	0.32	0.27	1.20	1.50	0.71	0.59	1.02	1.63
Spring 2017	Spring 2019	0.36	0.33	0.35	0.28	0.90	0.26	0.51	0.43	0.35	0.95	0.47	0.89	0.64	0.76	1.85
Spring 2017	Autumn 2020	0.34	0.37	0.45	0.18	1.46	0.33	0.46	0.33	0.19	1.47	0.94	0.80	0.65	0.76	1.71

[1]RMSE = root mean square error; RPIQ = ratio of performance to interquartile distance.

Table 5. Prediction accuracy of the models training and validating on the data collected in different years but analyzed by the same machine[1]

		BHB					Fatty acids					Urea				
Training set	Validation set	Bias	Slope	R^2	RMSE	RPIQ	Bias	Slope	R^2	RMSE	RPIQ	Bias	Slope	R^2	RMSE	RPIQ
Spring 2017	Spring 2018 – herd 1	0.22	0.54	0.54	0.11	1.41	0.16	0.44	0.33	0.29	0.98	0.84	0.72	0.71	0.91	1.43
Spring 2017	Spring 2018 – herd 2	0.23	0.70	0.65	0.16	1.33	0.29	0.58	0.15	0.27	0.40	−0.02	0.85	0.76	1.06	2.10

[1]RMSE = root mean square error; RPIQ = ratio of performance to interquartile distance.

data. The authors have not stated any conflicts of interest.

REFERENCES

Belay, T. K., B. S. Dagnachew, Z. M. Kowalski, and T. Ådnøy. 2017. An attempt at predicting blood β-hydroxybutyrate from Fourier-transform mid-infrared spectra of milk using multivariate mixed models in Polish dairy cattle. J. Dairy Sci. 100:6312–6326. https://doi.org/10.3168/jds.2016-12252.

Bell, M. J., R. J. Eckard, M. Haile-Mariam, and J. E. Pryce. 2013. The effect of changing cow production and fitness traits on net income and greenhouse gas emissions from Australian dairy systems. J. Dairy Sci. 96:7918–7931. https://doi.org/10.3168/jds.2012-6289.

Bellon-Maurel, V., E. Fernandez-Ahumada, B. Palagos, J.-M. Roger, and A. McBratney. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends Analyt. Chem. 29:1073–1081. https://doi.org/10.1016/j.trac.2010.05.006.

Benedet, A., M. Franzoi, M. Penasa, E. Pellattiero, and M. De Marchi. 2019. Prediction of blood metabolites from milk mid-infrared spectra in early-lactation cows. J. Dairy Sci. 102:11298–11307. https://doi.org/10.3168/jds.2019-16937.

Bonfatti, V., A. Fleming, A. Koeck, and F. Miglior. 2017. Standardization of milk infrared spectra for the retroactive application of calibration models. J. Dairy Sci. 100:2032–2041. https://doi.org/10.3168/jds.2016-11837.

Bonfatti, V., S. A. Turner, B. Kuhn-Sherlock, T. D. W. Luke, P. N. Ho, C. V. C. Phyn, and J. E. Pryce. 2019. Prediction of blood β-hydroxybutyrate content and occurrence of hyperketonemia in early-lactation, pasture-grazed dairy cows using milk infrared spectra. J. Dairy Sci. 102:6466–6476. https://doi.org/10.3168/jds.2018-15988.

Bresolin, T., and J. R. R. Dórea. 2020. Infrared spectrometry as a high-throughput phenotyping technology to predict complex traits in livestock systems. Front. Genet. 11:923. https://doi.org/10.3389/fgene.2020.00923.

Brunner, N., S. Groeger, J. Canelas Raposo, R. M. Bruckmaier, and J. J. Gross. 2019. Prevalence of subclinical ketosis and production diseases in dairy cows in Central and South America, Africa, Asia, Australia, New Zealand, and Eastern Europe. Transl. Anim. Sci. 3:84–92. https://doi.org/10.1093/tas/txy102.

Butler, W. R., J. J. Calaman, and S. W. Beam. 1996. Plasma and milk urea nitrogen in relation to pregnancy rate in lactating dairy cattle. J. Anim. Sci. 74:858–865. https://doi.org/10.2527/1996.744858x.

Chang, C.-W., D. A. Laird, M. J. Mausbach, and C. R. Hurburgh. 2001. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65:480–490. https://doi.org/10.2136/sssaj2001.652480x.

Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships between energy balance and health traits of dairy cattle in early lactation. J. Dairy Sci. 83:2683–2690. https://doi.org/10.3168/jds.S0022-0302(00)75162-9.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171–1186. https://doi.org/10.3168/jds.2013-6799.

Eskildsen, C. E., M. A. Rasmussen, S. B. Engelsen, L. B. Larsen, N. A. Poulsen, and T. Skov. 2014. Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables. J. Dairy Sci. 97:7940–7951. https://doi.org/10.3168/jds.2014-8337.

Esposito, G., P. C. Irons, E. C. Webb, and A. Chapwanya. 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. Anim. Reprod. Sci. 144:60–71. https://doi.org/10.1016/j.anireprosci.2013.11.007.

Gengler, N., A. Mineur, S. Vanderick, and H. Hammami.GplusE Consortium. 2018. Milk mid-infrared spectra based biomarkers contributing to genetic improvement for udder health, fertility and longevity. ICAR Tech. Ser. 23:221-227.

Grelet, C., C. Bastin, M. Gelé, J. B. Davière, M. Johan, A. Werner, R. Reding, J. A. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. J. Dairy Sci. 99:4816–4825. https://doi.org/10.3168/jds.2015-10477.

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci. 98:2150–2160. https://doi.org/10.3168/jds.2014-8764.

Grelet, C., J. A. F. Pierna, P. Dardenne, H. Soyeurt, A. Vanlierde, F. Colinet, C. Bastin, N. Gengler, V. Baeten, and F. Dehareng. 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci. 100:7910–7921. https://doi.org/10.3168/jds.2017-12720.

Grelet, C., A. Vanlierde, M. Hostens, L. Foldager, M. Salavati, K. L. Ingvartsen, M. Crowe, M. T. Sorensen, E. Froidmont, C. P. Ferris, C. Marchitelli, F. Becker, T. Larsen, F. Carter, and F. Dehareng. 2018. Potential of milk mid-IR spectra to predict metabolic status of cows through blood components and an innovative clustering approach. Animal 13:649–658.

Hewavitharana, A., and B. van Brakel. 1997. Fourier transform infrared spectrometric method for the rapid determination of casein in raw milk. Analyst (Lond.) 122:701–704. https://doi.org/10.1039/a700953d.

Ho, P. N., V. Bonfatti, T. D. W. Luke, and J. E. Pryce. 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 102:10460–10470. https://doi.org/10.3168/jds.2019-16412.

Hogeveen, H., K. Huijps, and T. Lam. 2011. Economic aspects of mastitis: New developments. N. Z. Vet. J. 59:16–23. https://doi.org/10.1080/00480169.2011.547165.

Ingvartsen, K. L., and J. B. Andersen. 2000. Integration of Metabolism and Intake Regulation: A Review Focusing on Periparturient Animals. J. Dairy Sci. 83:1573–1597. https://doi.org/10.3168/jds.S0022-0302(00)75029-6.

Iwersen, M., U. Falkenberg, R. Voigtsberger, D. Forderung, and W. Heuwieser. 2009. Evaluation of an electronic cowside test to detect subclinical ketosis in dairy cows. J. Dairy Sci. 92:2618–2624. https://doi.org/10.3168/jds.2008-1795.

Kohler, A., M. Zimonja, V. Segtnan, and H. Martens. 2009. 2.09 – Standard normal variate, multiplicative signal correction and extended multiplicative signal correction preprocessing in biospectroscopy. Pages 139–162 in Comprehensive Chemometrics. S. D. Brown, R. Tauler, and B. Walczak, ed. Elsevier, Oxford, UK.

LeBlanc, S. 2010. Monitoring metabolic health of dairy cattle in the transition period. J. Reprod. Dev. 56:S29–S35.

Luke, T. D. W., T. T. T. Nguyen, S. Rochfort, W. J. Wales, C. M. Richardson, M. Abdelsayed, and J. E. Pryce. 2019a. Genomic prediction of serum biomarkers of health in early lactation. J. Dairy Sci. 102:11142–11152. https://doi.org/10.3168/jds.2019-17127.

Luke, T. D. W., S. Rochfort, W. J. Wales, V. Bonfatti, L. Marett, and J. E. Pryce. 2019b. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. J. Dairy Sci. 102:1747–1760. https://doi.org/10.3168/jds.2018-15103.

McArt, J. A. A., D. V. Nydam, and G. R. Oetzel. 2012. Epidemiology of subclinical ketosis in early lactation dairy cattle. J. Dairy Sci. 95:5056–5066. https://doi.org/10.3168/jds.2012-5443.

McMurray, C. H., W. J. Blanchflower, and D. A. Rice. 1984. Automated kinetic method for D-3-hydroxybutyrate in plasma or serum. Clin. Chem. 30:421–425. https://doi.org/10.1093/clinchem/30.3.421.

McParland, S., G. Banos, B. McCarthy, E. Lewis, M. P. Coffey, B. O'Neill, M. O'Donovan, E. Wall, and D. P. Berry. 2012. Validation of mid-infrared spectrometry in milk for predicting body energy

status in Holstein-Friesian cows. J. Dairy Sci. 95:7225–7235. https://doi.org/10.3168/jds.2012-5406.
Melendez, P., A. Donovan, J. Hernandez, J. Bartolome, C. A. Risco, C. Staples, and W. W. Thatcher. 2003. Milk, plasma, and blood urea nitrogen concentrations, dietary protein, and fertility in dairy cattle. J. Am. Vet. Med. Assoc. 223:628–634. https://doi.org/10.2460/javma.2003.223.628.
Mevik, B.-H., and R. Wehrens. 2007. The pls package: Principal component and partial least squares regression in R. J. Stat. Soft. 18:23.
NHMRC (National Health and Medical Research Council). 2013. Australian Code for the Care and Use of Animals for Scientific Purposes. 8th ed. National Health and Medical Research Council.
Oltenacu, P. A., and D. M. Broom. 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. Anim. Welf. 19:39–49.
Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010. Evaluation of nonesterified fatty acids and β-hydroxybutyrate in transition dairy cattle in the northeastern United States: Critical thresholds for prediction of clinical diseases. J. Dairy Sci. 93:546–554. https://doi.org/10.3168/jds.2009-2277.
Pralle, R. S., K. W. Weigel, and H. M. White. 2018. Predicting blood β-hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network. J. Dairy Sci. 101:4378–4387. https://doi.org/10.3168/jds.2017-14076.
Pralle, R. S., and H. M. White. 2020. Symposium review: Big data, big predictions: Utilizing milk Fourier-transform infrared and genomics to improve hyperketonemia management. J. Dairy Sci. 103:3867–3873. https://doi.org/10.3168/jds.2019-17379.
Pryce, J. E., K. L. Parker Gaddis, A. Koeck, C. Bastin, M. Abdelsayed, N. Gengler, F. Miglior, B. Heringstad, C. Egger-Danner, K. F. Stock, A. J. Bradley, and J. B. Cole. 2016. Invited review: Opportunities for genetic improvement of metabolic diseases. J. Dairy Sci. 99:6855–6873. https://doi.org/10.3168/jds.2016-10854.
R Development Core Team. 2020. The GNU Project. The R Project for Statistical Computing. Accessed Jan. 4, 2020. http://www.rproject.org/.
Raboisson, D., A. Albaaj, G. Nonne, and G. Foucras. 2017. High urea and pregnancy or conception in dairy cows: A meta-analysis to define the appropriate urea threshold. J. Dairy Sci. 100:7581–7587. https://doi.org/10.3168/jds.2016-12009.
Savitzky, A., and M. J. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627–1639. https://doi.org/10.1021/ac60214a047.
Schwarz, D. 2019. New innovations for dairy herd management based on milk analysis. Proceedings of Herd 19 Conference. Bendigo Victoria, Australia. Datagene, Bundoora, Australia.
Shalloo, L., A. Cromie, and N. McHugh. 2014. Effect of fertility on the economics of pasture-based dairy systems. Animal 8(Suppl 1):222–231. https://doi.org/10.1017/S1751731114000615.
Shenk, J. S., and M. O. Westerhaus. 1995. Forage analysis by near infrared spectroscopy. Pages 111–120 in Forages. Vol. II. The Science of Grassland Agriculture. 5th ed. R. F. Barnes, D. A. Miller, and C. J. Nelson, ed., Iowa State University Press, Ames, IA.
Urdl, M., L. Gruber, W. Obritzhauser, and A. Schauer. 2015. Metabolic parameters and their relationship to energy balance in multiparous Simmental, Brown Swiss and Holstein cows in the periparturient period as influenced by energy supply pre- and post-calving. J. Anim. Physiol. Anim. Nutr. (Berl.) 99:174–189. https://doi.org/10.1111/jpn.12178.
van den Berg, I., P. N. Ho, T. D. W. Luke, M. Haile-Mariam, S. Bolormaa, and J. E. Pryce. 2021. The use of milk mid-infrared spectroscopy to improve genomic prediction accuracy of serum biomarkers. J. Dairy Sci. 104:2008–2017. https://doi.org/10.3168/jds.2020-19468.
Wang, Q., and H. Bovenhuis. 2019. Validation strategy can result in an overoptimistic view of the ability of milk infrared spectra to predict methane emission of dairy cattle. J. Dairy Sci. 102:6288–6295. https://doi.org/10.3168/jds.2018-15684.
Wathes, D. C., M. Fenwick, Z. Cheng, N. Bourne, S. Llewellyn, D. G. Morris, D. Kenny, J. Murphy, and R. Fitzpatrick. 2007. Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. Theriogenology 68(Suppl 1):S232–S241. https://doi.org/10.1016/j.theriogenology.2007.04.006.
Wilcox, A. A., W. E. Carroll, R. E. Sterling, H. A. Davis, and A. G. Ware. 1966. Use of the Berthelot reaction in the automated analysis of serum urea nitrogen. Clin. Chem. 12:151–157. https://doi.org/10.1093/clinchem/12.3.151.

ORCIDS

P. N. Ho https://orcid.org/0000-0001-9481-9612
T. D. W. Luke https://orcid.org/0000-0001-5070-2332
J. E. Pryce https://orcid.org/0000-0002-1397-1282