Preventive Veterinary Medicine 179 (2020) 105006

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Predicting physiological imbalance in Holstein dairy cows by three different sets of milk biomarkers

Leslie Foldager[a,b,*], Charlotte Gaillard[a,1], Martin T. Sorensen[a], Torben Larsen[a], Elizabeth Matthews[c], Roisin O'Flaherty[d], Fiona Carter[c], Mark A. Crowe[c], Clément Grelet[e], Mazdak Salavati[f,2], Miel Hostens[g], Klaus L. Ingvartsen[a], Mogens A. Krogh[a], GplusE Consortium[3]

[a] Department of Animal Science, Aarhus University, Blichers Allé 20, DK8830, Tjele, Denmark
[b] Bioinformatics Research Centre, Aarhus University, C.F. Møllers Allé 8, DK8000, Aarhus, Denmark
[c] University College Dublin (UCD), Dublin, Ireland
[d] NIBRT GlycoScience Group, National Institute for Bioprocessing, Research and Training, Mount Merrion, Blackrock, Co., Dublin, Ireland
[e] Walloon Agricultural Research Center (CRA-W), 5030, Gembloux, Belgium
[f] Royal Veterinary College, London, NW1 0TU, United Kingdom
[g] Department of Reproduction, Obstetrics and Herd Health, Faculty of Veterinary Medicine, Ghent University, 9820, Merelbeke, Belgium

ARTICLE INFO

Keywords:
Metabolites
Enzymes
FT-MIR
IgG *N*-glycans
Metabolic clusters
Random forests

ABSTRACT

Blood biomarkers may be used to detect physiological imbalance and potential disease. However, blood sampling is difficult and expensive, and not applicable in commercial settings. Instead, individual milk samples are readily available at low cost, can be sampled easily and analysed instantly. The present observational study sampled blood and milk from 234 Holstein dairy cows from experimental herds in six European countries. The objective was to compare the use of three different sets of milk biomarkers for identification of cows in physiological imbalance and thus at risk of developing metabolic or infectious diseases. Random forests was used to predict body energy balance (EBAL), index for physiological imbalance (PI-index) and three clusters differentiating the metabolic status of cows created on basis of concentrations of plasma glucose, β-hydroxybutyrate (BHB), non-esterified fatty acids (NEFA) and serum IGF-1. These three metabolic clusters were interpreted as cows in balance, physiological imbalance and "intermediate cows" with physiological status in between. The three sets of milk biomarkers used for prediction were: milk Fourier transform mid-IR (FT-MIR) spectra, 19 immunoglobulin G (IgG) *N*-glycans and 8 milk metabolites and enzymes (MME). Blood biomarkers were sampled twice; around 14 days after calving (days in milk (DIM)) and around 35 DIM. MME and FT-MIR were sampled twice weekly 1 − 50 DIM whereas IgG *N*-glycan were measured only four times. Performances of EBAL and PI-index predictions were measured by coefficient of determination (R^2_{cv}) and root mean squared error ($RMSE_{cv}$) from leave-one-cow-out cross-validation (cv). For metabolic clusters, performance was measured by sensitivity, specificity and global accuracy from this cross-validation. Best prediction of PI-index was obtained by MME (R^2_{cv} = 0.40 (95 % CI: 0.29 − 0.50) at 14 DIM and 0.35 (0.23 − 0.44) at 35 DIM) while FT-MIR showed a better performance than MME for prediction of EBAL (R^2_{cv} = 0.28 (0.24 − 0.33) vs 0.21 (0.18 − 0.25)). Global accuracies of predicting metabolic clusters from MME and FT-MIR were at the same level ranging from 0.54 (95 % CI: 0.39 − 0.68) to 0.65 (0.55 − 0.75) for MME and 0.51 (0.37 − 0.65) to 0.68 (0.53 − 0.81) for FT-MIR. R^2_{cv} and accuracies were lower for IgG *N*-glycans. In conclusion, neither EBAL nor PI-index were sufficiently well pre-

Abbreviations: BHB, β-hydroxybutyrate; cv, cross-validation; DIM, days in milk; EBAL, body energy balance; FT-MIR, Fourier transform mid-IR; IgG, immunoglobulin G; LDH, dehydrogenase; MME, metabolites and enzymes; NAGase, *N*-acetyl-β-D-glucosaminidase; NEFA, non-esterified fatty acids; PI-index, index for physiological imbalance; R^2, coefficient of determination; RMSE, root mean squared error; VIM, variable importance measures

* Corresponding author at: Department of Animal Science, Aarhus University, Blichers Allé 20, DK8830, Tjele, Denmark.
E-mail address: leslie@anis.au.dk (L. Foldager).
[1] Present address: PEGASE, INRAE Agrocampus Ouest, 35590, Saint-Gilles, France.
[2] Present address: Genetics and Genomics Division, The Roslin Institute, Easter Bush Campus, Midlothian, EH25 9RG, United Kingdom.
[3] All members of the GplusE Consortium are listed at the web site www.gpluse.eu.

https://doi.org/10.1016/j.prevetmed.2020.105006
Received 20 October 2019; Received in revised form 10 April 2020; Accepted 11 April 2020

dicted to be used as a management tool for identification of risk cows. MME and FT-MIR may be used to predict the physiological status of the cows, while the use of IgG *N*-glycans for prediction still needs development. Nevertheless, accuracies need to be improved and a larger training data set is warranted.

1. Introduction

Diseases at calving and during early lactation account for the majority of health and welfare problems in dairy production (Ingvartsen et al., 2003). These include production diseases such as fatty liver, ketosis, rumen acidosis and lameness. Most of such diseases in periparturient cows are argued to be the result of physiological imbalance (Ingvartsen, 2006). Correspondingly, infectious diseases such as mastitis and metritis may occur as the immune system is strongly interlinked with physiological imbalance via the endocrine system and metabolites that must accommodate to the demands for lactation facing the transition cow (Ingvartsen and Moyes, 2015). The consequences of subclinical and clinical diseases are suboptimal animal welfare, production and fertility. Thus, physiological imbalance leading to these subclinical and clinical diseases should have high priority of being addressed with regard to development of management tools. In particular, subclinical stages of diseases can be detected by biomarkers while the cow may appear completely healthy. A number of biomarkers in blood are well described but are currently less well characterized in milk. In the review of Ingvartsen (2006), it is documented that plasma concentrations of glucose, non-esterified fatty acids (NEFA) and β-hydroxybutyrate (BHB) are relevant indicators to determine subclinical ketosis. LeBlanc et al. (2005) also identified blood NEFA and BHB as relevant indicators of displaced abomasum in dairy cows. Piechotta et al. (2012) reported that concentrations of serum NEFA and plasma IGF-1 prepartum are associated with postpartum diseases, while IGF-1 postpartum was the best predictor of both left displaced abomasum and risk of culling (Lyons et al., 2014). However, collecting and analysing blood samples for measuring biomarkers is difficult and expensive, and not applicable in commercial settings. Instead, individual milk samples are readily available and milking systems even provide automatic sampling and measurement of e.g. milk conductivity. Such automatic systems can be expanded to measure e.g. milk BHB (e.g. Herd Navigator™, http://www.herdnavigator.com). Enjalbert et al. (2001) showed that subclinical ketosis can be identified by measuring BHB in milk with enzymatic analysis or with Ketolac test strips. Other studies also reported milk BHB to be a relevant indicator of subclinical and clinical ketosis (e.g., Nielsen et al., 2005). Free glucose, glucose-6-phosphate (Larsen and Moyes, 2015), and isocitrate (Larsen, 2014) reflect the nutrient availability and metabolic turnover in the mammary gland that are linked to the blood levels and therefore potentially indicators of physiological imbalance and risk of disease. Larsen et al. (2010) and Kitchen et al. (1978), respectively, reported that the milk enzymes lactate dehydrogenase (LDH) and *N*-acetyl-β-D-glucosaminidase (NAGase) performed equally with somatic cell count and acute phase proteins as inflammatory indicators of mastitis. In addition,

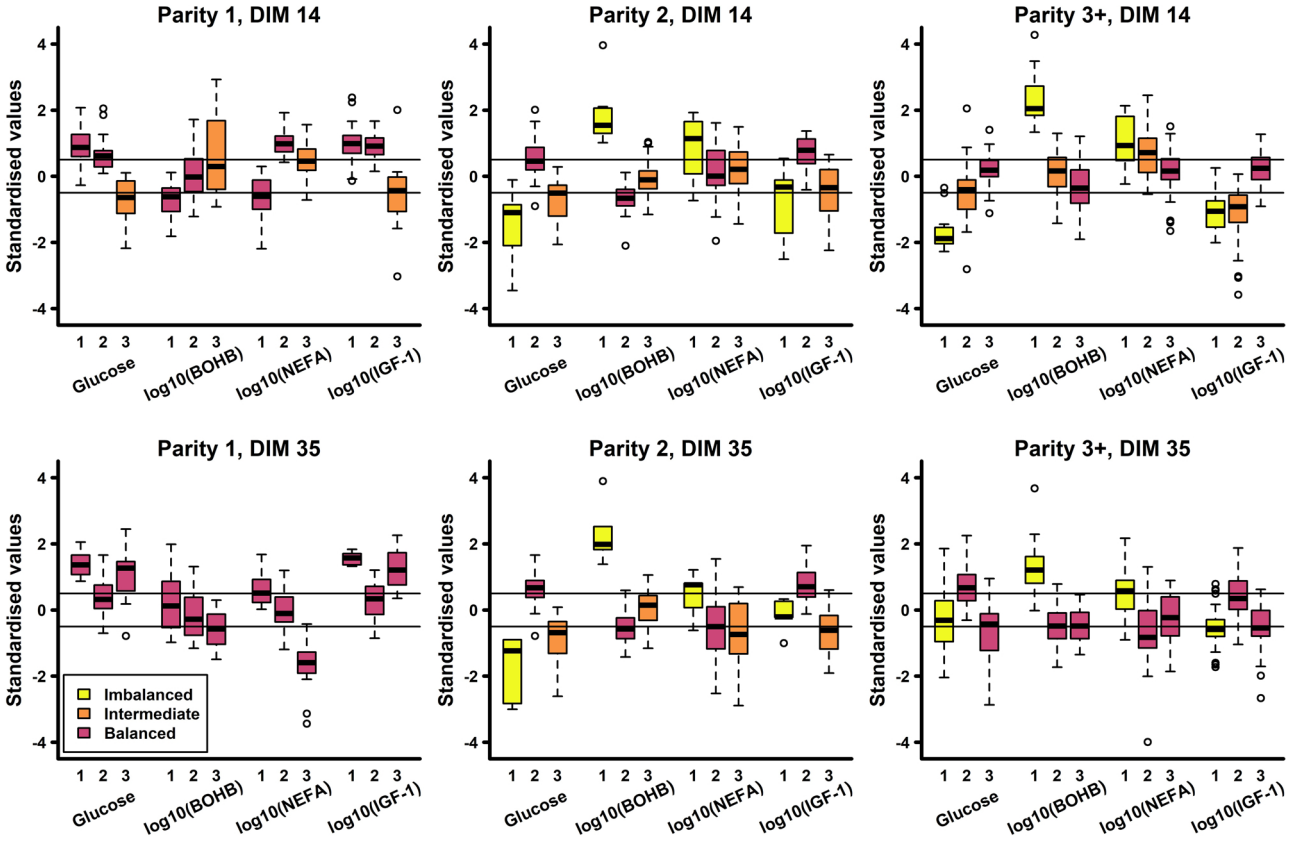

Fig. 1. Box-and-whiskers plots for graphical interpretation (note that bars are medians) of k-means clusters into metabolic clusters as indicated by colours: balanced cluster (magenta), intermediate cluster (orange) and physiological imbalanced cluster (yellow). Distribution of standardised blood metabolites and IGF-1 in each cluster (1, 2 and 3), at 14 DIM (first row), at 35 DIM (second row), for primiparous Holstein dairy cows (first column), second parity cows and for parity 3+ cows (last column). The horizontal lines indicate +/-0.5 SD. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Fourier transform mid-IR (FT-MIR) spectrum of milk is known to contain information on milk composition and cow status (Gengler et al., 2016), and measures of milk immunoglobulin G (IgG) *N*-glycans may be potential new biomarkers.

Based on the same data as here, two other papers (De Koster et al., 2019; Grelet et al., 2019) have considered the prediction of metabolic status (balanced/imbalanced) using metabolic clusters based on k-means clustering of four blood biomarkers; glucose, NEFA and BHB in plasma and IGF-1 in serum. Grelet et al. (2019) only considered FT-MIR, and De Koster et al. (2019) only used multiparous cows and both studies only considered prediction of clusters. The present paper supplements these studies by comparing random forests predictions from three different sets of milk biomarkers; metabolites and enzymes (MME), FT-MIR spectra and IgG *N*-glycans among all parities and by comparing predictions of body energy balance (EBAL) and index for physiological imbalance (PI-index) (Ingvartsen, 2006; Moyes et al., 2013a, 2013b). Additionally, the present paper focuses deeper on MME and also investigates the potential of IgG *N*-glycans as a set of milk biomarkers and contributes to the understanding of the clustering approach. The main objective was to compare the use of MME, FT-MIR and IgG *N*-glycans for identification of cows in physiological imbalance and thus at risk of developing a metabolic or infectious disease.

2. Material and methods

Study design, sampling and analysis of milk and blood have been described in De Koster et al. (2019), Grelet et al. (2019) and Krogh et al. (2019). In brief, cows were observed in the six research herds involved in the GplusE project. The location of the research herds were Northern Ireland (UK), Denmark (DK), Belgium (BE), Italy (IT), Germany (DE) and Ireland (IE) and for each herd at least 25 cows were planned to be included, attaining a total number of cows that was larger than most recent studies at that time. A total of 234 Holstein dairy cows (55 first parity, 66 second parity, and 113 in third or higher parity (3+)) were recorded for the first 50 days in milk (DIM), see Supplementary Table S1. Data from additionally 7 cows that were culled before 50 DIM were removed before analysis. In four locations, all cows were fed a standard diet typical for the particular country. In the UK and DK, a standard diet and two alternative diets were used. In total, the observations were done in six different production systems and with ten different diets to represent the variation in feeding and management in northern and western European countries. An overview of the diets is shown in Table 1 of Krogh et al. (2019).

2.1. Derived measures

The calculation of EBAL was described in De Koster et al. (2019) and Krogh et al. (2019). EBAL was only calculated if both morning and evening milk yield was available for that day. Afterwards, three days (i.e., +/- 1 days in milk (DIM)) moving averages of EBAL were calculated and used for the analyses. The average live body weights within calendar week was used to smooth large day-to-day variation and measurement errors of scales. Summary statistics of EBAL are shown in supplementary tables of Krogh et al. (2019).

PI-index was calculated as $[\log_{10}(\text{NEFA})] + [\log_{10}(\text{BHB})] - [\text{glucose}]$ (Moyes et al., 2013a), where plasma concentrations of the individual metabolites were standardised to an overall mean of zero and variance of one (as indicated by square brackets). Moyes et al. (2013a) used the natural logarithm (ln) but since $\log_{10}$ and ln are proportional, $\ln(y) = \ln(10)\log_{10}(y)$, the standardised values will be exactly equal, i.e. $[\ln(y)] = [\log_{10}(y)]$. Thus, since the manuscripts of Grelet et al. (2019) and De Koster et al. (2019) applied $\log_{10}$-transformations of NEFA and BHB this same approach was used here.

2.2. Metabolic clusters

As an alternative phenotype to negative EBAL and PI-index, clusters were created by use of the k-means method of Hartigan and Wong (1979) from standardised measures of plasma glucose, plasma $\log_{10}$(BHB), plasma $\log_{10}$(NEFA), and serum $\log_{10}$(IGF-1). As mentioned in the Introduction, these four blood biomarkers contained complementary information on the physiological status of the animal. Three clusters ($k = 3$) were constructed for each combination of three parities (1, 2 and 3+ lactations) and two periods in early lactation (around 14 and 35 DIM) as visualised in Fig. 1. Deciding on the number of clusters can be intricate but in the present sample $k = 3$ was found to be a fair compromise between size and similarity (in terms of the within cluster sum of squares, results not shown). Based on a graphical interpretation using boxplots of the standardised concentrations of plasma glucose, NEFA and BHB and serum IGF-1 (see Fig. 1), three metabolic clusters were defined as representing balanced, intermediate and imbalanced cows.

Criteria to define the imbalanced metabolic cluster are the most important. We defined the metabolic cluster as imbalanced if standardised plasma glucose and serum IGF-1 concentrations were both lower than those of plasma BHB and plasma NEFA, and in addition both median BHB and NEFA were above 0.5 SD (Fig. 1). Intermediate and balanced metabolic clusters had less sharp definitions: The intermediate metabolic cluster generally had lower standardised glucose and IGF-1 concentrations than BHB and NEFA, with NEFA and BHB boxes in the ± 0.5 SD area and glucose and IGF-1 around or below -0.5 SD. The balanced metabolic cluster had standardised glucose and IGF-1 concentrations around 0.5 SD and standardised NEFA and BHB concentrations below or equal to those of glucose and IGF-1, or all four approximately equal and around -0.5 SD. The metabolic cluster was also considered balanced if all four boxes were inside the ± 0.5 SD area.

2.3. Milk biomarkers

Three different sets of milk biomarkers (MME, FT-MIR spectra and IgG *N*-glycans) were considered as predictors. Metabolites and enzymes consisted of six milk metabolites (glycose-6-phosphate, free glucose, BHB, isocitrate, urea and uric acid) and two enzymes (NAGase and LDH). Milk Fourier transform mid-IR spectra were obtained from different instruments and consequently standardised into a common format. FT-MIR data consisted of absorbance values at 212 wavenumbers selected from a total of 1060 by removal of areas known to be non-reproducible between instruments or non-informative due to the water component in milk (Grelet et al., 2016). Finally, 19 peaks of IgG *N*-glycans were manually identified and integrated. Each peak's percentage of the total area under the 19 peaks was used as the measure for the statistical analyses. Further details on the laboratory analysis are given in De Koster et al. (2019).

2.4. Random forests predictions

Each of the three sets of milk biomarkers were used to predict the responses (EBAL, PI-index and metabolic clusters) separately for each parity and period by use of the random forests algorithm (see below), i.e. in total 54 predictions. In addition, each of the six plasma metabolites and serum IGF-1 were predicted. To make a more fair comparison with IgG *N*-glycans, we also made a comparison using only data that were complete across all three sets of milk biomarkers in relation to the two periods; around DIM 14 and DIM 35. Random forests belongs to the field of machine learning and is an ensemble of classification or regression trees (Breiman, 2001) with each tree being a set of decision rules. A short description of the algorithm is given below, whereas we refer to Breiman (2001) for a technical presentation and introduction to random forests. We generally used default settings of the implementation except that we used 2500 trees (instead of the default 500) to

stabilise estimates of accuracy.

2.5. Random forests algorithm

In summary, for each of a pre-specified number of trees a sample is drawn from the original data by sampling with replacement (bootstrap sample). These samples have the same size as the original data but contain on average approximately two thirds of the individual records, since some are selected more than once and some not at all. Each bootstrap sample is used for training an unpruned tree. At each node of the tree, a set of predictors (default for binary classification: square root number of predictors) are chosen at random as candidates for splitting the data present at the current (parent) node into two chunks (child nodes). The algorithm then selects the candidate (categorical) or cut–point (continuous) that give the largest reduction of the Gini index (Breiman et al., 1984), i.e., the most homogeneous child nodes. Each tree is grown as large as possible. The random selection of candidate predictors at each node protects from overfitting (Breiman, 2001) and pruning is not necessary. When the random forest of trees have been developed, new records are passed through each tree and a majority voting or averaging predicts their classes or values.

2.6. Statistical analysis

The statistical analyses were carried out using R version 3.6.2 (R Core Team, 2020). For k-means clustering the *kmeans* function of R was used. Random forests modelling was carried out by use of the *randomForest* package (Liaw and Wiener, 2002). We evaluated performance of random forests predictions for metabolic clusters by a leave-one-cow-out (internal) cross-validation strategy, i.e., in turn preserving data from one cow as test set and using data from the other cows for training of a random forests model. By use of the *confusionMatrix* function of the *caret* package (Kuhn, 2008), we calculated global accuracy (proportion of correctly classified samples, i.e., the diagonal of the 3 by 3 contingency table of predicted versus true cluster also known as the confusion matrix), sensitivity for each cluster (proportion correctly predicted to that cluster) and specificity (proportion correctly predicted not to be in that cluster). In addition, the precision of predictions for the individual blood biomarkers, EBAL and PI-index was measured by the coefficient of determination of cross-validation (R^2_{cv}) and the root mean squared error ($RMSE_{cv}$).

To explore the ranking of the individual MME biomarkers within parity and period, the variable importance measure (VIM) was calculated (Breiman, 2001) and plotted using *randomForests*. This measure is based on the internal out-of-bag samples, i.e., the third not picked to be included in each bootstrap sample, see Breiman (2001).

Characteristics and differences among metabolic clusters in milk metabolite concentrations, enzyme activities and daily milk yield were examined separately for parity 2 and 3+ at DIM 14 by ANOVA with F-tests. Since most health events and imbalances are expected to happen in the first part of the early lactation period, we only focused on DIM 14 for this part. First parity cows were not given further attention since none of these were classified to the imbalanced cluster at DIM 14 and all were in clusters classified as balanced at DIM 35.

3. Results

Summary statistics for production, blood biomarkers and MME can be found in tables and supplementary tables of Krogh et al. (2019).

3.1. Predictions of EBAL and PI-index by sets of milk biomarkers

The performances (R^2_{cv} and $RMSE_{cv}$) of predicting measures of EBAL and PI-index by the three sets of milk biomarkers as determined by leave-one-cow-out cross-validation are shown in Table 1. The best precision was obtained when predicting PI-index by MME with an R^2_{cv} of 0.40 (95 % CI: 0.29 − 0.50) at 14 DIM and 0.34 (0.23 − 0.44) at 35 DIM. For FT-MIR, the corresponding R^2_{cv} was 0.26 (0.16 − 0.36) and 0.19 (0.10 − 0.30). For EBAL, however, FT-MIR showed a better performance than MME with an R^2_{cv} of 0.28 (0.24 − 0.33) vs 0.21 (0.18 − 0.25). The RMSEs from MME and FT-MIR predictions were respectively 23.7 (22.6 − 24.7) and 23.4 (22.1 − 24.6) for EBAL and between 1.62 (1.44 − 1.78) and 1.93 (1.60 − 2.21) for PI-index. Predictions of EBAL and PI-index by IgG *N*-glycans had the lowest precisions, with R^2_{cv} ranging between 0.01 (0.00 − 0.08) and 0.06 (0.02 − 0.12) and with $RMSE_{cv}$ being 26.3 (23.7 − 28.6) for EBAL and 2.04 (1.75 − 2.30) for PI-index.

3.2. Predictions of individual blood biomarkers by sets of milk biomarkers

Predictions of individual blood biomarkers are shown in Table 2. The best precisions were obtained with MMEs for plasma urea (R^2_{cv} = 0.62 (95 % CI: 0.53 − 0.69) for 14 DIM and 0.59 (0.51 − 0.67) for 35 DIM) and for plasma BHB (R^2_{cv} = 0.46 (0.36 − 0.56) and 0.40 (0.30 − 0.50)). Precisions of serum IGF-1 were at the same level as plasma BHB for DIM 35 (R^2_{cv} = 0.40 (0.30 − 0.50)) and somewhat lower for DIM 14 (R^2_{cv} = 0.32 (0.22 − 0.42)). The precisions by IgG *N*-glycans were always the lowest whereas generally, FT-MIR were at the same

Table 1
Precision of random forests predictions of EBAL and PI-index with three sets of milk biomarkers (MME, FT-MIR and IgG *N*-glycans)[a] in Holstein dairy cows in six herds. The performance was measured by R^2_{cv} and $RMSE_{cv}$. Individual milk biomarkers were standardised using all available data before matching. In addition to sets of milk biomarkers, parity (1, 2 and 3+) as a factor and DIM (days in milk) as continuous covariate were included as predictors for EBAL, whereas only parity was added as predictor for PI-index. Number of cows (samples) are after removal of records excluded due to missing values.

Response	Period (DIM)	Sets of milk biomarkers	N_{cows} ($N_{samples}$)	R^2_{cv} (95 % CI)	$RMSE_{cv}$ (95 % CI)
EBAL (only using DK, IE and UK herds)	1 − 50	MME	132 (1608)	0.21 (0.18 − 0.25)	23.7 (22.6 − 24.7)
		FT-MIR	132 (1230)	0.28 (0.24 − 0.33)	23.4 (22.1 − 24.6)
		IgG	122 (328)	0.06 (0.02 − 0.12)	26.3 (23.7 − 28.6)
PI-index	14	MME	216	0.40 (0.29 − 0.50)	1.62 (1.44 − 1.78)
		FT-MIR	201	0.26 (0.16 − 0.36)	1.86 (1.59 − 2.09
		IgG	133	0.01 (0.00 − 0.08)	2.04 (1.68 − 2.35)
	35	MME	218	0.34 (0.23 − 0.44)	1.71 (1.48 − 1.91)
		FT-MIR	195	0.19 (0.10 − 0.30)	1.93 (1.60 − 2.21)
		IgG	134	0.05 (0.00 − 0.15)	2.04 (1.75 − 2.30)

[a] Milk biomarkers were matched with the EBAL closest in sampling date (+/- 3 days). For FT-MIR this matching strategy was also applied to PI-index for the period noted in the column denoted "Period (DIM)". If no perfect match (same day) was found, we proceeded as follows: Step 1 day backward first (day before milk biomarker sampling date), then 2 days forward (i.e. 1 day after the sampling data), then 3 days back (corresponding to 2 days before sampling), then 4 days forward, 5 days backward and 6 days forward. That is, closest match within 7 days (a week) centred in the milk biomarker's sampling date. For IgG *N*-glycans, the measure from the period noted was used for these two measurements. Averages of measures of milk metabolites and enzymes within the same week (Monday-Sunday) as blood sampling were used for PI-index.

Table 2
Precision (R^2_{cv} and $RMSE_{cv}$) of random forests predictions of plasma metabolites and serum IGF-1 with three sets of milk biomarkers (MME, FT-MIR and IgG *N*-glycans) in Holstein dairy cows. Individual milk biomarkers were standardised and the sample matching the blood sample date (+/- 3 days) was used. In addition, parity (1, 2 and 3+) was included as a predictor. Number of cows are after removal of those excluded due to missing values.

Blood biomarker	Period (DIM)	Sets of milk biomarkers	N_{cows}	R^2_{cv} (95 % CI)	$RMSE_{cv}$ (95 % CI)
Plasma fructosamine	14	MME	213	0.12 (0.05 – 0.21)	16.9 (15.1 – 18.6)
		FT-MIR	198	0.11 (0.04 – 0.20)	17.2 (15.6 – 18.7)
		IgG	131	0.03 (0.00 – 0.11)	17.6 (15.4 – 19.5)
	35	MME	214	0.18 (0.10 – 0.28)	16.4 (14.6 – 18.0)
		FT-MIR	191	0.02 (0.00 – 0.08)	18.5 (16.1 – 20.6)
		IgG	132	0.11 (0.03 – 0.23)	17.2 (14.8 – 19.3)
Plasma urea	14	MME	216	0.62 (0.53 – 0.69)	0.72 (0.63 – 0.80)
		FT-MIR	201	0.06 (0.01 – 0.13)	1.08 (0.97 – 1.18)
		IgG	133	0.01 (0.00 – 0.07)	1.07 (0.92 – 1.20)
	35	MME	218	0.59 (0.51 – 0.67)	0.78 (0.69 – 0.87)
		FT-MIR	195	0.13 (0.05 – 0.22)	1.13 (1.02 – 1.23)
		IgG	134	0.01 (0.00 – 0.07)	1.16 (1.00 – 1.29)
Plasma cholesterol	14	MME	216	0.09 (0.03 – 0.17)	0.68 (0.62 – 0.74)
		FT-MIR	201	0.01 (0.00 – 0.06)	0.72 (0.64 – 0.78)
		IgG	133	0.01 (0.00 – 0.07)	0.72 (0.63 – 0.79)
	35	MME	218	0.12 (0.05 – 0.20)	0.98 (0.90 – 1.06)
		FT-MIR	195	0.03 (0.00 – 0.10)	1.02 (0.92 – 1.12)
		IgG	134	0.04 (0.00 – 0.12)	1.02 (0.91 – 1.13)
Plasma log_{10}(NEFA)	14	MME	216	0.13 (0.06 – 0.23)	0.25 (0.22 – 0.27)
		FT-MIR	201	0.10 (0.03 – 0.19)	0.26 (0.23 – 0.28)
		IgG	133	< 0.01 (0.00 – 0.01)	0.26 (0.22 – 0.29)
	35	MME	218	0.09 (0.03 – 0.17)	0.30 (0.27 – 0.33)
		FT-MIR	195	0.03 (0.00 – 0.09)	0.31 (0.28 – 0.34)
		IgG	134	0.01 (0.00 – 0.06)	0.32 (0.28 – 0.36)
Plasma glucose	14	MME	216	0.29 (0.19 – 0.39)	0.41 (0.36 – 0.45)
		FT-MIR	201	0.23 (0.13 – 0.33)	0.43 (0.37 – 0.48)
		IgG	133	0.11 (0.03 – 0.23)	0.49 (0.41 – 0.56)
	35	MME	218	0.32 (0.22 – 0.43)	0.43 (0.37 – 0.47)
		FT-MIR	195	0.19 (0.10 – 0.29)	0.48 (0.42 – 0.53)
		IgG	134	0.17 (0.07 – 0.29)	0.49 (0.42 – 0.55)
Plasma log_{10}(BHB)	14	MME	216	0.46 (0.36 – 0.56)	0.16 (0.14 – 0.18)
		FT-MIR	201	0.27 (0.17 – 0.37)	0.20 (0.16 – 0.22)
		IgG	133	0.04 (0.00 – 0.12)	0.24 (0.19 – 0.27)
	35	MME	218	0.40 (0.30 – 0.50)	0.17 (0.14 – 0.19)
		FT-MIR	195	0.25 (0.15 – 0.36)	0.19 (0.16 – 0.22)
		IgG	134	< 0.01 (0.00 – 0.02)	0.22 (0.18 – 0.25)
Serum log_{10}(IGF-1)	14	MME	216	0.32 (0.22 – 0.42)	0.27 (0.24 – 0.30)
		FT-MIR	204	0.36 (0.26 – 0.47)	0.26 (0.22 – 0.30)
		IgG	136	0.24 (0.12 – 0.37)	0.29 (0.26 – 0.32)
	35	MME	216	0.40 (0.30 – 0.50)	0.21 (0.19 – 0.23)
		FT-MIR	197	0.35 (0.24 – 0.46)	0.22 (0.20 – 0.25)
		IgG	138	0.14 (0.05 – 0.26)	0.25 (0.22 – 0.28)

Table 3
Number of Holstein dairy cows per metabolic cluster (balanced, intermediate, imbalanced) at DIM 14 and 35. Furthermore, the last column shows which clusters the DIM 35 cows belonged at DIM 14.

Cluster and parity	Number of cows		Cluster affiliation at DIM 14 for DIM 35 cows
	DIM 14	DIM 35	
Parity 1			
Balanced	38	52	38 Balanced + 14 Intermediate
Intermediate	14	0	
Imbalanced	0	0	
Parity 2			
Balanced	23	32	21 Balanced + 11 Intermediate
Intermediate	28	21	1 Balanced +17 Intermediate + 3 Imbalanced
Imbalanced	7	5	1 Balanced + 4 Imbalanced
Parity 3+			
Balanced	38	70	31 Balanced + 39 Intermediate
Intermediate	54	0	
Imbalanced	11	33	7 Balanced +15 Intermediate + 11 Imbalanced
Total	213	213	

level as MME but in some cases much lower.

3.3. Metabolic cluster changes

The numbers of cows in each of the three metabolic clusters at DIM 14 and DIM 35 are reported in Table 3 with indications of changes between the two periods. All the 52 primiparous cows were interpreted balanced at DIM 35. Among the 28 parity 2 cows in the intermediate cluster at DIM 14, 17 (61 %) did not shift to a cluster deemed to be more "balanced" at DIM35, staying in an intermediate cluster, while the rest changed to a balanced cluster (N = 11). Most of the 23 parity 2 cows in the balanced cluster at DIM 14 stayed in a balanced cluster at DIM 35 (N = 21) with only two cows shifting; one to an imbalanced and one to an intermediate cluster at DIM 35. For 15 (4 + 11) out of 18 (7 + 11) (83 %) parity 2 and 3+ cows in the imbalanced cluster DIM 14, extra attention may be relevant as they were also in an imbalanced cluster DIM 35. Concerning parity 3+ cows in the balanced cluster DIM 14, 31 out of 38 (82 %) were still in a balanced cluster at DIM 35 while the rest changed to an imbalanced cluster. Of the 54 parity 3+ cows in the intermediate cluster DIM 14, 39 (72 %) changed to a balanced cluster at DIM 35, while the rest changed to an imbalanced cluster.

Table 4
Leave-one-cow-out cross-validation of performance for random forests predictions of metabolic clusters by MME, FT-MIR and IgG *N*-glycans. Clusters based on k-means clustering (k = 3) of standardised values of plasma glucose, log_{10}(BHB) and log_{10}(NEFA) and serum log_{10}(IGF-1) in Holstein dairy cows.

Period and parity	Cluster number[a]	Metabolic cluster[b]	Sensitivity			Specificity			Global accuracy[c] (95 % CI)		
			MME	FT-MIR	IgG	MME	FT-MIR	IgG	MME	FT-MIR	IgG
Parity 1											
DIM 14	1	Balanced	0.74	0.70	0.38	0.52	0.61	0.48	0.54	0.51	0.32
	2	Balanced	0.14	0.40	0.10	0.89	0.75	0.79	(0.39 − 0.68)	(0.37 − 0.65)	(0.17 − 0.51)
	3	Intermediate	0.60	0.31	0.45	0.84	0.87	0.70			
DIM 35	1	Balanced	0.63	0.25	0.00	0.98	0.90	1.00	0.62	0.68	0.43
	2	Balanced	0.68	0.83	0.69	0.63	0.71	0.21	(0.47 − 0.75)	(0.53 − 0.81)	(0.25 − 0.63)
	3	Balanced	0.53	0.69	0.18	0.73	0.87	0.68			
Parity 2											
DIM 14	1	Imbalanced	0.50	0.00	0.00	0.98	0.98	1.00	0.55	0.59	0.46
	2	Balanced	0.50	0.70	0.42	0.68	0.65	0.70	(0.42 − 0.68)	(0.45 − 0.72)	(0.29 − 0.63)
	3	Intermediate	0.61	0.70	0.61	0.53	0.68	0.29			
DIM 35	1	Imbalanced	0.00	0.00	0.00	0.98	0.96	1.00	0.58	0.55	0.53
	2	Balanced	0.79	0.69	0.71	0.50	0.52	0.53	(0.44 − 0.70)	(0.40 − 0.69)	(0.35 − 0.70)
	3	Intermediate	0.36	0.50	0.44	0.70	0.71	0.60			
Parity 3+											
DIM 14	1	Imbalanced	0.70	0.00	0.00	1.00	0.99	1.00	0.63	0.66	0.51
	2	Intermediate	0.74	0.76	0.74	0.51	0.63	0.17	(0.53 − 0.73)	(0.56 − 0.76)	(0.38 − 0.64)
	3	Balanced	0.46	0.70	0.17	0.78	0.76	0.74			
DIM 35	1	Imbalanced	0.71	0.59	0.10	0.87	0.74	0.73	0.65	0.59	0.44
	2	Balanced	0.71	0.63	0.71	0.68	0.82	0.70	(0.55 − 0.75)	(0.49 − 0.70)	(0.31 − 0.57)
	3	Balanced	0.50	0.56	0.45	0.90	0.83	0.73			

[a] The cluster numbers are arbitrary and cannot be compared among period/parity combinations.
[b] As interpreted from Fig. 1. The metabolic clusters are comparable among period/parity combinations.
[c] Proportion of correctly classified observations by the prediction, i.e. the diagonal of the confusion matrix.

3.4. Prediction of metabolic clusters

Accuracies to predict the clusters from sets of milk biomarkers with random forests models are presented in Table 4 for each combination of parity (1, 2 and 3+) and period (DIM 14 and 35). As in Grelet et al. (2019) and De Koster et al. (2019), including milk yield as a factor with the aim to help distinguishing between classes did not improve the accuracy (results not shown). Global accuracies from MME and FT-MIR were at the same level and ranged from 0.54 (95 % CI: 0.39 − 0.68) to 0.65 (0.55 − 0.75) for MME and 0.51 (0.37 − 0.65) to 0.68 (0.53 − 0.81)

Table 5
Characteristics[a] of milk yield, metabolites and enzymes and comparisons among the three metabolic clusters (balanced, intermediate and physiological imbalanced) of Holstein dairy cows at DIM 14 in parity 2 and 3+, respectively. Results of ANOVA F-tests for differences among metabolic clusters are indicated[b].

Milk measure and parity	Balanced (n = 24)[d]			Intermediate (n = 28)			Imbalanced (n = 9)[d]			
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	
Parity 2										
Glucose-6-P (mM)	0.17	0.22	0.28	0.14	0.18	0.20	0.16	0.18	0.23	*
Free glucose (mM)	0.18	0.25	0.28	0.17	0.22	0.26	0.07	0.12	0.15	**
log_{10} (BHB)[c]	1.56	1.63	1.72	1.66	1.76	1.85	1.98	2.06	2.40	***
Isocitrate (mM)	0.15	0.17	0.19	0.17	0.19	0.20	0.19	0.28	0.29	**
Urea (mM)	2.47	3.15	3.83	2.16	3.18	3.79	2.66	2.82	4.90	ns
Uric acid (μM)	161	176	204	154	164	203	139	173	181	ns
log_{10} (NAGase)[c]	0.24	0.35	0.46	0.18	0.26	0.41	0.41	0.42	0.46	ns
log_{10} (LDH)[c]	0.37	0.46	0.63	0.42	0.56	0.68	0.46	0.57	0.72	ns
Milk yield (kg/day)	30.5	32.4	36.8	26.3	31.6	35.9	28.2	30.5	34.4	ns
Milk measure and parity	Balanced (n = 39)[d]			Intermediate (n = 54)			Imbalanced (n = 11)			
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	
Parity 3+										
Glucose-6-P (mM)	0.15	0.19	0.24	0.15	0.17	0.22	0.16	0.18	0.20	ns
Free glucose (mM)	0.17	0.21	0.24	0.13	0.16	0.18	0.09	0.10	0.11	***
log_{10} (BHB)[c]	1.55	1.66	1.74	1.66	1.74	1.92	2.05	2.12	2.23	***
Isocitrate (mM)	0.14	0.16	0.19	0.15	0.18	0.21	0.22	0.26	0.28	***
Urea (mM)	2.26	3.12	3.63	1.87	2.76	3.57	2.96	3.17	4.62	ns
Uric acid (μM)	126	166	200	114	155	187	144	174	203	ns
log_{10} (NAGase)[c]	0.17	0.27	0.36	0.24	0.35	0.47	0.48	0.55	0.62	**
log_{10} (LDH)[c]	0.28	0.41	0.61	0.38	0.48	0.67	0.55	0.64	0.73	ns
Milk yield (kg/day)	34.3	36.4	40.6	32.1	34.6	38.6	29.9	33.0	36.7	ns

[a] Q1: first quartile, Q2: second quartile (median), Q3: third quartile, M: molar (mol/L).
[b] ns $P \geq 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
[c] BHB (μM), NAGase (units/L), LDH (units/L).
[d] The difference in totals compared to Table 3 is due to cows only having measures DIM 14.

for FT-MIR. Accuracies were lower for IgG *N*-glycans; ranging from 0.32 (0.17 − 0.51) to 0.53 (0.35 − 0.70). The sensitivity for prediction of the imbalanced cluster was better with MME than with FT-MIR and IgG *N*-glycans. Unfortunately, examples of zero sensitivity (none predicted correctly) were seen, likely due to a relatively low number of cows in the imbalanced clusters, see Table 3.

Results from predictions using only data that were complete across all three sets of milk biomarkers in each period are shown in Supplementary Table S2 and are less stable with confidence intervals that are bit wider due to the smaller number of observations. Nevertheless, predictions by IgG *N*-glycans tend to be less unfavourable compared with MME and FT-MIR when judged on this reduced data set, potentially giving a more fair comparison. Global accuracies tended to be lower with the reduced data set and ranged from 0.39 (95 % CI: 0.22 − 0.58) to 0.59 (0.45 − 0.72) for MME, 0.34 (0.19 − 0.53) to 0.67 (0.46 − 0.83) for FT-MIR and 0.19 (0.07 − 0.36) to 0.57 (0.37 − 0.76) for IgG *N*-glycans. Using this reduced data set, we also examined the pairwise agreement of predictions among the three sets of milk biomarkers, see Supplementary Table S3. The best agreement with a global accuracy of 0.76 (95 % CI: 0.62 − 0.87) was found between MME and FT-MIR for parity 3 + cows around DIM 14 but it should be noted that for these, none of the cows in the imbalanced cluster were correctly determined by FT-MIR. The lowest agreement was seen between FT-MIR and IgG *N*-glycans for parity 3 + cows around DIM 35 with a global accuracy of 0.27 (0.16 − 0.41). Generally, the agreements were at the same level among all three sets of milk biomarkers.

To ease comparison with table 6 in Grelet et al. (2019) and Fig. 5 in De Koster et al. (2019), we calculated the global accuracy for predicting the imbalanced cluster vs intermediate and balanced combined. For MME in parity 3 + this accuracy was 0.97 (0.92 − 0.99) and 0.82 (0.73 − 0.89) for DIM 14 and 35, respectively. For FT-MIR the corresponding accuracies were 0.89 (0.81 − 0.95) and 0.69 (0.59 − 0.78) and for IgG *N*-glycans 0.92 (0.82 − 0.97) and 0.53 (0.40 − 0.66). These accuracies tend to be higher DIM 14 and at the same level or lower DIM 35 than those found in Grelet et al. (2019) and De Koster et al. (2019). For parity 2, number of cows in the imbalanced clusters were quite low (see Table 3) and almost all sensitivity estimates were 0 and specificities at or close to 1 (see Table 4). Thus, parity 2 accuracies are high (e.g. 0.93 (0.83 − 0.98) for MME at 14 DIM) but driven by specificity.

3.5. Differences in milk metabolite contents among metabolic clusters

Considering further the characteristics of parity 2 and 3 + cows at DIM 14, Table 5 presents quartiles for milk yield, metabolites and enzymes for each of the three metabolic clusters. These results indicate that some of the milk metabolites and enzymes were significantly different between the three metabolic clusters. The concentration of free glucose was significantly lower in the imbalanced cluster while, generally, those of BHB and isocitrate were higher. For the parity 2 cows, glucose-6-phosphate, and free glucose concentrations were higher for the balanced cluster than for the imbalanced, while for BHB, isocitrate and NAGase the concentrations or activities were lower or tended ($P = 0.07$) to be lower for the balanced compared with the imbalanced cluster. For parity 3 + cows, glucose-6-phosphate did not differ between the metabolic clusters but otherwise the results were similar to those of second parity cows. For parity 3 + cows, the urea concentration also tended ($P = 0.07$) to be higher for the imbalanced cluster compared with the balanced cluster. To explore the ranking of importance within parity and period for the eight milk metabolites and enzymes in the MME set of milk biomarkers, VIM plots are shown in Supplementary Figs. S1–S4. BHB is among the most important for both the 14 and 35 DIM periods whereas isocitrate is important for both parity in the period around DIM 14 but only for the oldest (3 +) cows around DIM 35. For second lactation cows around DIM 35, free glucose and LDH were marginally more important than BHB which ranked third. For the oldest cows (3 +) free glucose was more important than isocitrate around DIM 14 whereas around DIM 35, uric acid and urea were also important for the prediction of the metabolic clusters.

4. Discussion

The objective was to compare the use of three different sets of milk biomarkers for identification of cows in physiological imbalance and thus at risk of developing a metabolic or infectious disease. The analysed data derived from six different countries with ten different diets, which should improve the external validity of the results.

IgG *N*-glycans performed really poorly compared with the other two sets of milk biomarkers for prediction of individual blood biomarkers, EBAL, PI-index and metabolic clusters. Since IgG *N*-glycans in humans appear to be associated mostly with ageing (Krištić et al., 2014; Yu et al., 2016) and inflammation (Dall'Olio et al., 2013) it is not surprising that the predictive power of the IgG *N*-glycans on measures associated with metabolic health in cattle was less promising. In addition, the poor predictive power may partly be due to a less dense sampling of this milk biomarker. Nevertheless, even when accounting for the difference in sampling density IgG *N*-glycans had lower prediction accuracies than MME, FT-MIR or both. In addition, the analytical procedure involves multiple steps and is more complex than other methods. Thus, also in that respect more work is needed to make this milk biomarker useful in herd health management.

The precision of predictions for the individual blood biomarkers, EBAL and PI-index was measured by the coefficient of determination of cross-validation (R^2_{cv}) and by the root mean squared error ($RMSE_{cv}$). These two measures of precision were interpreted with the recommendations from Alexander et al. (2015) in mind that as a rule of thumb the R^2 should higher than 0.6 and the RMSE within 10 % of the outcome's range.

To predict individual blood biomarkers, the best models were obtained by MME with R^2_{cv} of 0.62 (95 % CI: 0.53 − 0.69) and 0.59 (0.51 − 0.67) for plasma urea at 14 and 35 DIM, respectively. These were the only predictions reaching the 0.6 threshold mentioned above. Moreover, $RMSE_{cv}$ for MME predictions (0.72 (0.63 − 0.80) and 0.78 (0.69 − 0.87)) were below 10 % of the plasma urea range at 8.45 mM (supplementary tables of Krogh et al., 2019). The R^2_{cv} for FT-MIR models were generally lower than for MME and in some cases much lower, e.g. 0.06 (0.01 − 0.13) at DIM 14 and 0.13 (0.05 − 0.22) at DIM 35 for plasma urea. Correspondingly, the $RMSE_{cv}$ for FT-MIR were higher, e.g. 1.08 (0.97 − 1.18) and 1.13 (1.02 − 1.23) for plasma urea at 14 and 35 DIM. Based on the same data, Grelet et al. (2019) fitted FT-MIR models to predict blood glucose, IGF-1, NEFA and BHB that performed much better (R^2_{cv} 0.44, 0.61, 0.39 and 0.70) than found in the present study (R^2_{cv} 0.23, 0.36, 0.1, 0.27 at DIM 14 and 0.19, 0.35, 0.03 and 0.25 at DIM 35). This may be explained by different methodologies as Grelet et al. (2019) included milk yield and parity in the models and combined all DIM into one global model, increasing the ranges and thus the R^2 (Davies and Fearn, 2006). Further, the distribution of data were artificially modified in Grelet et al. (2019); a first derivative was applied to spectra and partial least squares regression was used instead of random forests. These differences were one of the reasons for redoing the FT-MIR predictions in the present paper. Comparison of FT-MIR models from different studies are even more difficult, due to differences in datasets characteristics and analytical and validation procedures. However, one of the central problems of predicting blood NEFA, BHB and glucose is that we want to predict the results of the few deviating samples with e.g., high BHB. Benedet et al. (2019) predicted blood BHB from milk FT-MIR. They found a R^2_{cv} of 0.64, which suggests the potential use of FT-MIR. However, when estimating the performance to detect cows with more than 1.2 mmol/L BHB in blood, they found a sensitivity of 0.28 and a specificity of 0.98. These results indicate that the prediction models appear to have a decent performance, but this is a result of predicting the vast majority of the samples within the normal range. The estimated sensitivity in the study by Benedet et al. (2019) is

too low to be useful for management at a cow level. Combined with results from other studies and the difficulties to interpret the performance of the predictive power of FT-MIR, this emphasizes the importance that such predictive models should be assessed based on a clear well-defined purpose of developing them (Kostoulas et al., 2017).

For EBAL, FT-MIR showed a better performance than MME with an R^2_{cv} of 0.28 (95 % CI: 0.24 − 0.33) vs 0.21 (0.18 − 0.25) whereas the opposite was the case when predicting PI-index with R^2_{cv} of 0.26 (0.16 − 0.36) vs 0.40 (0.29 − 0.50) at 14 DIM and 0.19 (0.10 − 0.30) vs 0.34 (0.23 − 0.44) at 35 DIM. Clearly these are below the 0.6 rule of thumb. The RMSEs from EBAL predictions (between 23.4 (22.1 − 24.6) and 26.3 (23.7 − 28.6)) were lower than 10 % of the absolute range, whereas for PI-index only RMSEs from MME predictions (1.62 (1.44 − 1.78) and 1.71 (1.48 − 1.91)) were around 10 % of the absolute range. That MME is not able to provide better prediction is somewhat surprising because it is close to using milk BHB and glucose to predict a composite measure of blood BHB, NEFA and glucose. However, other studies have also estimated the correlation between blood and milk BHB and found correlations of 0.64 (Enjalbert et al., 2001).

Metabolic clusters were created as alternative phenotypes. The global accuracy of predicting the metabolic clusters varied from 0.54 (0.39 − 0.68) to 0.65 (0.55 − 0.75) and 0.51 (0.37 − 0.65) to 0.68 (0.53 − 0.81) for MME and FT-MIR predictions, respectively. Thus, the performance of MME and FT-MIR was at an equal level. It should be noted that examples of sensitivity of zero and specificity close to one were seen and may have biased the accuracy upwards. There was no improvement when daily milk yield was included in the prediction models, as also concluded by Ingvartsen et al. (2003). It is not milk yield per se that increases the risk of diseases but rather physiological imbalance reflecting difficulties for some animals to adapt to the major physiological changes that occur particularly in the transition cow. Moreover, this is in accordance with results in Grelet et al. (2019) and De Koster et al. (2019) though comparison with these two studies is complicated by differences in examined periods and parities. The present study did notice differences in blood biomarker profiles among parities but more data would be desirable for such differentiation. In this study, work has focused on the first 7 weeks after calving and does not apply to cows at later stages. Since no clusters of primiparous cows were considered imbalanced, it generally seems that first parity cows do not require extra care and the attention should be on the multiparous cows, at least in the present study. Parity 2 also had relatively few cows in the imbalance clusters and sensitivity estimates of zero and specificities close to one. Thus, neither first nor second parity cows were really informative for the ability to predict the imbalanced cluster.

The purpose of the presented random forests algorithms were to identify cows in physiological imbalance at risk of developing subclinical or more severe stages of diseases. Such cows may need extra attention and potentially altered feeding or other management actions to prevent the physiological imbalance developing into subclinical or more severe disease states. Requirements for the accuracy of detection may be less demanding for this purpose since there is no risk of harm to the animal or use of medicine. The accuracies mentioned in this paper are likely too low for diagnosing diseases that require medical treatment with e.g., antibiotics. Generally, the required accuracy depends on the specific purpose and of e.g., disease prevalence, costs associated with treatment and possible side-effects. The required accuracy could be established by methods like decision trees (Rojo-Gimeno et al., 2018). Possibly, a larger data set for training prediction algorithms would improve the accuracies and the results presented here may be used to guide sample size decisions for future studies.

Presently, no sensors are available to measure e.g., free glucose, isocitrate and glucose-6-phosphate, but since FT-MIR algorithms tended to give as accurate predictions as MME, FT-MIR may give the same opportunities to make relevant classification of cows as balanced or in physiological imbalance (see also Grelet et al., 2019 and De Koster et al., 2019). Moreover, it would also be interesting to investigate direct prediction of udder inflammation from FT-MIR as opposed to the use of e.g., LDH and NAGase enzymes that constitute an alternative for somatic cell counts, helping in the detection of subclinical diseases (Kitchen et al., 1978; Larsen et al., 2010; Hovinen et al., 2016).

4.1. Conclusion

Neither EBAL nor PI-index were sufficiently precise to be used as a management tool for identification of risk cows. As an alternative, cows were divided into clusters based on measures of glucose, BHB and NEFA in plasma and IGF-1 in serum. These can be interpreted into metabolic clusters and the cluster of imbalanced cows can be predicted equally well by MME and FT-MIR. Nevertheless, accuracies still need to be improved and a larger data set for training the prediction algorithms would probably be needed. Free glucose, isocitrate, glycose-6-phosphate, BHB and NAGase measured in milk were significantly different among the three metabolic clusters (balanced, intermediate and physiological imbalanced). Thus, if MME is the preferred set of milk biomarkers to predict cows in physiological imbalance and at risk of developing a production or infectious disease, the above mentioned metabolites and enzyme should have high priority for inclusion. The use of IgG *N*-glycans for prediction still needs development. The prediction algorithms should be validated using an external data set although this was not possible in the present study.

Author's contribution

LF, CGa, MAK, MTS and KLI made the first draft of the paper. LF, CGr, MS, MH and other partners from GC undertook data handling and data quality control. LF, CGr, MS and MH did the major parts of the statistical analyses including the conception of the idea of using k-means clusters to combine selected blood biomarkers with contribution to the latter from MTS and KLI. LF, CGa, MAK, MTS and KLI collaboratively defined the metabolic interpretation of these clusters. MTS, MAC, KLI and other partners from GC did the conception and designed the study. TL handled storage of milk and blood samples and did lab analyses of milk metabolites, milk enzymes and blood metabolites and assisted during the data quality control of these biomarkers. CGr and other partners from GC undertook analyses and calibrations of FT-MIR. EM, ROF, FC, MAC and other partners from GC did lab analyses and interpretation of IgG *N*-glycans. All authors critically revised the first draft and approved the final version of the manuscript.

Funding

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 613689. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Ethics statement

The experiments were carried out in accordance with the standards recommended by the EU Directive 2010/63/EU for animal experiments.

Software and data repository resources

None of the data were deposited in an official repository.

Declaration of Competing Interest

There is no direct financial interest of the authors and affiliations in the subject matter discussed in the manuscript. All financial support is

identified in the Funding section.

Acknowledgements

The barn staff is acknowledged for their animal care work, Jens Clausen and Carsten Berthelsen, Aarhus University, for lab work and Martin Bjerring, Aarhus University, for data management. Dr L.J. Spicer, Oklahoma State University, is acknowledged for assistance with the IGF-1 radioimmunoassay and Dr Parlow, the National Hormone & Peptide Program (NHPP), for supplying the anti-hIGF-I, NHPP-NIDDK.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.prevetmed.2020.105006.

References

Alexander, D.L.J., Tropsha, A., Winkler, D.A., 2015. Beware of R^2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. J. Chem. Inf. Model. 55, 1316–1322. https://doi.org/10.1021/acs.jcim.5b00206.

Benedet, A., Franzoi, M., Penasa, M., Pellattiero, E., De Marchi, M., 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.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J.M., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman & Hall/CRC, Boca Raton, FL, USA.

Dall'Olio, F., Vanhooren, V., Chen, C.C., Slagboom, P.E., Wuhrer, M., Franceschi, C., 2013. N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. Ageing Res. Rev. 12, 685–698. https://doi.org/10.1016/j.arr.2012.02.002.

Davies, A., Fearn, T., 2006. Back to basics: calibration statistics. Spectrosc. Europe 18, 31–32,. accessed 8 Apr 2020. https://www.spectroscopyeurope.com/system/files/pdf/TD_18_2.pdf.

De Koster, J., Salavati, M., Grelet, C., Crowe, M., Opsomer, G., Foldager, L., GplusE Consortium, Hostens, M., 2019. Prediction of metabolic clusters in early lactation dairy cows using models based on milk biomarkers. J. Dairy Sci. 102, 2631–2644. https://doi.org/10.3168/jds.2018-15533.

Enjalbert, F., Nicot, M.C., Baourthe, C., Moncoulon, R., 2001. Ketone bodies in milk and blood of dairy cows: relationship between concentrations and utilization for detection of subclinical ketosis. J. Dairy Sci. 84, 583–589. https://doi.org/10.3168/jds.S0022-0302(01)74511-0.

Gengler, N., Soyeurt, H., Dehareng, F., Bastin, C., Colinet, F., Hammami, H., Vanrobays, M.L., Lainé, A., Vanderick, S., Grelet, C., Vanlierde, A., Froidmont, E., Dardenne, P., 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99, 4071–4079. https://doi.org/10.3168/jds.2015-10140.

Grelet, C., Fernández Pierna, J.A., Dardenne, P., Soyeurt, H., Vanlierde, A., Colinet, F., Gengler, N., Baeten, V., Dehareng, F., 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., Vanlierde, A., Hostens, M., Foldager, L., Salavati, M., Ingvartsen, K.L., Crowe, M., Sorensen, M.T., Froidmont, E., Ferris, C.P., Marchitelli, C., Becker, F., Larsen, T., Carter, F., GplusE Consortium, Dehareng, F., 2019. Potential of milk mid-IR spectra to predict metabolic status of cows through blood components and an innovative clustering approach. Animal 13, 649–658. https://doi.org/10.1017/S1751731118001751.

Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28, 100–108. https://www.jstor.org/stable/2346830.

Hovinen, M., Simojoki, H., Poso, R., Suolaniemi, J., Kalmus, P., Suojala, L., Pyorala, S., 2016. N-acetyl-beta-D-glucosaminidase activity in cow milk as an indicator of mastitis. J. Dairy Res. 83, 219–227. https://doi.org/10.1017/S0022029916000224.

Ingvartsen, K.L., 2006. Feeding- and management-related diseases in the transition cow: physiological adaptations around calving and strategies to reduce feeding-related diseases. Anim. Feed Sci. Technol. 126, 175–213. https://doi.org/10.1016/j.anifeedsci.2005.08.003.

Ingvartsen, K.L., Moyes, K.M., 2015. Factors contributing to immunosuppression in the dairy cow during the periparturient period. Jpn. J. Vet. Res. 63 (Suppl. 1), S15–S24. https://doi.org/10.14943/jjvr.63.suppl.s15.

Ingvartsen, K.L., Dewhurst, R.J., Friggens, N.C., 2003. On the relationship between lactational performance and health: is it yield or metabolic imbalance that cause production diseases in dairy cattle? A position paper. Livest. Prod. Sci. 83, 277–308. https://doi.org/10.1016/S0301-6226(03)00110-6.

Kitchen, B.J., Middleton, G., Salmon, M., 1978. Bovine milk N-acetyl-b-D-glucosaminidase and its significance in the detection of abnormal udder secretions. J. Dairy Res. 45, 15–20. https://doi.org/10.1017/S0022029900016149.

Kostoulas, P., Nielsen, S.S., Branscum, A.J., Johnson, W.O., Dendukuri, N., Dhand, N.K., Toft, N., Gardner, I.A., 2017. STARD-BLCM: standards for the reporting of diagnostic accuracy studies that use bayesian latent class models. Prev. Vet. Med. 138, 37–47. https://doi.org/10.1016/j.prevetmed.2017.01.006.

Krištić, J., Vučković, F., Menni, C., Klarić, L., Keser, T., Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K., Rudan, P., Novokmet, N., Sarac, J., Missoni, S., Kolčić, I., Polašek, O., Rudan, I., Campbell, H., Hayward, C., Aulchenko, Y., Valdes, A., Wilson, J.F., Gornik, O., Primorac, D., Zoldoš, V., Spector, T., Lauc, G., 2014. Glycans are a novel biomarker of chronological and biological ages. J. Gerontol. A Biol. Sci. Med. Sci. 69, 779–789. https://doi.org/10.1093/gerona/glt190.

Krogh, M.A., Hostens, M., Salavati, M., Grelet, C., Sorensen, M.T., Wathes, D.C., Ferris, C.P., Marchitelli, C., Signorelli, F., Napolitano, F., Becker, F., Larsen, T., Matthews, E., Carter, F., Vanlierde, A., Opsomer, G., Gengler, N., Dehareng, F., Crowe, M.A., Ingvartsen, K.L., Foldager, L., 2019. Between and within-herd variation in blood and milk biomarkers in Holstein cows in early lactation. Animal 14 (5), 1067–1075. https://doi.org/10.1017/S1751731119002659. e-pub ahead of print 7 Nov 2019.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28 (5). https://doi.org/10.18637/jss.v028.i05.

Larsen, T., 2014. Fluorometric determination of free and total isocitrate in bovine milk. J. Dairy Sci. 97, 7498–7504. https://doi.org/10.3168/jds.2014-8018.

Larsen, T., Moyes, K.M., 2015. Are free glucose and glucose-6-phosphate in milk indicators of specific physiological states in the cow? Animal 9, 86–93. https://doi.org/10.1017/S1751731114002043.

Larsen, T., Rontved, C.M., Ingvartsen, K.L., Vels, L., Bjerring, M., 2010. Enzyme activity and acute phase proteins in milk utilized as indicators of acute clinical E. coli LPS-induced mastitis. Animal 4, 1672–1679. https://doi.org/10.1017/S1751731110000947.

LeBlanc, S.J., Leslie, K.E., Duffield, T.F., 2005. Metabolic predictors of displaced abomasum in dairy cattle. J. Dairy Sci. 88, 159–170. https://doi.org/10.3168/jds.S0022-0302(05)72674-6.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22,. Accessed 8 April 2020. https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf.

Lyons, N.A., Cooke, J.S., Wilson, S., van Winden, S.C., Gordon, P.J., Wathes, D.C., 2014. Relationships between metabolite and IGF1 concentrations with fertility and production outcomes following left abomasal displacement. Vet. Rec. 174, 657. https://doi.org/10.1136/vr.102119.

Moyes, K.M., Bendixen, E., Codrea, M.C., Ingvartsen, K.L., 2013a. Identification of hepatic biomarkers for physiological imbalance of dairy cows in early and mid lactation using proteomic technology. J. Dairy Sci. 96, 3599–3610. https://doi.org/10.3168/jds.2012-5900.

Moyes, K.M., Larsen, T., Ingvartsen, K.L., 2013b. Generation of an index for physiological imbalance and its use as a predictor of primary disease in dairy cows during early lactation. J. Dairy Sci. 96, 2161–2170. https://doi.org/10.3168/jds.2012-5646.

Nielsen, N.I., Friggens, N.C., Chagunda, M.G.G., Ingvartsen, K.L., 2005. Predicting risk of ketosis in dairy cows using in-line measurements of beta-hydroxybutyrate: a biological model. J. Dairy Sci. 88, 2441–2453. https://doi.org/10.3168/jds.S0022-0302(05)72922-2.

Piechotta, M., Sander, A.K., Kastelic, J.P., Wilde, R., Heppelmann, M., Rudolphi, B., Schuberth, H.J., Bollwein, H., Kaske, M., 2012. Short communication: prepartum plasma insulin-like growth factor-I concentrations based on day of insemination are lower in cows developing postpartum diseases. J. Dairy Sci. 95, 1367–1370. https://doi.org/10.3168/jds.2011-4622.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria Accessed 8 April 2020. https://www.R-project.org.

Rojo-Gimeno, C., Fievez, V., Wauters, E., 2018. The economic value of information provided by milk biomarkers under different scenarios: case-study of an ex-ante analysis of fat-to-protein ratio and fatty acid profile to detect subacute ruminal acidosis in dairy cows. Livest. Sci. 211, 30–41. https://doi.org/10.1016/j.livsci.2018.02.001.

Yu, X., Wang, Y., Kristic, J., Dong, J., Chu, X., Ge, S., Wang, H., Fang, H., Gao, Q., Liu, D., Zhao, Z., Peng, H., Pucic Bakovic, M., Wu, L., Song, M., Rudan, I., Campbell, H., Lauc, G., Wang, W., 2016. Profiling IgG N-glycans as potential biomarker of chronological and biological ages: a community-based study in a Han Chinese population. Medicine (Baltimore). 95, e4112. https://doi.org/10.1097/MD.0000000000004112.