



Predicting dry matter intake in mid-lactation Holstein cows using point-in-time data streams available on dairy farms

W. E. Brown,¹ M. J. Caputo,¹ C. Siberski,² J. E. Koltes,² F. Peñagaricano,¹ K. A. Weigel,¹ and H. M. White^{1*}

¹Department of Animal and Dairy Sciences, University of Wisconsin-Madison 53706

²Department of Animal Science, Iowa State University, Ames 50011

ABSTRACT

Quantifying dry matter intake (DMI) in lactating dairy cows is important for determining feed efficiency; however, there are no methods for economically quantifying individual cow DMI on dairy farms where cows are group-fed. Attempts have been made to model DMI using cow factors, milk production, milk infrared spectra, and behavioral sensors with reasonable success. Other data streams are available on the farm that may contribute to DMI predictions. In this study, our objective was to model DMI with multiple linear regression using data from a single point-in-time that can easily be accessed on-farm. Candidate predictor variables included cow descriptors, milk yield and composition, milk fatty acid profile, and production and efficiency predicting transmitting abilities (PTA). Observations of DMI were obtained from 350 cows across 6 cohorts using individual feed bunks. The cow to bunk ratio was 2:1, with an overall bunk occupation rate of 32% throughout the day. The following models were developed sequentially with milk data obtained from a single morning milking and other data from the same day: model B (production, metabolic body weight, body condition score, lactation category, and week of lactation), model BC [model B + fatty acid (FA) content], model BY (model B + FA yield), model BPE (model B + production and efficiency PTA), model BYP (model BY + production PTA), model BYE (model BY + efficiency PTA), and model BYPE (model BY + production and efficiency PTA). Outcome variables predicted in these models were the DMI on the previous day or current day relative to the morning milk sample. The predictions for DMI on the previous day outperformed current day DMI in every model for which they were both determined. Addition of milk FA and PTA as candidate predictor variable types

to the models resulted in enhanced predictive ability, with incremental enhancements when combined. The most robust model (BYPE) included cow descriptors, protein and FA yields, and PTA for milk and residual feed intake. Model BYPE described 21 to 32% more of the variation in DMI (based on concordance correlation coefficient) than when other common DMI models were applied to the same data set. Overall, reasonable performance of models including single point-in-time cow descriptors, milk and FA production, and production and efficiency PTA commonly available to dairy farmers through dairy herd improvement programs offer an opportunity for on-farm prediction of DMI, yet further improvement may be possible.

Key words: feed intake, multiple linear regression, prediction models, feed efficiency, milk fatty acids

INTRODUCTION

Feed is the largest input expense on dairy farms (USDA-ERS, 2021), and seeking methods to reduce feed cost through greater feed efficiency are warranted. There are many ways to calculate or assess feed efficiency for the purpose of farm-level evaluation (VandeHaar et al., 2016); however, identifying feed efficiency at the individual cow-level is often limited to research farms, where accurate recording of daily feed intake is conducted via electronic or manual methods, often requiring individual feeding stations. These labor-intensive and costly feeding systems are not practical on farms employing group feeding strategies, which represents most cows in the United States. Therefore, a need exists to develop accurate methods of cow-level DMI prediction using readily available on-farm data streams.

Recent modeling work has demonstrated marginal to good ability to predict individual cow DMI from data readily available on dairy farms (Shetty et al., 2017; Dórea et al., 2018; Martin et al., 2021a). These models have been generated primarily using cow descriptive variables, milk production, milk mid-infrared

Received November 30, 2021.

Accepted July 21, 2022.

*Corresponding author: heather.white@wisc.edu

spectra, and behavioral sensors. A recent publication added prediction of limited milk fatty acids (FA) from milk mid-infrared spectra as candidate predictor variables (Tedde et al., 2021). In many cases, data used to generate DMI prediction models were developed from data averaged over a study period (Shetty et al., 2017; Dórea et al., 2018; Martin et al., 2021a). Although reducing the number of available samples may (Dórea et al., 2018) or may not (McParland et al., 2014) reduce prediction accuracy, collecting vast quantities of additional data are likely not feasible on dairy farms due to expense and labor. Farms that are enrolled in DHI programs have access to routine monthly milk sample analysis, which may include detailed reports of milk FA profile. Additional performance indicators, such as previous lactation performance and PTA, provide potential data for model development. Specifically, the recently released Feed Saved PTA (CDCB, 2020) provides the first estimate of the genetic predisposition for feed intake and feed efficiency of US Holstein cattle. Milk FA have been explored minimally (Tedde et al., 2021), whereas previous lactation performance and cow PTA have not yet been explored as potential predictor variables to enhance DMI predictions.

Efforts to model DMI have relied on multiple linear regression (Wallén et al., 2018; Martin et al., 2021a) and partial least squares regression (McParland et al., 2014; Shetty et al., 2017; Martin et al., 2021a). More recently, the machine-learning techniques of artificial neural networks (Dórea et al., 2018; Martin et al., 2021a; Tedde et al., 2021) and stacked ensemble (Martin et al., 2021a) have been tested. Artificial neural networks have an advantage in their ability to determine more intricate relationships between predictor and response variables, which could prove useful in situations where DMI cannot be modeled as a linear function. Nonetheless, similar results have been achieved in modeling DMI between artificial neural networks and linear regression approaches (Dórea et al., 2018; Martin et al., 2021a), and linear regression models are easier to implement computationally in real-world settings.

In this paper, we explore the possibility of multiple linear regression prediction of DMI using single point-in-time data collection from 1 milking (milk and component yields, FA profile), cow descriptors (lactation category, week of lactation, BW, and so on), previous lactation performance, and PTA for production and efficiency as candidate predictor variables. This research aimed to develop a suite of models for applied use on dairy farms to predict DMI, where specific models can be selected depending on the data streams available on each farm. Using data averaged over a period of time, other DMI prediction models have explained 70 to 80%

of the variation in DMI (R^2), including up to 78% for cross-validation data (Shetty et al., 2017; Dórea et al., 2018; Martin et al., 2021a). Our objective was to develop a DMI prediction model using data from only a single time point, but including novel predictors (i.e., milk FA and PTA) to achieve model performance as previously realized.

MATERIALS AND METHODS

Animal handling and sampling procedures were approved by the Animal Care and Use Committees of the University of Wisconsin-Madison College of Agricultural and Life Sciences and Iowa State University for the respective data sets generated at each institution.

Mid-lactation Holstein cows [$n = 315$; 111 ± 23 DIM (\pm SD) at sampling; 2.4 ± 1.3 lactations] in 6 cohorts spanning from July 2018 to December 2019 were housed in sand-bedded freestalls at the University of Wisconsin-Madison Emmons Blaine Dairy Cattle Research Center (Arlington, WI) as part of a national feed efficiency project. In total, there were 350 individual cow observations, given that some cows were enrolled in multiple cohorts. No treatments of any kind were applied to the animals. Cows were provided ad libitum access to a common diet within each cohort (Table 1), formulated to meet or exceed requirements, and milked twice daily, as described previously (Martin et al., 2021a,b). Briefly, the diet was mixed once daily, and feed was administered over 2 separate feedings into roughage intake control electronic feeding bins (Hokofarm Group), with target refusals of 10%. Cows had unlimited access to all feeding bins, with an approximate ratio of cows to bins of 2:1. Average feed bunk occupancy was 32.1% across all cohorts on the day of DMI observation. A bunk occupancy rate $\geq 90\%$ was only recorded for 33 min, corresponding to the time of feeding. Daily feed intake was monitored electronically throughout the cohort data collection period. Milk weight was recorded electronically for each cow at each milking, and milk samples were obtained weekly during 4 consecutive milkings. Samples were sent to a commercial laboratory (AgSource, Menominee, WI) for Fourier-transformed infrared spectrometric analysis of milk fat, true protein, lactose, and MUN using the FOSS MilkoScan FT6000 (FOSS Analytical). Somatic cell count was determined using a Fossomatic FC (FOSS Analytical). Body weight was measured on 3 consecutive days in the first, fourth, and seventh weeks of the data collection period, and BCS was recorded by 3 trained individuals on the same weeks, using a scale from 1 to 5, where 1 = thin and 5 = fat, with increments of 0.25 (Wildman et al., 1982). Previous lactation 305 mature-equivalent yield

Table 1. Parity breakdown, nutrient composition, and dietary ingredients of the diets offered to mid-lactation Holstein cows across the cohorts, comprising the training and validation data sets

Item	Training, cohort						Validation ¹
	1	2	3	4	5	6	
Cows							
Multiparous, n	32	35	48	39	50	45	20
Primiparous, n	30	24	13	15	10	9	32
Chemical composition ²							
DM, %	52.3	51.4	52.4	49.3	52.6	58.6	54.5
CP, %	17.3	18.1	16.9	17.4	17.6	20.4	15.4
NDF, %	29.3	30.8	25.5	30.1	29.6	28.5	29.0
ADF, %	21.8	22.1	17.8	21.3	21.7	21.5	21.0
Lignin, %	3.4	3.1	2.9	3.5	4.6	4.3	12.1
Starch, %	25.5	25.0	29.6	23.4	24.6	21.0	26.2
Fat, %	4.2	4.1	4.9	4.5	4.8	4.6	4.7
Ash, %	6.1	7.0	7.7	8.0	7.8	9.9	8.9
NE _L , Mcal/kg	1.71	1.69	1.74	1.67	1.65	1.61	1.70
Ingredient							
Corn silage	27.5	28.3	26.7	30.1	28.0	15.4	42.7
Alfalfa haylage	20.0	22.2	20.7	22.5	22.7	23.6	14.5
High-moisture corn	16.6	14.5	16.1	11.5	—	—	—
Shelled corn	—	—	—	—	12.9	—	—
Ground corn	—	—	—	—	—	12.0	—
Cottonseed	4.5	4.9	4.9	5.0	5.0	3.7	—
DDGS ³	3.3	1.5	2.9	2.5	2.3	1.7	—
Concentrate mix	27.8	28.3	28.7	28.3	29.1	43.7	42.8

¹Iowa State University external data set.

²Percent of total DM unless otherwise noted.

³DDGS = dried distillers grains with solubles.

of milk, fat, and protein, in addition to PTA for milk, fat, protein, and fluid merit, were obtained in May 2021 (AgSource). The genomic PTA for Feed Saved, residual feed intake (**RFI**), and BW composite were also obtained in May 2021 from the Council on Dairy Cattle Breeding, with data corresponding to the April 2021 official evaluations (CDCB, 2021).

Data Selection, Calculations, and Cleaning

For the purpose of this study, to model DMI on a single day using readily available data streams available on dairy farms, milk yield and composition data were used from 1 morning milking at the mid-point of the data collection period (approximately 23 d after initiation of data collection) to mimic monthly DHI testing procedures. Body weight and BCS data were also available for the same sample day. The dependent variable was DMI, which was modeled for the calendar day in which the milk sample was obtained (current day DMI; **CDI**), in addition to the day prior (previous day DMI; **PDI**). The candidate predictor variables (independent variables) and the models in which they were tested are outlined in Table 2, and they represent variables that can readily be obtained on-farm through routine management practices and DHI data streams.

These variables included descriptive cow factors (i.e., lactation category, week of lactation, BW, BCS), milk and component yield, milk FA content, and PTA.

Metabolic BW (**MBW**) was calculated as $MBW = BW^{0.75}$. Milk energy output was calculated as follows according to the NRC (2001; equation 2–15), with the CP coefficient adjusted for true protein:

$$\begin{aligned} \text{Milk energy output} &= (9.29 \times \text{fat, kg}) \\ &+ (5.63 \times \text{true protein, kg}) + (3.95 \times \text{lactose, kg}). \end{aligned}$$

Models in this paper were compared with the DMI prediction of the NRC (NRC, 2001) and de Souza et al. (2019). The seventh edition of the Nutrient Requirements of Dairy Cattle DMI equation (equation 1–2; NRC, 2001) requires daily milk and component yields for calculation of FCM; therefore, estimated daily yields of milk and milk fat were calculated using adjustments for milking interval and stage of lactation (ICAR, 2017; equations 1–6). The DMI prediction equation by de Souza et al. (2019; equation 6) uses milk energy output (Mcal/d) as an input variable based on daily yield of milk components, so data were calculated according to adjustments outlined by ICAR (2017). Protein and lactose are assumed to be similar between milkings

Table 2. Dependent variables and independent candidate variables used in stepwise prediction of DMI using a single milk sample analysis from a morning milking

Variable	Model ¹
Dependent variable ²	
CDI	B, BC, BY
PDI	B, BC, BY, BPE, BYP, BYE, BYPE
Candidate predictor variable ³	
Lactation category ⁴	B, BC, BY, BPE, BYE, BYPE
Week of lactation	B, BC, BY, BPE, BYP, BYE, BYPE
MBW	B, BC, BY, BPE, BYP, BYE, BYPE
BCS	B, BC, BY, BPE, BYP, BYE, BYPE
Milk yield, kg	B, BC, BY, BPE, BYP, BYE, BYPE
Fat, %	B, BY, BPE, BYP, BYE, BYPE
Fat yield, kg	B, BC
Protein, %	B, BC, BY, BPE, BYP, BYE, BYPE
Protein yield, kg	B, BC, BY, BPE, BYP, BYE, BYPE
MUN, mg/dL	B, BC, BY, BPE, BYP, BYE, BYPE
SCC, 1,000 cells/mL	B, BC, BY, BPE, BYP, BYE, BYPE
De novo FA, %	BC
De novo FA yield, g	BY, BPE, BYP, BYE, BYPE
Mixed FA, %	BC
Mixed FA yield, g	BY, BPE, BYP, BYE, BYPE
Preformed FA, %	BC
Preformed FA yield, g	BY, BPE, BYP, BYE, BYPE
SCFA, %	BC
SCFA yield, g	BY, BPE, BYP, BYE, BYPE
MCFA, %	BC
MCFA yield, g	BY, BPE, BYP, BYE, BYPE
LCFA, %	BC
LCFA yield, g	BY, BPE, BYP, BYE, BYPE
C14, %	BC
C14 yield, g	BY, BPE, BYP, BYE, BYPE
C16, %	BC
C16 yield, g	BY, BPE, BYP, BYE, BYPE
C16, %	BC
C16 yield, g	BY, BPE, BYP, BYE, BYPE
C18, %	BC
C18 yield, g	BY, BPE, BYP, BYE, BYPE
C18:1, %	BC
C18:1 yield, g	BY, BPE, BYP, BYE, BYPE
PTA candidate predictor variable	
Milk	BPE, BYP, BYPE
Fat	BPE, BYP, BYPE
Protein	BPE, BYP, BYPE
Residual feed intake	BPE, BYE, BYPE
BW composite	BPE, BYE, BYPE

¹B = candidate predictor variables included milk yield and components, MBW, BCS, lactation category, and week of lactation; BC = model B candidate predictor variables + milk fatty acid content; BY = model B candidate predictor variables + fatty acid yield; BPE = model B candidate predictor variables + PTA for production and efficiency; BYP = model BY candidate predictor variables + PTA production; BYE = model BY candidate predictor variables + efficiency; BYPE = model BY candidate predictor variables + PTA in BYP and BYE.

²Day for which DMI was predicted relative to the milking in which the morning milk sample was obtained; PDI = previous day intake relative to morning milk sample; CDI = current day intake relative to morning milk sample.

³MBW = metabolic BW; FA = fatty acid; SCFA = short-chain FA; MCFA = medium-chain FA; LCFA = long-chain FA.

⁴Primiparous or multiparous.

(ICAR, 2017); therefore, the content of protein and lactose in the morning milking were used to estimate daily protein and lactose yields.

One cow was excluded from the data set due to treatment for respiratory disease and concomitant low levels of DMI and milk production. All cows in the final data set had complete records for each candidate predictor variable.

Model Fitting

Multiple linear regression was performed using a stepwise approach in JMP Pro, version 15 (SAS Institute Inc.). Both forward and backward stepwise approaches were fitted automatically for the dependent variable (Allen et al., 2019), and the model with the lowest Akaike information criterion was chosen for further evaluation. Next, the variance inflation factor was determined for each independent variable remaining in the model to assess collinearity, as previously described (Chandler et al., 2018; Pralle et al., 2018; Allen et al., 2019). Briefly, variables with variance inflation factor ≥ 5 were removed from the model one at a time, and the model was refitted until all variance inflation factors were < 5 . Removal of predictor variables can alter the significance of other predictor variables; therefore, resulting nonsignificant variables were removed, and freedom of multicollinearity in the final model was confirmed a final time. Models in which co-derived variables were tested within the same model (i.e., protein content and yield) were additionally screened before model fitting to ensure absence of collinearity. The correlation was -0.04 for protein content and yield, and 0.48 for fat content and yield (model B only).

Several groups of models were tested and are described further in Table 3. Models were used to predict PDI and CDI relative to the morning milking in which the milk sample was obtained. The prediction of CDI was eventually terminated for models with more candidate predictor variables due to considerably poorer performance compared with PDI in the initial models tested. Model B (basic) evaluated the prediction of PDI and CDI relative to the morning milking in which the milk sample was obtained using milk macro-component content and yield, lactation category (primiparous vs. multiparous), week of lactation, and MBW and BCS as candidate predictor variables. Model BC (basic + milk FA content; **BC**) evaluated the addition of content of milk FA (g/100 g of milk) to model B, as candidate predictor variables to predict both PDI and CDI relative to the morning milking, from which the sample was obtained. Model **BY** (basic + milk FA yield) evaluated the addition of milk FA yield to model B as candidate predictor variables to predict both PDI and CDI. Model **BPE** (basic + production and efficiency PTA) added to model B for predicting only PDI. Model **BYP** (basic + milk FA yield + production PTA) built upon BY for predicting PDI with the addition of PTA for milk, protein, and fat. Model **BYE** (basic + milk FA yield + efficiency PTA) added to BY for predicting PDI, with the addition of PTA for RFI and BW composite as candidate predictor variables to the model. Model **BYPE** (basic + milk FA yield + production PTA +

Table 3. Description of types of candidate predictor variables included in each model set

Model ID ¹	Model constituent ²
B	Milk yield and components MBW and BCS
BC	Lactation category and week of lactation Milk yield and components MBW and BCS
BY	Lactation category and week of lactation +Fatty acid content Milk yield and components MBW and BCS
BPE	Lactation category and week of lactation +Fatty acid yield Milk yield and components MBW and BCS
BYP	Lactation category and week of lactation +Production and efficiency PTA Milk yield and components MBW and BCS
BYE	Lactation category and week of lactation +Fatty acid yield +Efficiency PTA Milk yield and components MBW and BCS
BYPE	Lactation category and week of lactation +Fatty acid yield +Production and efficiency PTA Milk yield and components MBW and BCS

¹B = candidate predictor variables included milk yield and components, MBW, BCS, lactation category, and week of lactation; BC = model B candidate predictor variables + milk fatty acid content; BY = model B candidate predictor variables + fatty acid yield; BPE = model B candidate predictor variables + PTA for production and efficiency; BYP = model BY candidate predictor variables + PTA production; BYE = model BY candidate predictor variables + efficiency; BYPE = model BY candidate predictor variables + PTA in BYP and BYE.

²MBW = metabolic BW.

efficiency PTA) added to BY for predicting PDI with the addition of all PTA factors from BYP and BYE as candidate predictor variables to the model. Descriptive statistics of all variables included in the models are shown in Table 4. The random effect of cohort was tested in each model, and subsequently removed when $P > 0.10$.

The final models were evaluated using 5-fold cross-validation R^2 , the concordance correlation coefficient (CCC; Lin, 1989), root mean squared error of prediction (RMSEP), and slope and mean biases as a percent of the mean square error of the prediction (Bibby and Toutenburg, 1977) using the Model Evaluation System (Mathematical Nutrition Models). To assess relative performance between models, the R^2 , CCC, and RMSEP for each of the 5-fold cross-validation testing data sets were calculated (R Core Team, version 4.0.3). The resulting parameters were compared between models

using a paired t -test (Microsoft Excel) with fold as the blocking factor.

The overall best model was further evaluated by leave-one-cohort-out cross validation. In this case, 1 cohort of the 6 was left out of the stepwise model development process using the same candidate predictor variables in the BYPE. The predicted values for the remaining cohort were then compared with the observed DMI, and model precision and accuracy were evaluated. The process was repeated for each cohort in the data set. An external data set from Iowa State University (Ames, IA) was used for validation of model BYPE (Validation Set; Siberski-Cooper et al., 2022). Cows ($n = 52$; 32 primiparous and 20 multiparous) were fed individually using a Calan Broadbent Feeding System (1 cow per bunk; American Calan). Diet chemical composition and ingredients are presented in Table 1. Milk weights were recorded electronically (BouMatic, LLC), and milk samples analyzed in the same manner and facility as those in the training data set (AgSource). The PTA data were also obtained (CDCB, 2021) using the April 2021 genetic evaluation. Descriptive statistics for the external data set are in Table 5. Finally, the DMI models from the NRC (2001) and de Souza et al. (2019) were compared with PDI to compare model performance with BYPE from the current study.

A secondary validation data set was obtained internally from 3 cohorts in which data were collected in 2021 at the University of Wisconsin-Madison dairy herd. The purpose of this data set was to assess whether RFI PTA is artificially inflating predictive ability in our training data set, given that the PTA is generated using the RFI phenotypes from the training data set (among thousands of other observations). Data for this validation set was only used if a cow had not previously been enrolled in a University of Wisconsin-Madison feed efficiency data collection period, and RFI phenotypes were not yet added to the national database. Therefore, a subset of cows for each of the 3 cohorts was used. The RFI PTA was obtained from the April 2021 genetic evaluation. Cows were handled and samples collected in a similar manner as described above.

RESULTS AND DISCUSSION

In this study, we sought to develop feed intake prediction models leveraging existing and easily accessible data streams. We included candidate predictor variables such as cow descriptive factors and milk components that have traditionally been used for DMI model development (Holter et al., 1997; Martin et al., 2021a; NRC, 2001), but also incorporated novel predictors like milk FA composition and yield, and PTA for production and efficiency (Tables 2, 3, and 4). Also

Table 4. Descriptive statistics of candidate predictor variables in mid-lactation Holstein cows included in the model training data set to predict DMI (n = 350)

Variable	Mean	SD	Minimum	Maximum
Descriptive ¹				
PDI	29.2	5.0	16.9	44.3
CDI	29.3	5.0	16.5	45.2
Lactation	2.4	1.3	1	7
Week of lactation	15.9	3.3	9.7	26.6
MBW	137.0	12.1	109.7	169.6
BCS	3.25	0.31	2.50	4.13
Performance ²				
Milk, kg	25.1	5.3	12.6	45.8
Fat, %	3.36	0.68	1.54	5.57
Fat, kg	0.82	0.17	0.32	1.34
True protein, %	3.06	0.24	2.46	3.67
True protein, kg	0.76	0.15	0.41	1.25
MUN, mg/dL	13.6	2.55	7.5	23.7
SCC, 1,000 cells/mL	79.5	230.1	4	3,000
Fatty acid length				
Short chain, %	0.35	0.11	0.00	0.67
Short chain, g	97	31.1	0.00	185
Medium chain, %	1.41	0.36	0.48	2.61
Medium chain, g	391	100.4	133	732
Long chain, %	1.18	0.23	0.35	1.82
Long chain, g	326	63.3	97	504
C14, %	0.34	0.08	0.15	0.60
C14, g	95	23.2	43	167
C16, %	0.97	0.26	0.27	1.71
C16, g	268	72.1	74	473
C18, %	0.35	0.08	0.12	0.61
C18, g	97	22.2	33	169
C18:1, %	0.81	0.17	0.24	1.26
C18:1, g	223	46.5	67	349
De novo, %	0.82	0.21	0.23	1.43
De novo, g	202	54.0	36.2	431.2
Mixed, %	1.06	0.25	0.46	1.78
Mixed, g	260	66.3	83	533
Preformed, %	1.20	0.22	0.39	1.95
Preformed, g	297	69.1	137	618
Production PTA				
Milk	281	518.8	-1,346	1,729
Fat	4	18.7	-43	57
Protein	7	13.8	-32	50
Efficiency PTA				
Residual feed intake	-7	89.3	-256	280
BW composite	0.1	0.77	-2.4	2.2

¹PDI = previous day intake relative to morning milk sample; CDI = current day intake relative to morning milk sample; MBW = metabolic BW.

²Performance candidate variables are from a single morning milking.

unique to our approach is that we developed these models use data from only a single point-in-time, in a manner that could be compatible with DHI sampling methods. The incorporation of various types of data as candidate predictor variables resulted in variable accuracy of predictions (Table 6). In general, adding milk FA and PTA enhanced model accuracy compared with basic cow factors traditionally used in DMI prediction equations (Tables 6 and 7).

The final variables included in each model are reported in Tables 8 and 9. Traditional cow descriptive predictive variables, such as lactation category, week of lactation, MBW, milk yield, and macro-components,

were all retained in at least 1 model. In fact, lactation category, BCS, milk protein, and SCC were all retained as predictor variables in at least 4 of the models tested, and MBW was retained in every model. In large part, the inclusion of these variables in the final models is to be expected because of their general relationship with energy demands.

Adding milk FA and PTA for production and efficiency as novel candidate predictor variables resulted in several variables consistently being included in final prediction models. With the addition of milk FA yield and concentration added as candidate predictor variables, final models consistently included de novo,

Table 5. Descriptive statistics of model BYPE predictor variables (basic factors + milk FA yield + production and efficiency PTA) in mid-lactation Holstein cows included in validation data set 1 to predict DMI

Variable ¹	Validation set ²			
	Mean	SD	Minimum	Maximum
PDI	25.1	4.51	16.2	36.1
Lactation	1.8	1.27	1	6
MBW	128.0	11.98	105.2	147.0
BCS	3.3	0.32	2.5	4.0
Milk, kg	22.1	4.60	13.2	33.7
True protein, kg	0.68	0.12	0.45	0.94
Preformed FA, g	288	88.1	125	579
C18:0, g	70	21.2	37	118
C18:1, g	169	47.9	86	318
PTA milk	396	624.7	-1,345	1,541
PTA RFI	-12.2	61.0	-141	94

¹PDI = previous day DMI relative to morning milk sample; MBW = metabolic BW; FA = fatty acid; RFI = residual feed intake.

²Iowa State University (Siberski-Cooper et al., 2022).

performed, and C18 and C18:1 FA (Tables 8 and 9). Furthermore, when PTA were included as candidate predictor variables, PTA for milk and RFI were consistently included in the final models (Tables 8 and 9).

To assist in comparing the performance of models, we conducted a paired *t*-test on the 5-fold R^2 , CCC, and RMSEP validation data set within each model, and the results are reported in Table 7. Overall, model BYPE had the most robust model performance based on R^2 , CCC, and RMSEP (Table 6). Model parameters for BYPE were also significantly different compared with the other models based on comparison of the 5-fold cross validation results ($P \leq 0.07$; Table 7). Overarching themes highlighting predictor variables that were retained across multiple models and in model BYPE will be discussed in the appropriate sections in this article.

Current Versus PDI Predictions

The initial models BC and BY demonstrate an advantage for predicting PDI versus CDI when milk samples are obtained from the morning milking (Tables 6 and 7). In models BC and BY, the prediction of PDI compared with CDI significantly increased CCC ($P \leq 0.03$; Tables 6 and 7) and tended to increase R^2 ($P \leq 0.07$). Using milk samples to predict production outcomes likely suffers from a certain lag time (McParland and

Table 6. Parameter evaluation of models predicting DMI in mid-lactation Holstein cows from single point-in-time data for milk and components, milk fatty acids, metabolic BW, BCS, previous lactation performance, and production and efficiency PTA

Item ¹	R^2 cross-validation ²	CCC ³	RMSEP, ⁴ kg/d	MSEP ⁵ decomposition, %		
				Mean bias	Slope bias	Random error
Model B						
PDI ⁶	0.55	0.72	3.31	0.00	0.00	100.00
CDI ⁷	0.49	0.67	3.52	0.00	0.00	100.00
Model BC						
PDI	0.64	0.79	2.95	0.00	0.01	100.00
CDI	0.50	0.69	3.46	0.00	0.00	100.00
Model BY						
PDI	0.63	0.78	3.00	0.00	0.00	100.00
CDI	0.50	0.69	3.43	0.00	0.00	100.00
Model BPE						
PDI	0.58	0.75	3.19	0.00	0.00	100.00
Model BYP						
PDI	0.65	0.80	2.90	0.00	0.00	100.00
Model BYE						
PDI	0.65	0.80	2.91	0.00	0.00	100.00
Model BYPE						
PDI	0.67	0.81	2.83	0.00	0.00	100.00

¹Model B = candidate predictor variables, including milk yield and components, metabolic BW, BCS, lactation category, and week of lactation; model BC = model B candidate predictor variables + milk fatty acid content; model BY = model B candidate predictor variables + fatty acid yield; model BPE = model B candidate predictor variables + PTA for production and efficiency; model BYP = model BY candidate predictor variables + PTA production; model BYE = model BY candidate predictor variables + efficiency; model BYPE = model BY candidate predictor variables + PTA in BYP and BYE.

²Five-fold cross-validation R^2 .

³CCC = concordance correlation coefficient (Lin, 1989).

⁴RMSEP = root mean squared error of the prediction.

⁵MSEP = mean squared error of the prediction.

⁶Prediction of DMI on the day before which the morning milk sample was obtained.

⁷Prediction of DMI on the day in which the morning milk sample was obtained.

Table 7. Comparison of DMI prediction models in mid-lactation Holstein cows using a paired *t*-test for R², CCC, and root mean squared error of the prediction (RMSEP) from 5-fold cross validation¹

Item	Model ²									
	B		BC		BY		BPE	BYP	BYE	BYPE
	PDI	CDI	PDI	CDI	PDI	CDI	PDI	PDI	PDI	PDI
R ²										
B-PDI	—	0.39	0.01*	0.47	0.01*	0.65	0.15	<0.01*	0.02*	0.01*
B-CDI		—	0.06*	0.39	0.06*	0.25	0.26	0.05*	0.04*	0.04*
BC-PDI			—	0.06*	0.19	0.07*	0.02*	0.77	0.26	0.02*
BC-CDI				—	0.06*	0.39	0.30	0.05*	0.04*	0.04*
BY-PDI					—	0.07*	0.13	0.13	0.16	0.06*
BY-CDI						—	0.38	0.07*	0.05*	0.05*
BPE-PDI							—	0.01*	0.01*	<0.01*
BYP-PDI								—	0.41	0.05*
BYE-PDI									—	0.05*
BYPE-PDI										—
CCC										
B-PDI	—	0.19	0.01*	0.26	0.02*	0.29	0.07*	0.01*	0.01*	0.01*
B-CDI		—	0.03*	0.06*	0.03*	0.18	0.13	0.02*	0.02*	0.02*
BC-PDI			—	0.03*	0.12	0.02*	0.03*	0.56	0.14	0.02*
BC-CDI				—	0.03*	0.51	0.17	0.02*	0.02*	0.02*
BY-PDI					—	0.02*	0.09*	0.10*	0.07*	0.04*
BY-CDI						—	0.17	0.02*	0.02*	0.02*
BPE-PDI							—	0.01*	0.01*	<0.01*
BYP-PDI								—	0.52	0.02*
BYE-PDI									—	0.04*
BYPE-PDI										—
RMSEP										
B-PDI	—	—	0.01*	0.41	0.01*	0.46	0.14	<0.01*	0.02*	0.01*
B-CDI		—	—	0.27	0.06*	0.30	0.28	0.04*	0.05*	0.04*
BC-PDI			—	—	0.21	0.05*	0.04*	0.53	0.29	0.01*
BC-CDI				—	—	0.64	0.31	0.04*	0.05*	0.04*
BY-PDI					—	—	0.13	0.10*	0.17	0.05*
BY-CDI						—	0.33	0.04*	0.04*	0.03*
BPE-PDI							—	0.02*	0.02*	<0.01*
BYP-PDI								—	0.75	0.07*
BYE-PDI									—	0.03*
BYPE-PDI										—

¹PDI = previous day DMI; CDI = current day DMI; CCC = concordance correlation coefficient.

²Model B = candidate predictor variables included milk yield and components, MBW, BCS, lactation category, and week of lactation; model BC = model B candidate predictor variables + milk fatty acid content; model BY = model B candidate predictor variables + fatty acid yield; model BPE = model B candidate predictor variables + PTA for production and efficiency; model BYP = model BY candidate predictor variables + PTA production; model BYE = model BY candidate predictor variables + efficiency; model BYPE = model BY candidate predictor variables + PTA in BYP and BYE.

**P* ≤ 0.10.

Berry, 2016) as DMI drives feed available for digestion, and the resulting metabolites that will become available to the mammary gland for milk synthesis. Considering that the milk samples were obtained at 0400 h, the milk yield would be less related to alterations in DMI that would occur over the next 20 h, for which the DMI was measured. In grazing dairy cows, for which it is notoriously difficult to determine DMI, prediction of averaged weekly effective energy intake was similar when using predictor variables from either a single morning or evening milking (McParland et al., 2014), which may be an artifact of the method in which they quantified DMI. Furthermore, if evening milk samples had been used in our data set, they may have increased

the accuracy of prediction, as has been noted in other studies with multiple milk samples (McParland et al., 2014), although this additional sampling adds practical limitations to implementation on most dairy farms. This was not done in the current research because our priority was to determine whether milk samples from a single milking could be used to predict DMI, as it is the most common method of DHI sampling and adding extra samples would add labor and expense for dairy farmers. Future work could evaluate covariate factors to be implemented in DMI prediction models to account for milk yield effects from diurnal variation and management factors (Lee and Wardrop, 1984; Rottman et al., 2014; Salfer and Harvatine, 2020), as has been

Table 8. Parameter estimates and SE (in parentheses) for variables included in multiple linear regression models predicting previous and current day DMI in mid-lactation Holstein cows¹

Item	Model ²					
	B		BC		BY	
	PDI	CDI	PDI	CDI	PDI	CDI
Intercept	-4.59 (3.28)	-12.08 (4.10)	-1.25 (2.19)	-12.7 (4.16)	-5.73 (2.66)	3.49 (3.68)
Lactation category ³	—	0.71 (0.31)	—	—	—	0.83 (0.30)
Week of lactation	—	—	—	0.13 (0.06)	—	—
MBW, ⁴ kg	0.14 (0.02)	0.13 (0.02)	0.13 (0.02)	0.16 (0.02)	0.14 (0.02)	0.15 (0.02)
BCS	-2.11 (0.68)	-2.04 (0.74)	—	-2.22 (0.70)	-1.92 (0.59)	-2.57 (0.71)
Milk, kg	—	0.32 (0.05)	—	0.37 (0.05)	—	—
Fat, %	—	—	—	—	—	-3.49 (0.56)
Fat, kg	—	—	3.34 (1.56)	—	—	—
True protein, %	2.94 (0.80)	5.73 (0.93)	—	3.97 (1.02)	—	3.41 (0.97)
True protein, kg	16.88 (1.56)	—	16.54 (1.84)	—	29.4 (1.92)	—
MUN, mg/dL	—	0.30 (0.08)	—	0.28 (0.08)	—	0.23 (0.08)
SCC, 1,000 cells/mL	-0.0018 (0.00080)	—	-0.0014 (0.00070)	—	-0.0018 (0.00072)	—
De novo FA, %	—	—	—	5.16 (1.41)	—	—
De novo FA, g	—	—	—	—	—	0.04 (0.007)
Preformed FA, %	—	—	-10.80 (1.16)	-3.25 (1.23)	—	—
Preformed FA, g	—	—	—	—	-0.03 (0.004)	0.07 (0.005)
C18, %	—	—	28.36 (3.32)	—	—	—
C18, g	—	—	—	—	0.09 (0.01)	0.07 (0.02)

¹Candidate predictor variables included cow basic descriptive factors, milk production and components, and milk fatty acid composition and yield.

²Model B = production, MBW, BCS, lactation category, and week of lactation; model BC = model B + fatty acid (FA) content; model BY = model B + FA yield; PDI = previous day DMI; CDI = current day DMI.

³Multiparous versus primiparous; the coefficient for primiparous cows is negative.

⁴MBW = metabolic BW.

done for accurately determining daily production records from a single milk sample (Duplessis et al., 2019). Overall, because of the improvement in models predicting PDI versus CDI in models B, BC, and BY, the remaining models evaluated only PDI, as in previous studies (Dórea et al., 2018).

Milk FA in Model Predictions

A main goal of this paper was to determine whether factors beyond basic animal descriptors, yield of milk, and macro-components could enhance the prediction of feed intake in lactating dairy cows. Because we

were seeking to use candidate variables that are readily available to dairy farmers, we chose to add milk FA yield and concentrations that are available from DHI organizations through routine on-farm milk testing programs. For the prediction of PDI, including the concentration of FA as candidate predictor variables in model BC (Tables 2 and 3), increased R^2 by 0.09 and increased CCC by 0.07, compared with model B ($P = 0.01$; Tables 6 and 7), and preformed and C18 FA were retained in the model. Additionally, it reduced the RMSEP by 0.36 kg/d for model BC versus model B ($P = 0.01$; Tables 6 and 7). For model BC compared with model B, we detected no change in mean bias, but

Table 9. Parameter estimates and SE for variables included in multiple linear regression models predicting previous day DMI (PDI) in mid-lactation Holstein cows¹

Item ²	Model, PDI ³			
	BPE	BYP	BYE	BYPE
Intercept	-9.18 (3.61)	-6.49 (2.75)	-8.95 (2.19)	-8.42 (2.45)
Lactation category ⁴	0.74 (0.29)	0.75 (0.27)	—	0.82 (0.25)
MBW, kg	0.10 (0.02)	0.13 (0.02)	0.13 (0.02)	0.11 (0.02)
BCS	—	-1.47 (0.60)	—	—
True protein, %	3.69 (0.98)	—	—	—
True protein, kg	16.9 (1.60)	29.2 (1.90)	29.7 (1.85)	30.0 (1.82)
MUN, mg/dL	—	—	—	—
SCC, 1,000 cells/mL	—	-0.0015 (0.0007)	—	—
Preformed FA, g	—	-0.03 (0.004)	-0.03 (0.004)	-0.03 (0.004)
SCFA, g	—	—	-0.02 (0.006)	—
C18, g	—	0.10 (0.01)	0.11 (0.01)	0.13 (0.014)
C18:1, g	—	—	-0.01 (0.006)	-0.01 (0.006)
PTA milk	0.002 (0.0006)	0.0015 (0.0003)	—	0.0016 (0.0003)
PTA fat	0.03 (0.01)	—	—	—
PTA protein	-0.07 (0.02)	—	—	—
PTA RFI	0.008 (0.002)	—	0.009 (0.002)	0.0085 (0.0017)

¹Candidate predictor variables included cow descriptive factors, milk production and components, milk fatty acid yield, and production and efficiency PTA.

²MBW = metabolic BW; FA = fatty acid; SCFA = short-chain fatty acid; RFI = residual feed intake.

³BPE = model B (production, MBW, BCS, lactation category, and week of lactation) + production and efficiency PTA; BYP = model BY (model B + FA yield) + production PTA; BYE = model BY + efficiency PTA; BYPE = model BY + production and efficiency PTA.

⁴Multiparous versus primiparous; the coefficient for primiparous cows is negative.

slope bias increased 0.01 for BC when predicting PDI (Table 6). For the addition of FA yield to the basic model set, improvements were similar to model BC over model B. Model BY increased R^2 by 0.08, increased CCC by 0.06, and reduced RMSEP by 0.31 kg/d, compared with model B, for PDI ($P \leq 0.02$; Tables 6 and 7). The model parameters between model BC and BY were negligible and not significantly different for R^2 , CCC, or RMSEP ($P \geq 0.12$; Tables 6 and 7), but fewer instances of multicollinearity in the model development process for BY compared with BC led us to continue our efforts using only FA yield in the remaining models. It should be noted that FA yields were reported as a

percentage of total milk yield, which may explain why we did not find greater differences between BC and BY.

Adding milk FA as candidate variables in models BC and BY may have helped improve accuracy and precision of DMI prediction compared with model B because milk FA can serve as an indicator of the energy status of the cow. For example, preformed FA content or yield was included as a variable in nearly every model for which it was added as a candidate predictor variable (Tables 7 and 8), with a negative parameter estimate indicating a reduction in DMI as preformed FA in milk increases. The preformed FA C18:1 also had a negative parameter estimate (Table 8). Conversely, de novo FA had a positive parameter estimate when included in the final model (BC, BY; Table 8). Milk FA less than C16 are synthesized in the mammary gland (de novo), those greater than C16 are absorbed from the diet or mobilized from body fat stores (preformed), and C16 FA are a mixture of de novo and preformed (Lanier and Corl, 2015). Taken together, milk FA types can provide insight into cow biology. The ratio of certain FA in milk has been shown to be a reasonable predictor of energy status in cows under the assumption that longer chain FA, such as C18:1, are mobilized from body fat reserves during negative energy balance (Dórea et al., 2017), and cows predicted to have hyperketonemia at the first milk test have greater preformed milk FA content (Pralle et al., 2021). Milk FA content changes over the lactation, with preformed and de novo FA displaying inverse prevalence throughout lactation as energy balance changes (Craninx et al., 2008). The changes noted in the ratio of preformed to de novo FA in milk are most pronounced in early lactation when the cow is adapting to negative energy balance, and the data used for model development herein were from mid-lactation cows that were not experiencing extreme negative energy balance.

The parameter estimate for C18:0 was positive in model BYPE (Table 8), despite it being a preformed FA. Due to rumen biohydrogenation of unsaturated long-chain FA, C18:0 is the primary FA absorbed (Loften et al., 2014); therefore, as DMI increases, more C18:0 could be absorbed and made available for secretion by the mammary gland. In general, dietary ingredients have a profound effect on milk FA. For example, greater dietary starch levels in the diet alter milk FA through depression of milk fat yields (Jenkins and McGuire, 2006), and increasing the FA content of the diet increases milk FA content and yield (Boerman and Lock, 2014). Dietary FA profile also directly alters milk FA profile (Kelly et al., 1998). The FA profile of the dietary ingredients in this paper are not available; however, the lack of major dietary changes between cohorts and the similarity of feedstuffs used in each

diet would not lend to great variability in FA profile across cohorts. Future work including data from cows fed a variety of starch levels and fat sources could prove useful in DMI prediction models using milk FA as predictor variables.

The use of milk FA in DMI predictions is a relatively new approach. Recently, Tedde et al. (2021) used concentration of predicted FA (g/100 g of fat) from milk mid-infrared spectra to predict DMI using partial least squares regression. They retained 6 FA as candidate variables [C8, C10, C12, C14, C18, and C18:1(9)], whereas we included 4 FA as candidate variables (C14, C16, C18, C18:1; Table 2), in addition to FA chain length and saturation. Adding FA candidate predictor variables in our study markedly improved DMI prediction accuracy and precision compared with the model including the basic animal factors (BY and BC vs. B; Table 6). Conversely, no improvement in outcomes was achieved when FA were added in Tedde et al. (2021), but this may have been because milk mid-infrared spectra were included in most of their basic models (Tedde et al., 2021). Using milk mid-infrared spectroscopy, several wavelengths associated with milk FA also contributed to DMI models generated by another laboratory group (Shetty et al., 2017).

A potential limitation of including FA data in future DMI prediction models is that there are likely differences between reported FA analysis due to a lack of industry standardization across laboratories and across equipment types (Pralle and White, 2020). Nonetheless, the fact that improved predictive ability was achieved by adding milk FA as candidate predictor variables is intriguing and merits further consideration, especially for early lactation cows.

PTA in Model Predictions

Additional insight could be garnered from genetic metrics such as PTA. Although models using PTA alone to predict DMI were poor (CCC = 0.08; data not shown), they could contribute novel information when added to the B and BY models. Compared with model B, BPE tended to increase CCC ($P = 0.07$; Tables 6 and 7). Compared with BY, the addition of PTA for milk, fat, and protein as candidate predictor variables in BYP tended to increase CCC (0.80 vs. 0.78; $P = 0.10$; Tables 6 and 7) and reduce RMSEP (2.90 vs. 3.00 kg/d; $P = 0.10$). We found similar improvements in model predictive ability by adding PTA for efficiency as candidate predictor variables (RFI and BW composite) in BYE (Tables 6 and 7). When all PTA were added as candidate predictor variables in model BYPE, the most robust model was produced ($R^2 = 0.67$; CCC = 0.81; RMSEP = 2.83 kg/d; mean bias = 0.00; slope bias =

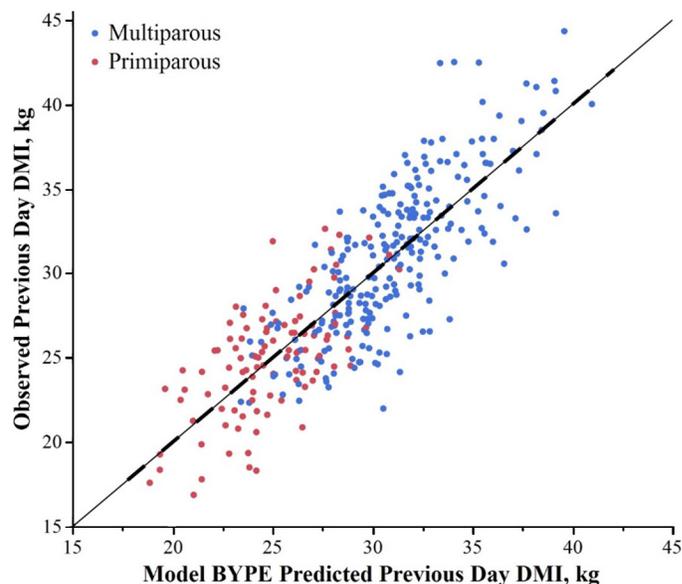


Figure 1. Comparison of model BYPE [model BY (production, metabolic BW, BCS, lactation category, and week of lactation + fatty acid yield) + production and efficiency PTA] predicted versus observed previous day DMI (dashed line) for primiparous (red) and multiparous (blue) Holstein cows. The regression line (dashed) and line of equivalence (solid) are included for reference. $R^2 = 0.68$; concordance correlation coefficient = 0.81; root mean squared error of prediction = 2.83 kg/d; mean bias = 0.00%; slope bias = 0.0%.

0.00; Table 6; Figure 1). Based on the 5-fold cross-validation comparison of model parameters, model BYPE had a greater R^2 , CCC, and a lower RMSEP compared with all other models ($P \leq 0.07$; Table 7).

In models BYPE, BYE, and BPE, PTA RFI was included in the final model with a positive relationship between PTA RFI and DMI (Table 8). Greater phenotypic and genetic RFI values represent greater DMI than predicted within a cohort, after accounting for major energy sinks (VandeHaar et al., 2016); thus, the relationship between PTA RFI and DMI in our models is expected. A secondary validation step was implemented to evaluate whether cows in the training data set, having an RFI phenotype that contributed to the calculation of the RFI PTA in the national database, inflated model accuracy. Using cows from the University of Wisconsin-Madison, for which the RFI phenotype had not yet been incorporated in the national genetic evaluation, DMI was predicted using model BYP and BYPE. The CCC was 0.07 greater for BYP than BYPE (data not shown), indicating no inflation of predictive ability by developing the model with data from cows that had already contributed to the RFI PTA. Furthermore, when comparing the original model performance built on the training data set, the addition of RFI PTA to models BPE and BYPE changed CCC negligibly compared with model BYP

(0.0 to 0.1 increase in CCC; Table 6). When considered together, using RFI PTA does not appear to artificially enhance model performance when determined from cows with an RFI phenotype in the national database. The inclusion of PTA RFI in the final models confirms the applicability of past research efforts to accurately identify genetic factors associated with feed intake in dairy cows. As additional RFI phenotypes continue to be recorded in research herds for use in national genetic evaluations, this will hopefully continue to enhance DMI prediction models.

Other Factors in Model Predictions

The fact that MBW was included as a positive predictor variable in all models in this paper (Tables 8 and 9) underscores the necessity to have capability to obtain BW measurements routinely on dairy farms. We recognize that many dairy farms do not have a method for weighing animals currently, and determining BCS requires a labor investment, so supplemental models without BW and BCS were developed (Supplemental Table S1; <https://data.mendeley.com/datasets/6f62gtxjt3/2>; Brown, 2022). However, walk-over scales have been implemented successfully on dairy farms for routine monitoring of animal performance (van Straten et al., 2008; Alawneh et al., 2011), and walk-over scales are an added benefit of automated milking systems (Thorup et al., 2012; Pszczola et al., 2018). There is also interest in determining BW using artificial intelligence, such as a 3-D vision system (Kuzuhara et al., 2015; Song et al., 2018; Ferreira et al., 2021), which would add a new option to fit farm management style and facility design. Ultimately, accurate prediction of DMI on farms will likely necessitate BW measurement in some capacity, and reasonable resources are available to implement those measurements on dairy farms.

Protein yield was a significant positive predictor variable for DMI in almost every model in this study (Tables 8 and 9). Milk protein yield was a significant predictor variable in previous DMI models (Holter et al., 1997; Roseler et al., 1997), and had a higher correlation with DMI than milk or fat yield (Roseler et al., 1997). We conducted a forward stepwise procedure with BYPE candidate predictor variables to determine the relative importance of all possible variables. Protein yield was the primary variable explaining variation in DMI, accounting for 48% of the variation in DMI alone (Supplemental Table S2; <https://data.mendeley.com/datasets/6f62gtxjt3/2>; Brown, 2022), and milk protein yield has been shown to be the predictor variable with the greatest influence on DMI in other studies (Holter et al., 1997). The biological relationship between milk protein yield and DMI is complex. Metabolizable pro-

tein supply is the main driver of milk protein yield (NRC, 2001), and is altered by forage quality, ruminal starch availability, and ruminal and intestinal protein digestibility. Cows with greater DMI likely have a greater supply of MP to support milk protein synthesis. Considering the consistent nature of the diets across cohorts used in the training data set, future models including data from diets that would alter the MP supply (and in turn the milk protein yield) would be useful.

A final effort to add additional options for farmers to choose from while developing these models was to incorporate previous lactation information for multiparous cows. The previous lactation mature-equivalent milk, protein, and fat yields, coupled with previous lactation length and days dry, were added to models BY, BY, and BYPE. Adding previous lactation candidate variables was only successful in BY, where 305-d mature-equivalent milk was retained, and resulted in poorer predictive ability than the original model B ($R^2 = 0.54$; CCC = 0.71; RMSE = 3.06; mean bias = 0.03; slope bias = 0.07; data not shown). In general, this suggests previous lactation information adds little benefit to current DMI prediction models.

Cow dynamics within the pen and at the bunk may influence DMI or feeding behavior. Within this feeding system, the calculated ratio of cows to bunks in the training data set was 2:1; however, the average occupancy of bunks was 32.1%, and 90% of bunks were only occupied concurrently for 33 min/d. Increasing feeding bunk stocking density from 0.75 to 3 cows/bunk quadratically increases standing time at peak feeding activity and for the entire day, and also increases feed bunk displacements (Huzzey et al., 2006). In cows managed with greater feed bunk density, the combination of social interactions, feed sorting, and more intense feeding bouts can theoretically alter ruminal pH and milk components (DeVries, 2019). Studies evaluating milk production and performance solely based on feeding bunk stocking density are limited because of confounding effects of freestall stocking density inherent in dairy facility design; however, evidence from an electronic feed bunk system suggests that increasing feed bunk competition (from 1 to 3 cows/bunk), absent changes in freestall stocking density, may alter milk composition (Crossley et al., 2017). Increasing stocking density does not appear to alter total DMI as cows alter their behavior to eat feed at different times throughout the day (Crossley et al., 2017). It is unclear how stocking density in our training data set may have altered predictor variables in our final model, but any variance introduced would be accounted for in the cross-validation performance (discussed below), given that the naive data set represents a feeding system with dedicated individual animal feeding bunks. Ongo-

ing research examining the effect of feeding behavior, competition, and social dynamics on feed efficiency will provide valuable insight and may identify behavioral traits or patterns that can contribute to future DMI prediction models (Baier et al., 2021).

Model Validation

An important step in ensuring that prediction models will be robust when applied to naive data sets is cross validation. Leave-one-cohort-out cross-validation resulted in $R^2 = 0.60$, $CCC = 0.63$, and $RMSEP = 3.86$ for model BYPE. A reduction of model performance in leave-one-out cross validation by randomly eliminating full cow records or entire trials from the training data set was observed by others (Shetty et al., 2017; Dórea et al., 2018). Model BYPE was also applied to an external data set obtained from a separate institution (validation set; Iowa State University; Siberski-Cooper et al., 2022). The advantage of this external data set was that they used the same data collection procedures for variables used in our data set, including the same source for milk analysis to eliminate issues with lack of standardization of milk FA characterization. The results of model BYPE applied to the external data set resulted in R^2 of 0.66, CCC of 0.74, and $RMSEP$ of 3.46 kg/d (Figure 2). Applying model BYPE to the external data set had a substantial mean and slope bias (35.96 and 8.46, respectively), where DMI was underpredicted at lesser DMI and overpredicted at greater DMI.

The NRC (2001) and de Souza et al. (2019) DMI models were also applied to the training data set (Figures 3 and 4) and the external validation data set (Supplemental Figure S1; <https://data.mendeley.com/datasets/6f62gtxtj3/2>; Brown, 2022). Compared with NRC (2001) for the training data set, BYPE had a greater R^2 (21%) and CCC (21%), and a smaller $RMSEP$ by 1.02 kg (Figure 3). The mean and slope bias for NRC (2001) were 9.42 and 0.24% greater, respectively, than for BYPE using the training data set. Furthermore, de Souza et al. (2019) had a 18% lesser R^2 , 32% lesser CCC , and a $RMSEP$ 1.8 kg/d greater than BYPE in the training data set (Figure 4). The slope and mean bias for de Souza et al. (2019) was 38.65 and 1.69%, respectively. Both the NRC (2001) and de Souza et al. (2019) models underpredicted DMI in the training data set, especially at greater DMI for the de Souza et al. (2019) model. Certainly, the data sets used in the development of the NRC (2001) and de Souza et al. (2019) DMI models were larger and likely from a more diverse population than the data set herein, and we are unsure of how BYPE would perform with a more heterogeneous data set. To account for variation between cohorts in our work, the random effect of

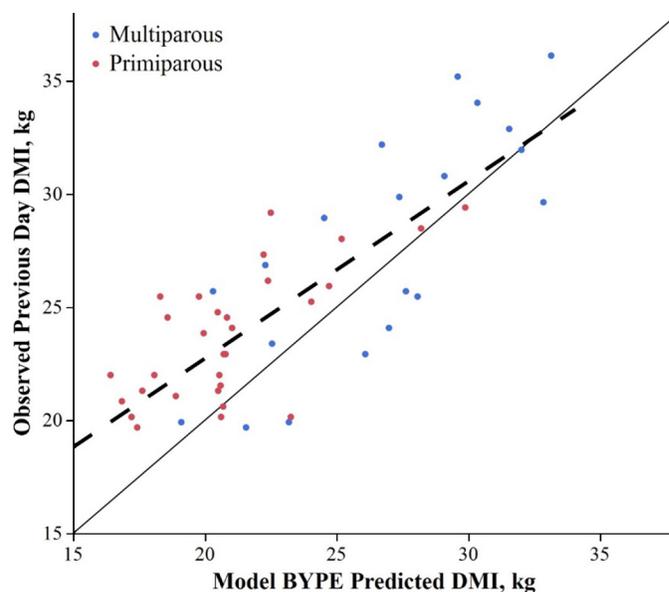


Figure 2. Comparison of model BYPE [model BY (production, metabolic BW, BCS, lactation category, and week of lactation + fatty acid yield) + production and efficiency PTA] predicted DMI versus observed previous day DMI for primiparous (red) and multiparous (blue) Holstein cows using an external data set consisting of 52 DMI observations. The regression line (dashed) and line of equivalence (solid) are included for reference. $R^2 = 0.66$; concordance correlation coefficient = 0.74; root mean squared error of prediction = 3.46 kg/d; mean bias = 35.96%; slope bias = 8.46%.

cohort was tested in all models; however, the random effect of cohort was not significant in any model ($P > 0.13$) and was subsequently removed. Another factor that may have contributed to the model performance was the use of adjustment factors to determine daily milk and component yields. Although these adjustment factors represent an estimate of daily production and may not be a perfect quantification of the cow milk yield, it is the most pragmatic approach to predicting DMI with other models when only 1 milk sample is available. Nonetheless, the comparison of BYPE to existing models underscores the benefit of inclusion of milk FA and PTA variables in productive, Midwest free stall herds, and gives confidence in the ability of model BYPE to accurately predict DMI when these additional variables are available.

CONCLUSIONS

Models predicting DMI using single point-in-time data to mimic a monthly DHI testing regimen resulted in moderately robust results in freestall-housed dairy cows with ad libitum access to feed via electronic feeding gates (2:1 bunk density; overall bunk occupancy rate of 32%). Model validation on an external data set revealed similarly robust results for freestall-housed

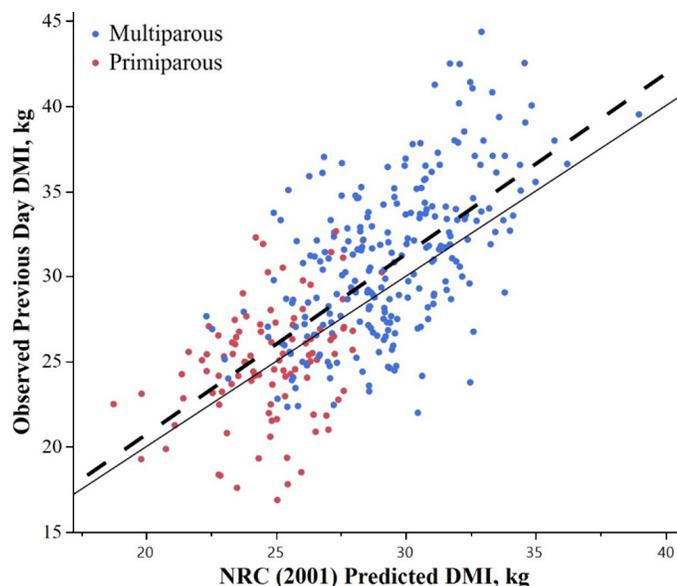


Figure 3. Comparison of NRC (2001) versus observed previous day DMI for primiparous (red) and multiparous (blue) Holstein cows. The regression line (dashed) and line of equivalence (solid) are included for reference. $R^2 = 0.47$; concordance correlation coefficient = 0.60; root mean squared error of prediction = 3.85 kg/d; mean bias = 9.42%; slope bias = 0.24%.

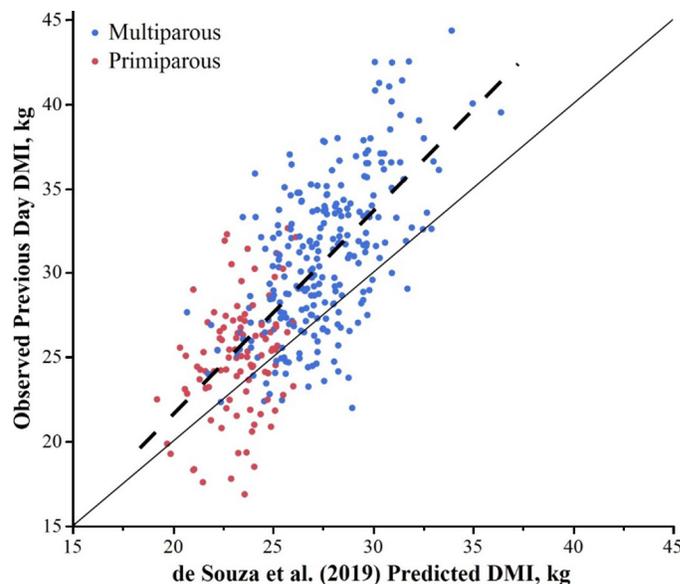


Figure 4. Comparison of de Souza et al. (2019) predicted daily DMI versus observed previous day DMI for primiparous (red) and multiparous (blue) Holstein cows. The regression line (dashed) and line of equivalence (solid) are included for reference. $R^2 = 0.50$; concordance correlation coefficient = 0.49; root mean squared error of prediction = 4.63 kg/d; mean bias = 38.65%; slope bias = 1.69%.

cows with dedicated individual feeding gates (1:1 bunk density). Prediction of PDI was improved over CDI when the milk sample was obtained in the early morning. Adding milk FA data enhanced the predictive ability of the models, perhaps due to patterns in preformed FA changes concomitant to alterations in energy balance. Novel inclusion of production and efficiency PTA were critical to the most robust model (BYPE) and indicates that PTA RFI has an important role in predicting DMI. Model BYPE showed promise when compared with DMI predictions from the NRC (2001) and de Souza et al. (2019). Overall, model BYPE was close to achieving our objective of explaining at least 70% of the variation in DMI in the training and external validation data sets ($R^2 = 0.67$ and 0.66, respectively). Future research should focus on developing models in early and late lactation cows using similar candidate predictor variables, and on identifying other accessible variables that can further improve DMI predictions.

ACKNOWLEDGMENTS

The authors thank Jordan Lambert and Linda Marty (AgSource, Menominee, WI), and Jessica Cederquist (University of Wisconsin-Madison) for retrieving historical data used in this paper. Sandy Bertics, Eric Ronk, and Sophia Kendall (University of Wisconsin-Madison) were instrumental in consistent data collec-

tion and processing. Luiz Ferraretto and Cole Diepersloot (University of Wisconsin-Madison) were generous in providing data for the secondary validation dataset. The projects generating the data used in these models were supported by the Foundation for Food and Agriculture Research (grant no. R109491; Washington, DC) with matching funds from the Council of Dairy Cattle Breeding (Bowie, MD), Zoetis (Kalamazoo, MI), and the Wisconsin Alumni Research Foundation (Madison, WI). W. E. Brown was supported by the Dairy Innovation Hub (University of Wisconsin-Madison). The authors have not stated any conflicts of interest.

REFERENCES

- Alawneh, J. I., M. A. Stevenson, N. B. Williamson, N. Lopez-Villalobos, and T. Otley. 2011. Automatic recording of daily walkover liveweight of dairy cattle at pasture in the first 100 days in milk. *J. Dairy Sci.* 94:4431–4440. <https://doi.org/10.3168/jds.2010-4002>.
- Allen, M. S., D. O. Sousa, and M. J. Vandehaar. 2019. Equation to predict feed intake response by lactating cows to factors related to the filling effect of rations. *J. Dairy Sci.* 102:7961–7969. <https://doi.org/10.3168/jds.2018-16166>.
- Baier, F. S., M. J. Martin, S. J. Erb, G. J. Combs, K. A. Weigel, H. M. White, and J. M. C. Van Os. 2021. Evaluation of parity impact on social competition and feed efficiency in lactating cows. *J. Dairy Sci.* (Suppl. 1):205. (Abstr.)
- Bibby, J., and H. Toutenburg. 1977. Prediction and Improved Estimation in Linear Models. John Wiley & Sons.
- Boerman, J. P., and A. L. Lock. 2014. Effect of unsaturated fatty acids and triglycerides from soybeans on milk fat synthesis and bihydrogenation intermediates in dairy cattle. *J. Dairy Sci.* 97:7031–7042. <https://doi.org/10.3168/jds.2014-7966>.

- Brown, W. 2022. Predicting dry matter intake in mid-lactation Holstein cows using point-in-time data streams available on dairy farms—Supplemental Files. Mendeley Data, V2. <http://www.doi.org/10.17632/6f62gtxtj3.2>.
- CDCB (Council of Dairy Cattle Breeding). 2020. Trait Reference Sheet – Feed Saved (FSAV). Accessed Sep. 6, 2021. https://www.uscdcb.com/wp-content/uploads/2020/11/CDCB-Reference-Sheet-Feed-Saved-12_2020.pdf.
- CDCB (Council of Dairy Cattle Breeding). 2021. Genetic Evaluations: April. CDCB.
- Chandler, T. L., R. S. Pralle, J. R. R. Dórea, S. E. Pooch, G. R. Oetzel, R. H. Fourdraine, and H. M. White. 2018. Predicting hyperketonemia by logistic and linear regression using test-day milk and performance variables in early-lactation Holstein and Jersey cows. *J. Dairy Sci.* 101:2476–2491. <https://doi.org/10.3168/jds.2017-13209>.
- Craninx, M., A. Steen, H. Van Laar, T. Van Nespén, J. Martín-Tereso, B. de Baets, and V. Fievez. 2008. Effect of lactation stage on the odd- and branched-chain milk fatty acids of dairy cattle under grazing and indoor conditions. *J. Dairy Sci.* 91:2662–2677. <https://doi.org/10.3168/jds.2007-0656>.
- Crossley, R. E., A. Harlander-Mataushek, and T. J. DeVries. 2017. Variability in behavior and production among dairy cows fed under differing levels of competition. *J. Dairy Sci.* 100:3825–3838. <https://doi.org/10.3168/jds.2016-12108>.
- de Souza, R. A., R. J. Tempelman, M. S. Allen, and M. J. VandeHaar. 2019. Updating predictions of dry matter intake of lactating dairy cows. *J. Dairy Sci.* 102:7948–7960. <https://doi.org/10.3168/jds.2018-16176>.
- DeVries, T. J. 2019. Feeding behavior, feed space, and bunk design and management for adult dairy cattle. *Vet. Clin. North Am. Food Anim. Pract.* 35:61–76. <https://doi.org/10.1016/j.cvfa.2018.10.003>.
- Dórea, J. R. R., E. A. French, and L. E. Armentano. 2017. Use of milk fatty acids to estimate plasma nonesterified fatty acid concentrations as an indicator of animal energy balance. *J. Dairy Sci.* 100:6164–6176. <https://doi.org/10.3168/jds.2016-12466>.
- Dórea, J. R. R., G. J. M. Rosa, K. A. Weld, and L. E. Armentano. 2018. Mining data from milk infrared spectroscopy to improve feed intake predictions in lactating dairy cows. *J. Dairy Sci.* 101:5878–5889. <https://doi.org/10.3168/jds.2017-13997>.
- Duplessis, M., R. Lacroix, L. Fadul-Pacheco, D. M. Lefebvre, and D. Pellerin. 2019. Assessment of the Canadian model predicting daily milk yield and milk fat percentage using single-milking dairy herd improvement samples. *Can. J. Anim. Sci.* 99:521–531. <https://doi.org/10.1139/cjas-2018-0215>.
- Ferreira, R. E. P., T. Bresolin, D. A. B. Oliverira, and J. R. R. Dorea. 2021. Early detection of ketosis in dairy cows using computer vision and machine learning. Page 124 in Conference of Research Workers in Animal Diseases, Chicago, IL.
- Holter, J. B., J. W. West, and M. L. McGilliard. 1997. Predicting ad libitum dry matter intake and yield of Holstein cows. *J. Dairy Sci.* 80:2188–2199. [https://doi.org/10.3168/jds.S0022-0302\(97\)76167-8](https://doi.org/10.3168/jds.S0022-0302(97)76167-8).
- Huzzey, J. M., T. J. DeVries, P. Valois, and M. A. G. Von Keyserlingk. 2006. Stocking density and feed barrier design affect the feeding and social behavior of dairy cattle. *J. Dairy Sci.* 89:126–133. [https://doi.org/10.3168/jds.S0022-0302\(06\)72075-6](https://doi.org/10.3168/jds.S0022-0302(06)72075-6).
- ICAR. 2017. Procedure 1 of Section 2 of ICAR Guidelines—Computing 24-hour yields. ICAR.
- Jenkins, T. C., and M. A. McGuire. 2006. Major advances in nutrition: Impact on milk composition. *J. Dairy Sci.* 89:1302–1310. [https://doi.org/10.3168/jds.S0022-0302\(06\)72198-1](https://doi.org/10.3168/jds.S0022-0302(06)72198-1).
- Kelly, M. L., J. R. Berry, D. A. Dwyer, J. M. Griinari, P. Y. Chouinard, M. E. Van Amburgh, and D. E. Bauman. 1998. Dietary fatty acid sources affect conjugated linoleic acid concentrations in milk from lactating dairy cows. *J. Nutr.* 128:881–885. <https://doi.org/10.1093/jn/128.5.881>.
- Kuzuhara, Y., K. Kawamura, R. Yoshitoshi, T. Tamaki, S. Sugai, M. Ikegami, Y. Kurokawa, T. Obitsu, M. Okita, T. Sugino, and T. Yasuda. 2015. A preliminarily study for predicting body weight and milk properties in lactating Holstein cows using a three-dimensional camera system. *Comput. Electron. Agric.* 111:186–193. <https://doi.org/10.1016/j.compag.2014.12.020>.
- Lanier, J. S., and B. A. Corl. 2015. Challenges in enriching milk fat with polyunsaturated fatty acids. *J. Anim. Sci. Biotechnol.* 6:26. <https://doi.org/10.1186/s40104-015-0025-0>.
- Lee, A. J., and J. Wardrop. 1984. Predicting daily milk yield, fat percent, and protein percent from morning or afternoon tests. *J. Dairy Sci.* 67:351–360. [https://doi.org/10.3168/jds.S0022-0302\(84\)81308-9](https://doi.org/10.3168/jds.S0022-0302(84)81308-9).
- Lin, L. I. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255–268. <https://doi.org/10.2307/2532051>.
- Loften, J. R., J. G. Linn, J. K. Drackley, T. C. Jenkins, C. G. Soderholm, and A. F. Kertz. 2014. Invited review: Palmitic and stearic acid metabolism in lactating dairy cows. *J. Dairy Sci.* 97:4661–4674. <https://doi.org/10.3168/jds.2014-7919>.
- Martin, M. J., J. R. R. Dórea, M. R. Borchers, R. L. Wallace, S. J. Bertics, S. K. DeNise, K. A. Weigel, and H. M. White. 2021a. Comparison of methods to predict feed intake and residual feed intake using behavioral and metabolite data in addition to classical performance variables. *J. Dairy Sci.* 104:8765–8782. <https://doi.org/10.3168/jds.2020-20051>.
- Martin, M. J., K. A. Weigel, and H. M. White. 2021b. Assessment of the relationship between postpartum health and mid-lactation performance, behavior, and feed efficiency in Holstein dairy cows. *Animals (Basel)* 11:1385. <https://doi.org/10.3390/ani11051385>.
- McParland, S., and D. P. Berry. 2016. The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows. *J. Dairy Sci.* 99:4056–4070. <https://doi.org/10.3168/jds.2015-10051>.
- McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. E. Pryce, and D. P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. *J. Dairy Sci.* 97:5863–5871. <https://doi.org/10.3168/jds.2014-8214>.
- NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th rev. ed. National Academy Press.
- Pralle, R. S., J. D. Amdall, R. H. Fourdraine, G. R. Oetzel, and H. M. White. 2021. Hyperketonemia predictions provide an on-farm management tool with epidemiological insights. *Animals (Basel)* 11:1291. <https://doi.org/10.3390/ani11051291>.
- 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>.
- Pszczola, M., M. Szalanski, K. Rzewuska, and T. Strabel. 2018. Short communication: Improving repeatability of cows' body weight recorded by an automated milking system. *Livest. Sci.* 214:149–152. <https://doi.org/10.1016/j.livsci.2018.04.016>.
- Roseler, D. K., D. G. Fox, L. E. Chase, A. N. Pell, and W. C. Stone. 1997. Development and evaluation of equations for prediction of feed intake for lactating Holstein dairy cows. *J. Dairy Sci.* 80:878–893. [https://doi.org/10.3168/jds.S0022-0302\(97\)76010-7](https://doi.org/10.3168/jds.S0022-0302(97)76010-7).
- Rottman, L. W., Y. Ying, K. Zhou, P. A. Bartell, and K. J. Harvatine. 2014. The daily rhythm of milk synthesis is dependent on the timing of feed intake in dairy cows. *Physiol. Rep.* 2:e12049. <https://doi.org/10.14814/phy2.12049>.
- Salfer, I. J., and K. J. Harvatine. 2020. Night-restricted feeding of dairy cows modifies daily rhythms of feed intake, milk synthesis and plasma metabolites compared with day-restricted feeding. *Br. J. Nutr.* 123:849–858. <https://doi.org/10.1017/S0007114520000057>.
- Shetty, N., P. Lovendahl, M. S. Lund, and A. J. Buitenhuis. 2017. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy.

- copy of milk. *J. Dairy Sci.* 100:253–264. <https://doi.org/10.3168/jds.2016-11609>.
- Siberski-Cooper, C. J., M. S. Mayes, M. Healey, B. M. Goetz, L. H. Baumgard, and J. E. Koltes. 2022. Associations of wearable sensor measures with feed intake, production traits, lactation, and environmental parameters impacting feed efficiency in dairy cattle. *Front. Anim. Sci.* 3:841797. <https://doi.org/10.3389/fanim.2022.841797>.
- Song, X., E. A. M. Bokkers, P. P. J. van der Tol, P. W. G. Groot Koerkamp, and S. van Mourik. 2018. Automated body weight prediction of dairy cows using 3-dimensional vision. *J. Dairy Sci.* 101:4448–4459. <https://doi.org/10.3168/jds.2017-13094>.
- Tedde, A., C. Grelet, P. N. Ho, J. E. Pryce, D. Hailemariam, Z. Wang, G. Plastow, N. Gengler, E. Froidmont, F. Dehareng, C. Bertozzi, M. A. Crowe, and H. Soyeurt. 2021. Multiple country approach to improve the test-day prediction of dairy cows' dry matter intake. *Animals (Basel)* 11:1316. <https://doi.org/10.3390/ani11051316>.
- Thorup, V. M., D. Edwards, and N. C. Friggens. 2012. On-farm estimation of energy balance in dairy cows using only frequent body weight measurements and body condition score. *J. Dairy Sci.* 95:1784–1793. <https://doi.org/10.3168/jds.2011-4631>.
- USDA-ERS. 2021. Milk Cost of Production Estimates. Accessed Oct. 14, 2021. <https://www.ers.usda.gov/data-products/milk-cost-of-production-estimates/>.
- van Straten, M., N. Y. Shpigel, and M. Friger. 2008. Analysis of daily body weight of high-producing dairy cows in the first one hundred twenty days of lactation and associations with ovarian inactivity. *J. Dairy Sci.* 91:3353–3362. <https://doi.org/10.3168/jds.2008-1020>.
- VandeHaar, M. J., L. E. Armentano, K. Weigel, D. M. Spurlock, R. J. Tempelman, and R. Veerkamp. 2016. Harnessing the genetics of the modern dairy cow to continue improvements in feed efficiency. *J. Dairy Sci.* 99:4941–4954. <https://doi.org/10.3168/jds.2015-10352>.
- Wallén, S. E., E. Prestløkken, T. H. E. Meuwissen, S. McParland, and D. P. Berry. 2018. Milk mid-infrared spectral data as a tool to predict feed intake in lactating Norwegian Red dairy cows. *J. Dairy Sci.* 101:6232–6243. <https://doi.org/10.3168/jds.2017-13874>.
- Wildman, E. E., G. M. Jones, P. E. Wagner, R. L. Boman, H. F. Troutt Jr., and T. N. Lesch. 1982. A dairy cow body condition scoring system and its relationship to selected production characteristics. *J. Dairy Sci.* 65:495–501. [https://doi.org/10.3168/jds.S0022-0302\(82\)82223-6](https://doi.org/10.3168/jds.S0022-0302(82)82223-6).

ORCID

- W. E. Brown  <https://orcid.org/0000-0002-8457-7741>
M. J. Caputo  <https://orcid.org/0000-0002-5766-628X>
C. Siberski  <https://orcid.org/0000-0003-0519-787X>
J. E. Koltes  <https://orcid.org/0000-0003-1897-5685>
F. Peñagaricano  <https://orcid.org/0000-0001-6661-3991>
K. A. Weigel  <https://orcid.org/0000-0002-2391-6260>
H. M. White  <https://orcid.org/0000-0001-5449-2811>