Predicting survival in dairy cattle by combining genomic breeding values and phenotypic information

E. M. M. van der Heide,[1]* **R. F. Veerkamp,**[1] **M. L. van Pelt,**[2] **C. Kamphuis,**[1] **and B. J. Ducro**[1]
[1]Wageningen University & Research Animal Breeding and Genomics, PO Box 338, 6700 AH, Wageningen, the Netherlands
[2]CRV BV, Animal Evaluation Unit, 6800 AL Arnhem, the Netherlands

ABSTRACT

Advances in technology and improved data collection have increased the availability of genomic estimated breeding values (gEBV) and phenotypic information on dairy farms. This information could be used for the prediction of complex traits such as survival, which can in turn be used in replacement heifer management. In this study, we investigated which gEBV and phenotypic variables are of use in the prediction of survival. Survival was defined as survival to second lactation, plus 2 wk, a binary trait. A data set was obtained of 6,847 heifers that were all genotyped at birth. Each heifer had 50 gEBV and up to 62 phenotypic variables that became gradually available over time. Stepwise variable selection on 70% of the data was used to create multiple regression models to predict survival with data available at 5 decision moments: distinct points in the life of a heifer at which new phenotypic information becomes available. The remaining 30% of the data were kept apart to investigate predictive performance of the models on independent data. A combination of gEBV and phenotypic variables always resulted in the model with the highest Akaike information criterion value. The gEBV selected were longevity, feet and leg score, exterior score, udder score, and udder health score. Phenotypic variables on fertility, age at first calving, and milk quantity were important once available. It was impossible to predict individual survival accurately, but the mean predicted probability of survival of the surviving heifers was always higher than the mean predicted probability of the nonsurviving group (difference ranged from 0.014 to 0.028). The model obtained 2.0 to 3.0% more surviving heifers when the highest scoring 50% of heifers were selected compared with randomly selected heifers. Combining phenotypic information and gEBV always resulted in the highest scoring models for the prediction of survival, and especially improved early predictive performance. By selecting the heifers with the highest predicted probability of survival, increased survival could be realized at the population level in practice.

Key words: dairy cow, survival, longevity, individual prediction

INTRODUCTION

Optimally, not all female calves born on a farm should be kept as replacements heifers to avoid unnecessary costs (Mohd Nor et al., 2012, 2015). To have insight into these costs, economic models have been developed to help farmers make the best choices in terms of heifer management (Mourits et al., 1997; Groenendaal et al., 2004). However, despite the availability of these models, many Dutch farmers keep more than the optimal number of replacement heifers (Mourits et al., 2000; Mohd Nor et al., 2015). Uncertainty about the survival and future performance of replacement heifers and dairy cows is one of the reasons for a farmer to keep a surplus of replacement heifers (Mohd Nor et al., 2015). Studies in the United Kingdom, France, and Sweden show that between 86 and 88% of heifer calves reach their first lactation (Hultgren et al., 2008; Brickell and Wathes, 2011; Raboisson et al., 2013), and between 83 and 96% of all first-lactation heifers survive to their second lactation (Dechow and Goodling, 2008; Bach, 2011; Brickell and Wathes, 2011). In the Netherlands, 86.6% of first lactation heifers born between 2009 and 2013 reached their second lactation (van Pelt et al., 2016b). Because it takes on average around 1.5 lactations to repay the rearing costs of a dairy cow, reducing the number of replacement heifers raised is important for farm profitability (Bach, 2011; Boulton et al., 2017). Surplus heifers may be sold, but this does not always cover the rearing costs (Mohd Nor et al., 2015), or may be used to replace older dairy cows that are voluntarily culled. However, because resilience and longevity of the dairy herd are becoming increasingly important from a societal and welfare point of view (Ortiz-Pelaez et al., 2008; LTO, 2011; Mohd Nor et al., 2014; Barkema et

al., 2015), it may become preferable to not cull older dairy cows simply because a younger replacement is available.

Despite the importance of survival, it is rarely included in heifer management models and even when survival is taken into account, there is no consideration for individual differences between heifers (Mohd Nor et al., 2015). Individual prediction of survival in heifers is often not attempted because of the difficulty of predicting phenotypic survival accurately. Survival traits are affected by a combination of production, fertility, and health traits (Heise et al., 2016), and environmental factors such as farm management. Furthermore, risk of culling is not constant over time. The complex nature of survival and the fact that true survival can only be measured in retrospect (the animal is culled or has died) has also led to many different definitions of survival traits being used (Essl, 1998; Fetrow et al., 2006). Estimated breeding values for at least one survival trait are available in most countries for breeding purposes (Forabosco et al., 2009). However, the heritability of these traits is almost invariably very low (van Pelt et al., 2015; Heise et al., 2016), making the phenotypic prediction of individual survival based on parent information alone not very accurate.

Combining available genomic estimated breeding values (**gEBV**) and phenotypic information might yield more accurate predictions for individual survival. Genotyping costs in dairy cattle have decreased to the point where it is economically feasible to obtain gEBV for production animals at birth (Pryce et al., 2012; Weigel et al., 2012; Calus et al., 2015). More phenotypic information on individual animals is also available on farms. The aim of this study is to investigate the possibility of combining phenotypic information and gEBV for the phenotypic prediction of survival. Five distinct moments were chosen for prediction of survival. These "decision moments" were moments in the lifetime of a dairy cow where new phenotypic information becomes available on which a management decision could be made.

MATERIALS AND METHODS

Data

A data set was obtained from cattle improvement cooperative CRV (Arnhem, the Netherlands) to investigate which gEBV and phenotypic variables could serve as predictors for survival. The data were available from animals born on 463 different Dutch and Flemish farms that participate in a "data plus" program, where additional information is gathered on dairy cows on commercial farms. All animals in this study were genotyped at birth, herd book registered, and had at least 87.5% Holstein blood. Survival was defined as the binary trait "survival until second calving, plus 2 wk," henceforth referred to as "survival to second lactation." This point in time was selected due to limitations of our data and because the lactation following second calving is economically significant (Bach, 2011). Two additional weeks were included to avoid counting animals that died during or as a consequence of their second calving. To have a known observation for survival, all cows included had to be born at least 46 mo before the end of data collection (March 2017), and were not exported abroad during this time. Animals exported abroad were excluded because they had unknown dates of death and could not be used in the analysis. Animals sold to other Dutch farms could be used because records from other farms were available in our data set. As the cause of death was unknown for all heifers, all deaths were included. This included deaths on farm, involuntary culling, and voluntary culling. Twenty-four animals were removed from the data set due to having only second calving records available. The final data set consisted of 6,847 female cows born between January 2012 and June 2013. Out of these cows, 5,872 (85.8%) cows survived until second lactation and 975 (14.2%) cows did not.

Five different data sets were created to predict survival with data available at 5 different "decision moments" during the life of a cow: at birth, at 18 mo of age, at first calving, 6 wk after first calving, and at 200 days after first calving. These moments were selected as they were points in time when more phenotypic information becomes available and where relevant decisions related to survival could be made. This includes not only the decision to cull or not cull an animal, but for example also if an animal should be inseminated again or not. The decision not to inseminate a heifer is effectively an early culling decision, as without getting pregnant it is not possible to enter the next lactation, even if the heifer is kept in the herd for another few months. Appendix Table A1 shows all variables, as well as the decision moment in which they become available, and if the variables were originally continuous. Information was cumulative; all records available on the first decision moment were also available during the subsequent decision moments. Each animal had 50 gEBV, scaled to a value between 0 and 10 where the largest value was set to 10 and the smallest value set to 0. These gEBV are direct genomic values, which did not include any own performance. If regular gEBV were used, own information included would have led to a residual covariance between the gEBV and survival. Using this regular gEBV in a survival prediction would have led to auto-correlation. Because historical gEBV for the vari-

ous decision moments were not available, only gEBV based solely on genotype were used. Phenotypic records were available on gestation duration and dam parity, herd book status, birth records, calving records, records on moves between different farms, insemination records of the first and second parity, and first parity milk records. Calving ease scores were scored 1 to 6, where 1 was an easy birth, 2 was normal, and scores 3 to 6 were considered difficult births, because they denoted long labor and various veterinarian interventions. The milk records at "6 wk postcalving" were of the most recent milk test day record before the end of 6 wk. Milk records in the decision moment "200 d postcalving" were an average of all milk records available of an animal up to that point. Fertility records were used to determine nonreturn status at the decision moment "18 mo of age" and "200 d postcalving." If a heifer had AI records, but received no AI in the 56 d before the 2 aforementioned decision moments, it was listed as nonreturn, because it had not received another insemination at least 56 d after the last insemination. Heifers with an insemination in the 56 d before the decision moment, no recorded inseminations at that point in time or subject to natural mating were listed as unknown. In these cases the exact insemination or conception date was unknown or it could not otherwise be determined if the animal was pregnant. Animals with missing records were always included in the data sets. Missing records could have reflected active management decisions (e.g., not inseminating an animal) or could reflect, for example, a fertility problem, and thus could have been useful in prediction. This was modeled by adding a class "unknown" to all factorial variables to identify missing values. Continuous variables with missing values were transformed into factors with between 5 and 8 classes, depending on the distribution of individual variables. We chose to select at least 5 levels to keep sufficient variation within each variable. In total, each animal had 50 gEBV (available at each decision moment), and up to 62 additional phenotypic variables accumulating throughout their lifetime.

Model and Analysis

To determine if survival to second lactation could be predicted, logistic regression models were constructed for each decision moment. Each of the 5 data sets (one for each decision moment) were split into sets of 70% training and 30% testing data, stratified by survival group to ensure a representative amount of both survival groups in the testing and training sets. Stratified sampling meant that the training and the testing data sets included identical proportions of nonsurviving and surviving heifers. We used the statistical program R (R Core Team 2016), version 3.3.1, and the package 'caret' (Kuhn, 2008) to select models for each decision moment. Both forward stepwise selection and stepwise selection combining forward and backward stepwise selection were tested on the 5 training data sets using the following general model:

$$\text{Logit } (P) = \beta_0 + \Sigma \beta_i X_i,$$

where logit (P) is the estimated probability of survival, β_0 is the population mean, and $\beta_i X_i$ is the set of predictor variables, consisting of phenotypic variables and gEBV. The Akaike information criterion (**AIC**) was used to determine the best possible model for each decision moment. The stepwise procedure combining forward and backward selection resulted in the models with the lowest AIC value and thus only the models derived using this procedure were used for analysis. The stepwise procedure was first used on the training sets containing both gEBV and phenotypic variables to identify which variables are significant for the prediction of survival. The models were tested on their corresponding test data set to get probabilities of survival for the heifers in the test set. By applying the model to the data, predicted probabilities of survival were obtained for each heifer. These were values between 0 and 1, where values close to 1 indicate a high probability to reach the second lactation, and values closer to 0 a lower probability to reach second lactation. To compare and validate the models, the prediction accuracy, specificity, sensitivity, positive predictive value, and negative predictive value were calculated. The predicted probability of survival was transformed into a survival prediction of 1 when equal to or above the average probability for survival in our data (0.858), and to a survival prediction of 0 below the average probability of survival. The accuracy was the proportion of correct predictions. The sensitivity was calculated as the true positive outcomes divided by the true positives plus the false negatives, and the specificity as the true negatives divided by the true negatives plus the false positives. The positive predictive value is the true positive value divided by the true positive value plus the false positive value, and similarly the negative predictive value is the true negative value divided by the true negative value plus the false negative value. We also calculated the balanced accuracy and the area under the receiver operating curve (**ROC**), the area under the curve (**AUC**) value. The balanced accuracy is the average of the accuracy for nonsurviving heifers and the accuracy of the surviving heifers. The AUC value is the predictive ability of the model including all possible cut-offs and was calculated

using the pROC package (Robin et al., 2011). These metrics were selected as survival was an imbalanced trait, with more survivors than nonsurvivors. Both balanced accuracy and AUC value are more robust against imbalanced predictors than accuracy.

To investigate if there was merit in including both gEBV and phenotypic traits, we also selected models for each decision moment using either only gEBV or only phenotypic information as input for the variable selection. For each decision moment, the 3 resulting models were tested on a new 70/30 split of training and testing data. The AIC was recorded as indication of model fit and the AUC value calculated using pROC as an indication of model performance.

RESULTS

Figure 1 shows the distribution of age in days at death for heifers in our data set. This figure indicates that in our data set most heifers reached 18 mo of age, with few early deaths. The number of culled and dead animals increased after 18 mo, most due to the availability of phenotypic records on which selective culling could take place.

An overview of the gEBV and phenotypic variables selected using stepwise selection for each decision moment when including only living heifers at a decision moment is shown in Table 1. The coefficients of the selected variables are shown in Appendix Tables A2, A3, A4, A5, and A6, separated by decision moment. The gEBV for longevity was selected at all decision moments, and was in each case positively associated

with survival. While different gEBV were selected at different decision moments, at each decision moment they broadly fell into the same categories: fertility, exterior score, udder score, udder health, and feet and leg gEBV. The gEBV for production variables were not prominent among the selected variables, although they were selected at the first 3 decision moments. At 200 d after first calving, several gEBV regarding longevity, exterior score, udder conformation, and health and fertility were still selected. The phenotypic variables selected at birth were season of birth and year of birth. Year of birth and season of birth were selected up to the last decision moment, where season of first calving was selected as an alternative for season of birth. Phenotypic information new to a decision moment was always selected, with some variables remaining important over the next decision moment(s). Age at first calving, for example, was strongly associated with survival in all decision moments after first calving (Appendix Tables A4 to A6). Earlier calving ages had a more positive association with survival. Phenotypic fertility information appears important in general, as nonreturn status and number of inseminations were selected in both decision moments when these variables became available (Appendix Tables A3 and A6). At 200 d after calving, both the number of inseminations at 18 mo and at 200 d after calving were selected. Phenotypic information on production traits was also important: kg of milk produced at the milk test day closest to 6 wk postcalving and the average milk production per test milk day at 200 d after calving were both significantly associated with survival (Appendix Tables A5 and A6).
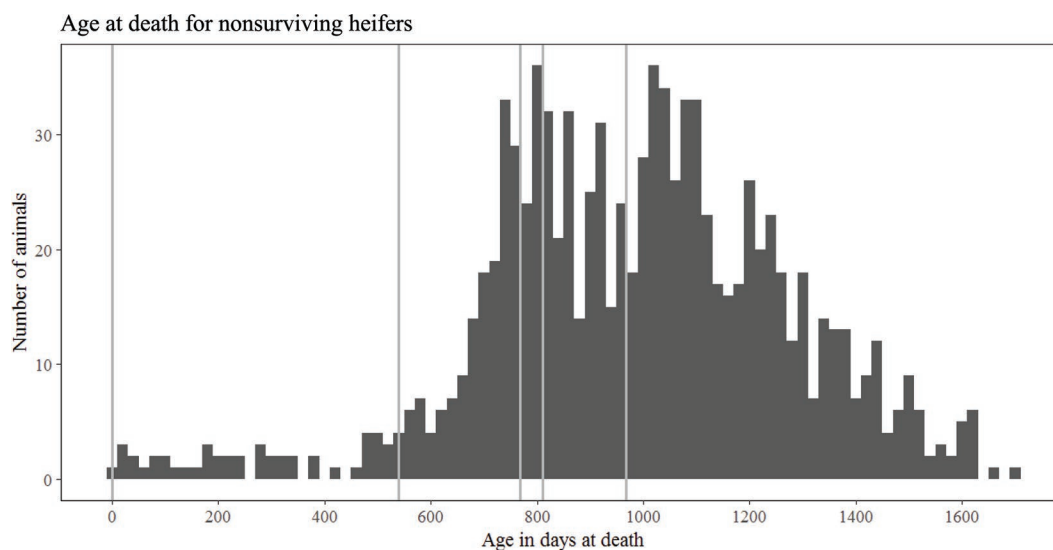

Figure 1. The distribution of age at death in days for nonsurviving heifers. Gray lines indicate the decision moments. The decision moment at first calving was set at the average age at first calving in the data set.

Table 1. Selected variables for the best model in each decision moment[1]

Variable	Estimate for selected variable decision moments				
	Birth	18 mo of age	At first calving	6 wk after first calving	200 d after first calving
gEBV					
Longevity	0.128	0.212	0.261	0.242	0.207
Foot angle	−0.115	−0.078	−0.086	−0.085	
Udder depth	−0.102	−0.137	−0.140		−0.106
Frame	0.074				
kg of fat	0.061	0.122			
Nonreturn at 56 d	−0.092			−0.088	
Interval first–last insemination	0.140			−0.213	0.106
Overall fertility			0.145	0.091	0.242
Milking speed		0.061			
NVI[2]		−0.122			
Overall exterior score		0.143			0.256
SCC			0.196	0.212	
Udder health			−0.283	−0.289	−0.153
Rear udder height			0.072		
Feet and legs			−0.066		
Chest width			0.099	0.131	
Rear legs hind view				−0.162	
Locomotion				0.125	
kg of lactose				−0.073	
Stature					−0.094
Udder support					0.074
Phenotypic variables					
Season of birth	0.434	0.474	0.423	0.600	
Year of birth	−0.372	−0.268	−0.270	−0.342	−0.239
Number of inseminations at 18 mo		1.305			1.553
Coat color			−0.208		
Nonreturn status at 18 mo		−0.426			
Season of first calving					0.725
Age in days at first calving			1.700	1.881	2.831
Sex of first calf			0.378	0.272	
Calf survival 1 wk after birth					14.392
kg of milk produced at milk test day closest to 6 wk postcalving				2.512	
Total number of transports at 200 d postcalving					15.215
Average kg of milk per test milk day at 200 d postcalving					1.899
Average cell count (× 1,000) per test milk day at 200 d postcalving					1.032
Average percentage of protein per test milk day at 200 d postcalving					1.414
Negative indication at test milking 6 wk postcalving					1.401
Number of negative indications at test milk days before 200 d postcalving					−0.788
Number of inseminations at 200 d postcalving					1.553
Nonreturn status at 200 d postcalving					−1.065

[1]The table shows the coefficients for the genomic estimated breeding values (gEBV). For the factorial phenotypic variables, this table shows the difference in coefficient between the highest and the lowest class for a variable. The coefficient of the referent class is 0. The maximum difference was used because class variables have more than one coefficient. For the exact coefficients for each variable, see Appendix Tables A2 to A6.

[2]NVI = the Dutch/Flemish merit index score.

Furthermore, the average percentage of protein per test milk day was selected at 200 d postcalving. Phenotypic traits on udder health were also selected as number of negative outcomes, which include mastitis and other illnesses, negative indication (yes/no) at test milking closest to 6 wk postcalving, and average cell count per test milk day at 200 d postcalving.

To determine if adding genotype information has additional value, we compared a model using only genotypes, a model using only phenotypes, and the results of the combined model (Table 2). The combined model has the highest AUC value at all decision moments, except at first calving where all methods perform equally, and at 200 d post-first calving, where it performs equal to the model containing only phenotypic variables. The AIC value was always highest for the combined model, as a model containing both gEBV and phenotypic variables was always selected through stepwise selection (Table 1).

On average, the predicted probability for the surviving group was between 0.014 (at birth) and 0.028 (at 200 d post-first calving) higher than the probability of survival for the nonsurviving group (Table 3). This means that while overlap occurred, a difference was

Table 2. The area under the curve (AUC) of the receiver operator curve (ROC) value for a model including both genomic estimated breeding values (gEBV) and phenotypic information, only phenotype information, and only gEBV

Decision moment	gEBV only	Phenotype only	gEBV and phenotype
Birth	0.557	0.562	0.584
18 mo	0.580	0.594	0.606
First calving	0.597	0.597	0.596
6 wk after first calving	0.560	0.646	0.677
200 d after first calving	0.573	0.731	0.731

observed between the 2 groups on average. The accuracy of the models increased from 0.562 to 0.776 in later decision moments (Table 4). The AUC values also increased from 0.578 at birth to 0.648 at 200 d postcalving. This means that the model improved at later decision moments. However, the balanced accuracy did not increase, remaining around 0.56 at all decision moments. It appears that the models improved by predicting surviving heifers better. This can be seen from an increase in the sensitivity of the models from 0.566 at birth to 0.816 at 200 d postcalving and an increase in positive predictive value from 0.880 to 0.933. This means that both a larger proportion of all surviving heifers in the data set were predicted correctly as surviving, and a larger proportion of the heifers predicted to survive actually survive. In contrast, although the model was able to better predict which heifers survived, predicting which heifers did not survive proved more difficult. The negative predictive value did not increase consistently, and was actually lowest at 200 d postcalving. Furthermore, the specificity of the model varied inconsistently, ranging from 0.295 to 0.534, meaning that the proportion of nonsurviving heifers that is identified did not improve.

Rather than focusing on individual prediction, a situation could be considered where the models were used to select 50% of heifer calves to become replacement heifers. Table 5 shows the average probability of survival in a decision moment and the probability

Table 3. Mean and SD of the predicted probability of survival for the survival and the nonsurvival heifer groups in the testing set of each decision moment, including only alive heifers at the start of the decision moment

Decision moment	Survival = no		Survival = yes	
	Mean	SD	Mean	SD
Birth	0.847	0.049	0.861	0.044
18 mo	0.842	0.075	0.868	0.060
At first calving	0.873	0.073	0.897	0.052
6 wk after first calving	0.869	0.072	0.896	0.052
200 d after first calving	0.867	0.064	0.896	0.054

of survival of the selected group. For example, in the first decision moment, 85.8% of calves reached second lactation in a random selection, compared with 88.6% of the calves selected through the model. This meant that out of the 1,026 selected calves in our testing data set, 909 of heifers selected by our model would reach second lactation, compared with 880 random heifers. In subsequent decision moments, using a model resulted in up to 3% more surviving heifers compared with random selection.

DISCUSSION

Our research showed that it was possible to predict the survival outcome of heifers at a population level. Predicting nonsurviving animals proved difficult even with large amounts of phenotypes and gEBV available. This was not unexpected because a cow could have been culled for a myriad of reasons that often influence each other, and the decision to cull is time dependent and based on decisions made by individual farmers (Hadley et al., 2006; Zijlstra et al., 2013). Our data set also included unpredictable causes of death such as random accidents (Brickell and Wathes, 2011). Prediction may have been improved by including the exact causes of death or culling (e.g., allowing us to remove deaths caused by accidents), but this information was not available.

The difficulty of predicting survival meant that individual predictions obtained by our models were too inaccurate for the purpose of identifying nonsurviving heifers. Because the negative predictive value ranged from 0.12 to 0.17, any heifer predicted as not surviving only had a 12 to 17% chance of not reaching second lactation. However, if applied on a large group of individuals, it was possible to use these models to select the heifers with the highest probability of survival. When 50% of the heifers with the highest probability of survival were selected, 2.0 to 3.0% more heifers reached second lactation compared with random selection. When selecting at birth, this would have resulted in a 2.8% increase in surviving heifers. Although a 2.8% increase does not seem like a large improvement, this represented a 15.5% reduction of nonsurviving heifers.

The selected variables gave insight into which variables were associated with survival at the various decision moments. The first variables selected were birth season and year of birth. Because these variables were cohort variables, these variables were not animal specific and did not distinguish between calves born in the same season and year. This meant they could not be used for an individual farmer to distinguish between calves born in the same season or year. Birth season and year are important correction factors at the population

Table 4. The accuracy, balanced accuracy, sensitivity, specificity, positive and negative predictive values, and area under the curve (AUC) values of the models at each decision moment[1]

Decision moment	Accuracy[2]	Balanced accuracy[3]	Specificity[4]	Sensitivity[5]	Neg. pred. value[6]	Pos. pred. value[7]	AUC[8]
Birth	0.562	0.550	0.534	0.566	0.170	0.880	0.578
18 mo	0.634	0.577	0.498	0.656	0.186	0.893	0.606
At first calving	0.756	0.545	0.276	0.813	0.150	0.904	0.594
6 wk after first calving	0.755	0.567	0.329	0.806	0.168	0.910	0.620
200 d after first calving	0.776	0.555	0.295	0.816	0.117	0.933	0.648

[1]Survival was scored as 1 when the predicted probability of survival was above the average probability of survival (0.858) and scores below the average probability of survival were scored as 0.

[2]The proportion of correct predictions.

[3]The average of the accuracy for nonsurviving heifers and the accuracy of the surviving heifers.

[4]The true positive outcomes divided by the true positives plus the false negatives.

[5]The true negatives outcomes divided by the true negatives plus the false positives.

[6]The true positive outcomes divided by the true positive outcomes plus the false positive outcomes.

[7]The true negative outcomes divided by the true negative outcomes plus the false negative outcomes.

[8]The AUC of the receiver operating characteristic.

level, however, and were selected in all but the last decision moment, where calving season was used instead of season of birth. Production gEBV were surprisingly not selected at every decision moment. This may be explained by the fact that longevity was uncorrected for production, and thus possibly served as a substitute for production gEBV. Phenotypic production traits such as (average) kg of milk produced were selected once available, because there is a known increased risk of culling for animals with low milk production (Hadley et al., 2006). Most of the other associations found were not surprising. The association between feet and leg traits and survival is well known in literature (Buenger et al., 2001; Caraviello et al., 2004). The gEBV for foot angle was associated with lameness specifically (Wells et al., 1993), which is an important reason for culling (Olechnowicz and Jaskowski, 2011; Zijlstra et al., 2013). Udder conformation traits (Caraviello et al., 2004; Kern et al., 2015) and udder health (Mohd Nor et al., 2014) are known to be strongly associated with survival. Udder and health traits are correlated (Carlström et al., 2016), and udder traits have even been

suggested as candidate traits for indirect prediction of longevity (Kern et al., 2015). Udder health phenotype variables were selected in the form of negative indication counts and SCC, which is associated with increased risk of culling (Beaudeau et al., 2000). The exterior score gEBV selected in this study were mostly related to size. Body size has been associated with longevity and efficiency (Getu and Misganaw, 2015; Kern et al., 2015); however, the direction of the associations found varied, and some studies find no effect at all (Sewalem et al., 2004). Even when studies find negative associations, culling decisions may be influenced by a (regional) farmer preference for larger cows (Hansen et al., 1999; Caraviello et al., 2004). Lastly, fertility had been reported as one of the most important reason for culling in Dutch Holstein cows (Zijlstra et al., 2013), and there is a well-documented relation between fertility and survival (Pritchard et al., 2013). Age at first calving, which is also an indication of fertility, was always selected when available. A higher age at first calving was associated with higher costs, lower fertility, and higher risk of culling overall (Sewalem et al., 2008; van Pelt et al., 2016a). Interestingly, despite the availability of 2 parities worth of phenotypic fertility records, fertility gEBV were still selected by the combined model in the fifth decision moment.

The selection criteria for our data were very stringent because they included only animals genotyped at birth. This was done to avoid a winners' bias; genotyping was more expensive previously, and so only promising heifers and proven cows were genotyped. This means that cows that were genotyped either already had reached second lactation, or were more likely to reach the second lactation than an average cow in the population. Less than 7% of heifers do not reach the second lacta-

Table 5. The average chance of survival at a decision moment, the average chance of survival for the animals in the top 50% predicted probabilities of survival, and the difference in percentage between the two

Item	Probability of survival (%)		Difference (%)
	Mean	Predicted top 50%	
Birth	85.8	88.6	2.8
18 mo	86.4	89.4	3.0
At first calving	89.0	91.4	2.4
6 wk after first calving	90.4	92.4	2.0
200 d after first calving	92.3	95.3	3.0

tion if we included all genotyped cows in our data set. As we did not include all genotyped heifers, the heifers in our data set were limited to those born within a period of around a year and a half. This meant it was impossible to investigate true survival or beyond the fourth lactation, since these lactations would require an opportunity group born before the heifers available in our data set. The limited number of heifers also explains why some variables such as "number of movements" and "calf survival at 2 wk" were selected. Both variables had very strong associations with survival for some classes (Appendix Table A6). These classes had less than 10 heifers each, all of which survived by chance. All heifers of a class surviving resulted in a strong association with survival for these variables. In future studies, classes with very few cases could be excluded or merged with other classes, if they prove problematic for the analysis. Increasing the amount of animals in the data set should also reduce the number of very small classes. In this study, as survival is difficult to predict and information was already limited on some causes of death, we chose not to merge or remove classes, resulting in some artifacts. Another effect of the limited number of heifers was that many farms only had a small number of calves. The average number of calves per farm was 15, with many farms having fewer calves and some farms having a much larger number of calves. Farm identification numbers could have functioned as a proxy for herd and farm management, which were not included in the data. However, farm identification numbers were not selected due to the large number of farms with very few calves. Next to requiring all heifers to be genotyped, this study included only phenotypic information that was readily available. This meant that information gaps still existed at each decision moment. For example, between birth and 18 mo of age only fertility variables and movements between farms were recorded on a large scale. Variables such as the occurrence of diseases (Svensson and Hultgren, 2008; Heinrichs and Heinrichs, 2011) or simple body (size) measurements of calves (Wathes et al., 2008) could have provided valuable information on the health and development of the animal during the first 18 mo of life. Weight or size of the heifer may also be useful for predicting fertility, because while earlier first calving ages seemed preferable, it may be beneficial for some heifers to be inseminated later as early inseminations may have detrimental effects (Hoffman and Funk, 1992; Heinrichs, 1993).

Because a multiple regression was used in this study, the direction and the strength of the association of the variables with survival was only valid in the context of the whole model. It is important in the interpretation that the values cannot be taken individually.

For example, gEBV for interval between first and last insemination had both a negative and a positive associations with survival depending on the decision moment. Appendix Table A7 shows an example of the differences in direction and strength of associations when some variables available at birth were tested in a single or multiple regression model. The differences between a single and multiple regression model could be explained by high correlations between some variables, because collinearity is known to cause issues with multiple regression (Whittingham et al., 2006; Yoo et al., 2014). Antagonistic relations between multiple gEBV that were both associated with survival (such as fertility and production; Zavadilová and Zink, 2013) could also explain some of the unusual relations found. The combination of various related gEBV could be somewhat mitigated by more stringent selection of the gEBV variables provided. Each decision moment appeared to included one or multiple selected gEBV of several groups: feet and legs, udder, udder health, exterior, and fertility. We tested the consistent selection of groups of gEBV by building a model that included gEBV for longevity, overall fertility, udder score, udder health, feet and legs, and overall exterior at all decision moments. These general models also included the phenotypic records normally selected at each decision moment. The general models performed identically to the models described in this paper (data not shown).

This study shows that there is merit in the combination of phenotypic information and gEBV for the prediction of survival, because gEBV were always selected in combination with phenotypic information. Accuracies for the combined model were also higher than for models using only gEBV or phenotype information in all but the last decision moment. The literature already shows an increased interest in multiple fields to develop methods to combine phenotypic and genomic information for various purposes (Javed et al., 2014; Blake et al., 2016; Haendel et al., 2016), and has proven valuable for example in disease prognosis in human diseases (Perlee et al., 2013; Javed et al., 2014). In cattle, a recent paper estimating the lifetime profitability of a dairy cow also combines gEBV with a small amount of phenotypic variables to obtain more accurate estimates (Kelleher et al., 2015). Because there is interest in combining genotype and phenotypic information, future research could explore the exact benefit of including genotypes for performance measures, as well as investigate other methods compared with multiple regression.

CONCLUSIONS

In this study, genomic information in the form of gEBV was combined with phenotypic information to

predict survival to second lactation in Holstein dairy cows at 5 different decision moments. A combination of gEBV and phenotypic information resulted in better models than using only one type of information. The addition of gEBV especially improved early prediction. A combination of gEBV and phenotypic information also resulted in the best predictive performance up to the last decision moment. While accurate individual prediction of survival outcome could not be achieved, surviving heifers were predicted to have a higher probability of survival than nonsurviving heifers on average. By selecting the heifers with the highest predicted probability of survival, increased survival could be realized at the population level in practice.

ACKNOWLEDGMENTS

REFERENCES

Bach, A. 2011. Associations between several aspects of heifer development and dairy cow survivability to second lactation. J. Dairy Sci. 94:1052–1057.

Barkema, H. W., M. Von Keyserlingk, J. Kastelic, T. Lam, C. Luby, J.-P. Roy, S. LeBlanc, G. Keefe, and D. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98:7426–7445.

Beaudeau, F., H. Seegers, V. Ducrocq, C. Fourichon, and N. Bareille. 2000. Effect of health disorders on culling in dairy cows: A review and a critical discussion. Pages 293–311 in Annales de Zootechnie. Vol. 49, No. 4. EDP Sciences.

Blake, V. C., C. Birkett, D. E. Matthews, D. L. Hane, P. Bradbury, and J.-L. Jannink. 2016. The Triticeae toolbox: Combining phenotype and genotype data to advance small-grains breeding. Plant Genome 9. https://doi.org/10.3835/plantgenome2014.12.0099.

Boulton, A., J. Rushton, and D. Wathes. 2017. An empirical analysis of the cost of rearing dairy heifers from birth to first calving and the time taken to repay these costs. Animal 1–9.

Brickell, J. S., and D. Wathes. 2011. A descriptive study of the survival of Holstein-Friesian heifers through to third calving on English dairy farms. J. Dairy Sci. 94:1831–1838.

Buenger, A., V. Ducrocq, and H. Swalve. 2001. Analysis of survival in dairy cows with supplementary data on type scores and housing systems from a region of Northwest Germany. J. Dairy Sci. 84:1531–1541.

Calus, M. P., P. Bijma, and R. Veerkamp. 2015. Evaluation of genomic selection for replacement strategies using selection index theory. J. Dairy Sci. 98:6499–6509.

Caraviello, D. Z., K. Weigel, and D. Gianola. 2004. Analysis of the relationship between type traits and functional survival in US Holstein cattle using a Weibull proportional hazards model. J. Dairy Sci. 87:2677–2686.

Carlström, C., E. Strandberg, K. Johansson, G. Pettersson, H. Stålhammar, and J. Philipsson. 2016. Genetic associations of in-line recorded milkability traits and udder conformation with udder health. Acta Agric. Scand. A Anim. Sci. 66:84–91.

Dechow, C. D., and R. Goodling. 2008. Mortality, culling by sixty days in milk, and production profiles in high-and low-survival Pennsylvania herds. J. Dairy Sci. 91:4630–4639.

Essl, A. 1998. Longevity in dairy cattle breeding: A review. Livest. Prod. Sci. 57:79–89.

Fetrow, J., K. Nordlund, and H. Norman. 2006. Invited review: Culling: Nomenclature, definitions, and recommendations. J. Dairy Sci. 89:1896–1905.

Forabosco, F., J. Jakobsen, and W. Fikse. 2009. International genetic evaluation for direct longevity in dairy bulls. J. Dairy Sci. 92:2338–2347.

Getu, A., and G. Misganaw. 2015. The role of conformational traits on dairy cattle production and their longevities. Open Access Library Journal 2:7.

Groenendaal, H., D. Galligan, and H. Mulder. 2004. An economic spreadsheet model to determine optimal breeding and replacement decisions for dairy cattle. J. Dairy Sci. 87:2146–2157.

Hadley, G. L., C. Wolf, and S. Harsh. 2006. Dairy cattle culling patterns, explanations, and implications. J. Dairy Sci. 89:2286–2296.

Haendel, M. A., M. G. Kann, and N. L. Washington. 2016. Innovative approaches to combining genotype, phenotype, epigenetic, and exposure data for precision diagnostics. Pages 93–95 in Biocomputing 2016: Proceedings of the Pacific Symposium.

Hansen, L. B., J. Cole, G. Marx, and A. Seykora. 1999. Productive life and reasons for disposal of Holstein cows selected for large versus small body size. J. Dairy Sci. 82:795–801.

Heinrichs, A. J. 1993. Raising dairy replacements to meet the needs of the 21st century. J. Dairy Sci. 76:3179–3187.

Heinrichs, A. J., and B. Heinrichs. 2011. A prospective study of calf factors affecting first-lactation and lifetime milk production and age of cows when removed from the herd. J. Dairy Sci. 94:336–341.

Heise, J., Z. Liu, K. F. Stock, S. Rensing, F. Reinhardt, and H. Simianer. 2016. The genetic structure of longevity in dairy cows. J. Dairy Sci. 99:1253–1265.

Hoffman, P., and D. A. Funk. 1992. Applied dynamics of dairy replacement growth and management. J. Dairy Sci. 75:2504–2516.

Hultgren, J., C. Svensson, D. O. Maizon, and P. A. Oltenacu. 2008. Rearing conditions, morbidity and breeding performance in dairy heifers in southwest Sweden. Prev. Vet. Med. 87:244–260.

Javed, A., S. Agrawal, and P. C. Ng. 2014. Phen-Gen: Combining phenotype and genotype to analyze rare disorders. Nat. Methods 11:935.

Kelleher, M. M., P. Amer, L. Shalloo, R. Evans, T. Byrne, F. Buckley, and D. P. Berry. 2015. Development of an index to rank dairy females on expected lifetime profit. J. Dairy Sci. 98:4225–4239.

Kern, E. L., J. A. Cobuci, C. N. Costa, C. M. McManus, and J. Braccini Neto. 2015. Genetic association between longevity and linear type traits of Holstein cows. Sci. Agric. 72:203–209.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28:1–26.

LTO. 2011. Melkveehouderij: Midden in de maatschappij–Visie. LTO Nederland Vakgroep Melkveehouderij, Den Haag, the Netherlands.

Mohd Nor, N., W. Steeneveld, and H. Hogeveen. 2014. The average culling rate of Dutch dairy herds over the years 2007 to 2010 and its association with herd reproduction, performance and health. J. Dairy Res. 81:1–8.

Mohd Nor, N., W. Steeneveld, M. Mourits, and H. Hogeveen. 2012. Estimating the costs of rearing young dairy cattle in the Netherlands using a simulation model that accounts for uncertainty related to diseases. Prev. Vet. Med. 106:214–224.

Mohd Nor, N., W. Steeneveld, M. Mourits, and H. Hogeveen. 2015. The optimal number of heifer calves to be reared as dairy replacements. J. Dairy Sci. 98:861–871.

Mourits, M. C., A. Dijkhuizen, R. Huirne, and D. Galligan. 1997. Technical and economic models to support heifer management decisions: Basic concepts. J. Dairy Sci. 80:1406–1415.

Mourits, M. C., H. Van der Fels-Klerx, R. Huirne, and M. Huyben. 2000. Dairy-heifer management in the Netherlands. Prev. Vet. Med. 46:197–208.

Olechnowicz, J., and J. M. Jaskowski. 2011. Reasons for culling, culling due to lameness, and economic losses in dairy cows. Med. Weter. 67:618–621.

Ortiz-Pelaez, A., D. Pritchard, D. Pfeiffer, E. Jones, P. Honeyman, and J. Mawdsley. 2008. Calf mortality as a welfare indicator on British cattle farms. Vet. J. 176:177–181.

Perlee, L. T., A. T. Bansal, K. Gehrs, J. S. Heier, K. Csaky, R. Allikmets, P. Oeth, T. Paladino, D. H. Farkas, and P. L. Rawlings. 2013. Inclusion of genotype with fundus phenotype improves accuracy of predicting choroidal neovascularization and geographic atrophy. Ophthalmology 120:1880–1892.

Pritchard, T., M. Coffey, R. Mrode, and E. Wall. 2013. Genetic parameters for production, health, fertility and longevity traits in dairy cows. Animal 7:34–46.

Pryce, J., B. Hayes, and M. Goddard. 2012. Genotyping dairy females can improve the reliability of genomic selection for young bulls and heifers and provide farmers with new management tools. Proceedings of the 38th ICAR Session 28.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raboisson, D., F. Delor, E. Cahuzac, C. Gendre, P. Sans, and G. Allaire. 2013. Perinatal, neonatal, and rearing period mortality of dairy calves and replacement heifers in France. J. Dairy Sci. 96:2913–2924.

Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12:77.

Sewalem, A., G. Kistemaker, F. Miglior, and B. Van Doormaal. 2004. Analysis of the relationship between type traits and functional survival in Canadian Holsteins using a Weibull proportional hazards model. J. Dairy Sci. 87:3938–3946.

Sewalem, A., F. Miglior, G. J. Kistemaker, P. Sullivan, and B. J. Van Doormaal. 2008. Relationship Between Reproduction Traits and Functional Longevity in Canadian Dairy Cattle. J. Dairy Sci. 91:1660–1668. https://doi.org/10.3168/jds.2007-0178.

Svensson, C., and J. Hultgren. 2008. Associations between housing, management, and morbidity during rearing and subsequent first-lactation milk production of dairy cows in southwest Sweden. J. Dairy Sci. 91:1510–1518.

van Pelt, M., G. De Jong, and R. Veerkamp. 2016a. Changes in the genetic level and the effects of age at first calving and milk production on survival during the first lactation over the last 25 years. Animal 10:2043–2050.

van Pelt, M., V. Ducrocq, G. De Jong, M. Calus, and R. Veerkamp. 2016b. Genetic changes of survival traits over the past 25 yr in Dutch dairy cattle. J. Dairy Sci. 99:9810–9819.

van Pelt, M., T. Meuwissen, G. de Jong, and R. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117–4130.

Wathes, D., J. Brickell, N. Bourne, A. Swali, and Z. Cheng. 2008. Factors influencing heifer survival and fertility on commercial dairy farms. Animal 2:1135–1143.

Weigel, K. A., P. Hoffman, W. Herring, and T. Lawlor. 2012. Potential gains in lifetime net merit from genomic testing of cows, heifers, and calves on commercial dairy farms. J. Dairy Sci. 95:2215–2225.

Wells, S., A. Trent, W. Marsh, P. McGovern, and R. Robinson. 1993. Individual cow risk factors for clinical lameness in lactating dairy cows. Prev. Vet. Med. 17:95–109.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behaviour? J. Anim. Ecol. 75:1182–1189.

Yoo, W., R. Mayberry, S. Bae, K. Singh, Q. P. He, and J. W. Lillard Jr.. 2014. A study of effects of multicollinearity in the multivariable analysis. Int. J. Appl. Sci. Technol. 4:9.

Zavadilová, L., and V. Zink. 2013. Genetic relationship of functional longevity with female fertility and milk production traits in Czech Holsteins. Czech J. Anim. Sci. 58:554–565.

Zijlstra, J., M. Boer, J. Buiting, K. Colombijn-Van der Wende, and E.-A. Andringa. 2013. Rapport 668: Routekaart Levensduur; Eindrapportage van het project "Verlenging levensduur melkvee". Wageningen UR Livestock Research, Wageningen, the Netherlands.

ORCIDS

E. M. M. van der Heide ⦿ https://orcid.org/0000-0002-8781-7560

R. F. Veerkamp ⦿ https://orcid.org/0000-0002-5240-6534

M. L. van Pelt ⦿ https://orcid.org/0000-0003-1315-1329

C. Kamphuis ⦿ https://orcid.org/0000-0001-5552-036X

B. J. Ducro ⦿ https://orcid.org/0000-0002-0866-0199

APPENDIX

Table A1. All 62 phenotypic variables and 50 genomic estimated breeding values (gEBV) available and the decision moment in which each variable is available

Item	Decision moment	Continuous
Phenotypic variable		
Animal identification number	Birth	No
Year of birth	Birth	No
Birth farm identification number (UBN)	Birth	No
Month of birth	Birth	No
Birth season	Birth	No
Parity dam	Birth	Yes
Breed	Birth	No
Holstein %	Birth	No
Red factor	Birth	No
Calving ease dam	Birth	No
Gestation duration dam	Birth	Yes
Birth weight	Birth	Yes
Insemination farm	18 mo	No
Insemination season	18 mo	No
Countable inseminations	18 mo	Yes
Nonreturn status at 18 mo	18 mo	No
No insemination information at 18 mo	18 mo	No
Number of farm movements at 18 mo	18 mo	Yes
Age at first insemination	18 mo	Yes
Type of first insemination	18 mo	No
Number of inseminations	18 mo	Yes
Raised at a specialty calf-rearing farm	First calving	No
Calving season	First calving	No
Total number of farm movements at calving	First calving	Yes
Age at first calving	First calving	Yes
Calving farm UBN	First calving	No
Calf sex	First calving	No
Calf survival first 24 h	First calving	No
Calving ease	First calving	No
Gestation duration	First calving	Yes
Birthweight calf	First calving	Yes
Calf survival first week	First calving	No
Calf survival second week	First calving	No
Twins	First calving	No
kg of milk at 6 wk	6 wk after first calving	Yes
Fat percentage milk at 6 wk	6 wk after first calving	Yes
Protein percentage milk at 6 wk	6 wk after first calving	Yes
Cell count milk at 6 wk	6 wk after first calving	Yes
Urea milk at 6 wk	6 wk after first calving	Yes
Lactose percentage milk at 6 wk	6 wk after first calving	Yes
Cow status indicator at 6 wk	6 wk after first calving	No
Number of negative indications at 6 wk	6 wk after first calving	Yes
Number of days in lactation on milk test day	6 wk after first calving	Yes
Complete milk measurement available at 6 wk	6 wk after first calving	No
First parity insemination farm UBN	200 d after first calving	No
First parity insemination season	200 d after first calving	No
First parity first insemination type	200 d after first calving	No
Number of inseminations in first parity	200 d after first calving	Yes
Nonreturn status at 200 d postcalving	200 d after first calving	No
Age at 200 d postcalving	200 d after first calving	Yes
Insemination known in the first parity	200 d after first calving	No
Age at first insemination in the first parity	200 d after first calving	Yes
Number of farm movements at 200 d postcalving	200 d after first calving	Yes
Number of known milk testings 200 d postcalving	200 d after first calving	Yes
Average kg of milk	200 d after first calving	Yes
Average fat percentage of milk	200 d after first calving	Yes
Average protein percentage of milk	200 d after first calving	Yes
Average cell count of milk	200 d after first calving	Yes
Average urea of milk	200 d after first calving	Yes
Average lactose percentage of milk	200 d after first calving	Yes

Continued

Table A1 (Continued). All 62 phenotypic variables and 50 genomic estimated breeding values (gEBV) available and the decision moment in which each variable is available

Item	Decision moment	Continuous
Number of negative indications at 200 d postcalving (mastitis, abortion, other illness, teat disorders)	200 d after first calving	Yes
Number of farm movements in the first parity	200 d after first calving	Yes
gEBV		
NVI[1] Dutch breeding goal standard	Birth	Yes
kg of milk	Birth	Yes
kg of fat	Birth	Yes
kg of protein	Birth	Yes
kg of lactose	Birth	Yes
"Inet" Dutch production index	Birth	Yes
Cell count	Birth	Yes
Subclinical mastitis	Birth	Yes
Clinical mastitis	Birth	Yes
Udder health	Birth	Yes
Lifespan	Birth	Yes
Lifespan with predictors	Birth	Yes
Birth index	Birth	Yes
Calving ease	Birth	Yes
Postcalving ease	Birth	Yes
Livability calving (maternal)	Birth	Yes
Livability birth (direct)	Birth	Yes
Overall fertility	Birth	Yes
Nonreturn status at 56 d	Birth	Yes
Interval calving–first insemination	Birth	Yes
Calving interval	Birth	Yes
Interval first–last insemination	Birth	Yes
Conception ratio	Birth	Yes
Claw health	Birth	Yes
Calf vitality 3–365 d	Birth	Yes
Milking speed	Birth	Yes
Dairy strength	Birth	Yes
Stature	Birth	Yes
Chest width	Birth	Yes
Body depth	Birth	Yes
Angularity	Birth	Yes
Body condition	Birth	Yes
Rump angle	Birth	Yes
Rump width	Birth	Yes
Rear legs hind view	Birth	Yes
Rear leg side view	Birth	Yes
Foot angle	Birth	Yes
Locomotion	Birth	Yes
Fore udder attachment	Birth	Yes
Front teat placement	Birth	Yes
Teat length	Birth	Yes
Udder depth	Birth	Yes
Rear udder height	Birth	Yes
Udder support	Birth	Yes
Rear teat placement	Birth	Yes
Frame	Birth	Yes
Robustness	Birth	Yes
Overall udder score	Birth	Yes
Feet and legs	Birth	Yes
Overall exterior score	Birth	Yes
Milking robot efficiency	Birth	Yes

[1]NVI = the Dutch/Flemish merit index score.

Table A2. Variables selected by stepwise selection for the decision moment "birth"[1]

Item	Coefficient
gEBV and phenotypic variable selected	
gEBV longevity	0.128
gEBV foot angle	−0.115
gEBV udder depth	−0.102
gEBV frame	0.074
gEBV interval first-last insemination	0.140
gEBV nonreturn at 56 d	−0.092
gEBV kg of fat	0.061
Year of birth (referent = 2012)	
2013	−0.372
Birth season (referent = fall)	
Spring	0.338
Summer	0.286
Winter	0.434

[1]Estimated coefficients are shown for the genomic estimated breeding values (gEBV) and for the individual classes of each phenotypic variable with the referent class in parentheses.

Table A3. Variables selected by stepwise selection for the decision moment "18 mo of age"[1]

Item	Coefficient
gEBV and phenotypic variable selected	
gEBV overall fertility	0.145
gEBV kg of fat	0.122
gEBV longevity	0.212
gEBV milking speed	0.061
gEBV foot angle	−0.078
gEBV udder depth	−0.137
NVI[2]	−0.122
gEBV overall exterior score	0.143
Birth season (referent = fall)	
Spring	0.411
Summer	0.474
Winter	0.410
Year of birth (referent = 2012)	
2013	−0.268
Nonreturn status at 18 mo (referent = nonreturn)	
Unknown	−0.426
Number of inseminations at 18 mo (referent = 0)	
1	0.206
2	0.317
3	0.096
4	−0.346
5+	−0.988
Unknown	−0.130

[1]Estimated coefficients are shown for the genomic estimated breeding values (gEBV) and for the individual classes of each phenotypic variable with the referent class in parentheses.
[2]NVI = the Dutch/Flemish merit index score.

Table A4. Variables selected by stepwise selection for the decision moment "first calving"[1]

Item	Coefficient
gEBV and phenotypic variable selected	
gEBV longevity	0.261
gEBV udder depth	−0.140
gEBV overall fertility	0.091
gEBV foot angle	−0.086
gEBV SCC	0.196
gEBV udder health	−0.283
gEBV rear udder height	0.072
gEBV feet and legs	−0.067
gEBV chest width	0.099
Year of birth (referent = 2012)	
2013	−0.270
Birth season (referent = fall)	
Spring	0.423
Summer	0.341
Winter	0.121
Age at first calving in days (referent = >1,000)	
<650	1.531
650–700	1.700
700–750	1.549
750–800	1.417
800–850	1.307
850–900	1.053
900–950	1.005
950–1,000	1.110
Sex of calf (referent = male)	
Female	0.206
Unknown	−0.172
Coat color (referent = red)	
Black	−0.208

[1]Estimated coefficients are shown for the genomic estimated breeding values (gEBV) and for the individual classes of each phenotypic variable with the referent class in parentheses.

Table A5. Variables selected by stepwise selection for the decision moment "6 wk after first calving"[1]

Item	Coefficient
gEBV and phenotypic variable selected	
gEBV longevity	0.242
gEBV overall fertility	0.242
gEBV rear legs hind view	−0.162
gEBV locomotion	0.124
gEBV chest width	0.131
gEBV foot angle	−0.085
gEBV SCC	0.212
gEBV udder health	−0.289
gEBV interval first–last insemination	−0.213
gEBV kg of lactose	−0.073
gEBV nonreturn at 56 d	−0.088
Year of birth (referent = 2012)	
2013	−0.342
Birth season (referent = fall)	
Spring	0.459
Summer	0.600
Winter	0.423
Age at first calving in days (referent = >1,000)	
<650	1.520
650–700	1.881
700–750	1.741
750–800	1.660
800–850	1.456
850–900	1.309
900–950	0.849
950–1,000	1.097
Sex of calf (referent = male)	
Female	0.156
Unknown	−0.116
kg of milk produced at milk test day closest to 6 wk postcalving (referent = <15)	
15–20	0.354
20–25	0.825
25–30	1.359
30–35	1.521
35–40	1.688
40+	2.512
Unknown	0.049

[1]Estimated coefficients are shown for the genomic estimated breeding values (gEBV) and for the individual classes of each phenotypic variable with the referent class in parentheses.

Table A6. Variables selected by stepwise selection for the decision moment "200 d after first calving"[1]

Item	Coefficient
gEBV and phenotypic variable selected	
gEBV longevity	0.207
gEBV udder health	−0.153
gEBV interval first–last insemination	0.106
gEBV leg and feet	−0.247
gEBV overall exterior score	0.256
gEBV stature	−0.094
gEBV udder support	0.074
gEBV udder depth	−0.106
Year of birth (referent = 2012)	
2013	−0.239
Age at first calving in days (referent = >1,000)	
<650	2.831
650–700	1.844
700–750	1.792
750–800	1.958
800–850	1.763
850–900	1.554
900–950	0.835
950–1,000	1.385
Calving season (referent = fall)	
Spring	0.510
Summer	0.725
Winter	0.331
Calf survival at 1 wk of age (referent = alive)	
Died within 1 wk	14.250
Dead 24 h after calving	−0.142
Unknown	0.165
Number of farm moves at 200 d	
1	−0.998
2	13.527
3	12.102
4	14.217
Average kg of milk per test milk day at 200 d postcalving (referent = <20)	
20–25	0.335
25–30	0.884
30–35	1.311
35–40	1.060
40+	1.339
Unknown	−0.560
Average cell count (× 1,000) per test milk day at 200 d postcalving (referent = <25)	
25–50	−0.145
50–75	−0.348
75–100	0.448
100–125	−0.437
125+	−0.584
Number of inseminations at 200 d postcalving (referent = 0)	
1	0.598
2	0.132
3	0.094
4	−0.308
5+	−0.091
Unknown	−0.955
Nonreturn status at 200 d postcalving (referent = nonreturn)	
Unknown	−1.065
Average percentage of protein per test milk day at 200 d postcalving (referent = <3.0)	
3.00–3.25	0.456
3.25–3.50	0.784
3.50–3.75	0.705
3.75–4.00	1.414
4.00+	0.953
Negative indication at test milking 6 wk postcalving (referent = no)	
Yes	1.401
Unknown	0.946
Negative indication count	−0.788

Continued

Table A6 (Continued). Variables selected by stepwise selection for the decision moment "200 d after first calving"[1]

Item	Coefficient
Number of inseminations at 18 mo (referent = 0)	
1	0.633
2	0.523
3	0.270
4	0.027
5+	−0.581
Unknown	0.783

[1]Estimated coefficients are shown for the genomic estimated breeding values (gEBV) and for the individual classes of each phenotypic variable with the referent class in parentheses.

Table A7. Variables and estimated coefficients selected by bi-directional stepwise selection for the decision moment at birth in both a multiple and single variable regression

Item	Multiple regression coefficient	Single regression coefficient
gEBV[1] and phenotypic variable selected		
gEBV longevity	0.159	0.153
gEBV foot angle	−0.109	−0.120
gEBV udder depth	−0.068	−0.060
gEBV frame	0.062	0.017
gEBV kg of fat	0.063	0.087
gEBV nonreturn at 56 d	−0.086	0.009
gEBV dairy strength	0.057	0.043
gEBV interval first–last insemination	0.218	0.085
gEBV calving interval	−0.106	0.048
Season of birth (referent = fall)		
Spring	0.304	0.118
Summer	0.231	0.205
Winter	0.334	0.119
Year of birth (referent = 2012)		
2013	−0.360	−0.220

[1]gEBV = genomic estimated breeding value.