

BULLETIN NO. 62

Proceedings of the 2026 Interbull Meeting

31 May – 1 June 2026, Verona, Italy

INTERNATIONAL BULL EVALUATION SERVICE

A permanent subcommittee of the International Committee for Animal Recording

Office: Interbull Centre, Department of Animal Biosciences,
Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden

Variance Component Re-Estimation and Model Simplification in a Single-Step Random Regression Test-Day Model for Nordic Red Dairy Cattle

T. Pitkänen[1], M.H. Lidauer[1], M. Taskinen[1], J. Pösö[2], A. Kudinov[3], K. Byskov[4], H. Splittorff[5], M. Koivula[1]

[1]Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland

[2] Faba co-op, P.O. BOX 40, FIN-01301 Vantaa, Finland

[3]Växa, Box 288, 75105 Uppsala, Sweden

[4] SEGES Innovation P/S, Agro Food Park 15, 8200 Aarhus N, Denmark

[5] NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus N, Denmark

Corresponding author: e-mail address: timo.j.pitkanen@luke.fi

Abstract

The Nordic Red dairy cattle random regression test-day model (RRM) has been updated to reflect increased genetic similarity among the original country subpopulations (Finland, Sweden and Denmark) and changes in data structure over the past two decades. Data prior to 2005 were excluded, and previously country-specific models were harmonized into a simplified nine-trait multi-trait single-step RRM using a metafounder approach. The objective of this study was to describe the implementation of the updated model. This included re-estimation of variance components for parameter-reduced covariance functions for random effects identified as optimal in a previous study, accounting for heterogeneous measurement error (co)variances across eight milk recording schemes, and evaluation of genomic estimated breeding value reliability using forward cross-validation. Variance components were estimated from a harmonized subset of data sampled across Finland, Sweden and Denmark, applying the updated model. Compared with previously used variance components derived from Finnish-only data, heritability estimates for 305-day yields increased by 8 to 46% across all traits. Substantial differences in measurement error variances and residual correlations among daily milk, protein, and fat yields were observed across the eight milk recording schemes, reflecting heterogeneous milking systems and historical protocol changes in automated milking systems in Finland. For Finland, four residual covariance matrices were required to model measurements from conventional milking systems and three automated milking system protocols. In contrast, two residual classes (automated and conventional milking systems) were sufficient for Denmark and Sweden. The selected covariance functions enabled a substantial reduction in variance component parameters while maintaining model stability and predictive ability. The model was solved using a fully component-wise ssGTBLUP approach with metafounders. The updated model provides a computationally efficient basis for routine genetic evaluation of Nordic Red dairy cattle.

Key words: Nordic Red dairy cattle, test-day model, yield evaluation, variance components

Introduction

In 2006, the Nordic countries, Finland (FIN), Denmark (DNK), and Sweden (SWE), introduced across-country genetic yield evaluations for their dairy cattle populations (Lidauer et al. 2015).

Revising the random regression test-day model (RRM) for Red dairy cattle (RDC) evaluation was considered necessary as the genetic similarity among the original country subpopulations has increased and changes in data structure over the past two decades have made the original model outdated.

The objective of this study was to describe the implementation of the updated Nordic RDC yield evaluation model. This included re-estimation of variance components, deriving parameter-reduced covariance functions for random effects (Pitkänen et al. 2026), accounting for heterogeneous measurement error (co)variances across eight milk recording schemes, updating pedigree and including metafounders. Finally, the impact of the new variance components on genetic parameters are validated.

Materials and Methods

Data

The model update started with the data and pedigree modifications (Table 1). We decided to truncate the test-day (TD) data from year 2005 as older data had much more variation in the countries. From Finnish data, also Holstein records were removed, as well as later lactation (3+) records. This simplifies the model. Including Holstein and later lactation data were important in the past to increase contemporary group sizes for Finnish herds. As herd sizes have increased, this is not needed anymore.

Automated milking system recording protocols have changed historically in Finland but remained stable in Denmark and Sweden (Pitkänen et al. 2024). In Finland, three AMS recording protocols were used: 1) the daily average of a month's milkings, 2) a 24-h method based on two successive milkings (Peeters and Galesloot, 2002), and 3) a 96-h method (Lazenby et al., 2002). The first protocol was used until 2015, and the second from January 2015 to June 2022. However, from January 2017 onwards, AMS herds gradually moved to the 96-h method which is the current method for AMS records. Information on the milking system used for test-day recording was corrected for a small proportion of records because of database issues in Finland and Sweden.

Finally, unknown parent groups were redefined to support the metafounder approach. The revised grouping structure was designed to better reflect the historical relationships among the Nordic RDC subpopulations and provide a more realistic representation of the genetic base population.

Model

The new updated multiple-trait RRM describes TD milk, protein, and fat yields for the first three lactations using nine model traits instead of 27 in the previous model. Each trait has a random regression function for additive genetic and permanent environmental effects. The model includes fixed effects nested within countries, heterosis and recombination loss adjustments, and correction for heterogeneous variance (HV). Due to rank reduction, 16 cow-specific coefficients define all nine breeding value curves (Pitkänen et al. 2026).

First steps in the model revision were to recalculate heterosis and recombination effects (Pitkänen et al. 2025) and correct the calving age effects and new herd-year season effects. Because of moving from 27 country specific traits to nine traits, country interactions were included in some fixed effects.

Table 1: Main changes implemented in the updated Nordic RDC yield evaluation model.

Area	Main update
Data	TD data truncated from 2005, Finnish HOL and lactations 3+ removed
Pedigree	Metafounder approach implemented
Traits	27 traits reduced to 9 traits
Fixed effects	Updated heterosis, recombination, calving age and herd-year-season effects
Recording schemes	AMS/CMS classes revised
Variance components	Nordic RDC specific variance components and parameter-reduced covariance functions
Evaluation	ssGTBLUP with metafounders

Revising the variance components

Variance component estimation was based on a representative sample of Nordic Red dairy cattle from Finland, Denmark, and Sweden,

including TD records from both conventional milking systems (CMS) and automated milking systems (AMS).

Animals were selected according to strict criteria. All records had to originate from a single herd, and complete lactation sequences were required. Cows could contribute data from only the first lactation, first and second lactations, or the first three lactations, with no missing lactations before the final observed lactation. Herds were required to have at least eight but fewer than 100 suitable first-lactation cows per year on average and at least 12 years of data beginning in 2010.

The final sample size was approximately 100,000 animals. Herds were sampled independently within each country × milking-system class, with allocations proportional to the number of first-lactation cows in each class. Variance components were estimated using a pedigree-based random regression TD model implemented in MiX99 (Gao et al., 2026). The statistical model for VCE was defined as:

$$y = \mathbf{X}b + \mathbf{Z_h}h + \mathbf{Zc}c + \mathbf{Z_p}p + \mathbf{Z_a}a + e,$$

The model included fixed effects of heterosis, recombination, herd-year-season, year-month × milking system × country, days carried calf, and a linear herd effect. The random effects included herd-test-day (HTD), herd-lactation curve (HC), permanent environmental (PE), additive genetic (A), and residual effects. The HTD effect was fitted as an intercept effect. The HC effect was modeled using a second-order Legendre polynomial without intercept but combined with a Wilmink function exponential term, with common regression coefficients for the second and third lactations. Both additive genetic and permanent environmental effects were modeled using third-order Legendre polynomials combined with a Wilmink function exponential term. Residual variances were assumed to be heterogeneous across 12 lactation intervals.

To account for differences in recording protocols among countries and milking systems, residual covariance matrices were estimated separately for each country × milking-system class. In total, eight residual classes were defined, including one CMS and three AMS classes in Finland corresponding to different recording protocols, and one AMS and one CMS class in both Denmark and Sweden. This structure allowed heterogeneous measurement error variances and residual correlations to be modelled consistently across recording environments.

Parameter reduction of covariance functions

The estimated variance components were subsequently reduced for implementation in routine single-step genetic evaluations. Covariance functions were fitted to the variance components. The genetic covariance function was reduced from a full rank of 36 to rank 16 while retaining approximately 99% of the original genetic variation. Similarly, the permanent environmental covariance function was reduced from rank 36 to rank 21. The permanent environmental covariance function was designed to absorb variation associated with permanent environmental effects, herd-lactation curve effects, and within-lactation residual variation. For more details, see Pitkänen et al. (2026). The parameter-reduced covariance functions were used in the updated Nordic RDC evaluation model.

Validation

The impact of the updated variance components was assessed with TD data from May 2026. The data included approximately 34.7 million records from 1.78 million animals, and a pedigree with 3.1 million animals. Genotype data for RDC included 315 448 animals. The single-step RRM was solved by preconditioned conjugate gradient iteration on data (Strandén and Lidauer, 1999) implemented in MiX99. The model was solved using a fully component-wise formulation of ssGTBLUP (Mäntysaari et al., 2017) combined with metafounders. Metafounders were used to account for genetic

differences among base populations and to model relationships among founder groups.

Validation was based on a four-year data truncation. The linear regression (LR) method (Legarra and Reverter, 2018) was used for validation. Validation statistics included prediction bias, regression coefficients (b_1), and coefficients of determination (R^2) for milk, protein, and fat yield traits.

Validation bulls were required to have at least 20 daughters in the complete dataset and no daughters in the reduced dataset. For cows, the validation set included animals with at least five TD records in the full data and no records in the reduced data.

Results

The updated variance components resulted in clearly higher heritability estimates compared with the previous model (Table 2). For milk yield, heritability increased from 0.39 to 0.44 in first lactation, from 0.28 to 0.35 in second lactation, and from 0.25 to 0.31 in third lactation. Similar increases were observed for protein and fat yield traits.

The largest relative increase was observed for third-lactation fat yield, where heritability increased by approximately 46%. For the other traits, increases ranged from 12% to 26%.

Table 2: Old and new 305d heritability estimates for milk, protein and fat in lactation 1-3, and change in heritability as %.

Trait	New	Old	Change %
Lactation1			
Milk	0.44	0.39	13
Protein	0.37	0.33	12
Fat	0.38	0.34	12
Lactation 2			
Milk	0.35	0.28	26
Protein	0.32	0.27	19
Fat	0.30	0.25	20
Lactation 3			
Milk	0.31	0.25	24
Protein	0.29	0.24	21
Fat	0.29	0.20	46

Substantial differences in residual variances were observed among milking-system classes (Figure 1). In Finland, the residual variance for milk yield differed markedly between CMS and the three AMS recording protocols. The highest residual variances were generally observed for the 24-h AMS protocol, whereas the monthly average showed reduced residual variation. In Denmark and Sweden, residual variance profiles were more stable because recording protocols remained unchanged throughout the study period.

These results indicate that a single residual covariance structure is insufficient for describing the heterogeneous recording environments present in the Nordic RDC population. Accounting for country-specific AMS and CMS recording schemes improved the biological interpretation of residual variation and provided a more realistic description of measurement error in the updated evaluation model.

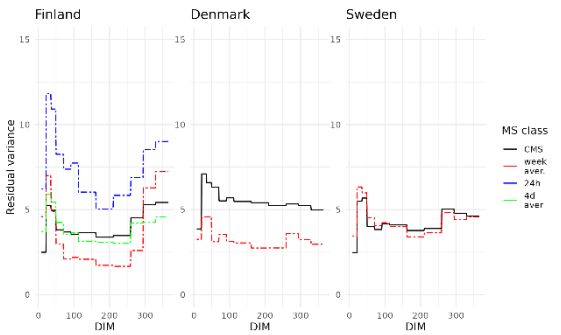


Figure 1. Residual variances for first-lactation milk yield across country × milking-system classes. Finnish AMS observations were divided according to recording protocol (week average, 24-h equivalent and 96-h average).

Validation results showed comparable performance between the old and new variance components (Table 3). The old variance components resulted in slightly better regression coefficients for all traits, differences were small. Also, coefficients of determination were nearly same for the old and the new variance components, and overall, predictive ability remained at a similar level for both models.

The updated variance components resulted in slightly higher genetic trends and larger genetic standard deviations for both bulls and cows (Figure 2). Consequently, greater differentiation among animals was observed, leading to some re-ranking of breeding candidates compared with evaluations based on the previous variance components.

Table 3: Linear regression (LR) validation in bulls and cows with new (NewVCE) or old (OldVCE) variance components. The values in the table are: bias(kg) = mean ($GEBV_{Full} - GEBV_{Reduced}$) ±SD, b_1 regression coefficient and R^2 coefficient of determination for milk, protein and fat, respectively.

	Model	bias (kg)	b_1	R^2
Bulls	OldVCE	-267.21 ±377.8	1.06	0.75
	NewVCE	-296.67 ±397.7	1.02	0.73
	OldVCE	-9.48 ±13.4	0.99	0.64
	NewVCE	-11.00 ±13.9	0.96	0.63
	OldVCE	-11.12 ±15.8	0.97	0.67
	NewVCE	-13.60 ±16.8	0.97	0.68
	Model	**bias(kg)**	**b_1**	**R^2**
Cows	OldVCE	-120.82 ±279.1	1.03	0.83
	NewVCE	-147.94 ±303.6	1.01	0.81
	OldVCE	-3.75 ±9.8	0.98	0.78
	NewVCE	-4.92 ±10.1	0.96	0.77
	OldVCE	-3.84 ±11.8	0.97	0.79
	NewVCE	-5.69 ±12.5	0.97	0.80

Discussion and conclusions

The updated Nordic RDC evaluation model includes several important improvements in data structure, pedigree handling, trait definitions, and variance component estimation. The use of newly estimated Nordic-specific variance components provides a more realistic representation of the current Nordic RDC population and production environment. In addition to updated variance components, the implementation of metafounders provides a more appropriate framework for modelling the increasingly connected Nordic RDC population.

Higher heritability estimates suggest that the previous model underestimated the proportion of genetic variation, particularly in later lactations. The estimation of separate residual covariance matrices for different country × milking-system classes revealed substantial heterogeneity in measurement error structures. This was particularly evident in Finland, where changes in AMS recording protocols over time resulted in distinct residual variance patterns. These findings support the use of multiple residual covariance structures in Nordic evaluations and suggest that part of the residual variation previously attributed to unexplained environmental effects was related to differences in recording procedures.

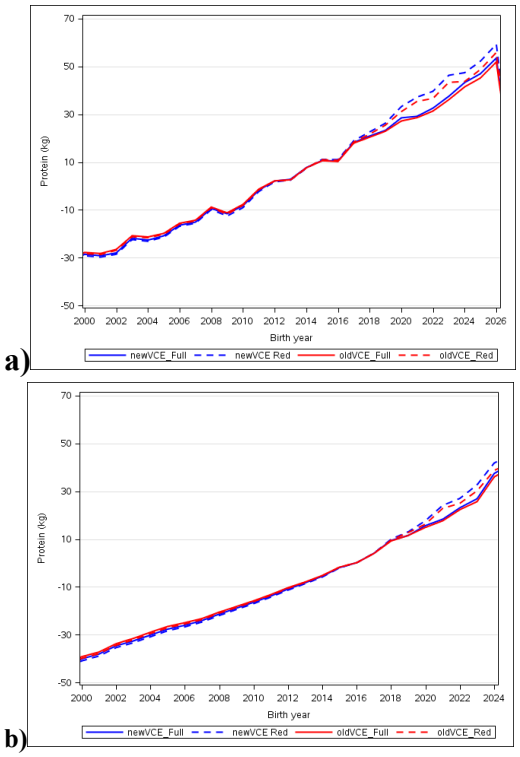


Figure 2. Genetic trends for protein a) bulls and b) cows by birth year. The model ssGTBLUP with MF using old (OldVCE) or new variance components (NewVCE).

Parameter reduction successfully simplified the covariance functions while preserving nearly all biologically relevant variation. This reduction is important for maintaining computational efficiency in routine single-step evaluations involving large datasets and genomic information.

Acknowledgements

This work was a part of the Nordic project "Modelling of phenotypes and ancestral relationships in genomic prediction for yield traits in Nordic Red dairy cattle". The Nordic breeding organizations (Viking Genetics, Randers, Denmark; Nordic Cattle Genetic Evaluation (NAV), Aarhus, Denmark), are acknowledged for providing the genotype data and the TD data, and for cofounding the project.

References

Gao, H., Taskinen, M., Pitkänen, T.J., Ylitalo, A.-K., Vargas Jurado, N., Lidauer, M.H., Strandén, I. 2026. Estimation of variance components in single-step models using MiX99. Proc. 13th WCGALP, Madison, Wisconsin, USA, July 12-17, 2026. doi: 10.31274/wcgalp.23537.

Lazenby, D., Bohlsen, E., Hand, K.J., Kelton, D.F., Miglior, F. and Lissemore, K.D. 2002. Methods to estimate 24-hour yields for milk, fat and protein in robotic milking herds. In: *Performance Recording of Animals: State of the Art 2002*. EAAP/ICAR Technical Series, pp. 65-71. doi: 10.3920/9789086865116_010

Legarra, A., Reverter, A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol. 50*, 53. doi: 10.1186/s12711-018-0426-6.

Lidauer, M., Pösö, J., Pederson, J., Lassen, J., Madsen, P., Mäntysaari, E.A., Nielsen, U., Eriksson, J-Å., Johansson, K., Pitkänen, T., Strandén, I., Aamand, G.P. 2015. Across-country test-day model evaluations for Nordic Holstein, Red Cattle and Jersey. *J. Dairy Sci. 98*, 1296–1309. doi: 10.3168/jds.2014-8307.

Mäntysaari, E.A., Evans, R.D., Strandén, I. 2017. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci. 95*, 4728-4737. doi: 10.2527/jas2017.1912.

Peeters R, Galesloot P. 2002. Estimating Daily Fat Yield from a Single Milking on Test Day for Herds with a Robotic Milking System. *J. Anim. Sci.,* 85, 682-688. doi: 10.3168/jds.S0022-0302(02)74124-6.

Pitkänen, T, Lidauer, M, Pösö,J, Fikse, W.F., Nielsen, U.S. Aamand, G.P., Koivula, M. 2024. Revision of random regression test-day model has improved genomic prediction for Nordic Red dairy cattle. *Interbull Bulletin,* 60, 20-21. https://journal.interbull.org/index.php/ib/article/view/1932.

Pitkänen, Koivula, M, Taskinen, M, Pösö, J, Nielsen, U.S, Kudinov, A, Lidauer, M.H. 2025. Dynamic breed definition for admixture Nordic Red dairy cattle – An ad hoc approach for modelling heterosis in genomic evaluation for production traits. In: *Book of abstracts of the 67th annual meeting of EAAP*, Innsbruck, Austria. 25- 29 Aug. p. 584.

Pitkänen, T, Taskinen, M, Koivula, M, Pösö, J, Kudinov, A, Nielsen, U.S., Byskov, K, Lidauer, M.H. 2026. Deriving covariance functions (CF) for Nordic Red dairy cattle yield evaluation model. *Proc. 13th WCGALP*, Madison, Wisconsin, USA, July 12-17, 2026. doi: 10.31274/wcgalp.23946.

Strandén, I., Lidauer, M 1999. Solving large mixed models using preconditioned conjugate gradient iteration. *J. Dairy Sci*. 82, 2779-2787. doi: 10.3168/jds.S0022-0302(99)75535-9.

Augmented Monte Carlo AI-REML for (co)variance component estimation in single-step models

H. Gao[1] and I. Strandén[1]

[1] Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland

Corresponding author: hongding.gao@luke.fi

Abstract

Accurate (co)variance component estimation (VCE) is essential for reliable genetic evaluation. Analytical restricted maximum likelihood (REML) estimation in single-step genomic models is often computationally demanding because mixed model equations (MME) have dense genomic relationship blocks. In this study, we extended the augmented AI-REML to an augmented Monte Carlo (MC) AI-REML approach, in which MC sampling was used to approximate the trace terms required for REML, thereby avoiding explicit inversion of the coefficient matrix of MME. The method was implemented for three single-step models: ssGBLUP, ssGTABLUP, and ssSNPBLUP. Data were simulated for 44,280 individuals over 20 generations, with 9,000 markers and three scenarios comprising 10,000, 20,000, or 30,000 genotyped individuals. Variance component estimates were consistent across single-step models and genotyping scenarios. The augmented MC AI-REML converged within 11-18 REML iterations. Computing time increased with the number of genotyped individuals for all models. Based on the current datasets, ssGBLUP was the fastest in all scenarios but showed a substantial increase in memory requirements as the number of genotyped individuals increased. ssSNPBLUP required the longest computing time. ssGTABLUP showed intermediate computing times with substantially lower memory requirements than ssGBLUP. Overall, augmented MC AI-REML provides an efficient approach for VCE in single-step models and has potential for application to large-scale genomic datasets.

Key words: Genetic parameters, ssGBLUP, ssGTABLUP, ssSNPBLUP, genomic prediction

Introduction

Genomic prediction and genetic evaluation require accurate (co)variance component estimation (VCE). Restricted maximum likelihood (REML) has long been the preferred method for performing VCE (Misztal et al., 2020). Compared with Bayesian approaches based on Gibbs sampling, REML avoids the need for Markov chain Monte Carlo (MCMC) sampling and its associated practical complications, making it an attractive and computationally efficient approach for routine VCE in genetic evaluation.

Analytical REML methods generally require factorization or inversion of the coefficient matrix of the mixed model equations (MME) to form the trace terms (Meyer, 1991; Gilmour et al., 1995; Johnson and Thompson, 1995; Jensen et al., 1997). In pedigree-based evaluations, this can be performed efficiently because the sparse inverse of the numerator relationship matrix results in sparse MME, enabling efficient computation even for large datasets. However, in single-step models, analytical REML can become computationally prohibitive because MME have dense genomic relationship blocks.

Monte Carlo (MC) REML methods, which approximate the trace terms using random samples generated from the same distribution as the observed data and the current VC estimates, can overcome the computational limitation (Matilainen et al., 2012, 2013). Recently, Bermann et al. (2025) extended MC REML to single-step genomic BLUP (ssGBLUP) by developing MC-ss-GREML. They showed that

the resulting VC estimates were in agreement with those from exact ss-GREML, while requiring only 14% of the computing time and 1% of the memory compared to the exact ss-GREML.

In this study, the aims were: (1) to extend the augmented AI-REML (Stranden et al., 2024) to augmented MC AI-REML for VCE under three single-step models (ssGBLUP, ssGTABLUP, and ssSNPBLUP); and (2) to compare the computational performance of these implementations.

Materials and Methods

Augmented MC AI-REML

In the augmented MC AI-REML, the update vector ($\boldsymbol{\Delta}$) can be obtained by solving the following augmented MME (Stranden et al., 2024):

$$\begin{bmatrix} \mathbf{C} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{F} \\ \mathbf{F}'\mathbf{R}^{-1}\mathbf{W} & \mathbf{F}'\mathbf{R}^{-1}\mathbf{F} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{s}}^{*} \\ \boldsymbol{\Delta} \end{bmatrix} = \begin{bmatrix} \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{F}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{t} \end{bmatrix} \quad (1)$$

where $\mathbf{C}$ is the original coefficient matrix, $\mathbf{W} = [\mathbf{X} \quad \mathbf{Z}]$ is the design matrix, $\mathbf{F}$ is the working matrix, $\mathbf{R} = \mathbf{R}_0 \otimes \mathbf{I}$, $\mathbf{R}_0$ is the residual variance covariance matrix, $\mathbf{I}$ is an identity matrix, $\otimes$ denotes the Kronecker product, $\mathbf{y}$ is the vector of observations, $\widehat{\mathbf{s}}^{*}$ is the new solution vector of fixed and random effects, and $\mathbf{t}$ is the trace vector, containing the trace contribution for each variance and covariance component in the REML score vector (i.e., the first derivatives of the REML log-likelihood).

Monte Carlo sampling was used to approximate the trace vector $\mathbf{t}$ in (1), thereby avoiding explicit inversion of the coefficient matrix. The approach follows the strategy of García-Cortés et al. (1995) and its applications (Matilainen et al., 2012, 2013). At each REML iteration, data were simulated using the current estimates of the (co)variance matrices to form the simulated phenotypes. Then simulated phenotypes were subsequently solved using the same MME as the observed data.

For the genetic part in the trace vector:

$$\mathbf{t}_{\mathrm{G}} = \mathbf{G}_0^{-1}(q\mathbf{G}_0 - \mathbf{PEV}_{\mathrm{G}})\mathbf{G}_0^{-1} \quad (2)$$

where $\mathbf{G}_0$ is the current genetic (co)variance matrix, q is the number of levels in the genetic effects, and the vector of prediction error variance (PEV) can be computed as

$$\mathbf{PEV}_{\mathrm{G}} = q\mathbf{G}_0 - \frac{1}{s}\sum_{h=1}^{s} \widehat{\tilde{\mathbf{u}}}^{h\prime}\mathbf{K}^{-1}\widehat{\tilde{\mathbf{u}}}^{h} \quad (3)$$

or

$$\mathbf{PEV}_{\mathrm{G}} = \frac{1}{s}\sum_{h=1}^{s}\left(\tilde{\mathbf{u}}^{h} - \widehat{\tilde{\mathbf{u}}}^{h}\right)'\mathbf{K}^{-1}\left(\tilde{\mathbf{u}}^{h} - \widehat{\tilde{\mathbf{u}}}^{h}\right) \quad (4)$$

where s is the number of MC samples, each $h = 1, \ldots, s$ represents one independently simulated dataset under the current VC estimates. $\tilde{\mathbf{u}}^{h}$ is the vector of simulated genetic effects for MC replicate h, and $\widehat{\tilde{\mathbf{u}}}^{h}$ is the vector of estimated genetic effects based on the simulated phenotypes. $\mathbf{K}$ is the relationship matrix used in the model, and thus $\mathbf{K} = \mathbf{H}$ for ssGBLUP model (Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). Similar strategy can be applied for other random effects in the model to form the corresponding parts in the trace vector.

For the residual part in the trace vector:

$$\mathbf{t}_{\mathrm{R}} = \mathbf{R}_0^{-1}(n\mathbf{R}_0 - \mathbf{PEV}_{\mathrm{R}})\mathbf{R}_0^{-1} \quad (5)$$

where $\mathbf{R}_0$ is the current residual (co)variance matrix, n is the number of observations, and the vector of PEV can be computed as

$$\mathbf{PEV}_{\mathrm{R}} = n\mathbf{R}_0 - \frac{1}{s}\sum_{h=1}^{s} \hat{\tilde{\mathbf{e}}}^{h\prime}\hat{\tilde{\mathbf{e}}}^{h} \quad (6)$$

or

$$\mathbf{PEV}_{\mathrm{R}} = \frac{1}{s}\sum_{h=1}^{s}\left(\tilde{\mathbf{e}}^{h} - \hat{\tilde{\mathbf{e}}}^{h}\right)'\left(\tilde{\mathbf{e}}^{h} - \hat{\tilde{\mathbf{e}}}^{h}\right) \quad (7)$$

where $\tilde{\mathbf{e}}^{h}$ is the vector of simulated residuals for MC replicate h, $\hat{\tilde{\mathbf{e}}}^{h}$ is the vector of estimated residuals. A detailed description of the computational strategy can be found in Bermann et al., (2025).

In this study, a MC sample size of five was used. Thus, at each REML iteration, the original MME were solved six times (once for the observed data and five times for the simulated data) and the augmented MME were solved once. Following (Matilainen et al., 2019), the same MC samples were reused across REML iterations to reduce the MC noise. Convergence of the augmented MC AI-REML was assessed based on the round-to-round change in VC estimates, with the convergence threshold value set to 1.0E-10. All analyses were conducted

using the MiX99 software (Pitkänen et al., 2022).

Single-step models

Three single-step models (ssGBLUP, ssGTABLUP, and ssSNPBLUP) were applied for VCE. In all models, genomic and pedigree information were blended to account for residual polygenic effects with $w = 0.1$.

1) ssGBLUP (Legarra et al., 2009, Christensen and Lund, 2010)

In ssGBLUP, the genomic relationship matrix (GRM) was built using Method 1 from VanRaden (2008).

$$\mathbf{G} = (1-w)\mathbf{Z}\mathbf{B}\mathbf{Z}' + w\mathbf{A}_{22} \quad (8)$$

where $\mathbf{Z}$ is the centered marker matrix of size $n \times m$, defined as $\mathbf{Z} = \mathbf{M} - 2\mathbf{P}$, and $\mathbf{M}$ is original $n \times m$ marker covariates matrix (coded as 0,1,2), $\mathbf{P}$ is an $n \times m$ centering matrix, defined as $\mathbf{P} = \mathbf{1}\mathbf{p}'$, and $\mathbf{p}$ is a vector of allele frequencies of base population, m is the number of markers, n is the number of genotyped individuals. $\mathbf{B}$ is a scaling matrix of size $m \times m$, defined as $\mathbf{B}^{-1} = k\mathbf{I}$, with $k = \sum_{l=1}^{m} 2p_l(1-p_l)$, $\mathbf{A}_{22}$ is the pedigree relationship matrix among the genotyped individuals. Allele frequencies in the base population were estimated from all available genomic data using the Bpop software (Strandén and Mäntysaari, 2020). The inverse of the GRM ($\mathbf{G}^{-1}$) was precomputed and stored in memory.

2) ssGTABLUP (Vandenplas et al., 2023)

The ssGTABLUP model used Woodbury matrix identity to avoid explicitly building $\mathbf{G}$ and $\mathbf{G}^{-1}$. The $\mathbf{G}^{-1}$ can be decomposed as

$$\mathbf{G}^{-1} = \frac{1}{w}\mathbf{A}_{22}^{-1} - \frac{1}{w}\mathbf{A}_{22}^{-1}\mathbf{Z}\mathbf{K}^{-1}\mathbf{Z}'\frac{1}{w}\mathbf{A}_{22}^{-1} \quad (9)$$

where $\mathbf{K} = \frac{1}{w}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z} + \frac{1}{1-w}\mathbf{B}^{-1}$ and was precomputed and stored in memory.

3) ssSNPBLUP (Liu et al., 2014)

The ssSNPBLUP model directly operated on the marker covariates matrix ($\mathbf{M}$) and thus required no precomputation.

Data

Data were simulated using AlphasimR (Gaynor et al., 2021) for a single trait with genetic and residual variances of $\sigma_u^2 = 50$ and $\sigma_e^2 = 50$, respectively, in the base population. The breeding scheme was as follows: in each generation, 2,000 females and 200 males were selected based on their phenotypes and randomly mated with each selected female producing one offspring. The simulated population comprised 44,280 individuals with single record each across 20 generations. For VCE, three genotyping scenarios were investigated, in which 10,000, 20,000, or 30,000 youngest individuals were genotyped for 9.000 markers.

Results & Discussion

Variance component estimates were consistent across models and genotyping scenarios (Table 1). As expected, the estimates differed from the initial simulated parameters because phenotypic selection over 20 generations reduced the genetic variance. Moreover, in single-step models, additive genetic variance is estimated on the scale of the combined relationship matrix (**H**) rather than on the scale of the pedigree-based relationship matrix (**A**). The VC estimates were also in agreement with those reported by Gao et al. (2026), in which analytical and MC EM-REML were applied to the same datasets.

Across all analyses, augmented MC AI-REML converged within 11-18 REML iterations. The number of iterations required showed only minor variation among models and genotyping scenarios. In contrast, MC EM-REML required approximately 200-300 REML iterations based on the same datasets (Gao et al., 2026). This difference is consistent with the faster convergence expected for AI-REML, which uses both first (gradient) and second order derivative (curvature) information of the log likelihood compared with the slower but generally more robust EM-REML algorithm.

The computational efficiency of MC AI-REML for ssGBLUP model was also shown by Bermann et al. (2025), who reported that MC-ss-GREML required only 14% of the computing time and 1% of the memory compared to the exact ss-GREML while producing similar VC estimates. The substantial reductions in computing time and memory were attributed mainly to avoiding storage and factorization of the coefficient matrix of MME. Their findings demonstrated the potential of MC AI-REML for computationally efficient VCE in large single-step genomic models.

Computing time increased with the number of genotyped individuals for all models. Substantial differences in computational performance were observed among the models (Table 1). ssGBLUP required the least computing time in all scenarios, whereas ssSNPBLUP was the most computationally demanding model. ssGTABLUP showed intermediate computing times between ssGBLUP and ssSNPBLUP. This ranking was expected given the relatively small numbers of genotyped individuals in the present study. However, as the number of genotyped individuals increased, computing time increased proportionally more for ssGBLUP than for ssGTABLUP or ssSNPBLUP, indicating less favorable computational scaling of ssGBLUP. This pattern is consistent with the computational complexity of breeding value estimation, which scale as $O(n^2)$ for ssGBLUP but approximately as $O(mn)$ for ssGTABLUP and ssSNPBLUP (Vandenplas et al., 2023).

Peak memory usage for ssGBLUP increased sharply with the number of genotyped individuals, whereas the increase was much more moderate for ssGTABLUP and ssSNPBLUP. The high memory requirement of ssGBLUP was mainly due to the precomputation and in-memory storage of the dense $\mathbf{G}^{-1}$. In contrast, ssGTABLUP and ssSNPBLUP substantially reduced memory requirements, although at the expense of increased computing time. Because these methods work on marker covariates matrix. Based on the current datasets, ssGTABLUP provided a compromise between computing time and memory usage, particularly as the number of genotyped individuals increased.

Table 1: Estimated variance components (standard errors), number of REML iterations (N), REML computing time, and peak RAM usage across single-step models and genotyping scenarios

Data[1]	M[2]	σ_u^2	σ_e^2	N	Time	RAM
10	G	35.0 (0.97)	52.2 (0.59)	14	3m18s	2.4
	T	34.1 (0.95)	52.6 (0.59)	14	12m7s	2.2
	S	34.1 (0.95)	52.6 (0.59)	18	39m54s	1.6
20	G	34.7 (0.93)	52.9 (0.52)	11	7m:4s	7.1
	T	34.2 (0.92)	53.3 (0.52)	11	20m9s	2.9
	S	34.2 (0.92)	53.3 (0.52)	18	1h54m	2.3
30	G	36.8 (0.98)	53.0 (0.47)	13	19m30s	15.5
	T	36.6 (0.97)	53.2 (0.47)	12	28m59s	3.6
	S	36.6 (0.97)	53.2 (0.47)	12	1h45m	2.9

[1] 10, 20, and 30 indicate 10,000, 20,000, and 30,000 genotyped individuals, respectively.
[2] G, T, and S indicate ssGBLUP, ssGTABLUP, and ssSNPBLUP, respectively.

Peak memory usages were similar between augmented MC AI-REML and MC EM-REML for the corresponding single-step models (Gao et al., 2026). This was expected because memory usage was determined primarily by the implementation of each single-step model rather than by the REML optimization algorithm. In both studies, memory requirements increased with the number of genotyped individuals for ssGBLUP, whereas they remained considerably lower for ssGTABLUP and ssSNPBLUP. These results suggest that augmented MC AI-REML retains the memory advantages of MC-based VCE while converging substantially faster than MC EM-REML.

Conclusions

We extended augmented AI-REML to augmented MC AI-REML for VCE under three single-step models: ssGBLUP, ssGTABLUP, and ssSNPBLUP. All models produced similar VC estimates within relatively few REML iterations. Computational performance differed among the single-step models. ssGTABLUP offered a better trade-off between computing time and memory requirements. Overall, augmented MC AI-REML provides an efficient approach for VCE in single-step models and has potential for application to large-scale genomic datasets.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci.* 93(2):743-752.

Bermann, M., A. Legarra, I. Aguilar, A. Alvarez-Munera, I. Misztal, and D. Lourenco. 2025. Estimation of (co)variance components for very large datasets and complex single-step genomic models. *Genet. Sel. Evol.* 57(1):63.

Christensen, O. F. and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet Sel Evol.* 42:2.

Gao, H., M. Taskinen, T. J. Pitkänen, A. Ylitalo, N. Vargas Jurado, M. Lidauer, and I. Strandén. 2026. Estimation of variance components in single-step models using MiX99. *in Proc. 13th World Congr. Genet. Appl. Livest. Prod.*, Wisconsin, USA.

García-Cortés, L., C. Moreno, L. Varona, and J. Altarriba. 1995. Estimation of prediction-error variances by resampling. *J Anim Breed Genet.* 112(1-6):176-182.

Gaynor, R. C., G. Gorjanc, and J. M. Hickey. 2021. AlphaSimR: an R package for breeding program simulations. *G3* 11(2).

Gilmour, A. R., R. Thompson, and B. R. Cullis. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51(4):1440-1450.

Jensen, J., E. A. Mäntysaari, P. Madsen, and R. Thompson. 1997. Residual maximum likelihood estimation of (co) variance components in multivariate mixed linear models using average information. *J Indian Soc Agric Stat.* 49:215-236.

Johnson, D. L. and R. Thompson. 1995. Restricted Maximum-Likelihood-Estimation of Variance-Components for Univariate Animal-Models Using Sparse-Matrix Techniques and Average Information. *J Dairy Sci.* 78(2):449-456.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J Dairy Sci.* 92(9):4656-4663.

Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effects. *J Dairy Sci.* 97(9):5833-5850.

Matilainen, K., E. A. Mantysaari, M. H. Lidauer, I. Stranden, and R. Thompson. 2012. Employing a Monte Carlo algorithm in expectation maximization restricted maximum likelihood estimation of the linear mixed model. *J Anim Breed Genet.* 129(6):457-468.

Matilainen, K., E. A. Mantysaari, M. H. Lidauer, I. Stranden, and R. Thompson. 2013. Employing a Monte Carlo algorithm in Newton-type methods for restricted maximum likelihood estimation of genetic parameters. *PLoS One* 8(12):e80821.

Matilainen, K., E. A. Mantysaari, and I. Stranden. 2019. Efficient Monte Carlo algorithm for restricted maximum likelihood estimation of genetic parameters. *J Anim Breed Genet.* 136(4):252-261.

Meyer, K. 1991. Estimating Variances and Covariances for Multivariate Animal-Models by Restricted Maximum-Likelihood. *Genet. Sel. Evol.* 23(1):67-83.

Misztal, I., D. Lourenco, and A. Legarra. 2020. Current status of genomic evaluation. *J Anim Sci.* 98(4).

Pitkänen, P. J., H. Gao, A. Kudinov, M. Taskinen, E. A. Mäntysaari, M. H. Lidauer, and I. Strandén. 2022. From data to genomic breeding values with the MiX99 software suite. *in Proc. 12th World Congr. Genet. Appl. Livest. Prod.,* Rotterdam, The Netherlands.

Strandén, I. and E. A. Mäntysaari. 2020. Bpop: an efficient program for estimating base population allele frequencies in single and multiple group structured populations. *Agric. Food Sci.* 29(3):166–176.

Stranden, I., E. A. Mantysaari, M. H. Lidauer, R. Thompson, and H. Gao. 2024. A computationally efficient algorithm to leverage average information REML for (co)variance component estimation in the genomic era. *Genet. Sel. Evol.* 56(1):73.

Vandenplas, J., J. Ten Napel, S. N. Darbaghshahi, R. Evans, M. P. L. Calus, R. Veerkamp, A. Cromie, E. A. Mantysaari, and I. Stranden. 2023. Efficient large-scale single-step evaluations and indirect genomic prediction of genotyped selection candidates. *Genet. Sel. Evol.* 55(1):37.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci.* 91(11):4414-4423.

A Sire-Dam Threshold Model for U.S. Single-step Calving-Ease Evaluation

S. Savoia*[*1], J. Maskal[1, 2], E. Nicolazzi[1], J. Cole[1], and A. Legarra[1,3]

[1] Council on Dairy Cattle Breeding, Bowie (MD), USA

[2] Department of Animal Sciences, Purdue University, West Lafayette (IN), USA

[3] Department of Animal and Dairy Science, University of Georgia, Athens (GA), USA

Corresponding author: simone.savoia@uscdcb.com

Abstract

The traditional U.S. calving ease (CE) evaluation currently uses a single-trait sire-maternal grandsire (sire-MGS) threshold model implemented in a stand-alone pipeline separate from other trait groups, while the genomic evaluation uses a multi-step approach with de-regressed proofs as input. As the Council on Dairy Cattle Breeding (CDCB) transitions national evaluations toward single-step genomic BLUP (ssGBLUP), a sire-dam threshold model using animal-model pedigree was investigated to better represent maternal information while accommodating incomplete calf identification in U.S. data. CE records from 2002 onward were analyzed using single- and multi-trait threshold models, with first and later calvings modeled either jointly or as correlated traits. Variance components were estimated from an informative subset using Gibbs sampler, and breeding values were subsequently obtained at national scale under both traditional pedigree-based BLUP (pBLUP) and ssGBLUP threshold models using a Newton-Raphson algorithm. Genetic correlations between first and later calvings were high for direct (0.95) and maternal (0.84) effects, supporting adoption of the single-trait model. In the final model, direct and maternal heritabilities were 0.092 and 0.115, respectively, with a genetic correlation of −0.40. Genetic trends from pBLUP and ssGBLUP were very similar, whereas comparisons with the current domestic sire-MGS pBLUP evaluation showed breed- and component-dependent re-ranking, particularly for the maternal effect. The sire-dam threshold model provides a practical framework for integrating CE into the CDCB single-step evaluation system while maintaining continuity with the current U.S. interpretation of the published Sire Calving Ease and Daughter Calving Ease traits.

Key words: Calving Ease; Threshold Model; Sire-Dam; Sire-MGS

Introduction

Genetic evaluations for calving ease (CE) in U.S. Holsteins (HO), date back to 1978 (Berger, 1994). In 2003, the evaluation transitioned to a single-trait sire-maternal grandsire (MGS) threshold model (Van Tassell et al., 2003), incorporating MGS information to partially account for differences in maternal genetic ability and the genetic merit of mates. The evaluation was subsequently extended to consider Brown Swiss (BS) animals in 2005 (Cole et al., 2005) and, in 2020, underwent a broader revision of model effects, data edits, and genetic parameter estimates (Gaddis, 2020). The single-trait sire-MGS structure and its dedicated stand-alone computational pipeline have, however, remained in place and are part of the evaluations currently conducted by the Council on Dairy Cattle Breeding (CDCB; Bowie, Maryland, USA).

For international evaluation, liability-scale solutions are provided to Interbull for MACE evaluation of direct and maternal CE, whereas national evaluations are published as predicted transmitting abilities (PTA) for Sire Calving Ease (SCE) and Daughter Calving Ease (DCE), respectively.

CDCB's transition of the U.S. national genetic evaluation system toward single-step

genomic evaluation provided an opportunity to improve the statistical modeling and reconsider the computational implementation of CE, with the goal of integrating CE into a common evaluation pipeline across trait groups.

The revision process initially considered ICAR guidelines for harmonization of national calving trait evaluations (ICAR, 2026), which favor, in brief, an animal-model formulation accounting for direct and maternal genetic effects and treating first and later parities as distinct but correlated traits.

However, characteristics of the U.S. dairy data precluded direct implementation of the recommended approach. Calf identification is not consistently available in the U.S. National Cooperator Database stewarded by CDCB. Reconstructing dummy calf identities across the entire data set was considered operationally infeasible and would not allow those calves to be linked subsequently to the same animals when they entered the evaluation as calving cows. As an alternative to the current sire-MGS model, a sire-dam formulation was therefore investigated, allowing the dam to contribute directly to the maternal component without requiring identification of the calf as an evaluated animal. Moreover, despite the categorical nature of the CE phenotype, a linear-model approach was initially investigated, prioritizing operational efficiency and facilitating future integration of MACE proofs at national level. Difficulties in transforming and interpreting EBV from the linear scale as expected percentages of difficult calvings redirected the evaluation toward a threshold-model formulation. Maintaining the established definition and interpretation of U.S. CE PTAs was considered an important requirement for the industry; therefore, retaining the threshold model framework allowed continuity in the interpretation and publication of CE evaluations while changing the representation of the maternal component.

The objective of this study was to develop and evaluate a sire-dam threshold model for CE suitable for integration into the CDCB single-step evaluation framework, using the ICAR guidelines as a reference while accommodating the data structure and operational requirements of the U.S. dairy cattle evaluation system.

Materials and Methods

Data

CE records and pedigree information were obtained from the U.S. National Cooperator Database. CE is scored by herd owners on a 5-point scale, where 1 = no problem, 2 = slight problem, 3 = assistance needed, 4 = considerable force, and 5 = extreme difficulty. For genetic evaluation, scores 4 and 5 are combined into a single category. Records were restricted to calvings occurring from 2002 onward to reduce the proportion of observations with missing dam information. To align the CE evaluation with other trait groups, phenotypic records from all dairy breeds and the most common crossbred groups were included, whereas beef-on-dairy crossbreds were excluded. Records were retained for parities 1 through 6 and required known service-sire and dam identification, with dam age at calving falling within predefined parity-specific limits and service-sire required to be older than 12 mo at calving. Observations from multiple birth calvings were excluded, and gestation length was required to fall within predefined limits. Herds were screened based on the distribution of reported CE scores. Contemporary groups were defined as herd-years and were required to contain more than 10 calvings from at least two different service-sires. The resulting dataset used for breeding-value estimation included approximately 23.3 million records, of which about 35% were from first-parity calvings, with a pedigree containing 17.8 million animals. Genotypes were available for approximately 965,000 informative animals, of which 38,968 were included in the APY core (Misztal et al., 2014).

Sire-dam Threshold Models

CE was analyzed using sire-dam threshold models in which service-sires and dams were included in the pedigree as animals together with their ancestors. Missing parent groups were represented by 103 metafounders, defined by sex, breed, and year of birth, and subsequently merged within breed and sex following Legarra and Aguilar (2025). Fixed effects included sex of the calf, dam age within parity, year-season, and covariates accounting for calf heterosis based on ordered sire and dam breed fractions (VanRaden, 1992). Random effects included herd-year contemporary group, correlated genetic effects of the service-sire and dam, and a permanent environmental effect of the dam. Both single-trait and multi-trait formulations were evaluated. In the multi-trait model, first and later calvings were treated as distinct but genetically correlated traits, whereas in the single-trait model were considered expressions of the same trait and analyzed jointly.

Variance components and Breeding Value Estimation

Variance components were estimated from a subset of the data using Gibbs sampling as implemented in GIBBSF90+ v.3.27 (Sorensen et al., 1995; Lourenco et al., 2022), using pedigree information only. For this work and to facilitate biological interpretation, variance and covariance components estimated under the sire-dam parameterization were transformed to the equivalent direct-maternal scale, from which the corresponding genetic parameters were derived. The transformation follows from equating the genetic contributions of the sire-dam and direct-maternal representations of the calving phenotype.

Breeding values were estimated with the sire-dam threshold model at the national scale under both pedigree BLUP (pBLUP) and single-step genomic BLUP (ssGBLUP) using maximum a posteriori estimation with the Newton-Raphson algorithm (Bermann et al., 2026). Estimates on the underlying liability scale from the investigated models were compared with those from the current sire-MGS model, thereby avoiding differences associated with definitions of the phenotypic bases and with transformation to the observed PTA scale. To facilitate comparison of maternal effects, dam-effect solutions (d) from the new model were expressed as $d/2$, placing them on the same scale as the current MGS solutions.

Results & Discussion

Figure 1 shows the phenotypic trends in the incidence of difficult calvings (scores 4 and 5) for first and later parities over the last 20 years. The strong reduction in dystocia likely reflects both genetic selection for improved CE and changes in herd management and reproductive practices, including increased use of sexed semen. This marked phenotypic change, together with the very low incidence of difficult calvings observed in recent years, may have contributed to the poor performance of the linear-model approach, particularly when translating EBV to the observed frequency scale, because the transformation assumes a fixed rate of incidence of dystocia over time.

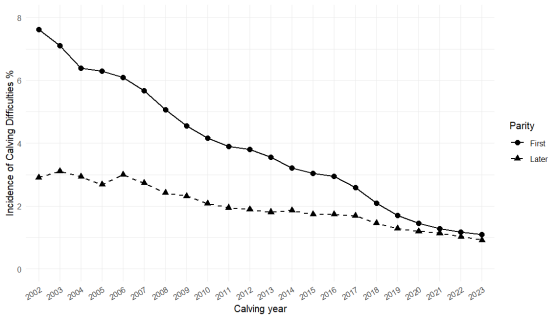


Figure 1. Incidence of difficult calvings (original CE scores 4 and 5, combined as category 4 for evaluation) by calving year in first and later parities.

Table 1 summarizes the genetic parameters obtained from the multi-trait sire-dam threshold model after transformation into the direct-maternal scale to make them comparable to literature. Direct heritability estimates were lower than maternal estimates in both parity groups (first calvings: 0.068 vs. 0.128; later

calvings: 0.048 vs. 0.065). This pattern contrasts with previous studies, in which direct heritability for CE has generally exceeded maternal heritability. For example, Eaglen and colleagues (2012) reported direct and maternal heritabilities of approximately 0.11 and 0.08, respectively, at first parity, compared with 0.03 and 0.02 at later parities.

The direct-effect genetic correlation across parities was close to unity (0.945), whereas the maternal-effect correlation was lower (0.835), but remained indicative of substantial genetic similarity across parity groups. These estimates supported treating first and later calvings as a single trait for subsequent analyses, a choice also consistent with the increasingly similar phenotypic trajectories of the two parity groups in recent years.

Table 1: Direct and maternal genetic parameters for CE from the multi-trait sire-dam threshold model. Heritabilities are shown on the diagonal and genetic correlations below the diagonal

Genetic Effect *	Dir. 1st	Dir. Later	Mat. 1st	Mat. Later
Dir. 1st	0.068			
Dir. Later	0.945	0.048		
Mat. 1st	-0.254	-0.143	0.128	
Mat. Later	-0.208	-0.116	0.835	0.065

* Dir. 1st: direct first calving; Dir. Later: direct later calvings; Mat. 1st: maternal first calving; Mat. Later: maternal later calvings.

Genetic parameters obtained from the single-trait model showed direct and maternal heritabilities of 0.092 and 0.115, respectively, compared with 0.072 and 0.053 under the current single-trait sire-MGS model.

The genetic correlation between direct and maternal effects was -0.403, indicating an antagonistic genetic relationship between the two components. Such negative direct-maternal correlation is consistent with previous CE studies; Eaglen and colleagues (2012) reported estimates of -0.47 and -0.40 for first and later parities, whereas the current U.S. sire-MGS evaluation was originally parameterized with a direct-maternal correlation of approximately -0.12.

The genetic trends from the proposed pBLUP evaluation were broadly consistent with those from the current CDCB evaluation for the direct component, with both showing a gradual decline (reduction of calving difficulties) over time in HO and a similar, although more variable, pattern in BS (Figure 2a). Larger differences were observed for the maternal component. In particular, the new pBLUP model produced a markedly steeper decline in the maternal effect for recent HO birth years than the current sire-MGS evaluation, with a similar but less pronounced divergence in BS (Figure 2b). This is likely to reflect the change in maternal representation from MGS to dam and the resulting increased information and redistribution of genetic signal between direct and maternal components.

Comparisons between pBLUP and ssGBLUP under the new sire-dam threshold model showed very similar genetic trends for proven bulls, as expected, indicating that the main differences relative to the current evaluation were driven primarily by the revised model structure rather than by the incorporation of genomic information.

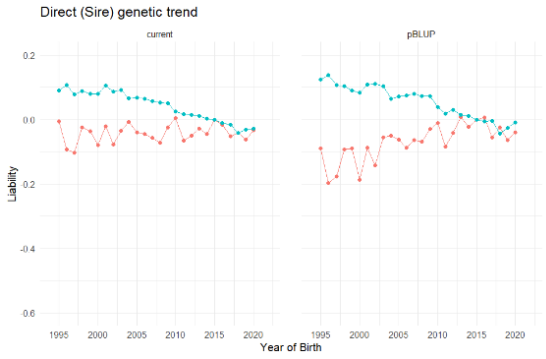


Figure 2a. Genetic trends for CE direct genetic effect under the current CDCB evaluation (left panel) and the proposed sire-dam threshold model (right panel) for HO (cyan) and BS (pink). Solutions were expressed on liability scale relative to a genetic base defined within-breed by bulls born in 2015. All bulls were required to have at least 25 observations distributed across at least two herds.

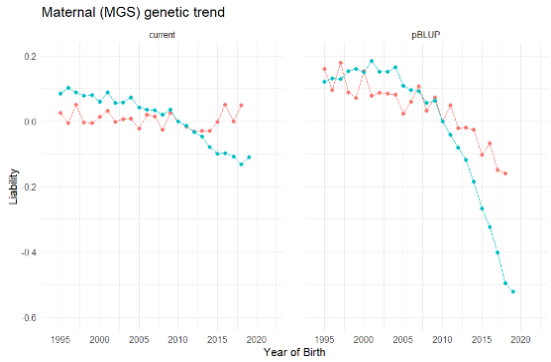


Figure 2b. Genetic trends for CE MGS genetic effect under the current CDCB evaluation (left panel) and the proposed sire-dam threshold model (right panel) for HO (cyan) and BS (pink). Solutions were expressed on liability scale relative to a genetic base defined within-breed by bulls born in 2010. All bulls were required to have at least 25 daughters' observations distributed across at least two herds.

Spearman correlations between solutions from the current CDCB evaluation and the proposed sire-dam pBLUP model indicated re-ranking, with differences across breeds and genetic components. For the direct effect, correlations were 0.79 for BS and 0.84 for HO, whereas for the maternal effect they were 0.72 and 0.87, respectively. The extent of re-ranking may reflect breed-specific differences in data structure, including the amount of information available for individual sires and their use in crossbreeding. These differences may be particularly relevant because the proposed model evaluates a broader population than the current evaluation, which is restricted to HO, BS, and HO × BS crosses. These comparisons were based exclusively on national evaluations; consequently, the magnitude of re-ranking may differ after incorporation of Interbull MACE information, particularly for breeds and bulls with limited domestic information.

Conclusions

The proposed sire-dam threshold model provides a practical framework for integrating CE into the CDCB single-step evaluation system while preserving the established interpretation of SCE and DCE PTAs. High genetic correlations across parities supported a single-trait formulation, and results indicated a stronger maternal contribution than under the current sire-MGS model. Genetic trends from pBLUP and ssGBLUP were very similar, suggesting that differences from the current evaluation were driven mainly by the revised model structure. Overall, the sire-dam threshold model appears suitable for routine national implementation, with future work focused on MACE integration.

Acknowledgments

The authors acknowledge the support of Dairy Herd Improvement (DHI) organizations and Dairy Records Processing Centers (DRPC) for processing and relaying the information to CDCB, as well as purebred breed associations for providing pedigree data.

References

Berger, P.J. 1994. Genetic prediction for calving ease in the United States: Data, models, and use by the dairy industry. *J. Dairy Sci.*, 77, 1146-1153. https://doi.org/10.3168/jds.S0022-0302(94)77051-X

Bermann, M., Legarra, A., Misztal, I., and Lourenco, D. 2026. Methods for joint genetic prediction of multiple ordinal categorical and continuous traits. *Genetics* 233 (1), iyag086. https://doi.org/10.1093/genetics/iyag086

Cole, J.B., Goodling, R.C., Wiggans, G.R., and VanRaden, P.M. 2005. Genetic evaluation of calving ease for Brown Swiss, Jersey, and Holstein bulls from purebred and crossbred calvings. *J. Dairy Sci.*, 88, 1529-1539. https://doi.org/10.3168/jds.S0022-0302(05)72822-8

Eaglen, S.A.E., Coffey, M.P., Woolliams, J.A., and Wall, E. 2012. Evaluating alternate models to estimate genetic parameters of calving traits in United Kingdom Holstein-Friesian dairy cattle. *Genet. Sel. Evol.*,

44:23. https://doi.org/10.1186/1297-9686-44-23

Gaddis, K. P. 2020. Calving Ease, Stillbirth Evaluations Enhanced. Council on Dairy Cattle Breeding, Bowie, MD, USA. https://uscdcb.com/wp-content/uploads/2024/05/CDCB-Parker-Gaddis-Calving-Enhancements-03_2020-1.pdf

ICAR. 2026. Section 09 - Dairy Cattle Genetic Evaluation. ICAR Guidelines. International Committee for Animal Recording. Accessed May 2026. https://wiki.icar.org/index.php/Section_09_%E2%80%93_Dairy_Cattle_Genetic_Evaluation

Legarra, A., and Aguilar, I. 2025. Tools to refine unknown parent group definitions. *J. Dairy Sci.* 108:1747-1756. https://doi.org/10.3168/jds.2024-25408

Lourenco, D., Tsuruta, S., Aguilar, I., Masuda, Y., Bermann, M., Legarra, A. and Misztal, I. 2022. Recent updates in the BLUPF90 software suite. *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, Rotterdam, The Netherlands. 1530-1533.

Misztal, I., Legarra, A., and Aguilar, I. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.*, 97, 3943-3952. https://doi.org/10.3168/jds.2013-7752

Sorensen, D.A., Andersen, S., Gianola, D., and Korsgaard, I.R. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.*, 27, 229-249. https://doi.org/10.1051/gse:19950303

VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J. Dairy Sci.*, 75, 3136-3144. https://doi.org/10.3168/jds.S0022-0302(92)78077-1

Van Tassell, C.P., Wiggans, G.R., and Misztal, I. 2003. Implementation of a sire-maternal grandsire model for evaluation of calving ease in the United States. *J. Dairy Sci.*, 86, 3366-3373. https://doi.org/10.3168/jds.S0022-0302(03)73940-X

Genetic correlations between juvenile mortality traits and traits in selection on French dairy cattle breeds

S. Aguerre[1,2*], A. Baur[1,3], L. Chapard[1,2], R. Vallée[1,2], H. Leclerc[1,3], F. Besnard[1,3], S. Fritz[1,3], S. Mattalia[1,2], P. Croiseau[1]

[1] Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France

[2] IDELE, Paris, France

[3] ELIANCE, Paris, France

Corresponding author: sophie.aguerre@idele.fr

Abstract

Reducing juvenile mortality is a pressing challenge for the dairy cattle industry due to its economic, environmental, and animal welfare impacts. While juvenile mortality has multiple causes, genetics plays a substantial role. The implementation of a routine genomic evaluation could therefore contribute to achieving this goal. In France, a genomic evaluation for mortality within the first 48 hours of life is already available for dairy cattle breeds. As part of the SEVITAL project, new juvenile mortality traits covering the first year of life are being studied for inclusion, in the breeding programs of nine dairy breeds (Abondance, Brown Swiss, Holstein, Jersey, Montbéliarde, Normande, Simmental, Tarentaise and Vosgienne). Life history information, i.e. date of birth, date and cause of death (natural death or slaughter) were extracted from the French national cattle database for females born from 2010 to 2024. Three traits have been defined based on the age at which the animal died: 3-30 days, when the leading cause of calf mortality is diarrhea; 31-365 days, when calf mortality is mainly due to respiratory diseases; 3-365 days, to cover the whole period. These traits were defined as binary traits: 0 was assigned to animals still alive during the considered period, and a value of 100 was assigned to animals that died. For each trait, single-step genomic evaluations were carried out with HSSGBLUP (Tribout *et al.*, 2020) using a univariate animal model that included *herd x year-of-birth* and *region x month-of-birth x year-of-birth* as fixed effects. Estimated breeding values (EBVs) of bulls were calculated using the performances of a total number of 4,750 to 2,236,362 animals depending on the breed, among which 2,683 to 608,585 were genotyped. Correlations between juvenile mortality traits and traits routinely evaluated (milk production traits, udder health, fertility, mortality within the first 48 hours) were estimated using Calo's method (Calo et al., 1973). The correlations between EBVs adjusted for their reliability estimated with this method provided an approximation of the genetic correlations. By offering a better understanding of the genetic relationships between traits, these results will help breeding companies determine how to incorporate juvenile mortality traits in their overall breeding objectives.

Key words: Dairy cattle, juvenile mortality, genetic correlations

Introduction

Juvenile mortality is an important challenge for the cattle sector as it represents a major animal welfare issue and is associated with important economic losses and impacts on selection opportunities.

In addition to management practices, genetic selection can be a useful tool to improve calf survival. Several countries have developed genomic evaluations to (i) reduce calf mortality such as the Netherlands (van Pelt et al., 2012), the Nordic countries (Nordic Cattle Genetic Evaluation, 2025) or the United Stats (Neupane

et al., 2021) or (ii) improve resistance to its main underlying causes, i.e. digestive and respiratory diseases such as Canada (Lynch et al., 2024) or the United States (Neupane et al., 2025).

In France, a routine genomic evaluation exists only for perinatal mortality for dairy cattle breeds (GenEval, 2026). To go further, a genomic evaluation for juvenile mortality traits covering the first year of life is being developed. As part of this work, relationships between juvenile mortality traits and traits currently selected for were investigated.

Materials and Methods

Juvenile mortality traits

Datasets

Life-history information including date of birth and date and cause of death (slaughter or natural death) were extracted from the French national bovine database for nine dairy breeds (Abondance, Brown Swiss, Holstein, Jersey, Montbéliarde, Normande, Simmental, Tarentaise and Vosgienne).

The first year of life was divided into two periods to reflect the principal underlying cause of mortality: 3–30 days, during which digestive diseases are the most frequent cause of death, and 31–365 days, during which mortality is mainly due to respiratory diseases. The overall period of 3–365 days was also studied. Perinatal mortality, which includes stillbirth and death within the first 48 hours of life, is already evaluated. Therefore, the first two days of life were excluded from the analyses.

Juvenile mortality was analyzed as a binary trait. For a given period, animals were considered alive if they were alive at the end of the period or slaughtered during that period, and dead if they died of natural causes during the period.

Data of singletons females born in France between 2010 and 2025 were considered (Table 1). Females that did not stayed in the farm where they were born farm during the considered period and males were not taken into account in this study. Indeed, these animals can face health challenges during transport and variable conditions from one farm to another. At least 5 females were included for each herd x birth year and region x birth month x birth year combination. At least 10 paternal half-sisters born in a minimum of 5 herds were retained.

Table 1: Number of calf records available for each period in the nine French dairy cattle breeds

Breed	**n phenotypes**		
	3-30 days	**31-365 days**	**3-365 days**
Abondance	60 508	49 785	53 695
Brown Swiss	58 969	49 183	54 003
Holstein	2 784 101	2 118 154	2 236 354
Jersey	42 509	31 335	36 745
Montbéliarde	1 725 075	1 477 134	1 583 405
Normande	945 345	839 798	894 760
Simmental	64 062	55 219	58 660
Tarentaise	23 221	18 512	20 503
Vosgienne	5 423	4 498	4 750

Model

Estimated breeding values (EBVs) were estimated for each period with a hybrid single step model (Fernando et al., 2016) using HSSGBLUP program (Tribout et al., 2020):

$$\begin{bmatrix} y_n \\ y_g \end{bmatrix} = X\beta + \begin{bmatrix} Z_n & 0 \\ 0 & Z_g M_g \end{bmatrix} \begin{bmatrix} u_n \\ \alpha \end{bmatrix} + e$$

where y is the vector with phenotypes; n refers to non-genotyped animals and g to genotyped animals; β is the vector of fixed effects of herd x birth year and of region x birth month x birth year; M_g are centered SNP genotypes of genotyped animals; $u_n \sim N(0, A_{nn}\sigma_g^2)$ corresponds to EBVs of non-genotyped animals; $\alpha \sim N(0, I\sigma_\alpha^2)$ is the vector of markers effects with $\sigma_\alpha^2 = \frac{\sigma_g^2}{\sum_{1st\ SNP}^{n^{th}\ SNP} 2p_k(1-p_k)}$; $e \sim N(0, I\sigma_e^2)$ is the vector of residual effects; X and Z are incidence matrices for fixed and random animal effects, respectively, and I is the identity matrix.

EBVs were then inverted so that a positive EBV indicates better survival.

Genetic correlations

Genetic correlations between juvenile mortality traits and traits in selection including milk yield (MY), fat yield (FY), protein yield (PY), fat content (FC), protein content (PC), somatic cell score (SCS), survival at birth (SURVIVAL_B), survival at calving (SURVIVAL_C), cows and heifers fertility (FERT_COW, FERT_HEIFER), interval between calving and first artificial insemination (CALV_AI1), non-return rate for cows and heifers (NR_COW, NR_HEIFER) were approximated using Calo's method (Calo et al., 1973):

$$r_g = \frac{\sqrt{\sum_{i=1}^{n} \rho_{i,EBV1} \times \sum_{i=1}^{n} \rho_{i,EBV2}}}{\sum_{i=1}^{n} \rho_{i,EBV1} \times \rho_{i,EBV2} \times r_{EBV1,EBV2}}$$

where $\rho_{i,EBV1}$ is the reliability of the EBV for trait 1 for the bull i, $\rho_{i,EBV2}$ is the reliability of the EBV for trait 2, and $r_{EBV1,EBV2}$ is the Pearson correlation between the EBV for trait 1 and EBV for trait 2. EBVs with a reliability lower than 0.4 were not included.

Results & Discussion

Low to moderate estimates were obtained for the genetic correlations between juvenile mortality traits and traits currently selected for (Table 2). Higher estimates were obtained for the longest periods (31-365days and 3-365 days) in comparison with the estimates obtained for 3-30days. The positive values of estimated correlations indicate favourable relationships between traits as EBVs were inverted for juvenile mortality traits.

In the literature, genetic correlations between calf survival or resistance to diarrhea and respiratory diseases are generally low to moderate (Axford et al., 2024; Jamrozik et al., 2025; Gonzales-Peña et al., 2019, 2020; Neupane et al., 2021, 2025). Our estimated correlations between juvenile mortality traits and milk production traits (between 0.038 and 0.272) were in accordance with the values obtained by Neupane et al. (2021) who found

Table 2: Estimated genetic correlations between juvenile mortality traits and traits in selection in Holstein and Montbéliarde breeds.

	Holstein			Montbéliarde		
Trait	**3-30 days**	**31-365 days**	**3-365 days**	**3-30 days**	**31-365 days**	**3-365 days**
Milk yield	0.122	0.141	0.149	0.494	0.480	0.542
Fat yield	0.145	0.272	0.248	0.484	0.476	0.535
Protein yield	0.190	0.267	0.263	0.599	0.577	0.645
Fat content	0.038	0.155	0.123	0.023	0.026	0.011
Protein content	0.099	0.186	0.169	0.257	0.211	0.241
SCS	0.121	0.231	0.228	0.363	0.299	0.344
Survival at birth	0.329	0.294	0.309	0.514	0.542	0.562
Survival at calving	0.391	0.420	0.473	0.327	0.358	0.297
Cow fertility	0.261	0.424	0.355	0.542	0.577	0.601
Heifer fertility	0.179	0.339	0.260	0.317	0.288	0.343
Interval between calving and 1rst AI	-0.046	0.032	-0.002	0.009	-0.011	-0.058
Cow non return rate	0.356	0.436	0.409	0.278	0.289	0.370
Heifer non return rate	0.168	0.213	0.175	0.071	-0.018	0.114

EBVs for juvenile mortality traits were inverted so that positive EBVs are associated with better survival.

correlations ranging from 0.18 and 0.22 between heifer liveability and milk yields in Holstein. Although, our estimates for correlations with fertility traits (between -0.002 and 0.436) were mostly higher than those of Neupane et al. (2021)

Same trends could be observed in Holstein and Montbéliarde breeds. However, it can be noted that the correlations were substantially higher in Montbéliarde for milk production traits.

For local breeds, reasonable estimation of genetic correlations could not be achieved. This is probably due to the low reliabilities associated with EBVs for juvenile mortality traits as a result of low heritabilities and limited number of performances

Conclusions

Low to moderate favourable correlations were estimated between juvenile mortality traits and traits currently selected for. These results suggest that juvenile mortality can be reduced without impairing selection for other economically important traits.

Acknowledgments

The SEVITAL project received funding from APIS-GENE. IDELE is grateful for the financial support that it received from the French Ministry of Agriculture (Paris, France) through the CASDAR program.

The authors are grateful to the breeding societies and organizations involved in the SEVITAL project.

References

Axford, M.M., Khansefid, M., Haile-Mariam, M. Godddard, M.E., Pryce, J.E. 2024. Genetic evaluation for stillbirth and preweaning mortality in Australian dairy cattle. *J Dairy Sci*. 107(9), 6994-7008. 10.3168/jds.2023-23891

Calo, L. L., McDowell, R.E., VanVleck, L.D., and Miller P. D. 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. *J. Anim. Sci.* 37, 676–682. https://doi.org/10.2527/jas1973.373676x

Fernando, R.L., Cheng, H., Golden, B.L., Garrick, D.J. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. *Genet. Sel. Evol*. 48, 96. https://doi.org/10.1186/s12711-016-0273-2

GenEval. 2026. Evaluation génétique des bovins laitiers Production laitière – Morphologie – Caractères fonctionnels. https://geneval.fr/methode-evaluation-BL

Gonzales-Peña, D., Vukasinovic, N., Brooker, J.J., Przybyla, C.A., DeNise, S.K. 2019. Genomic evaluation for calf wellness traits in Holstein cattle. *J Dairy Sci.* 102, 2319-2329 https://doi.org/10.3168/jds.2018-15540Get rights and content

Gonzales-Peña, D., Vukasinovic, N., Brooker, J.J., Przybyla, C.A., Baktula, A., DeNise, S.K. 2020. Genomic evaluation for wellness traits in US Jersey cattle. *J Dairy Sci*. 103(2), 1735-1748) https://doi.org/10.3168/jds.2019-16903

Jamrozik, J., Lynch, C., Sullivan, P.G., Fleming, A. 2025. Genomic evaluation for calf health in Canada. *Interbull bulletin 61*, 10-15 https://journal.interbull.org/index.php/ib/article/view/1986/1954

Lynch, C., Schenkel, F.S., van Staaveren, N., Miglior, F., Kelton, D., Baes, C.F. 2024. Investigating the potential for genetic selection of dairy calf disease traits using management data. *J. Dairy Sci*. 107, 1022-1034 https://doi.org/10.3168/jds.2023-23780

Neupane, M., Hutchison, J.L., VanTassel, C.P., VanRaden, P.M. 2021. Genomic evaluation of dairy heifer livability. *J. Dairy Sci.* 104,

8959-8965. https://doi.org/10.3168/jds.2020-19687

Neupane, M., Parker Gaddis, K.L., Toghiani, S., Miles, A.M., Graham, J.R., Burchard, J.F., Dürr, J.W., Cole, J.B., O'Connell, J.R., Van Tassell, C.P., VanRaden, P.M. 2025. Improving dairy calf health through inclusion of diarrhea and respiratory health data into a US national genetic evaluation. *J. Dairy Sci*. 108, 11164-11172 https://doi.org/10.3168/jds.2025-26497

Nordic Cattle Genetic Evaluation. 2005. NAV official genetic evaluation of Dairy Cattle-Data and genetic models. https://nordicebv.info/wp-content/uploads/2025/11/NAV-routine-genetic-evaluation-November-2025.pdf

van Pelt, M.L., Eding, H., Vessies, P., de Jong, G. 2012. Developing a genetic evaluation for calf survival during rearing in the Netherlands. *Interbull bulletin 46*, 61-65 https://journal.interbull.org/index.php/ib/article/view/1741

Tribout, T., Ducrocq, V., Boichard, D. 2020. HSSGBLUP : A singlestep SNP-BLUP genomic evaluation software adapted to large livestock populations. In Proc. 6th International congress of quantitative genetics. https://hal.inrae.fr/hal-03155472

Re-ranking of Imported Angus Sires Following Integration of Foreign Genotypes into a National Genomic Evaluation

H.N Kruger[1], J.J. Reding[1] and R.R. van der Westhuizen[1]

[1]South African Stud Book and Animal Improvement Association, Bloemfontein, South Africa

Corresponding author: hannah@studbook.co.za

Abstract

In South Africa, imported Angus sires are frequently introduced into the national genetic evaluation with little to no local progeny records, limited pedigree depth, and typically without own or progeny performance data recorded under local production conditions. Historically, interim evaluations incorporated country-of-origin genetic merit information, expressed as percentile rankings and blended them into the South African evaluation until sufficient local progeny records became available. This avoids an initial underestimation of their EBVs. Such approaches may not fully account for genotype-by-environment interactions, population-specific allele frequencies, differences in genetic parameters, or divergent selection pressures between source and recipient populations. Single-step multiple-trait genomic BLUP (MT ssGBLUP) enables direct integration of SNP genotypes, allowing imported sires to be evaluated relative to a locally genotyped reference population. This study compared EBVs from the national pedigree-based BLUP evaluation incorporating percentile adjustments with EBVs from MT ssGBLUP without percentile adjustments for 194 imported Angus sires across pre-weaning traits. Pearson correlations between evaluation methods were 0.92 for Birth Weight Direct (BWD), 0.80 for Weaning Weight Direct (WWD), and 0.71 for Weaning Weight Maternal (WWM). Despite the high correlation for BWD, individual re-ranking averaged 21 positions. Greater re-ranking was observed for WWD (maximum −117 to +139 positions) and WWM (−145 to +147 positions), likely reflecting lower data density and the delayed accumulation of maternal information. Across traits, incorporation of genomic information resulted in substantial redistribution of ranking, with most imported sires improving relative position, while others declined depending on genomic relationships with locally measured and genotyped animals. These results indicate that MT ssGBLUP does not uniformly inflate evaluations but reallocates merit according to realized genomic similarity within the South African reference population. Strengthening genomic connectedness between imported sires and locally measured animals is therefore essential to reduce uncertainty during early evaluation stages and to improve the reliability of selection decisions under production environments distinct from those of origin.

Key Words: genetic evaluation, single-step GBLUP, imported sires

Introduction

The genetic improvement of beef cattle populations increasingly depends on the exchange of genetic material and genetic information across national borders. In South Africa, as in many countries with comparatively small registered seedstock populations, breeders routinely import semen and live animals from established Angus populations in North America, Australia and elsewhere to broaden the genetic base and accelerate the rate of genetic gain (Guarini et al., 2019; Adekale et al., 2025). These imported sires typically enter the national pedigree with little to no local progeny records, shallow pedigree ties to the resident population, and, in many cases, no performance data recorded under South African production conditions. To avoid the systematic underestimation of breeding values that would

otherwise result from evaluating such animals on parent-average information alone, national genetic evaluation systems have historically relied on interim procedures that translate country-of-origin genetic merit, often expressed as within-breed percentile rankings, into the local evaluation until sufficient local progeny records accumulate.

While practical, percentile-based blending of foreign genetic merit does not explicitly account for several biological and statistical realities that distinguish the source and recipient populations. Chief among these is genotype-by-environment (G×E) interaction, which has repeatedly been shown to cause the re-ranking of sires and a decline in the genetic correlation between the same trait expressed in different environments as those environments diverge. Breeders must carefully assess the performance of sires and their offspring in the target environment before making selection decisions, because the average performance of progeny sired by bulls selected under more favourable conditions can deviate substantially from expectation when commercial conditions differ from those under which the sires were originally evaluated. In beef cattle, sire-by-region and sire-by-system interactions have been reported for growth and reproductive traits using both classical multi-trait and reaction-norm approaches, with several studies showing that these interactions can substantially alter the ranking of sires for growth traits based on their estimated breeding values. Multi-trait analyses of fertility traits in tropical beef populations have similarly shown that genetic correlations for the same trait measured in different environments tend to decline as those environments diverge, with correlations below approximately 0.8 taken as evidence of genotype-by-environment interaction and an indication that splitting selection by environment would be beneficial for the breeding program. More recently, reaction-norm studies in Nellore cattle have reported genotype-by-environment interactions for feed intake traits with genetic correlations between extreme environmental levels as low as 0.22–0.26, resulting in selection coincidences of only 40–53% between environments, underscoring how substantially sire merit can be redistributed when evaluated across divergent production environments (Silva Neto et al., 2023). Beyond G×E, differences in population-specific allele frequencies, linkage disequilibrium patterns, genetic parameters, and historical selection pressure between the exporting and importing populations further complicate the direct transfer of foreign genetic merit into a national evaluation (Silva Neto et al., 2024; Adekale et al., 2025).

Multi-step approaches that address these issues by deregressing foreign estimated breeding values (EBVs) and blending them with national information, as implemented for Multiple Across-Country Evaluation (MACE) proofs in dairy cattle and Interbeef proofs in beef cattle, have been shown to improve reliability and reduce bias relative to pedigree-only evaluations, but they remain vulnerable to double-counting of information and to loss of accuracy inherent in the deregression process itself (Guarini et al., 2019; Adekale et al., 2025). Single-step genomic BLUP (ssGBLUP) offers an alternative that avoids these intermediate steps by combining pedigree, phenotypic, and genomic information into a single evaluation, connecting genotyped imported sires directly to the local reference population through realized genomic relationships rather than through expected pedigree relationships or externally derived proofs (Aguilar et al., 2010; Christensen and Lund, 2010). This approach has been implemented successfully in large national beef cattle evaluations, where the increase in ability to predict future performance for birth weight, weaning weight, and post-weaning gain with single-step genomic BLUP was 25% in a 2014 dataset and 36% in a 2017 dataset compared with traditional BLUP, with genetic trends revealing that multistep evaluations had overestimated merit, especially for traits with fewer phenotypic records (Lourenco et al., 2018).

Despite these developments, it remains unclear how imported Angus sires with limited South African performance records would be re-ranked if evaluated through a multi-trait ssGBLUP system that draws directly on realized genomic connectedness with the local reference population, rather than through the percentile-adjusted pedigree-based system currently in use. Understanding the magnitude and direction of this re-ranking, and the traits for which it is most pronounced, is essential for assessing whether the current interim evaluation approach remains fit for purpose and for guiding future investment in genomic connectedness between imported and locally measured animals. The objective of this study was therefore to compare EBVs for pre-weaning growth traits obtained from the current national pedigree-based evaluation, with EBVs obtained from a multi-trait single-step genomic evaluation, for a sample of 194 imported Angus sires within the South African Angus population.

Materials and Methods

Data

The phenotypic data were acquired from the LOGIX Genetic Evaluation System (SA Stud Book). Records for birth weight (BW) and weaning weight (WW) used in the evaluation are summarised in Table 1. The Angus cattle breeders' society of South Africa has a recording rate of 99% for birth weight and 74% for weaning weight.

Table 1: Total number of weight records for birth weight (BW) and weaning weight (WW).

Trait	Number of Male Records	Number of Female Records	Total Number of Records
BW	55 892	83 281	139 173
WW	41 690	64 530	106 220

Pedigree information on 280 194 animals dating back to 1931, of which 20 853 were registered as active animals, consisting of 6 244 males and 14 609 females.

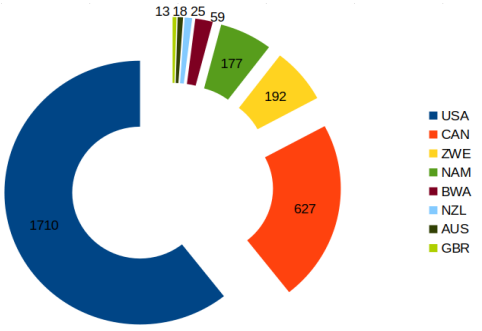


Figure 1. The top eight countries of imported Angus sires (USA: United States of America; CAN: Canada; ZWE: Zimbabwe; NAM: Namibia; BWA: Botswana; NZL: New Zealand; AUS: Australia; GBR: Great Britain)

The pedigree contained 2,847 imported sires from 18 countries, with the top eight countries shown in Figure 1. Of these, 194 imported Angus sires had genomic profiles available and were included in the evaluation.

Furthermore, the genomic profiles of 1,427 South African born Angus animals, comprising 63% males and 37% females, were used in this study. Animals were genotyped across four commercial single nucleotide polymorphism (SNP) arrays of varying densities. Genotyping was done on commercial variants of the Illumina BovineSNP50 v.3; namely the SASB 50k, ICBF IDB v.2, Versa 50k and the newer Versa 80k. Quality control of genomic SNP data, performed in PLINK v1.90b6.10 (Purcell et al. 2007), retained only autosomal SNPs with a known base-pair position, a call rate ≥0.90, a MAF ≥0.01, and no significant deviation from Hardy–Weinberg equilibrium ($p>0.001$). Animals were required to have a call rate ≥0.90 while individuals with ≥0.95 identical genotypes were discarded. Genotypes were imputed with FImpute v3 (Sargolzaei et al. 2014), using pedigree information, to a density of approximately 51,000 SNPs.

Models

In order to predict estimated breeding values in multiple scenarios the same multi-trait animal linear model was used. The model was defined as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where **y** is the vector of phenotypes, **b** is the vector of fixed effects, **u** is a vector of random genetic effects, including direct and maternal additive genetic effects; the model also accounted for the relevant (co)variances among genetic effects and other random effects, with **u** ~ N(0,A σ_u^2), where A is the pedigree-based matrix and σ_u^2 is the direct-genetic variance, **e** represents the residual, where e ~ N(0,I σ_e^2), with σ_e^2 representing the residual variance, **I** the identity matrix while **X** and **Z** are incidence matrices for **b** and **u** respectively.

Fixed effects in **b** for the model included herd x year x season x treatment group x birth status with age adjustment within contemporary group, sex, dam parity (1 or >1) and linear (α) and quadratic (α^2) regression coefficients for age of dam. Estimation of variance components for the above model was calculated using restricted maximum likelihood (REML) optimised with quasi-Newton procedure using analytical gradients in Variance Component Estimation (VCE) (Groeneveld, 2010) software. MiX99 (MiX99 Development Team, 2017) was used to predict both traditional EBVs and GEBVs using the same model. The ssGBLUP model utilises the inverse of the joint relationship matrix $\mathbf{H^{-1}}$ (Aguilar et al. 2010; Legarra et al. 2014).

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix},$$

where $\mathbf{A}^{-1}$ is the inverse of the pedigree-based matrix, A_{22} is the overlapping part of A for the genotyped animals and **G** is the genomic relationship matrix (GRM). The GRM was constructed using VanRaden Method 1 of (VanRaden, 2008). This was achieved using a weighting factor of 0.1, with the resulting matrix blended with the pedigree-based relationship matrix (A_{22}) using the RelaX2 HGInv program (Strandén, 2014) to produce the G^{-1} matrix. The resulting matrix was then used in MiX99 for the internal construction of H^{-1} in ssGBLUP.

Results & Discussion

A total of 194 imported Angus sires were included in the comparison of estimated breeding values (EBVs) derived from pedigree-based BLUP and multi-trait single-step GBLUP (MT ssGBLUP) across three traits: Birth Weight Direct (BWD), Weaning Weight Direct (WWD), and Weaning Weight Maternal (WWM).

The mean EBV difference revealed a consistent upward shift across all three traits upon the integration of genomic information, with BWD (Figure 2) increasing by a mean of 0.82 kg (SD = 0.41), WWD (Figure 3) by 21.87 kg (SD = 3.79), and WWM (Figure 4) by 3.34 kg (SD = 2.11). This systematic upward shift is consistent with the genomic reference population capturing realised genetic relationships between animals that were not accounted for in the pedigree BLUP run. Where pedigree-based evaluation assumes Mendelian sampling is zero on average across non-relatives, the genomic relationship matrix reflects the actual proportion of the genome shared between individuals, thereby capturing realised genetic relationships that pedigree information alone cannot resolve.

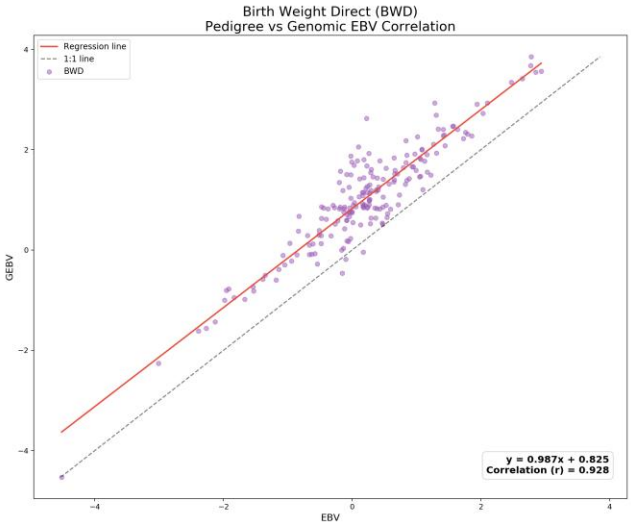


Figure 2. Pearson correlation for birth weight

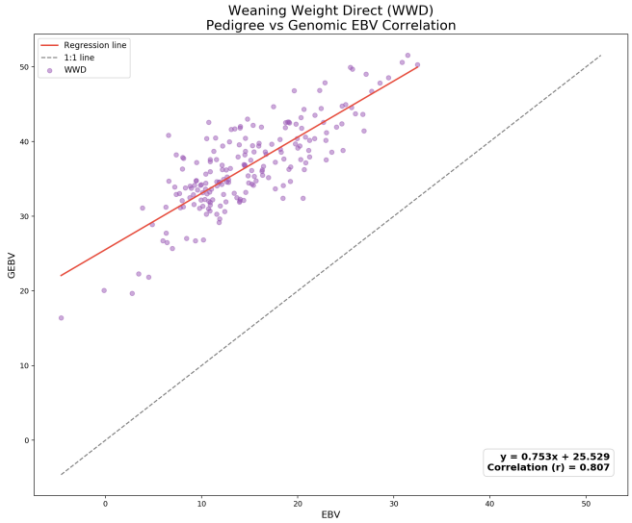


Figure 3. Pearson correlation for weaning weight direct

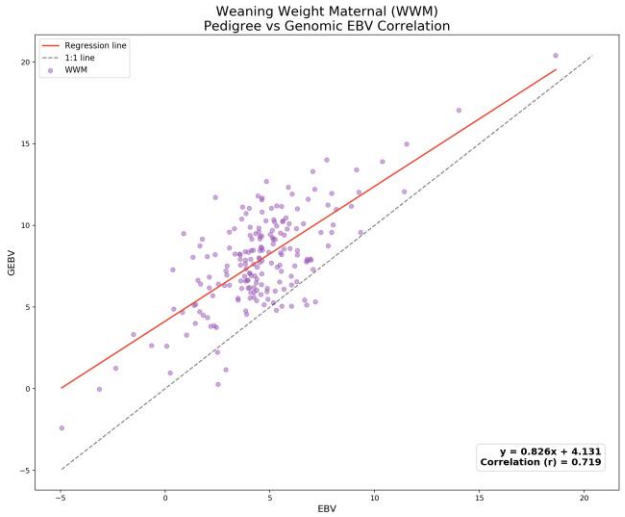


Figure 4. Pearson correlation for weaning weight maternal

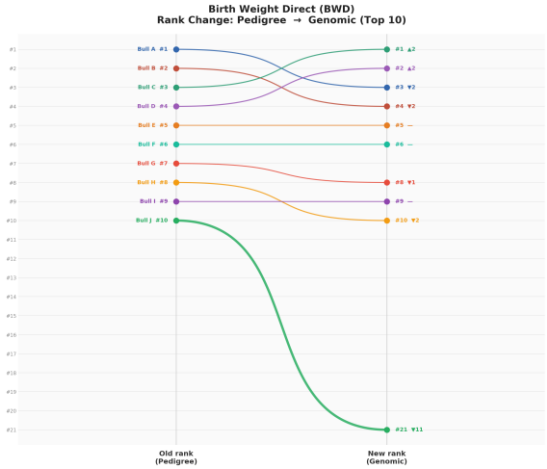


Figure 5. Rank shuffle for top 10 animals by old rank for birth weight

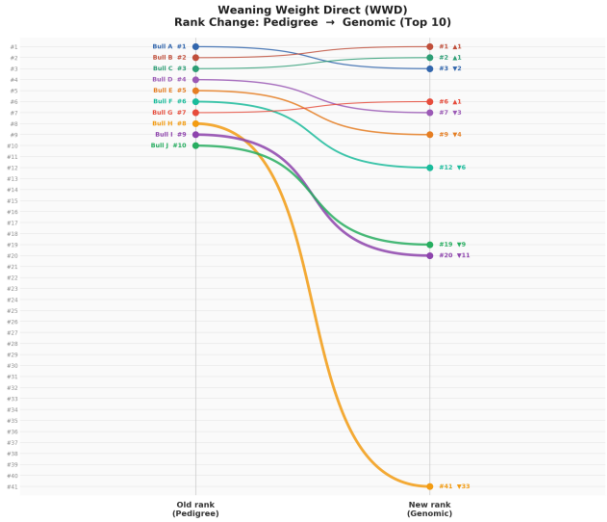


Figure 6. Rank shuffle for top 10 animals by old rank for weaning weight direct

Pearson correlations between pedigree and genomic EBVs were high and significant for all traits (BWD: r = 0.928; WWD: r = 0.807; WWM: r = 0.719; all p<0.0001), indicating broad consistency between the two evaluation systems. The lower correlation for WWM compared with the direct growth traits is consistent with greater dependence of maternal EBV prediction on the availability and structure of maternal information.

The objective of this study was to determine how the rankings of imported Angus sires would change with the inclusion of genomic information. Despite the high EBV correlations, substantial re-ranking of sires occurred across all traits following the transition to ssGBLUP. Mean absolute rank movement was 20.8 positions for BWD (Figure 5), 29.7 for WWD (Figure 6), and 39.7 for WWM (Figure 7), with maximum movements of 91, 139, and 147 positions respectively. Approximately half of all animals declined in rank for BWD (50.8%) and WWD (50.3%), while 51.3% improved in rank for WWM, suggesting a slight directional shift in maternal rankings under the genomic evaluation.

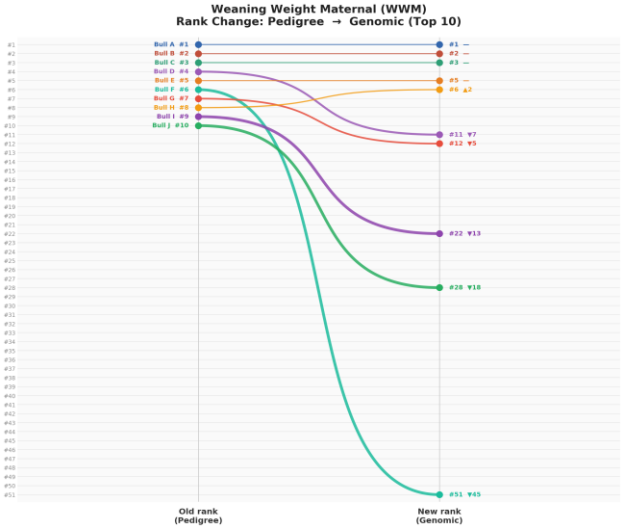


Figure 7. Rank shuffle for top 10 animals by old rank for weaning weight maternal

Movement of the 194 imported bulls within the entire population of genotyped males showed varied results across traits. For BWD, the group overall remained in the bottom two-thirds of the population, although three additional bulls entered the top 25% under the genomic evaluation, indicating some upward movement while confirming that these imported bulls generally transmit lighter birth weights. The most notable improvement was observed for WWD, where the group shifted from the bottom half to the top half of all genotyped males in the population. Bulls in the top 10% more than doubled in number, as did those in the top 25%, with 159 out of 194 bulls improving their rank for WWD under the genomic run. For WWM, the group moved closer to the population median, with 140 out of 194 bulls improving their rank, suggesting a broad upward re-evaluation of maternal genetic merit when genomic information was included.

These results indicate that while the overall genetic landscape was broadly preserved between evaluations, individual sire rankings were considerably affected, with practical implications for sire selection decisions made based on pedigree BLUP rankings alone.

A significant negative relationship between number of progeny and absolute rank movement was observed for BWD (Spearman p=-0.449, p<0.0001) and WWD (p=-0.296, p<0.0001), as well as for the average movement across traits (p=-0.313, p<0.0001). Animals with no recorded progeny showed the largest mean absolute rank movement across traits (35.7 positions on average), nearly double that of animals with 51 or more progeny (19.5 positions) as seen in Table 2.

Table 2: Average change in rank according to the number of progeny with performance records

Progeny group	Number in group	Average rank movement (positions)
0	67	35.7
1 – 5	32	32.8
6 – 20	44	29.8
21 - 50	28	22.2
51 +	22	19.5

This pattern is consistent with the theoretical expectation of ssGBLUP, where the genomic relationship matrix effectively provides additional information for animals with limited phenotypic data in young or unproven animals by capturing realised allele sharing that the pedigree cannot account for (Misztal et al., 2013). In the pedigree-only evaluation, EBVs for animals without progeny are derived almost entirely from parent averages and pedigree relationships, which carry considerably less information than direct performance records. The inclusion of genomic data therefore disproportionately updates EBVs, and consequently rankings, for these animals. The majority of the ten largest rank movers across traits carried two or fewer progeny, further supporting this interpretation and highlighting the particular susceptibility of imported bulls with limited local performance records to re-ranking under genomic evaluation.

In contrast to the direct growth traits, WWM showed no significant relationship between progeny count and rank movement (ρ=-0.051, p=0.482). Mean absolute rank movement for WWM remained consistently high across all progeny groups, ranging from 37.3 positions in the zero-progeny group to 32.0 positions in animals with 51 or more progeny, a far smaller decline than observed for BWD, where movement dropped from 29.3 to 4.2 positions

across the same progeny range. This suggests that an increase in direct progeny records does not stabilise maternal EBVs the way it does for direct growth traits.

This trait-specific pattern reflects the biological complexity of estimating maternal ability in bulls. Maternal EBVs rely heavily on female relatives' performance records, such as those from dams, grandams, and sisters with calves recorded in the local evaluation environment (Mrode, 2014). Following first importation, many of the relevant records of their female relatives reside in foreign databases that the local evaluation does not have access to. This means that these sires have limited access to the specific daughter and female-relative information needed for reliable maternal estimation, even when their number of direct progeny increases. This effect is further compounded by the composition of the South African Angus genomic reference population, which comprises 37% females and may therefore provide less information for maternal traits, potentially contributing to the persistently high rank instability observed for WWM.

The large rank movements observed for WWM should therefore be interpreted with caution, as they may reflect instability in the estimation process as much as genuine genetic re-ranking and underscore the need for an expanded and more sex-balanced genomic reference population to improve the reliability of maternal EBV prediction for imported sires in the South African Angus evaluation.

Conclusions

In conclusion, the results show that the inclusion of genomic information via ssGBLUP resulted in a systematic upward shift in EBVs across all three traits, reflecting the capture of realised genetic relationships that pedigree-based BLUP alone was unable to account for. Genomic information had the greatest impact on animals with limited progeny records. The results also showed that maternal EBVs proved uniquely resistant to stabilisation by progeny count, reflecting the biological requirement for locally recorded daughter performance data, and highlighting the need for continued development of the maternal genomic reference population.

Overall, incorporation of genomic information resulted in substantial re-ranking of the imported sires within the South African Angus population, with the magnitude and direction of change varying among traits.

.

Acknowledgements

The authors thank the Angus Cattle Breeders' Society of South Africa for providing the data used in this study.

References

Adekale, D., Z. Liu, R. Evans, T. Pabiou, R. Reents, D. Segelke, and J. Tetens. 2025. The impact of deregressed foreign breeding values on national beef cattle single-step genomic evaluation. *Genet. Sel. Evol.* 57:37. https://doi.org/10.1186/s12711-025-00982-2

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752. https://doi.org/10.3168/jds.2009-2730

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. https://doi.org/10.1186/1297-9686-42-2

Groeneveld, E. 2010. VCE User's Guide and Reference Manual Version 6.0.

Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, S. Tsuruta, I. Misztal, and F. S. Schenkel. 2019. Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle. *J. Dairy Sci.* 102:8175–8183. https://doi.org/10.3168/jds.2018-15819

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. *Livest. Sci.* 166:54–65.

Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, S. Miller, D. Moser, and I. Misztal. 2018. Single-step genomic BLUP for national beef cattle evaluation in US: From initial developments to final implementation. Invited paper 11.495 in *Proc. 11th World Congress on Genetics Applied to Livestock Production*, Auckland, New Zealand.

MiX99 Development Team. 2017. MiX99: A Software Package for Solving Large Mixed Model Equations. Release 17.11. Natural Resources Institute Finland (Luke), Jokioi, Finland.

Misztal, I., S. E. Aggrey, and W. M. Muir. 2013. Experiences with a single-step genome evaluation. *Poult. Sci.* 92:2530–2534. https://doi.org/10.3382/ps.2012-02739

Mrode, R. A. 2014. *Linear Models for the Prediction of Animal Breeding Values*. 3rd ed. CABI, Wallingford, UK.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81:559–575.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478.

Silva Neto, J. B., L. F. M. Mota, S. T. Amorim, E. Peripolli, L. F. Brito, C. U. Magnabosco, and F. Baldi. 2023. Genotype-by-environment interactions for feed efficiency traits in Nellore cattle based on bi-trait reaction norm models. *Genet. Sel. Evol.* 55:93. https://doi.org/10.1186/s12711-023-00867-2

Silva Neto, J. B., L. F. M. Mota, M. Londoño-Gil, P. I. Schmidt, G. R. D. Rodrigues, V. A. Ligori, L. M. Arikawa, C. U. Magnabosco, L. F. Brito, and F. Baldi. 2024. Genotype-by-environment interactions in beef and dairy cattle populations: A review of methodologies and perspectives on research and applications. *Anim. Genet.* 55:871–892. https://doi.org/10.1111/age.13483

Strandén, I. 2014. RelaX2 Program for Pedigree Analysis, User's Guide for Version 1.65.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

Estimation of Genetic Parameters for Lactation Persistency and Productive Maturity in the Italian Brown Swiss Breed

A. Rossoni[1], C. Nicoletti[1] and E. Santus

[1] ANARB (Associazione Nazionale Allevatori Razza Bruna Italiana), Bussolengo (VR), Italy

Corresponding author: attilio.rossoni@anarb.it

Abstract

Modern dairy cattle breeding increasingly emphasizes functional efficiency and lifetime productivity to enhance herd sustainability. The Italian Brown Swiss breed, characterized by relatively flat lactation curves, represents a suitable population for selection on lactation persistency and productive maturity. This study aimed to estimate genetic parameters for protein yield (kg) across four lactation phases using a multi-trait repeatability animal model. Traits were defined as early (0–150 days) and late (150–300 days) production in first and subsequent lactations. After stringent data editing procedures, approximately 500,000 test-day derived records were included in the analysis. Heritability estimates ranged from 0.11 to 0.25 across phases, in agreement with values commonly reported for protein yield in Brown Swiss populations. Genetic correlations between phases were high (0.75–0.88), indicating substantial shared genetic background across lactation stages and parities. Correlations with the official protein index were above 0.90, supporting the consistency of the proposed model with the current evaluation system. Model validation, based on residual diagnostics and sire ranking comparisons, confirmed its ability to discriminate animals according to differences in lactation curve shape. Based on these results, synthetic indices for persistency and productive maturity are under evaluation to provide new selection tools aimed at reducing metabolic stress and improving lifetime efficiency.

Key words: Brown Swiss, Persistency, Maturity, Genetic parameters, Repeatability model.

Introduction

In modern dairy cattle breeding, moving beyond peak yield toward balanced production cycles is fundamental for establishing functional and long-lived herds. Lactation persistency and productive maturity play central roles in three major domains: animal welfare, biological efficiency, and economic sustainability.

Post-peak stability drastically reduces metabolic stress, particularly during early lactation when energy demands outpace dry matter intake (Collard et al., 2000; Pryce et al., 2004).

Flat lactation curves align with the natural genetic propensity of the Brown Swiss breed, optimizing roughage utilization.

Growth across lactations (maturity) maximizes lifetime profitability, lowers replacement costs, and extends herd longevity (Van Arendonk, 1985).

While random regression (RR) models are widely used to generate continuous genetic lactation curves (Schaeffer and Dekkers, 1994; Jamrozik and Schaeffer, 1997), they present significant limitations when applied to highly stratified datasets with small cohort sizes (e.g., Italian mountain herds), leading to statistical instability and biased early genomic evaluations (Vicario and Degano, 2005).

To overcome these challenges, a multi-trait repeatability approach was developed to deliver stable, unbiased genetic proofs tailored to the Italian population structure.

Materials and Methods

The dataset comprised approximately 500,000 test-day protein yield (kg) records from Italian Brown Swiss cows. The productive career was partitioned into a 4-phase matrix defined by parity and stage of lactation:

- F_early: First lactation, initial phase (0–150 Days in Milk, DIM).
- F_late: First lactation, final phase (150–300 DIM).
- S_early: Second and following lactations, initial phase (0–150 DIM).
- S_late: Second and following lactations, final phase (150–300 DIM).

Genetic parameters were estimated using a linear multi-trait repeatability animal model:

$$y = htd + Ye \times L \times nlat \times age \times dim \times prg + pe + a + e$$

where:

- htd: Herd-Test-Day fixed effect.
- Ye: Year fixed effect (grouped by quinquennium).
- L: Herd production level effect (representing the gap between primiparous and multiparous cows).
- $nlat$: Lactation number.
- age: Calving age effect.
- dim: Days in milk effect.
- prg: Days pregnant effect.
- pe: Permanent environmental random effect.
- a: Additive genetic random effect.
- e: Random residual error.

To condense complex multi-trait genetic evaluations into practical selection tools for breeders, synthetic indices and evaluation traits were formulated to capture lactation curve shape (persistency) and career trajectory (productive maturity) in descriptive functional terms:

- Persistency Ratio (Per÷): Evaluates persistency as the average proportional ratio of protein yield produced in the late phase (150–300 DIM) relative to the early phase (0–150 DIM), averaged across first and following lactations. Higher values identify cows that retain a higher proportion of peak production into late lactation.
- Persistency Percentage (Per%): Expresses early-lactation protein production (0–150 DIM) as a percentage of total 300-day combined output across both early and late phases in all lactations. Lower percentage values correspond to flatter, more persistent lactation curves.
- Persistency Difference (Per-): Quantifies persistency as the average absolute physical difference (in kilograms of protein) between late-phase and early-phase yields across lactations.
- Maturity Ratio Index (Mat÷): Measures productive maturity as the average proportional ratio of first-lactation yield relative to subsequent lactation yields, evaluated across both early and late lactation stages
- Maturity Percentage (Mat%): Expresses total first-lactation protein yield as a percentage of total combined output produced across lactations.
- Maturity Difference (Mat-): Quantifies maturity as the average difference (in kilograms of protein) between first-lactation production and following lactation production.

Results & Discussion

Heritability (h^2) and repeatability (R) estimates across the four lactation phases are presented in Table 1.

Table 1: Estimates of heritability (h^2) and repeatability (R) for protein yield across lactation phases

Lactation	Days in Milk	h^2	R
1	0–150	0.107	0.264
1	150–300	0.253	0.611
>1	0–150	0.149	0.360
>1	150–300	0.198	0.465

Across all phases, heritability estimates ranged from 0.11 to 0.25. Peak heritability occurred during the late phase of first lactation (25.3%). While subsequent lactations exhibited a similar pattern (14.9% early vs. 19.8% late), the differential between early and late phases was substantially more pronounced in first lactation. These estimates are consistent with those of the official evaluation performed by ANARB (0.18) (Interbull Centre, 2026).

Genetic correlations between the four lactation phases are structured as a symmetric correlation matrix in Table 2.

Table 2. Genetic correlation matrix between protein yield traits in the four lactation phases.

Trait	F_late	S_early	S_late
F_early	0.85	0.83	0.75
F_late		0.80	0.86
S_early		—	0.88

Genetic correlations between phases were high overall, ranging from 0.75 to 0.88, confirming substantial shared genetic background and high model stability. Crucially, the lowest correlation observed (0.75 between F_early and S_late) demonstrates that the initial phase of the first lactation possesses a degree of genetic independence from late-career maturity, validating the necessity of a 4-phase matrix architecture.

Correlations between sire estimated breeding values (EBVs) for total protein yield (P.Kg.), individual phase EBVs, and the synthetic persistency and maturity indices are reported in Table 3.

Table 3. Sire EBV correlations between synthetic persistency/maturity indices, total protein yield (P.Kg.), and individual phase traits.

Index	P.Kg.	F_early	F_late	S_early	S_late
Pers÷	0.18	0.02	0.20	0.07	0.25
Pers%	0.08	0.24	0.26	0.13	0.30
Pers−	0.49	0.63	0.67	0.54	0.72
Matu÷	-0.30	-0.22	-0.26	-0.32	-0.34
Matu%	-0.25	-0.38	-0.30	-0.41	-0.44
Matu−	-0.83	-0.75	-0.81	-0.88	-0.90

Daughter average lactation curves for the top and bottom 10% sires confirmed that the persistency indices effectively discriminate animals based on post-peak curve slope.

However, productive maturity indices showed strong negative genetic correlations with overall protein yield (-0.25 to -0.83). This marked antagonism indicates that direct selection for higher maturity under current index formulations would result in a substantial penalty on overall protein production, restricting its direct application in selection programs.

Conclusions

The multi-trait repeatability model developed in this study provides a robust and statistically stable methodology for evaluating lactation persistency and productive maturity in the Italian Brown Swiss breed.

High genetic correlations across the four lactation phases (0.75 to 0.88) confirm the structural consistency of the evaluation system, while the relative genetic independence observed during the initial phase of the first lactation underscores that distinct career stages are governed by partially independent biological drivers.

The synthetic persistency indices successfully discriminate animals according to lactation curve geometry, providing practical tools to select cows capable of sustained post-peak output and reduced metabolic stress. Conversely, the marked genetic antagonism observed between productive maturity indices and overall protein yield restricts the immediate, direct application of current maturity formulas in commercial selection programs.

To overcome these constraints, ongoing research is focused on evaluating direct phenotypic ratios and simple yield differences between phases, analyzing individual residuals from single-trait evaluation models to capture curve deviations, and identifying the most stable and heritable alternative traits for

integration into the national single-step genomic evaluation pipeline.

References

Collard, B. L., Boettcher, P. J., Dekkers, J. C. M., Petitclerc, D., and Schaeffer, L. R. (2000). Relationships between energy balance and health traits in dairy cattle in early lactation. *J. Dairy Sci.*, 83(11), 2683-2690.

Interbull Centre. 2026. PREPdb - Performance Recording, Evaluation and Publication database [Online]. Interbull Centre-Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden. Available: https://prep.interbull.org/ [Accessed date 2026/07/23].

Jamrozik, J., and Schaeffer, L. R. (1997). Estimates of genetic parameters for a test day model with random regressions for yield traits of Holsteins. *J. Dairy Sci.*, 80(4), 762-770.

Pryce, J. E., Royal, M. D., Garnsworthy, P. C., and Mao, I. L. (2004). Fertility in the high-producing dairy cow. *Livest. Prod. Sci.*, 83(2), 115-129.

Schaeffer, L. R., and Dekkers, J. C. M. (1994). Random regression models for test day production of dairy cows. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Guelph, Ontario, Canada. 18, 443-446.

Van Arendonk, J. A. M. (1985). Studies on the replacement policy in dairy cattle. *Livest. Prod. Sci.*, 13(2), 101-121.

Vicario D., and Degano L. 2005. Test-Day Model for National Genetic Evaluation of Somatic Cell Count in Italian Simmental Population *Interbull Bulletin* 33, 171-175.

Refinement of U.S. Reproductive Performance Evaluations for the Modern Dairy Herd

T.M. McWhorter[1], J.R. Graham[2] E. Nicolazzi[1], P.M. VanRaden[3], A. Ling[1], S. Savoia[1], and A. Legarra[1]

[1] Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

[2] Holstein Association USA, Inc., Brattleboro, VT 05302-0808, USA

[3] Animal Genomics and Improvement Laboratory, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350, USA

Corresponding author: taylor.mcwhorter@uscdcb.com

Abstract

United States (U.S.) national genetic evaluations for reproductive performance include Daughter Pregnancy Rate (DPR), Cow Conception Rate (CCR), Heifer Conception Rate (HCR), and Early First Calving (EFC). Since their introduction, 2003, 2009, 2009, and 2019, respectively, changes in reproductive management (e.g., variable voluntary waiting periods, sexed semen, synchronization protocols, and beef-on-dairy usage) have altered the biological and data structure underlying these traits. In addition, undesirable run-to-run behavior was observed in recent years, including seasonal fluctuations and directional shifts in young bull predicted transmitting ability (PTA). A comprehensive review of the phenotype pipeline and evaluation model was conducted. Key updates implemented for the April 2026 evaluation include: (1) herd-year and lactation-specific voluntary waiting period adjustment in DPR; (2) introduction of a new trait, First Service to Conception (FSC), defined as days from first insemination to confirmed pregnancy (capped at 200 d with penalty for non-conception); (3) revised CCR and HCR pre-adjustments accounting for service sire breed, mating type, and short-cycle inseminations; (4) stricter daughter age-related inclusion rules to reduce outcome-dependent bias in young bulls; (5) updated contemporary group edits; (6) new variance component estimation; and (7) stricter convergence criteria. DPR, CCR, HCR, and FSC are evaluated in a multi-trait model, while EFC is analyzed separately. Genetic correlations among DPR, CCR, and FSC exceed 0.94, with moderate correlation to HCR. Across seven consecutive triannual runs from December 2023 to December 2025, the revised system demonstrated improved genetic trend behavior, elimination of systematic downward drift in young bull PTA, and more symmetric run-to-run changes. Traditional PTA distributions by birth cohort better reflect expected selection response. Reranking is trait-dependent and most pronounced for CCR and DPR in low-reliability animals, whereas rankings among young high-reliability bulls (REL ≥95) remain highly consistent (>0.94 Spearman). Interbull validation requirements were met. The revised evaluation provides improved biological alignment with modern reproductive management while maintaining appropriate stability and interpretability for international comparison.

Key words: estimated breeding values, female fertility, management changes, reproduction, trait stability, voluntary waiting period

Introduction

Reproductive performance is an important component of dairy cattle profitability and herd sustainability. In the United States (U.S.), genetic selection for female fertility is supported by national evaluations for Daughter Pregnancy Rate (DPR), Cow Conception Rate (CCR), Heifer Conception Rate (HCR), and Early First Calving (EFC). These evaluations are provided by the Council on Dairy Cattle Breeding (CDCB; Bowie, MD, USA) and

describe related but distinct aspects of reproductive performance. DPR reflects the rate at which nonpregnant lactating cows become pregnant and was introduced as a national genetic evaluation in 2003 (VanRaden et al., 2004). CCR and HCR measure conception success following insemination in cows and heifers, respectively; development of heifer fertility evaluations was described by Kuhn et al. (2006). The reproductive performance traits have subsequently been incorporated into multi-trait national evaluations to utilize information among correlated traits (VanRaden et al., 2014). Most recently, EFC was implemented in 2019 (Hutchison et al., 2017) and measures genetic differences in age at first calving.

Management practices affecting reproduction have continued to change since these traits were implemented. Greater use of sexed semen, synchronization programs, beef semen in dairy herds, and herd-specific reproductive strategies has altered both insemination patterns and the structure of the data used in national evaluations. Changes in reproductive technologies, including the rapid increase in embryo transfer, have previously required modifications to U.S. fertility evaluations (Miles et al., 2023). Voluntary waiting periods also vary among herds and across lactations, which is particularly relevant for DPR because its phenotype is derived from days open, the interval between calving and conception. These changes raised questions about whether existing phenotype definitions, adjustments, and data edits continued to represent reproductive performance appropriately in modern dairy herds.

At the same time, routine evaluation monitoring identified unexpected behavior in some reproductive performance traits. Genetic trends showed seasonal fluctuations and predicted transmitting abilities for some groups of young bulls exhibited directional changes between consecutive evaluations. These patterns and the initial investigation of the U.S. female fertility evaluations were described by McWhorter et al. (2025). Although PTA changes are expected as additional daughter information becomes available, changes should generally occur in both directions rather than display a persistent systematic shift. These observations motivated a broader review of the reproductive performance evaluation system.

The objective of this work was to refine the U.S. reproductive performance evaluations to better reflect current management practices and improve run-to-run behavior. This paper describes the revisions implemented in the national evaluation and summarizes their effects on genetic trends, PTA stability, and ranking consistency.

Materials and Methods

National reproductive records used for U.S. genetic evaluations were obtained from the National Cooperator Database managed by CDCB. Phenotypes included records for lactating cows and heifers. The evaluation analyzed DPR, CCR, HCR, FSC, and EFC for Ayrshire, Brown Swiss, Guernsey, Holstein, Jersey, and Milking Shorthorn. Tests were performed using the routine U.S. evaluation system over time. National evaluation data was tested across seven consecutive triannual runs from December 2023 through December 2025, including every April, August, and December evaluation in between. The legacy system refers to the evaluations before any updates were applied and the revised system refers to the evaluations after updates and the new trait, FSC, were applied. Record counts for the December 2025 evaluation are shown in Table 1.

Trait Definitions

DPR is a non-linear transformation from days open to pregnancy rate and predicts the proportion of nonpregnant cows expected to conceive during a 21-day estrous cycle (VanRaden et al., 2004). CCR and HCR evaluate conception success following insemination in lactating cows and maiden heifers, respectively. They can be understood as

Table 1: The number of phenotypes and cows/heifers available in the December 2025 evaluation run after all updates have been applied.

Trait*	Records	Cows/Heifers
DPR	93,692,573	37,064,183
CCR	45,286,331	19,071,905
HCR	14,136,039	14,136,039
FSC	45,286,331	19,071,905
EFC	34,634,436	34,634,436

*DPR=daughter pregnancy rate; CCR=cow conception rate; HCR=heifer conception rate; FSC=first service to conception; EFC=early first calving

the sum of adjusted inseminations (0=failed, 1=successful with an adjustment), divided by the total number of attempts, up to seven inseminations. Adjusted inseminations refer to accounting for effects on individual inseminations prior to combining into a single conception rate per lactation. EFC is the number of days between an animal's first calving and her birth. FSC was introduced to evaluate the interval between first insemination and conception and complements the existing fertility traits by characterizing reproductive performance after the first breeding event. In the phenotype calculation for FSC, conception is the date of the successful insemination confirmed through pregnancy diagnosis or subsequent calving.

Evaluation Updates

Several revisions were implemented across the reproductive performance traits evaluation system. DPR was updated to account for herd-year and lactation-specific voluntary waiting periods. Pre-adjustments for CCR and HCR were revised to account for service sire breed, mating type, and short-cycle inseminations. Data edits and inclusion criteria were also updated, including stricter daughter age requirements and contemporary group edits. Additional changes included updated variance components, stricter convergence criteria, and the introduction of FSC.

DPR, CCR, HCR, and FSC were evaluated jointly in a multi-trait animal model. In the legacy system, EFC was included in the multi-trait analysis with DPR, CCR and HCR, it was treated as uncorrelated. Therefore, in the revised system, EFC was evaluated separately using a single-trait animal model for computational efficiency. Updated variance components were estimated before implementation.

Results & Discussion

The revisions described in this study represent a comprehensive update to the U.S. reproductive performance evaluations following the initial review of evaluation behavior described by McWhorter et al. (2025). The revised reproductive performance evaluations produced more consistent behavior across routine national evaluations while maintaining biologically plausible genetic trends.

Genetic trends (mean PTA by birth year) in the legacy system demonstrated patterns more consistent with expectations (Figures 1 and 2). The legacy system showed downward shifts over time for DPR and CCR. In the revised system, the trends slowly started increasing from the onset of implementation of their genetic evaluations. Therefore, the revised system genetic trends more closely reflected the expected response to ongoing selection.

Run-to-run genetic trends were also more consistent under the revised evaluation system (Figure 3). Whereas the legacy evaluation occasionally produced shifts in average PTA for younger bulls between consecutive evaluations, the revised system resulted in a more consistent trend from one run to the next.

Implementation of the revised evaluation resulted in reranking of animals, with the greatest changes occurring among animals with lower reliability. However, rankings among young bulls with high reliability remained highly consistent across evaluation systems. Interbull validation requirements were satisfied for the revised evaluations prior to national implementation.

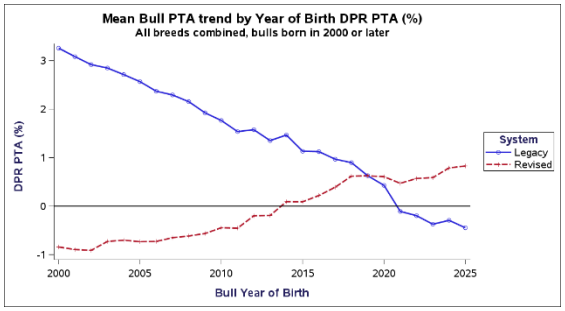


Figure 1. Daughter pregnancy rate (DPR) genetic trend for all bulls born 2000 or after for the legacy system (blue) and the revised system (red).

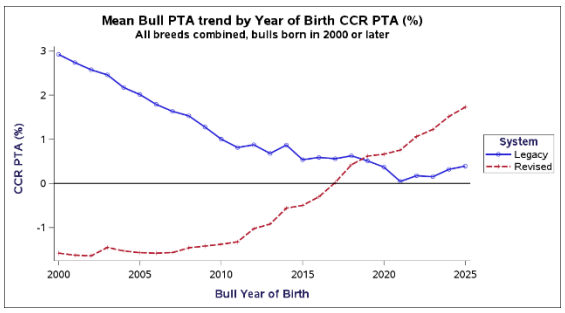


Figure 2. Cow conception rate (CCR) genetic trend for all bulls born 2000 or after for the legacy system (blue) and the revised system (red).

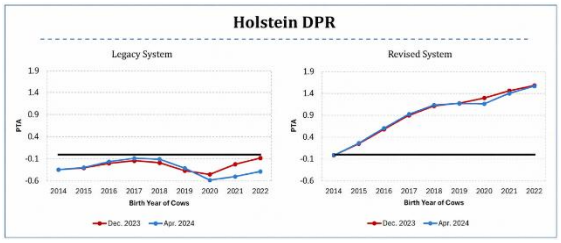


Figure 3. Daughter pregnancy rate (DPR) average sire PTA of cows with phenotypes by birth year of cows for Holstein in the legacy system (left) and the revised system (right) when comparing a previous run, December 2023 (red) to a subsequent run, April 2024 (blue).

The revised evaluations produced genetic trends and run-to-run PTA behavior that were more consistent with biological expectations while maintaining stability among highly reliable animals. Although some reranking occurred following implementation, the limited changes observed for high-reliability bulls indicate that the revisions primarily affected animals for which additional information or revised modeling had the greatest impact. These results suggest that the revised evaluation more appropriately reflects both current management practices and the accumulation of new performance data.

The addition of FSC further expands the U.S. reproductive performance evaluation by providing a measure of reproductive performance beginning at first insemination, regardless of voluntary waiting period.

Conclusions

The U.S. reproductive performance evaluations were revised to better reflect current dairy management and to address undesirable behaviors from one evaluation to the next. The revised system incorporated updated phenotype processing, editing, modeling procedures, and the new FSC trait. Across consecutive evaluations, the revisions produced more biologically plausible trends, more balanced PTA changes in young bulls, and high-ranking consistency among animals with high reliability. The revised evaluations met Interbull validation requirements and were implemented for the August 2026 evaluation.

Acknowledgments

The authors would like to thank U.S. Dairy Producers, member sectors and collaborators, as well as the USDA Animal Genomics and Improvement Laboratory (AGIL) for contributions to this project. The support and input from an internal CDCB team were essential throughout the review process.

The mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by CDCB. CDCB is an equal opportunity provider and employer.

References

Hutchison, J.L., P.M. VanRaden, D.J. Null, J.B. Cole, and D.M. Bickhart. 2017. Genomic evaluation of age at first calving. *J Dairy Sci.*

100(8), P6853-6861. https://doi.org/10.3168/jds.2016-12060

Kuhn, M.T., Hutchison, J.L., and Wiggans, G.R. 2006. Characterization of Holstein Heifer Fertility in the United States. *J Dairy Sci*. 89 (12), P4907-4920. https://doi.org/10.3168/jds.S0022-0302(06)72541-3.

Miles, A.M., Hutchison, J.L., and VanRaden, P.M. 2023. Improving national fertility evaluations by accounting for the rapid rise of embryo transfer in US dairy cattle. *J Dairy Sci*. 106 (7), P4836-4846. https://doi.org/10.3168/jds.2022-22298.

McWhorter, T.M., J. Graham, E. Nicolazzi, P. VanRaden, G. Dutta, & A. Legarra. 2025. Foundational Review of US Female Fertility Trait Evaluations. *Interbull Bulletin vol. 61*. https://journal.interbull.org/index.php/ib/article/view/1971

VanRaden, P.M., Sanders, A.H., Tooker, M.E., Miller, R.H., Norman, H.D., Kuhn, M.T., and Wiggans, G.R. 2004. Development of a National Genetic Evaluation for Cow Fertility. *J Dairy Sci*. 87 (7), P2285-2292. https://doi.org/10.3168/jds.S0022-0302(04)70049-1.

VanRaden, P.M., Tooker, M.E., Wright, J.R., Sun, C., and Hutchison, J.L. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J Dairy Sci*. 97 (12), P7952-7962. https://doi.org/10.3168/jds.2014-8489.

2025 National Breeding Objective review for Australian dairy cattle

T.T.T. Nguyen[1], P. Thurn[1], L. Monks[1], G. Nieuwhof[1], P. Williams[1], M. Axford[1,2], M. Shaffer[1,2]

[1] DataGene Ltd., AgriBio, 5 Ring Road, Bundoora, VIC 3083, Australia

[2] Current address: Dairy Australia, HWT Tower, Southbank, VIC 3006, Australia

Corresponding author: GNieuwhof@datagene.com.au

Abstract

DataGene reviews the Australian National Breeding Objective (NBO) for dairy cattle approximately every 5 years to ensure national genetic selection remains aligned with farmer priorities, economic signals and industry needs. The 2025 review combined stakeholder consultation, economic and genetic review, and scenario testing of national selection indices. Stakeholder consultation included an online survey of 217 dairy farmers and 38 herd improvement (HI) industry personnel, 20 qualitative in-depth interviews with dairy farmers, and discussion of findings at 28 stakeholder events involving 138 participants. The Balanced Performance Index (BPI) remained the most widely used and influential index across regions, breeds and production systems. Consultation supported retaining a simple national index framework while recognising demand for a seasonal pasture-based focus additional emphasis on calving traits in the Health Weighted Index (HWI). Economic updates used forecast milk and feed pricing, including a revised farmgate protein-to-butterfat price ratio that increased the relative value of fat. Genetic parameters were updated, and more than 30 index scenarios were tested using BPI24 as a baseline and August 2024 public-release Australian Breeding Values (ABVs) for Holstein, Jersey and Red breed bulls. Updated BPI, HWI and Sustainability Index (SI) options were recommended and implemented in December 2025. Correlations between 2024 and 2025 BPI values were high across breeds (0.980 to 0.982), although some reranking occurred due mainly to the revised fat and protein emphasis. The updated BPI was predicted to deliver faster genetic gain in fat and slower gain in protein than BPI24, while retaining balanced improvement in profitability, health and fertility traits. HWI25 increased expected responses for survival and fertility and incorporated calving ease and gestation length. SI25 maintained stronger emphasis on production traits to reduce emissions intensity through improved productivity.

Key words: dairy cattle; selection index; breeding objective; stakeholder consultation; genetic evaluation; sustainability

Introduction

The Australian National Breeding Objective (NBO) defines breeding priorities for dairy cattle to improve long-term profitability and ensure genetic selection remains aligned with farmer and industry needs. To remain relevant, the NBO must respond to changes in production systems, market signals, farmer preferences, breeding technologies and trait information.

DataGene reviews the NBO and associated national selection indices approximately every 5 years, with additional reviews undertaken when required. The NBO is delivered through indices that combine Australian Breeding Values (ABVs) into ranking tools for bulls, cows and herds. The 2014 review led to three national indices in 2015: the Balanced Performance Index (BPI), Health Weighted Index (HWI) and Type Weighted Index (TWI) (Byrne et al. 2016; Axford et al. 2021). Following the 2020 review, BPI and HWI trait weightings were revised and TWI was discontinued (Axford et al. 2021). DataGene released the Sustainability Index (SI) in 2022 to accelerate reductions in greenhouse-gas

emission intensity (Nguyen et al. 2023) and removed the milk volume penalty later that year to better reflect Australian milk pricing.

The BPI is the primary economic index and is designed to improve net profit through a balance of production, functionality and type. The HWI places greater emphasis on health and fertility, and the SI places greater emphasis on traits that reduce greenhouse-gas emission intensity through improved productivity. These indices are important decision-support tools for farmers, breeding companies and advisers.

The 2025 NBO review combined three linked components: stakeholder consultation, economic and genetic review, and scenario testing of alternative index options. Stakeholder consultation identified farmer and industry priorities and tested the relevance of existing indices and traits. The technical review updated the assumptions that underpin trait economic weights and expected responses to selection. This paper reports the stakeholder findings and final index updates implemented in the December 2025 Australian Breeding Value release.

Materials and Methods

Stakeholder consultation

Stakeholder consultation timing and purposes are presented in Table 1. DataGene commissioned Down To Earth Research (DTER) to implement a mixed-method consultation process. A nationwide online survey was live between late June and early July 2024. Survey invitations were distributed to Dairy Australia levy payers with available email addresses and DataGene Genemail members, and the survey was also promoted through generic links. The survey examined current use of Australian indices, the influence of indices on artificial insemination (AI) sire-selection decisions, attitudes to genetics and breeding, and individual trait preferences.

A total of 296 farmers and industry personnel participated. After partial and poor-quality responses were removed, 255 high-quality responses remained, comprising 217 dairy farmers and 38 herd improvement (HI) industry personnel. Online survey data were analysed by DTER using IBM Statistical Package for Social Sciences (SPSS) and Q research software.

Qualitative in-depth interviews were used to explore themes that emerged from the survey. The first 10 interviews targeted dairy farmers representing Total Mixed Ration (TMR), Jersey, pasture-based, Holstein and subtropical farm segments. A further 10 interviews were added to strengthen understanding of TMR farmer views, because survey participation from TMR farms was low. Interviews were conducted in August 2024, averaged approximately 20 minutes, and used a discussion guide rather than a structured questionnaire to encourage open feedback. Results from the survey and interviews were then presented and discussed at 28 stakeholder events involving 138 participants, including bull companies, resellers, breed organisations, farmers, industry representatives and researchers.

Table 1: Stakeholder consultation activities used in the 2025 National Breeding Objective review

Component	Timing	Sample	Purpose
Online survey	Late June to early July 2024	255 high-quality responses: 217 dairy farmers and 38 HI personnel	Assess use of Australian indices and ABVs, influence on breeding decisions, attitudes to genetics and trait preferences.
In-depth interviews	August 2024	20 dairy farmer interviews	Explore emerging themes for production-system, regional and breed-specific needs, with additional interviews for TMR farms.
Stakeholder meetings	August to October 2024	28 events with 138 participants	Test survey and interview findings with bull companies, resellers, breed organisations, farmers, industry and research stakeholders.

Economic and genetic review

Updated economic and genetic parameters were applied to the bio-economic model used to calculate trait weightings. Economic weights for production traits were derived from evidence-based forecast prices rather than historical prices, reflecting the approximate 3-year lag between index implementation and the first calving of animals bred using updated information. DataGene commissioned EverAg to provide a 5-year historical analysis and 10-year forecast of milk price relative to feed cost and the farmgate protein-to-butterfat price ratio. An intermediate 5-year average milk price scenario, based on projected prices from 2025 to 2029, was selected.

The assumed milk solids price increased from $6.18/kg, based on 2020 prices, to $8.43/kg. The updated protein-to-butterfat price ratio increased the relative emphasis on fat compared with previous indices. Feed costs were updated using Dairy Australia's Dairy Farm Monitor Project data (Dairy Australia, 2024) while maintaining a constant feed-cost-to-milk-income ratio. Genetic parameters were reviewed by Agriculture Science and Technology, Agriculture Victoria. The Survival ABV was replaced with a new Survival trait, also named Survival, defined as an animal's ability to produce daughters that last in the herd for many lactations.

Index modelling and scenario testing

Stakeholder consultation indicated that traits included in the 2024 BPI remained relevant for the updated indices. Consultation also identified a need to include calving ease and gestation length in the HWI to improve its relevance to seasonal calving systems. BPI24 was used as the baseline for developing and testing updated index options. AbacusBio modelled the indices using ABVs from the August 2024 public release for 28,152 Holstein, 1,574 Jersey and 182 Red breed bulls. More than 30 alternative scenarios were tested to address consultation themes and stakeholder feedback. Table 2 summarises testing scenarios for indices that were updated and published in December 2025.

Results & Discussion

Stakeholder use of indices and ABVs

The survey specifically targeted farmers using AI. As a result, respondents may have had stronger views on herd genetics than the broader dairy farming population. Nevertheless, the consultation captured diverse farm types. Among dairy farmer respondents, 54% had at least some registered cattle and 46% were solely commercial dairy farms. Holstein was the most common main breed, followed by Jersey, crossbred and Red breed cattle.

Most respondents used at least one Australian index or breeding value when selecting AI sires. This included 88% of dairy farmers and 92% of HI industry personnel. BPI was the most commonly used index, followed by HWI and SI, and remained influential across regions, breeds and feeding systems. Among dairy farmers, 43% reported that BPI had a large influence on AI sire selection and a further 36% reported that it had some influence. Among HI personnel, 61% reported that BPI had a large influence and 29% reported that it had some influence. HWI had greater influence among respondents with commercial dairy farms than among those with registered cattle.

Table 2: Recommended 2025 national selection index updates

Index	Recommended 2025 update
BPI25	Updated BPI for non-Jersey breeds using projected milk values for 2024-25 to 2028-29 and revised fat, protein, feed, labour and other economic parameters.
BPIJe25	Updated BPI for Jersey using the same economic assumptions as BPI25, with Feed Saved excluded.
HWI25	Updated HWI for seasonal calving herds, using the same economic assumptions as BPI25, with Overall Type set to zero, Milking Speed and Temperament adjusted for shorter lactations, and Calving Ease and Gestation Length included.
SI25	Updated SI using projected milk values for 2024-25 to 2028-29 and revised economic assumptions, retaining emphasis on reduced emissions intensity through productivity.

Consultation provided evidence of demand for a seasonal pasture-based index among farmers operating pasture-based systems. Views on a TMR-specific index were more polarised. Some interview respondents saw value in distinguishing between housed and pasture-based production systems, whereas others considered that well-fed cows should perform across systems. Jersey breeders showed strong reliance on existing Australian indices, and a Jersey-specific index was generally viewed as a useful addition rather than an essential requirement. There was also some interest in indices tailored to warmer regions and TMR systems; however, the level of demand was not considered sufficient to warrant publication of these indices. Nevertheless, these options were included in the development and testing process to ensure that stakeholder feedback was appropriately evaluated, but not reported here.

Mammary system, daughter fertility and production were rated among the most important individual traits when selecting AI sires. Farmers in seasonal calving or pasture-based systems placed greater priority on fertility, and subtropical farmers placed greater emphasis on heat tolerance. Stakeholder meetings reinforced four practical communication themes: production remains central to farm income; simplicity is preferred; Feed Saved ABV is important but not well understood. These findings informed the final index recommendations and highlighted areas where extension support remains important.

Updated economic weights

The main economic changes were the higher milk solids price, the revised protein-to-butterfat price ratio and the corresponding increase in the economic weight for fat (Table 3). BPI25 was calculated for non-Jersey breeds, BPIJe25 for Jersey, and HWI25 and SI25 for all breeds. Compared with the 2024 indices, the updated indices placed greater emphasis on fat and survival. HWI25 also introduced non-zero weights for calving ease and gestation length and set Overall Type to zero to align more closely with the intended seasonal-calving health and fertility focus

Correlations between indices

Correlations between 2024 and 2025 BPI values were high for Holstein, Jersey and Red breed bulls (0.980 to 0.982; Table 4). Some reranking is expected, mainly because the revised protein-to-butterfat price ratio favours bulls with superior fat ABVs. Correlations between BPI, HWI and SI were lower, reflecting their different emphasis on production, health, fertility and emissions intensity.

Table 3: Economic weights for traits in 2024 and 2025 national selection indices

Trait	BPI24	BPI25	BPIJe24	BPIJe25	HWI24	HWI25	SI24	SI25
Protein	6.76	7.95	6.76	7.95	4.36	5.12	17.49	13.17
Fat	2.08	5.67	2.08	5.67	1.35	3.68	2.82	7.61
Milk	-0.08	-0.09	-0.08	-0.09	-0.06	-0.05	-0.08	-0.09
Survival	7.20	14.29	7.20	14.29	7.20	14.29	20.21	23.22
Fertility	6.94	14.02	6.94	14.02	14.11	22.86	6.94	8.25
Somatic cell count	0.69	1.26	0.69	1.26	0.69	1.26	0.69	1.26
Mastitis resistance	6.75	10.89	6.75	10.89	6.75	10.89	8.70	12.61
Milking speed	5.02	6.52	5.02	6.52	5.02	6.46	5.02	6.52
Temperament	3.60	0.97	3.60	0.97	3.60	0.97	3.60	0.97
Mammary system	2.76	3.59	2.76	3.59	3.59	4.66	2.79	3.59
Udder depth	0.82	1.08	0.82	1.08	0	0	0.82	1.08
Overall type	1.36	2.70	1.36	2.70	1.36	0	1.36	2.70
Pin set	0.78	1.55	0.78	1.55	0.78	1.55	0.78	1.55
Feed saved	0.19	0.21	0	0	0.39	0.42	0.72	0.74
Calving ease	0	0	0	0	0	0.90	0	0
Gestation length	0	0	0	0	0	-3.14	0	0

Table 4: Selected correlations among 2024 and 2025 index values by breed

Breed	BPI24 with 2025 BPI	2025 BPI with HWI25	2025 BPI with SI25	HWI25 with SI25
Holstein	0.982	0.929	0.970	0.847
Jersey	0.981	0.923	0.980	0.804
Red breeds	0.980	0.842	0.957	0.715

Predicted response to selection

Predicted responses for a 1 standard deviation increase in index value are shown in Table 5. Across all breeds, BPI25 was predicted to deliver faster genetic gain in fat and slower gain in protein than BPI24. This again reflected the revised protein-to-butterfat price ratio and increased fat emphasis. In Jersey and Red breeds, BPI25 was also predicted to deliver greater improvement in survival, somatic cell count and mastitis resistance.

Compared with HWI24, HWI25 was predicted to achieve higher gains in protein, fat and survival. In Jersey and Red breeds, HWI25 was also predicted to deliver substantially greater gains in daughter fertility. Compared with SI24, SI25 was predicted to deliver a faster reduction in emissions intensity in Holstein cattle, at 8.53% compared with 7.64%. The predicted reduction was lower in Jersey and Red breeds, at 6.08% and 3.29%, respectively, compared with 8.96% and 5.52% under SI24. These differences illustrate the value of testing index options across breeds rather than relying only on an average herd or single breed profile

Implementation considerations

The 2025 review supported retaining a national index framework while refining the indices to better reflect current economic conditions, stakeholder priorities and trait information. BPI25 retained a balanced selection objective for profitability, health, fertility, production and type. BPIJe25 acknowledged Jersey-specific limitations for Feed Saved. HWI25 was broadened to improve relevance for seasonal calving herds through inclusion of calving ease and gestation length. SI25 retained its role as the index most directly aligned with reduced emissions intensity through improved productivity.

The consultation also highlighted that technical improvements need to be accompanied by clear communication. Negative weight on milk volume continues to be poorly understood because it reflects feed and other costs rather than a preference for low production. Similarly, Feed Saved was viewed as important but its reliability and interpretation require extension support. These findings suggest that future NBO reviews should continue to integrate formal stakeholder consultation, economic and genetic analysis, and transparent scenario testing so that index changes are both technically robust and usable in commercial breeding decisions.

Table 5: Predicted response to selection for a 1 standard deviation increase in index value

Trait	Hol BPI25	Hol HWI25	Hol SI25	Jer BPIJe25	Jer HWI25	Jer SI25	Red BPI25	Red HWI25	Red SI25
Protein	0.40	0.22	0.55	3.58	1.37	4.95	0.53	0.28	0.70
Fat	0.63	0.37	0.69	7.79	4.65	8.44	0.64	0.27	0.70
Milk	-0.08	-0.14	0.04	21.07	-19.82	63.99	0.21	0.03	0.39
Survival	0.63	0.70	0.59	0.98	1.13	0.95	0.48	0.46	0.43
Daughter fertility	0.54	0.80	0.37	1.22	2.48	0.57	-0.06	0.41	-0.21
Somatic cell count	0.63	0.62	0.58	10.31	12.29	8.97	0.52	0.61	0.38
Mastitis resistance	0.51	0.56	0.44	1.50	1.97	1.27	0.52	0.61	0.38
Milking speed	0.05	0.02	0.02	0.20	0.03	0.19	0.06	0.01	0.04
Temperament	-0.05	-0.11	-0.04	0.04	-0.26	0.04	0.06	-0.02	0.08
Mammary system	-0.02	-0.15	-0.10	-0.03	-1.04	-0.39	0.24	0.05	0.17
Udder depth	0.20	0.25	0.08	0.42	0.97	-0.05	0.22	0.12	0.13
Overall type	-0.04	-0.20	-0.12	0.18	-1.03	-0.08	0.32	-0.08	0.27
Pin set	-0.06	-0.05	-0.08	0.50	0.06	0.44	-0.01	0.11	0.01
Feed saved	-0.07	0.17	0.01	-4.43	7.67	0.80	-0.27	0.06	-0.20

Note: Hol = Holstein; Jer = Jersey; Red = Red breeds. Responses are expressed in the units of the relevant ABV or index trait

Conclusions

The 2025 NBO review updated the BPI, HWI and SI to reflect stakeholder feedback, forecast milk and feed economics, current genetic information and contemporary breeding priorities. Stakeholder consultation confirmed strong ongoing support for the BPI and identified practical needs for simple, broadly applicable breeding tools, with additional relevance for seasonal pasture-based systems. The largest technical change was the revised protein-to-butterfat price ratio, which increased the relative emphasis on fat while maintaining balanced selection for profitability, health, fertility and sustainability. HWI25 was broadened through inclusion of calving ease and gestation length. The updated indices were implemented in the December 2025 ABV release and provide a contemporary national framework for ranking Australian dairy cattle and supporting genetic improvement for profitable, resilient and sustainable herds.

Acknowledgments

The authors acknowledge the contributions of survey respondents, interviewees and participants in stakeholder meetings. The authors also thank DataGene's Genetic Evaluation Standing Committee, Down To Earth Research, EverAg, AbacusBio, Dairy Australia, Agriculture Science and Technology (Agriculture Victoria), the Dairy Farm Monitor Project tea for input, analysis and technical support.

References

Axford M., Santos B., Stachowicz K., Quinton C., Pryce J.E. and Amer P. 2021. Impact of a multiple-test strategy on breeding index development for the Australian dairy industry. Anim. Prod. Sci. 61, 1940-1950. https://doi.org/10.1071/AN21058.

Byrne T.J., Santos B.F.S., Amer P.R., Martin-Collado D., Pryce J.E. and Axford M. 2016. New breeding objectives and selection indices for the Australian dairy industry. J. Dairy Sci. 99, 8146-8167.

Dairy Australia. 2024. Dairy Farm Monitor Project. https://www.dairyaustralia.com.au/industry-reports/dairy-farm-monitor-project

Nguyen T.T.T., Richardson C.M., Post M., Amer P.R., Nieuwhof G.J., Thurn P. and Shaffer M. 2023. The Sustainability Index: a new tool to breed for reduced greenhouse-gas emissions intensity in Australian dairy cattle. Anim. Prod. Sci. 63, 1126-1135. https://doi.org/10.1071/AN23026

Genetics of increased importance of health traits

***I. Misztal*[1], *M. Bermann*[1], *and D. Lourenco*[1]**
[1] *University of Georgia, Athens, GA, USA*
Corresponding author: ignacy@uga.edu

Abstract

The purpose of this study is to present mechanisms of possible negative effects of genomic selection in secondary traits, especially in health traits, and present steps to mitigate those effects. Selection on production traits causes corresponding declines in antagonistic fitness traits. Under traditional selection, declines were initially compensated or slowed down by continuously improving management. When the deterioration was too strong to ignore, the fitness traits were added to a selection index. Under genomic selection the genetic progress for production traits accelerates causing accelerated declines in fitness traits, and management modifications are not fast enough to compensate. Including the fitness traits in the selection index is only partially effective because of lower accuracy of evaluation due to lower heritability and fewer phenotypes. In particular, genomic selection can give rise to a sudden onset of categorical traits such as diseases. If deteriorating traits are underlying values for categorical traits such as a disease status, such traits are not expressed until a threshold is crossed. Then the incidence can increase rapidly. Mitigating the negative effects of genomic selection for health traits will require several steps. First, the selection pressure on major traits that are antagonistic to the deteriorating health traits needs to be reduced, allowing for management modifications to catch up. Second, selection for health traits needs to intensify, with weight on each trait calculated from recent genetic parameters. As genetic parameters can change rapidly under genomic selection, a challenge is to estimate these parameters for the last generation using complete data, as required for unbiased estimates. Third, methods for evaluating categorical traits can be improved. If the evaluation for health traits is not true multitrait with the production traits, the index weights need to be modified to account for ignored genetic correlations.

Key words: genomic selection, health traits, negative effects

Introduction

Genomic selection was originally perceived as an all-positive technology that facilitates fast genetic improvement at a low cost (Meuwissen et al., 2013). Recent reports indicate declines in secondary traits across species (Misztal and Lourenco, 2024). One of the most visible declines is the deterioration of hatchability in broiler chickens (Cash et al., 2025). While hatchability before genomic selection increased from 81% in 1998 to 85% in 2012, it dropped to 79% by 2024 under genomic selection (*https://innovateanimalag.org/the-hatchability-crisis*). The purpose of this study is to present mechanisms of possible negative effects of genomic selection in secondary traits, especially for categorical health traits, and suggest steps to slow or reverse the decline.

Materials and Methods

Misztal and Lourenco (2024) presented a theory on potential negative effects in genomic selection, as illustrated in Figure 1. Selection on production traits causes negative responses on antagonistic fitness traits. While fitness traits deteriorated under a single-trait selection, the decline was smaller than predicted because it was partly compensated by constantly improving management. When the deterioration could no longer be ignored, fitness traits were added to a selection index, slowing or even reversing the decline. Under

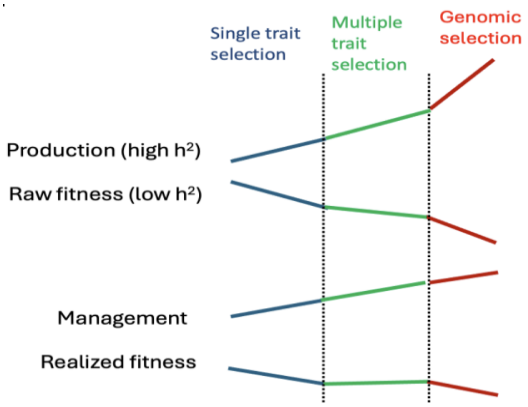


Figure 1. Trend in production and fitness under three types of selection; single trait selection is for the production trait and management applied to the fitness trait.

genomic selection, selection outcomes accelerate and the negative outcomes intensify.

Response to selection for unrecorded trait

Some theory on why we experience increased incidence of problems under selection for new traits can be derived from the theory of the selection index. Assume that selection is for an index of trait $I = \mathbf{a}'\hat{\mathbf{u}}$, where $\mathbf{a}$ is vector of economic weights and $'\hat{\mathbf{u}}$ is vector of EBV. Assume that fitness traits exist that are not yet problematic but may become problematic later. The response to selection for an unrecorded trait u_n is (Van Der Werf, 2025):

$$\begin{bmatrix} R_I \\ R_n \end{bmatrix} = i_I \, acc_I \begin{bmatrix} 1 & \\ & corr_{In} \end{bmatrix} \begin{bmatrix} \sigma_I \\ \sigma_n \end{bmatrix},$$

where R_I is a response to selection for the index, R_n is a response to selection for the unselected trait, i_I is selection intensity for the index, acc_I is accuracy of the index, $corr_{In}$ is correlation between the index and the unrecorded trait, and σ_I and σ_n are standard deviations for the index and the unrecorded trait, respectively. The equation can be generalized to many unrecorded traits.

When the correlation between the index and unrecorded traits is negative, selection on the index elicits a negative response on unrecorded traits. Although the change can be small in a single generation, it accumulates over time. The value of the correlation depends on composition of the index. When the index is composed mainly of production traits, the correlation with fitness traits is likely to be negative, as shown in Figure 2.

When the index I contains many fitness traits ("balanced"), the correlations of the index with fitness traits may be small and even positive, although the genetic progress for the production traits would be lower. In such a case, the unrecorded trait would not deteriorate.

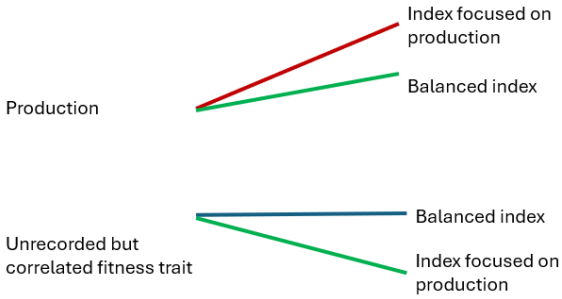


Figure 2. Impact of selection on an unrecorded correlated fitness trait with different indices.

Nonlinearity because of resource allocation issues

Usually, genetic analyses assume that the genetic parameters are static. However, under selection for production, resources allocation issues (van der Waaij, 2004) can cause negative genetic correlations to intensify, with changes accelerating under genomic selection (Hidalgo et al., 2020). If antagonistic traits are included in the selection index that were derived with old parameters, the response to selection for these traits will be progressively lower than expected, as illustrated in Figure 3.

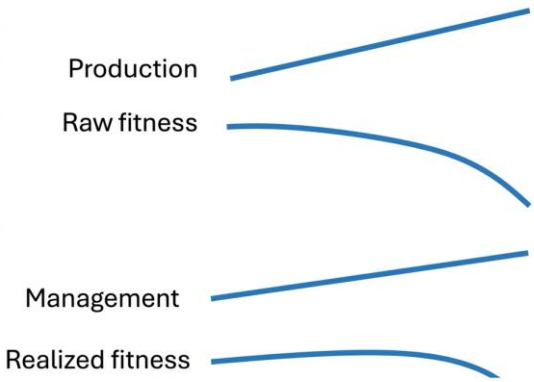


Figure 3. Relationship between production and fitness traits under changing genetic correlation.

Underlying values

A special case in animal breeding is a disease where the incidence increases rapidly, e.g., prolapse in pigs (Supakorn et al., 2017). Assume that the disease status is an effect of a threshold response to an underlying value, like in the threshold model (Gianola, 1982). Assume that the underlying variable is unselected and is antagonistic to the index. Subsequently, changes in the underlying variable will not be visible until that variable approaches the threshold. As is illustrated in Figure 4, then the incidence increases rapidly.

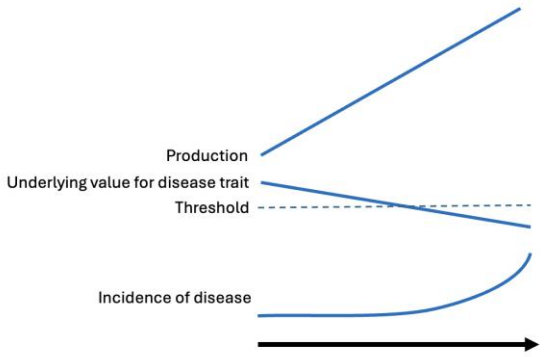


Figure 4. Incidence of a disease when its underlying variable is antagonistic with Production.

Use of incomplete multiple-trait model in the selection index

A selection index using EBV from a multiple-trait BLUP index is composed of only the economic values and EBVs (Van der Werf, 2025). Due to computational efficiency, usually production and fitness traits are evaluated separately. If a trait has little information and its genetic correlation with other traits is ignored in computing EBVs, the index based on such EBVs generates less response for that trait (Van der Werf, 2014). Corrections in such a case result in a custom index with weights adjusted for accuracies and genetic correlations.

Discussion

Mitigating negative effects

Mitigating the negative effects of genomic selection will require several steps, although, due to low heritability of fitness traits, a noticeable improvement can take several cycles of selection. First, the selection pressure on production traits that are antagonistic to the deteriorating traits may have to be reduced, allowing for management modifications to catch up. Second, selection for recorded fitness traits needs to intensify, with weight on each trait calculated from recent genetic parameters preferably by a multiple-trait model. Third, as genetic parameters can change rapidly under genomic selection, these need to be re-estimated for the last generation using the complete data. While traditional methods for parameter estimation are unfeasible with large data, new methods are emerging that are unrestricted by data size, such as genetic estimation by predictivity (GPP) (Misztal and Gopal, 2025) or REML with traces calculated by Monte Carlo methods (Bermann et al., 2025). Fourth, methods for evaluating fitness traits can be improved. While many fitness traits are categorical, current methods allow for the evaluation of a single categorical trait and multiple linear traits, and corresponding computations are expensive. New methods under development will potentially analyze any number of categorical and linear traits simultaneously by a threshold-linear model at a reasonable computing cost (Bermann et al, 2026). Similar advancements would also be helpful for censored traits such as survival.

Conclusions

Genomic selection can cause deterioration of secondary traits as an accelerated selection on production traits causes a correlated decline on antagonistic traits; improving management is too slow to mitigate the decline. Additional factors intensifying the decline is lack of the deteriorating trait in the selection index, genetic correlations become more antagonistic, and possibly a suboptimal index if EBV are not calculated by a complete multiple-trait model. The decline can occur rapidly for categorical traits. Sustaining the long-term progress of genomic selection while managing and

preventing potential negative effects requires including the problematic traits in the selection index at an appropriate weight, possibly with higher weights if the model ignores the genetic correlations, improving the accuracy of fitness traits, and adopting cutting-edge models and computational methods.

Acknowledgements

We gratefully acknowledge discussion on selection indices with Paul VanRaden and Julius van der Werf.

References

Bermann, M., Legarra, A., Aguilar, I., Munera, A.A., Misztal, I. and Lourenco, D. 2025. Estimation of (co)variance components for very large datasets and complex single-step genomic models. *Genet. Sel. Evol.* 57 (1), 63. https://doi.org/10.1186/s12711-025-01006-9

Bermann, M., Legarra, A., Misztal, I. and Lourenco, D. 2026. Methods for joint genetic prediction of multiple ordinal categorical and continuous traits. *Genetics* 233 (1), iyag086. https://doi.org/10.1093/genetics/iyag086

Cash, C., Witherspoon, K. and Athrey, G. 2025. How concerned should we be about broiler breeder fertility declines? *Poult. Sci.* 104 (5), 104992. https://doi.org/10.1016/j.psj.2025.104992

Gianola, D. 1982. Theory and Analysis of Threshold Characters. *J. Anim. Sci.* 54 (5), 1079-1096. https://doi.org/10.2527/jas1982.5451079x

Hidalgo, J., Tsuruta, S., Lourenco, D., Masuda, Y., Huang, Y., Gray, K.A. and Misztal, I. 2020. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. *J. Anim. Sci.* 98 (2), skaa032. https://doi.org/10.1093/jas/skaa032

Meuwissen, T., Hayes, B. and Goddard, M. 2013. Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1 (1), 221-237. https://doi.org/10.1146/annurev-animal-031412-103705

Misztal, I. and Lourenco, D. 2024. Potential negative effects of genomic selection. *J. Anim. Sci.* 102, skae155. https://doi.org/10.1093/jas/skae155

Misztal, I. and Gopal, G. 2025. Estimation of heritabilities and genetic correlations by time slices using predictivity in large genomic models. *Genetics* 230 (2), iyaf066. https://doi.org/10.1093/genetics/iyaf066

Supakorn, C., Stock, J., Hostetler, C. and Stalder, K. 2017. Prolapse incidence in swine breeding herds is a cause for concern. *Open J. Vet. Med.* 7 (8), 85-97. https://doi.org/10.4236/ojvm.2017.78009

Van der Waaij, E.H. 2004. A resource allocation model describing consequences of artificial selection under metabolic stress. *J. Anim. Sci.* 82 (4), 973-981. https://doi.org/10.2527/2004.824973x

Van der Werf, J. 2014. Index weights or Economic Values for a dairy index based on single trait ABVs. Mimeo, University of New England. Available at https://jvanderw.une.edu.au/Index%20weight%20for%20single%20traits%20EBVs%20in%20multiple%20trait%20indexesv4.pdf

Van der Werf, J. 2025. Selection Index Theory: Information from Relatives. Available at https://jvanderw.une.edu.au/Chapter03_GENE422_SelInd.pdf.

Interaction between adaptation and mitigation traits in Canadian Holsteins

G.R. Dodd[1], C.F. Baes[1], F.S. Schenkel[1], C. Machefert[1], I. Ampofo[1], B. Amini[1], V. Storing[1], A. Fleming[2], K. Lee[2], F. Malchiodi[1,3], C. Richardson[4], and F. Miglior[1,2]

[1] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, N1G 2W1 Ontario, Canada

[2] Lactanet, Guelph, N1K 1E5 Ontario, Canada

[3] Semex, 5653 Hwy 6, Guelph, N1H 6J2 Ontario, Canada

[4] AbacusBio International Ltd, Roslin Innovation Centre, Easter Bush Campus, University of Edinburgh, EH25 9RG Edinburgh, United Kingdom

Corresponding author: gdodd@uoguelph.ca

Abstract

In response to climate change, the dairy industry has shifted focus to reducing environmental impact and strengthening resilience. However, it is unclear how adaptation to environmental challenges and mitigation of environmental effects interact. This project aims to quantify the impact of selecting traits related to adaptation on profitability, health, efficiency, resilience, and environmental sustainability to inform selection of climate-smart dairy. Mitigation traits were defined as methane production (g/d), body maintenance requirements, and dry matter intake while adaptation traits were defined as heat tolerance with regards to production, fertility, and health. The first objective will be to quantify the direct effects of heat stress on production, fertility, health, and mitigation traits using a reaction norm approach to estimate heat tolerance. The second objective will be to estimate genetic correlations between mitigation traits and heat tolerance defined by production, fertility, and health traits. The third objective is to estimate genetic correlations among heat tolerance defined by all trait groups. Preliminary results have identified trait-specific heat stress thresholds for production, fertility, and mitigation traits ranging from THI 55 to 67, where fat yield was the most sensitive to environmental pressures and first service to conception was the least sensitive. Genetic correlations estimated between performance under thermoneutral conditions and heat tolerance for production traits were found to be moderately unfavorable (-0.27 to -0.31), indicating a trade-off between productivity and resilience. Weak and favorable genetic correlations were estimated between methane emissions and heat tolerance of production traits, with the strongest relationships observed between heat tolerance of fat yield and methane emissions (-0.05 to -0.16). These results highlight the need for balanced selection strategies to improve both productivity and heat tolerance, while suggesting that selection for reduced methane emissions is unlikely to adversely affect heat tolerance. These findings will inform climate-smart dairy selection strategies balancing profitability, resilience, and environmental impact.

Key words: environment, methane, heat tolerance

Introduction

Improving the environmental efficiency of dairy production is an important component of sustainable livestock production. Genetic selection provides a cumulative and permanent means of improving efficiency through traits related to greenhouse gas emissions and resource use, including methane efficiency, feed efficiency, and body maintenance requirements. In Canada, these traits have been incorporated into the Environmental Impact Sub-Index of the national Lifetime Performance Index (LPI), providing a genetic selection

framework for reducing the environmental footprint of dairy production (Van Doormaal et al., 2025).

Climate change presents a complementary challenge for the dairy industry. Increasing temperatures and more frequent heat stress events can negatively affect production, fertility, and health in dairy cattle (Bush et al., 2019; Polsky and von Keyserlingk, 2017; Yan et al., 2021; Campos et al., 2022b; Ishida et al., 2024; Dodd et al., 2026a). These effects vary among animals and traits, partly due to genetic differences in their ability to maintain performance under heat stress, commonly referred to as heat tolerance (Campos et al., 2022a; McWhorter et al., 2023). Previous research has demonstrated genetic variation in heat tolerance that could be exploited through selection for improved resilience. This is particularly relevant for Canadian Holsteins given increasing temperatures across Canada and the susceptibility of high-producing cattle to heat stress (Zhang et al., 2019; Hansen, 2004).

Despite opportunities to improve both environmental efficiency and heat tolerance, these objectives have largely been studied independently. Consequently, the genetic relationships between mitigation and adaptation traits remain poorly understood in Canadian Holsteins. Characterizing these relationships is necessary to determine whether selection for improved environmental efficiency could influence heat tolerance and to identify potential trade-offs or opportunities for balanced selection. Integrating these traits could therefore support the development of breeding strategies that improve both environmental efficiency and resilience to changing environmental conditions.

The objective of this project was to characterize the genetic basis of adaptation and mitigation in Canadian Holsteins and evaluate relationships among these traits. Specifically, the project aimed to 1) quantify direct effects of heat stress on production, health, fertility, and mitigation traits; 2) assess genetic relationships among adaptation and mitigation traits to identify potential trade-offs and opportunities for balanced selection; and 3) evaluate genetic relationships of heat tolerance among trait groups.

Materials and Methods

No approval by an Institutional Animal Care and Use Review Board was required because data were obtained from an existing database.

Data:

Phenotype data were provided by Lactanet Canada (Guelph, ON) comprising test-day records from first-lactation Canadian Holstein cows. Production traits included milk (MY; kg/d), fat (FY; g/d), and protein (PY; g/d) yields. Health traits included β-hydroxybutyrate (BHB; mmol/L) and somatic cell score (SCS). The fertility trait evaluated was days from first service to conception (FSTC; d). Mitigation traits included methane production predicted from mid-infrared spectral data (MP_{MIR}; g/d), body maintenance requirements (BMR; body weight^{75}), and dry matter intake (DMI; kg/d).

Weather data were extracted from the NASA POWER Database. Data included hourly records of ambient temperature (T; ˚C) and relative humidity (H; %) for each herd across the study period. Meteorological estimates in POWER are derived primarily from the NASA Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), which combines observations from multiple sources, including satellites, surface observations, aircraft, radiosondes, ships, and buoys, with an atmospheric model through data assimilation. The data are provided at a spatial resolution of 0.5˚ latitude by 0.625˚ longitude. Herd coordinates were matched to the existing database of weather records, and date ranges from test dates and heat stress windows evaluated were used to extract the relevant records. Hourly temperature and

humidity index (THI) values were calculated using the equation reported by NRC (1971):

$$THI = (1.8*T) + 32 - (0.55 - (0.0055*H))*((1.8*T) - 26) \quad (1)$$

Daily THI values were generated by averaging across each 24-hour day and rounding to the tenths decimal place.

Phenotypic Analysis:

Phenotypic analyses were conducted to evaluate the impact of heat stress on production, fertility, health, and mitigation traits. Single trait animal models were fit for each trait to remove all known sources of environmental and genetic effects following the general structure of,

$$y = Xb + Za + Wu + Vo + e \quad (2)$$

where y is the phenotype, X is the design matrix for fixed effects in vector b, Z is the design matrix for additive genetic effects in vector a, W is the design matrix for additional random effects in vector u, V is the design matrix for permanent environment effects in vector o, and e is the vector of residuals. Fixed effects, presence of additional random effects, and presence of permanent environment effects varied between traits. Models implemented were based on the national evaluation models implemented by Lactanet Canada. The residual terms of these models capture variation in the phenotypes due to unknown sources, such as THI, and reflects the unexpected deviation in the phenotype after accounting for fixed and random effects.

For each trait, several temporal windows of effects were tested surrounding the test-day of the phenotype. Range of windows were based on previous studies and biological reasoning. Daily THI was averaged across each window and associated with the phenotypic record with windows ranging from ten days prior to two days after the test-day. Model residuals were plotted against the average THI generated from each of these windows to approximate the number and location of breakpoints. Thresholds of heat stress effects were identified using a segmented regression approach, where the relationship between THI and residuals is modeled by multiple linear functions. The breakpoints for each model were identified based on model fit, assessed by the adjusted coefficient of determination (R^2), visual inspection of the fitted regression line, and evaluation of associated pre- and post-breakpoint slopes.

Genetic Analysis:

Genetic analyses of heat tolerance were performed to estimate the heritability of traits under thermoneutral and heat stress conditions as well as to estimate the genotype by environment effect for the traits. The general model fit was

$$y=Xb+ Z_0a_0+ Z_1a_1+ Wu+ V_0o_0+ V_1o_1+ e \quad (3)$$

where y, X, b, W, u, and e are as defined above. The additive genetic and permanent environment effects were adjusted to include a slope and intercept term such that Z_0 is the incidence matrix for the random additive genetic intercept term in a_0, Z_1 is the incidence matrix for the random additive genetic slope term in a_1, V_0 is the incidence matrix for the random permanent environment intercept term in o_0, V_1 is the incidence matrix for the random permanent environment slope term in o_1. Slope terms (a_1 and o_1) were fit using a heat stress function,

$$HSF(Tr) = \max(0, Tr - th_t) \quad (4)$$

where r is each test-day record, Tr is the THI average associated with record r, and th_t is the THI threshold associated with each trait. Thresholds and windows for THI averages used in the genetic analysis were based on the best fit model identified in the phenotypic analysis. In this model, intercept terms (a_0 and o_0) represent thermoneutral production and slope terms (a_1

and o_1) represent heat tolerance, defined as the linear change per unit THI above the threshold.

Genetic correlations between the thermoneutral genetic intercept (a_0) and heat tolerance genetic slope (a_1) were estimated to evaluate the relationship between performance under thermoneutral conditions and the response to increasing THI above the trait-specific threshold defined by the heat stress function in Equation 4.

Correlation Analysis of Adaptation and Mitigation:

Genetic correlations were estimated among adaptation and mitigation traits to evaluate relationships between environmental resilience and mitigation potential and identify potential trade-offs in selection. Pairwise models were used to estimate correlations between the additive genetic effects of mitigation traits and the genetic intercept and slope terms of adaptation traits. Adaptation traits included production, health, and fertility traits and were modeled using the reaction norm approach described in Equation 3. Mitigation traits included MP_{MIR}, BMR, and DMI and were modeled using repeatability animal models. Thus, correlations with the adaptation intercept (a_0) represented relationships with thermoneutral performance, while correlations with the adaptation slope (a_1) represented relationships with heat tolerance. Correlations were considered significant when the absolute value of the estimate exceeded twice its standard error.

Preliminary Results & Discussion

Heat Stress Effects Varied Among Traits:

The THI threshold and window of susceptibility to heat stress differed among traits (Table 1). Production traits thresholds ranged from THI 55 for FY to THI 65 for MY. The highest estimated threshold was found for FSTC at THI 67. Windows of susceptibility also varied among traits, ranging from 2 d before through the test day for production traits to 3 d before through 2 d after the test day for FSTC. These differences demonstrate that the timing and magnitude of heat stress responses are trait-specific and support the use of trait-specific THI thresholds and exposure windows when evaluating heat tolerance. Detailed results for production and fertility traits are presented in Rockett et al. (2023b) and Dodd et al. (2026a), respectively. Results for MP_{MIR} (Amini et al., *in preparation*) and health traits (Campos et al., 2026; Storing et al., *in preparation*) are in preparation.

Table 1: Trait-specific temperature-humidity index (THI) thresholds and windows of susceptibility to heat stress. FSTC=first service to conception; MY=milk yield; FY=fat yield; PY=protein yield.

Trait	Units	Threshold	Window
FSTC	days	67	d-3 to d2
MY	kg/d	65	d-2 to d0
FY	g/d	55	d-2 to d0
PY	g/d	60	d-2 to d0

Genetic Variation in Heat Tolerance:

Genetic variation was identified for both thermoneutral performance and heat tolerance across the evaluated traits (Table 2). Heritability estimates for production traits ranged from 0.30 ± 0.01 to 0.22 ± 0.00 under thermoneutral conditions and from 0.24 ± 0.01 to 0.27 ± 0.01 under heat stress. These estimates indicate that genetic selection could be used to improve both overall production and the ability of animals to maintain performance under heat stress. Overall, the moderate heritability estimates under heat stress are consistent with previous studies in North American Holsteins (Campos et al., 2025; McWhorter et al., 2023) and indicate sufficient genetic variation to select for improved heat tolerance.

Genetic correlations between thermoneutral performance and heat tolerance were negative for all three production traits (Table 2). The strongest antagonistic relationship was observed for MY, followed by PY and FY. These results indicate an antagonistic genetic relationship between production potential and the ability to maintain performance under heat

stress. This is consistent with previous reports in Canadian and U.S. Holsteins (Campos et al., 2022a; Rockett et al., 2023a; McWhorter et al., 2023), which also reported a trade-off between production potential and heat tolerance. Thus, selection solely for increased production may compromise heat tolerance unless it is explicitly incorporated into breeding objectives.

Table 2: Heritabilities (h^2) and genetic correlations $r_{g(TN,HT)}$ among production traits under thermoneutral conditions and heat tolerance. Standard errors are shown in parentheses. Heritability of heat tolerance estimated at temperature-humidity index (THI) 75. TN = thermoneutral conditions; HT = heat tolerance; MY = milk yield; FY = fat yield; PY = protein yield.

Trait	TN h^2	HT h^2	$r_{g(TN,HT)}$
MY	0.30 (0.01)	0.26 (0.01)	-0.29 (0.05)
FY	0.22 (0.00)	0.24 (0.01)	-0.22 (0.03)
PY	0.22 (0.00)	0.27 (0.01)	-0.23 (0.03)

Genetic Correlations Between Adaptation and Mitigation:

Genetic correlations between MP_{MIR} and production traits were generally weak under both thermoneutral and heat stress conditions (Table 3). Under thermoneutral conditions, correlations were negligible for MY (-0.02 ± 0.02) and PY (-0.11 ± 0.02) and moderately unfavorable for FY (0.38 ± 0.02). Under heat stress, correlations were weak and favorable for all traits, ranging from -0.03 ± 0.04 for PY to -0.16 ± 0.04 for FY. These results indicate that selection for lower MP_{MIR} is unlikely to substantially affect MY or PY under either condition. The stronger association with FY suggests that selection for lower MP_{MIR} may be associated with a slight reduction in thermoneutral FY and a small improvement in the ability to maintain FY under heat stress. Results of this analysis are discussed in further detail in Dodd et al. (2026b).

Table 3: Genetic correlations between MIR-predicted methane production and production traits under thermoneutral conditions and heat tolerance. Standard errors are shown in parentheses. TN = thermoneutral conditions; HT = heat tolerance; MY = milk yield; FY = fat yield; PY = protein yield; MP_{MIR} = MIR-predicted methane production.

Trait	TN	HT
MY	-0.02 (0.02)	-0.10 (0.05)
FY	0.38 (0.02)	-0.16 (0.04)
PY	-0.11 (0.02)	-0.03 (0.04)

Conclusions

Heat stress responses were trait-specific, with differences in both the THI breakpoint and window of susceptibility among trait groups. Genetic variation was identified for both thermoneutral performance and heat tolerance of production traits, indicating potential for genetic selection to improve resilience to heat stress. The antagonistic relationship between thermoneutral production and heat tolerance further highlights the importance of including heat tolerance directly in breeding objectives rather than relying solely on selection for production. Positive genetic correlations among production traits indicate shared genetic control while supporting the inclusion of all traits in selection objectives. Generally weak relationships between heat tolerance of production traits and MP_{MIR} suggest that adaptation and mitigation can be improved with limited correlated responses. This project is ongoing and analyses are being completed across production (MY, FY, PY), health (BHB, SCS), fertility (FSTC), and mitigation traits (MP_{MIR}, BMR, DMI). Integration of these results will provide a comprehensive assessment of the genetic relationships among adaptation and mitigation traits to support the development of breeding strategies for more resilient and sustainable dairy cattle.

Acknowledgments

This study was part of Net-Zero Dairy Genome Project (NDGP; ICT-19401), Dairy Cluster 4 projects (Activities 8 & 13), AFC project (AFC-2025F3716R), AdaptMitigate Dairy Genome Project (ADGP; GAPP-1545773). We gratefully acknowledge all funding and support, as administered by Genome Canada (Ottawa, ON, Canada), Ontario Genomics (Toronto, ON, Canada), Genome Quebec (Montréal, QC, Canada), Genome Alberta (Calgary, ON Canada), Lactanet (Guelph, ON, Canada) and Dairy Farmers of Canada (Ottawa, ON, Canada) and contributions of Dairy Farmers of Ontario (Mississauga, ON, Canada), Dairy Farmers of Quebec (Montréal, QC, Canada), Alberta Milk (Edmonton, AB, Canada), Agricultural Funding Consortium (Edmonton, AB, Canada), Alberta Innovates (Edmonton, AB, Canada), Alberta Ministry of Technology and Innovation (Edmonton, AB, Canada), and Natural Sciences and Engineering Research Council of Canada (NSERC; Ottawa, ON, Canada). This research is supported by a contribution from the Dairy Research Cluster 4 partners (Lactanet Canada, Dairy Farmers of Canada, Semex Alliance, Holstein Canada and Agriculture and Agri-Food Canada) under the Sustainable Canadian Agricultural Partnership AgriScience Program. As per the research agreement, researchers maintain independence in conducting their studies, own their data, and report the outcomes regardless of the results. The decision to publish the findings rests solely with the researchers. No approval by an Institutional Animal Care and Use Review Board was required because data were obtained from an existing database. The authors have not stated any conflicts of interest.

References

Bush, E., N. Gillett, E. Watson, J. Fyfe, F. Vogel, and N. Swart. 2019. Understanding Observed Global Climate Change. Pages 24–72 in Canada's Changing Climate Report. Government of Canada.

Campos, I.L., T.C.S. Chud, G.A.O. Junior, C.F. Baes, Á. Cánovas, and F.S. Schenkel. 2022a. Estimation of Genetic Parameters of Heat Tolerance for Production Traits in Canadian Holsteins Cattle. Animals 12:3585. https://doi.org/10.3390/ani12243585.

Campos, I.L., T.C.S. Chud, H.R. Oliveira, C.F. Baes, A. Cánovas, and F.S. Schenkel. 2022b. Using publicly available weather station data to investigate the effects of heat stress on milk production traits in Canadian Holstein cattle. *Can. J. Anim. Sci.* 102:368–381. https://doi.org/10.1139/cjas-2021-0088.

Campos, I.L., R.E. Jahnel, F. Miglior, Z. Feng, A. Canovas, C.F. Baes, and F.S. Schenkel. 2025. Modeling heat tolerance for production traits in Canadian Holstein cattle. *J. Dairy Sci.* 108:9956–9966. https://doi.org/10.3168/jds.2025-26433.

Campos, I.L., R.E. Jahnel, H.R. Oliveira, F. Miglior, C.F. Baes, and F.S. Schenkel. 2026. Phenotypic and genetic analysis of somatic cell score response to heat stress in first lactation Holstein cattle. *J Anim Sci* skag252. https://doi.org/10.1093/jas/skag252.

Dodd, G.R., F. Miglior, F.S. Schenkel, I.L. Campos, R.E. Jahnel, and C.F. Baes. 2026a. Modeling heat stress effects on first service to conception rates in Canadian Holstein dairy cattle. *J. Dairy Sci.* 109:2727–2742. https://doi.org/10.3168/jds.2025-27171.

Dodd, G.R., F.S. Schenkel, C.F. Baes, F. Malchiodi, C. Richardson, and F. Miglior. 2026b. Genetic associations between mid-infrared-predicted methane production and heat tolerance of production traits in dairy cattle. *J. Dairy Sci.* 109:9677-9686. https://doi.org/10.3168/jds.2026-28568.

Hansen, P.J. 2004. Physiological and cellular adaptations of zebu cattle to thermal stress. *Animal Reproduction Science* 82–83:349–360. https://doi.org/10.1016/j.anireprosci.2004.04.011.

Ishida, S., A. Nishiura, T. Yamazaki, H. Abe, S. Nakagawa, Y. Nakahori, S. Yamaguchi, Y. Masuda, Y. Saito, R. Tatebayashi, T. Osawa, C.-H. Huang, and K. Hagiya. 2024. Effects of heat stress on predicted energy balance, fat-to-protein ratio, and milk β-hydroxybutyrate in first-lactation Holstein cattle in Hokkaido, Japan. *Animal Science Journal* 95:e70013. https://doi.org/10.1111/asj.70013.

McWhorter, T.M., M. Sargolzaei, C.G. Sattler, M.D. Utt, S. Tsuruta, I. Misztal, and D. Lourenco. 2023. Single-step genomic predictions for heat tolerance of production yields in US Holsteins and Jerseys. *J. Dairy Sci.* 106:7861–7879. https://doi.org/10.3168/jds.2022-23144.

NRC, (National Research Council). 1971. A Guide to Environmental Research on Animals. National Academy of Sciences.

Polsky, L., and M.A.G. von Keyserlingk. 2017. Invited review: Effects of heat stress on dairy cattle welfare. *J. Dairy Sci.* 100:8645–8657. https://doi.org/10.3168/jds.2017-12651.

Rockett, P.L., I.L. Campos, C.F. Baes, D. Tulpan, F. Miglior, and F.S. Schenkel. 2023a. Genetic evaluation of heat tolerance in Holsteins using test-day production records and NASA POWER weather data. *J. Dairy Sci.* 106:6995–7007. https://doi.org/10.3168/jds.2022-22776.

Rockett, P.L., I.L. Campos, C.F. Baes, D. Tulpan, F. Miglior, and F.S. Schenkel. 2023b. Phenotypic analysis of heat stress in Holsteins using test-day production records and NASA POWER meteorological data. *J. Dairy Sci.* 106:1142–1158. https://doi.org/10.3168/jds.2022-22370.

Van Doormaal, B., C. Jaton, H. Sweett, P.G. Sullivan, A. Fleming, F. Miglior, and C.M. Richardson. 2025. Modernizing Canada's Lifetime Performance Index (LPI). Interbull Bulletin 203–212.

Yan, G., K. Liu, Z. Hao, Z. Shi, and H. Li. 2021. The effects of cow-related factors on rectal temperature, respiration rate, and temperature-humidity index thresholds for lactating cows exposed to heat stress. *Journal of Thermal Biology* 100:103041. https://doi.org/10.1016/j.jtherbio.2021.103041

Zhang, X., G. Flato, M. Kirchmeier-Young, L. Vincent, H. Wan, X. Wang, R. Rong, J. Fyfe, G. Li, and V.V. Kharin. 2019. Changes in Temperature and Precipitation Across Canada. Pages 112–193 in Canada's Changing Climate Report. Government of Canada.

Walloon strategies targeting new traits for efficiency, mitigation, adaptation and resilience in dairy cattle

K. Wijnrocx[1], P. Lemal[1], H. Atashi[1], Y. Chen[1], H. Bérat[1] and N. Gengler[1]

[1] TERRA Teaching and Research Centre, ULiège-Gembloux Agro-Bio Tech, 5030 Gembloux, Belgium

Corresponding author: nicolas.gengler@uliege.be

Abstract

The transition towards sustainable dairy production requires the development and implementation of novel traits addressing environmental impact, resource efficiency, animal welfare, adaptation, and resilience. In the Walloon Region of Belgium, these developments are supported by long-term investments in milk mid-infrared (MIR) phenotyping, routine milk recording, genomic evaluation infrastructure, and digital data integration. The objective of this paper is to present a unified framework enabling rapid implementation of novel traits into routine genomic evaluations. The proposed framework relies on predicting novel phenotypes from routinely collected MIR milk spectra, followed by genomic evaluation using existing random regression single-step genomic BLUP (RR-ssGBLUP) procedures originally developed for production traits. Methane emissions predicted from MIR spectra (PCH4) represent the first trait already implemented in routine evaluations. The methane evaluation currently includes more than 4.5 million test-day records from over 328,000 Holstein cows and 9,631 genotyped animals. Genomic breeding values and methane efficiency indexes are routinely generated. The same framework is currently being expanded towards adaptation and resilience traits. A recent example is Heat Stress Response (HSR), developed from MIR-based phenotypes generated from data collected during periods of thermal challenge. Genomic evaluations resulted in heritability estimates of approximately 0.10 and generated more than 6,700 genomic breeding values with moderate to high reliabilities. Preliminary analyses indicate favorable opportunities for selection aimed at improving heat tolerance while maintaining productive performance. Additional developments include welfare-related traits, feed efficiency indicators, nitrogen use efficiency, and the integration of sensor-derived information through the WALLeSmart digital platform. These developments demonstrate that routine phenotyping infrastructures can support rapid deployment of novel sustainability traits. The presented strategy illustrates how a common genomic evaluation architecture can accelerate implementation of emerging traits and facilitate future international collaborations and genomic exchange.

Key words: genomic evaluation, methane emissions, heat stress, mid-infrared spectra, resilience, sustainability

Introduction

The dairy sector faces increasing societal, environmental, and economic pressures requiring breeding objectives that go beyond traditional production traits. Sustainable dairy production requires simultaneous improvement of efficiency, mitigation of greenhouse gas emissions, adaptation to changing climatic conditions, welfare, and resilience.

The Walloon breeding program has invested for more than fifteen years in the development of milk MIR-based phenotypes through the Futurospectre consortium (the Walloon Breeders Association - Elevéo, the Milk Committee - CdL, the Walloon Agricultural Research Center - CRA-W and ULiège-Gembloux Agro-Bio Tech). This infrastructure provides access to routine MIR spectra for nearly all milk-recorded cows and offers a

unique opportunity to derive large-scale phenotypes for novel traits. This regional organization of dairy recording in Wallonia further facilitates interactions among breeders, breed organizations and research institutions. Through long-term collaborations not only standardized MIR spectral information (Grelet et al., 2015) is routinely available for virtually all milk-recorded cows but also numerous MIR prediction equations, including fatty acid profiles, methane emissions and heat stress related traits, thereby creating a favorable environment for rapid implementation of novel phenotypes into breeding programs.

The objective of this paper is to describe a unified genomic evaluation strategy for MIR-based novel traits and to illustrate its implementation using methane emissions as a routine application and heat stress response as an advanced prototype. Perspectives for welfare, efficiency, and resilience traits are also discussed.

Materials and Methods

Walloon Phenotyping Infrastructure and Common Genomic Evaluation Pipeline

The Walloon phenotyping infrastructure is built on the routine collection of milk samples through dairy herd improvement and milk payment schemes, providing MIR spectra for the vast majority of dairy cows in Wallonia. Spectra are standardized following CRA-W procedures (Grelet et al., 2015), and novel phenotypes are generated through prediction equations linking MIR spectra with directly measured biological traits (Grelet et al., 2021) but also physiological status as heat stress (Lemal et., al 2026). This cost-effective approach enables large-scale phenotyping while avoiding the expense and logistical constraints of direct measurements.

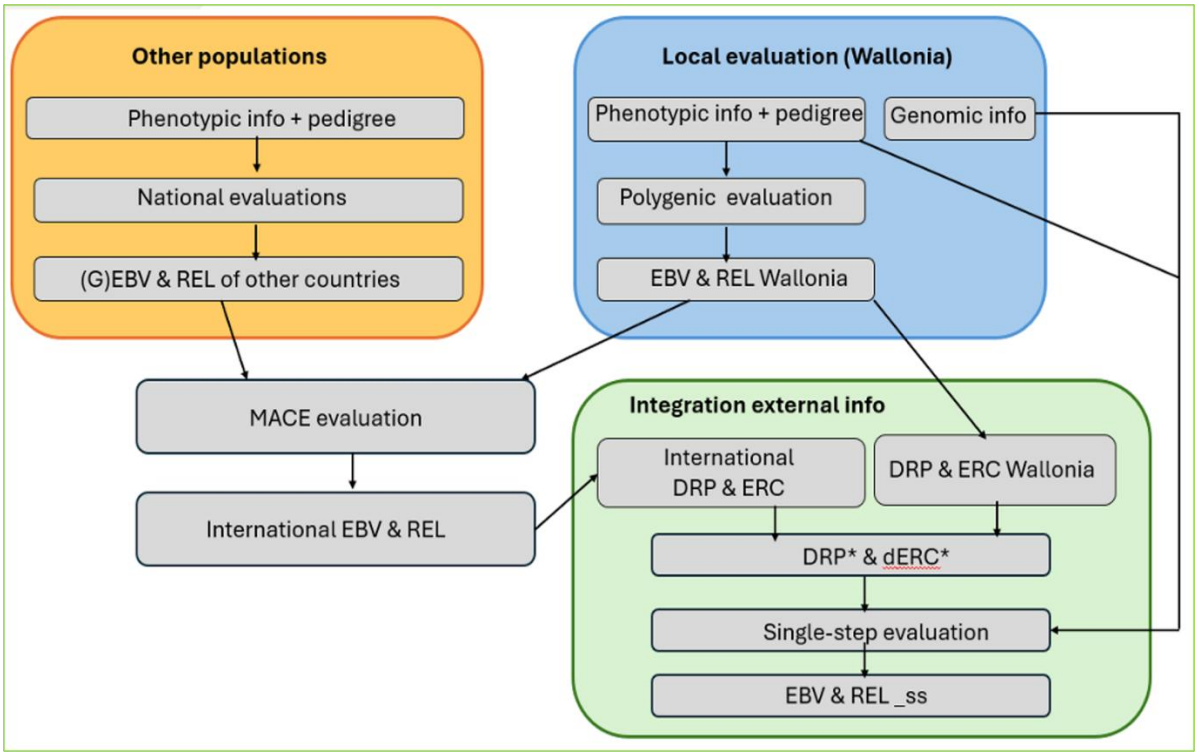


Figure 1. Generic ssGBLUP framework used in Wallonia showing the general evaluation pipeline, here illustrated for a trait which has a MACE evaluation (Vanderick et al., 2025).

All MIR-derived traits can be integrated into a common genomic evaluation pipeline (Figure 1.) based on (i) derivation of phenotypes from MIR spectra, (ii) test-day random regression modelling, (iii) integration of pedigree and genomic information through single-step GBLUP (ssGBLUP), (iv) estimation of genomic breeding values and reliabilities, and (v) dissemination of breeding values and derived indexes.

In practice, conventional milk, fat, and protein traits are replaced in the genetic evaluation framework by MIR-predicted traits while maintaining the same statistical evaluation framework, allowing new traits to be implemented with minimal modifications once a sufficiently robust prediction equation is available.

An alternative to replacement of milk, fat and protein is outlined by Chen et al. (2026) where a highly correlated trait is added as a trait to the system. In this specific application somatic cell score was added.

The entire system is supported by the WALLeSmart platform, which integrates phenotypic, genomic, management, and sensor-derived information, facilitates collaboration among breeders, breed organizations, and scientists, and accelerates the validation, deployment, and routine use of innovative phenotypes and precision breeding tools.

Methane Emissions

An already implemented trait is methane emissions. More details are reported by Chen et al. (2025). Methane emission were predicted using an MIR calibration equation developed from 1,089 reference observations collected on 299 cows originating from Belgium, Ireland, Switzerland, the United Kingdom, France, Denmark and Germany. The published calibration achieved a cross-validation coefficient of determination of approximately 0.68 with a prediction error of approximately 57 g CH4/day (Vanlierde et al., 2021).

Methane emissions were represented by MIR-predicted methane output (PCH4). The current evaluation includes

- approximately 4.5 million test-day records,
- 328,000 Holstein cows,
- 1,520 herds,
- 9,631 genotyped animals.

Routine genomic evaluations generate breeding values for expected methane emissions together with alternative methane-efficiency indexes.

Heat Stress Response

Heat Stress Response (HSR) was developed using an MIR-based prediction equation derived from 399 Holstein cows recorded during 2024 (Lemal et al., 2026). The approach was designed to characterize the biological response of cows to thermal challenges independently of direct THI measurements. The resulting equation was subsequently applied to more than one million MIR spectra collected from Holstein cows between 2020 and 2022 to generate large-scale phenotypes for genomic evaluation.

The genomic evaluation dataset included approximately:

- 64,000 cows with phenotypes,
- 146,000 pedigree animals,
- 7,170 genotyped animals.

Genomic breeding values were estimated using the same ssGBLUP framework as used for methane emissions.

Future Trait Developments

Additional developments currently under investigation include:

- Welfare-related traits;
- Feed efficiency indicators;
- Nitrogen-use efficiency;
- Traits derived from automatic milking systems;
- Sensor-based resilience indicators.

They will not be described in detail except welfare-related traits (Chen et al., 2026)

Results & Discussion

Generic Implementation of MIR-Based Genomic Evaluations

The principal advantage of the proposed framework is that once an acceptable MIR prediction equation becomes available, implementation within the genomic evaluation system requires only minimal modifications to existing procedures. This substantially reduces development costs and facilitates rapid deployment of novel traits.

Routine Methane Evaluation

Methane emission evaluations represent the first routine implementation of this strategy (Chen et al., 2025). Therefore, Methane emissions provided the first operational example of how MIR-based phenotyping can be translated into routine genomic selection. Rather than directly measuring methane emissions on each cow, methane output is predicted from routinely collected milk mid-infrared (MIR) spectra using a previously validated calibration equation. This approach transforms MIR information already generated through milk recording into repeated predicted methane phenotypes (PCH4) at the test-day level, which can subsequently enter the same random regression genomic evaluation framework used for other dairy traits.

The current Walloon evaluation includes more than 4.5 million test-day records from approximately 328,000 Holstein cows, illustrating how routine MIR recording can provide the phenotypic depth required for genomic evaluation of a trait that would otherwise be difficult and costly to measure directly at population scale.

As illustrated in Figure 1 and 2, the Walloon methane evaluation combines MIR-predicted methane records with pedigree and genomic information to generate genomic breeding values for methane emissions. In parallel, existing genomic evaluations for milk production and composition provide information on the methane emissions expected from an animal's production potential. Contrasting these two components allows methane efficiency to be evaluated while accounting for productive performance, rather

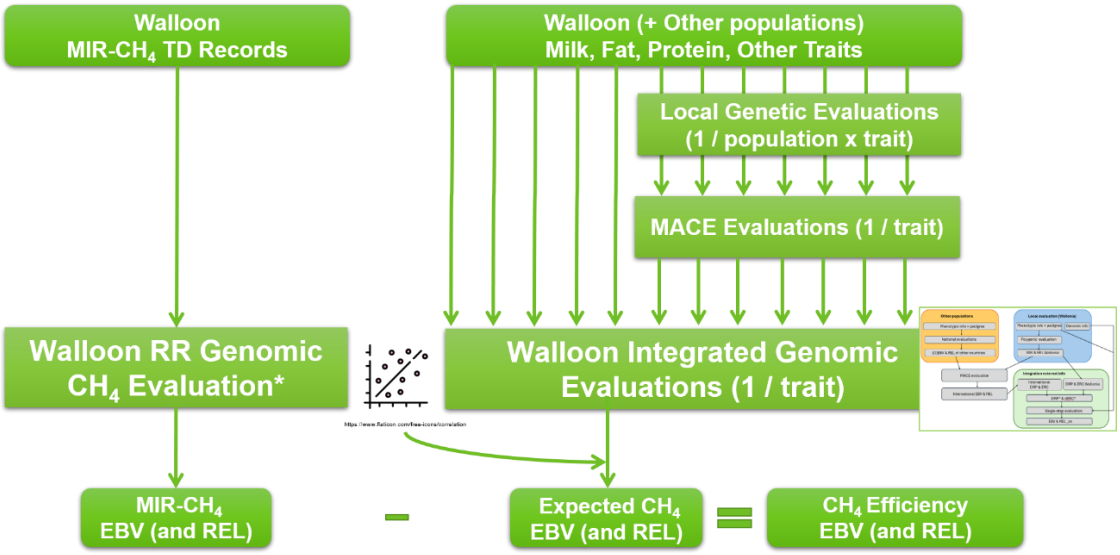


Figure 2. Walloon genomic evaluation framework for methane emissions and methane efficiency. MIR-predicted methane test-day records are used in a random regression genomic evaluation to derive methane breeding values. Information from routine genomic evaluations for production traits is used to derive expected methane emissions, and the contrast between MIR-predicted and expected methane breeding values provides an indicator of methane efficiency.

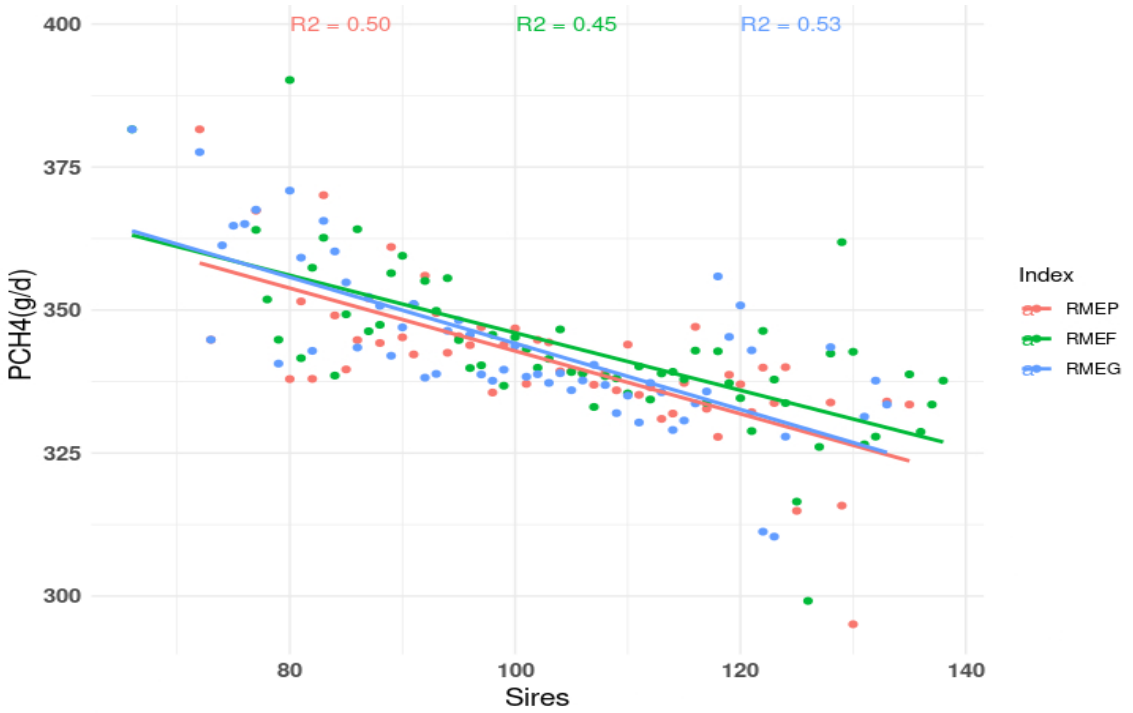


Figure 3. Relationship between sire methane-efficiency indexes and MIR-predicted methane emissions (PCH4) pf daughter groups. Lines represent linear regressions for the three alternative methane-efficiency indexes; corresponding coefficients of determination (R^2) are shown above the graph (Chen et al., 2025).

than simply favoring animals with lower methane output because of lower production.

Alternative definitions of methane efficiency can therefore lead to somewhat different sire rankings. Their relationship with observed PCH4 of daughter groups is shown in Figure 3.

Furthermore, comparisons with foreign methane evaluations based on sniffer and GreenFeed technologies showed positive indications of consistency between breeding values despite substantial differences in phenotyping approaches. These results provide indirect evidence that MIR-predicted methane phenotypes capture biologically meaningful variation associated with methane production and may serve as a useful proxy phenotype for large-scale genetic evaluation (Chen et al, 2025).

Heat Stress Response as an Advanced Prototype

Heat stress response illustrates the extension of the framework beyond environmental mitigation towards climate adaptation. We present here the ongoing developments as reported by Lemal et al. (2024, 2026). Figure 4. illustrates the ability of the developed HSR indicator to detect that animals are in heat stress generating conditions.

Preliminary analyses indicate that genetic variability exists for HSR and that genomic evaluations can be produced with practical

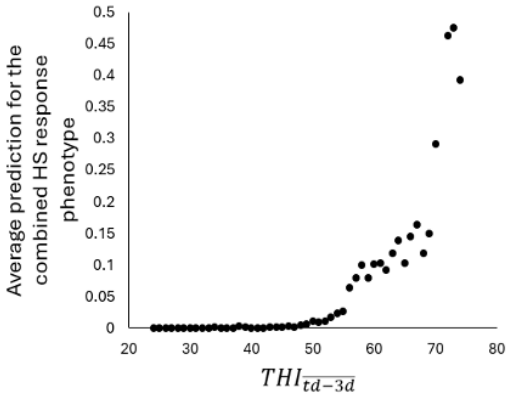


Figure 4. MIR-predicted combined heat stress response compared to observed THI 3 days before test-day (Lemal et al., 2026).

reliability levels. Approximate correlations derived from genomic breeding values suggest modest associations with production traits (0.15 for milk, 0.04 for fat and −0.14 for protein) and more pronounced relationships with functional traits such as longevity (−0.33) and fertility (−0.39). Although preliminary, these results support further investigation of HSR within future breeding objectives.

The growing availability of sensor-derived data may further improve the definition and validation of heat stress phenotypes. Sensor-derived behavioral and activity data are expected to play a particularly important role in both validating reference heat stress phenotypes and improving the timeliness of detecting responses to thermal challenges (Lemal et al., 2024).

Towards Integrated Sustainability Breeding Goals

Methane emissions, feed efficiency, welfare indicators, adaptation, and resilience should not be viewed as independent objectives.

The unified Walloon framework aims to create a portfolio of compatible genomic evaluations that can ultimately be combined through selection indexes according to future economic, environmental, and societal priorities.

In this context several very innovative local research efforts are ongoing. As an example, we can cite here Chen et al. (2026).

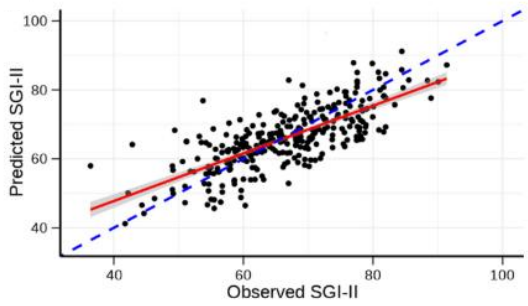


Figure 5. Random regression test-day model (RR-TDM) predicted versus observed "individual welfare score". The shaded area represents the 95% CI. (Chen et al., 2026).

Current general developments include the estimation of SNP effects and the derivation of direct genomic values for selected traits. These technical advancements directly facilitate data exchanges with external partners, thereby supporting collaborative genomic evaluations beyond regional populations. Consequently, this approach opens new pathways for international collaboration through common trait definitions and genomic exchanges, establishing a solid framework for future contributions to Interbull activities focused on emerging traits.

Conclusions

The Walloon breeding program has established a unified framework for implementing MIR-based novel traits in routine genomic evaluations. While methane emissions represent the first fully operational application, heat stress response has reached an advanced prototype stage, with further developments underway for welfare, efficiency, and resilience indicators. By combining routine MIR phenotyping, ssGBLUP evaluations, and digital data integration, this methodology provides a highly effective pathway to accelerate the deployment of sustainability traits in dairy cattle. Ultimately, this proposed framework can serve as a scalable bases for future international strategies aiming to seamlessly integrate mitigation, adaptation, and resilience into modern genomic selection programs.

Acknowledgments

The authors gratefully acknowledge the Walloon Futurospectre consortium (CRA-W, elevéo, Milk Committee) and the WALLeSmart initiative. We thank the dairy farmers of Wallonia and the wider European networks for providing herd data. We acknowledge Elevéo (awé groupe) for providing phenotypic, genealogical, and genomic data. We thank the Walloon Agricultural Research Centre (CRA-

W) and the Comité du Lait for their technical cooperation. This work was supported by the Public Service of Wallonia (SPW). Funding was received from the National Fund for Scientific Research (F.R.S.-FNRS) through grants T.W005.23 (WEAVE-DFG project HTwoTHI) and T.0095.19 (PDR project DEEPSELECT). This research was also funded by the European Interreg IVB North-West Europe program through the OptiMIR project, and the Interreg NWE program through the HoliCow project (grant NWE0100132).

References

Chen, Y., Atashi, H., Wijnrocx, K., Vandewalle, M., Bérat, H., Vanlierde, A., Dehareng, F., Soyeurt, H., Gillon, A. and Gengler, N. 2025. Advancing genomic evaluation for methane efficiency in Walloon Holstein cattle towards implementation. *Interbull Bull.* 61, 182-188. https://journal.interbull.org/index.php/ib/article/view/1974/1962

Chen, Y., Atashi, H., Lemal, P., Wijnrocx, M. N., Tran, M. N., Evrard, J., Christiaens, B. and Gengler, N. 2026. Prediction of welfare scores throughout lactation by random regression test-day models in Walloon Holstein dairy cows. *J. Dairy Sci.* 109(6), 6376-6386. https://doi.org/10.3168/jds.2025-27722.

Grelet, C., Fernández Pierna, J.A., Dardenne, P., Baeten, V. and Dehareng, F. 2015. Standardization of milk mid-infrared spectra from a European dairy network. *J. Dairy Sci.* 98(4), 2150-2160. https://doi.org/10.3168/jds.2014-8764

Grelet, C., Dardenne, P., Soyeurt, H., Fernández Pierna, J.A., Vanlierde, A., Stevens, F., Colinet, F., Becker, G., McParland, S., Lewis, E., Dehareng, F. and Gengler, N. 2021. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. *Methods.* 186, 97-111. https://doi.org/10.1016/j.ymeth.2020.07.012

Lemal, P., Tran, M-N., Atashi, H., Schroyen, M. and Gengler, N. 2024. Adding behavior traits to select for heat tolerance in dairy cattle. JDS Communications 5(5), 368–373. https://doi.org/10.3168/jdsc.2023-0421.

Lemal, P., Grelet, C., Dehareng, F., Soyeurt, H., Schroyen, M. and Gengler, N. 2026. Prediction of heat stress response in dairy cows using milk mid-infrared spectra. *Sci. Rep.* 16, 14258. https://doi.org/10.1038/s41598-026-39287-9

Vanderick, S., Wijnrocx, K., Vandenplas, J., Gillon, A. and Gengler, N. 2025. Implementation of a single-step genomic evaluation system for dairy cattle in Wallonia, Belgium. *Interbull Bull.* 61. https://journal.interbull.org/index.php/ib/article/view/1961/1982

Vanlierde, A., Dehareng, F., Gengler, N., Froidmont, E., McParland, S., Kreuzer, M., Grandl, F., Kuhla, B., Lund, P., Olijhoek, D., Eugène, M., Martin, C., Bell, M.J. and Soyeurt, H. 2021. Improving robustness and accuracy of predicted daily methane emissions of dairy cows using milk mid-infrared spectra. *J. Sci. Food Agric.* 101(8), 3394-3403. https://doi.org/10.1002/jsfa.10969

Genetic residual feed intake and its genetic association to other breeding goal traits

K.A. Bakke[1], L.S. Eikje[1], G.F. Difford[2] and B. Heringstad[1,2]

[1] Geno Breeding and A.I. Association, Hamar, Norway.

[2] Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway.

Corresponding author: karoline.bakke@tine.no

Abstract

Selecting for improved feed efficiency in dairy cows is relevant for optimising the utilization of local feed resources, improving sustainability, while reducing the climate footprint from dairy production. The current study analyses feed efficiency based on real data on feed intake and methane in Norwegian Red cows (NR). The goal is to include feed efficiency in the current NR breeding program. Genetic residual feed intake (gRFI) is a measure of the difference between actual an expected feed intake. It has been proposed as a promising feed efficiency trait as it is genetically uncorrelated with the energy driving processes such as milk production and body weight of the cow. Therefore, the objective of this study was to calculate gRFI, predict breeding values (EBVs) and estimate genetic associations between gRFI and other traits. Daily feed intake, body weight, milk yield, and methane emissions from 2,412 NR cows, with on average 433 repeated records, from 14 herds were measured between 2019 and 2024. Two multivariate linear animal repeatability models in DMU were used for estimation of (co)variance components between all traits. Genetic (co)variances for daily traits such as feed intake measured as daily dry matter intake (dDMI) or daily energy intake (dEI), milk yield or energy-corrected yield (ECM), and bodyweight were further used to obtain regression coefficients of the energy sinks (milk and bodyweight on feed intake) and compute gRFI. Heritability of gRFI was low, 0.07. The EBVs for gRFI showed variation and there were differences between animals, with BLUP breeding values ranging from -10.9 to 12.5 for dEI (daily energy-intake from both roughage and concentrate). The correlation between EBVs of gRFI and EBVs of daily body weight and milk yield were close to 0, as expected. The EBVs for gRFI was not standardized, low gRFI was therefore desirable. The correlation of gRFI EBVs to indexes from Geno's routine genetic evaluation ranged from -0.1 to 0.19 and were in general low. The strongest unfavourable correlations were found between gRFI and 305-days milk production traits and exterior. Somatic cell count, calving ease, calf size, silent heat, and number of inseminations had favourable correlations with gRFI (-0.10 to -0.06). These results indicate that improving feed efficiency defined as gRFI in the Norwegian Red population are possible, but before implementing gRFI as a novel trait in the breeding program the genetic correlations must be established and the definition should be improved in order for gRFI to be uncorrelated with current production traits in the TMI.

Key words: Feed efficiency, EBV, methane, residual feed intake

Introduction

Improving feed efficiency of dairy cows has been of great interest for many years and improving cows individual feed utilization is possible by improving the diet composition (management) or by improving genetics for feed efficiency of a population. As routine phenotype collection of feed intake is already implemented in many countries and dairy breeds, knowledge about the associations to traits of economic importance is necessary before feed efficiency or feed intake are to be included in a breeding goal. Along with this, different definitions of cows' ability to utilize feed are also important to investigate. Genetic

residual feed intake (gRFI) is an example of a trait that might be a good alternative to include in the breeding goal for improving feed efficiency of dairy cows. It is a promising efficiency trait for selection because it is genetically uncorrelated to energy driving processes in the cow and thus has no genetic correlation to milk yield and body weight. By calculating breeding values for gRFI, we can get information about cows eating less than expected according to their body weight and milk production, and they are then expected to have higher EBVs for this trait.

The aim of current study was to compute gRFI for Norwegian Red (NR) cows. The secondary aim was to predict gRFI EBVs and estimate genetic associations between gRFI and traits included in Geno's routine genetic evaluation for the NR population.

Materials and Methods

Data

The data used in this study was collected from Geno's feed efficiency database and included records from fourteen herds with daily feed intake, body weight and enteric methane records. Daily milk yield was collected from the automatic milking system. For the genetic analysis only NR dairy cows with a known proven sire from the official breeding program were included.

From the national herd recording system we collected calving dates and pedigree, and information on fat, protein and lactose content in milk from routine test days. Energy corrected milk was calculated after Sjajuna et al. (1990).

Traits

All the traits included for further analysis were defined with one record per day per cow and they were the following:

- Feed intake as sum daily dry matter intake from silage and concentrate in kg per day (dDMI)
- Feed intake as sum daily energy intake from silage and concentrate in MegaJoule NEL per day (dEI)
- Daily milk yield from AMS in kg per day (dYield)
- Average methane in grams per day (dCH_4)
- Energy corrected milk (dECM) in kg per day
- Average daily body weight in kg per day (dBW)

Feed intake records were from direct measurements of silage intake from feed bins provided by BioControl, body weight were measured either by weight scales in the milking robot or by additional scales also provided by BioControl. Methane was measured by GreenFeed provided by C-Lock inc. Milk yield was collected from either a DeLaval or Lely AMS.

Edits of data

All the traits had a phenotypic normal distribution, and records were deleted when they were outside four standard deviations from the mean. In addition to this, a visual inspection of data was done and non-logical records deleted. Each cow had at least five records for each trait. Test days within the herd with less than 10 cows with known phenotypes were removed from further analysis. The pedigree was constructed by keeping relatives on average 7 generations back in time, and only animals with breed code for Norwegian Red was included for constructing the A-matrix. The pedigree included 17 706 animals in total.

The final dataset included 1 046 629 daily records from 2019 to 2024, and 2 412 NR cows with data after data cleaning. Summary statistics of the final dataset are in table 1.

Statistical model

The statistical models used were mixed linear repeatability animal models to estimate (co)variance components and BLUP breeding values (EBVs). Two different rounds of multivariate models (containing four-traits) that differed for which feed intake and milk yield traits that were included. Variance components were estimated with DMUAI (Madsen and Jensen, 2013). The model included the fixed

Table 1: Summary statistics for traits used for calculating breeding values of gRFI for 2 412 Norwegian Red cows with records on feed intake, milk yield, body weight, and methane. The traits are measured as sum or daily average in kg.

Trait	Rec.	Avg	Sd	Min	Max
Dry matter intake (dDMI),	499 866	20.14	4.75	2	40
Yield (dYield)	814 527	28.63	10.18	5	65
Body Weight (dBW)	565 398	615	76.93	400	850
Methane (dCH_4)	257 973	411	102.2	100	800
ECM (dECM)	788 510	29.15	9.95	1.7	60
Energy Intake (dEI)	495 399	130	31.2	30	230

regression of days in milk (DIM), herd year (HY), calving age nested within parity (CA_p), and calving year and season (CYS).

We used two 4-trait models for the analysis and to estimate the (co)variance components and the heritability for the traits; dDMI, dYield, dBW, dCH_4 in the first analysis. The traits dEI, dECM, dBW, dCH_4 in the second analysis. All analyses were performed using DMU (Madsen and Jensen 2013). The model used was as follows:

$$y_{ijklm} = \mu + HY_i + DIM_j + CYS_k + CA_p_l + a_m + pe_m + htd_m + e_{ijklm},$$

where y_{ijklm} is the trait (dDMI or dEI, dYield or dECM, dBW, dCH_4), μ is the intercept, *HY* is the *i*th year within herd, *DIM* is the *j*th day in milk, *CYS* is the *k*th calving year season (2 seasons), CA_p_l is the *l*th age at calving nested within parity, a_l is the random additive effect of the *m*th animal, pe_l is the permanent environmental effect for the animal due to repeated records, htd_l is the herd-test day variance and *e* is the random residual.

The method we used to calculate genetic RFI and phenotypic RFI is similar to what is explained in Difford et al (2020), from the 4x4 additive genetic and phenotypic (co) variance matrices we retrieved genetic regression coefficients for the two "energy sinks" on feed intake: $\hat{b}_g dDMI|dYield, dBW|$ for the first model resulting in $\hat{b}_{g1}$ and $\hat{b}_{g2}$ for dYield on dDMI and dBW on dDMI, respectively. For the second model the multivariate genetic partial regression coefficients where computed for dECM on Energy intake (dEI) and for dBW on dEI: $\hat{b}_g dEI|dECM, dBW|$. The genetic regression coefficients were further used together with the phenotypic regression coefficients in a covariance function B, $B\hat{G}B'$ including *I dDMI* (or *I dEI*), *I dYield* (or *I dECM),* and *I dBW* where *I* means the identity matrix. Methane dCH_4 was not included in the gRFI calculation directly but included ascriptive to the (co)variance as a correlated or helping trait.

Genetic correlations between measured traits

The genetic correlations among the measured traits collected in this study were estimated by running multivariate model (two 4-trait models). From the multivariate models with measured traits, we calculated the heritabilities of the traits by dividing the additive genetic variance component for each trait on the total phenotypic variance that included the additive genetic variance, permanent environmental variance, and residual error variance.

Associations with other traits

We calculated correlations between EBVs for gRFI to other traits included in Genos routine breeding value estimation. For the 2 412 cows with phenotypes for traits, we collected their breeding values and used the Proc Corr procedure in SAS (2013). Indexes from one of the routine genetic evaluations in February 2025 were provided by Geno and used for calculating correlations to all indexes for all cows included in the study.

Results & Discussion

Breeding values estimation for gRFI

The breeding values for gRFI showed a gaussian distribution and ranged from -10.9 to 12.5 when gRFI were calculated from dEI (figure 2) measured as daily NEL, and from -1.65 to 2.59 kg of daily dry matter intake when dDMI (figure 1 and 2) were used for gRFI calculation. The breeding values in figure 1 are not standardized and negative gRFI breeding values must be interpreted as being favorable (eating less than expected) and positive values as being unfavorable (genetic predisposed for eating more than the expectations). We did not calculate the standard error of the predicted gRFI breeding values.

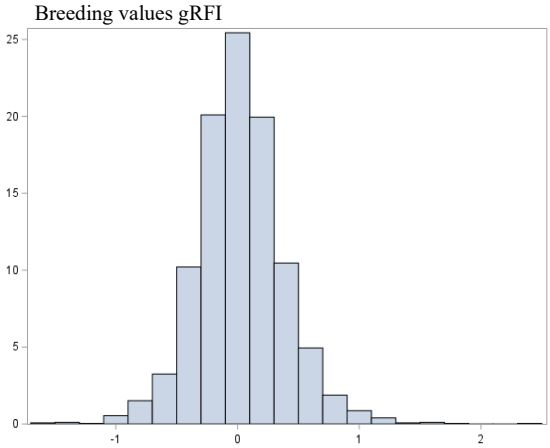


Figure 1. Distribution of BLUP breeding values of genetic residual feed intake (gRFI) for 2 412 Norwegian Red cows with feed intake records, given in kg of daily dry matter intake (dDMI).

Heritabilities

Estimated heritabilities are given in table 2. Heritability of gRFI was on 0.07. Comparing our results with reports from previous studies, gRFI heritability for NR cows was in line with other studies on other breeds. Our results are lower compared to a heritability of 14 % in the study by Difford et al. (2020), but similar to 5% in an older study by Veerkamp et al. (1995). A relatively high heritability of 0.30 to 0.38 was found in Veerkamp et al.'s (1995) study when they used the phenotypic level to estimate regression coefficients of energy requirement and calculate phenotypic residual feed intake. Our results were in the lower range of what Stephansen et al. (2023) presented in a across-country calculation with data on Holstein cows, when they used a random regression model and estimated heritabilities for gRFI within and across parities. One explanation for their higher heritability in other studies might be due to more controlled settings as cows were measured at research facilities, while our current study is based on data from commercial Norwegian farms with less standardized environments. Another explanation might be that they used non-linear (random regression) models compared to our linear models.

Table 1: Heritabilities for the measured traits, daily dry matter intake in kg (dDMI), daily energy intake (dEI), in addition to heritability for calculated daily genetic residual feed intake (gRFI).

Trait	Heritability
dDMI	0.14
dEI	0.13
gRFI	0.07

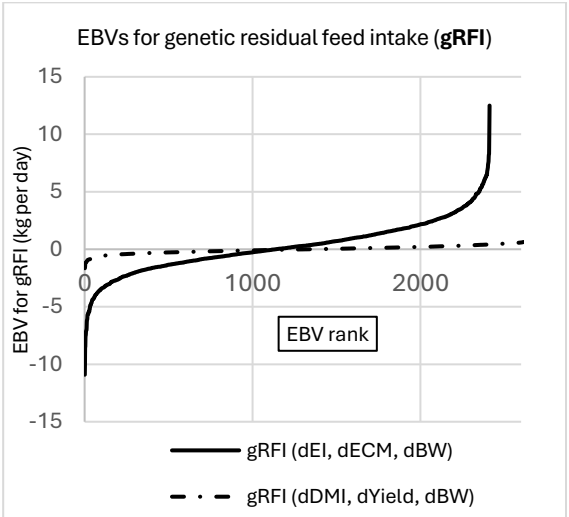


Figure 2. Breeding values (EBV) for gRFI for 2 413 Norwegian Red cows, sorted from lowest to highest rank (x-axis) and EBV on the y-axis. EBVs of gRFI calculated from daily energy intake (dEI) and energy corrected milk (dECM) - black line. EBVs of gRFI calculated from daily dry matter intake (dDMI) and "raw" milk (dYield) - dotted line.

Table 3: Genetic correlations between the measured feed intake traits included in gRFI calculations, to body weight (dBW), methane (dCH_4), milk yield (dYield), and energy corrected milk yield (dECM).

Trait	dDMI	dEI
dBW	0.66 (0.07)	0.64 (0.07)
dCH_4	0.82 (0.06)	0.80 (0.06)
dYield	0.64 (0.07)	NA
dECM	NA	0.69 (0.08)

Genetic correlations

The two measured feed intake traits, dDMI and dEI, had relatively high genetic correlations to the other "ingredients" in the gRFI calculation ranging from 0.64 to 0.82 with strongest correlation to CH_4 (table 3).

Index correlations

The correlations between breeding values for gRFI (not standardized) to traits included in the routine genetic evaluation are in table 4. Because the index correlations showed similar trends for the two different gRFI calculations (dDMI / dEI), we only give the results on index correlations for gRFI based on dDMI including dBW and dYield explaining energy driving processes. The indexes for body (size), udder conformation traits, 305-days milk production traits, calving ease and milk fever showed the strongest unfavorable genetic associations to gRFI. Udder exterior traits and 305-days yield had highest correlation to gRFI ranging from 0.13 to 0.19.

Table 4: Index correlations to gRFI calculated from dDMI in kg, traits with strongest unfavorable correlation to gRFI in Norwegian Red cows.

Trait	Correlation to EBV for gRFI
Udder Exterior traits	0.13 to 0.19
Milk yield, kg fat 305d	0.19
Milk Protein % 305d	0.18
Milk Fat % 305d	0.17
Milk yield, kg protein 305d	0.16
Body Total Score*	0.14
Ketosis (Parity 1-5)	0.08 to 0.12
Angularity*	0.11
Stature*	0.11
Calving Ease (Maternal Parity 1)	0.10
Milk Fever (Parity 3-5)	0.06 to 0.08

*Trait not included in the Norwegian total merit index. P values < .001.

Table 5: Index correlations between gRFI calculated from dDMI in kg, and traits with strongest favorable correlation to gRFI in Norwegian Red cows.

Trait	Correlation to EBV for gRFI
Somatic Cell Count (305d)	-0.10
Silent Heat (Parity 1)	-0.06 to -0.10
Calf Size (Direct)[1,2]	-0.10
Milk Leakage	-0.10
Calving Ease (Direct)	-0.07
Nmbr Inseminations (Parity 1)	-0.07

[1]Trait not included in the Norwegian total merit index

[2] Calf size, direct: high score is small calf

P values < .0001

Although these correlations are relatively weak, they are interesting and can give us an indication of the genetic correlations we can expect between current indices to gRFI. Because we used milk yield from the AMS (daily trait) for gRFI, the trait was not uncorrelated to milk production in current breeding goal. Daily AMS yield was uncorrelated with gRFI, -0.02 (0.3), as expected.

The traits with strongest favorable correlation to gRFI are given in table 5, and were direct calving traits, cow fertility, and udder health. Indexes for somatic cell count, silent heat, calf size, and milk leakage had correlations ranging from -0.06 to -0.10. The correlations were not very strong but indicate a favorable association to improved feed efficiency, and might be related to the energy status of the cow. Low reliabilities of the gRFI EBVs might be one reason for relatively weak correlations, this uncertainty can be accounted for in future work with a "Calo" correction of the EBVs and to approximate the genetic correlations between gRFI to breeding goal traits.

Conclusions

We demonstrated that genetic variation for feed efficiency defined as gRFI in NR cows exists. Udder health, calving traits, and cow fertility had favorable associations with the calculated gRFI breeding values. Although selection for

improved genetic RFI is possible, the current definition are correlated with 305 days milk production traits. Further improvement to define gRFI as independent of production traits in the breeding goal is necessary but a potential for further refining gRFI is possible, once the genetic correlations to the current breeding goal traits are established.

Acknowledgement

We would like to acknowledge Norwegian Research Fund for Agriculture and Food Industry (FFL/JA, 309451), and Innovation Norway for funding the *"Feed Efficiency project*", as well as the farmers that are project partners for contributing with data.

References

Difford, G. F. Løvendahl, P. Veerkamp, R. Bovenhuis, H. Visker, M.H.P.W. Lassen, J., and de Haas, Y. 2020. Can greenhouse gases in breath be used to genetically improve feed efficiency of dairy cows? *J. Dairy Sci*. 103 (3), 2442-2459. Doi: 10.3168/jds.2019-16966.

Madsen, P. and Jensen, J. 2013. DMU Ver. 6, rel 5.2.https://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf.

SAS Institute. 2013. SAS User's Guide: Statistics (Version 9.4). SAS Institute Inc., Cary, NC.

Sjaunja, L.O. Baevre, L. Junkkarinen, L. Pedersen, J., and J. Setälä. 1990. A Nordic proposal for an energy corrected milk (ECM) formula. *Proc. 27th Biennial Session of Int. Comm. for Anim. Recording* (ICAR), Paris, France, 156-157.

Stephansen, R.B. Martin, P. Manzanilla-Pech, C.I.V. Gredler-Grandl, B. Sahana, G. Madsen, P. Weigel, K. Tempelman, R.J. Peñagaricano, F. Parker Gaddis, K.L. White, H.M. Santos, J.E.P. Koltes, J.E. Schenkel, F. Hailemariam, D. Plastow, G. Abdalla, E. VandeHaar, M. Veerkamp, R.F. Baes, C., and Lassen, J. 2023. Novel genetic parameters for genetic residual feed intake in dairy cattle using time series data from multiple parities and countries in North America and Europe. *J. Dairy Sci*. 106 (12), 9078-9094. Doi: 10.3168/jds.2023-23330.

Veerkamp, R., G. Emmans, A. Cromie, and G. Simm. 1995. Variance components for residual feed intake in dairy cows. *Livest. Prod. Sci.* 41 (2), 111-120. Doi: https://doi.org/10.1016/0301-6226(94)00056-D.

Using sensor-derived estrus dates to improve the genetic evaluation of fertility in the New Zealand dairy industry

Melissa Stephen[1], Ahmed Ismael, Ric Sherlock, Bevin Harris
Livestock Improvement Corporation, Waikato, NZ
[1]Corresponding author: melissa.stephen@lic.co.nz

Abstract

Reproductive performance remains a cornerstone for farm profitability in New Zealand's pasture-based dairy systems. Most NZ farmers maintain a strictly annual and tightly condensed calving period to synchronize herd feed demand with seasonal pasture growth. The interval from calving to first estrus (post-partum anestrus interval; PPAI) has been recognized as a useful candidate predictor trait for fertility EBVs, but widespread phenotype collection has historically been hindered by the expense and logistical challenges associated with serial blood testing or vigilant visual observation. With the advent and rapid adoption of activity sensors, LIC now has available a large and expanding dataset of estrus events, enabling efficient and scalable derivation of a range of fertility-related traits, including PPAI. Preliminary analyses demonstrate that PPAI has a heritability of around 0.12 and exhibits a moderate genetic correlation of -0.57 with timing of pregnancy (i.e. conception success within the first 42-d of the breeding period is associated with shorter PPAI). These preliminary results support PPAI as a useful predictor phenotype for fertility EBVs. Furthermore, the availability of routinely recorded estrus dates for individual cows creates new opportunities for investigating other novel fertility phenotypes, such as length and regularity of estrous cycles, early confirmation of pregnancy and differentiation of conception-failure and early pregnancy loss. Integrating sensor-based estrus detection into the national fertility evaluation is likely to increase the rate of genetic improvement for fertility in the NZ dairy herd, driving improved reproductive performance, and therefore cow longevity.

Key words: Fertility, Cattle, post-partum, anestrus

Introduction

New Zealand's seasonal dairy production system places considerable emphasis on reproductive efficiency. Herds are typically managed to calve over a concentrated period in spring, allowing lactation-driven energy requirements to coincide with the seasonal availability of pasture (Macdonald et al., 2010). Maintaining this annual pattern requires cows to return to reproductive cyclicity quickly following calving and become pregnant within a restricted mating period. Fertility is therefore an important component of the national breeding objective (Harris, 2005). Genetic improvement of fertility is challenging, however, because conventional fertility phenotypes, including pregnancy outcomes, have low heritability and are not available until comparatively late in an animal's life (Pryce et al., 2014).

The interval from calving to the resumption of ovarian activity is an important component of reproductive performance. Cows that return to estrus earlier after calving may be on their second or third estrus cycle at the start of the seasonal breeding period, making conception to their first insemination more likely (Friggens and Labouriau, 2010). Postpartum anestrus interval (PPAI), defined here as the interval from calving to first detected estrus, may therefore provide an informative predictor of subsequent fertility outcomes. Historically, however, routine collection of PPAI phenotypes at the scale required for national genetic evaluation has been impractical. Methods such as serial progesterone testing or visual

observation of estrus have provided evidence of the value of PPAI as a predictor of fertility (Royal et al., 2002; Nyman et al., 2014), but are labor-intensive and costly, and consequently have not generated sufficiently large datasets for widespread genetic evaluation.

The increasing adoption of wearable technologies on New Zealand dairy farms provides a new opportunity to overcome this limitation. Activity-monitoring systems can detect behavioral changes associated with estrus and generate individual-animal estrus dates, allowing PPAI phenotypes to be derived. The volume of these data is increasing rapidly: the number of cows for which LIC received wearable-derived estrus records increased from approximately 42,000 in 2022 to more than 680,000 in 2025. This growing data resource provides an opportunity to evaluate reproductive physiology at a scale that was previously not feasible and to investigate whether sensor-derived phenotypes can add useful information to existing fertility evaluations.

The objective of this study was to provide a preliminary assessment of the potential for sensor-derived PPAI to improve genetic evaluation of fertility in New Zealand dairy cattle. We characterized the available PPAI phenotype, and estimated genetic covariances between PPAI and pregnancy success phenotypes (pregnancy rate within the first 42d of the seasonal breeding period) using our current dataset of wearable-derived estrus event phenotypes.

Materials and Methods

Data

The estrus data described and analyzed in this study were obtained from the Livestock Improvement Corporation (LIC) animal and herd recording database. This database contains routinely collected dairy cattle records, including calving dates, insemination dates, and milk production and composition data. Much of this information is recorded directly by New Zealand dairy farmers through LIC's herd management software 'MINDA', while additional data are received from third-party providers. Wearable-derived estrus events are one example of data supplied to LIC by third parties. Farmers using wearable technologies can elect to integrate these data with their existing MINDA herd records. In 2025, New Zealand had approximately 4.68 million dairy cows (10,370 herds), of which 681,558 (1,117 herds) had wearable-derived estrus records integrated with MINDA. In total 813,330 unique cows had wearable-derived estrus records integrated with MINDA between 2021 and 2025.

Prior to our analysis, we applied quality-control filtering to remove herds with limited wearable data or records indicative of poor data quality. Herds were excluded if the average number of recorded estrus events was fewer than five per cow (n = 345,964 cows) or if the average duration of wearable-derived records was less than 90 days (n = 355,058 cows). The biological plausibility of intervals between consecutive estrus events was also assessed. Each interval was assigned a score of 1 if it was either 16–27 days, representing a plausible interval between successive estrus events, or 60–390 days, representing a biologically plausible prolonged absence of recorded estrus events (i.e. a likely pregnancy). All other intervals were assigned a score of 0. Interval scores were averaged within cow and subsequently across cows within herd, and herds with an average score <0.5 were excluded (n = 137,630 cows). Finally, PPAI phenotypes of less than 10 days or greater than 150 days were excluded, and contemporary groups were required to contain at least five animals. Many herds failed more than one quality-control criterion. To maintain alignment with phenotypes used in the national fertility evaluation, only records from the first three parities were retained. Following these filtering steps, the final dataset comprised PPAI phenotypes from 298,335 cows.

Wearable-derived estrus event data are received from third party providers as the date

on which an animal was predicted to have been in estrus, based on changes in her behavior. Several wearable technology providers offer integration with MINDA, and the proprietary algorithms used to detect estrus events likely differ among providers. Postpartum anestrus interval (PPAI) was defined as the number of days between an animal's calving date and the date of her first recorded postpartum estrus event. PPAI records were restricted to intervals between 10 and 150 days, animals with PPAI outside these intervals had their phenotype set to missing. To increase confidence that an identified event represented a true estrus event, the first postpartum estrus event was required to be followed by a subsequent estrus event 18–24 days later.

Pregnancy rate in 42 days (PR42) is a binary trait denoting whether a cow calved within the first 42 days of her herds seasonal calving period. Pregnancy records were coded as 1 for pregnancies attained within 42d and 0 for pregnancies attained after this period. Non-pregnant cows were also coded as 0. While our analysis was restricted to herds recording with LIC, pregnancy phenotypes were obtained from the national fertility extract, which contains all phenotypes used in the national fertility evaluation. The PR42 phenotypes are calculated based on an animal's mating dates, with conception confirmed by a calving date (accounting for gestation length of the sire). Therefore, many animals with only first lactation PPAI phenotypes were missing PR42 phenotypes when our data was extracted. A total of 76,870 cows had both a PPAI phenotype and a PR42 phenotype.

The national dairy herd in NZ consists predominantly of Holstein-Friesian and Jersey and Holstein-Friesian Jersey cross cattle. Our dataset was reflective of these demographics, with 55,899 Holstein-Friesian, 11,149 Jersey, 222,526 Holstein-Friesian Jersey cross, 514 Ayrshire and 8,247 other cattle.

Genetic analysis

Phenotypic data were pre-processed in Python using the Polars library. Genetic analyses were conducted using APEX, a standalone software package for fitting linear mixed models (Boerner, 2026). Pedigree-based animal models were used to estimate genetic parameters for postpartum anestrus interval (PPAI) and pregnancy within the first 42 days of mating (PR42). Variance components and heritabilities were estimated using univariate models, while genetic and residual covariances between PPAI and PR42 were estimated using a bivariate model.

For each trait, the model was:

$$y = Xb + Za + Wp + e \qquad \text{(equation 1)}$$

where y is the vector of phenotypic observations, b is a vector of fixed effects, a is the vector of additive genetic animal effects, p is the vector of permanent environmental animal effects, and e is the vector of residual effects. Additive genetic effects were modelled using the pedigree-derived numerator relationship matrix. Fixed effects included contemporary group, age at calving, Friesian, Jersey, Holstein and Red Dairy Cattle breed proportions, pairwise breed heterosis proportions, and inbreeding coefficient.

Results & Discussion

Adoption of wearable devices on New Zealand dairy farms

Wearable technologies are becoming increasingly widespread on New Zealand dairy farms. In 2025, LIC received wearable-derived estrus records from 681,558 individual cows, up from just 42,045 cows in 2022 (Table 1). These figures likely underestimate the true scale of adoption, as they exclude herds that do not record animals with LIC or that use wearable technologies without integrating the resulting data with their LIC herd records.

The increasing availability of routinely recorded estrus events provides an opportunity to derive PPAI at a scale that has not previously been practical. Beyond PPAI, longitudinal estrus records may also enable the development of additional fertility phenotypes, including

measures of estrous cycle length and regularity, conception failure, and early pregnancy loss. These traits may provide complementary information on different components of reproductive performance and warrant further investigation.

Table 1: Count of cows with wearable-derived estrus events recorded in LIC's herd recording database, by year.

Year	Animal Count
2022	42,045
2023	231,773
2024	378,606
2025	681,558

(Co)Variance parameters of PPAI and pregnancy rate

PPAI was more heritable than PR42 ($h^2 = 0.12$ vs 0.017), while the two traits showed negative genetic and residual correlations (−0.57 and −0.18, respectively), indicating that shorter PPAI was associated with greater six-week pregnancy success (Table 2).

The heritability estimate for PPAI indicates that there is meaningful additive genetic variation in the interval from calving to resumption of estrous activity. Our result is consistent with previous heritability estimates for PPAI (Ismael et al., 2015; Lefebvre et al., 2021) and suggests that PPAI would respond to genetic selection. In contrast, the substantially lower heritability of PR42 is consistent with the generally low heritability of conventional fertility traits derived from calving, mating and pregnancy data (Amer et al., 2016; Stephen et al., 2024), which are influenced by a wide range of genetic and non-genetic factors.

The moderate negative genetic correlation between PPAI and PR42 ($r = -0.57$) indicates that animals genetically predisposed to resume estrous activity sooner after calving also tend to have superior genetic merit for pregnancy within the first 42 days of the breeding period. This relationship is biologically plausible, as earlier resumption of ovarian activity provides cows with more opportunity to establish regular estrus cycles, be detected in estrus, inseminated, and become pregnant during a restricted seasonal breeding period. Shorter PPAI may also reflect more rapid recovery of reproductive function following calving, including uterine recovery and restoration of normal hypothalamic–pituitary–ovarian activity. Consequently, some of the physiological processes contributing to a shorter PPAI may also favor subsequent conception and establishment of pregnancy (Gautam et al., 2010). The correlation was, however, substantially less than unity, indicating that PPAI and PR42 are related but genetically distinct traits. PPAI therefore captures one component of reproductive performance rather than serving as a direct replacement for PR42.

Table 2: Estimated heritabilities and correlations between postpartum anestrus interval (PPAI) and six-week pregnancy rate (PR42). Heritability estimates are presented on the diagonal, with the genetic correlation below the diagonal and the residual correlation above the diagonal.

Trait	PPAI	PR42
PPAI	**0.12**	-0.18
PR42	-0.57	**0.017**

Conclusions

These results support PPAI as a promising predictor trait for fertility evaluation in New Zealand dairy cattle. PPAI was more heritable than PR42 and showed a moderate genetic correlation with PR42. As wearable technologies continue to generate increasing volumes of routinely recorded estrus data, PPAI offers a practical opportunity to strengthen genetic evaluation for reproductive performance. Further work is required to quantify the potential improvement in breeding value accuracy by incorporating PPAI into the national fertility evaluation.

Acknowledgments

Financial support for this project has been received from the NZ Ministry of Primary Industries, SFF Futures Programme: Resilient Dairy – Innovative breeding for a sustainable dairy future (Grant Number: PGP06-17006).

We also acknowledge the New Zealand dairy farmers whose herd-recording and wearable-derived data contributed to this research.

References

Amer, P.R., Stachowicz, K., Jenkins, G.M., Meier, S. 2016. Short communication: Estimates of genetic parameters for dairy fertility in New Zealand. *J. Dairy Sci.* 99, 8227-8230. https://doi.org/10.3168/jds.2015-10429.

Boerner, V. 2026. APEX: high-performance computing solution for commercial genetic evaluation and research and development. In: World Congress on Genetics Applied to Livestock Production Digital Archive. Iowa State University Digital Press.

Friggens, N.C., Labouriau, R. 2010. Probability of pregnancy as affected by oestrus number and days to first oestrus in dairy cows of three breeds and parities. Anim. Reprod. Sci. 118, 155-162. https://doi.org/10.1016/j.anireprosci.2009.08.009

Gautam, G., Nakao, T., Yamada, K., Yoshida, C. 2010. Defining delayed resumption of ovarian activity postpartum and its impact on subsequent reproductive performance in Holstein cows. *Theriogenology* 73, 180-189. https://doi.org/10.1016/j.theriogenology.2009.08.011

Harris, B.L. 2005. Breeding dairy cows for the future in New Zealand. *N.Z. Vet. J.* 53, 384-389. https://doi.org/10.1080/00480169.2005.36582

Ismael, A., Strandberg, E., Kargo, M., Fogh, A., Løvendahl, P. 2015. Estrus traits derived from activity measurements are heritable and closely related to the time from calving to first insemination. *J. Dairy Sci.* 98, 3470-3477. https://doi.org/10.3168/jds.2014-8940

Lefebvre, R., Larroque, H., Barbey, S., Gallard, Y., Colleau, J.J., Laine, A.L., Boichard, D., Martin, P. 2021. Genome-wide association study for age at puberty and resumption of cyclicity in a crossbred dairy cattle population. *J. Dairy Sci.* 104, 5794-5804. https://doi.org/10.3168/jds.2020-18228

Macdonald, K.A., Glassey, C.B., Rawnsley, R.P. 2010. The emergence, development and effectiveness of decision rules for pasture based dairy systems. In: *Proceedings of the 4th Australasian Dairy Science Symposium*, pp. 199-209.

Nyman, S., Johansson, K., de Koning, D.J., Berry, D.P., Veerkamp, R.F., Wall, E., Berglund, B. 2014. Genetic analysis of atypical progesterone profiles in Holstein-Friesian cows from experimental research herds. *J. Dairy Sci*. 97, 7230-7239. https://doi.org/10.3168/jds.2014-7984

Pryce, J.E., Woolaston, R., Berry, D.P., Wall, E., Winters, M., Butler, R., Shaffer, M. 2014. World trends in dairy cow fertility. In: *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production,* pp. 1-10.

Royal, M.D., Pryce, J.E., Woolliams, J., Flint, A.P.F. 2002. The genetic relationship between commencement of luteal activity and calving interval, body condition score, production, and linear type traits in Holstein-Friesian dairy cattle. *J. Dairy Sci.* 85, 3071-3080.

Stephen, M.A., Burke, C.R., Steele, N., Pryce, J.E., Meier, S., Amer, P.R., Phyn, C.V.C., Garrick, D.J. 2024. Genome-wide association study of age at puberty and its (co)variances with fertility and stature in growing and lactating Holstein-Friesian dairy cattle. *J. Dairy Sci.* 107, 3700-3715. https://doi.org/10.3168/jds.2023-23963

Assessing the impact of hormonal interventions on genetic evaluation of fertility in New Zealand

K. Stachowicz[1], T. Valente[1], E.C. Ooi[1], F. Fikse[2] and S. Meier[3]

[1]AbacusBio Limited, 333 Princes Street, PO Box 5585, Dunedin 9058, New Zealand

[2]AbacusBio Int. Limited, Roslin Innovation Centre, Edinburgh EH25 9RG, United Kingdom

[3] DairyNZ Limited, Private Bag 3221, Hamilton 3240, New Zealand

Corresponding author: kstachowicz@abacusbio.co.nz

Abstract

Reproductive efficiency is crucial for productivity and profitability in the New Zealand seasonal dairy system. It requires compact calving patterns to align lactation with pasture availability. Achieving a high 6-week in-calf rate is essential for maintaining a 365-day calving interval and optimising farm efficiency. Hormonal reproductive interventions are often being used to manage anoestrus cows, synchronise oestrus in cycling cows and heifers, and improve AI submission rates. Accurate genetic evaluation of fertility requires, therefore, reliable recording of hormonal synchronisation interventions. The aim of this study was to assess if the current level of reporting of fertility interventions is adequate for inclusion in the national genetic evaluation of cow fertility, and to evaluate alternative approaches for incorporating synchronisation information into genetic evaluation. After extensive data filtering requiring consistent synchrony reporting and herd continuity, a high-quality subset was created for variance component estimation. Adjusting fertility phenotypes for synchrony through either setting synchronised records to missing or applying a penalty substantially increased heritabilities, revealing additional additive genetic variation masked by intervention use. Synchronised cows consistently exhibited poorer fertility performance, with 1.5% lower 3-week submission rates and 11% lower 6-week in-calf rates compared to non-synchronised contemporaries. Use of synchronisation itself as a new trait showed heritability of 0.07–0.1, which is significantly higher than any fertility trait evaluated currently in New Zealand. Moderate to strong negative genetic correlations between synchrony use and fertility traits (ranging from -0.38 to -0.78) demonstrate that synchronisation captures biologically meaningful information about underlying reproductive merit. Overall, the study shows that intervention data can substantially improve fertility trait modelling and breeding value accuracy. However, only 7% of cows have treatment records available in recent years, which is not sufficient for implementation. Improved industry-wide synchrony recording is essential for integrating these findings into routine genetic evaluations.

Key words: fertility, genetic evaluation, dairy cows, reproductive interventions

Introduction

Dairy cow fertility has a major influence on farm production and profitability, particularly in seasonal grazing systems. In New Zealand, compact spring calving is needed to align lactation with pasture growth and helps minimise feed costs (Verkerk, 2003; Neal and Roche, 2019). In addition, maintaining a 365-day calving interval also requires cows to re-establish pregnancy within a maximum of 12 weeks. Mating season performance is commonly assessed using 3-week submission rate and 6-week in-calf rate, which correspond to fertility traits used in genetic evaluation (Stachowicz *et al*., 2023).

The current New Zealand fertility genetic evaluation includes heifer calving season day (CSD0 – number of days from planned start of calving to calving), presented for mating in the

first 21 days of the mating season (PM21 – binary, with 0 = not presented and 1 = presented), and pregnancy rate within the first 42 days of mating (PR42 – binary, with 0 = not pregnant and 1 = pregnant). PM21 and PR42 data from the first three parities are evaluated. The breeding value of PR42 is the only fertility trait included in the Breeding Worth Index (DairyNZ, 2026).

In New Zealand, anoestrus may affect a substantial proportion of lactating cows at the beginning of the breeding season, with estimates around 20% reported in earlier studies (Rhodes *et al*., 2003). The evaluated phenotypes are intended to capture genetic differences in fertility under comparable management conditions (Stachowicz *et al*., 2023). However, hormonal reproductive interventions are commonly used to treat cows that have not resumed cyclicity before mating, synchronise oestrus and/or ovulation in cycling animals, and improve early submission rates. As a result, these interventions can alter submission and pregnancy rate phenotypes by changing the timing of artificial insemination. If hormonal intervention records are unavailable for the genetic evaluation, phenotypes from treated cows introduce bias because the model effectively considers them as untreated animals.

The objective of this study was to assess whether the current level of reporting of fertility interventions is adequate for inclusion in the national genetic evaluation of cow fertility, and to evaluate alternative approaches for incorporating synchronisation information into fertility evaluation.

Materials and Methods

Data

Data for this study were derived from the national fertility genetic evaluation extracts combined with additional fertility treatment records provided by the Livestock Improvement Corporation (LIC). The hormonal treatment records were available from the LIC health treatment dataset, which contained records from 1997, although reporting increased substantially from the 2008–2010 seasons onward. Therefore, data from 2008 were included in this study. Treatment records were merged with pregnancy (PR42) and mating (PM21) phenotypes following the data pre-processing practice currently used in New Zealand for fertility genetic evaluation extracts.

A cow was marked as synchronised in a given season if she had a treatment record that included a fertility-related hormone product category and a health condition category recorded as Non-Cycling/Anoestrus. Fertility-related products included luteolytic hormones (e.g. prostaglandin), treatments associated with fixed-time insemination (e.g. CIDRs, DIBs and GnRH), and other reproductive hormones (e.g. FSH and oestradiol).

The full initial fertility genetic evaluation dataset for the 2008–2023 seasons contained 11.7 million cows from approximately 18,200 herds. For variance component estimation, multiple filtering steps were applied to ensure that included herds recorded calving and mating data in all seasons and the majority of cows had known parents in the pedigree. Herds were retained when synchronisation was reported in at least 6 of the 16 breeding seasons. This generated a dataset with synchronisation reporting levels closer to the rates reported by Laven *et al*. (2025).

The resulting filtered dataset contained 131,775 cows from 92 herds, representing approximately 1.1% of cows and 0.5% of herds in the full 2008–2023 genetic evaluation dataset. For further analyses, a random sample of 31,537 cows from 24 herds was used for variance component estimation. Within this random sample there were 1,150 sires represented. Results presented in this paper are based on this sample.

Alternative phenotype definitions

Three alternative definitions of PM21 and PR42 phenotypes were analysed. First, the current national genetic evaluation phenotype definitions were used. Second, phenotypes were set to missing when synchronisation was

reported. Third, records were penalised by setting PM21 and PR42 to 0 when synchronisation was reported. In addition, synchronisation use was analysed directly as binary phenotypes (0 = no, 1 = yes): PSYNC1, PSYNC2 and PSYNC3, indicating whether synchronisation was reported in parities 1, 2 and 3, respectively.

Variance components estimation models

Variance component estimation was carried out using ASReml 4.2 (Gilmour *et al*., 2023). Models were based on the current national fertility evaluation framework, with fixed effects for age at calving, contemporary group (herd-year), breed composition and heterosis. Seven-trait multivariate models were fitted including CSD0, PM21 and PR42 records from parity 1 to 3. Three variants were analysed: 1) the current definition of PM21 and PR42 phenotypes; 2) PM21 and PR42 phenotypes set to missing when synchronisation was reported; and 3) PM21 and PR42 phenotypes set to 0 when synchronisation was reported. Each parity was modelled as a separate trait, as in the current national fertility evaluation, and residuals between traits and parities were modelled as correlated.

Additionally, the 7-trait model (using the current definitions for PM21 and PR42) was expanded to a 10-trait model by adding synchronisation use (PSYNC) as three additional binary traits, one for each of parity. This model retained the current fertility trait definitions while allowing synchronisation use to be analysed as a correlated trait with CSD0, PM21 and PR42.

For EBV level comparisons data from bulls with at least 50 daughters with fertility and synchronisation phenotypes were used. There were 150 such sires available for analyses.

Results & Discussion

Data suitaility and phenotypic differences

The data quality assessment showed that reproductive intervention records were incomplete at the national level. Across all cows in the fertility extracts, 93% had no treatment records and 7% had between 1 and 44 treatments recorded per lactation. For stricter fertility treatment definitions based on health product and heath condition categories, only 1.7% of cows and 0.07% of heifers had synchronisation records. These percentages of hormonal treatment were lower than the expected level of intervention use reported by Laven *et al*. (2025).

According to these authors, 60% of New Zealand farmers used interventions, and 38% reported using them every year. Assuming representativeness of survey respondents, we can infer that in the dataset available for this study, hormonal interventions are very underreported, with only 10% of herds each year having any synchrony records. The subset of data created for variance component estimation had an average percentage of herds using synchrony of 70% (range 50-80% in 16 years), i.e. close to the value reported by Laven *et al*. (2025).

Laven *et al*. (2025) noted that approximately 20-30% of cows are not detected in oestrus before the start of the breeding season and may therefore be candidates for treatment. If 60% of farms use synchrony then we would expect treatment records for around 12-15% of cows on average. In the dataset used for variance component estimation, this average was 17% in lactation 1 and 2 and 13% in lactation 3. Therefore, the subset of data generated for variance component estimation seems to be of sufficient suitability.

Although individual level recording was inconsistent and incomplete, contemporary group (herd-year) level patterns indicated that after quality control data from 2008 to 2023 were suitable for exploratory analyses. The average percentage of cows with fertility

treatment records within reporting contemporary groups generally ranged from 15% to 25% in parities 1 and 2 and was slightly lower in parity 3. Heifer data were much less suitable for analysis because few contemporary groups contained usable on-farm intervention records.

The contrast between cow and heifer data was important. Heifer intervention records were too sparse to support reliable analysis, even though heifer synchronisation may be common for management purposes. The present analyses therefore focused on lactating cows, where reporting contemporary groups were numerous enough for exploratory work leading to a better understanding of bias due to fertility interventions. This distinction also means that the conclusions should not be extended directly to heifer fertility traits without improved heifer synchronisation recording by dairy producers.

The quality control strategy used created a dataset that resembled what was reported by Laven *et al*. (2025) and, therefore, suitable for variance component estimation, but it also highlighted a key implementation limitation. Herds retained for analysis were those with more consistent synchronisation recording, and therefore they were not expected to be a random sample of the population. Consequently, the subset was appropriate for testing whether synchronisation records contained useful information, but not sufficient by itself to support immediate national implementation, which requires consistent recording practices across producers.

Phenotypic comparisons showed that synchronised cows had poorer fertility performance than non-synchronised contemporaries (Table 1). In the filtered dataset, non-synchronised cows had higher average PM21 and PR42 phenotypes in parities 1-3. Differences between synchronised and non-synchronised cows were 0.02 to 0.08 for PM21 and 0.11 to 0.12 for PR42. This is consistent with reports of how treatment and synchronisation outcomes differ depending on cycling status and treatment protocol (McDougall, 2010).

Table 1: Average PM21 and PR42 phenotypes for synchronised and non-synchronised cows in the dataset filtered for synchrony reporting in at least 6 seasons

Trait	Non-synchro-nised	Synchro-nised	Difference
PM21, parity 1	0.86	0.84	0.03
PM21, parity 2	0.84	0.79	0.06
PM21, parity 3	0.86	0.78	0.08
PR42, parity 1	0.74	0.66	0.11
PR42, parity 2	0.74	0.65	0.12
PR42, parity 3	0.76	0.68	0.09

Phenotypic statistics of the dataset used for variance component estimation are presented in Table 2. Synchronisation was reported for 17% of cows in parities 1 and 2 and 13% in parity 3. The mean PM21 values were 0.84 to 0.86 across parities, while mean PR42 values were 0.74 to 0.76, indicating that the sampled dataset retained substantial variation in both submission and pregnancy phenotypes, respectively.

When PM21 and PR42 were set to missing due to synchronisation use being reported, the average PM21 changed only slightly, which indicates that similar proportion of synchronised cows was presented and not presented for mating in first 21 days of mating season (Table 2). The average PR42 increased by 2-4% after setting phenotypes of synchronised cows to missing. This indicates that cows that do not require synchronisation have better pregnancy rates than their synchronised contemporaries. On the other hand, a greater effect was observed applying the penalty approach, i.e. phenotype equal to zero, which resulted in PM21 and PR42 averages decreasing by 10-22% and 9-19%, respectively.

Table 2: Phenotypic means of data used for variance component estimation

Trait	Current pheno-type	Synchro-nised records set to missing	Synchro-nised records penalised
PM21, parity 1	0.86	0.86	0.72
PM21, parity 2	0.84	0.85	0.70
PM21, parity 3	0.86	0.86	0.75
PR42, parity 1	0.74	0.76	0.61
PR42, parity 2	0.74	0.76	0.61
PR42, parity 3	0.76	0.77	0.66

Seven-trait model variants for synchronisation adjustment

Heritabilities of PM21 and PR42 were lower than in the official national fertility evaluation, indicating that the subset was not fully representative of the national population. However, when PM21 and PR42 phenotypes were set to missing or penalised, heritabilities slightly increased and were 0.036-0.042 for PM21 and 0.033-0.045 for PR42 (Table 3). Penalising synchronised phenotypes resulted in the largest increase, with estimates approximately doubling for PM21 and PR42 in the 7-trait model (previously 0.07-0.088 for PM21 and 0.049-0.082 for PR42). These results indicate that penalising synchronised cows allows capturing more of the additive genetic variation in fertility phenotypes. Thus, proper recording of synchronisation would further lead to more accurate estimated breeding values for fertility and faster genetic progress for those traits.

Estimated genetic correlations among the seven fertility traits also changed when synchronised PM21 and PR42 records were penalised (Table 4). Genetic correlations between CSD0 and PM21 or PR42 became less negative in the penalised variant than in the current model. In contrast, correlations within and between the PM21 and PR42 trait groups increased when synchronised cows were penalised. For example, the correlation between first parity PM21 and PR42 increased from 0.71 to 0.81. This indicates that the penalised phenotype strengthened the genetic relationship between submission and pregnancy traits across parities, consistent with synchronisation records identifying cows with poorer underlying fertility performance. This might not be a desirable outcome as PM21 and PR42 represent different aspects of cows' fertility.

Ten-trait model with synchronisation trait

Synchronisation use itself had a measurable genetic component. In the 10-trait model, heritabilities for synchronisation use were 0.099, 0.078 and 0.057 for parities 1, 2 and 3, respectively (Table 5). These were higher than estimates for the individual fertility traits analysed in the same dataset indicating that synchronisation use contains important additive genetic signal that are relevant to fertility evaluation. Heritability of synchronisation use decreased from parity 1 to parity 3, from 0.099 to 0.057, indicating that the genetic variability captured by synchronisation use was greatest in first parity and declined across later parities.

Genetic correlations between synchronisation use and CSD0 were low and positive in all three parities (Table 5). Moreover, they were consistently negative with PM21 and PR42, with parity specific estimates

Table 3: Heritability estimates for the current 7-trait model and alternative phenotypes adjusted for synchronisation

Trait	Current phen-otype	Synchro-nised records set to missing	Synchro-nised records penalised
PM21, parity 1	0.022	0.038	0.085
PM21, parity 2	0.023	0.042	0.088
PM21, parity 3	0.025	0.036	0.070
PR42, parity 1	0.031	0.045	0.082
PR42, parity 2	0.027	0.033	0.070
PR42, parity 3	0.031	0.033	0.049

Table 4: Genetic correlations from 7-trait model variants. Current model correlations are above the diagonal; penalty-based model correlations are below the diagonal

Trait	CSD0	PM21 P1	PM21 P2	PM21 P3	PR42 P1	PR42 P2	PR42 P3
CSD0		-0.42	-0.32	-0.22	-0.50	-0.27	-0.14
PM21 P1	-0.31		0.62	0.55	0.71	0.59	0.20
PM21 P2	-0.26	0.88		0.57	0.94	0.75	0.54
PM21 P3	-0.10	0.80	0.86		0.40	0.48	0.24
PR42 P1	-0.37	0.81	0.86	0.71		0.67	0.50
PR42 P2	-0.31	0.84	0.79	0.80	0.90		0.53
PR42 P3	-0.04	0.62	0.72	0.62	0.80	0.83	

of -0.60, -0.78 and -0.61 between PSYNC and PM21, and -0.61, -0.62 and -0.43 between PSYNC and PR42 in parities 1, 2 and 3, respectively. These results indicate that cows with greater genetic propensity for synchronisation use tend to have lower genetic merit for submission and pregnancy rate traits, further supporting the need for proper recording of fertility interventions and their incorporation in the national fertility evaluation.

Using synchronisation as an additional correlated trait may be more practical under incomplete recording than directly adjusting PM21 or PR42 phenotypes. The 10-trait model retained the current fertility phenotypes while allowing synchronisation use to contribute correlated information. This approach avoids assigning an automatic missing or penalty to all synchronised cows, while still allowing the evaluation to use information from recorded interventions.

Table 5: Synchronisation trait heritabilities and parity specific genetic correlations with fertility traits in the 10-trait model

Estimate	Parity 1	Parity 2	Parity 3
Heritability of PSYNC	0.099	0.078	0.057
Genetic correlation: PSYNC with CSD0	0.17	0.22	0.08
Genetic correlation: PSYNC with PM21	-0.60	-0.78	-0.61
Genetic correlation: PSYNC with PR42	-0.61	-0.62	-0.43

A remaining interpretation issue is that synchronisation records do not always identify why an intervention was carried out. Treatment of anoestrous cows, synchronisation of cycling cows for compact mating, and heifer synchronisation for management efficiency may have different implications for genetic evaluation. Future recording systems would therefore need to distinguish intervention purpose, timing and animal physiological status as far as practical.

Overall, the results support the value of synchronisation data for understanding fertility phenotypes but also show that recording quality and consistency are the main barriers to implementation. Improvements in consistent herd-season reporting, sufficient coverage within contemporary groups, and clear definitions of treatment categories are needed before intervention records could be incorporated routinely into national evaluations.

Sire EBV correlations

The sire-level EBV correlations provided an initial indication of how alternative ways of accounting for synchronisation might affect ranking. The correlation between current PR42 EBV and PR42 EBV from the model with synchronisation set to missing was 0.95, indicating little reranking. In contrast, the correlation between current PR42 EBV and PR42 EBV with synchronisation penalty was 0.87, leading to a greater degree of sire reranking when synchronised records were penalised.

Conclusions

This study showed that the current level of reporting of fertility interventions is not adequate for direct inclusion in the national genetic evaluation of cow fertility. Incomplete and inconsistent recording means that intervention information is not reliably available across herds and seasons, creating a source of potential bias in the evaluation because treated animals are effectively analysed as untreated since synchronisation use is ignored.

Among the trait definitions and models tested, the 7-trait model with synchronised PM21 and PR42 records penalised, and the 10-trait model including synchronisation use as an additional correlated trait, appear to be the most suitable approaches for further exploration. The penalty approach captured additional additive genetic variation in fertility phenotypes, while the 10-trait model retained the current fertility definitions and used synchronisation as correlated information.

The next step should be a validation study to quantify the effect of unrecorded or unused fertility intervention information on the current evaluation. This should include assessment of recording completeness across herds and seasons, minimum data requirements for reliable use, sire re-ranking under alternative models, and separation of interventions used for poor fertility from those used for management convenience. Longer term work should focus on improving industry-wide recording standards and assessing whether synchronisation information, including the development of a synchronisation use (PSYNC) EBV and the estimation of all hormonal intervention costs, could be incorporated into the National Breeding Objective or a fertility sub-index.

Acknowledgments

This work was supported by funding from New Zealand dairy farmers through the Dairy Levy, and the Ministry of Primary Industries through the Resilient Dairy project, part of the Innovative Breeding for a Sustainable Future Programme. Livestock Improvement provided the reproductive treatment data for this study.

References

Dairy NZ. 2026. Breeding Worth. Accessed 11 June 2026. https://www.dairynz.co.nz/animal/breeding-decisions/breeding-worth/

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. 2023. *ASReml User Guide Release 4.2*. VSN International Ltd., Hemel Hempstead, UK.

Laven, R.A., Cuttance, E.L. and Chambers, G. 2025. Use of hormones on New Zealand dairy farms: an analysis of the results from a survey of farmers and a survey of veterinarians. *New Zealand Veterinary Journal*.

McDougall, S. 2010. Effects of treatment of anoestrous dairy cows with gonadotropin-releasing hormone, prostaglandin, and progesterone. *J. Dairy Sci.* 93, 1944–1959.

Neal, M. and Roche, J.R. 2019. Profitable and resilient pasture-bases dairy farm businesses in New Zealand. *Anim. Prod. Sci.* 60, 169-174.

Rhodes, F.M., McDougall, S., Burke, C.R., Verkerk, G.A. and Macmillan, K.L. 2003. Invited review: Treatment of cows with an extended postpartum anoestrous interval. *J. Dairy Sci.* 86, 1876–1894.

Stachowicz, K., Ooi, E.C., Santos, B. and Amer, P.R. 2023. Conception-based fertility trait for genetic evaluation of New Zealand dairy cattle. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics* 25, 362–365.

Verkerk, G. 2003. Pasture-based dairying: challenges and rewards for New Zealand producers. *Theriogenology* 59, 553-561.

Combining environmental and genomic information in artificial neural networks for phenotypic prediction

J. Vandenplas[1], M.P.L. Calus[1], M. Schrauf[1], R. Bonifazi[1], J. ten Napel[1], T. Pook[1], M.C.A.M. Bink[2] and M. Curcic[3,4]

[1] Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands

[2] Hendrix Genetics, P.O. Box 114, 5830 AC, Boxmeer, The Netherlands

[3] Department of Ocean Sciences, Rosenstiel School of Marine, Atmospheric and Earth Science, University of Miami, Miami, FL, USA

[4] Frost Institute for Data Science and Computing, University of Miami, Coral Gables, FL, USA

Corresponding author: jeremie.vandenplas@wur.nl

Abstract

Artificial neural networks (ANNs) have received increasing attention for phenotypic and genomic prediction because they can accommodate non-linear relationships and high-dimensional inputs. Their performance has nevertheless been inconsistent when using genomic data, and many reported studies have relied on phenotypes pre-corrected for fixed and non-genetic random effects. The objective of this study was to evaluate the feasibility and predictive ability of ANNs that use environmental and genomic information jointly for phenotypic prediction, thereby avoiding phenotype pre-correction while allowing non-linear relationships among effects to be captured. Three levels of complexity were investigated. The first baseline model (ANN 1) was an ANN including a single input layer combining environmental and genomic data followed by a dropout layer with a 30 % rate, and then a fully connected (FC) layer with a single node and a linear activation function. The second model (ANN 2) consisted in a modular three-network architecture that separated genomic and environmental inputs, each with one FC layer and one node, before combining their predictions to predict phenotypes with a third ANN, also with one FC layer. The third model (ANN 3) was built on the structure of ANN 2 and included additional locally connected layers and non-linear activation functions. The predictive abilities of the ANNs were evaluated using data from the Dutch routine genomic evaluations for dry matter intake (DMI) and enteric methane (CH4) emissions, partitioned into training, validation and test subsets. As benchmark, phenotypes were also predicted with Single Nucleotide Polymorphism Best Linear Unbiased Prediction (SNPBLUP) applied on the training sets only. Modelling environmental and genomic data separately (ANNs 2 and 3) improved predictive ability relative to the single combined-input ANN 1. For DMI, ANN 3 achieved a Pearson correlation of 0.85 with the observed phenotypes for the test subset but did not exceed SNPBLUP (0.87). For CH4 and the test subset, ANN 3 attained the highest Pearson correlation (0.30), exceeding SNPBLUP (0.22). These results indicate that ANN architectures with networks dedicated to a source of information may improve phenotypic prediction, particularly for CH4, although the computational cost of our implementation with the most complex model was substantial compared to SNPBLUP.

Key words: artificial neural networks, phenotypic prediction, environmental effects, DMI, CH4

Introduction

Artificial neural networks (ANNs) have been proposed as an alternative to linear models for phenotypic and genomic prediction because they offer flexibility for modelling non-linear relationships and can accommodate large numbers of predictor variables (Pérez-Enciso

and Zingaretti, 2019; Pook et al., 2020). However, when using genomic data, their predictive performance has not always been favourable, and many published studies rely on phenotypes that were pre-corrected for environmental effects prior to model fitting (Montesinos-López et al., 2018; Waldmann, 2018; Pook et al., 2020).

This study investigated whether ANNs can be used to integrate environmental and genomic data directly for phenotypic prediction. The principal motivation was to remove the need for phenotype pre-correction while allowing potentially non-linear relationships among sources of information to be represented within the prediction model. The comparison was performed using two routine genomic evaluation datasets for the Dutch Holstein population, one for dry matter intake (DMI) and one for enteric methane (CH4) emissions. The Single Nucleotide Polymorphism Best Linear Unbiased Prediction (SNPBLUP) was used as the benchmark method.

Materials and Methods

Data sets

The predictive abilities of the ANNs were evaluated using datasets extracted from the Dutch Holstein genomic evaluations for DMI (E-40-Breeding value dry matter intake, 2022) and CH4 (E-48-Breeding value methane, 2026). After extraction, the DMI dataset included 1 415 first-parity genotyped cows with 25 153 DMI records. The records included in this study were weekly averages and had a mean of 26.59 kg dry matter per day (kg/d) and a standard deviation of 7.29 kg/d. Genotypes included 50 008 SNPs. Environmental effects included in ANNs and SNPBLUP, and therefore contributing to the phenotypic prediction, were experiment, herd, month, year, parity by age at calving, days-in-milk (DIM) fitted with a second-order Legendre polynomial, and permanent environment. Training, validation, and test sets were partitioned by herd and year of measurement, and included 1149, 122 and 144 animals and 20 388, 2253 and 2512 phenotypes, respectively.

The CH4 dataset included 7 284 first-parity genotyped cows with 166 224 sniffer CH4 records. These records were standardized and centered CH4 phenotypes that were calculated from average weekly concentrations (expressed in parts per million), with the mean and standard deviation calculated across the entire dataset. The phenotypes included in this study had a mean of 0.01 and a standard deviation of 1.01. Genotypes included 50 373 SNPs. Environmental effects included in ANNs and SNPBLUP were herd, year, week, parity by DIM fitted with a second-order Legendre polynomial, breed composition, permanent environment, and permanent environment within parity. Training, validation and test sets were defined by year of measurement, with 6453 cows with 45 995 records before 2022 assigned to training, 2675 cows with 105 442 records from 2023 to validation and 4186 cows with 14 787 records from 2024 to testing.

ANN architectures

Three ANN architectures were evaluated for phenotypic prediction. The architectures were designed to assess whether predictive ability could be improved by processing environmental and genomic data separately before combining them. The first baseline ANN (ANN 1) represented a direct approach in which both sources of information were combined at the input stage, whereas the second (ANN 2) and third ANN (ANN 3) used separate subnetworks for environmental and genomic data followed by a third network that combined their outputs. For all ANN architectures, model parameters were optimized using the Adam optimization algorithm.

The ANN 1 consisted of a single network in which the environmental and genomic inputs

were concatenated. The combined input was followed by a dropout layer with a dropout rate of 30%, and subsequently by a fully connected (FC) layer containing a single node with a linear activation function.

The ANN 2 used a modular architecture comprising three subnetworks: a genomic ANN, an environmental ANN, and a final combining ANN. The genomic ANN processed the genomic data using a dropout layer with a dropout rate of 30%, followed by a single-node FC layer with a linear activation function. The environmental ANN consisted of a single-node FC layer with a linear activation function. The outputs of the genomic and environmental ANNs were subsequently used as inputs to the combining ANN, which consisted of a single-node FC layer with a linear activation function to generate the final phenotypic prediction.

The ANN 3 retained the modular structure of ANN 2 but introduced additional layers and non-linear activation functions. The genomic ANN processed the genotypic data through a locally connected layer with three inputs per local receptive field, 32 units, and a ReLU activation function. This layer was followed by a dropout layer with a dropout rate of 30% and a single-node FC layer with a linear activation function. The environmental ANN consisted of a two-node FC layer with a ReLU activation function, followed by a single-node FC layer with a linear activation function. Finally, the combining ANN processed the outputs of the genomic and environmental subnetworks through a two-node FC layer with a Leaky ReLU activation function, followed by a single-node FC layer with a linear activation function to obtain the final phenotypic predictions.

The progression from ANN 1 to ANN 3 therefore allowed the effect of two design principles to be evaluated: first, the separation of environmental and genomic information into dedicated subnetworks and, second, the introduction of additional model complexity and non-linear relationships within and between these sources of information.

Software and benchmark comparison

The ANNs were trained on the training sets using mini-batches of 32 records. Their predictive ability was assessed at each epoch using mean squared errors (MSE) between predicted and observed phenotypes in the validation set. The model with the lowest MSE across up to 50 epochs was used for predictive ability comparison among ANNs using the three subsets. As benchmark, phenotypes of the three subsets were also predicted with SNPBLUP trained on the training sets only.

The SNPBLUP analyses were performed with MiXBLUP (Vandenplas et al., 2022), using the variance component estimates of the Dutch Holstein genomic evaluations. The ANN analyses were performed using the Fortran program *mixnn* developed by Wageningen University and Research and built on the *neural-fortran* library (Curcic, 2019). Environmental data were provided to *mixnn* as sparse incidence matrices, as constructed for mixed models, and SNP genotypes were provided in PLINK 1 binary (BED) format (Chang et al., 2015). All analyses were performed using the same computational environment, with 10 CPU cores available for each analysis.

Results & Discussion

Predictive ability of ANNs

The predictive abilities of the ANNs differed substantially according to the architecture and trait (Tables 1 and 2). Overall, ANN 1, which directly combined environmental and genomic information in a single input layer, showed the poorest predictive ability among the ANN approaches. This pattern was particularly clear for DMI, for which ANN 1 had a considerably higher MSE than ANN 2 and ANN 3 across all subsets. For example, on the test set, the MSE was 49.19 kg^2/d^2 for ANN 1, compared with

14.80 kg^2/d^2 and 13.36 kg^2/d^2 for ANN 2 and ANN 3, respectively (Table 1). Similarly, the Pearson correlation on the test set was only 0.46 for ANN 1, compared with 0.83 and 0.85 for ANN 2 and ANN 3, respectively (Table 2). Therefore, these results suggest that separating environmental and genomic information into dedicated subnetworks can improve predictive performance.

Furthermore, additional flexibility allowed by the subnetworks could be beneficial for phenotypic prediction. For DMI, ANN 3 showed an improvement over ANN 2 on the validation and test sets, with correlations of 0.85 compared with 0.82 and 0.83, respectively (Table 2). The corresponding test set MSE values were 13.36 kg^2/d^2 for ANN 3 and 14.80 kg^2/d^2 for ANN 2 (Table 1). A similar advantage was observed for CH4. On the test set, ANN 3 showed the best performance among the ANN architectures, with an MSE of 0.77 $units^2$ and a correlation of 0.30. In comparison, ANN 2 had an MSE value of 0.93 $units^2$ and a correlation of 0.17, respectively. Thus, although modelling environmental and genomic data separately in ANN 2 improved substantially the predictive ability over combining both sources in a single network, the additional flexibility included in ANN 3 provided an additional advantage for phenotypic prediction.

Comparison with SNPBLUP

For DMI, SNPBLUP provided the best overall predictive performance. On the test set, SNPBLUP achieved a correlation of 0.87 and a MSE of 12.58 kg^2/d^2, compared with a correlation of 0.85 and MSE value of 13.36 kg^2/d^2 for the best ANN (ANN 3; Tables 1 and 2). Thus, although the modular ANN architectures substantially improved upon ANN 1, neither ANN approach exceeded

Table 1. Mean square errors between observed and predicted phenotypes for dry matter intake (DMI) and enteric methane emissions (CH4).

Trait	Subset	SNPBLUP[1]	ANN[2] 1	ANN 2	ANN 3
DMI	Training	4.29	44.24	5.57	6.78
	Validation	11.69	42.90	15.72	13.30
	Test	12.58	49.19	14.80	13.36
CH4	Training	0.32	1.25	1.00	0.56
	Validation	1.03	1.51	0.89	0.91
	Test	1.22	1.47	0.93	0.77

[1] SNPBLUP = Single Nucleotide Polymorphism Best Linear Unbiased Prediction.
[2] ANN = Artificial Neural Network

Table 2. Pearson correlations between observed and predicted phenotypes for dry matter intake (DMI) and enteric methane emissions (CH4).

Trait	Subset	SNPBLUP[1]	ANN[2] 1	ANN 2	ANN 3
DMI	Training	0.96	0.65	0.95	0.94
	Validation	0.87	0.47	0.82	0.85
	Test	0.87	0.46	0.83	0.85
CH4	Training	0.86	0.50	0.60	0.75
	Validation	0.30	0.20	0.30	0.31
	Test	0.22	0.16	0.17	0.30

[1] SNPBLUP = Single Nucleotide Polymorphism Best Linear Unbiased Prediction.
[2] ANN = Artificial Neural Network.

SNPBLUP for DMI. The difference between ANN 3 and SNPBLUP was nevertheless relatively small on the validation and test sets.

For CH4, the results differed from those obtained for DMI. ANN 3 achieved the highest predictive ability on the test set, with a correlation of 0.30, compared with 0.22 for SNPBLUP (Table 2). ANN 3 also had the lowest test set MSE (0.77 units^2), compared with 1.22 units^2 for SNPBLUP (Table 1). On the validation set, however, ANN 3 and SNPBLUP showed similar predictive ability, with correlations of 0.31 and 0.30, respectively (Table 2), and comparable MSE values of 0.91 units^2 and 1.03 units^2 (Table 1). These results indicated that the flexible ANN architecture was particularly advantageous for predicting CH4 phenotypes.

The contrasting results for DMI and CH4 suggested that the benefit of increased model flexibility depends on the trait and prediction scenario. For DMI, the SNPBLUP approach remained competitive and slightly outperformed the ANNs on independent data. In contrast, ANN 3 provided improved prediction for CH4, suggesting that the relationships between the available environmental and genomic information and CH4 phenotypes may benefit from the additional flexibility provided by non-linear and locally connected layers. However, differences in the definition and size of the training, validation and test sets between the two traits, as well as differences in the trait definitions themselves (e.g., CH4 was centered and scaled), should also be considered when comparing results across traits.

Computational performance

The improvement in predictive ability obtained with the more flexible ANNs was accompanied by a substantial increase in computational cost for the training phase. For DMI, wall clock times were 111 seconds (s) for SNPBLUP, 1 455 s for ANN 1, 1 393 s for ANN 2 and 55 708 s for ANN 3. For CH4, the corresponding wall clock times were 741 s, 12 488 s, 12 114 s and 220 859 s. These results underline the trade-off between predictive ability and feasibility. Although ANN 3 provided the strongest overall ANN performance, its computational burden for the training phase was greater than that of the simpler ANN alternatives and SNPBLUP.

Perspective

The results of this study showed the potential of ANNs to predict phenotypes directly from environmental and genomic information, without requiring phenotypes to be pre-corrected before training the ANN. In particular, the results suggest that a modular architecture, in which environmental and genomic information are processed separately before being combined, may provide a useful framework for integrating heterogeneous sources of information.

The extent to which additional network complexity improves prediction appears to depend on the trait and the data structure. While the most complex architecture provided only limited improvements for DMI and did not outperform SNPBLUP, it achieved the highest predictive ability for CH4 on the test set. Further research should therefore investigate which characteristics of traits and datasets determine when more flexible ANN architectures provide an advantage over conventional linear approaches.

An important limitation of the present implementation was the substantial computational cost associated with ANN 3. Future work should therefore focus on improving computational efficiency and systematically evaluating alternative architectures and hyperparameter settings. In addition, the modular framework could be extended to incorporate other sources of information relevant to phenotypic prediction. Such developments could help identify situations in which the additional flexibility of ANNs provides sufficient gains in predictive performance to justify their computational requirements.

Conclusions

ANNs can be applied to phenotypic prediction from environmental and genomic information without relying on pre-corrected phenotypes. The most favourable results were obtained when separate ANNs were assigned to different sources of information, and their predictions were combined in a final network. This modular strategy improved predictive ability relative to a single combined-input network and showed the greatest potential for CH4 prediction. The principal limitation of the most flexible architecture in this study was its substantial computational cost for the training phase.

Acknowledgments

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food Project LWV20054) and the Breed4Food partners CRV (Arnhem, the Netherlands), Hendrix Genetics (Boxmeer, the Netherlands), and Topigs Norsvin (Den Bosch, the Netherlands). The use of the HPC cluster has been made possible by the RegioDeal Foodvalley (through Shared Research Facilities Wageningen UR).

References

Chang, C.C., C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, and J.J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7. https://doi.org/10.1186/s13742-015-0047-8.

Curcic, M. 2019. A parallel Fortran framework for neural networks and deep learning. arXiv:1902.06714.

E-40-Breeding value dry matter intake. 2022. Accessed September 14, 2026. https://www.cooperatie-crv.nl/wp-content/uploads/2023/06/E_40-Voeropname-Oktober-2022-Engels.pdf

E-48-Breeding value methane. 2026. Accessed September 14, 2026. https://www.cooperatie-crv.nl/wp-content/uploads/2026/04/E-chapter-methane-april-2026.pdf.

Montesinos-López, O.A., J. Martín-Vallejo, J. Crossa, D. Gianola, C.M. Hernández-Suárez, A. Montesinos-López, P. Juliana, and R. Singh. 2018. A Benchmarking Between Deep Learning, Support Vector Machine and Bayesian Threshold Best Linear Unbiased Prediction for Predicting Ordinal Traits in Plant Breeding. *G3: Genes Genomes Genet.* g3.200998.2018. https://doi.org/10.1534/g3.118.200998.

Pérez-Enciso, M., and L.M. Zingaretti. 2019. A Guide on Deep Learning for Complex Trait Genomic Prediction. *Genes* 10:553. https://doi.org/10.3390/genes10070553.

Pook, T., J. Freudenthal, A. Korte, and H. Simianer. 2020. Using Local Convolutional Neural Networks for Genomic Prediction. *Front. Genet.* 11:561497. https://doi.org/10.3389/fgene.2020.561497.

Vandenplas, J., R. f. Veerkamp, M. p. l. Calus, M. h. Lidauer, I. Strandén, M. Taskinen, M. Schrauf, and J. ten Napel. 2022. MiXBLUP 3.0 - software for large genomic evaluations in animal breeding programs. Pages 1498–1501 in Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP). Wageningen Academic Publishers. https://edepot.wur.nl/586651

Waldmann, P. 2018. Approximate Bayesian neural networks in genomic prediction. *Genet. Sel. Evol.* 50:70. https://doi.org/10.1186/s12711-018-0439-1.

A Machine Learning Framework for Concurrent Identification of Favorable, Deleterious, and Lethal Variants in Bovine Populations

N. Blair[1], D. Null[1], J. Cole[1,2,3,4], and E. Nicolazzi[1]

[1] Council on Dairy Cattle Breeding, Bowie, Maryland, 20716, USA

[2] Department of Animal Science, North Carolina State University, Raleigh, North Carolina, 27695-7621, USA

[3] Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Program, and the Genetics Institute, University of Florida, Gainesville, Florida, 32611-0910, USA

[4] Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

Corresponding author: john.cole@uscdcb.com

Abstract

Accurate genomic screening in livestock is essential for managing a complex spectrum of traits, including favorable production-related alleles, deleterious variants, and recessive lethal alleles. However, most organizations rely on legacy analytic pipelines that tend to be computationally slow and which suffer from high technical debt. These legacy systems are often difficult to maintain, lack modularity, and create significant bottlenecks as marker panels expand. We developed a high-throughput machine learning framework for identifying carrier status for many different haplotypes which will replace a legacy SAS program. By transitioning to a modular, Python-based architecture utilizing XGBoost and Polars, we streamlined the codebase, significantly reducing maintenance overhead and improving system extensibility. The new framework uses an automated feature-engineering pipeline to process genotype data, allowing for the concurrent classification of multiple favorable, sub-lethal deleterious, and recessive lethal variants in a single, reproducible environment. The ML-based approach demonstrated robust performance across all haplotypes tested, achieving an average concordance of 99.8% relative to the legacy system. In addition to matching the accuracy of the legacy system, the framework produced comparable runtimes on traditional CPU hardware and significantly reduced runtimes on modest GPU hardware. Crucially, the modernization of the codebase resulted in a 50% reduction in lines of code, enabling markedly faster updates as new genetic markers are discovered and integrated into the screening panel. Modernizing genomic screening with machine learning provides high-performance trait identification and a sustainable, maintainable engineering foundation with minimal drawbacks. By reducing technical debt and increasing throughput, this framework allows genomic labs to pivot quickly to new research requirements while maintaining rigorous operational standards for population health and genetic progress.

Key words: genomic screening, machine learning, XGBoost, haplotype carrier identification, dairy cattle, polled, recessive lethal haplotype

Introduction

Genomic screening is a routine component of dairy cattle breeding programs, enabling identification of animals which carry alleles of economic, health, or fertility relevance before mating and selection decisions are made. Variants of interest cover a spectrum:

favorable alleles such as polled (hornlessness), deleterious recessive conditions which reduce viability or performance, and recessive lethal haplotypes that cause embryonic loss when inherited in the homozygous state.

The Council on Dairy Cattle Breeding (CDCB; Bowie, Maryland, USA) currently assigns haplotype carrier status from an animal's genotype using a custom SAS-based software package called Hapcheck. Hapcheck, which was developed by scientists in the Animal Genomics and Improvement Laboratory of the United States Department of Agriculture (Beltsville, Maryland, USA), combines genotype and pedigree information to determine carrier status and has served reliably for years, but it has accrued substantial operational limitations. It is sensitive to segmentation in the input data; predictions for newly discovered haplotypes typically require additions to the SNP list and re-imputation of all genotypes; haplotypes are assigned serially, one after another; and numerous special cases have accumulated in the code that are increasingly difficult to validate as the number of genotypes grows.

To address these limitations we developed HapchIQ (pronounced "hap-sheek"), a modular, Python-based machine learning framework which reproduces Hapcheck's carrier calls, reduces technical debt, and scales gracefully as the number of haplotypes tracked grows. HapchIQ uses gradient-boosted decision trees (XGBoost) to classify carrier status for each haplotype, trains an independent model for each haplotype so that predictions can be produced in parallel, and returns a calibrated probability with each call. This paper describes the framework, evaluates it against Hapcheck and, where available, against direct gene-test results for favorable, deleterious, and recessive lethal variants in multiple breeds.

Materials and Methods

Genotype data

Evaluations were performed on CDCB genotype data across the Brown Swiss (BS), Holstein (HO), Jersey (JE), and Ayrshire (AY) breeds. Analyses reported here draw on populations ranging from roughly 20,000 Ayrshire animals to millions of Holstein genotype records, with training and evaluation runs spanning December 2024 through December 2025. The genotype records come from over 60 SNP chips of varying densities and missing SNP were imputed using findhap.f90 (VanRaden et al., 2015), our standard pedigree-based imputation process (Wiggans and Carrillo, 2022).

Target variants

The screening panel includes three categories of variant. Favorable: the polled locus in Brown Swiss (BHP), Holstein (HHP), and Jersey (JHP) (e.g., Aldersey et al., 2020). Recessive lethal haplotypes: Brown Swiss Haplotype 6 (BH6; Häfliger et al., 2021), Holstein Haplotypes 1, 2, and 5 (HH1, -2, -5; VanRaden et al., 2011; Cooper et al., 2013), and others. Deleterious recessive conditions: Holstein Cholesterol Deficiency (HCD; Kipp et al., 2015), Holstein Muscle Weakness (HMW; Dechow et al, 2022), Brown Swiss Weaver Syndrome (BHW; Leipold et al., 1973), Brown Swiss Spinal Dysmyelination (BHD; Hafner et al., 1993), Brown Swiss Haplotype 14 (BH14; Häfliger et al., 2021), Ayrshire Haplotype 2 (AH2; Null et al., 2017), Ayrshire Arthrogryposis Multiplex Congenita (AHC; Agerholm et al., 2016), and Jersey Neuropathy with Splayed Forelimbs (JNS; Al-Khudhair et al., 2022).

BH6 illustrates a key challenge and opportunity. Affected embryos die, reducing fertility; the causal allele lies on chromosome 2 (Ortega et al., 2022) and is not itself on the SNP list, but it can be detected through linkage disequilibrium because it falls between two SNPs that are tracked. Polled is challenging for a different reason: multiple mutations produce hornless animals, more than one variant segregates in some breeds, and many laboratory gene tests examine only the Friesian mutation.

Legacy system (Hapcheck)

Hapcheck assigns carrier status from genotype and pedigree information. In addition to non-carrier and carrier calls it emits ambiguous status codes ("3" and "4") when a call cannot be made confidently. Because it depends on the SNP list and on imputed genotypes, incorporating a newly discovered marker requires re-imputing all genotypes, and its serial, special-case-heavy design becomes progressively harder to maintain and validate as the database grows. Hapcheck is the primary baseline against which HapchIQ is compared.

Framework architecture

HapchIQ is implemented in Python 3.12 with a modular architecture that separates data ingestion, feature engineering, model inference, and reporting. Classification uses gradient-boosted decision trees (XGBoost 3.2.0), chosen because tree ensembles perform well on the large volumes of structured, tabular genotype data involved; one-dimensional convolutional neural networks were also evaluated and showed promise but were more resource-intensive. Large genotype flat files are parsed with Polars 1.36.1 for memory-efficient processing within reasonable RAM limits.

Automated feature engineering

An automated pipeline transforms raw genotype calls into model-ready features, encoding genotypes as allele dosages and selecting the markers relevant to each haplotype. Input is typically restricted to one or two chromosomes for efficiency, though this is not required. In practice, data engineering accounts for roughly 80% of the effort, driven by the difficulty of parsing very large data files efficiently.

Model training and class imbalance

A separate classifier is trained per haplotype using Hapcheck calls (and, where available, gene-test results) as labels. Training is sensitive to class imbalance: homozygous ("2") genotypes are very infrequent relative to non-carriers and heterozygotes for some haplotypes (e.g., HH2, BHW HCD), and the algorithm can discard them as likely genotyping or imputation errors. The data was segregated into train/validation/test partitions at 81%/9%/10%, respectively. Genotype rows exhibiting extreme class imbalances were discarded altogether.

Concurrent classification and certainty

Because a separate model is produced for each haplotype, predictions for all haplotypes can be generated in parallel within a single reproducible run. Each prediction carries a calibrated probability, providing a measure of certainty that removes the need for the ambiguous "3" and "4" status codes used by Hapcheck.

Validation and concordance

HapchIQ calls were compared with Hapcheck calls and, for BH6 and polled, with direct gene-test results, both overall and per haplotype. Concordance is reported as the percentage of animals receiving matching calls. Stability was also assessed by contrasting run-to-run variation (same

training set, later evaluation) against training-set-to-training-set variation (model retrained on newer data).

Hardware and runtime benchmarking

Runtimes were compared with Hapcheck on comparable inputs. GPU hardware produced a substantial reduction in training and inference time relative to CPU, at higher hardware cost. CPU comparisons were performed on an Intel Xeon Gold 6226 with 1TB of RAM; GPU hardware was a NVIDIA RTX 4000.

Results & Discussion

Recessive lethal detection: Brown Swiss Haplotype 6

On the official SNP list, HapchIQ reproduced gene-test BH6 status with 90.7% concordance, close to the 92.2% achieved by Hapcheck only after adding new DNA markers and re-imputing all Brown Swiss genotypes, and far above the 46.5% achieved by Hapcheck on the official SNP list without re-imputation (Table 1). HapchIQ thus recovers most of the BH6 signal from the existing SNP panel — via linkage disequilibrium with the untracked causal variant — without the costly re-imputation Hapcheck requires.

HapchIQ agreed with Hapcheck's BH6 calls for the large majority of animals, with disagreements confined to a small number of cases near the classification boundary (Table 2).

Table 1: Concordance of BH6 calls from Hapcheck and HapchIQ with gene-test results.

Program	SNP list	Concordance
Hapcheck	Official	46.5%
Hapcheck	Modified[1]	92.2%
HapchIQ	Official	90.7%

[1] Adding new DNA markers to the official SNP list requires that all genotypes be re-imputed; this was done only for Brown Swiss animals, for internal testing.

Table 2: Comparison of Hapcheck (official) and HapchIQ (research) results for Brown Swiss Haplotype 6.

		Hapchiq (research)	
		Non-carrier	Carrier
Hapcheck (official)	Non-carrier	87,932	32
	Carrier	51	4,656
	Homozygous	0	0

Predictions were stable over successive evaluation runs on a fixed training set, with only a handful of status changes; larger movement appeared only when the model was retrained on a substantially newer and larger training set (Table 3).

Favorable variant: polled in Brown Swiss, Holstein, and Jersey

Polled illustrates HapchIQ's largest advantage. On the official SNP list, HapchIQ reached 99.9%, 99.1%, and 99.3% concordance in Brown Swiss, Jersey, and Holstein, respectively, whereas Hapcheck performed well only in Holstein (95.4%) and poorly in Brown Swiss (85.3%) and Jersey (63.9%). After a correction was made to the algorithm used to infer polled carrier status from animal names ("Name fix"), all methods converged to near-perfect agreement (Table 4).

Table 3: Run-to-run variation versus training-set-to-training-set variation in BH6 calls[a].

TS1	ER1	TS2	ER2	Animals	Non-carrier to carrier	Carrier to non-carrier
2412	2412	—	2504	82,553	3	0
2412	2504	—	2508	83,829	1	1
2412	2508	2512	2512	90,062	78	46

[a]Training sets (TS) and evaluation runs (ER) are identified by year and month of evaluation, e.g., "2412" is the December 2024 run.

Table 4: Concordance of polled calls by breed, SNP list, and program (BS = Brown Swiss, JE = Jersey, HO = Holstein)[a].

Breed	SNP list	Program	Matches	Mismatches	Concordance
BS	Official	Hapcheck	17,679	3,047	85.3%
BS	Official	HapchIQ	20,697	29	99.9%
BS	Name fix	Hapcheck	20,661	65	99.7%
BS	Name fix	HapchIQ	20,723	3	100.0%
JE	Official	Hapcheck	17,559	9,928	63.9%
JE	Official	HapchIQ	27,231	256	99.1%
JE	Name fix	Hapcheck	26,784	671	97.6%
JE	Name fix	HapchIQ	27,470	15	99.9%
HO	Official	Hapcheck	55,943	2,714	95.4%
HO	Official	HapchIQ	58,248	409	99.3%
HO	Name fix	Hapcheck	58,079	577	99.0%
HO	Name fix	HapchIQ	58,590	66	99.9%

[a]"Name fix" indicates improvements made to the use of information from animal names to identify polled status (e.g., "P" to indicate heterozygous polled and "PP" to indicate homozygous polled.

HapchIQ therefore extracts the polled signal from the existing panel where Hapcheck cannot, without requiring panel updates or re-imputation.

Table 5: Comparison of Hapcheck and HapchIQ results for Holstein Haplotype 1 (HH1), Holstein Haplotype 2 (HH2), Holstein Haplotype 5 (HH5), Holstein Cholesterol Deficiency (HCD), and Holstein Muscle Weakness (HMW) using August 2025 data.

Haplotype	Mismatches / genotypes	Concordance
HH1	1,233 / 9,088,709	99.99%
HH2	5,963 / 9,089,160	99.93%
HH5	2,170 / 9,088,709	99.98%
HCDy	44,410 / 8,914,901	99.50%
HMW	67,024 / 8,173,668	99.18%

Table 6: Comparison of Hapcheck and HapchIQ results for other Ayrshire Haplotype 2 (AH2), Ayrshire Arthrogryposis Multiplex Congenita (AHC), Brown Swiss Weaver Syndrome (BHW), Brown Swiss Spinal Dysmyelination (BHD), Brown Swiss Haplotype 14 (BH14), and Jersey Neuropathy with Splayed Forelimbs (JNS) , using August 2025 data.

Haplotype	Mismatches / genotypes	Concordance
AH2	109 / 20,448	99.47%
AHC	48 / 20,448	99.77%
BHW	101 / 90,081	99.89%
BHD	136 / 90,081	99.85%
BH14	104 / 90,081	99.88%
JNS	479 / 930,718	99.95%

Other haplotypes across breeds

Across additional Holstein haplotypes, HapchIQ agreed with Hapcheck 99.18–99.99% of the time (Table 5). Holstein Muscle Weakness was the lowest at 99.18%, consistent with the difficulty of haplotypes with incomplete penetrance. Concordance was similarly high for haplotypes in other breeds, ranging from 99.47 to 99.95% (Table 6).

Codebase, runtime, and maintainability

Modernization reduced the codebase by approximately 50% in lines of code relative to Hapcheck, and the per-haplotype, parallelizable design markedly lowered the effort to add new markers. GPU hardware substantially reduced runtimes relative to CPU.

Discussion

HapchIQ matches Hapcheck on well-behaved haplotypes and clearly outperforms it in situations where Hapcheck is limited by the SNP list. For polled in Brown Swiss and Jersey, and for BH6, HapchIQ recovers most of the signal from the existing panel through linkage disequilibrium, avoiding the SNP-list

updates and full re-imputation Hapcheck needs. This is the central practical advantage: new or hard-to-genotype variants can be screened without the costly re-imputation cycle.

The engineering benefits reinforce the analytical ones. A modular, Python-based design with independent per-haplotype models supports parallel execution and rapid extension; calibrated probabilities supply a certainty measure that eliminates the ambiguous “3” and “4” calls; and Polars-based ingestion handles large flat genotype files within practical RAM limits. Together these reduce technical debt and shorten the path from a newly discovered marker to a deployed screen.

Several limitations remain. Training is sensitive to class imbalance, and rare homozygous genotypes for some haplotypes (HH2, BHW, HCD) can be discarded as likely genotyping or imputation errors. Training is computationally intensive — GPUs help considerably but add cost — and the rules encoded in the decision trees can be difficult to interpret, a consideration for a regulated screening context. Haplotypes with incomplete penetrance, such as HMW (99.18%), are harder to predict. Because HapchIQ is trained largely on Hapcheck calls, it can also inherit systematic errors from that baseline where gene-test labels are unavailable.

Ongoing work addresses these issues through multi-chromosome inputs and hyperparameter tuning, and research is underway to determine whether pedigree information — used by Hapcheck but not yet by HapchIQ — can be incorporated to further improve predictions.

Conclusions

HapchIQ, a modular XGBoost-based framework using Polars for efficient data handling, predicts haplotype carrier status as well as Hapcheck in some cases and better in others, notably for polled in Brown Swiss and Jersey and for BH6 on the official SNP list. A separate model is trained for each haplotype, so predictions run in parallel, and each prediction carries a measure of certainty that eliminates the need for ambiguous “3” and “4” calls. By recovering signal from the existing SNP panel without re-imputation, halving the codebase, and enabling GPU acceleration, HapchIQ delivers high-performance trait identification on a sustainable, maintainable engineering foundation, letting genomic laboratories integrate new markers quickly while maintaining rigorous standards for population health and genetic progress. Research is underway to incorporate pedigree information into HapchIQ predictions.

Acknowledgments

The authors thank the U.S. Dairy Producers, member sectors and collaborators for contributions to this project. The support and input from an internal CDCB team were essential throughout the review process.

The mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by CDCB. CDCB is an equal opportunity provider and employer.

References

Agerholm, J.S., McEvoy, F.J., Menzi, F., Jagannathan, V., and Drögemüller, C. 2016. A CHRNB1 frameshift mutation is associated with familial arthrogryposis multiplex congenita in Red dairy cattle. *BMC Genomics* 17, 479. https://doi.org/10.1186/s12864-016-2832-x.

Al-Khudhair, A., Null, D.J., Cole, J.B., Wolfe, C.W., Steffen, D.J., and VanRaden, P.M. 2022. Inheritance of a mutation causing neuropathy with splayed forelimbs in Jersey cattle. *J. Dairy Sci.* 105, 1338–1345. https://doi.org/10.3168/jds.2021-20600.

Aldersey, J.E., Sonstegard, T.S., Williams, J.L., and Bottema, C.D.K. 2020. Understanding the effects of the bovine POLLED variants. *Animal Genet.* 51, 166–176. https://doi.org/10.1111/age.12915.

Cooper, T.A., Wiggans, G.R., VanRaden, P.M., Hutchison, J.L., Cole, J.B., and Null, D.J. 2013. Genomic evaluation of Ayrshire dairy cattle and new haplotypes affecting fertility and stillbirth in Holstein, Brown Swiss and Ayrshire breeds. *J. Dairy Sci.* 96(E-Suppl. 1):74(Poster T206).

Dechow, C.D., Frye, E., and Maunsell, F.P.. 2022. Identification of a putative haplotype associated with recumbency in Holstein calves. *JDS Comm.* 3, 412–415. https://doi.org/10.3168/jdsc.2022-0224.

Häfliger, I.M., Seefried, F.R., Spengeler, M., and Drögemüller, C. 2021. Mining massive genomic data of two Swiss Braunvieh cattle populations reveals six novel candidate variants that impair reproductive success. *Genet. Sel. Evol.* 53, 95. https://doi.org/10.1186/s12711-021-00686-3.

Hafner, A., Dahme, E., Obermaier, G., Schmidt, P., and Dirksen, G. 1993. Spinal dysmyelination in new-born Brown Swiss x Braunvieh calves. *Zentralbl Veterinarmed B* 40, 413-22. https://doi.org/10.1111/j.1439-0450.1993.tb00158.x.

Kipp, S., Segelke, D., Schierenbeck, S., Reinhardt, F., Reents, R., Wurmser, C., Pausch, H., Fries, R., Thaller, G., Tetens, J., Pott, J., Piechotta, M., and Grünberg, W. 2015. A new Holstein haplotype affecting calf survival. *Interbull Bulletin* 49, 49-53.

Leipold, H.W., Blaugh, B., Huston, K., Edgerly, CG., and Hibbs, CM. 1973. Weaver syndrome in Brown Swiss cattle: clinical signs and pathology. *Vet. Med. Small Anim. Clin.* 68, 645-647.

Null, D.J., Hutchison, J.L., Bickhart, D.M., VanRaden, P.M., and Cole, J.B. 2017. Discovery of a haplotype affecting fertility in Ayrshire dairy cattle and identification of a putative causal variant. *J. Dairy Sci.* 100(Suppl. 2):199(abstr. 205).

Ortega, M.S., Bickhart, D.M., Lockhart, K.N., Null, D.J., Hutchison, J.L., McClure, J.C., and Cole, J.B. 2022. Truncation of IFT80 causes early embryonic loss in Holstein cattle associated with Holstein haplotype 2. *J. Dairy Sci.* 105, 9001–9011. https://doi.org/10.3168/jds.2022-21853.

VanRaden, P.M., Olson, K.M., Null, D.J., and Hutchison, J.L. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* 94, 6153–6161. https://doi.org/10.3168/jds.2011-4624.

VanRaden, P.M., C. Sun, and J.R. O'Connell. 2015. Fast imputation using medium or low-coverage sequence data. *BMC Genetics* 16:82. https://doi.org/10.1186/s12863-015-0243-7.

Wiggans, G.R., and J.A. Carrillo. 2022. Genomic selection in United States dairy cattle. *Front. Genet.* 13:994466. https://doi.org/10.3389/fgene.2022.994466

Guidelines on combining genomic data across populations

R. Bonifazi[1], S. Aivazidou[1], A. Bowman[1], F. Macedo[2], R.F. Veerkamp[1], K. Quigley[3], T. Pabiou[3], R. Evans[3], H. Bunning[4], G.M. Tarekegn[5], P. Davoudi[4], K. Matthews[4], J. Layton[7], R.G.M.F. Damai[7], R. Finocchiaro[7], M. Cassandro[6,7,12], A. Quaglia[8], D. Vicario[9], L. Degano[9], M. Piazza[10], T. Roozen[2], M. Coffey[5], M. Burke[11] and J. Vandenplas[1]

[1] Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands.

[2] Interbull Centre, Department of Animal Biosciences, SLU, Box 7023, S-75007 Uppsala, Sweden.

[3] Irish Cattle Breeding Federation, Link Rd Ballincollig, P31 D452 Cork, Ireland.

[4] Agriculture and Horticulture Development Board, Middlemarch Business Park, Siskin Parkway East, CV3 4PE Coventry, England.

[5] Scotland's Rural College, West Mains Road, EH9 3JG Edinburgh, Scotland, United Kingdom.

[6] National Federation of National Breeders Associations, XXIV Maggio 43, 00187 Roma, Italy.

[7] Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana, Via Bergamo 292, 26100 Cremona, Italy.

[8] Associazione Nazionale Allevatori Bovini Italiani da Carne, Strada del Vio Viscioloso 21, 06132 Perugia, Italy.

[9] Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana, Via Ippolito Nievo 19, 33100 Udine, Italy.

[10] Associazione Nazionale degli Allevatori delle razze bovine Charolaise e Limousine Italiane, Via Ventiquattro Maggio 44/45, 00187 Roma, Italy.

[11] International Committee for Animal Recording, Arthur van Schendelstraat 550, 3511 MH Utrecht, The Netherlands.

[12] University of Padova, DAFNAE, Agripolis, viale dell'Università, 16 – 35020 – Legnaro Padova, Italy

Corresponding author: renzo.bonifazi@wur.nl

Abstract

Genomic selection has been widely implemented across several livestock species. Due to reduced genotyping costs, genomic data are being generated at increasing rates. Various genotyping arrays are available with single-nucleotide polymorphisms (SNPs) genotyping densities ranging from few hundred to hundreds of thousands of single-nucleotide polymorphisms (SNPs). Moreover, advancements in genome sequencing enable the generation of newer and more accurate reference genomes. Genotyping arrays are typically updated to remain close to newer reference genomes, thereby increasing the accuracy of genome-wide association studies and, potentially, genomic predictions. Thus, genomic data have been generated using different genotyping chips, i.e., arrays, and assemblies, creating the need to merge them into a unified map for downstream analyses. However, SNPs do not completely overlap across chips. The number of overlapping SNPs decreases even further when chips come from different manufacturers, from different organisations, which may use customised chips, or when the corresponding assembly versions differ. Moreover, even when chips refer to the same genome assembly, SNPs do not always point to the same physical positions, and metadata, such as SNP names, may differ. Therefore, merging genomic data across populations presents significant challenges which are further complicated by the discontinuation of public mapping resources. The aim of this study is to provide guidelines for combining SNP genotyping data collected across different genotyping chips and populations based on the experience of the M3GE project. More than 3.1 million individual genotypes from purebred and crossbred cattle were collected across six national breeding organisations from three countries: Ireland, Great Britain, and Italy. Genotypes were collected on 11 different genotyping chips with densities ranging from 30K to 777K SNPs. The corresponding assemblies were either UMD3.1 or ARS-UCD1.2. Genotypes were either in raw format or imputed to a custom set of SNPs as used in routine evaluations. Genotypes were collected through the Interbull Centre GenoEx-GDE database

following the Illumina A/B coding and counting the number of copies of the A allele. All genomic data were converted to the UMD3.1 assembly, validated for inconsistencies or errors in allele coding, physical positions on the genome, and allele frequency distributions, both within and across breeds. For maps alignment, we advise storing and exchanging maps together with their metadata (rsID, assembly version, and manifest), so that SNP positions can be linked or converted across chips and assemblies. For genotypes alignment, we advise using a consistent A/B allele coding and using allele frequencies as a diagnostic tool, combining correlation and root mean squared error and comparing each chip with at least two other chips. These alignment steps, together with the required file formats, are needed to ensure valid downstream analyses of combined genomic data.

Key words: genomic data, SNP chips, allele coding, cattle, multi-breed

Introduction

Genomic selection (Nejati-Javaremi et al., 1997; Meuwissen et al., 2001) has been routinely implemented in many livestock species (e.g., Wiggans et al., 2017; Georges et al., 2019; Song et al., 2023). With decreasing genotyping costs, the number of genotyped animals keeps increasing across all livestock species. Individual genotypes are being collected on different single nucleotide polymorphism (SNP) chips, i.e., genotyping arrays, with different densities, ranging from a few hundred to hundreds of thousands of SNPs. Furthermore, reference genome assemblies are updated as sequencing technologies improve. For instance, in cattle, the UMD3.1 assembly (Zimin et al., 2009) was followed by ARS-UCD1.2 (Rosen et al., 2020). Genotyping arrays are typically updated to remain close to newer reference genomes, thereby increasing the accuracy of genome-wide association studies and, potentially, genomic prediction. Therefore, genomic data available within and across organisations have been generated on different chips and refer to different assemblies. This heterogeneity creates the need to merge genomic data into a unified map for downstream analyses.

Combining genotypes collected on different chips requires identifying the same SNP across chips. However, chips only partly overlap. Higher-density chips do not necessarily include all SNPs present on lower-density chips (Nicolazzi et al., 2014). The number of overlapping SNPs decreases even further when chips come from different manufacturers, from different organisations, which may use customised chips, or when the corresponding assembly versions differ. Indeed, genomic data from different sources may differ in several ways. SNP's physical position may refer to different genome assemblies, the same SNP may have different names across genotyping chips even when mapped to the same assembly, and alleles may be coded using different formats and strands (e.g., A/B or A/C/G/T alleles, TOP/BOT or forward/reverse strands; Illumina, 2006; Nicolazzi et al., 2014). Even when two chips refer to the same assembly, the positions reported for the same SNP may not always agree due to manufacturers using different procedures to map probe sequences to the genome (e.g., Fadista & Bendixen, 2012; Schmitt et al., 2010). Errors in any of these elements do not prevent genotype files from being merged, and they may remain unnoticed until they affect downstream analyses, such as imputation and genomic predictions.

Public resources have been developed to link SNP names, positions, and allele coding across chips. For instance, SNPchiMp (Nicolazzi et al., 2014, 2015) collected information on commercial chips for six livestock species directly from the chip manufacturers and linked it to public databases, including SNP positions on the latest assembly available at the time. However, SNPchiMp

(https://webserver.ibba.cnr.it/SNPchimp/) is no longer updated and maintained.

Combining genotypes across organisations is needed, for instance, in projects requiring collaborations across countries, such as the M3GE project. The M3GE project aims to develop multi-trait, multi-breed, multi-country genomic evaluations for beef cattle, focusing on sustainability traits and small populations (Bonifazi et al., 2025). Within M3GE, genotypes of purebred and crossbred cattle were collected from national breeding organisations in three countries, on 11 different chips referring to two assemblies. Therefore, genomic data had to be combined across chips, assemblies, breeds, and countries before any downstream analysis.

The aim of this study is to provide guidelines for combining SNP genotyping data collected across different chips and multiple populations, based on the experience of the M3GE project. In particular, we describe the alignment of map information across chips and assemblies, the validation of genotype information using allele frequencies, and the resulting guidelines.

Materials and Methods

Data

Genotypes were collected within the M3GE project (Bonifazi et al., 2025) from six national breeding organisations in three countries: Ireland, Great Britain, and Italy. In total, about 3.1 million individual genotypes were collected on purebred beef cattle, beef × beef and beef × dairy crossbred animals, and purebred dairy AI selection candidates. Genotypes were available at medium (30K SNPS) to high density (777K SNPs), either in raw format or imputed to the custom set of SNPs used in the national routine evaluation. Genotypes were collected through the Interbull Centre GenoEx-GDE platform (Interbull Centre, 2025).

For each chip, a map file was collected with the SNP name, chromosome, and base-pair position, the rsID when available, and the genome assembly version. Genotypes were coded following the Illumina A/B allele coding (Illumina, 2006) and counting the number of copies of the A allele, as required by GenoEx-GDE, i.e., 0 = BB, 1 = AB, 2 = AA, and 5 = missing genotype. Table 1 summarises the files and information used in this study to combine genomic data across chips.

In total, 11 chips were collected, with SNP positions referring to either the UMD3.1 (Zimin et al., 2009) or the ARS-UCD1.2 (Rosen et al., 2020) assembly. Different breeds were genotyped on different chips, and the same chip could be used by more than one organisation. Chips and countries were anonymised (e.g., COU1_chip1), and the overlap between chips was computed using only autosomal SNP names.

To combine genomic data across chips, we ensured that i) maps were aligned across chips, and ii) genotypes were aligned with the corresponding map and allele coding was consistent across chips.

Maps alignment across chips

Two alternative approaches to align SNP information across assemblies and maps were investigated. The first approach converts positions between any two assemblies, whereas the second assigns positions on the UMD3.1 assembly only, as the SNP dictionary used was built from UMD3.1 positions.

In the first approach, the maps referring to the ARS-UCD1.2 assembly were aligned to the UMD3.1 assembly. An in-house script was developed to convert map information between any two assemblies, hereafter referred to as "source" and "target" assemblies. This first approach is hereafter referred to as the "rsID-bridging" approach. First, variant call format (VCF) files corresponding to the source and target assemblies were downloaded from Ensembl (https://jun2026.archive.ensembl.org/Bos_taurus/Info/Index). Then, for each SNP, the chromosome and base-pair position in the map file were matched to the VCF file of the source assembly to retrieve the SNP identifier (rsID).

The rsID was then searched in the VCF file of the target assembly to retrieve the chromosome and base-pair position on the target assembly (Figure 1a).

In the second approach, hereafter called the "dictionary-based approach", SNP names were first harmonised across all 11 chips by capitalising all letters. Regardless of the assembly version each chip referred to, the corresponding SNP physical positions on the UMD3.1 assembly were then assigned from a SNP dictionary based on the commercial SNP names. The dictionary was built by combining two sources of SNP positions on UMD3.1: SNPchiMp v.3 (Nicolazzi et al., 2015), with ~1.9 million SNPs, and a SNP database provided by ANAPRI (Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana, Udine, Italy), with ~0.79 million SNPs. The two sources were merged into a dictionary of unique SNP names. When the two sources reported conflicting positions for the same SNP name, the position from SNPchiMp was retained. Figure 1b summarises the workflow.

Table 1. Files and information used to combine genomic data across chips[1].

File	Information	Notes
Map file (one per chip)	SNP name; chromosome; base-pair position; rsID (if available)	• Assembly version of the positions stored next to the corresponding map
Genotype file	Genotypes coded 0, 1, 2, or 5	• Raw or imputed genotypes • Number of copies of the A allele following the Illumina A/B coding: 0 = BB, 1 = AB, 2 = AA, 5 = missing (as for GenoEx-GDE) • SNP order as in the map file
Source and target assembly VCF	CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO	• CHROM, POS, and ID; used for the rsID bridging approach (Figure 1a)
SNP dictionary	SNP name; chromosome; position on UMD3.1	• Combined from SNPchiMp v.3 and the ANAPRI database; used for the dictionary-based approach (Figure 1b)

[1] CHROM = chromosome; POS = base-pair position; ID = SNP identifier (rsID); REF = reference allele; ALT = alternative allele; VCF = variant call format.

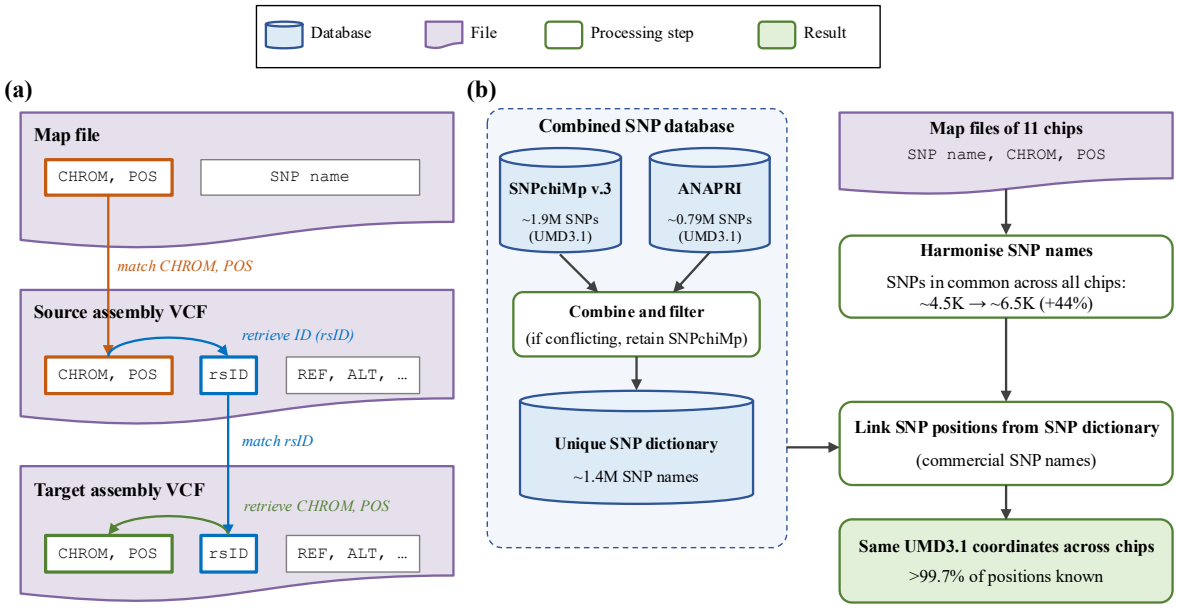


Figure 1. Workflow for the alignment of maps across chips: (a) conversion of positions between assemblies following the rsID-bridging approach; (b) harmonisation of SNP names and assignment of positions from a SNP dictionary combining SNPchiMp v.3 and the ANAPRI database. CHROM = chromosome; POS = base-pair position; ID = SNP identifier (rsID); VCF = variant call format.

Genotypes alignment with the map and allele coding

Genotype information was validated for agreement between genotypes and the corresponding map, and for consistency of allele coding across chips both within and across countries. Allele frequencies (AF) were used as a diagnostic tool. For each SNP in common between two chips, the AF of the A allele was computed as the sum of the genotype codes divided by twice the number of non-missing genotypes, using a custom AWK script to handle the large genotype files. For each pair of chips, the AF of overlapping SNPs were compared using the Pearson correlation (ρ) and the root mean squared error (RMSE), computed as: $\text{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(p_{1\text{j}} - p_{2\text{j}}\right)^2}$, where m is the number of SNPs in common between the two chips, and $p_{1\text{j}}$ and $p_{2\text{j}}$ are the AF of SNP j on the first and on the second chip, respectively. No thresholds were set on ρ or RMSE, and each comparison was evaluated by inspecting the scatter plot of AF (made using "ggplot2" package in R) together with the corresponding ρ and RMSE.

Genomic information was compared at two levels: i) within-breed, i.e., using genotypes of the same breed, either within-country (when more than one chip for a country was available) or across-country; and ii) across-breed and across-country. In all comparisons, AF were computed using different animals since not enough individuals were genotyped for more than one chip at the same time to allow for meaningful comparisons. Finally, a stronger agreement between AF was expected within breed than across breeds, with across-breed comparisons expected to be noisier but still informative.

Results & Discussion

Overlap between chips

Table 2 shows the number of autosomal SNPs on each chip and in common between each pair of chips. The number of autosomal SNPs per chip ranged from 28,289 (COU1_chip2) to 667,044 (COU3_chip2). Across all chips, there were ~815K unique SNP names, of which ~160K were present on two or more chips, and only ~6.5K were in common across all chips. The overlap between pairs of chips varied widely. For instance, COU1_chip2 shared between 9,582 and 26,952 SNPs with the other chips, while COU2_chip3 and COU2_chip4 shared 48,862 SNPs. Most SNPs present on a single chip were on the highest-density chip: 535,040 SNPs on COU3_chip2.

Maps alignment across chips

Using the rsID-bridging approach, two chips that were based on the ARS-UCD1.2 assembly (COU1_chip3 and COU3_chip2) were converted to the UMD3.1 assembly. This approach matched 92.7% and 85.7% of the SNPs for COU1_chip3 and COU3_chip2, respectively. For these two chips, respectively, 1.8% and 8.4% of the SNPs were not matched in the source VCF (ARS-UCD1.2), and 5.5% and 5.9% of the SNPs were not matched in the target VCF (UMD3.1). The rsID-bridging approach is straightforward, fast, and flexible, and it works well for chips whose SNPs have a dbSNP rsID. However, this approach relies on the availability and consistency of rsIDs across assemblies, and it requires an exact match of chromosome and position in the source VCF. Coordinate liftover tools, such as UCSC liftOver (Hinrichs et al., 2006) and CrossMap (Zhao et al., 2014), convert positions between assemblies using a pre-computed alignment of the two assemblies and do not rely on rsIDs. Therefore, liftover could be used for the SNPs that remained unmatched. However, when applied to map files that contain only chromosome and position, liftover converts positions without checking alleles, and errors in the original positions are carried over to the target assembly.

In the dictionary-based approach, the combined SNP dictionary included a total of ~1.4 million unique SNP names. Of these, 776,922 had the same position in both the

Table 2. Number of autosomal SNPs on each chip (diagonal), number of SNPs in common between pairs of chips (below the diagonal), and number of SNPs present only on that chip (unique SNPs)[1].

Chip	COU1 chip1*	COU1 chip2*	COU1 chip3	COU1 chip4	COU2 chip1	COU2 chip2	COU2 chip3	COU2 chip4	COU2 chip5	COU3 chip1	COU3 chip2
COU1_chip1*	132,665										
COU1_chip2*	26,952	28,289									
COU1_chip3	60,874	14,812	64,888								
COU1_chip4	37,420	9,582	37,555	42,172							
COU2_chip1	34,670	11,716	23,964	14,091	41,855						
COU2_chip2	37,863	10,491	38,040	37,685	14,745	44,460					
COU2_chip3	39,280	11,016	38,940	38,640	15,154	44,165	48,862				
COU2_chip4	39,280	11,016	38,940	38,640	15,154	44,165	48,862	48,862			
COU2_chip5	37,863	10,491	38,040	37,685	14,745	44,460	44,165	44,165	51,828		
COU3_chip1	37,796	10,511	37,990	37,564	14,689	44,210	43,922	43,922	50,390	50,493	
COU3_chip2	116,260	24,961	60,118	35,716	37,420	38,064	41,877	41,877	38,064	38,065	667,044
Unique SNPs	491	139	44	127	-	-	-	-	1,188	-	535,040

[1] Chips are anonymised by country (COU1 to COU3) and chip number. Counts are based on autosomal SNP names; * Chip used by more than one national breeding organisation; - = none.

SNPchiMp and ANAPRI databases, 612,189 were only present in the SNPchiMp database, 10,224 were only present in the ANAPRI database, and 1,568 had conflicting positions between the two databases. For these ~1.5K SNPs, the SNPchiMp database information was retained. Finally, a total of 46,721 SNP names had no chromosome and base-pair position in either database. After harmonising SNP names, the number of autosomal SNPs in common across all 11 chips increased from ~4.5K to ~6.5K, corresponding to an increase of 44%. SNP positions on UMD3.1 were then assigned to each chip using the built SNP dictionary by matching the SNP names. On each chip, the SNP position was known for more than 99.7% of the SNPs (Figure 1b). The main advantage of the dictionary-based approach is that the same SNP name receives the same coordinates on all chips.

Genotypes alignment with the map and allele coding

Comparing AF between chips revealed different types of misalignment. Figure 2 shows an example of an allele coding switch. The AF of chip 1 were negatively correlated with those of chip 2 for the same breed within country (Figure 2a), and with those of chip 3 for the same breed in another country (Figure 2b). In contrast, the AF of chip 2 and chip 3 agreed (Figure 2d). A negative correlation, especially when as strong as that observed in Figure 2a indicates a complete switch of the allele coding (i.e., the A allele has been counted in one chip, while the B allele was counted in another chip). However, it is not possible from a single comparison to determine which chip had the allele coding switch; comparing three chips allowed us to attribute the switch to chip 1. The issue was solved by recoding the genotypes of chip 1, i.e., 0 became 2 and vice versa. The across-breed comparison between chip 1 and chip 3 (Figure 2c) showed the same negative trend, although with a larger spread of points around the regression line, as expected for AF of different breeds. Thus, across-breed comparisons were noisier but still useful to confirm allele coding switches.

Figure 3a shows the comparison between two chips of the same breed within country, in which chip 2 was misaligned. The AF of the two chips were not correlated (ρ = 0.01), and the RMSE was 0.34. Importantly, the marginal distribution of chip 2 AF was closer to a normal distribution than a uniform distribution. Such a pattern indicated a potential misalignment between the genotypes and their corresponding map for chip 2. However, AF checks alone cannot indicate whether the error lies in the map

or in the genotypes. In this case, both map and genotypes were flagged and re-sent by the data provider.

Figure 3b shows a less evident case of misalignment, in which the maps of chip 1 and chip 2 were offset by one SNP, such that the AF of a SNP was assigned to the next SNP. Despite the shift, ρ remained high (0.90), likely because the shift led to comparing AF of adjacent SNPs, which, due to linkage disequilibrium, may have similar AF. However, the RMSE was 0.12, with many SNPs deviating from the diagonal (Figure 3b).

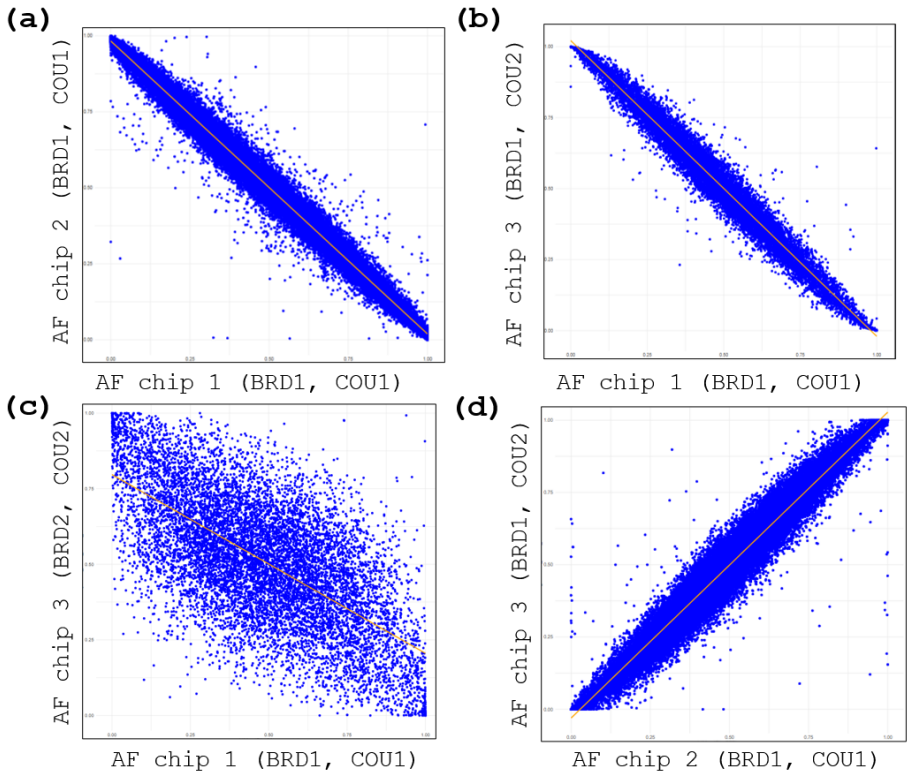


Figure 2. Scatterplot of allele frequencies (AF) for SNPs in common between pairs of chips. (a) chip 1 vs chip 2, same breed (BRD1) and country (COU1); (b) chip 1 vs chip 3, same breed (BRD1) in different countries (COU1 and COU2); (c) chip 1 vs chip 3, different breeds (BRD1 and BRD2) and countries; (d) chip 2 vs chip 3, same breed (BRD1) in different countries.

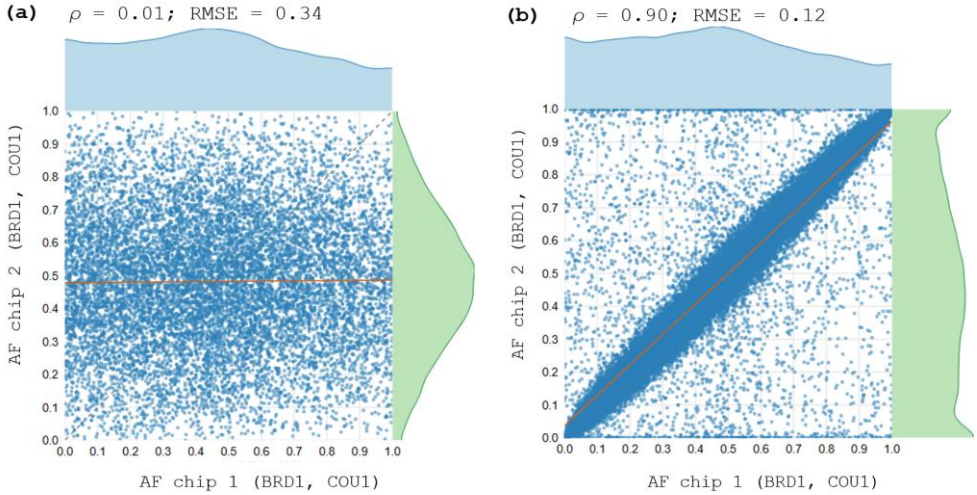


Figure 3. Scatterplot of allele frequencies (AF) of SNPs in common between two chips of the same breed (BRD1) and country (COU1): (a) example of misaligned chip 2; (b) maps of both chips offset by one SNP relative to the genotypes. Marginal distributions of AF for chip 1 and chip 2 are shown above and to the right of each panel, respectively. ρ = Pearson correlation; RMSE = root mean squared error.

Table 3. Guidelines for combining genomic data across populations[1].

Step	Guideline	Reason
Maps alignment	Use public resources when possible (e.g., SNPchiMp) to assign positions from commercial SNP names	Same coordinates across chips; >99.7% of positions known (Figure 1b)
	Store the rsID and the assembly version together with the map	Needed to convert positions between assemblies (Figure 1a)
	Keep complete manifest files of each chip	Manifest information on older chips is not always available or in the public domain
	Add support for metadata (rsID, assembly version, manifest) to GenoEx	These metadata are needed to link and convert positions across chips and assemblies (Figure 1)
Genotypes alignment	Use a consistent Illumina A/B coding, i.e., always count the same allele (A or B). A/B alleles are designated from the actual or contextual sequence of each SNP (TOP/BOT strand) and therefore remain the same even if genome assemblies or the submitted strand change (Illumina, 2006)	An allele coding switch reverses the AF of a chip (Figure 2)
	Use AF checks as a diagnostic tool, combining both ρ and RMSE	A shift by one SNP still gave $\rho = 0.90$, while RMSE = 0.12 (Figure 3b)
	Compare each chip with at least two other chips, using either genotypes from the same breed (within or across countries) or, if not available, from different, preferably related, breeds	Needed to identify which chip may cause a misalignment (Figure 2)
	Automate AF checks (ρ and RMSE) at submission, using already-approved genotypes that act as a reference, if available	Issues can be detected before genotypes are combined

[1] AF = allele frequency; ρ = Pearson correlation; RMSE = root mean squared error.

The RMSE indicated that the alignment between map and genotype was not correct. Once correctly aligned, ρ was 0.99 and RMSE was 0.04. Thus, ρ and RMSE should be used together to detect misalignments, as ρ alone may not be sufficient.

Guidelines

Table 3 summarises our guidelines for combining genomic data across populations, with the reason for each guideline based on the results of this study. The guidelines are grouped into maps alignment and genotypes alignment. For maps alignment, the main guideline is that map information should be stored and exchanged together with its metadata, i.e., rsID, assembly version, and manifest, such that positions can be linked or converted across chips and assemblies. For genotypes alignment, the main guideline is to use AF as a diagnostic tool, comparing chips for which AF are expected to agree (i.e., the same or closely related breeds), with checks done on submission based on both ρ and RMSE.

Implications

These guidelines apply whenever genotypes collected by different organisations are to be combined. This is the case for many collaborative projects, such as the development of international genomic evaluations (e.g., M3GE (Bonifazi et al., 2025), InterGenomics (Jorjani et al., 2011; Savoia, 2021) SNP-MACE (Goddard et al., 2019)), studies on relationships among different populations (Vandewalle et al., 2026), and the exchange of genotypes through GenoEx for genomic evaluations or parentage verification (Interbull Centre, 2025). Public resources, such as SNPchiMp (Nicolazzi et al., 2015) and the SNP coordinates made available by Schnabel (Schnabel, 2018), were of great help in this study. However, these resources are

either no longer updated or incomplete. Therefore, new and maintained tools are needed to link (commercial) SNP names, SNP positions, and allele coding across chips and assemblies.

Conclusions

Combining genomic data collected on different chips and assemblies across populations is challenging and prone to mistakes, which may go unnoticed after merging. In the M3GE project, harmonising SNP names and assigning positions from a common SNP dictionary provided the same coordinates across 11 chips and increased the number of SNPs in common across all chips by 44%. Comparing AF of overlapping SNPs between chips detected allele coding switches and misalignments between maps and genotypes, and highlighted that, next to visual inspections, both ρ and RMSE should be considered. Public resources were useful for this study but are outdated. New and updated tools are needed to reliably combine genomic data across populations.

Acknowledgments

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food Project LWV23062) and the M3GE partners ICAR (Utrecht, the Netherlands), SLU (Uppsala, Sweden), AHDB (Coventry, United Kingdom), SRUC (Scotland, United Kingdom), ICBF (Cork, Ireland), and FedANA (Rome, Italy), including its member organisations ANABIC (Associazione Nazionale Allevatori Bovini Italiani da Carne, Perugia, Italy), ANACLI (Associazione Nazionale degli Allevatori delle razze bovine Charolaise e Limousine Italiane, Rome, Italy), ANAFIBJ (Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana, Cremona, Italy), and ANAPRI (Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana, Udine, Italy) for providing the data. We also acknowledge the Interbull Centre (Uppsala, Sweden) for providing the infrastructure to perform the analyses.

References

Bonifazi, R., S. Aivazidou, R.F. Veerkamp, F. Macedo, R. Evans, T. Pabiou, K. Quigley, K. Matthews, H. Bunning, G.M. Tarekegn, M. Cassandro, R. Finocchiaro, J. Layton, R.G.M.F. Damai, A. Quaglia, D. Vicario, L. Degano, M. Piazza, F. Tiezzi, T. Roozen, M. Coffey, M. Burke, and J. Vandenplas. 2025. The M3GE project: developing beef cattle multi-trait multi-breed multi-country genomic evaluations for sustainability traits and small populations. *Interbull Bulletin* 61:231–241.

Fadista, J., and C. Bendixen. 2012. Genomic Position Mapping Discrepancies of Commercial SNP Chips. PLoS ONE 7:e31025. https://doi.org/10.1371/journal.pone.0031025.

Georges, M., C. Charlier, and B. Hayes. 2019. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* 20:135–156. https://doi.org/10.1038/s41576-018-0082-2.

Goddard, M.E., A. Jighly, H. Benhajali, H. Jorjani, and Z. Liu. 2019. SNPMace – A meta-analysis to estimate SNP effects by combining results from multiple countries. *Interbull Bulletin* 54:1–6.

Hinrichs, A.S., D. Karolchik, R. Baertsch, G.P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T.S. Furey, R.A. Harte, F. Hsu, J. Hillman-Jackson, R.M. Kuhn, J.S. Pedersen, A. Pohl, B.J. Raney, K.R. Rosenbloom, A. Siepel, K.E. Smith, C.W. Sugnet, A. Sultan-Qurraie, D.J. Thomas, H. Trumbower, R.J. Weber, M. Weirauch, A.S. Zweig, D. Haussler, and W.J. Kent. 2006. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res 34:D590–D598. https://doi.org/10.1093/nar/gkj144.

Illumina. 2006. Technical Note: "TOP/BOT" Strand and "A/B" Allele. Accessed July 26, 2026. https://www.illumina.com/documents/products/technotes/technote_topbot.pdf.

Interbull Centre. 2025. GenoEx - International Genotype Exchange Platform for Animal

Breeding. Accessed July 24, 2026. https://genoex.org/.

Jorjani, H., J. Jakobsen, M.A. Nilforooshan, E. Hjerpe, B. Zumbach, and V. Palucci. 2011. Genomic Evaluation of BSW Populations InterGenomics: Results and Deliverables. *Interbull Bulletin*. 43:5–8.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Nejati-Javaremi, A., C. Smith, and J.P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Animal Sci* 75:1738. https://doi.org/10.2527/1997.7571738x.

Nicolazzi, E.L., M. Picciolini, F. Strozzi, R.D. Schnabel, C. Lawley, A. Pirani, F. Brew, and A. Stella. 2014. SNPchiMp: A database to disentangle the SNPchip jungle in bovine livestock. *BMC Genom.* 15:1–6. https://doi.org/10.1186/1471-2164-15-123

Nicolazzi, E.L., A. Caprera, N. Nazzicari, P. Cozzi, F. Strozzi, C. Lawley, A. Pirani, C. Soans, F. Brew, H. Jorjani, G. Evans, B. Simpson, G. Tosser-Klopp, R. Brauning, J.L. Williams, and A. Stella. 2015. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genom.* 16:283. https://doi.org/10.1186/s12864-015-1497-1.

Rosen, B.D., D.M. Bickhart, R.D. Schnabel, S. Koren, C.G. Elsik, E. Tseng, T.N. Rowan, W.Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S.D. McKay, F. Thibaud-Nissen, J. Hoffman, B.M. Murdoch, W.M. Snelling, T.G. McDaneld, J.A. Hammond, J.C. Schwartz, W. Nandolo, D.E. Hagen, C. Dreischer, S.J. Schultheiss, S.G. Schroeder, A.M. Phillippy, J.B. Cole, C.P. Van Tassell, G. Liu, T.P.L. Smith, and J.F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9:giaa021. https://doi.org/10.1093/gigascience/giaa021

Savoia, S. 2021. Interbull genomic evaluation of small Holstein populations: InterGenomics-Holstein (IG-HOL). *Interbull Bulletin* 56:180–187.

Schmitt, A.O., R.H. Bortfeldt, and G.A. Brockmann. 2010. Tracking chromosomal positions of oligomers - a case study with Illumina's BovineSNP50 beadchip. *BMC Genom.* 11:80. https://doi.org/10.1186/1471-2164-11-80

Schnabel, R.D. 2018. ARS1.2 coordinates for 25 bovine SNP chips available. Accessed. https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/.

Song, H., T. Dong, X. Yan, W. Wang, Z. Tian, A. Sun, Y. Dong, H. Zhu, and H. Hu. 2023. Genomic selection and its research progress in aquaculture breeding. *Reviews in Aquaculture* 15:274–291. https://doi.org/10.1111/raq.12716.

Vandewalle, M., R. Bonifazi, J. Vandenplas, S. Hiemstra, G. De Jong, D. Hinrichs, N. Gengler, and H. Wilmot. 2026. Unravelling the Genetic Structure of Local and Mainstream Red-Pied Cattle Breeds Using Genomics and Extended Pedigree Analysis. *J. Anim. Breed. Genet.* 143:705–718. https://doi.org/10.1111/jbg.70054.

Wiggans, G.R., J.B. Cole, S.M. Hubbard, and T.S. Sonstegard. 2017. Genomic Selection in Dairy Cattle: The USDA Experience. *Annu. Rev. Ani. Biosci.* 5:309–327. https://doi.org/10.1146/annurev-animal-021815-111422.

Zhao, H., Z. Sun, J. Wang, H. Huang, J.-P. Kocher, and L. Wang. 2014. CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics 30:1006–1007. https://doi.org/10.1093/bioinformatics/btt730.

Zimin, A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.P. Van Tassell, T.S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J.A. Yorke, and S.L. Salzberg. 2009. A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol.* 10:R42. https://doi.org/10.1186/gb-2009-10-4-r42.

European Bovine Evaluation – building joint genetic evaluations across Europe

J. Heise[1], Z. Liu[1], I. Croué[1,2], A. Michenet[1,2], R. Reents[1,3], J. Baltissen[1,4], A. Levring[1,5], H. Splittorff[1,8], L. Schibler[1,9], and J.-Y. Dréau[1,10]

[1] European Bovine Evaluation F.M.B.A., Agro Food Park 12, 8200 Aarhus N, Denmark

[2]Eliance, 149 rue de Bercy, 75012 Paris, France

[3] Vereinigte Informationssysteme Tierhaltung w.V. (vit), Heinrich-Schröder-Weg 1, 27283 Verden, Germany

[4] Rinder-Union West eG, Schiffahrter Damm 235a, 48147 Münster, Germany

[5] Landbrug & Fødevarer F.m.b.A, Vesterbrogade 4A, 4. sal, 1620 København V, Denmark

[6] Växa Sverige ek. för, Box 288, 751 05 Uppsala, Sweden

[7] Faba co-op, P.O. Box 40, FIN-01301 Vantaa, Finland

[8] NAV – Nordic Cattle Genetic Evaluation, Agro Food Park 15, 8200 Aarhus N, Denmark

[9] Gènes Diffusion, 3595 Route de Tournai, 59500 Douai, France

[10] Synetics, 4 Rue Eric Tabarly, 35530 Noyal-sur-Vilaine, France

Corresponding author: Johannes Heise (johannes.heise@ebe-centre.eu)

Abstract

In 2025, Partners from France, Germany and the Nordics (Denmark, Finland, Sweden) founded European Bovine Evaluation FMBA (EBE) to develop joint single-step genetic evaluations for Holsteins by 2028. These evaluations cover 9 countries across Europe, i.e., additional to the Partners' countries: Austria, Belgium (Walloon region), Luxembourg, and Norway. Number of offspring of bulls in the different countries that participate in Interbull MACE are routinely reported by the countries. A comparison of these figures between the EBE countries on the one side and the currently largest Holstein population under joint routine genetic evaluations worldwide on the other side shows that the Holstein population under joint routine genetic evaluations at EBE will become one of the largest worldwide as soon as routine evaluations are established at EBE.

Key words: Holstein, European collaboration, single-step, genetic evaluation, joint evaluation, large populations

Introduction

With the establishment of European Bovine Evaluation FMBA (EBE) in 2025, Partners from France, Germany and the Nordics (Denmark, Finland, Sweden) have joined forces to develop joint routine single-step genetic evaluations covering cattle populations from 9 European countries: in addition to the Partners' own countries, the collaboration covers populations from Austria, Luxembourg, Norway, and Wallonia (Belgium). The development of joint single-step genetic evaluations will serve three main goals:

1) make GEBVs for individual traits directly comparable across countries,
2) increase the reliabilities of GEBVs
3) leverage cross-country synergies to deliver better genetic evaluations more efficiently, including the development of genetic evaluations for new traits.

As its first mission until 2028, EBE will develop joint single-step evaluations for the Interbull traits, additional health traits (claw health, metabolic health, reproductive disorders), and young stock survival for Holsteins. After this, joint single-step evaluations will follow for other breeds, e.g.,

Jersey, Red Dairy Cattle, Normande, as well as for other traits that are not covered by the first mission. Furthermore, EBE will be open to new Partners.

In the following paper, we present the expected size of the Holstein population under genetic evaluation at EBE for different Interbull traits and compare it to the currently largest Holstein population under genetic evaluation worldwide.

Materials and Methods

EBE will merge phenotypic, genomic, and pedigree data of Holsteins from currently four different national genetic evaluation centres (NGEC) to develop and conduct joint single-step genetic evaluations. Currently, these NGECs provide data from Holsteins to Interbull MACE (Interbull 2026; Schaeffer, 1994) independently: DEU (Germany, Austria and Luxembourg; VIT, 2026), FRA (France; GenEval, 2026), DFS (Denmark, Finland, Sweden, Norway; NAV, 2026), and BEL (Walloon region of Belgium; Vanderick et al. 2025). Due to the early stage of development of genetic evaluations at EBE, analyses from joint data were not available at the time of this study. Therefore, we used information provided to Interbull by the different countries to give an indication of the future Holstein population size under joint genetic evaluation at EBE once those are established.

For Holstein bulls, we analysed number of offspring reported by the different countries in the Interbull bull evaluations (MACE) as of April 2026. Within birth year classes of bulls, we compared number of offspring reported by DEU, DFS, FRA (incl. BEL) with the USA and EBE, i.e., the sum of DEU, DFS, FRA (incl. BEL). This comparison was conducted for all 39 Interbull traits evaluated in Interbull MACE separately. In this paper, we only present numbers for protein yield and calving ease direct.

Results & Discussion

Comparison of number of offspring for protein yield across countries

Figure 1 shows the total number of offspring with phenotypes of Holstein bulls for the trait protein yield (PRO) in comparison for EBE, USA, DEU, FRA and DFS. EBE countries have reported higher numbers of offspring for most historic sire birth years until 2018, e.g., for sires born in 2012, EBE countries have reported

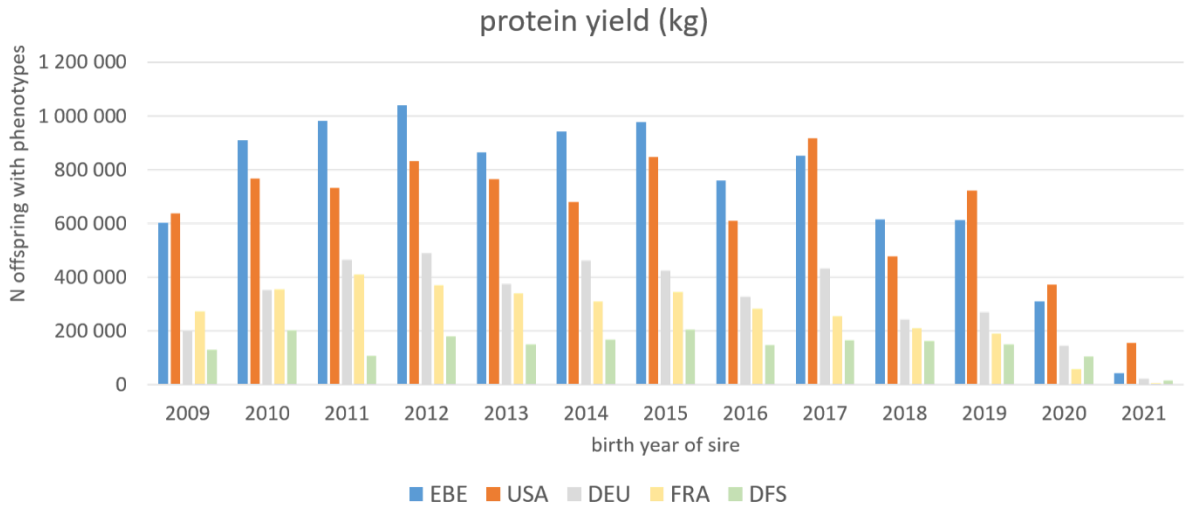


Figure 1. Total number of offspring with phenotype for protein yield of all Holstein bulls included in Interbull MACE as of April 2026 for EBE (sum of DEU, DFS, FRA, BEL), USA, DEU, FRA (incl. BEL), DFS.

1,039,877 offspring with phenotypes in total, while the USA have reported 832,473 offspring with phenotypes. For the most recent sire birth years (2019 to 2021), a slightly higher number of offspring was reported by the USA compared to the EBE countries. Similar patterns were also found for the trait groups somatic cell scores, female fertility and longevity (results not shown).

Comparison of calf counts for direct genetic effect of calving ease between countries

In contrast to most dairy cattle traits, where phenotypes are recorded on cows, for calving ease direct (DCE) phenotypes of calves contribute to estimated breeding values. Figure 2 therefore shows calf counts by birth years of their sires for EBE, USA, DEU, FRA and DFS for the trait DCE. EBE countries have reported substantially larger numbers of calves with phenotypes than USA, e.g., for sire birth year 2012 (year with the highest number of calves), the total number of calves of Holstein bulls reported by the EBE countries is 2,109,781, while only 1,311,709 calves are reported by the USA, which is 62% amount of calves reported by the EBE countries.

Similarly, a much lower number of calves was also observed for stillbirth direct in the USA population than in the EBE population (results not shown).

Interpretation of number of offspring of bulls reported by different countries

NGECs only send EBVs of bulls that have offspring in at least 10 herds in their national evaluation to Interbull MACE, but national evaluations also include sires with offspring in less herds, and often also include natural service sires. Therefore, the actual number of animals with phenotypes included in the national evaluations are usually somewhat higher than the total number of offspring with phenotypes reported for bulls that participate in MACE at Interbull. Although the actual gap might differ between countries due to differences in sire usage patterns across herds and proportions of natural mating, it remains small compared to the size of the population. Therefore, the number of offspring of bulls participating in Interbull MACE which are reported by the different countries can serve as a good indicator for the actual population size.

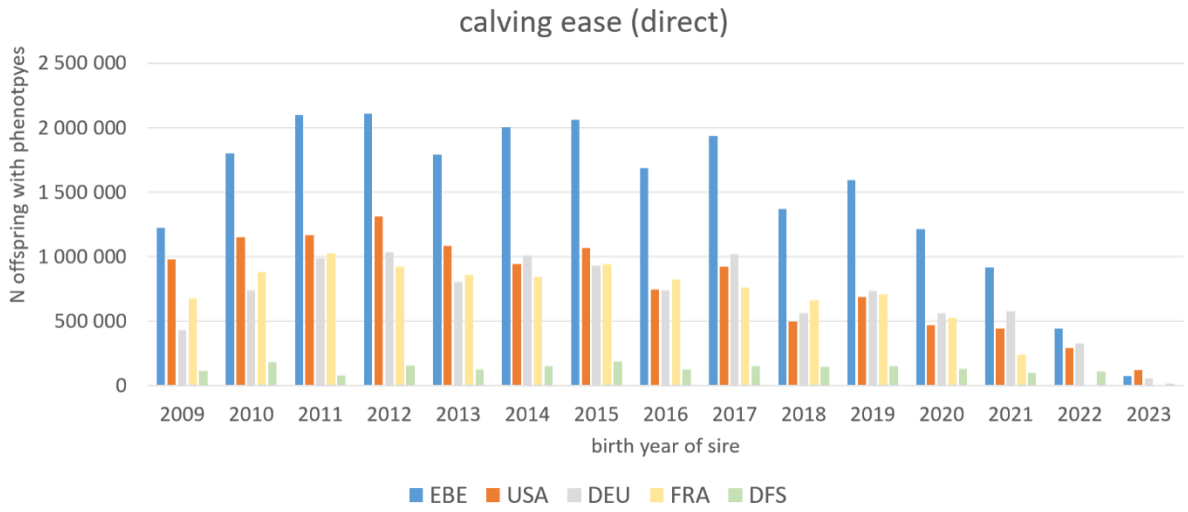


Figure 2. Total number of offspring with calving ease (direct) phenotype of all Holstein bulls included in Interbull MACE as of April 2026 for EBE (sum of DEU, DFS, FRA, BEL), USA, DEU, FRA (incl. BEL), DFS.

Conclusions

With EBE, the EBE Partners aim to set the European Standard for genetic evaluations of cattle. Our analysis shows that the expected size of the Holstein population under joint routine single-step genetic evaluations at EBE will be mostly comparable to the currently largest Holstein population under joint genetic evaluations worldwide. In addition, EBE builds upon a long history of high-quality data recording in the European countries, which includes also well-established data recording for many relatively new traits. Equipped with these core foundations, EBE has the potential to match and even exceed its aim to set the European Standard for genetic evaluations of cattle.

References

GenEval, 2026. Evaluation genetique des bovins laitiers, production laitiere – morphologie – caracteres functionnels. https://geneval.fr/wp-content/uploads/2026/07/2610-15_Methodes_evaluation_Races_BL.pdf (23 Aug 2026)

Interbull, 2026. International Evaluations – MACE. https://interbull.org/ib/mace (24 Aug 2026)

NAV, 2026. Nordic cattle genetic evaluation: NAV official genetic evaluation of dairy cattle – data and genetic models. https://nordicebv.info/wp-content/uploads/2025/11/NAV-routine-genetic-evaluation-November-2025.pdf (28 Aug 2026)

Vanderick, S., K. Wijnrocx, J. Vandenplas, A. Gillon and N. Gengler. 2025. Implementation of a Single-Step genomic evaluation system for dairy cattle in Wallonia, Belgium. *Interbull Bulletin* 61:48-56.

VIT, 2026. Estimation of breeding values for dairy breeds. https://www.vit.de/fileadmin/DE/Zuchtwertschaetzung/vit_zws_milchrind_beschreibung_eng.pdf (28 Aug 2026)

Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. *J Dairy Sci*. 77:2671-2678.

Females in MACE: Resolving double counting using the reverse reliability approximation.

J. Mugambe[1], M. Johnsson[2], H. Gao[3], Z. Liu[4], V. Palucci[1], S. Andonov[2]

[1]Interbull Centre, Department of Animal Biosciences, Swedish University of Agricultural Sciences, Box 7023, S-75007 Uppsala, Sweden

[2]Department of Animal Biosciences, Swedish University of Agricultural Sciences, Box 7023, S-75007 Uppsala, Sweden

[3]Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland

[4]IT-Solutions for Animal Production, Heinrich-Schroeder-Weg 1, 27283 Verden / Aller, Germany

Corresponding author: julius.mugambe@slu.se

Abstract

Including females' estimated breeding values (EBVs) in Multiple Across-Country Evaluation (MACE) requires a way to prevent information shared through relatives from being counted more than once. This paper evaluates a reverse reliability approximation that converts submitted reliability into an adjusted effective daughter contribution (EDC) after accounting for pedigree information. The results focus on five Ayrshire populations contributing with female data: Australia, Canada, Great Britain, the United States, and South Africa. For cows, mean adjusted EDCs and reliabilities were compared with submitted national values for milk, fat, and protein. Adjusted mean EDCs were equal to or lower than national means in almost all population and trait combinations, with the largest reductions observed for Great Britain. Mean female MACE (FMACE) reliabilities were generally close to, but often lower than, national reliabilities; South Africa was the exception, with higher FMACE mean reliability for all three traits. Bull comparisons between conventional MACE and FMACE showed that adding female information increased international reliability across the participating populations. These results support the use of the reverse reliability approach as a practical mechanism for controlling double-counting when female information is incorporated into MACE. The results also show why EDC is central to the procedure: it represents the amount of information retained for an animal after reliability attributable through relatives has been accounted for.

Key words: International evaluations, MACE, small populations, females in MACE

Introduction

Multiple Across-Country Evaluation (MACE) combines national genetic evaluations across countries and provides international breeding values expressed on each participating country's scale (Schaeffer, 1994). The females in MACE work extend this framework by allowing cow EBVs to contribute to the international evaluation (Mugambe et al., 2025). The principal technical concern is double-counting: A cow EBV contains information from the cow itself and from relatives; some of those relatives may be A.I. bulls included in the MACE evaluation. If the full reliability of every submitted proof is treated as independent information, shared information can receive more weight than intended.

The approach examined here addresses this overlap through a reverse reliability approximation. Instead of using submitted effective daughter contributions (EDC) directly, reliability is reversed to estimate the contribution needed after information attributable to relatives has been considered. The adjusted EDC is therefore intended to represent the information that remains available to weigh an animal in female MACE (FMACE)

after any possible overlapping through pedigree has been accounted for. The objective of this paper is to present the results for cows and bulls and to assess whether the resulting pattern is consistent with a useful correction for double-counting.

Materials and Methods

Production trait data for milk, fat, and protein were evaluated for five Ayrshire populations contributing with female proofs from Australia (AUS), Canada (CAN), Great Britain (GBR), the United States (USA), and South Africa (ZAF). The analysis used the existing MACE framework with female records added to create the FMACE evaluation.

Sources of double-counting

The inclusion of female EBVs in MACE poses a risk of double-counting. In conventional MACE, a bull's EDC already represents information contributing to his national evaluation, including information from his daughters. The same problem occurs among females because a cow's EDC can contain information from her own records, progeny and further descendants, while those related animals may also enter the evaluation with their own EDCs. For example, information from a bull may contribute to the EDC of his dam while the bull simultaneously contributes his own EDC to MACE. Consequently, simply combining the EDCs supplied for bulls and cows would treat overlapping information as if it were independent, potentially inflating the amount of information entering the evaluation and affecting the resulting reliabilities. The purpose of deriving adjusted EDCs (EDCadj) is therefore to partition the available information among related animals so that information already represented through relatives is not counted again when both bulls and cows are included in MACE.

The reverse reliability procedure

The reverse reliability procedure uses reliability, pedigree, and trait heritability to approximate an animal's effective contribution after accounting for reliability attributable to relatives. This follows the logic of the Harris and Johnson (1998) approximate animal model reliability calculation, applied in reverse. Reliability from two information sources, x and y, is combined as shown below.

$$R(x+y) = \frac{[R(x) + R(y) - 2R(x)R(y)]}{[1 - R(x)R(y)]}$$

And this can be reversed to recover the reliability attributable to one source:

$$R(x) = \frac{[R(x+y) - R(y)]}{[R(x+y)R(y) + 1 - 2R(y)]}$$

This algorithm was implemented iteratively, looping through countries in the MACE pipeline one at a time, using APAX99 to obtain adjusted EDCs free of double counting. The resulting adjusted EDC is then used as the information weight in FMACE. Cow results were summarized as population means for national EDC, adjusted FMACE EDC, national reliability, and FMACE reliability. Bull results were summarized by comparing the mean reliability of conventional MACE with the corresponding FMACE reliability after including female information.

Results & Discussion

The cow results (Table 1) show directly how the reverse reliability adjustment changes the amount of information assigned to female proofs. Across the five populations, adjusted mean EDCs were never higher than national mean EDCs except for South African milk, where both means were 9. The size of the adjustment differed considerably among populations, indicating that the amount of reliability attributable through relatives is population-specific rather than a fixed proportion of the submitted proof.

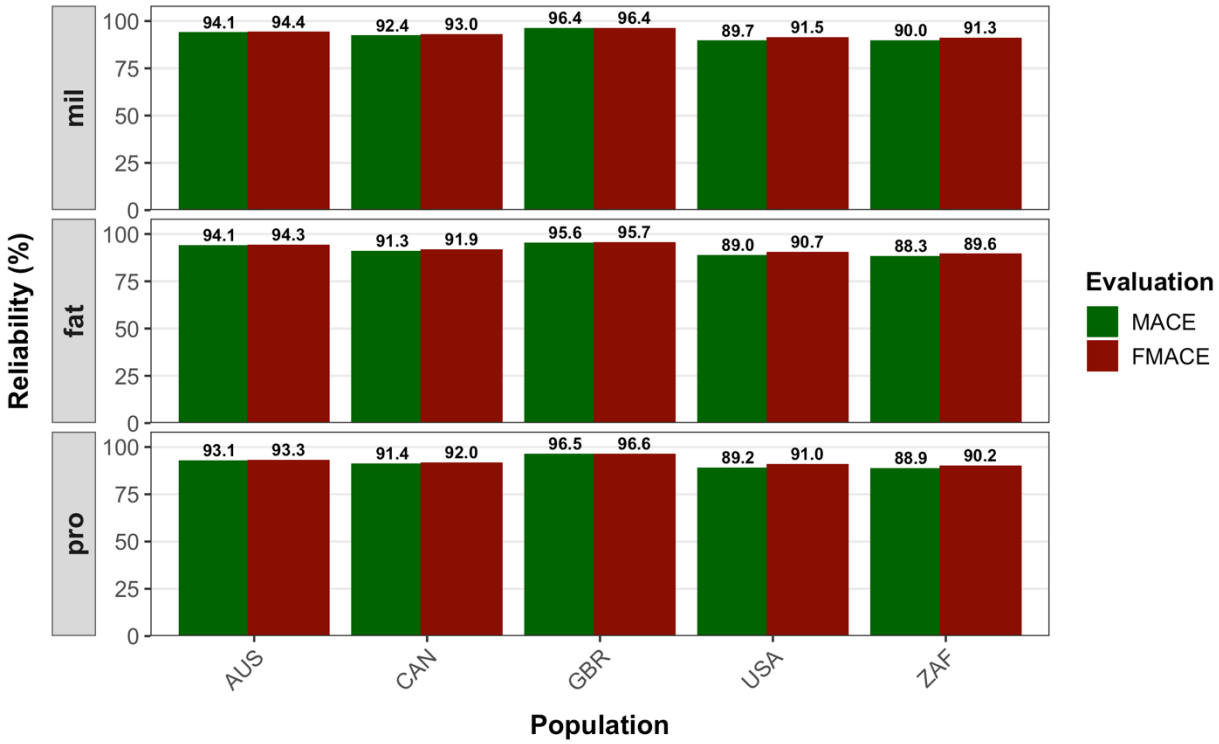


Figure 1. Comparison of bull reliability in conventional MACE and FMACE

Table 1: Mean EDC and reliability of cows before and after the reverse reliability adjustment for milk, fat and protein.

Trait	Pop	N	EDC_{nat}	EDC_{adj}	rel_{nat}	$Frel_{int}$
Milk	AUS	25,990	10	10	63.39	63.31
	CAN	10,817	16	12	63.94	63.48
	GBR	84,731	11	6	59.96	57.77
	USA	108,496	8	8	48.22	46.61
	ZAF	38,495	9	9	51.18	52.10
Fat	AUS	25,973	10	10	64.45	64.32
	CAN	10,821	16	15	63.93	63.47
	GBR	84,767	14	7	59.95	57.77
	USA	107,991	7	7	46.20	44.39
	ZAF	38,564	13	12	47.14	48.35
Protein	AUS	25,709	7	7	55.13	54.62
	CAN	10,814	14	14	63.93	63.46
	GBR	84,750	15	6	59.95	57.79
	USA	108,613	9	9	49.46	47.99
	ZAF	38,708	10	10	49.69	50.77

Pop; Population, **EDC_{nat};** National EDC, **EDC_{adj}**;Adjusted EDC, **rel_{nat};** National reliabilities, **$Frel_{int}$;**FMACE reliability

The largest EDC reductions were observed for GBR. For milk, mean EDC declined from 11 to 6; for fat, from 14 to 7; and for protein, from 15 to 6. CAN declined from 16 to 12 for milk and from 16 to 15 for fat, while protein remained at 14. AUS changed little, with equal, rounded mean EDCs for milk and fat and no change for protein. The USA also showed no change in the rounded mean EDCs for the three traits. These results illustrate the purpose of the correction: adjusted EDC is not intended to reproduce the submitted EDC, but to represent the information remaining after overlap through relatives is accounted for.

Mean cow reliabilities (Figure 1) were close between the submitted national values and FMACE, but the direction and magnitude differed across populations.

The reliability results reinforced the importance of interpreting FMACE using the information weights derived from the reverse procedure. Lower mean FMACE reliability does not by itself demonstrate a poorer evaluation because the procedure deliberately removes information represented through relatives from the independent contribution assigned to a cow. Possible causes of these discrepancies may include variations in data quality across countries, such as incomplete pedigrees, and inconsistencies in reliability estimation methods.

The bull results provided the complementary test of whether female information contributes to the international evaluation after the double-counting correction. In Figure 1, FMACE reliability is higher than conventional MACE reliability for the displayed participating populations across milk, fat, and protein. The increases are generally modest. This is consistent with the objective of the reverse reliability approach: only information that remains after accounting for overlap should receive additional weight. The value of the result is therefore not the size of the increase alone, but that female information can add reliability to bulls after the information weights have been adjusted to control double-counting.

Together, the cow table and bull figure place EDC at the centre of the results. The cow table shows how much the information weight can change after reversal, while the bull figure shows the consequence of allowing the resulting female information into MACE. The combination supports an interpretation in which reverse reliability is used first to separate overlapping information and adjusted EDCs, and then determine how much information each proof contributes to FMACE.

Conclusions

These results show that the reverse reliability approximation provides a practical way to control double-counting when female EBVs are introduced into MACE. Its most important output is the adjusted EDC, which represents the information retained for an animal after reliability attributable through relatives has been considered. The cow results demonstrate that this adjustment can be substantial in some populations and small in others.

Across milk, fat, and protein, FMACE cow reliabilities remained close to submitted national reliabilities, although most population and trait means were slightly lower after the correction, and ZAF was consistently higher. The bull comparison showed higher FMACE reliability than conventional MACE in the displayed participating populations. Together, these results indicate that female information can contribute to MACE if all information contained in a cow proof is treated as independent. The key to the approach is therefore not simply the inclusion of more cows, but the use of adjusted EDCs to determine how much non-overlapping information each proof contributes.

Acknowledgments

The author acknowledges the research partners who have contributed data on females in the MACE project, the Interbull Technical Committee for their technical input, and the

LUKE team for supporting the development and evaluation of the reverse reliability procedure.

References

Harris, B. and Johnson, D. 1998. Approximate reliability of genetic evaluations under an animal model. *J. Dairy Sci.* 81 (10), 2723–2728. https://doi.org/10.3168/jds.S0022-0302(98)75829-1

Mugambe, J., Palucci, V., Dürr, J., Van Doormaal, B., & Andonov, S. (2025). Female Inclusive MACE for Improved Genetic Evaluations in Small Populations: A Special Case for Ayrshire Dairy Herds. *Interbull Bulletin*, 61, 146–150.

Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77 (9), 2671–2678. https://doi.org/10.3168/jds.S0022-0302(94)77209-X

Visualizing Relationships between Conformation Traits over Time

M. Misztal[1], S. Tsuruta[1], I. Misztal[1], J. Graham[2], S. Comstock [2], J. Bewley[2], and D. A. L. Lourenco[1]

[1] *Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602*

[2] *Holstein Association USA Inc., Brattleboro, VT 05302*

Corresponding author: Maciej@uga.edu

Abstract

Genetic evaluations often include many correlated traits, which are necessary in selection indices. The genetic parameters underlying their estimation have the potential to change at an accelerated pace under genomic selection. Standard methods for their evaluation are computationally more expensive with large numbers of animals over several years. This study aims to present a methodology to visualize relationships between conformation traits over time. Genetic Parameters via Predictivity is a method that allows this study to graph correlations among 18 linear conformation traits over yearly times slices from 2010 through 2020 using all available genomic, pedigree and phenotypic information For the initial time slice of 2010, the validation population included cows born in 2009 through 2011 while the reference population included cows born from 2000 through 2008. Each subsequent time slice increases the year by one. Using the most recent time slice, with a validation subset of cows born between 2019 through 2021, each trait was regressed on every combination of 2, 3, and 4 remaining traits, choosing the combination with the highest explained variance (R^2). The chosen combinations were then used to compute and graph the predictor traits for all time slice going back to 2010. No more than 4 predictor traits were needed for each target trait, as evidenced by the low 4^{th} eigenvalue in 2010 and 2020, where principal component analysis was used to graph biplots.

Key words: genetic correlation structures, genetic parameters via predictivity, genomic selection, selection index

Introduction

For use in breeding programs, 18 linear conformation traits (also known as type traits) are recorded by the Holstein Association USA Inc. (Brattleboro, VT). These are combined with functional and production traits to the Total Production Index (Holstein Association USA Inc., 2017). Weigel at al. (1998) used conformation traits for the prediction of productive life, defined as the length of time between first calving and culling. Higher reliability was obtained with predictions that used correlated traits, including conformation traits. Tsuruta et al. (2004) examined slow changes in genetic parameters when calculating genetic correlations over 14 years. Cole (2024) discussed the need for periodic recalculation of correlation and covariance matrices in population under selection. Misztal et al. (2021) discussed how such estimation may no longer be feasible with larger genomic datasets. A random regression model over multiple time slices would add computation costs (Tsuruta et al., 2021). Misztal and Gowane (2025) developed a method that addresses these constraints by calculating productivity within and across traits over time slices using Genetic Parameters via Predictivity. Given the high correlation among the traits, visualization of changes over time can be difficult and is addressed using the method in this paper.

Materials and Methods

Data

Tsuruta et al. (2024) used the method outlined in Misztal and Gowane (2025) to compute yearly genetic parameters over the 18-time traits used in this paper. The method combines the deterministic formula for predicting the accuracy of genomic breeding values (Daetwyler et al., 2008) combined with the within- and across-trait predictivity formula developed by Legarra et al. (2008). The correlation between traits i and j is

$$Corr_{ij} = \frac{Corr\left(\mathbf{y}_i - \mathbf{Xb}_i, \hat{\mathbf{u}}_j\right)}{h_i acc_j}$$

where $\mathbf{y}_i$ is the phenotypic vector for trait i, $\mathbf{Xb}_i$ corrects for fixed effects, $\hat{\mathbf{u}}_j$ is the genomic EBV (GEBV) for trait j, h_i is the square root of trait i heritability, and acc_j is the GEBV accuracy for trait j.

Validation and reference datasets were defined annually using a sliding window. For 2010, the validation set comprised cows born 2009–2011 and the reference set comprised cows born 2001–2008. The evaluation was advanced forward by one year through 2020, ending with a validation set born 2019–2021 and a reference set born 2001–2018 (Tsuruta et al., 2024). Genomic estimated breeding values were computed using adjusted phenotypes from validation and reference populations. The reference population ranged from 2,725,000 records for 367,000 genotyped animals born from 2001-2008 to 4,661,000 records from 625,000 genotyped animals for animals born from 2001 through 2018. The validation data were composed of 515,000 animals and 33,000 genotyped animals with a total of 685,000 records for cows born from 2009 through 2011 to 203,000 animals and 67,000 genotyped animals with a total of 191,000 records for cows born from 2019 through 2021.

Methods

Dimensionalities were calculated using the eigenvalues of correlation matrices of the 18 conformation traits in each time slice. Negative or small eigenvalues were set to 10^{-6}. Regressions were then run for each of the 18 target traits using every combination of 2, 3, 4 of the remaining 17 predictor traits.

Out of the possible 136, 680 and 2,380 combinations of 2, 3 and 4 predictor traits in 2020, the combinations yielding the highest R^2 were kept. These combinations were estimated for the previous 10 time slices for each target trait. These were graphed to illustrate the temporal stability of the traits over time. Biplots of the first two principal components were constructed for 2010 and 2020 to inspect geometric shifts in trait vectors. Dendrograms were created assigning distances of 0 to 2 for traits that were perfectly correlated to perfectly negatively correlated, respectively. All statistical analyses were executed in Julia using GLM.jl, MultivariateStats.jl, and Clustering.jl.

Results and Discussion

Rear udder width as a target trait in the case of three predictor traits is used as an example of applying the best fit. Rear udder height, udder depth and rear teat placement are the combination with the highest % variance explained (R^2) of 84. The next best combination keeps the first two traits but replaces rear teat placement with body depth. R^2 does not decrease significantly. The third best combination replaces body depth with rump width, with R^2 dropping to 0.83. Finally, the worst combination of 3 traits R^2 of 0.08. These traits are the least correlated with rear udder width and have the least exploratory power.

Table 1: Coefficient of regression and R^2 for rear udder width as a target trait for three best and the worst combination of remaining 17 traits

Model	1	2	3	680
Udder Height	0.91	0.82	0.83	
Udder Depth	-0.51	-0.39	-0.43	
R Teat Place	0.31			
Body Depth		0.30		
Rump Width			0.30	
Rump angle				0.01
Foot Angle				0.27
Teat length				0.06
R^2	0.84	0.84	0.83	0.08

The combination chosen would be the first combination (udder height, udder depth, and rear teat placement) which had the highest variance explained. This analysis was performed for 2020 and then these same three traits would be estimated for all 11 time slices and then graphed.

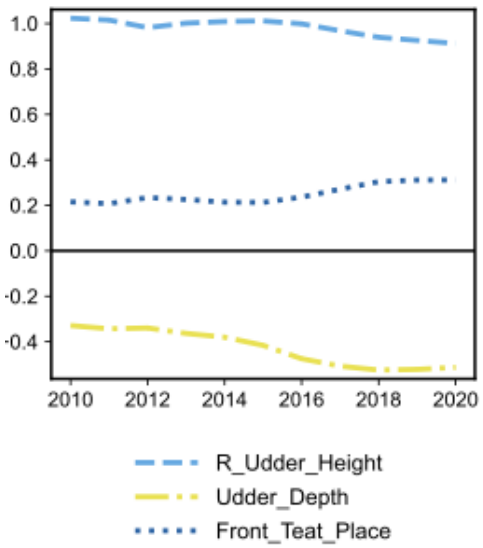


Figure 1. The coefficients for the best-fit model for rear udder width with optimal three predictor traits in 2020 over all 11 time slices

Eigenvalue analysis of the 2020 genetic correlation matrix yielded 18 values ranging from 6.92 down to -0.05. After setting non-positive eigenvalues to 10^{-6}, 8 principal components accounted for 90% of total variance, and 14 components explained 99%. This highlights substantial statistical redundancy across the 18 linear traits.

For best fit combinations of predictor traits, adding extra traits does not increase R^2 considerably (Table 1). Average change from 2 to 3 predictor traits was 0.04 for all traits and 0.03 from 3 to 4 predictor traits. Combined with low 4th eigenvalues, 4 predictor traits suggest possible overfitting and more predictor traits are unnecessary.

Table 2: Highest proportions of variance explained (%) for selected target traits and the corresponding optimal 2 ,3 and 4 trait models

	Proportion of variance (R^2)		
Target Trait	2-trait	3-trait	4-trait
Body Depth	96	97	97
Strength	95	96	97
Stature	92	95	95
Dairy Form	88	92	93
Udder Depth	81	89	95
R Udder Height	79	84	88
R Udder Width	75	84	88
Foot Angle	69	76	78
Rump Width	52	55	57
Teat Length	28	33	33
Rump Angle	9	17	23

Discussion

Highly correlated traits were better explained than poorly correlated traits. Body depth, strength, and stature all have explained variances of above 92 percent even with only two predictor traits. Their best predictor traits were frequently each other. They all measured body capacity traits and the most frequent predictor traits for all 18 target traits. Udder depth, rear udder height and rear udder depth were also highly correlated and were the best predictor traits for each other but were less frequently the optimal predictor traits for other traits.

Teat length and rump angle had the lowest variance explained and were the least correlated with other traits and the least likely to be optimal predictor traits for any other target traits. They measure unique and distinct

biological features. In the Trait Performance Index, dairy form, strength and stature are featured in more than one composite while rump angle is the one trait that is not featured in any of the composites.

Comparing all the traits with each other, the trait with the most significant change in relationships between 2010 and 2020 was dairy form. In 2010, it was clustered tightly with rear udder height and rear udder width. In 2020, dairy form was most more closely correlated with strength, body depth, and rump width. This may be related to breeding objectives, as dairy form is correlated negatively with fertility (Dechow et al., 2004), whereas strength is generally regarded as desirable by the industry.

Conclusions

This study provides a framework to convert correlation matrices derived by the Genetic Parameters via Predictivity method into visuals that present complex relationships between related traits over time. The best fit coefficient graphs, biplots, and dendrograms allow an intuitive interpretation for an otherwise overwhelming amount of subtle changes. The 2- and 3- trait models summarize linear conformation traits in Holsteins that exhibit high statistical redundancy. Rump angle and teat length are structurally independent, and desired selection pressure should be applied to them individually. This framework provides an objective approach to adjust to shifts in genetic parameters while including all genomic, pedigree and phenotypic information, which is crucial for trait selection in selection indices.

Acknowledgements

This research was primarily supported by grants from Holstein Association USA (Brattleboro, VT) and the U.S. Department of Agriculture's National Institute of Food and Agriculture (Agriculture and Food Research Initiative competitive grant 2015-67015-22936. This proceeding presents a summary of research accepted for publication in The Journal of Dairy Science

References

Cole, J.B. 2024. Are there too many traits in our selection indices? *J. Dairy Sci.* 108 (11), 12483–12505. https://doi.org/10.3168/jds.2024-25833

Daetwyler, H.D., Villanueva, B. and Woolliams, J.A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, 3 (10), e3395. https://doi.org/10.1371/journal.pone.0003395

Dechow, C.D., Rogers, G.W., Klei, L., Lawlor, T.J. and VanRaden, P.M. 2004. Body condition scores and dairy form evaluations as indicators of days open in US Holsteins. *J. Dairy Sci.* 87 (10), 3534–3541. https://doi.org/10.3168/jds.S0022-0302(04)73490-6

Holstein Association USA. 2017. Updates to the Total Performance Index (TPI) and type composites. Accessed Mar. 2, 2026. https://www.holsteinusa.com/pdf/Upcoming_Changes_aug17.pdf

Legarra, A., Robert-Granié, C., Manfredi, E. and Elsen, J.M. 2008. Performance of genomic selection in mice. *Genetics*, 180 (1), 611–618. https://doi.org/10.1534/genetics.108.088575

Misztal, I., Aguilar, I., Lourenco, D., Ma, L., Steibel, J. and Toro, M. 2021. Emerging issues in genomic selection. *J. Anim. Sci.* 99 (9), skab092. https://doi.org/10.1093/jas/skab092

Misztal, I. and Gowane, G. 2025. Estimation of heritabilities and genetic correlations by time slices using predictivity in large genomic models. *Genetics*, 230 (2), iyaf066. https://doi.org/10.1093/genetics/iyaf066

Tsuruta, S., Misztal, I. and Lawlor, T.J. 2004. Genetic correlations among production, body size, udder, and productive life traits over time in Holsteins. *J. Dairy Sci.* 87 (5), 1457–1468. https://doi.org/10.3168/jds.S0022-0302(04)73297-X

Tsuruta, S., Lawlor, T.J., Lourenco, D.A.L. and Misztal, I. 2021. Bias in genomic predictions by mating practices for linear type traits in a large-scale genomic evaluation. *J. Dairy Sci.* 104, 662–677. https://doi.org/10.3168/jds.2020-18668

Tsuruta, S., Lourenco, D., Lawlor, T. and Misztal, I. 2024. Over-time genetic correlations based on accuracy in GEBV for 18 linear conformation traits in US Holsteins. Page 785 in Book of Abstracts of the 75th Annual Meeting of the European Federation of Animal Science. Florence, Italy. https://docs.eaap.org/boa/2024_Florence_EAAP_Book_Abstracts.pdf

Weigel, K.A., Lawlor, T.J., VanRaden, P.M. and Wiggans, G.R. 1998. Use of linear type and production data to supplement early predicted transmitting abilities for productive life. *J. Dairy Sci.* 81 (7), 2040–2044. https://doi.org/10.3168/jds.S0022-0302(98)75778-9

Estimation of genetic parameters for production traits in Polish Simmental cattle using a random regression test-day model

Dawid Słomian[1], Klaudia Gogola[1], Monika Skarwecka[1], Kacper Żukowski[1]

[1]National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland

Abstract

In Poland, routine genetic evaluation of Simmental cattle for production traits is based on a multi-trait animal model using total yields for each of the first three lactations. The purpose of the study was to select a reference population for the Simmental breed, which is the second largest population of cattle in Poland, and to estimate variance components for production traits, using a single-trait, multi-lactation random regression test-day model (RR-TDM). Analyses were performed for the first three lactations within the range of 5 to 305 days in milk (DIM). Random regression was fitted using Legendre polynomials, and variance components were estimated by Gibbs sampling. The collected production traits included: milk yield in kilograms; fat, protein, and lactose expressed in kilograms and percentages; and somatic cell score. Phenotypic data from 13,966 cows that were purebred Simmentals, comprising 256,341 test-day records from the first three lactations, were used to estimate variance components. The dataset was prepared without crossbred animals and records from herds with only one cow. Based on additive genetic, permanent environmental, and residual variance components for each trait across lactation, daily heritability was calculated. The average heritability estimates were: 0.20 for milk yield, 0.18 for fat yield, 0.25 for fat content, 0.19 for protein yield, 0.37 for protein content, 0.22 for lactose yield, 0.32 for lactose content, and 0.13 for somatic cell score. The highest average heritability was observed for protein percentage, whereas the lowest was observed for somatic cell score. Depending on the trait, minimum and maximum heritability occurred at different stages of lactation. For most traits, maximum heritability occurred in the late stage of lactation. The results confirm that the RR-TDM approach is suitable for modeling production traits and for describing changes in genetic variation during lactation.

Key words: pedigree BLUP, test-day model, simmental cattle, routine evaluation

Introduction

The routine evaluation of the breeding value of Simmental cattle in Poland for production traits is currently based on a multi-trait animal model that uses aggregated production data from the first three lactations. However, advances in statistical methods and computational capabilities enable models that use test-day data, enabling a more accurate description of changes in yield and variance components throughout the entire lactation.

Test-day models with random regression allow for the treatment of performance on successive days of lactation as a continuous trait and for the modelling of individual genetic and environmental curves for animals. The advantage of this approach is that it allows for changes in genetic variance and permanent environment across days of lactation, rather than assuming a homogeneous variance structure throughout lactation (Schaeffer, 2004).

The significance of this approach for the Polish Simmental cattle population was confirmed by Otwinowska-Mindur et al. (2025), who, using a multi-lactation test-day model with random regression, demonstrated significant variability in heritability over the course of lactation and, for most of the analyzed traits, the highest heritability values during the final stage of lactation.

The objective of this study was: (1) to prepare a reference dataset comprising purebred Simmental cows, (2) to estimate variance components and daily heritability estimates for production traits using a single-trait, multi-lactation random regression test-day model, and (3) to prepare the first validation of breeding values obtained using a pedigree BLUP model.

Materials and Methods

To estimate the variance components, a dataset was prepared using animals with purebred Simmental breed composition. After data quality control, crossbreds and observations from herds with only one cow were removed. The final dataset included 13,966 cows and 256,341 test-day observations from the first three lactations. Observations between the 5th and 305th days of lactation were included.

Eight traits were analyzed: milk yield, fat yield, fat content, protein yield, protein content, lactose yield, lactose content, and SCS. For each trait, a single-trait, multi-lactation test-day model was used.

The pedigree data were checked and prepared using the Relax2 program (Strandén and Vuori, 2006). The prepared pedigree and phenotypic data were used as input to the RENUMF90 program, which renumbered the animals and created input files for further analysis in GIBBSF90 (Lourenco et al., 2022). Before running RENUMF90, the day of lactation was rescaled to the range for Legendre polynomials, and for each record, three normalized values of Legendre polynomials were calculated. These values corresponded to the three regression coefficients subsequently used to model the random additive genetic and permanent environmental effects.

The variance components were estimated using Gibbs sampling in the GIBBSF90 program. For each trait analyzed, 200,000 Markov chain iterations were performed. The first 20,000 iterations were discarded as the burn-in period. Afterwards, every tenth iteration was saved, resulting in a total of 18,000 samples from the posterior distribution used for further analysis. The posterior distributions and the results of the variance component estimates were then analyzed using the POSTGIBBSF90 program. Based on the estimated (co)variance matrices of the regression coefficients, the results and summary statistics were calculated using the R package.

The calculated variance components were then used as model parameters in the analysis of a larger dataset of production records. Breeding values were estimated using the MiXBLUP 3.0 software (Vandenplas et al., 2022). The pedigree BLUP test-day animal model can be written as:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Wf} + \boldsymbol{Vp} + \boldsymbol{Vu} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is the vector of cows' test-day records for production traits from the first three lactations, $\boldsymbol{b}$ is the vector of fixed effects of herd-test-day-parity-milking frequency, $\boldsymbol{f}$ is a fixed lactation curve effect, $\boldsymbol{p}$ is a vector of permanent environmental effects expressed as random regression coefficients of the Legendre polynomials, $\boldsymbol{u}$ is a random additive genetic effect based on pedigree.

The phenotype data from 49,564 cows were used for the full evaluation, while the pedigree included 121,963 cows and 61,350 bulls.

The linear regression (LR) method proposed by Legarra and Reverter (2018) was used for the validation. This method is based on comparing breeding values obtained from two consecutive evaluations conducted with the same model but using different phenotypic information. The full data set contains all available information, while the truncated data set includes all phenotypic data except the youngest animals from the last four years. The LR method provides statistics describing bias, dispersion, and predictive performance (Legarra and Reverter, 2018; Bermann et al., 2024)

Validation cows were defined as animals whose observations were removed when preparing the truncated dataset. Validation bulls were young sires of cows in the validation group. Validation was performed separately for

6,355 cows and 160 bulls. For both groups, breeding values estimated based on the full and truncated datasets were compared using regression:

$$\boldsymbol{EBVf} = b_0 + b_1\boldsymbol{EBVt} + e,$$

where **EBVf** represents the vectors of breeding values estimated based on the full data set, while **EBVt** represent the vector of breeding values predicted by the truncated data set, b_0 represent the intercept, which indicated a systematic bias in the model's prediction, and b_1 represents the regression slope, the dispersion of prediction compared to actual results. The R^2 coefficient is one of the results of linear regression and serves as a measure of prediction accuracy; it indicates the percentage of variance in the **EBVt** explained by **EBVf**.

Results & Discussion

Variance components

The average heritability was calculated based on results from the first three lactations; the highest heritability was obtained for protein content (0.37) and lactose content (0.32), followed by fat content (0.25). For traits expressed in kilograms, the average heritability values were lower, at 0.20 for milk yield, 0.177 for fat yield, 0.19 for protein yield, and 0.22 for lactose yield. The lowest heritability was recorded for SCS (0.13). The repeatability values were higher than the heritability values for all analyzed traits. The highest repeatability values were obtained for lactose yield (0.72), milk yield (0.72), and protein yield (0.68), while the lowest average repeatability was recorded for fat content (0.38). Detailed results on heritability and repeatability are presented in Table 1.

The estimates obtained were in general agreement with those previously reported for Simmental cattle. Dong et al. (2026) have likewise recently reported similar findings for Chinese Simmental cattle, with heritabilities estimated at 0.28 for fat content, 0.52 for protein content, and 0.35 for lactose content using a random regression test-day model. The differences in estimate sizes across studies might be due to differences in population structure.

Validation results

For the validation cows, correlation between full and truncated datasets ranged from 0.54 for fat content to 0.77 for protein content. Relatively high correlation values were also obtained for lactose content (0.74), protein yield (0.73), milk yield (0.71), and lactose yield (0.71). The lowest correlation was observed for fat content (0.54) and SCS (0.59). For the

Table 1: Heritability and repeatability estimates for production traits across the firs three lactations.

Trait	Lactation	heritability	repeatability
Fat yield	1st	0.18	0.56
	2nd	0.20	0.61
	3rd	0.16	0.60
	Total	0.18	0.59
Fat content	1st	0.27	0.40
	2nd	0.23	0.37
	3rd	0.25	0.37
	Total	0.25	0.38
Lactose yield	1st	0.17	0.67
	2nd	0.24	0.73
	3rd	0.22	0.74
	Total	0.22	0.72
Lactose content	1st	0.32	0.53
	2nd	0.34	0.61
	3rd	0.31	0.62
	Total	0.32	0.59
Milk yield	1st	0.19	0.67
	2nd	0.21	0.72
	3rd	0.20	0.73
	Total	0.20	0.72
Protein yield	1st	0.17	0.63
	2nd	0.20	0.69
	3rd	0.19	0.70
	Total	0.19	0.68
Protein content	1st	0.40	0.61
	2nd	0.34	0.63
	3rd	0.38	0.63
	Total	0.37	0.62
SCS	1st	0.09	0.49
	2nd	0.14	0.54
	3rd	0.15	0.59
	Total	0.13	0.54

validation bulls, the results were lower than those for validation cows. The correlations ranged from 0.40 for SCS to 0.79 for protein content. High correlations were also observed for the lactose content (0.73) and protein yield (0.61). Detailed correlation and coefficient of determination values are shown in Figure 1.

The regression coefficient b_1 for the validation cows was very close to the expected value of 1, ranging from 0.92 (fat content) to 1.05 (protein content). The results closest to the ideal value were those for lactose content (1.01), protein yield (0.98), milk yield (0.97), and lactose yield (0.97). In the context of validation bulls, the b1 values were more variable, ranging from 0.59 for SCS to 1.03 for protein content. Only protein content and lactose content were values obtained that were relatively close to the expected values. Detailed results for the regression coefficient b1 are shown in Figure 2.

In the absence of a systematic shift and with proper centering of the breeding values, the expected value of b_0 should be close to 0. For the validation cows, the b_0 values ranged from -2.26 for protein content to 2.18 for protein yield, except for milk yield, for which b_0 was 51.67. For the validation bulls, greater variation was observed, ranging from -26.90 for SCS to 90.13 for milk yield. However, the breeding values were analysed on their original scale and were neither centered nor standardised before regression. Detailed results for the bias b_0 are shown in Figure 3.

Conclusions

The results obtained confirm the usefulness of the multi-lactation random regression test-day model for evaluating production traits in the Polish Simmental cattle population. The estimated heritability and repeatability values indicate varying degrees of genetic determination for the traits analysed, with milk composition traits generally exhibiting higher heritability than traits in kilograms. Pedigree BLUP validation showed satisfactory stability for validation cows, as indicated by moderate-

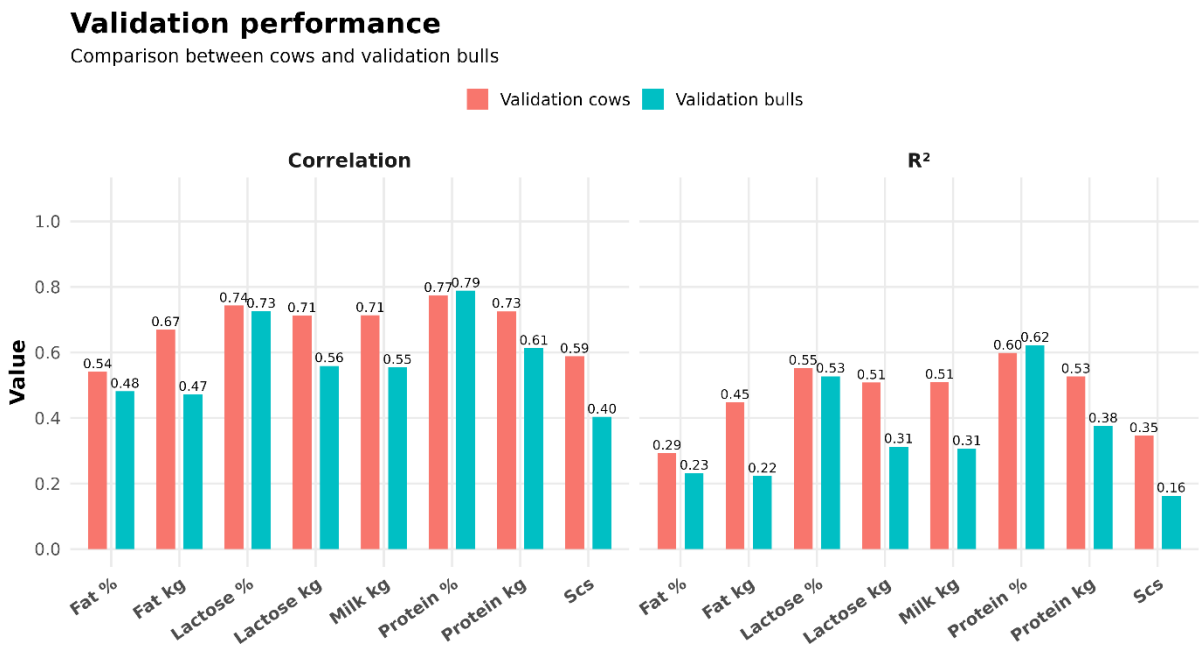


Figure 1. Correlation and coefficient of determination (R^2) between breeding values estimated using the full and truncated datasets for validation cows and bulls across production traits.

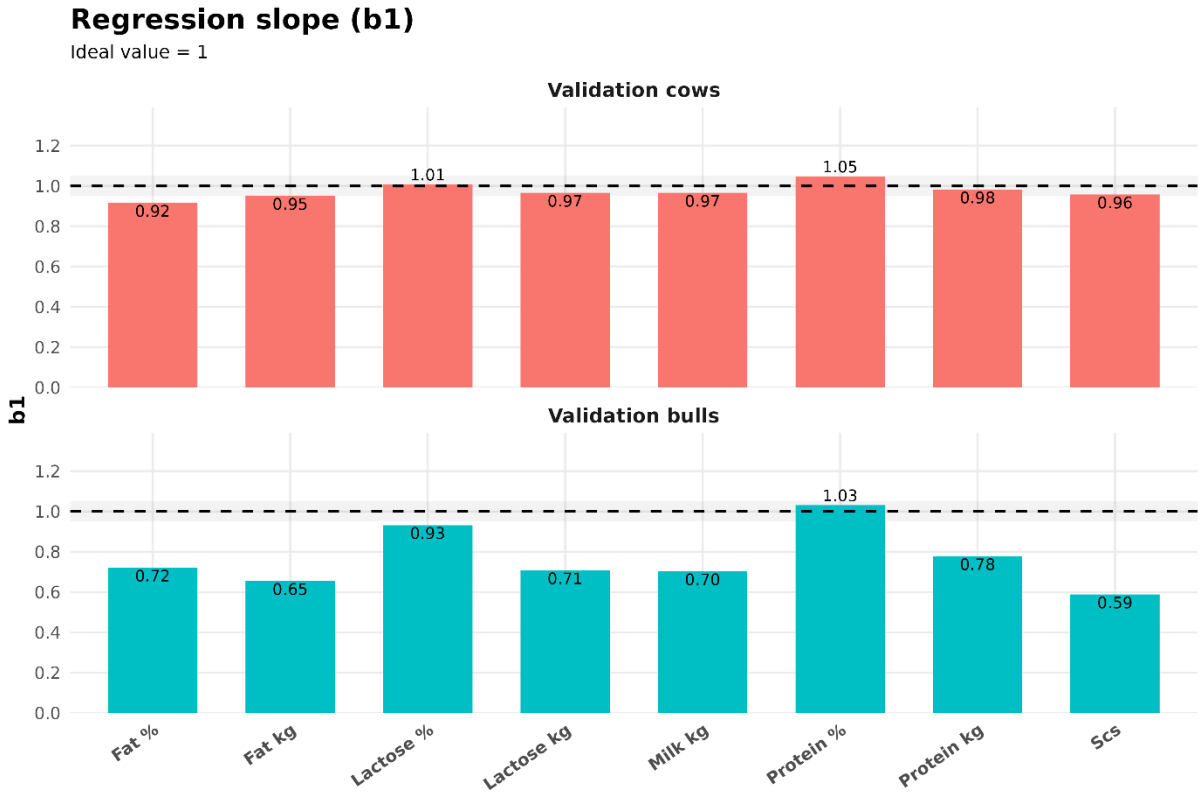


Figure 2. Regression slope (b_1) for validation cows and bulls across production traits

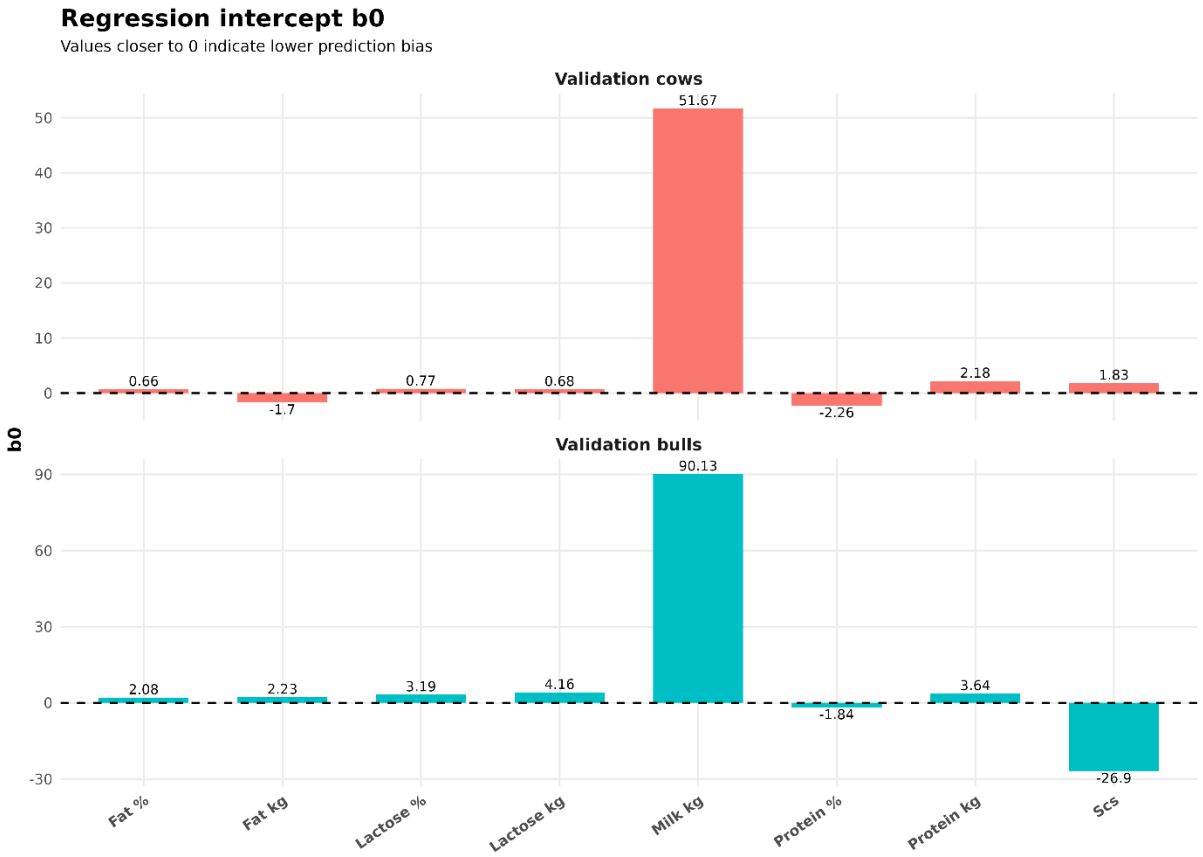


Figure 3. Regression intercept (b_0) for validation cows and bulls across production traits.

to-high correlations between full and truncated evaluations and a regression coefficient (b_1) close to 1. For the validation bulls, evaluation consistency was lower and more variable across traits, which may be related to the smaller validation group size and the limited available information. The b_0 values, however, showed greater variation due to the lack of standardization of breeding values. Overall, the results obtained provide a basis for the random regression test-day model for Polish Simmental cattle and for the further development of the system toward genomic evaluation using the single-step approach.

References

Bermann, M., Legarra, A., Alvarez Munera, A., Misztal, I., and Lourenco, D. 2024. Confidence intervals for validation statistics with data truncation in genomic prediction. *Genet. Sel. Evol.*, 56, 18. https://doi.org/10.1186/s12711-024-00883-w.

Dong, C., Ma, P., Tang, Y., Han, H., Chen, S., Yang, J., Sun, Z., Wang, W., Yu, Y., and Li, S. 2026. Genetic parameter estimation and fine-mapping of milk-production traits and somatic cell score in Chinese Simmental cattle. *J. Dairy Sci.*, 109(4), 4195–4208. https://doi.org/10.3168/jds.2025-27039.

Legarra, A., and Reverter, A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.*, 50, 53. https://doi.org/10.1186/s12711-018-0426-6.

Lourenco, D., Tsuruta, S., Aguilar, I., Masuda, Y., Bermann, M., Legarra, A., and Misztal, I. 2022. Recent updates in the BLUPF90 software suite. *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production*, 1530–1533. https://doi.org/10.3920/978-90-8686-940-4_366.

Otwinowska-Mindur, A., Ptak, E., Jagusiak, W., and Zarnecki, A. 2025. Genetic parameters for milk production traits of Simmental cows with random regression test-day model. *Animal,* 19(2), 101395. https://doi.org/10.1016/j.animal.2024.101395.

Schaeffer, L.R. 2004. Application of random regression models in animal breeding. *Livest. Prod. Sci.*, 86(1–3), 35–45. https://doi.org/10.1016/S0301-6226(03)00151-9.

Strandén, I., and Vuori, K. 2006. RelaX2: Pedigree analysis program. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, Brazil, 27–30.

Vandenplas, J., Veerkamp, R., Calus, M., Lidauer, M., Strandén, I., Taskinen, M., Schrauf, M., & Napel, J. T. 2022. 358. MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs. *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production.* https://doi.org/10.3920/978-90-8686-940-4_358

Pilot genomic evaluations of French crossbred dairy cattle

Audrey A.A. Martin[1,2], Didier Boichard[2], Sandra Dominique[3], Manon Guillerm[3], Sophie Mattalia[2,3], Sébastien Fritz[1,2], Pascal Croiseau[2]

[1] *Eliance, 149 Rue de Bercy, 75595, Paris cedex 12, France*

[2] *Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France*

[3] *Institut de l'Élevage, 149 Rue de Bercy, 75595, Paris cedex 12, France*

Corresponding author: audrey.martin@eliance.fr

Abstract

The use of crossbreeding in dairy cattle has been increasing, reflecting breeders' interest in new strategies to address current challenges in the dairy industry. Crossbred animals benefit from both parental breeds abilities and from heterosis, making them attractive to producers. However, genetic tools commonly used in purebred populations, such as pedigree or genomic estimated breeding values (GEBV), are not yet available for crossbred populations, limiting herd management options and the prediction of future phenotypes. The TransEvagenoc project aimed to develop and test a genomic evaluation tool adapted to current crossbred populations. Genotypes from 4,004 individuals resulting from crosses among the three main French dairy breeds (Normande, Montbéliarde, Holstein) were used. Breed-of-origin of alleles (BOA) and heterosis percentages were estimated based on reference populations of 5,000 individuals per pure breed. SNP effects estimated in pure breeds were then applied to crossbred genotypes according to BOA, and breed differences and heterosis effects were then added to estimate breeding values. Evaluations were performed for milk, fat and protein yields (MY, FY, PY), somatic cell score (SCS) and interval from calving to first service (ICFS). Methodological choices, such as GEBV standardisation, and results preparation were discussed in focus groups (FG) with industry stakeholders to better align to the field needs. Correlations between breeding values and first-lactation phenotypes were 0.35, 0.27, 0.28, 0.31 and 0.08 for MY, FY, PY, SCS and ICFS, respectively. Based on focus group discussions, GEBV were standardised using a representative crossbred population. This tool provides a first genomic evaluation framework for the French crossbred dairy cattle encouraging the collect of additional crossbred genotypes and phenotypes.

Key words: Dairy - Crossbreeding – Genomic evaluation.

Introduction

Historically, French cattle populations are composed of pure breeds. However, dairy crossbreeding has recently gained interest due to its adaptation to current environmental challenges, thanks to a more balanced trait profile and the added benefit of heterosis.

Nowadays, we are observing an increase in the number of inseminations resulting in dairy crossbred cattle in France (IDELE, 2024). In 2024, the proportion of dairy cross inseminations reached 6.1% of whole inseminations. In addition, the increase in the use of crossbreeding for the first insemination (up by 1.5%) indicates a direct breeding choice, rather than positioning dairy crossbreeding as an alternative following potential failure of purebred insemination.

This increase happened while no adapted genetic tools, such as genomic indices, were available for crossbred females, restricting crossbred replacement heifers' selection and mating plans.

Research groups and breeding organisations around the world have developed advanced and appropriate methods to perform accurate genomic evaluations for crossbred animals

(examples: Brøndum et al., 2011; Madalena et al., 1990; Wiggans et al., 2021). However, due to the lack of adequate management tools, there have been little efforts to phenotype and genotype crossbreds over the past decades. Therefore, none of the proposed advanced methods can be practically applied to the French crossbred population at the moment.

It is important to propose an intermediate approach, applicable at a small population scale and with outputs practically usable for breeders. This would promote performance recording and genotyping for crossbreds, bringing it to an adequate scale for more advanced and accurate methods. The TransEvagenoc project aimed to develop and test a pipeline based on a proposed intermediate approach (Clasen et al., 2023; Eiríksson et al., 2022, 2021), and to produce the first genetic evaluation for dairy crossbreds. The number of available performances being limited in crossbreds and in order to avoid any change in the pure breed evaluations, the crossbred individuals were evaluated by their average breed composition and the SNP effects estimated in pure breeds, as proposed by Eiríksson et al (2021).

Material and Methods

The crossbred evaluation is composed of two steps: the determination of the Breed of Origin of Alleles (BOA), and the actual crossbred evaluation. In-house programs were developed to perform the different steps. The pipeline requires both purebred and crossbred information.

Datasets

The pilot evaluation was conducted using crosses from any combination of the three main French dairy breeds: Holstein, Montbéliarde, and Normande. This initial selection of crossbred animals was performed by tracing back pedigree origins using Pedig (Boichard, 2002). Additionally, all individuals in the evaluation had to be genotyped (54,205 SNP) and phenotyped.

Five traits were considered: milk yield (MY), fat yield (FY), protein yield (PY), somatic cell score (SCS), and calving to first service interval (CTFS). To retain as many animals as possible, the pilot evaluation was performed using first-lactation phenotypes only. A total of 4,004 crossbred individuals were selected.

Methods description

Breed of Origin of Alleles: BreedOrigin

The BOA was determined using the in-house program BreedOrigin. Its performances in terms of accuracy and computational efficiency is described in the benchmark study of Liu et al. (2026). Its principle is simple: BreedOrigin analyses haplotypes. In a first step, the haplotype frequencies are estimated within each pure breed with a sufficiently large number of animals from each pure breed.

For the pilot evaluation, a sample of 5,000 individuals was selected to represent each pure breed. The first 3,000 individuals were the most widely used bulls (based on their number of offspring in the population), while the remaining 2,000 individuals were randomly selected from the most recent generation (born from 2023 to 2025) to capture the latest haplotypes present in the population.

The haplotypes of crossbreds were then compared, position by position, to those in purebreeds. A BOA was assigned when the haplotypic frequency was substantially higher in one breed; otherwise, the haplotype origin remained unknown. In subsequent iterations, unassigned segments flanked by segments assigned to the same breed were attributed to that breed. Otherwise, no BOA was assigned at this step. This iterative procedure progressively filled the gap, starting with haplotypes of 16 markers, then 8, 4, 2, and finally 1 marker.

Among other outputs, BreedOrigin provided crossbred genotypes including BOA information, together with the breed proportions and heterosis level of each crossbred individual.

Genetic evaluation: CrossSG

CrossSG program implements an extension of the GBLUP model designed for the genomic evaluation of crossbred animals. This approach accounts for the BOA when applying purebred marker effects to crossbred genotypes according to their BOA.

For the pilot model, and for each trait, the genetic value of each crossbred individual was estimated as follows:

$$\boldsymbol{y_i} = \sum_{\boldsymbol{b=1}}^{\boldsymbol{\#_breed}} \left((\boldsymbol{p_{i,b}}\mu_{\boldsymbol{b}}) + \sum_{\boldsymbol{j=1}}^{\boldsymbol{\#_{SNP}}} (\boldsymbol{\beta_{i,j,b}}\boldsymbol{X_{i,j,b}}) \right)$$

where y_i is the genetic value of individual i; μ is the genetic difference of breed b; $p_{i,b}$ is the proportion of breed b of individual i; $\beta_{i,j,b}$ is the breed of origin b of marker j of individual i; and $X_{i,j,b}$ is the effect of marker j, carried by individual i, centred on the allele frequency of marker j within breed b, such that :

$$X_{i,j,b} = (k_{i,j,b} - n_{i,j,b}f_{j,b})$$

where $k_{i,j,b}$ is the number of copies of the second allele at marker j originating from breed b in individual i; $n_{i,j,b}$ is the total number of alleles at marker j originating from breed b in individual i; and $f_{j,b}$ is the frequency of the second allele at marker j in breed b.

A heterosis term, ph_iH, could be added to this breeding value, where ph_i is the fraction of heterosis of individual i calculated by BreedOrigin as the proportion of loci with alleles originating from two different breeds; and H is the heterosis effect associated with the considered trait.

The marker effects and allele frequencies used in this pilot were obtained from the national purebred evaluations carried out by GenEval. The breed genetic differences (μ_b; expressed with the Holstein as standard) and heterosis effects (H) for all traits were obtained from Dezetter et al. (2015).

Discussion with the dairy breeding industry

Focus groups with the different stakeholders of the dairy cattle breeding industry were organised to discuss the future implementation and practical use of the crossbred evaluation.

The traits to be evaluated required careful consideration due to the difficulty of using multiple breeds, resulting in diverse measuring protocols, different trait definitions, or simply in the absence or presence of some traits in specific breeds.

Several scenarios were proposed for index standardisation. For ease of use, fairness across breeds, and technical considerations, the indices will be expressed on a crossbred scale in this pilot evaluation. The subset of the crossbred population used for the standardisation was selected to be representative of the breed composition observed in the overall current crossbred population. Additionally, raw indices will be shared with breeding organisations so that they can standardised them on their breed of choice.

Results & Discussion

Predictive ability

Including heterosis had little effect on EBV rankings (Pearson correlations > 0.99 for all traits) and slightly increased the correlation between EBV and phenotypes (Table 1).

Predictive ability, assessed as the Pearson correlation between genomic indices and phenotypes, was moderate but encouraging for the first crossbred indices. Our results showed that genomic prediction for crossbred dairy cattle is feasible, which is consistent with the conclusion of Steyn et al. (2021).

Table 1: Pearson correlations between EBV, EBV including heterosis (EBV+Het), and phenotype (pheno).

	Pearson correlation	
	pheno, EBV	pheno, EBV+Het
MY	0.35	0.36
PY	0.27	0.27
FY	0.28	0.28
SCS	0.31	0.31
CTFS	0.08	0.09

MY: milk yield; PY: protein yield; FY: fat yield; SCS: somatic cell score; and CTFS: calving to first service.

Out of the five traits, only CTFS showed a slightly different pattern for the results due to its low heritability and reversed optimum, i.e. lower values for the interval are desirable. Therefore, only results for MY and CTFS are detailed in this paper.

Genetic progress

The indices were standardised on a subset of the crossbred population. The genetic progress for MY and CTFS is shown in Figures 1.a and 1.b, respectively.

Before interpreting these figures, the overall breed proportions within the crossbred population across birth years were assessed to disentangle genetic progress from potential temporal changes in the representation of different crossbred types. For example, an increasing proportion of Holstein ancestry in the crossbred population could have mimicked the trend of genetic progress. As breed composition remained relatively constant across birth years, the observed trends can be attributed to genetic progress rather than changes in breed composition.

Accordingly, MY showed a clear upward genetic trend across birth years, indicating ongoing genetic improvement in the crossbred population, reflecting the genetic improvement achieved in the parental purebred populations (Figure 1a). In contrast, CTFS exhibited a weaker trend and greater year-to-year variability, consistent with its lower heritability (Figure 1b). The limited number of animals born after 2018 reflects the availability of crossbred genotypes, most of which originated from the Evagenoc project (2019–2022).

As shown in Figure 2, animals with higher genomic indices performed better in the first lactation. Animals with higher MY indices had better production and those with higher CTFS indices had shorter calving-to-first-service intervals, although the difference between index classes was less pronounced.

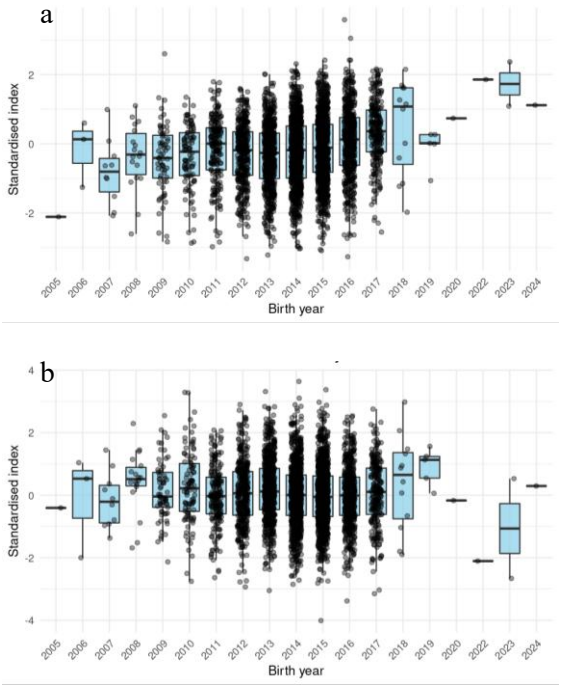


Figure 1. Genetic trend in crossbreds: Distribution of the genomic indices by birth year. a. Milk yield (MY) b. Calving to first service interval (CTFS) - note that lower values are desirable.

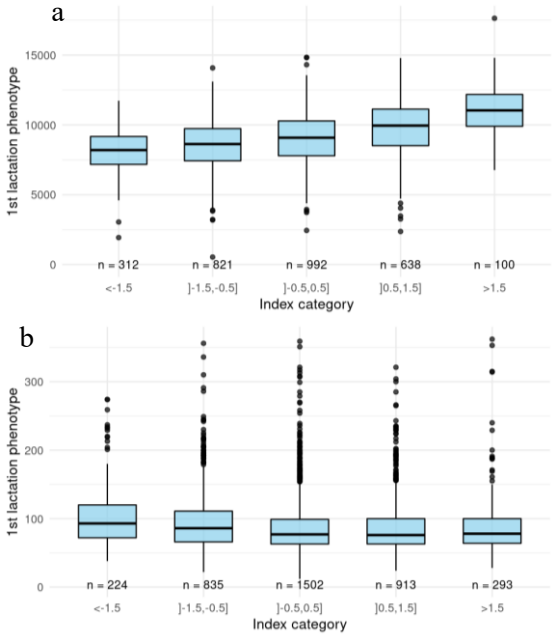


Figure 2. Distribution of first-lactation phenotypes across categories of genomic indices including heterosis effect. Boxplots represent the distribution of first-lactation phenotypes within each category. n corresponds to the category sample size. a. milk yield (MY). b. calving to first service (CTFS).

Evaluation strategy

Previous studies have demonstrated that genomic evaluations for crossbred dairy cattle can be implemented by combining genomic information from the parental breeds to the crossbred information (VanRaden et al., 2020; Wiggans et al., 2021). From a methodological perspective, the main challenge for a common evaluation comes from the difference of population sizes: purebred populations are much larger than crossbred populations and their information strongly dominate in the estimation of the SNP effects, leaving no opportunity for the crossbred information to impact estimates. From a practical perspective, implementing a common evaluation would also affect the organisation of national purebred evaluations and require substantial computational resources.

Because French crossbred resources are not yet enough developed, we used an intermediate method to obtain a genomic index for crossbred animals, as suggested by Eiriksson et al (2021). The observed associations between the indices and first-lactation performance suggest that this approach provides a useful tool until a dedicated crossbred genomic evaluation can be proposed.

Conclusions

Two programs, BreedOrigin and CrossSG, were developed within the TransEvagenoc project to implement genomic evaluations in the current French crossbred dairy population. They cope with the limited phenotypic and genomic data available in French dairy crossbreds, hampering the use of more advanced and adequate crossbred evaluation methods. This intermediate approach enabled the production and diffusion of the first genomic indices for crossbred dairy cattle, released at the end of 2025. Although only three breeds are presented in this paper, the routine evaluation includes the six breeds the most used in dairy crossbreeding in France.

By providing practical genomic evaluations for crossbred animals, this approach supports the choice of breeders using crossbreeding in France while encouraging the collection of additional phenotypic and genotypic data, paving the way for the future implementation of more comprehensive crossbred evaluation methods.

Acknowledgments

The work presented was supported by the Evagenoc and TransEvagenoc projects, funded by Apis-Gene, and by the contribution of several breeding companies that coordinated the genotyping of crossbred animals.

References

Boichard, D., 2002. Pedig: a fortran package for pedigree analysis suited for large populations. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*.

Brøndum, R.F., Rius-Vilarrasa, E., Strandén, I., Su, G., Guldbrandtsen, B., Fikse, W.F., Lund, M.S., 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. *J. Dairy Sci*. 94, 4700–4707. https://doi.org/10.3168/jds.2010-3765

Clasen, J. B., W. F. Fikse, G. Su, E. Karaman. 2023. Multibreed genomic prediction using summary statistics and a breed-origin of-alleles approach. *Heredity* 131, 33–42. https://doi.org/10.1038/s41437-023-00619-4.

Dezetter, C., Leclerc, H., Mattalia, S., Barbat, A., Boichard, D., Ducrocq, V., 2015. Inbreeding and crossbreeding parameters for production and fertility traits in Holstein, Montbéliarde, and Normande cows. *J. Dairy Sci.* 98, 4904–4913. https://doi.org/10.3168/jds.2014-8386

Eiríksson, J. H., E. Karaman, G. Su, and O. F. Christensen. 2021. Breed of origin of alleles and genomic predictions for crossbred dairy cows. *Genet. Sel. Evol.* 53, 84.

https://doi.org/10.1186/s12711-021-00678-3.

Eiríksson, J. H., K. Byskov, G. Su, J. R. Thomasen, and O. F. Christensen. 2022. Genomic predictions for crossbred dairy cows by combining solutions from purebred evaluation based on breed origin of alleles. *J. Dairy Sci.* 105, 5178-5191. https://doi.org/10.3168/jds.2021-21644.

IDELE, 2024. Statistiques des inséminations sur femelles laitières : le croisement laitier en 2024. Bilan des inséminations animales bovines en 2024. Rapport Résultats No. 0025203011.

Liu, H., Christensen, O.F., Eiríksson, J.H., Boichard, D., Milkevych, V., Thomasen, J.R., Karaman, E., 2026. Accuracy of local ancestry inference and its impact on genomic prediction in admixed dairy cattle populations. *J. Dairy Sci.* 109, 2743–2758. https://doi.org/10.3168/jds.2025-27019

Madalena, F.E., Teodoro, R.L., Lemos, A.M., Monteiro, J.B.N., Barbosa, R.T., 1990. Evaluation of Strategies for Crossbreeding of Dairy Cattle in Brazil. *J. Dairy Sci*. 73, 1887–1901. https://doi.org/10.3168/jds.S0022-0302(90)78869-8

Steyn, Y., Gonzalez-Pena, D., Bernal Rubio, Y.L., Vukasinovic, N., DeNise, S.K., Lourenco, D.A.L., Misztal, I., 2021. Indirect genomic predictions for milk yield in crossbred Holstein-Jersey dairy cattle. *J. Dairy Sci.* 104, 5728–5737. https://doi.org/10.3168/jds.2020-19451

VanRaden, P.M., Tooker, M.E., Chud, T.C.S., Norman, H.D., Megonigal, J.H., Haagen, I.W., Wiggans, G.R., 2020. Genomic predictions for crossbred dairy cattle. *J. Dairy Sci.* 103, 1620–1631. https://doi.org/10.3168/jds.2019-16634

Wiggans, G., VanRaden, P., Null, D., Nicolazzi, E., Jansen, G., Megonigal, J., 2021. Genomic evaluation of crossbred dairy cattle in the United States – an update. *Interbull Bulletin 56:17-21.*

Reassessing Wiggans' Method for Estimating Daily Yields in 3× Milking Systems over Four Decades: What Have We Learned?

Xiao-Lin Wu [1,2*], George Wiggans [1], Malia J. Caputo [1], Curtis P. Van Tassell [3], Ransom L. Baldwin VI [3], Steven Sievert [4], Jay Mattison [4], John B. Cole [1,5,6,7], Javier Burchard [1], and João Dürr [1]

[1] Council on Dairy Cattle Breeding, Bowie, MD
[2] University of Wisconsin, Madison, WI
[3] USDA Animal Genomics and Improvement Laboratory, Beltsville, MD
[4] National Dairy Herd Information Association, Verona, WI
[5] North Carolina State University, Raleigh, NC
[6] University of Florida, Gainesville, FL
[7] University of Guelph, Ontario, Canada
Corresponding author: nick.wu@uscdcb.com

Abstract

More than forty years have passed since Wiggans (1986) proposed a least-squares approach to derive multiplicative yield correction factors from proportional yields in cows milked three times daily. In the present study, we re-evaluated these correction factors using 3× milking data collected from 2023 to 2025 across four Holstein dairy farms in three U.S. states. The objectives were twofold: to assess the validity of the existing yield factors and to examine the robustness of the associated modeling strategies. We evaluated the accuracy of estimated daily yields using out-of-sample validation, allocating approximately 50% of records to training and the remaining 50% to testing. Accuracy measures included the correlation between observed and estimated 24-h yields, mean absolute error, and mean squared error. We observed clear shifts in model parameters over the past four decades, indicating greater dynamic responsiveness of proportional yield to interval length, while baseline section-specific effects have become less important. Nevertheless, improvements in the accuracy of estimated daily yields were modest. Substantial variability was observed among farms, suggesting that herd-level factors—such as management practices, milking system precision, and data consistency—contribute more to predictive performance than temporal changes in model parameters. Expanding the training dataset, particularly by including more representative herds and cow-day records, may enhance predictive stability and generalizability. Although imposing a 24-h normalization constraint improves model interpretability and ensures the coherence of proportional estimates, it does not necessarily yield meaningful gains in predictive accuracy. Overall, this study reaffirms the continued relevance of proportional-yield modeling and existing correction factors as practical benchmarks, while providing updated empirical guidance for estimating daily yields in three-times-daily milking systems.

Keywords: accuracy, Holstein, fat, milk, protein, yield correction factors

Introduction

Milking frequencies vary among farms and occasionally by season. In the U.S., the Dairy Herd Improvement (DHI) program routinely records only one (or sometimes two) of the day's milkings. However, herds that adopt twice- or thrice-daily milking still require consistent, unbiased estimates of 24-hour yields from partial milking records for accurate recordkeeping, genetic evaluations, and management decisions.

To address this need, Wiggans (1986) proposed an ordinary least-squares (OLS) approach with *a posteriori* rescaling of the intercepts to estimate yield correction factors by modeling proportional yields in cows milked

three times daily. Notably, this method is among those officially adopted in the U. S. through the National Dairy Herd Improvement (DHI) program. It has also been formally incorporated into the ICAR Guidelines (ICAR, 2022) as a standard procedure for computing 24-hour milk, fat, and protein yields from partial milking records (DeLorenzo and Wiggans, 1986; Liu et al., 2000).

In this study, we re-evaluated yield correction factors using 3X milking data collected from 2023 to 2025 at four Holstein dairy farms across three U.S. states. The objectives are twofold: to assess the validity of the existing yield factors and to examine the robustness of the associated statistical methods.

Materials and Methods

Statistical models

Let y_{ij} denote the proportional daily yield of the *j*-th milking for the *i*-th cow, and t_{ij} represent the corresponding milking interval (h). Wiggans (1986) assumed a linear relationship between these two variables:

$$y_{ij} = \alpha_j + \beta t_{ij} + e_{ij}. \quad (1)$$

where α_j is the intercept for milking *j*, β is the common regression coefficient across all three milkings, and e_{ij} is the residual effect.

The non-zero, intercept α_j values represent section-specific offsets that quantify the model-estimated yield proportion at $t = 0$ for each milking, which is merely a mathematical construct rather than a biological reality.

The common regression coefficient β indicates that the rate of change in proportional yield per additional hour of interval is the same across all milkings. Hence, biological differences among milkings are captured by their distinct intercepts.

Furthermore, Wiggans (1986) proposed correcting the intercepts to a 24-hour basis after fitting the model. For $t_{i1} + t_{i2} + t_{i3} = 24$,

$$\sum_{j=1}^{3} \hat{\alpha}_j = 1 - 24\hat{\beta}. \quad (2)$$

Mathematically, the difficulty with this form of *a posteriori* rescaling lies in the non-uniqueness of the solution. The imposed scaling constraint yields only one equation, yet three intercept parameters must be estimated, even when the regression coefficients are fixed.

One arbitrary approach is to define equal deviation adjustment of the intercepts across the three milkings, that is,

$$\delta = \frac{1}{3}\left(1 - \sum_{j=1}^{3} \hat{\alpha}_j - 24\hat{\beta}\right),$$

such that

$$\alpha_j^* = \hat{\alpha}_j + \delta = \hat{\alpha}_j + \frac{1}{3}\left(1 - \sum_{j=1}^{3} \hat{\alpha}_j - 24\hat{\beta}\right) \quad (3)$$

Then, summing the three intercepts confirms that the equality constraint (2) is satisfied.

Nevertheless, the above adjustment of the intercepts is statistically suboptimal. The key issue is that the constraint is not imposed during model estimation, but only afterward. As a result, the final adjusted parameters do not minimize the residual sum of squares under the constraint. Instead, we propose an alternative approach that incorporates the constraint directly during model fitting. This specification avoids the unrealistic extrapolation to $t = 0$ and reduces collinearity between the intercept and interval terms.

Specifically, in a 3× milking system, we define an 8-h interval as the physiological and management baseline. Deviations from this interval are interpreted as positive or negative shifts in proportional yield. Under this model setting, the intercepts reflect systematic differences among milkings at the baseline interval, whereas the slope quantifies the change in yield proportion per hour of deviation from the 8-h reference.

When the totality of a 24-hour daily milking interval holds precisely, i.e., $t_{i1} + t_{i2} + t_{i3} = 24$, the constraint for the intercepts becomes:

$$\sum_{j=1}^{3} \hat{\alpha}_j = 1. \quad (4)$$

In matrix notation, the reparameterized model can be expressed as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{b} + \boldsymbol{e}, \quad (5)$$

where $\boldsymbol{y}$ is a vector of proportional daily yields and $\boldsymbol{e}$ is the residual vector.

The design matrix (X) and the parameter vector (b) are defined as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & t_{i1} - 8 \\ 0 & 1 & 0 & t_{i2} - 8 \\ 0 & 0 & 1 & t_{i3} - 8 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}; \boldsymbol{b} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \end{bmatrix}.$$

The equality constraint on the intercepts is implemented as follows:

$$\boldsymbol{cb} = d, \quad (6)$$

where $\boldsymbol{c} = (1 \quad 1 \quad 1 \quad 0)$ and $d = 1$.

Mathematically, this formulation represents a quadratic programming problem with a linear equality constraint, and its solution is given by:

$$\widehat{\boldsymbol{b}} = \widehat{\boldsymbol{b}}_0 - (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{c}'(\boldsymbol{c}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{c}')^{-1} \times \left(\boldsymbol{c}\widehat{\boldsymbol{b}}_0 - d\right) \quad (7)$$

where $\widehat{\boldsymbol{b}}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$ is the unconstrained OLS estimator, and

$$\delta = (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{c}'(\boldsymbol{c}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{c}')^{-1}\left(\boldsymbol{c}\widehat{\boldsymbol{b}}_0 - d\right)$$

is the adjustment term.

The variance-covariance matrix of the constrained estimator is:

$$var\left(\widehat{\boldsymbol{b}}\right) = \sigma^2[(\boldsymbol{X}'\boldsymbol{X})^{-1} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}'(\boldsymbol{c}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}')^{-1}c(X'X)^{-1}]$$

Here, σ^2 is the residual variance from the constrained OLS model.

The constrained OLS model can also be implemented through reparameterization. Let

$$\alpha_3 = 1 - \alpha_1 - \alpha_2. \quad (8)$$

Substituting Equation (8) into model (1) and rearranging the terms to group coefficients by unknown parameter yields:

$$y_{ij} = (x_{i1} - x_{i3})\alpha_1 + (x_{i2} - x_{i3})\alpha_2 + x_{i3} + \beta t_{ij} + e_{ij}.$$

where x_{ij} is a dummy variable which equal 1 if the partial yield is observed or 0 otherwise.

Moving x_{i3} to the left-hand side as a part of the data gives:

$$y_{ij} - x_{i3} = (x_{i1} - x_{i3})\alpha_1 + (x_{i2} - x_{i3})\alpha_2 + \beta t_{ij} + e_{ij}.$$

The re-parameterized model can thus be expressed as:

$$y_{ij}^* = x_{i1}^*\alpha_1 + x_{i2}^*\alpha_2 + t_{ij}\beta + e_{ij}, \quad (9)$$

where $y_{ij}^* = y_{ij} - x_{i3}$, $x_{i1}^* = (x_{i1} - x_{i3})$, and $x_{i2}^* = (x_{i2} - x_{i3})$.

The parameter values, α_1, α_2, and β, in model (9) can then be obtained using OLS. Finally, the third intercept (α_3) is recovered from Equation (8).

Analyzing 3X milking data

Milking data were collected from 2023 to 2025 at four 3× milking Holstein dairy farms (A, B, C, and D) across three U.S. states. Milkings were classified by start time: 1 (02:00 to 09:59 h), 2 (10:00 to 17:59 h), or 3 (18:00 to 01:59 h). Milkings were recorded weekly for each cow up to 120 days in milk, then monthly until 305 days or the end of lactation. Data cleaning removed duplicate, missing, and incomplete records, leaving 15,702 × 3 (Farm A), 13,574 × 3 (Farm B), 8,515 × 3 (Farm C), and 10,821 × 3 (Farm D) milking observations.

Four models were evaluated: constrained OLS (CLS); reparameterized OLS (RLS), which is algebraically equivalent to CLS; the Wiggans (1986) model fitted with historical parameter estimates (WLS0); and the Wiggans (1986) model refitted to the current data (WLS1). The accuracy of the estimated yields was assessed using out-of-sample validation, with approximately 50% of the records allocated to training and the remaining 50% to testing; each validation was repeated ten times. Accuracy measures included the correlation between observed and predicted 24-h yields, mean absolute error (MAE), and mean squared error (MSE).

Results and Discussion

Estimation of model parameters

The CLS and RLS produced identical model parameters across all three traits, confirming that the two formulations are mathematically equivalent (Table 1). Also see Figure 1. Under the CLS specification, we consistently observed that $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 \approx 1$, indicating that the estimated daily yields are properly confined to a 24-h period.

In contrast, WLS1, which implemented the Wiggans (1986) approach, yielded parameter estimates substantially different from those of CLS and RLS, even though all three models fit the same data. Nevertheless, WLS1 can still be viewed as an approximately equivalent representation of the CLS model after an

algebraic transformation. To illustrate, consider the CLS equation for proportional milk yield at the first milking:

$$y = 0.334 + 0.0381 \times (t - 8)$$
$$= (0.334 - 0.0381 \times 8) + 0.0381 \times t$$
$$\approx 0.0292 + 0.0381t.$$

Table 1: Estimated model parameters from recently collected milking data, compared with the old estimates by Wiggans in 1986 [1,2]

Method	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}$
		Milk yield		
RLS	0.334	0.332	0.334	0.0381
CLS	0.334	0.332	0.334	0.0381
WLS1	0.029	0.028	0.029	0.0381
WLS0	0.077	0.068	0.066	0.0329
		Fat yield		
RLS	0.336	0.338	0.326	0.0246
CLS	0.336	0.338	0.326	0.0246
WLS1	0.139	0.142	0.129	0.0246
WLS0	0.186	0.186	0.182	0.0186
		Protein yield		
RLS	0.332	0.334	0.334	0.0390
CLS	0.332	0.334	0.334	0.0390
WLS1	0.020	0.022	0.021	0.0390
WLS0	0.072	0.065	0.063	0.0333

[1] CLS = constrained ordinary least squares; RLS = reparameterized CLS; WLS0 = Wiggans (1986) model fitted to historical data; WLS1 = Wiggans (1986) model fitted to current data.
[2] Values in the table are means across ten replicates.

This equation closely matched the WLS1 equation for the first milking, yielding identical intercepts and regression coefficients. The same was true for the equations for fat and protein.

When comparing the Wiggans (1986) model fitted to the current dataset (WLS1) with the model fitted to the historical data (WLS0), we observed clear shifts in model parameters over the past four decades: intercept estimates based on the recent data were much smaller, and the regression coefficients were larger than those reported in 1986.

The intercept parameter α_j represents a section-specific offset that captures systematic milking effects not explained by interval length (e.g., diurnal management routines or preparation differences). In the Wiggans (1986) model, because the intercept corresponds to the predicted proportional yield at $t = 0$, which lies outside the biologically meaningful range of milking intervals, its absolute magnitude should not be interpreted literally. Instead, the relative differences among intercepts reflect persistent structural differences among milkings that are not attributable to interval length.

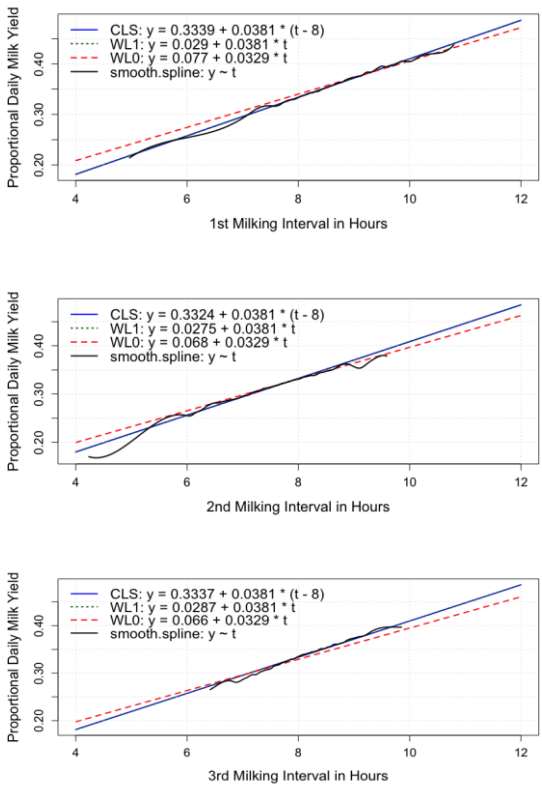


Figure 1. Illustration of linear regression models and smoothing splines for proportional daily yields over the milking interval (in hours): 1st milking (left); 2nd milking (middle); 3rd milking (right).

Across all three traits, intercept estimates for the three milkings were substantially lower (27.9–76.4%) in the current dataset than in the historical dataset. This reduction in between-section variability suggests that structural differences among the three milkings may have diminished over time, possibly due to more standardized management routines, improved milking-system precision, and more consistent cow handling.

It is noted that part of the observed reduction in intercept magnitudes is also mathematically consistent with an increase in the regression coefficient $\hat{\beta}$ under the imposed 24-h constraint,

$\sum_{j=1}^{3}\hat{\alpha}_j = 1 - 24\hat{\beta}$. Increasing the regression coefficient $\hat{\beta}$ necessarily reduces the sum of intercepts $\sum_{j=1}^{3}\hat{\alpha}_j$ in order to maintain the 24-h equility constraint. Thus, a portion of the smaller intercept value may reflect an algebraic consequence of the higher slope estimate: a larger regression coefficient mechanically forces a lower intercept.

The regression coefficient $\hat{\beta}$ represents the marginal change in proportional yield per additional hour of the milking interval. For milk yield, this coefficient rose from 3.29% per hour in 1986 to 3.81% per hour in the current dataset. Similarly, the per-hour change increased from 1.86% to 2.46% for fat yield and from 3.33% to 3.90% for protein yield. In relative terms, the largest increase was in protein yield (132.3%), followed by fat (117.1%) and milk (115.8%).

From a biological perspective, a larger $\hat{\beta}$ value indicates that proportional yield has become more responsive to interval length than it was four decades ago. Several factors likely contribute to this shift. Improvements in milking preparation, adoption of automatic take-off systems based on milk flow, and enhanced milking completeness may have reduced residual milk, thereby strengthening the proportional relationship between milk accumulation and elapsed time. In addition, long-term genetic selection and improved nutrition have increased milk secretion rates and cisternal storage capacity, making milk accumulation more linear and more tightly coupled to interval length.

A mathematical explanation is as follows. Let $S(t)$ denote the cumulative milk secretion since the previous milking, and let the daily yield be $Y = S(t_1) + S(t_2) + S(t_3)$. The proportional yield for a given milking is

$$y(t) = \frac{S(t)}{Y}.$$

If secretion were strictly linear over the relevant interval (6–12 h), then $S(t) \approx rt$, where r is the secretion rate. In that case,

$$y(t) \approx \frac{rt}{r(t_1+t_2+t_3)} = \frac{t}{24},$$

Hence, under the linear proportional yield model, the "theoretically perfect" slope is $1/24$ (4.17%) per hour for all three traits. In practice, however, milk accumulation rates deviate from perfect linearity due to residual milk, incomplete udder emptying, and temporary storage limitations. In the present study, the observed slopes were all smaller, ranging from 59.0% (fat) to 91.4% (milk) and 93.5% (protein) relative to the "perfect" slope.

Instead, a more general representation is

$$S(t) = rt - \delta(t),$$

where $\delta(t)$ captures nonlinear effects such as carryover or incomplete removal. Then, the derivative of proportional yield with respect to interval length is

$$\frac{dy}{dt} = \frac{S'(t)}{Y} = \frac{rt-\delta(t)}{24r-\Delta}$$

where $t_1 + t_2 + t_3 = 24$ and $\Delta = \delta_1(t_1) + \delta_2(t_2) + \delta_3(t_3)$.

Now, expand around the "ideal" case $\delta'(t) = 0$ and $\Delta = 0$. Using

$$\frac{1}{24r-\Delta} = \frac{1}{24r} \cdot \frac{1}{1-\Delta/(24r)} \approx \frac{1}{24r}\left(1+\frac{\Delta}{24r}\right)$$

for small $\Delta/(24r)$, we get

$$\frac{dy}{dt} \approx (r - \delta'(t)) \cdot \frac{1}{24r}\left(1+\frac{\Delta}{24r}\right)$$
$$= \frac{1}{24}\left(1-\frac{\delta'(t)}{r}\right)\left(1+\frac{\Delta}{24r}\right). \qquad (10)$$

In the above, the factor $\left(1-\frac{\delta'(t)}{r}\right)$ captures local nonlinear effects over interval t. If $\delta'(t) > 0$, meaning that nonlinear losses (e.g., residual milk or incomplete removal) increase with interval length, this term is less than 1, and the slope is reduced below $1/24$. Conversely, if nonlinear distortion diminishes, the slope approaches the ideal value.

The factor $\left(1+\frac{\Delta}{24r}\right)$ captures global day-level effects, where $\Delta = \sum_{j=1}^{3}\delta_j(t_j)$ represents total nonlinear deviation across the day. Because Δ enters through the denominator $Y = 24r - \Delta$, it adjusts the proportional scaling of secretion relative to total daily yield.

Hence, assuming ideal linear secretion with no nonlinear effects ($\delta'(t) = 0$ and $\Delta = 0$), we obtain $\frac{dy}{dt} = \frac{1}{24} = 0.0417$, meaning that the proportional yield increases exactly in proportion to time. In reality, the observed slope depends on the contrast between $\frac{\delta'(t)}{r}$ and $\frac{\Delta}{24r}$.

Our results suggest $\frac{\delta'(t)}{r} > \frac{\Delta}{24r}$ for the three traits. The larger the contrast, the smaller the regression coefficient will be. Consider a special case where $\frac{\delta'(t)}{r} \equiv \frac{\Delta}{24r}$. Then, we will always observe that $\frac{dy}{dt} \leq 0.417$.

Accuracy of estimated daily yields

The accuracy of yield estimates from different models is shown in Table 2. When fitted to the recent data, all models produced nearly identical accuracy measures for the three traits. The correlation between actual and estimated daily yields was 0.906 for milk, 0.833 for fat, and 0.885 for protein yield. MSE was approximately 3.03 kg for milk, 0.183 kg for fat, and 0.094 kg for protein.

MSE followed a pattern similar to MAE across models. Because the actual total daily milking interval was close to 24 hours (23.9 h for Herd 1, 24.2 h for Herd 2, 23.6 h for Herd 3, and 24.2 h for Herd 4), the 24-h normalization constraint had little practical impact on predictive performance, regardless of whether it was imposed during model fitting or applied a posteriori. Therefore, when the sum of observed intervals already approximates 24 h, enforcing the unity constraint on proportional yields produces only negligible adjustments to the fitted values.

From a statistical perspective, the constrained OLS estimator retains the best linear unbiased estimator (BLUE) property when the imposed constraint is correctly specified, which, however, does not guarantee a marginally globally minimized MSE. However, if the constraint is misspecified, the resulting estimator becomes biased because the true parameter vector does not lie within the constrained parameter space. Even when the constraint is valid, our results indicate that empirical gains in predictive accuracy are minimal. Under these considerations, the 24-h adjustment of intercepts may not be strictly necessary in practice, as its efficiency gains are modest and contingent upon the correctness of the imposed restriction. Omitting the 24-h constraint in the Wiggans (1986) framework would simplify implementation protocols (e.g., ICAR, 2022) for computing daily milk, fat, and protein yields from partial milking records.

Table 2. Accuracy of estimated milk, fat, and protein yields using different methods [1,2,3]

Traits	Metrics	RLS	CLS	WLS1	WLS0
MY	COR	0.906	0.906	0.906	0.904
	MAE	3.03	3.03	3.03	3.08
	MSE	21.7	21.7	21.7	22.1
FY	COR	0.833	0.833	0.833	0.831
	MAE	0.183	0.183	0.183	0.184
	MSE	0.071	0.071	0.071	0.072
PY	COR	0.885	0.885	0.885	0.883
	MAE	0.094	0.094	0.094	0.957
	MSE	0.020	0.020	0.020	0.022

[1,2] CLS, RLS, WLS0, and WLS1: see Table 1 for the abbreviations and explanations;
[3] Correlation = correlation between the estimated and actual daily yields; MAE = mean absolute error of the estimated yield; MSE = mean squared error of the estimated yield.

Over the past four decades, models fitted to current data generally achieved better fit on the training data, but only slightly higher corrections and lower errors (MAE and MSE) than those based on historical data. These results likely reflect the inherent two-step nature of such a proportional-yield modeling strategy. A better-fitting model yields a more accurate average projection of proportional yield over the milking interval. However, the estimated yield is obtained in the second step using the so-called "multiplicative yield correction factor" (Wu et al., 2022), and its accuracy depends on variations in partial yields.

Direct evidence comes from the marked differences in estimated yields across farms. Across ten replicate runs, the correlation between observed and estimated 24-h milk yields was 0.946 - 0.948 for Farm B, 0.905 - 0.910 for Farm A, and 0.890 – 0.895 for Farms C and D. In contrast, Farm A had a higher accuracy for fat yields (0.861–0.862) than the other three herds (0.792–0.804). Farm A also had higher correlations for fat protein yields (0.920–0.921) than the other three farms (0.859 – 0.874).

These large differences among farms indicate that model performance is strongly shaped by herd-level factors, including management consistency, milking-system precision, and data quality. Herds with more uniform milking intervals, accurate sensor calibration, and consistent cow-handling practices tend to produce more stable proportional-yield patterns, resulting in higher correlations and lower prediction errors. Conversely, greater variability in milking timing, equipment performance, or cow movement introduces additional noise into the proportional-yield relationship, thereby reducing model accuracy.

Given the substantial heterogeneity across the four Holstein herds, it is not surprising that models fitted to the current data only marginally outperformed those based on historical parameters. Between-herd variability in management practices, milking systems, and data quality likely contributed more to prediction error than temporal shifts in the proportional-yield relationship, particularly with a limited number of herds. In contrast, the historical dataset (Wiggans, 1986) comprised 47,495 cow-days across 65 herds. Hence, including more herds may have produced more stable parameter estimates by averaging over a wider range of management conditions.

Conclusions

We observed clear shifts in model parameters over the past four decades. The intercept estimates from recent data were smaller, whereas the regression coefficients were larger than those reported in 1986. This pattern indicates that the dynamic responsiveness of proportional yield to interval length has increased over time, while baseline section-specific effects have become relatively smaller.

Despite these changes in parameter estimates, improvements in daily yield accuracy were modest. However, substantial variability across farms suggests that herd-level factors—such as management practices, milking system precision, and data consistency—contribute more to predictive performance than temporal changes in model structure. Expanding the training dataset, particularly by including more representative herds and cow-day records, may further improve predictive stability and generalizability. Imposing a 24-h normalization constraint improves model interpretability and ensures coherent proportional estimates, but it does not necessarily improve predictive accuracy.

Overall, this study reaffirms the ongoing relevance of proportional-yield modeling and existing correction factors as practical benchmarks and offers updated empirical guidance for estimating daily yields in three-times-daily milking systems.

References

DeLorenzo, M. A., and Wiggans, G. R. 1986. Factors for Estimating Daily Yield of Milk, Fat, and Protein from a Single Milking for Herds Milked Twice a Day. *J. Dairy Sci.* 69, 2386–2394.

ICAR. 2022. ICAR Guidelines: Section 2 – Computation of 24-hour Yields (Procedure 1). International Committee for Animal Recording, Rome, Italy. Available from: https://wiki.icar.org/index.php/Guidelines

Liu, Z., Reents, R., Reents, R., Reinhardt, F., and Kuwan, K. (2000). Approaches to Estimating Daily Yield from Single Milk Testing Schemes and Use of a.m.-p.m. Records in Test-Day Model Genetic Evaluation in Dairy Cattle. *J. Dairy Sci*. 83, 2672–2682.

Wiggans, G. R. 1986. Estimating Daily Yields of Cows Milked Three Times a Day. *J. Dairy Sci.* 69, 2935–2940.

Wu XL, Wiggans GR, Norman HD, Miles AM, Van Tassell CP, Baldwin Vi RL, Burchard J, Durr J. 2022. Daily milk yield correction factors: What are they? *JDS Commun*. 4:40-45

Estimation of across-breed metafounder parameters in all-breed US dairy cattle.

A. Legarra[1,2] and M. Bermann[2]

[1] *Council on Dairy Cattle Breeding, Bowie, MD 20716, United States*

[2] *Animal and Dairy Science, University of Georgia, Athens, GA 30602, United States*

Corresponding author: andres.legarra@uscdcb.com

Abstract

Metafounders relationships **Γ** are useful to model missing pedigrees and different breeds. They are estimated from markers and pedigree, yet their estimation is difficult. Here we explore the behavior of two algorithms: General Least Squares (GLS) and Pseudo-Expectation Maximization (Pseudo-EM) to estimate **Γ** . We use a reduced set of US genotypes and pedigree, consisting in six breeds: Ayrshire, Brown Swiss, Guernsey, Holstein and Jersey, with number of genotypes (69 200 SNP markers) equal to 350, 681, 237, 13492 y 3541, with 370 307 animals in the overall pedigree. **Γ** estimates were also transformed to Fst differentiation indexes for ease of interpretation. Algorithms were fast. GLS resulted in biased estimates, with too high diagonal values of **Γ** for small breeds, and too high differentiation indexes across breeds. Algorithm Pseudo-EM resulted in more sensible estimates. Further, estimates of allele frequencies by Pseudo-EM agreed well with upstream estimates of base allele frequencies based on imputation. We recommend use of Pseudo-EM.

Key words: Metafounders, relationships, missing pedigree, breeds

Introduction

Disclaimer: a more complete manuscript has been submitted for peer-review publication. We provide this version as a short reference for the Interbull community. We also present some different results.

Metafounders (MF) are a conceptual framework to model relationships of unknown ancestors, both across breeds, and within breeds across time and selection pathway. They are particularly useful in single-step implementations, in particular for all-breed evaluations like the prospective US one. These relationships are condensed in a matrix **Γ**. Here we describe experiences on the estimation of **Γ** using two algorithms: Maximum Likelihood by Pseudo-Expectation-Maximization (Pseudo-EM) and General Least Squares (GLS). We used the newly released software gammaf90, a part of the blupf90 suite of programs (https://nce.ads.uga.edu/wiki/doku.php?id=start). We characterize a simple case where MF are essentially defined by breed and pathway (males or females ancestors). For this work we focus on results from the females ancestors (i.e. "dam" MF), as these are more representative of the general population.

Materials and Methods

Both algorithms need pedigree and markers. We used pedigree and markers extracted for the April 2025 evaluation of gestation length, which uses a sire model. This included 18 301 genotyped animals with a pedigree including ancestors up to 370 307 animals. All genotyped animals are purebred. Breeds included and genotype counts were Ayrshire (AY, 350), Brown Swiss (BS, 681), Guernsey (GU, 237), Holstein (HO, 13942) and Jersey (JE, 3541). We defined 11 metafounders: dam and sire MF for AY, BS and JE; a single one for GU; dam and sire for HO pre-1994 and dam and sire for

HO post-1994. The latter two were essentially for imported animals and are essentially ignored in the following results. SNP genotypes were the official 69 200 SNP CDCB list, after imputation.

Because we do not split MF across years, we are using a somehow fuzzy definition of base population. The mode of the year of birth of animals whose parents are missing is essentially 1950 to 1960 for all breeds, so our **Γ** estimates refer to that point.

Algorithms

Algorithms were as detailed in Legarra et al. (2024) although that article has some typos. Therefore, here we detail them again for completion. The definition of **Γ** if we knew the allele frequencies of two populations *i* and *j* is $\Gamma_{(i,j)} = \left(\frac{2}{k}\right)(2\boldsymbol{p}_i - 1)\left(2\boldsymbol{p}_j - 1\right)'$. Note that the same expression works for a population with itself.

In GLS, we estimate base allele frequencies using BLUE for $\widehat{\boldsymbol{p}}$ or equivalently, its GLS estimator (Gengler et al., 2007; Garcia-Baccino et al., 2017). Then we plug the estimates to obtain $\widehat{\boldsymbol{\Gamma}}_{GLS} = \frac{2}{k}\left(2\widehat{\mathbf{P}} - \mathbf{11}'\right)\left(2\widehat{\mathbf{P}} - \mathbf{11}'\right)'$. There is no iteration involved.

In Pseudo-EM we start with a guess of $\boldsymbol{\Gamma} = 0.1\mathbf{I}$ and iterate on (at iteration t):

1. Build $\mathbf{A}_{\Gamma(t)}$
2. For each marker i
   - $\widehat{2\boldsymbol{p}}_{i(t)} - \mathbf{1} = \mathbf{A}_{\Gamma(2,mf)(t)}\mathbf{A}^{-1}_{\Gamma 22(t)}\mathbf{z}_i$
3. $\widehat{\boldsymbol{\Gamma}}_{(t)} = \widehat{\boldsymbol{\Gamma}}_{(t-1)} - \mathbf{A}_{\Gamma(t)(mf,2)}\mathbf{A}^{-1}_{\Gamma(t)(2,2)}\mathbf{A}_{\Gamma(t)(2,mf)} + \frac{2}{k}\left(2\widehat{\mathbf{P}}_t - \mathbf{11}'\right)\left(2\widehat{\mathbf{P}}_t - \mathbf{11}'\right)'$

Where **z** contains genotypes coded as {-1,0,1} The likelihood is monitored. At convergence of the algorithm, we have estimates for **P** and **Γ**.

Results & Discussion

Computation took reasonable time – 13 hours for Pseudo-EM, and 3 minutes for GLS. Pseudo-EM took 950 iterations to reach a convergence criterion (squared relative difference of successive estimates) of 10^{-11}. Convergence was slow for elements with little information, e.g. Ayrshire (the least represented breed) with itself, but fast for well represented elements, e.g. Jersey with Holstein. Likelihood increased monotonically for Pseudo-EM (Figure 1), with the exception of the very first rounds, where there's a decrease – we believe that this is due to numerical round-off.

We report values only for the dam pathway (and in HO, for the dam pathway pre-1994). Table 1 and 2 contain estimates of Gamma coefficients. Table 3 contains Fst (Wright's fixation index) of differentiation across breeds, which is obtained as a direct transformation of **Γ** estimates using (Legarra et al., 2024): $F_{ST(i,j)} = \frac{\Gamma_{i,i} + \Gamma_{j,j} - 2\Gamma_{i,j}}{2+2-2\Gamma_{i,j}}$.

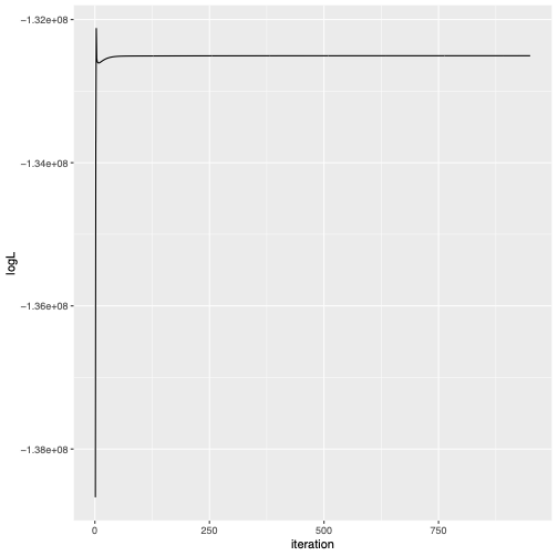


Figure 1. Evolution of the log-Likelihood during iterations of Pseudo-EM.

Table 1: Estimated Gamma values across breeds, General Least Squares method.

	AY	BS	GU	HO	JE
AY	1.20	0.17	0.21	0.18	0.18
BS	0.17	1.05	0.26	0.21	0.23
GU	0.21	0.26	0.60	0.26	0.31
HO	0.18	0.21	0.26	0.60	0.24
JE	0.18	0.23	0.31	0.24	0.85

Table 2: Estimated Gamma values across breeds, Pseudo-Expectation Maximization

	AY	BS	GU	HO	JE
AY	0.57	0.29	0.30	0.27	0.26
BS	0.29	0.59	0.31	0.25	0.29
GU	0.30	0.31	0.53	0.26	0.33
HO	0.27	0.25	0.26	0.39	0.26
JE	0.26	0.29	0.33	0.26	0.58

Table 3: Estimated Fst values across breeds. Upper triangular: General Least Squares method. Lower triangular: Pseudo-Expectation Maximization

	AY	BS	GU	HO	JE
AY		0.52	0.38	0.39	0.47
BS	0.17		0.32	0.34	0.41
GU	0.15	0.15		0.19	0.24
HO	0.12	0.14	0.11		0.27
JE	0.18	0.17	0.14	0.13	

The values of **Γ** in Tables 1 and 2 are hard to interpret. However, the values in the diagonal indicate homozygosity, and for GLS we see high values for small breeds such as AY or BS. This is unexpected. In Table 2 the values in the diagonal smaller and more similar across breeds, which agrees better with our knowledge of the breeds. Fst coefficients in Table 3 are in a better understood scale, with previously published reports (e.g. Gibbs et al., 2009). Typical Fst values across dairy European breeds fall in the 0.10 – 0.20 interval, which we found in estimates by Pseudo-EM. However, GLS gives very high values, comparable to differentiations between *Bos Indicus* and *Bos Taurus*. This concerns even breeds with abundant information such as HO and JE. We conclude that the GLS algorithm is generally biased.

Finally, Figure 2 presents a scatterplot of allele frequencies in BS as estimated by the imputation process at CDCB (upstream of the files used here) using findhap.f90 (VanRaden et al. 2013), and by Pseudo-EM. The correlation between Pseudo-EM and Imputation-based frequencies is 0.95. Note that Imputation uses a more correct model considering haplotype tracing, recombinations, and linkage disequilibrium patterns (VanRaden et al. 2013), whereas Pseudo-EM uses an approximated linear model which ignores the physical nature of the chromosome (Legarra et al., 2024). It is satisfactory to see that both methods largely agree.

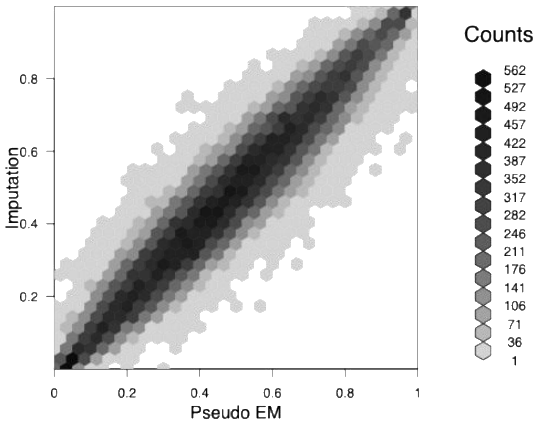


Figure 2. Hexbin scatterplot of base allele frequencies in BS, estimated by Pseudo-EM and Imputation.

Conclusions

For this data set, Pseudo-EM resulted in seemingly unbiased estimates of the MF matrix of relationships **Γ**, unlike GLS which gave biased estimates as observed in Fst values. Results concerning allele frequencies from Pseudo-EM agree well with Imputation results. Computing time was acceptable. We recommend Pseudo-EM.

Acknowledgments

This work uses the National Cooperator Database stewarded by CDCB. We would like to acknowledge the participating dairy producers for supplying data, DHI

organizations and DRPCs for processing and relaying the information to the Council on Dairy Cattle Breeding (CDCB) and the purebred breed associations for providing pedigree data.

References

Garcia-Baccino, C.A., A. Legarra, O.F. Christensen, I. Misztal, I. Pocrnic, Z.G. Vitezica, and R.J. Cantet. 2017. Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. *Genet. Sel. Evol*. 49:1,14.

Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* 1:21,28.

The Bovine HapMap Consortium, R.A. Gibbs, et al. 2009. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* 324,528–532. https://doi.org/10.1126/science.1167936.

Legarra, A., M. Bermann, Q. Mei, and O.F. Christensen. 2024. Redefining and interpreting genomic relationships of metafounders. *Genet. Sel. Evol.* 56,34. https://doi.org/10.1186/s12711-024-00891-w.

VanRaden, P.M., D.J. Null, M. Sargolzaei, G.R. Wiggans, M.E. Tooker, J.B. Cole, T.S. Sonstegard, E.E. Connor, M. Winters, J.B.C.H.M. van Kaam, A. Valentini, B.J. Van Doormaal, M.A. Faust, and G.A. Doak. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* 96:668,678. https://doi.org/10.3168/jds.2012-5702 .

Effects of weighting schemes and base adjustment on genomic validation outcomes in the Interbull GEBVtest framework

R.R. Mota[1], S. Savoia[1], A. Legarra[1], T.M. McWhorter[1], E. Nicolazzi[1]

[1] Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

Abstract

Genomic validation is required to ensure the reliability and international acceptance of genomic estimated breeding values (GEBV). Recent updates to the Interbull GEBVtest software introduced alternative options for checking, including different regression equations, weights, and base-adjustment procedures. The objective of this study was to evaluate the effects of these options on genomic validation outcomes in U.S. dairy cattle populations. Genomic validations were performed using the Interbull GEBVtest software and August 2025 U.S. genomic evaluations. A reduced dataset was generated by truncating domestic information after 2021 and combining the resulting national evaluations with foreign evaluations from Truncated Multiple Across Country Evaluation (TMACE) information from Interbull. Five dairy breeds and nine production, health, longevity, and fertility traits were evaluated. All validation scenarios used predicted deregressed genomic estimated breeding values (dGEBV) as the validation target. Four scenarios were compared, differing in regression methodology [Interbull weighted regression (ITB) or unweighted Legarra-Reverter (LR) regression] and application [yes/no] of base adjustment. A total of 42 breed-trait combinations were evaluated under four validation scenarios, resulting in 168 scenario-specific validation analyses. Across all analyses, 91 resulted in PASS outcomes, 18 in FAIL outcomes, 12 in WARN outcomes, and 47 in HiSE outcomes. WARN indicates validation statistics close to predefined acceptance thresholds, whereas HiSE indicates inconclusive validation results due to high slope standard errors. Validation results were generally consistent across scenarios, indicating that the genomic validation framework was robust for most breed-trait combinations. However, a limited number of analyses produced different validation outcomes depending on the methodological configuration applied. These differences were associated with interactions among regression methodology, base adjustment, and validation-group definition rather than any single factor alone. However, for traits that failed validation in some scenarios, validation statistics classified as WARN remained close to the threshold for FAIL. The results demonstrate that genomic validation outcomes are sensitive to details when validation statistics are close to predefined acceptance thresholds. A WARN result close to a FAIL threshold should prompt further investigation of the factors contributing to potential bias. The observed FAIL outcomes may be associated with limitations of the current multiple-step evaluation framework. CDCB is transitioning to single-step evaluation, which should impact validation performance.

Key words: base adjustment, candidate-bull selection, GEBVtest, validation target, weighting schemes

Introduction

Genomic estimated breeding values (GEBV) are widely used in dairy cattle breeding programs for the early selection and international marketing of young bulls. To ensure confidence in genomic predictions, countries participating in international evaluations are required to validate their genomic evaluation systems before GEBV can be used as input for Interbull services and accepted for routine publication. Since its introduction, the Interbull GEBV test has become the standard framework for genomic validation and has undergone several modifications to accommodate changes in genomic evaluation methodologies and data

structures (Mäntysaari et al., 2011; Interbull, 2021).

The rapid adoption of genomic selection has substantially altered dairy cattle populations and evaluation systems. Genomic preselection, increasing use of foreign information, and the implementation of single-step genomic evaluation models have created challenges for traditional validation procedures. In response, several studies have proposed methodological improvements to better account for these changes and to provide validation procedures that remain robust under modern genomic evaluation systems (Legarra and Reverter, 2018; Legarra and Reverter, 2019; VanRaden, 2021).

Recent updates to the Interbull validation software introduced new validation targets, alternative weighting schemes, and procedures for aligning evaluation scales between full and reduced datasets (Sullivan, 2023). Among these updates, the use of deregressed genomic estimated breeding values (dGEBV) as validation targets and the implementation of base-adjustment procedures were intended to reduce potential biases associated with genomic preselection and scale differences between datasets and evaluation time points. The software also allows, for exploration, the application of either the Interbull weighted regression approach or an unweighted regression framework similar to the Legarra-Reverter (LR) methodology described by Legarra and Reverter (2018, 2019).

Previous studies conducted in the United States demonstrated that genomic validation results are generally more stable for large dairy populations and highly heritable traits, whereas smaller populations and more complex traits remain challenging due to small numbers of validation animals yielding increased variability in validation statistics (Mota et al., 2023). More recently, the use of Truncated Multiple Across Country Evaluation (TMACE) evaluations as reference information improved validation consistency by reducing the effects of model and data changes accumulated over time and by allowing validation of traits implemented after the traditional four-year validation window (Mota et al., 2024).

Although the updated validation software provides several methodological options, limited information is available regarding the sensitivity of validation outcomes to different combinations of weighting schemes and base-adjustment procedures. Understanding the magnitude of these effects is important because validation results can influence the interpretation of genomic evaluation performance and, in some cases, may alter validation conclusions when statistics are close to predefined acceptance thresholds.

Therefore, the objective of this study was to evaluate the effects of weighting schemes and base adjustment on genomic validation outcomes in U.S. dairy cattle populations using the current Interbull GEBV validation framework. Emphasis was placed on comparing alternative validation scenarios and assessing the consistency of validation outcomes across methodological configurations.

Materials and Methods

Genomic validations were conducted using the Interbull GEBVtest software (gebvtest.py), corresponding to the version available from the Interbull Centre (Interbull, 2026). The software incorporates the most recent updates to validation statistics, candidate-bull definitions, and validation procedures described by Sullivan (2023).

The genomic prediction datasets in US dairy cattle populations were GEBV extracted from the August 2025 genomic evaluation (full dataset), which included MACE input, and truncated GEBV to the year of 2021 plus TMACE input (truncated dataset). TMACE follows the same principles as conventional MACE evaluations but uses truncated national information as input, thereby providing international information consistent with the reduced-data scenario (Jorjani and Dürr, 2011).

Five dairy breeds were included in the analyses: Holstein (HOL), Jersey (JER), Brown Swiss (BSW), Red Dairy Cattle (RDC), and

Guernsey (GUE). Nine traits representing production, health, longevity, and fertility were evaluated: milk yield (MIL), fat yield (FAT), protein yield (PRO), longevity (DLO), somatic cell score (SCS), clinical mastitis (MAS), heifer conception rate (HCO), cow conception rate (CC1), and daughter pregnancy rate (INT). Heifer conception rate was evaluated for all breeds except GUE, whereas clinical mastitis was evaluated for BSW, HOL, and JER, reflecting the breeds for which routine national evaluations were available.

All validation scenarios used predicted deregressed genomic estimated breeding values (dGEBV) as the validation target. Validation using GEBV-derived targets is easier to communicate to stakeholders and allows straightforward application in national evaluation systems that have adopted single-step methodologies (VanRaden, 2021). In addition, the Council on Dairy Cattle Breeding (CDCB) is currently transitioning toward single-step genomic evaluations with direct GEBV output, making dGEBV a more appropriate validation target for future implementation. Therefore, the validation target was kept constant across all scenarios so that the effects of weighting schemes and base adjustment could be evaluated independently.

Four validation scenarios were investigated (Table 1). The scenarios differed according to regression methodology and application of base adjustment. Scenarios 1 and 2 used the Interbull weighted regression approach (ITB), whereas Scenarios 3 and 4 used the unweighted LR regression approach described by Legarra and Reverter (2018, 2019). Base adjustment was not applied in Scenarios 1 and 3, while Scenarios 2 and 4 included base adjustment to align evaluation scales between full and reduced datasets as described by Sullivan (2023). Scenario 2 corresponds to the current official Interbull weighting and base-adjustment configuration, although dGEBV rather than EBV was used as the validation target in the present study.

Foreign bulls were included as candidate animals for the smaller breeds to increase the size of the validation population. Candidate and tested bull definitions followed the criteria implemented within the GEBVtest software. Validation outcomes were evaluated using the statistics provided by the software, including regression slope (b1), prediction accuracy (R^2), and overall PASS or FAIL status based on the official Interbull validation criteria. To focus on methodological effects rather than individual evaluation outcomes, breed-trait combinations are presented in aggregate form throughout the Results and Discussion section.

Table 1: Validation scenarios evaluated

S	*Target*	*Reg.*	*BA*	*Wt.*
S1	dGEBV	ITB	No	Yes
S2	dGEBV	ITB	Yes	Yes
S3	dGEBV	LR	No	No
S4	dGEBV	LR	Yes	No

S = validation scenario; **Reg.** = regression method; **BA** = base adjustment; **Wt.** = weighting applied; **ITB** = Interbull weighted regression; **LR** = unweighted regression approach; **dGEBV** = deregressed genomic estimated breeding value.

Results & Discussion

A total of 42 breed-trait combinations were evaluated under four validation scenarios, resulting in 168 scenario-specific validation analyses. Across all analyses, 91 (54.2%) resulted in a PASS outcome, whereas 18 (10.7%) resulted in FAIL, 12 (7.1%) resulted in WARN, and 47 (28.0%) resulted in HiSE (Table 2). WARN indicates validation statistics close to predefined acceptance thresholds, whereas HiSE indicates inconclusive validation results due to high slope standard errors. Most validation outcomes were therefore consistent with acceptable genomic validation performance, although a substantial proportion of analyses were classified as HiSE due to limited validation information available for some breed-trait combinations.

Overall, validation outcomes were generally consistent across scenarios, supporting

conclusions from previous studies that genomic validation performance is largely stable across traits and populations when sufficient information is available (Mota et al., 2023; Mota et al., 2024). Nevertheless, differences among scenarios were observed for a limited number of breed-trait combinations, indicating that choices can influence validation statistics and, in some cases, alter validation conclusions.

Table 2 : Validation outcomes across scenarios

Outcome	*Number of analyses*	*Percentage (%)*
PASS	91	54.2
FAIL	18	10.7
WARN	12	7.1
HiSE	47	28.0
TOTAL	168	100.0

Warn: validation statistics close to predefined acceptance thresholds; HiSE: high standard error that indicates an inconclusive statistical PASS due to high SE(b1) while FAILING the practical test: b1<0.8 or b1>1.2.

To illustrate the effect of alternative validation settings, Table 3 presents an example of a breed-trait combination for which validation outcomes differed among scenarios. Breed and trait identities are intentionally omitted to focus on methodological effects rather than individual evaluation outcomes.

Table 3: Example of validation outcomes under four validation scenarios

S	*Reg.*	*BA*	*Wt.*	*b1*	*Pass*
S1	ITB	No	Yes	1.17	PASS
S2	ITB	Yes	Yes	1.22	FAIL
S3	LR	No	No	1.15	PASS
S4	LR	Yes	No	1.20	PASS

S = validation scenario; **Reg.** = regression method; **BA** = base adjustment; **Wt.** = weighting applied; **ITB** = Interbull weighted regression; **LR** = unweighted regression approach; **dGEBV** = deregressed genomic estimated breeding value; b1 = regression slope.

The example demonstrates that alternative validation settings can produce different validation conclusions despite using the same underlying data and validation target. The only failing scenario corresponded to the ITB regression with base adjustment, whereas the same breed-trait combination passed under the remaining three scenarios. Although the differences in regression slopes were relatively small, they were sufficient to alter the final validation outcome because the statistics were close to the predefined acceptance thresholds. This observation highlights the sensitivity of genomic validation results to methodological choices when validation statistics are near decision boundaries. However, even scenarios classified as PASS but with statistics close to FAIL thresholds warrant further investigation.

The observed differences were not attributable solely to weighting schemes or base adjustment. Instead, the results suggest interactions among regression methodology, scale alignment, and validation-group definition. Sullivan (2023) reported that modifications to validation targets, candidate-bull selection, and base-adjustment procedures can affect the scale of reduced genomic evaluations and consequently influence regression statistics. The present results support those observations and demonstrate that alternative validation configurations may yield different outcomes even when the same genomic evaluations are analyzed.

A key observation from this study is that the effect of base adjustment depended on the regression framework being applied. Base adjustment altered the regression slope under both ITB and LR approaches; however, the resulting validation outcome was not identical across methodologies. This finding suggests that base adjustment should not be interpreted as an isolated factor but rather as one component of a broader validation framework. The combined effects of regression methodology, weighting strategy, and candidate-bull definition ultimately determine the validation statistics produced by the software.

These findings have practical implications for the interpretation of genomic validation results. PASS and FAIL outcomes are often viewed as definitive indicators of evaluation quality; however, the present study demonstrates that validation conclusions can be influenced by slight differences in configuration. Consequently, interpretation of GEBVtest results should consider the validation scenario employed, particularly when regression statistics are close to the acceptance thresholds. The results also emphasize the importance of transparency and consistency when comparing genomic validation outcomes across countries and evaluation systems.

The use of predicted dGEBV as the validation target aligns with recent recommendations for modern genomic evaluation systems and is particularly relevant given the ongoing transition of CDCB toward single-step genomic evaluations. As additional countries adopt single-step methodologies, validation procedures based on genomic predictions rather than traditional daughter-based measures may provide a more consistent framework for evaluating genomic performance. Continued evaluation of validation methodologies will therefore remain important to ensure that genomic validation procedures evolve alongside advances in genetic evaluation systems.

Conclusions

Validation outcomes were generally consistent across scenarios, although a limited number of analyses produced different conclusions depending on weighting scheme and base adjustment. The results indicate that genomic validation outcomes are influenced by interactions among methodological choices rather than individual components alone. Therefore, interpretation of exploratory GEBVtest results should consider the validation scenario employed. The use of dGEBV as the validation target provides a framework compatible with future single-step genomic evaluations. CDCB is currently transitioning toward single-step genomic evaluations with direct GEBV output, and future validation studies will assess whether this framework impacts scale consistency and genomic validation performance.

Acknowledgements and disclaimers

The authors acknowledge the support of Pete Sullivan of Lactanet (https://lactanet.ca/en/auteur/dr-pete-sullivan/) for providing an advance copy of the GEBV validation software, of participating dairy producers for data supply, Dairy Herd Improvement (**DHI**) organizations and Dairy Records Processing Centers (**DRPCs**) for processing and relaying the information to the CDCB. Purebred breed associations for providing pedigree data.

The mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by CDCB. The CDCB is an equal opportunity provider and employer.

References

Interbull Centre. 2021. Interbull CoP – Appendix VIII – Interbull validation test for genomic evaluations – GEBV test. https://wiki.interbull.org/public/CoPAppendixVIII?action=print&rev=48

Interbull. 2026. GEBVtest software. Interbull Centre, Uppsala, Sweden. Available at: https://interbull.org/ib/gebvtest_software

Jorjani, H., and Dürr, J. 2011. Truncated MACE for validation purposes. Interbull Technical Workshop, March 2011, Guelph, Canada.

Legarra, A., and Reverter, A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* 50:53.

Legarra, A., and Reverter, A. 2019. Predicting the accuracy of genomic breeding values. *Genet. Sel. Evol.* 51:55.

Mäntysaari, E., Liu, Z., and VanRaden, P. 2011. Interbull Validation Test for Genomic Evaluations. *Interbull Bulletin* 41:17-21

Mota, R.R., Nicolazzi, E.N., McWhorter, T.M. and VanRaden, P.M. 2023. Genomic validation software: USA results. *Interbull Bulletin*. 58:27–32

Mota, R.R., Sullivan, P., Nicolazzi, E., McWhorter, T.M., Legarra, A., and VanRaden, P. 2024. Genomic validation software: USA update including truncated MACE. *Interbull Bulletin* 60:20–21.

Sullivan, P.G. 2023. Updated Interbull software for genomic validation tests. *Interbull Bulletin* 58:7-16.

VanRaden, P.M. 2021. Improved genomic validation including extra regressions. *Interbull Bulletin*. 56:65–69

Using Interbull validation test results to improve national genomic evaluations

P.G. Sullivan
Lactanet, 660 Speedvale Ave, Guelph, Ontario, Canada, N1K 1E5
Corresponding author: psullivan@Lactanet.ca

Abstract

The Interbull GEBV test software was recently updated, with new methods to account for genomic pre-selection of young AI bulls, and to add diagnostic features that help national evaluation units isolate and resolve issues if sub-optimal validation test results are observed. Applications of the updated Interbull tests to recent Canadian data confirmed genomic evaluations published by Lactanet are accurate and unbiased, particularly for traits with highest emphasis in current selection programs. Average GEBV validation slopes for Holstein were 1.06 across the three production traits, 1.06 for twenty conformation traits validated for internal use in Canada, and 1.07 for the five production and conformation traits requiring validation by Interbull. Validation slopes were notably lower for other traits, and especially for longevity, fertility, and calving traits when the test was based on a mixed population of foreign and local bulls and while ignoring expected bias in the early progeny-based evaluations of validation test bulls with a limited amount of progeny data. New features in the Interbull validation software were used to explore modeling improvements for traits with sub-optimal validation test results, and with direct longevity as an example trait to demonstrate proper uses of new diagnostic features now available in the updated Interbull validation software. Relative to initial validation test slopes for the mixed population of local plus foreign bulls, which averaged 0.73, 0.71, and 0.77 for the direct longevity, fertility, and calving traits evaluated by Interbull in GMACE, validation slopes for these same three groups of MACE traits were 0.98, 0.99 and 0.89 respectively, for local validation bulls born in Canada, when using an unbiased GEBV validation target. Low validation test slopes for GEBV were almost exclusively attributable to biases in the parent average contributions in GEBV, rather than genomic modeling of the young bull genotype effects, secondarily to the inclusion of foreign bulls born and pre-selected outside of Canada and then progeny-tested within Canada, and thirdly due to genomic pre-selection bias that is expected in a validation bull's first GEBV that are based on only one year of progeny data. This research is ongoing, with plans to further explore validation of foreign bulls in Single Step evaluation systems that will include an integration of MACE information.

Key words: genomic validation, software, GEBV, GMACE

Introduction

The Interbull Centre recently released new standards and software, to validate the accuracy and unbiasedness of national GEBV used for international selection and marketing of young genomic bulls, and for input to the Interbull GMACE evaluation service. The newly added software features were described in detail by Sullivan (2023). The purpose of the present paper was to demonstrate proper use of new diagnostic features, with applications of the new Interbull software to Canadian data.

Materials and Methods

Official GEBV tests were applied to the traits in GMACE that included national GEBV from Lactanet Canada in August 2025. The full data sets were national GEBV published in August

2025, and the truncated data sets were national GEBV generated with current evaluation systems, current and full data sets for pedigree and genotypes, and truncated phenotypic data sets, reduced in size by excluding the four most recent years of phenotypes recorded after August 2021.

For foreign bulls with progeny four years ago in countries outside of Canada only, truncated-data MACE proofs from the Interbull TMACE evaluation service were used. The TMACE results from October 2025 were based on truncated-data national EBV from all countries participating in TMACE, including the truncated-data EBV for local bulls in Canada that were used in the present study. The TMACE results for foreign bulls in Canada are thus on the same scale as truncated-data EBV of local bulls that were submitted to TMACE by Lactanet. The integration of TMACE evaluations into national genomic evaluation of truncated data is preferred for validation testing, over the integration of MACE evaluations published four years ago, based on currently outdated methods for any countries making changes to evaluation systems for any of the traits evaluated in MACE over the previous four years. It would be extremely rare that no country participating in MACE would make changes for any trait, during a period of four years. The theoretical advantages of using TMACE for GEBV validation testing are clear, but multiple practical benefits have also been demonstrated by Mota et al. (2023; 2024; 2026).

The following strategy is recommended to identify and make improvements to national GEBV systems, while making best use of new features in the updated Interbull validation software:

1. Apply the official GEBV validation test as described in the Interbull code of practice (https://interbull.org/ib/cop_appendix8), by using only the default software options.
2. Identify traits with the greatest potential for improvement (i.e. with largest detected biases that could be reduced).
3. Use exploratory features of the Interbull validation software to diagnose the biases. Modified validation tests are encouraged in this step, to most efficiently find and isolate underlying causes of bias.
4. Consider improvements to the evaluation model and/or data pipelines to address the identified causes of bias, in a suitably revised national evaluation system.
5. Confirm better diagnostic results from step 3, when testing GEBV from the modified national system.
6. Repeat step 1 to confirm official Interbull test results are also improved by using the modified national system.

Traits with potential for improvement

For Canadian Holstein traits, genomic reference populations are large, and the Interbull validation test results for GEBV accuracy are consequently always acceptable by Interbull standards. The accuracy test is to confirm ranking of young bulls is improved by genotyping, a condition that is usually met if the genomic reference population is large.

The validation slope test is to confirm no bias in the scale, or variance, of GEBV used to pre-select genomic young bulls. The variance of GEBV for young bulls should be consistent with that of progeny-proven bulls. When the variance of GEBV for young genomic bulls is too high or too low, and it is therefore misaligned with the variance of full-data GEBV after progeny testing, this is detected as a bias with the Interbull validation slope test. A GEBV scale bias is detected when the slope of regression from the validation test model deviates significantly from its expected value, which is equal to 1 for any trait (Legarra and Reverter, 2019), if the trait definition is not changing over time for the validation bulls. Traits with largest deviation from an expected validation slope of 1 have largest relative degrees of bias in the scale of variance for GEBV of young genomic bulls. Traits with largest magnitudes of bias have the greatest potential for improvement, with changes to the

national genomic evaluation system, and the greatest potential for successful diagnosis of underlying causes of the bias.

New diagnostics

One of the important updates in the latest version of Interbull software was a feature to guarantee genetic bases of expression are the same between national GEBV from full versus truncated data sets. Methods and the general need to align genetic bases of expression between the two different data sets used for validation testing were fully described by Sullivan (2023). After aligning the scales of expression for bulls already progeny-proven in reduced data, unexpected changes in scale that remain for young genomic bulls who were not yet progeny-tested four years ago, but who are progeny proven in the full data, can be more easily analyzed, and with a better and faster diagnosis of the reasons for bias.

One approach to find underlying causes of bias is to examine either individuals or groups of bulls with largest differences between GEBV from full versus reduced data sets, or to explore plots of base-aligned GEBV from full versus reduced data to identify outliers with strong impacts on the validation slope. Another approach is to add extra variables to the validation test regression model, as proposed by VanRaden, (2021). The use and benefit of these different approaches will be demonstrated by example below, for a Canadian trait with relatively large magnitude of detected bias.

Results & Discussion

Validation test results for accuracy of Lactanet GEBV were acceptable to Interbull for all traits. The accuracy test results are not reported here because our main interest was in diagnosing biases for Lactanet GEBV that are used in Interbull GMACE calculations.

The validation slope test results after three iterations applying the new Interbull software, with official bias tests using the default software options, are shown in Table 1. Details of the investigations between each iteration of testing are described more fully below.

Table 1: Average validation test slope values, by MACE trait group.

Validation test		#1	#1	#2	#3
With foreign bulls		Y	Y	N	N
Years of progeny*		1-4	1-4	1-4	2-4
Trait Group	Trait tests	PA**	GPA**		
Production	3	0.97	1.06	1.10	1.06
Conf'n	2	1.10	1.09	1.17	1.16
Udder	1	0.84	1.14	1.16	1.20
Longevity	1	0.73	0.74	0.88	0.98
Calving	4	0.77	0.77	0.83	0.89
Fertility	5	0.71	0.84	0.95	0.99
Work'y	1	0.63	0.85	0.83	0.80

*Genomic preselection bias is largest in a bull's first year of progeny proving. Excluding the youngest group of validation bulls limits the effects of GPS-bias on validation slope estimates.

**Regression estimates for GEBV from full data on PA and GPA, denoting the EBV and GEBV from truncated data, respectively.

Traits with highest potential for improvement

For production and conformation traits, all validation test slopes (b_1) for Canadian data were within the acceptable levels of tolerance set by Interbull (i.e. $0.90 < b_1 < 1.20$), for all three iterations of validation testing. In the first set of tests, average validation slopes were 1.06 for the three production traits, and 1.09 for the two conformation traits, stature and median udder support, designated by Interbull as reference traits to validate national evaluation systems for all legacy conformation traits as a group. Validation test slopes for all other Canadian conformation traits in MACE were consistent with slope tests of the reference traits. The average validation slope was 1.06 across twenty conformation traits. For the remaining trait groups, validation test slopes for GPA were low due to slopes that were equally low or lower from regressions of full-data GEBV on the young bull's truncated-data PA. The genotype of a young genomic bull has no expected impact on the evaluations of the bull's parents, and much of the bias detected in GPA (the GEBV test) is due to an over-estimated variance of the differences between parent

averages, rather than due to the bull's own genotype.

The direct longevity trait submitted to MACE by Lactanet was among those with largest detected biases. Investigations of underlying causes of bias for direct longevity were examples of efficient ways to use new diagnostics now provided by the updated Interbull validation software.

New diagnostics

Firstly, from a view of the simple plot of base-aligned GEBV from full versus reduced data sets (Figure 1), we observed an especially high number of outlier points (circled) were for a group of foreign bulls who had very high GEBV before progeny testing (G_{red}=105) followed by average GEBV after progeny testing (G_F=100). A majority of these were foreign young bulls born in the USA, who in truncated data had parental information (PA) heavily influenced by ancestors with MACE proofs, rather than Canadian EBV based on local daughters in Canada. In USA, the direct longevity trait includes the effects of high production on survival (USA productive life), while in Canada the direct longevity trait is independent of production (Canadian functional herd life). By effectively changing the trait definition for validation bulls born in USA, as they transition from information on USA ancestry to Canadian progeny, it should be expected that production effects will be reduced in the Canadian GEBV, and thus the validation slope will have an expected value lower than 1 for high-production USA validation bulls who are progeny-proven in Canada.

Excluding foreign bulls from the validation test group had a dramatic effect on the estimated validation slopes, not only for direct longevity, but also for fertility and calving traits. Validation slopes increased notably with the second iteration of validation testing, based on only local bulls, from 0.74 to 0.88 for direct longevity, from 0.84 to 0.95 for the average of five fertility traits, and from 0.77 to 0.83 for the average for four calving traits (Table 1). As with

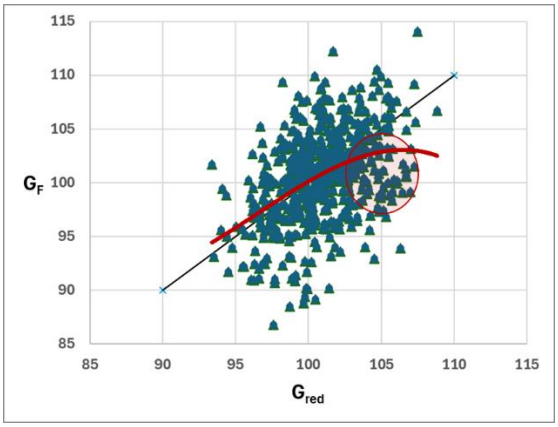


Figure 1. Non-linear polynomial regression of GEBV for direct longevity from full data (G_F) on base-aligned GEBV from data sets reduced by truncation of phenotypes (G_{red}). The highlighted group of points circled, for bulls with high G_{red} and average G_F, caused significant downward pressure on the linear validation test slope.

longevity, the fertility group includes traits that are defined differently in USA versus Canada. The harmonization of calving trait definitions has also been complicated internationally, by different models being used to separate direct and maternal effects of calf and dam respectively, and with categorical data for calving ease and calf survival.

The MACE system accounts for trait definition and modeling differences for bull re-rankings among countries at the population level, but not necessarily at the individually selected bull level. Validation test bulls and their sires are routinely being pre-selected, based on genomics, and based on intense levels of genomic pre-selection (GPS). If the GPS rule has a different correlation with the national definitions of a MACE validation trait defined differently among countries, then the level of correlated GPS effect will change as a validation bull's GEBV evolves, with respect to the relative contributions of information from each country. A pre-selected foreign young bull will transition from having GEBV information that is heavily based on the correlated GPS effects of ancestors in the selecting country, for the truncated-data PA and GPA, to a GEBV based on progeny in the local country in full data. The expectation of regression slopes for

full on truncated data can therefore be lower than 1 for foreign bulls in these cases. In Canadian Holstein genomic evaluation systems, approximately 2/3 of the genomic reference populations are foreign bulls with a MACE proof in Canada versus 1/3 being local bulls with a domestic EBV from progeny data recorded in Canada. The issue of foreign bulls is therefore especially important for validation tests by Lactanet, and potentially for other countries with similar mixtures of foreign versus local genomic reference bulls as Canada.

From a theoretical point of view, it is easily argued that expected values of validation slopes are lower than 1 for foreign bulls, due to changes in the respective trait definitions that drive GEBV information in full versus truncated data. This is the case for direct longevity, fertility, and calving traits in Canada. Interbull has never required the inclusion of foreign bulls in validation test groups, but it has been an optional way to increase validation group sizes that are otherwise too small based on only local bulls.

For Canadian data, Lactanet has historically included foreign bulls for GEBV validation testing, with a continuing interest to validate all progeny-tested bulls in Canada, and because there are many young genomic bulls born in USA who are used for breeding in both USA and Canada. Validation test results for the foreign bulls, however, can be very misleading for traits with expected slopes lower than 1 due to changes in the trait definitions among countries. The inclusion of foreign Holstein bulls is unnecessary for Lactanet because local bulls alone are sufficient for reliable validation tests with the Canadian data.

Further investigations made use of the VanRaden suggestion to add extra regressions in the Interbull validation test model. Of particular interest were the effects of local versus foreign bulls, and the number of years with progeny data in a bull's full-data GEBV. The number of years with accumulated progeny data was considered important for two reasons. Firstly, GPS biases in traditional EBV are highest in a bull's first year of progeny testing, and these biases quickly decrease to small levels as additional progeny data accumulate. Secondly, if a trait definition includes measurement of the trait across multiple parities, the data recorded after first parity are not available until the bull's second and later years with an official progeny-based GEBV. This is the case for direct longevity at Lactanet, which includes daughter survival rates to first, second, and third parity.

A variety of different validation models with extra regression variables were studied, and a good example for direct longevity, fitted separately for local versus foreign bulls was:

$$Y = b_0 + B_1 * (X_1 - \bar{X}_1),$$
$$\text{where } B_1 = b_1 * (1 + b_2 * Z(X_2))$$

and where Y and X_1 are the base-aligned national GEBV from full and reduced data sets, respectively, b_1 is the linear regression used for the Interbull slope test, b_2 is an extra regressions on sire's relative year of birth (Z ranging from 0 to 3), fitted by category of X_2 denoting low versus high levels of expected GPS bias in Y. Variable Z was used as a proxy for the number of years with progeny data in the bull's full-data GEBV. A combined data set was also analysis, fitting separate estimates by bull type, local versus foreign. Estimates were practically the same from combined versus separate analyses by bull type.

The solutions for B_1, the sum of estimated regression effects fitted as interactions with X_1, are in Table 2. Although validation slopes for direct longevity were initially very far from 1 (first test result in Table 1), the slopes were nearly equal to 1 for local bulls with a validation test based on a low-bias validation target (X_2=1, and the third test in Table 1). For both local and foreign bulls, the validation slopes were closer to 1 with a low-bias validation target, but always much lower than 1 for foreign bulls, and this was most likely due to the trait definition in

Table 2: Validation test slopes (B_1) by type of bull, from a validation test model with extra regressions, for the trait direct longevity.

Validation			Variable*		
Bulls	Born	n	X_2	Z	B_1
Foreign	2016	89	1	0	0.67
Foreign	2017	209	1	1	0.68
Foreign	2018	99	1	2	0.70
Foreign	2019	57	0	3	0.58
Local	2016	29	1	0	0.99
Local	2017	86	1	1	0.98
Local	2018	61	1	2	0.97
Local	2019	34	0	3	0.69

*As defined in the text.

change, from including in USA to excluding Canada, the impacts of high production effects on survival.

Although results are not shown here, another factor causing problems with our initial set of validation test results for direct longevity was the inclusion of multiple copies of a cloned USA donor bull's GEBV in the validation test data, once for the donor bull and again for each of the clones. In Canada, the GEBV generated by Lactanet for identical twins, which can include clones, split embryos or naturally occurring identical twins, are a copy for all twins of the single predicted GEBV from pooled data across each set of identical twins in the same identical family. It is correct to include the pooled-data GEBV as a single observation in the validation test data set, but incorrect to include replicated copies of the pooled-data observation, as if they were differently predicted GEBV for each identical twin. The correction to prevent using multiple copies of the same pooled-data GEBV had a remarkably large effect on the slope test results for direct longevity. This discovery came only after reviewing the individual outlier bulls identified in Figure 1.

Across all trait groups, validation slopes summarized in Table 1 were closer to expected values of 1 when based on local bulls only (the second set of tests), and when excluding GEBVs from full data with high expected levels of GPS bias (the third set of tests). In the third set of tests, validation slopes were close to 1 for all trait groups except Udder (Somatic Cell Score) and Workability (Milking Speed). The reasons for bias in national GEBV can differ among traits, and among the national genomic evaluation systems and data sets used for GEBV predictions in different countries.

The GEBV predicted differences for production, conformation, and Udder (SCS) traits are conservative for young, genomic, local bulls born in Canada, as indicated by validation test slopes that are higher than 1 for these trait groups. These results are consistent with historical efforts at Lactanet to limit GEBV bias and to avoid over-prediction of young bulls across all traits, and for all progeny-tested bulls in a mixed population of both local and foreign (mainly USA) young genomic bulls being used for breeding in Canada (e.g. Sullivan et al, 2015). The present study improved our understanding and expectation of different validation test results for local versus foreign bulls. While focusing on local bulls for official validation tests required by Interbull, we also continue efforts to optimize national GEBV for foreign bulls.

Single Step evaluation systems are used by Lactanet to evaluate newer traits not yet evaluated in MACE, while a multi-step genomic evaluation system is used for MACE traits. Methods to integrate foreign progeny information from MACE can differ between multi-step and Single Step genomic evaluation systems, and it has been unclear which of these two systems works best with MACE integration. Better validation methods and improved understanding of validation test results for both foreign and local bulls will help determine if MACE integration into Single Step systems can be equally or more effective than in multi-step genomic evaluation systems. Concerns about GPS bias in a bull's first year progeny proof will be lower with Single Step than with multi-step genomic evaluation systems. Development of Single Step systems that include integration of MACE information is currently underway at Lactanet.

Conclusions

Detecting underlying causes of bias in scale was previously complicated by differing scales of expression between national GEBV from full versus reduced data sets, especially for GEBV expressed on a variance-standardized scale, like the relative breeding value scale of GEBV published in Canada. The complexity of different scales of expression was eliminated by a new feature in the updated Interbull validation software, which was designed to detect and account for scale mis-alignments that can exist in validation test data sets, and to thereby ensure full-data and truncated-data GEBV are adjusted when necessary, and always directly comparable after scale alignment, for all traits and countries. The advantages of base-aligned scales of expression, combined with new software tools and methods to diagnose bias, are more efficient paths to both validating and improving national genomic evaluation models.

The underlying factors shown to cause bias in GEBV tests with Lactanet data, in the present study, were previously difficult to isolate and identify. The new Interbull validation software was invaluable for these new discoveries, and with the software being freely available to all, others can take similar advantage of the new diagnostic tools.

Acknowledgments

Contributions from past and present members of the Interbull validations working group are gratefully acknowledged, in the development and review of new validation methods and software, including Esa Mäntysaari, Paul VanRaden, Zengting Liu, Raphael Mrode, Andres Legarra, Ignacio Aguilar, and Valentina Palucci.

References

Legarra, A. and Reverter, A. 2019. Correction to: Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Gen. Sel. Evol. 51:69. https://doi.org/10.1186/s12711-019-0511-5.

Mota, R.R., Nicolazzi, E, McWhorter, T. and VanRaden, P.M. 2023. Genomic validation software: USA results. Interbull bulletin 58, 27-32.

Mota, R.R., Sullivan, P.G., Nicolazzi, E, McWhorter, T.M., Legarra, A. and VanRaden, P.M. 2024. Genomic validation software: USA update including truncated MACE. Interbull bulletin 60, 27-32.

Mota, R.R., Savoia, S., Legarra, A., McWhorter, T.M. and Nocolazzi, E.L. 2026. A comprehensive GEBV validation framework using Truncated MACE and new traits in U.S. dairy cattle. Proceedings 13th WCGALP 23414.

Sullivan, P.G, Jamrozik, J. and Kistemaker, G.J. 2015. De-regressing MACE versus domestic EBV for genomics. Interbull bulletin 49: 97-102.

Sullivan, 2023. Updated Interbull software for genomic validation tests. Interbull bulletin 58: 7-16.

VanRaden, P.M. 2021. Improved genomic validation including extra regressions. Interbull bulletin 56: 65-69.