J. Dairy Sci. TBC
https://doi.org/10.3168/jds.2025-27502

Comparison of approximation methods for genomic estimated breeding values from observed to liability scales in dairy cattle health traits

Denyus Augusto de Oliveira Padilha,[1]* Daniela Lourenco,[1] Daniel Gianola,[2] Fernando Bussiman,[1] Ignacy Misztal,[1] Hélène Wilmot,[1] Dianelys González-Peña,[3] Giovana Vargas,[3] Gerson A. Oliveira, Jr.[3] Miguel A. Sánchez-Castro,[3] Natascha Vukasinovic,[3] and Jorge Hidalgo[1]*

[1]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602
[2]Deparment of Animal and Dairy Sciences, University of Wisconsin−Madison, Madison, WI 53706
[3]Zoetis Genetics and Precision Animal Health, Kalamazoo, MI 49007

ABSTRACT

The GEBV for health traits are typically published as probabilities obtained using threshold models. While these models benefit from theoretical properties, they require substantial computational resources and may face convergence issues. Linear models can be a good alternative, but solutions need to be approximated to the liability scale before converting the GEBV into probabilities. Recently, an approximation from observed to liability scales was presented with limited success for traits with low prevalence (<5%). Our objective was to compare a new approximation with the previous one using health traits with very low (<1%) to moderate prevalences (up to 25%). We used data from Jersey cows for lameness, mastitis, retained placenta, ketosis, metritis, and displaced abomasum (up to 800k phenotypes per trait). Genotypes for 45k SNP were available for 200k animals. The GEBV were predicted using single-trait threshold and linear models implemented in the BLUPF90 software suite. Both approximations involved scaling the GEBV in the observed scale. The scaling factor was 1) the square root of the product of the residual variance and the proportion of unexplained variance in the linear model or 2) the height of the ordinate of a standard normal distribution evaluated at the threshold (new approximation). We used rank correlations, regression parameters, the overlapping of the distributions, mean squared error (MSE), and classification accuracy to compare GEBV from linear and threshold models on the probability scale for both approximations. Correlations between GEBV from threshold and linear models across approximations ranged from 0.87 (very low prevalence) to 0.99 (moderate prevalence). Although no differences were observed in the correlations across approximations, regression parameters, the overlapping of the distributions, MSE, and classification accuracy were improved with the new approximation method. Therefore, the new approximation provides greater consistency for large-scale evaluations using linear models for categorical traits with prevalences ranging from very low to moderate.

Key words: binary traits, liability scale, linear model, threshold model

INTRODUCTION

Increasing productivity has been a central goal in dairy cattle populations. However, selection for higher milk yield has been associated with negative effects on animal health and longevity (Chesnais et al., 2016; Misztal and Lourenco, 2024). In addition to economic losses, diseases have also raised public concerns about antibiotic use and animal welfare, reinforcing the importance of including health traits in breeding programs (Miglior et al., 2017). Therefore, modern systems also prioritize other breeding goals, such as improving health traits. These traits are commonly recorded as binary events, where animals are classified as healthy (1) or sick (2) based on visual or clinical assessments. Although health traits generally exhibit low heritability (Koeck et al., 2012; Vukasinovic et al., 2016), genetic improvement is cumulative and permanent. Furthermore, the incorporation of genomic data allows accelerated genetic gains through greater selection intensity, improved accuracy, and reduced generation intervals (García-Ruiz et al., 2016; Guinan et al., 2023).

Given the binary nature of health traits, the prediction of GEBV has been traditionally conducted using threshold models, where GEBV are expressed on the liability scale (Falconer, 1965). These models rely on the assumption of an underlying continuous variable (liability), which is assumed to follow a normal distribution and determines the observed categories based on fixed thresholds (Gianola and Foulley, 1983). Despite their

Received August 27, 2025.
Accepted November 13, 2025.
*Corresponding author: denyusaugustp@gmail.com

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-25. Nonstandard abbreviations are available in the Notes.

theoretical appropriateness, threshold models are computationally demanding, particularly with large-scale datasets, and may face convergence issues (Misztal et al., 1989). As an alternative, linear models are often favored due to their simplicity and computational efficiency in routine genetic evaluations (Malchiodi et al., 2017).

Although binary traits do not fully meet the statistical assumptions for linear models, several studies have shown a strong correlation between breeding values from linear and threshold models (Koeck et al., 2010; Hidalgo et al., 2024a). However, linear models yield GEBV in the observed scale, limiting their broader application. The GEBV for health traits are commonly published on a probability scale, which enhances interpretability and facilitates on-farm decision-making. In threshold models, GEBV are expressed on the liability scale, and a direct transformation to the probability scale is straightforward using the normal cumulative distribution function (Gianola and Foulley, 1983).

Although binary traits do not fully meet the statistical assumptions for linear models, several studies have shown a strong correlation between breeding values from linear and threshold models (Koeck et al., 2010; Hidalgo et al., 2024a). However, linear models yield GEBV in the observed scale, limiting their broader application. The GEBV for health traits are commonly published on a probability scale, which enhances interpretability and facilitates on-farm decision-making. In threshold models, GEBV are expressed on the liability scale, and a direct transformation to the probability scale is straightforward using the normal cumulative distribution function (Gianola and Foulley, 1983).

MATERIALS AND METHODS

Data, Prevalence, and Variance Components

Data and variance components (on the liability scale) were provided by Zoetis Genetics and Precision Animal Health (Kalamazoo, MI). Six clinical health traits (recorded as 1 = healthy, 2 = sick) were analyzed: displaced abomasum (**DA**, N = 358,681), retained placenta (**RETP**, N = 539,536), ketosis (**KETO**, N = 456,869), metritis (**METR**, N = 469,677), lameness (**LAME**, N = 692,933), and mastitis (**MAST**, N = 799,905). The pedigree included 678,366 animals, of which 199,480 were genotyped for 45,245 SNPs. Details regarding description of the data, trait definitions, and editing criteria can be found in Gonzalez-Peña et al. (2020). After quality control using predefined parameters for exclusion (duplicates, minor allele frequency <0.05, call rate <0.90, monomorphic, Hardy-Weinberg equilibrium with a maximum difference between observed and expected heterozygous frequency >0.15, located at the sex chromosomes, with unknown position or Mendelian conflicts) by preGSf90 (Misztal et al., 2014a), 30,950 autosomal SNP from 199,394 animals remained in the dataset for final analyses.

The general prevalence (α, %) for each trait was determined by considering the ratio between the number of sick events and the total number of events analyzed. The variance components were transformed from the liability to the observed scale as follows (Dempster and Lerner, 1950):

$$p\left(GEBV_{l_i} < \delta|\widehat{GEBV_{l_i}}\right) = p\left(GEBV_{O_i} < \delta z|\widehat{GEBV_{O_i}}\right) = p\left(GEBV_{O_i}\left\langle\delta^*\left|\widehat{GEBV_{O_i}}\right.\right\rangle\right). \quad [1]$$

where h_o^2 is the estimate for the heritability on the observed scale, z^2 is the square of the height of the ordinate of a standard normal probability density function evaluated at the threshold (calculated from observed prevalence) between categories, and h_l^2 is the estimate of the heritability on the liability scale. Thus, $z^2h_l^2$ corresponds to the additive genetic variance (σ_u^2), and $\alpha(1-\alpha)$ is the phenotypic variance (σ_y^2). The implemented statistical model considered the random effects of permanent environment (**pe**) and contemporary group (**cg**, defined by herd-year-season), as described in Gonzalez-Peña et al. (2020). Thus, these variance components on the observed scale σ_{pe}^2 and σ_{cg}^2, respectively) were obtained based on their proportional contribution to σ_y^2 on the liability scale. Finally, the residual variance on the observed scale was computed as $\sigma_e^2 = \sigma_y^2 - \left(\sigma_u^2 + \sigma_{pe}^2 + \sigma_{cg}^2\right)$.

Data Analyses

The GEBV were predicted using single-step GLUP single-trait linear and threshold models. The computations for linear models were performed with the BLUPF90IOD3 program, and for threshold models, with the CBLUP90IOD2 program, both from the BLUPF90 family (Misztal et al., 2014a). The general statistical model was:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{pe} + \mathbf{Z}_3\mathbf{cg} + \mathbf{e}, \quad [2]$$

where $\mathbf{y}$ is the vector of phenotypic observations (DA, RETP, KETO, METR, LAME, or MAST); $\boldsymbol{\beta}$ is the vector of fixed effects (parity: 1, 2, 3, 4, or ≥5); $\mathbf{u}$ is the vector of random animal additive genetic effects; **pe** is the vector of random permanent environmental effects; **cg** is the vector of random contemporary group effects; $\mathbf{e}$ is the vector of random residuals; and $\mathbf{X}$ and $\mathbf{Z}_i$ are the incidence matrices relating levels of $\boldsymbol{\beta}$, $\mathbf{u}$, **pe**, and **cg** to $\mathbf{y}$, respectively. For all

the analyses, multivariate normal distributions with zero mean were assumed for the random effects (**u**, **pe**, **cg**, and **e**). Furthermore, all random effects were uncorrelated; thus, their joint distribution was a multivariate normal distribution with the following covariance structure:

$$Var\begin{bmatrix}\mathbf{u}\\ \mathbf{pe}\\ \mathbf{cg}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{H}\sigma_u^2 & 0 & 0 & 0\\ 0 & \mathbf{I}\sigma_{pe}^2 & 0 & 0\\ 0 & 0 & \mathbf{I}\sigma_{cg}^2 & 0\\ 0 & 0 & 0 & \mathbf{I}\sigma_e^2\end{bmatrix}, \quad [3]$$

where **H** is the relationship matrix combining pedigree and genomic information (Legarra et al., 2009; Christensen and Lund, 2010); **I** denotes an identity matrix of proper order; and $\tilde{A}_u^2$, $\tilde{A}_{pe}^2$, $\tilde{A}_{cg}^2$, and $\tilde{A}_e^2$ are the additive genetic, permanent environmental, contemporary group, and residual variance components, respectively.

The inverse of **H** was computed as in Aguilar et al. (2010). This computation requires the inverse of the genomic relationship matrix ($\mathbf{G}^{-1}$). Because of the number of genotyped animals, the algorithm for proven and young (**APY**, Misztal et al., 2014b) was used to compute a sparse $\mathbf{G}^{-1}$ ($\mathbf{G}_{\text{APY}}^{-1}$; Misztal et al., 2020). Core animals were set to 13,684 according to the dimensionality of the genomic information (Pocrnic et al., 2016a,b). That is, the number of core animals was calculated as the number of eigenvalues explaining at least 99% of the variability in **G**. The group of core animals was the same in all the analyses, that is, across all traits and across threshold and linear models.

For the linear model, the GEBV were predicted on the observed scale directly using Equation 2, whereas for the threshold model, binary phenotypes y_i (i = 1, 2,…, n) were assumed to follow a Bernoulli distribution: $y_i|p_i \sim Bernoulli(p_i)$, where p_i represents the probability of being healthy ($y_i = 1$), and $1 - p_i$ represents the probability of being sick ($y_i = 2$). The observed phenotypes were assumed to be the response of an underlying variable, the liability (l_i) which, conditional on the model parameters, follows a normal distribution: $\mathbf{l}\left|\boldsymbol{\beta},\mathbf{u},\mathbf{pe},\mathbf{cg},\ \boldsymbol{\sigma}_l^2\right. \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}_1\mathbf{u}+\mathbf{Z}_2\mathbf{pe}+\mathbf{Z}_3\mathbf{cg},\mathbf{I}\boldsymbol{\sigma}_l^2\right)$, $\sigma_l^2 = \sigma_u^2 + \sigma_{pe}^2 + \sigma_{cg}^2 + \sigma_e^2$, and $\sigma_e^2 = 1$ to make the model identifiable (Sorensen and Gianola, 2002). The conditional response of the binary phenotypes was modeled with the following distribution:

$$p\left(\mathbf{y}\mid\mathbf{l},t\right)=\prod_{i=1}^{n}\left[I\left(l_i\le t\right)p\left(y_i=1\mid l_i,t\right)+I\left(l_i>t\right)p\left(y_i=2\mid l_i,t\right)\right], \quad [4]$$

where t is the threshold defining the categories, $I(\cdot)$ is an indicator function, taking the value of 1 if the specified condition is true; otherwise, it takes the value of 0. For example, if $l_i \le t$, then $p(y_i = 1|l_i, t)$ is used; otherwise, it takes the value 0, and $p(y_{ij} = 2|l_i, t)$ is applied.

Approximations of GEBV on the Liability Scale

Two approximations were tested. From here on, they will be referred to as approximation 1 (**AP1**, Hidalgo et al., 2023) and approximation 2 (**AP2**, our proposal). Let the observed scale be represented by the subscript "*o*," whereas the liability scale is represented by the subscript "*l*." Using AP1, the GEBV from linear models was approximated to the liability scale as:

$$GEBV_l \approx \frac{GEBV_o}{\sqrt{\sigma_{e_o}^2\left(1-\frac{h_o^2}{h_l^2}\right)}}. \quad [5]$$

In Equation 5, the $GEBV_o$ were scaled by the square root of the product of the residual variance on the observed scale and the proportion of additive variance unexplained by the linear model (i.e., the ratio of heritabilities; Dempster and Lerner, 1950).

In our proposal, first, Equation 1 is used to obtain $z = \sqrt{\frac{h_o^2}{h_l^2}\alpha\left(1-\alpha\right)}$.. Then $GEBV_o$ is approximated to the liability scale as (AP2):

$$GEBV_l \approx \frac{GEBV_o}{\sqrt{\frac{h_o^2}{h_l^2}\alpha\left(1-\alpha\right)}}. \quad [6]$$

The reasoning behind AP2 is that, near the population mean, $GEBV_o$ can be viewed as a rescaled version of $GEBV_l$, with the scaling factor determined by the height of the ordinate of a standard normal distribution at the threshold (Dempster and Lerner, 1950). In terms of the linear model, the liability takes the form

$$l_i = l_0 + u_i + pe_i + cg_i + e_i,$$

where l_o is the mean liability of the population (after absorbing fixed effects). Recall that $e \sim N(0,1)$. The probability of an animal being healthy (i.e., the probability mass function of the random variable y_i) is

$$p_i(y_i = 1|l_0, u_i, pe_i, cg_i) = p_i(l_i \le t|l_0, u_i, pe_i, cg_i)$$

$$= p_i(l_i - l_0 - u_i - pe_i - cg_i \leq t - l_0 - u_i - pe_i - cg_i | l_0, u_i, pe_i, cg_i)$$

$$= p_i(e_i \leq t - l_0 - u_i - pe_i - cg_i | l_0, u_i, pe_i, cg_i)$$

$$= \int_{-\infty}^{t-l_0-u_i-pe_i-cg_i} p(e_i)\, de_i = \Phi(t - l_o - u_i - pe_i - cg_i)$$
$$= \int_{l_0+u_i+pe_i+cg_i-t}^{\infty} p(e_i)\, de_i = 1 - \Phi(t - l_o - u_i - pe_i - cg_i)$$

$$= -\Phi(t - l_0 - u_i - pe_i - cg_i) = \Phi(l_0 + u_i + pe_i + cg_i - t),$$

where Φ is the cumulative distribution function of a standard normal. As the liability is unobserved, t can be set to 0 as an origin to make the model identifiable (Sorensen, 2023); therefore,

$$p_i(y_i = 1 | l_0, u_i, p_i, cg_i) = \Phi(l_0 + u_i + pe_i + cg_i).$$

Following Sorensen (2023), a first-order Taylor series expansion about $u_i = pe_i = cg_i = 0$ (i.e., about their expectations) can be applied to approximate p_i as a linear function of the random effects in the model:

$$p_i(y_i = 1 | l_0, u_i, pe_i, cg_i) \approx \Phi(l_0) + \frac{\partial \Phi(l_0 + u_i)}{\partial u_i} u_i \Big|_{u_i=0} + \frac{\partial \Phi(l_0 + pe_i)}{\partial pe_i} pe_i \Big|_{pe_i=0} + \frac{\partial \Phi(l_0 + cg_i)}{\partial cg_i} cg_i \Big|_{cg_i=0}$$

$$\approx \Phi(l_0) + \Phi(l_0)u_i + \Phi(l_0)pe_i + \Phi(l_0)cg_i$$
$$\approx \Phi(l_0) + \Phi(l_0)\,(u_i + pe_i + cg_i),$$

where Φ is the probability density function of a standard normal. Therefore, because $E(y_i|p_i) = p_i$, in the observed scale, we can express the observation as the probability of being healthy plus a random residual (Equation 7), that is,

$$y_i = p_i + r_i, \qquad [7]$$

$$y_i \approx \Phi(l_0) + \Phi(l_0)u_i + \Phi(l_0)pe_i + \Phi(l_0)cg_i + r_i$$
$$\approx F_i + U_i + PE_i + CG_i + r_i,$$

where $F_i = \Phi(l_0)$ represents the "fixed" part or mean; $U_i = \Phi(l_0)u_i$ is the $GEBV_o$; $PE_i = \Phi(l_0)pe_i$ and $CG_i = \Phi(l_0)cg_i$ are the permanent environmental and contemporary group estimates, respectively, on the observed scale; and r_i is a residual in the observed scale. Hence, if U_i is the $GEBV_o$, then the $GEBV_l$ would be

$$u_i \approx \frac{U_i}{\phi(l_0)} \therefore GEBV_l \approx \frac{GEBV_o}{\phi(l_0)} \approx \frac{GEBV_o}{z}. \qquad [8]$$

This means that under a first-order approximation near the population mean (i.e., applying the linear model), $GEBV_l$ can be computed as the ratio between $GEBV_o$ and the height of the ordinate of a standard normal density function evaluated at the threshold, which is determined by the mean liability of the population. Please notice that in practice, l_0 is calculated from the mean observed prevalence in the population (α).

GEBV on the Probability Scale

The GEBV was transformed from liability to probability according to Gianola and Foulley (1983):

$$p_i = \Phi(t - u_{ij}), \qquad [9]$$

where p_i is the probability of animal i being healthy; Φ and t were defined above; u_{ij} is the mean liability of animal i in group j (in our study, this was the sum of the solution for parity 4 and the $GEBV_l$ for animal i). We chose parity 4 because it corresponds to the definition of 305-d mature equivalent milk production (305ME).

Additionally, AP1 and AP2 were compared to the GEBV from threshold model on probability scale. This comparison used Spearman rank correlation, regression parameters, the overlap of their distributions, and mean squared error (**MSE**). For the regression model, a simple linear regression of the GEBV on probability scale from threshold models on the GEBV on probability scale from linear models was performed. For MSE, GEBV from the threshold model were the benchmark, whereas GEBV from the linear model after applying either AP1 or AP2 were treated as proxy estimates.

Finally, to assess the precision of both approximations (in terms of ranking and GEBV scale) relative to the benchmark (threshold model), we investigate the approximations' ability to select the same animals as the threshold solutions assuming a cutoff at the top 10% animals in the genotyped population. The cutoff was defined for each trait based on the threshold model solutions. Animals were then categorized into 4 groups: (1) selected by both models; (2) selected by the linear model only; (3) selected by the threshold model only; and (4) unselected by either model. The classification accuracy

(C_{acc}) between the threshold and linear models (AP1 and AP2) was then calculated as follows:

$$C_{acc} = \frac{selected\,in\,both + unselected\,in\,both}{all\,animals} \times 100.$$

For simplicity, traits were grouped according to the prevalence: moderate (9% to 25%), low (1% to 3%), and very low (< 1%). All postprocessing of the solutions in our study was done using in-house R (R Core Team, 2024) scripts and for plots the ggplot2 R package (Wickham, 2016) was used.

RESULTS AND DISCUSSION

The number of records and animals, prevalences (%), and heritability estimates on the observed and liability scales for each trait are provided in Table 1. Overall, prevalence ranged from moderate (24.29% for MAST) to very low (0.60% for DA). The heritability was low for all traits, ranging from 0.003 to 0.05 in the observed scale, whereas it was slightly higher in the liability scale (from 0.06 to 0.12). Figures 1, 2, and 3 show the distribution of GEBV for each trait, according to the model (threshold or linear), prevalence range (moderate to very low), and approximation method (AP1 or AP2). In summary, these figures present the distributions in their respective original scale: observed (linear model) and liability (threshold model); the approximated liability from AP1 and AP2 is also presented; finally, the transformation to the probability scale is plotted for AP1 and AP2. An ideal approximation method would result in sufficient overlap of the GEBV distribution on the adjusted liability and probability scales. For traits with moderate prevalence (Figure 1), as expected, AP1 by Hidalgo et al. (2023) performed well, as did AP2, and no major differences were observed. For traits with low to very low prevalence (Figures 2 and 3), AP1 was suboptimal, meaning that GEBV means and variances were different. In contrast, AP2 effectively overlapped the distributions for these traits (Figures 2 and 3), ensuring a more robust approximation.

Additionally, neither AP1 nor AP2 caused changes in the Spearman correlation between GEBV from different models (linear or threshold). This indicates that both approximations are linear transformations and yield equivalent rankings for selection, resulting in no practical differences in genetic selection outcomes. Rank correlations ranged from 0.87 (DA) to 0.99 (MAST); therefore, the linear model's ability to approximate the threshold models decays with the extreme prevalences. The correlation between linear and threshold models' solutions was moderate for DA, whereas it was high for the other traits indicating that the GEBV ranking from threshold and linear models would differ when the trait present extreme prevalence. In a trait with not extreme prevalence, i.e., ≥1.64%, the correlations were ≥0.97. This implies that an animal with a higher GEBV in one model tends to have a higher GEBV in the other, ensuring consistency in selection decisions. Consequently, selection decisions will not differ significantly between linear and threshold models for traits with no extreme prevalence. Moreover, the MSE was consistently lower for AP2 (Table 2), indicating that this new approximation yields GEBV more consistent with those from the threshold model across all traits. Scatterplots comparing GEBV on probability scale from both approximations and the threshold model is shown in Figures 4, 5, and 6. Regression parameters were added to the plots; ideal coefficients are $b_0 = 0.00$ and $b_1 = 1.00$. The regression parameters improved with increased prevalences for both approximations; however, these were always closer to the ideal parameters for AP2 (Figures 4, 5 and 6). As an illustration, for lameness, the AP1 yielded $b_0 = -34$ and $b_1 = 1.4$, whereas AP2 yielded $b_0 = -25$ and $b_1 = 1.3$. The superiority of AP2 was more apparent with lower prevalences. The AP2 efficacy was less limited by the trait's prevalence, unlike AP1. Additionally, Figures 4, 5, and 6 show the proportion of animals commonly selected by the threshold and linear models (after applying either AP1 or AP2). In general, AP2 showed better agreement with the threshold model

Table 1. Number of records, animals, prevalence (%), and heritability estimates on the observed $\left(h_o^2\right)$ and liability $\left(h_l^2\right)$ scales per trait

Trait	No. of records	No. of Animals	Prevalence (%)	h_o^2	h_l^2
Mastitis	799,890	345,219	24.29	0.05	0.09
Lameness	692,933	301,890	9.74	0.04	0.12
Metritis	469,660	223,244	2.55	0.02	0.12
Ketosis	456,853	230,888	1.69	0.01	0.10
Retained placenta	539,536	240,205	1.64	0.01	0.10
Displaced abomasum[1]	358,681	200,879	0.60	0.00	0.06

[1]For displaced abomasum, the heritability on the observed scale was 0.003.

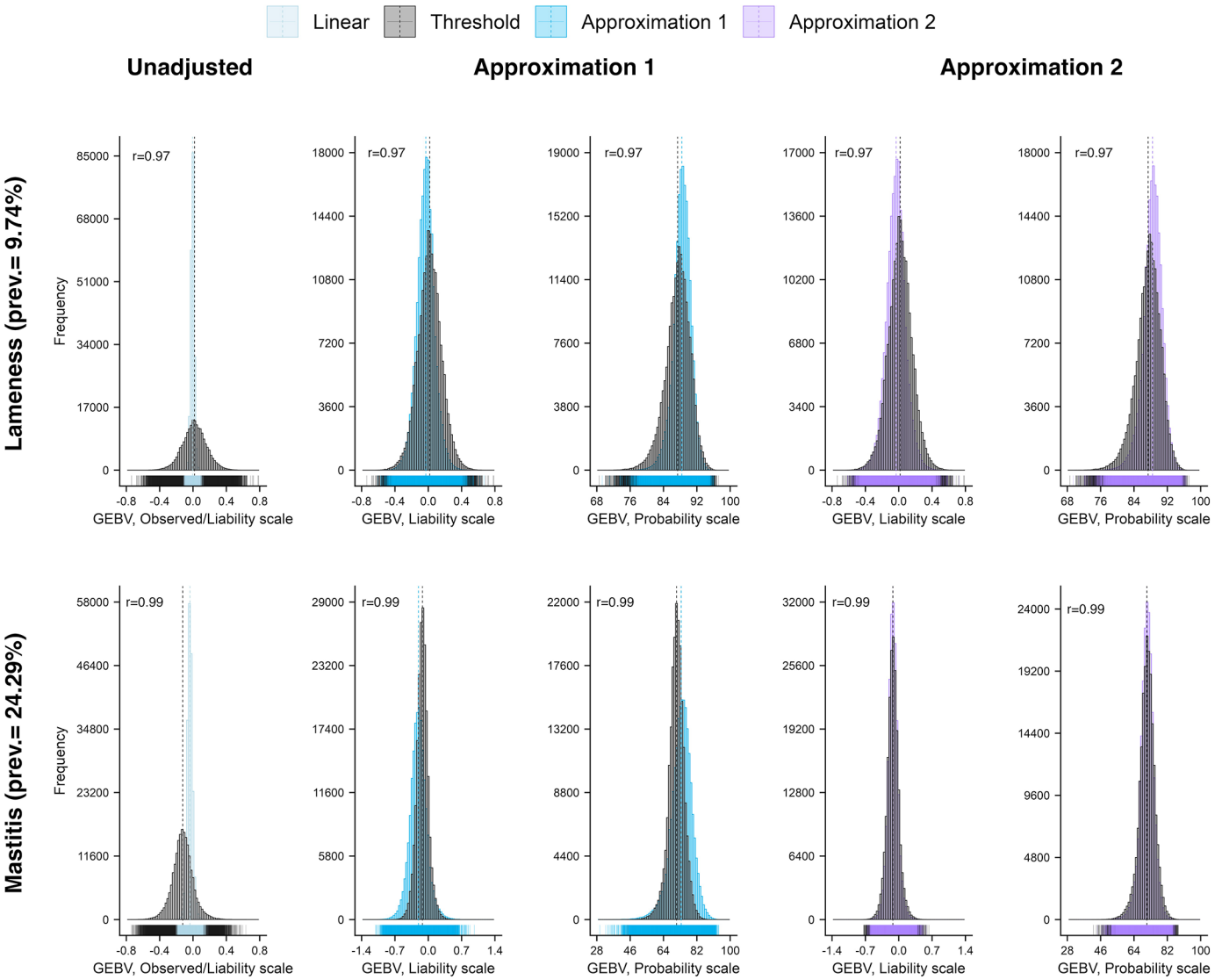

Figure 1. Distributions of GEBV on the observed and liability scales, after approximation to the liability scale, and on the probability scale across different approximation methods for traits with moderate prevalence (9%–25%). prev. = prevalence (%). r= correlation.

(for both selected and unselected animals) than AP1. This is further supported by the classification accuracy (Table 3), which ranged from 76.14% to 96.63% for AP1, and from 92.15% to 97.97% for AP2. These results highlight the ability of AP2 to achieve better agreement with the threshold model, not only in ranking but also in scale. However, as expected, this agreement decreased as trait prevalence declined. It is worth noting that because of how we obtained the classification accuracy, its value is equivalent to the rank correlation considering only individuals selected and unselected in both models (threshold and linear—after the approximations were applied).

Theoretically, threshold models are preferred for the genetic evaluation of binary traits. Their key advantage is that heritability on the liability scale remains independent of the prevalence (Falconer and Mackay, 1996). Also, they allow a transformation of GEBV to probabilities providing a simplified interpretation (Bennewitz et al., 2007). In contrast, when GEBV are predicted on the observed scale, disease susceptibility assessment becomes less precise. This occurs because part of the variation is unexplained, increasing measurement error due to environmental variance. The extent of this unexplained variation depends on prevalence, with the lowest unexplained variation observed for prevalence around 0.50 (Falconer, 1995). For traits with extreme prevalences (close to zero or to one), the prediction becomes more challenging due to reduced explained variation. Since phenotypic variance on the observed scale varies with prevalence, heritability also becomes

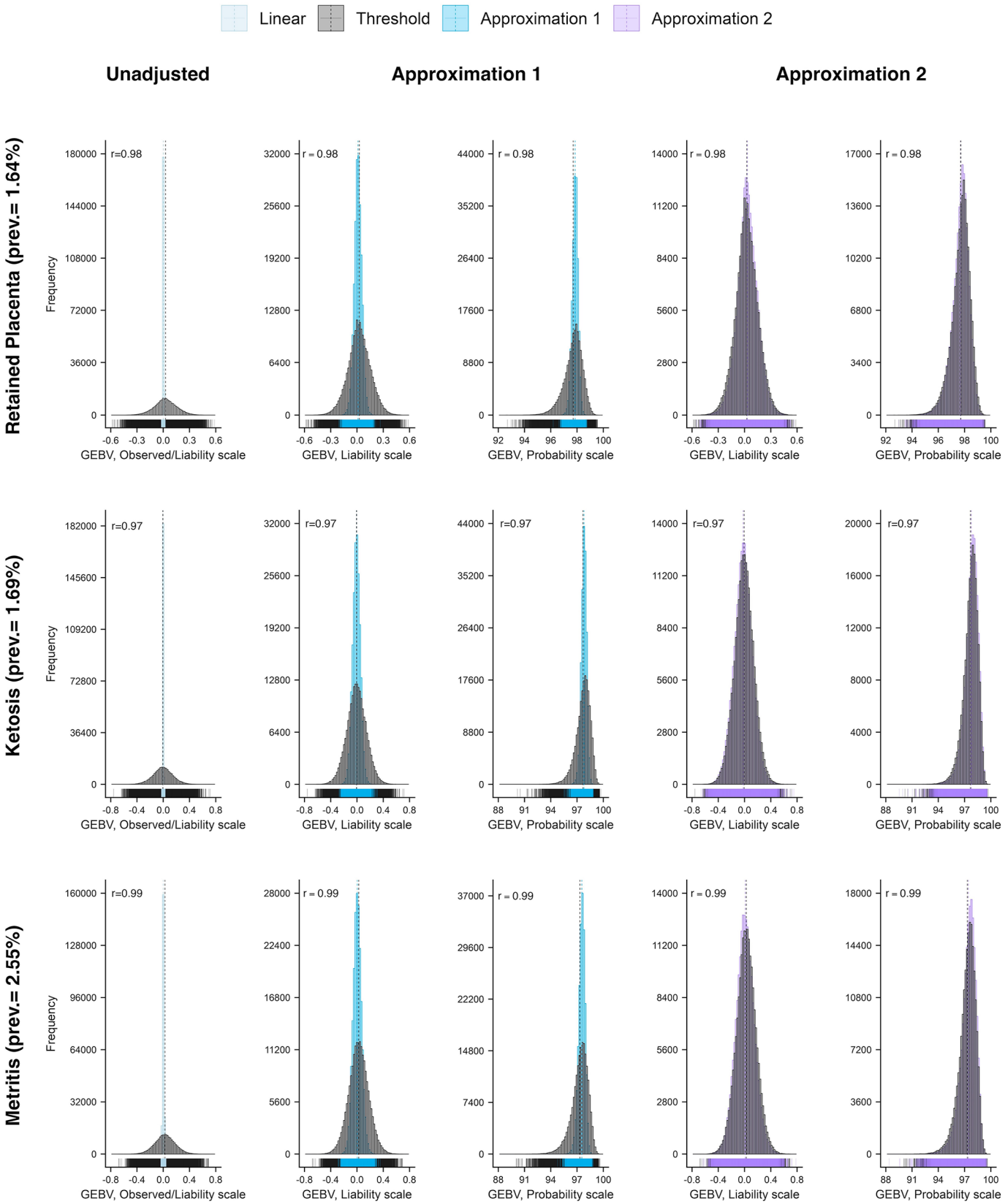

Figure 2. Distributions of GEBV on the observed and liability scales, after approximation to the liability scale, and on the probability scale across different approximation methods for traits with low prevalence (1%–3%). prev. = prevalence (%). r= correlation.

Table 2. Squared error between GEBV on the probability scale from the threshold model and GEBV from the linear models (after AP1 or AP2) across traits

	Squared error			
	Mean (SD)		Min-Max[1]	
Trait	AP1[2]	AP2[3]	AP1	AP2
Mastitis	0.18 (0.15)	0.01 (0.02)	0.00–3.08	0.00–0.93
Lameness	0.03 (0.04)	0.03 (0.04)	0.00–0.87	0.00–0.71
Metritis	0.00 (0.01)	0.00 (0.00)	0.00–0.33	0.00–0.11
Ketosis	0.00 (0.01)	0.00 (0.00)	0.00–0.32	0.00–0.07
Retained placenta	0.00 (0.01)	0.00 (0.00)	0.00–0.25	0.00–0.05
Displaced abomasum	0.00 (0.00)	0.00 (0.00)	0.00–0.32	0.00–0.03

[1]Min-Max = minimum and maximum values.
[2]AP1 = approximation 1.
[3]AP2 = approximation 2.

prevalence-dependent (Ojavee et al., 2022). On the liability scale, however, prevalence-related variation in residual variance is avoided and heritability reflects only additive genetic effects. Heritability estimates on the liability scale and then transformed to the observed scale preserve this property and account for the unexplained variation in the observed scale, that is why the heritability in the observed scale is always lower.

Despite the theoretical benefits of threshold models, linear models are widely used. Nonetheless, in linear models, the residuals should be independent and follow a normal distribution with homogeneous variance (Barker and Shaw, 2016). For categorical traits, although this assumption does not hold, the central limit theorem states that regardless of the original distribution, if the number of observations is sufficiently large, their mean follows a normal distribution (Kwak and Kim, 2017). According to Bulmer (1979), a good approximation to normality holds when $na(1 - \alpha) \approx 4$; that is, the required sample size (n) depends on the trait prevalence. Hidalgo et al. (2024b) argued that one of the limitations of AP1 (for DA in a study with US Holstein cattle) was the low prevalence (2%), which required ~200 "effective" records per animal to correctly predict breeding values. Applying the same reasoning in our study, since DA prevalence is even lower (0.60%), at least twice as many records would be needed.

In general, AP2 performed slightly better, showing closer agreement with threshold model liabilities as evidenced by MSE (Table 2) for low-prevalence traits than for those with moderate prevalence. This indicates that in addition to prevalence, other factors such as heritabil-

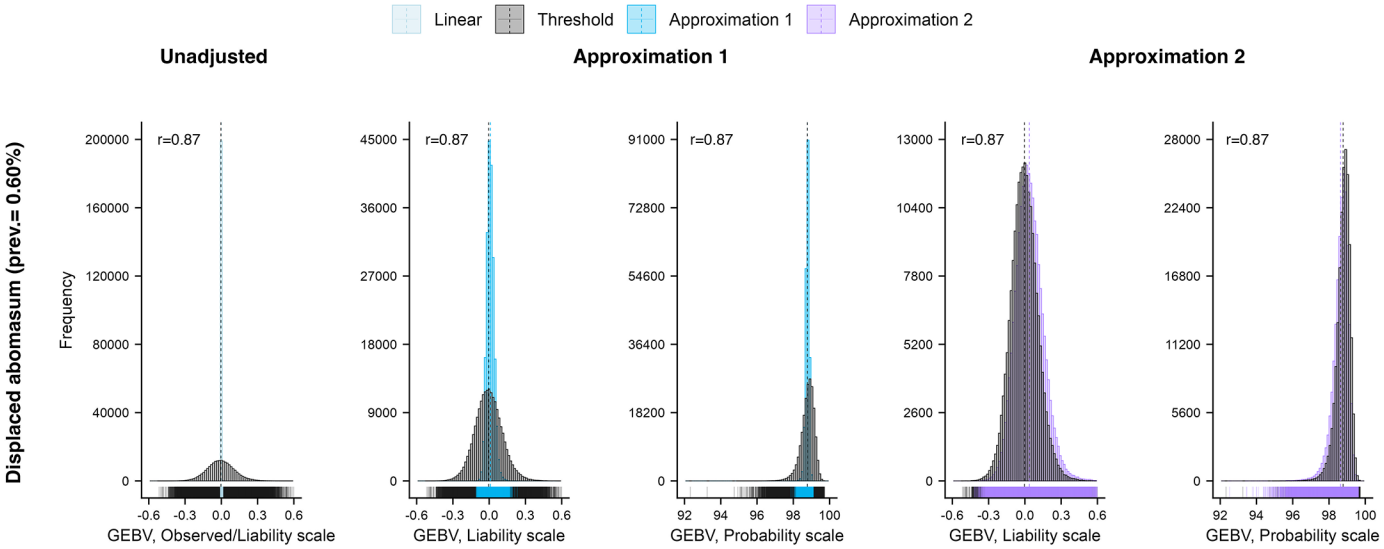

Figure 3. Distributions of GEBV on the observed and liability scales, after approximation to the liability scale, and on the probability scale across different approximation methods for trait with very low prevalence (<1%). prev. = prevalence (%). r= correlation.

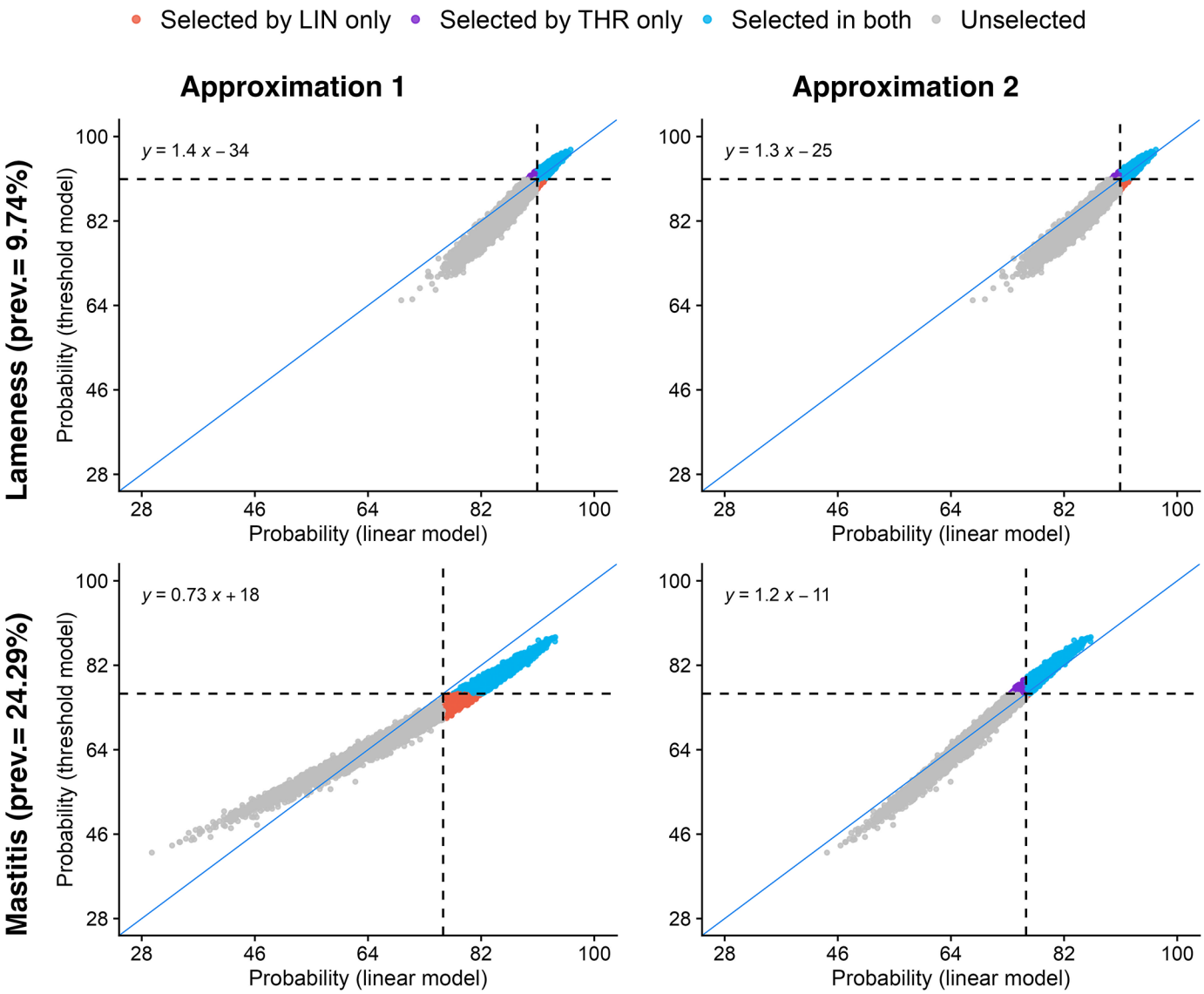

Figure 4. Scatter plots comparing GEBV on the probability scale (threshold vs. linear model) for traits with moderate prevalence (9%–25%) across approximation methods, and commonly selected animals across models and approximations. The dashed line represents the cutoff used to select the top 10% of animals for both models. Animals are colored according to their selection status. prev. = prevalence (%); LIN = linear model; and THR = threshold model.

ity and variance heterogeneity among subpopulations (in this case, contemporary groups) also influence the scaling in both approximations, as they are components of Equations 5 and 6. Overall, prevalence is the key factor connecting these elements, and it is influenced by indirect disease-related factors. In our study, all traits were obtained from producer-recorded data, which can be subjective and less precise. This makes genetic parameter estimation challenging. Additionally, discrepancies between sample and population prevalences can introduce bias. This occurs because datasets represent only a subset of the population, often underestimating the prevalence compared to the actual population prevalence (Ojavee et al., 2022).

The reason for no differences and a high correlation between GEBV for AP1 and AP2 (Figures 1, 2, and 3) is that z is associated with a linear transformation. In other words, $GEBV_l$ is linearly related to $GEBV_o$ by a constant (z). For any arbitrary δ (where $\delta^* = \delta z$) and mean prevalence $\Phi(l_0)$, the probability of $GEBV_l$ being lower than δ, given an estimate of $GEBV_l\left(\widehat{GEBV_l}\right)$, is $p\left(GEBV_{l_i} < \delta|\widehat{GEBV_{l_i}}\right) = p\left(GEBV_{O_i} < \delta z|\widehat{GEBV_{O_i}}\right)$ This $= p\left(GEBV_{O_i}\left\langle\delta^*\middle|\widehat{GEBV_{O_i}}\right.\right)$.

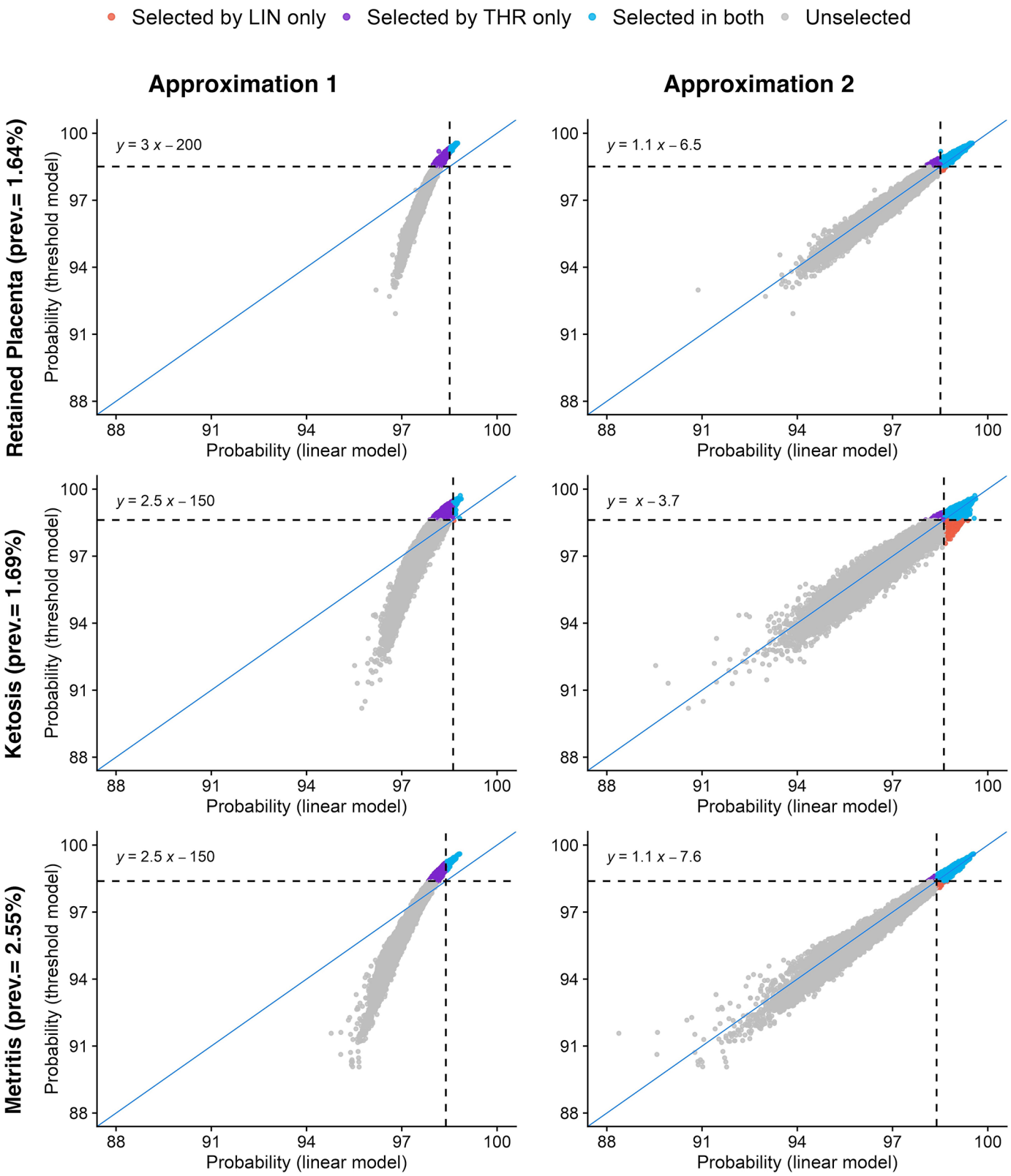

Figure 5. Scatter plots comparing GEBV on the probability scale (threshold vs. linear model) for traits with low prevalence (1%–3%) across approximation methods, and commonly selected animals across models and approximations. The dashed line represents the cutoff used to select the top 10% of animals for both models. Animals are colored according to their selection status. prev. = prevalence (%); LIN = linear model; and THR = threshold model.

Table 3. Classification accuracy (%) between GEBV on the probability scale from the threshold model and those from the linear model (after AP1 and AP2) across traits

	Classification accuracy (%)	
Trait	AP1[1]	AP2[2]
Mastitis	76.14	96.97
Lameness	96.63	95.97
Metritis	90.47	97.97
Ketosis	90.11	96.31
Retained placenta	90.11	96.98
Displaced abomasum	90.00	92.15

[1]AP1 = approximation 1.

[2]AP2 = approximation 2.

means that provided that a linear transformation is applied from observed to liability scales, the probability structure is preserved, and results are the same. However, such approximations rely on 2 assumptions: writing p_i as a linear function of liability and assuming that r_i is uncorrelated with the liability. This illustrates that while linear models may yield high correlations with outcomes from a threshold model, several approximations are involved. It also suggests that linear model's ability to approximate threshold models relies on the convergence of the binomial distribution to the normal distribution for large sample sizes (Hidalgo et al., 2024b).

Like AP1, AP2 assumes a fixed scaling factor across all individuals, implying constant prevalence within subgroups. However, the average prevalence across subpopulations is also a critical factor, given the assumption of a constant liability variance among groups (Falconer, 1965). To alleviate this limitation, we used only contemporary groups with at least one affected individual and modeled this effect as random. Future studies should be conducted to refine the method by incorporating group-specific scaling factors, which will allow the model to account for different prevalences, and heritability estimates across contemporary groups. In addition, we hypothesize that AP2 might perform even better when the GEBV were predicted using multitrait models, particularly due to the advantages of such models in analyzing low heritability but genetically correlated traits, such as RETP and METR (Koeck et al., 2012). This, however, remains yet to be tested. Multitrait approaches may also improve predictive performance for traits with substantial missing data when correlated traits have complete records, as demonstrated by Guo et al. (2014).

CONCLUSIONS

Our approximation (AP2), which assumes as a scaling factor the height of the ordinate of the standard normal distribution evaluated at the threshold, introduces a novel way to approximate GEBV from the observed scale to the

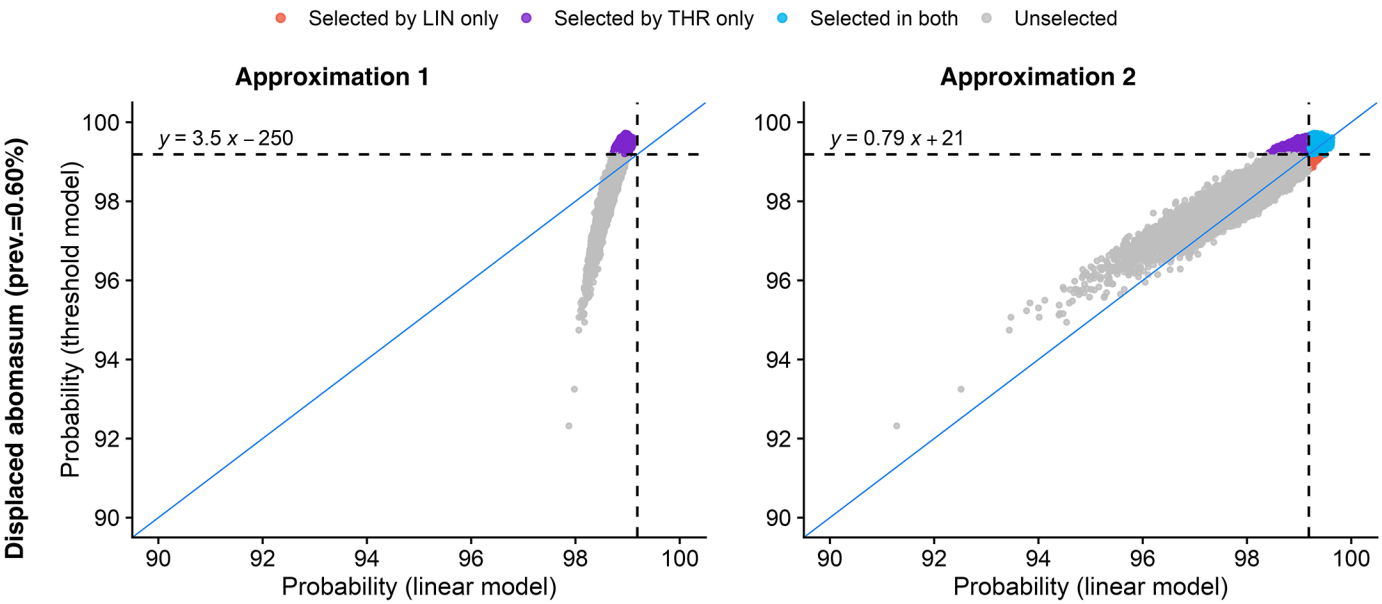

Figure 6. Scatter plots comparing GEBV on the probability scale (threshold vs. linear model) for trait with very low prevalence (<1%) across approximation methods, and commonly selected animals across models and approximations. The dashed line represents the cutoff used to select the top 10% of animals for both models. Animals are colored according to their selection status. prev. = prevalence (%); LIN = linear model; and THR = threshold model.

liability scale, enabling GEBV estimation through linear models as probabilities comparable to those obtained from threshold models. This strategy provides improved robustness and precision for traits with very low to moderate prevalences, making it a viable alternative for large-scale analyses of binary traits using linear models.

NOTES

This study was partly supported by Zoetis Genetics and Precision Animal Health (Kalamazoo, MI); the availability of data and feedback is also appreciated. Animal care and use committee approvals were unnecessary, as data were obtained from pre-existing databases. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: AP1 = approximation 1; AP2 = approximation 2; APY = algorithm for proven and young; C_{acc} = classification accuracy; cg = contemporary group; DA = displaced abomasum; KETO = ketosis; LAME = lameness; MAST = mastitis; METR = metritis; MSE = mean squared error; pe = permanent environment; RETP = retained placenta.

REFERENCES

Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90-POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Pages 1–3 in Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, BC, Canada.

Barker, L. E., and K. M. Shaw. 2015. Best (but oft-forgotten) practices: Checking assumptions concerning regression residuals. Am. J. Clin. Nutr. 102:533–539. https://doi.org/10.3945/ajcn.115.113498.

Bennewitz, J., O. Morgades, R. Preisinger, G. Thaller, and E. Kalm. 2007. Variance component and breeding value estimation for reproductive traits in laying hens using a Bayesian threshold model. Poult. Sci. 86:823–828. https://doi.org/10.1093/ps/86.5.823.

Bulmer, M. G. 1979. Principles of Statistics. Dover Publications, Garden City, NY.

Carlén, E., U. Emanuelson, and E. Strandberg. 2006. Genetic evaluation of mastitis in dairy cattle using linear models, threshold models, and survival analysis: A simulation study. J. Dairy Sci. 89:4049–4057. https://doi.org/10.3168/jds.S0022-0302(06)72448-1.

Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011. Effect of different genomic relationship matrices on accuracy and scale1. J. Anim. Sci. 89:2673–2679. https://doi.org/10.2527/jas.2010-3555.

Chesnais, J. P., T. A. Cooper, G. R. Wiggans, M. Sargolzaei, J. E. Pryce, and F. Miglior. 2016. Using genomics to enhance selection of novel traits in North American dairy cattle. J. Dairy Sci. 99:2413–2427. https://doi.org/10.3168/jds.2015-9970.

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.

Dempster, E. R., and I. M. Lerner. 1950. Heritability of threshold characters. Genetics 35:212–236. https://doi.org/10.1093/genetics/35.2.212.

Falconer, D. S. 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann. Hum. Genet. 29:51–76. https://doi.org/10.1111/j.1469-1809.1965.tb00500.x.

Falconer, D. S., and T. F. C. Mackay. 1996 Introduction to Quantitative Genetics. 4th ed. Addison Wesley Longman, Harlow.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc. Natl. Acad. Sci. USA 113. https://doi.org/10.1073/pnas.1519061113.

Gianola, D., and J. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201. https://doi.org/10.1186/1297-9686-15-2-201.

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

Guinan, F. L., G. R. Wiggans, H. D. Norman, J. W. Dürr, J. B. Cole, C. P. Van Tassell, I. Misztal, and D. Lourenco. 2023. Changes in genetic trends in US dairy cattle since the implementation of genomic selection. J. Dairy Sci. 106:1110–1129. https://doi.org/10.3168/jds.2022-22205.

Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du, and G. Su. 2014. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet. 15:30. https://doi.org/10.1186/1471-2156-15-30.

Hidalgo, J., I. Misztal, S. Tsuruta, M. Bermann, K. Retallick, A. Garcia, F. Bussiman, and D. Lourenco. 2024a. Transforming estimated breeding values from observed to probability scale: how to make categorical data analyses more efficient. J. Anim. Sci. 102:skae307. https://doi.org/10.1093/jas/skae307.

Hidalgo, J., S. Tsuruta, D. Gonzalez, G. de Oliveira, M. Sanchez, A. Kulkarni, C. Przybyla, G. Vargas, N. Vukasinovic, I. Misztal, and D. Lourenco. 2023. Converting linear breeding values to probabilities for health traits in dairy cattle. J. Dairy Sci. 106(Suppl. 1):187–188. (Abstr.)

Hidalgo, J., S. Tsuruta, D. Gonzalez, G. De Oliveira, M. Sanchez, A. Kulkarni, C. Przybyla, G. Vargas, N. Vukasinovic, I. Misztal, and D. Lourenco. 2024b. Converting estimated breeding values from the observed to probability scale for health traits. J. Dairy Sci. 107:9628–9637. https://doi.org/10.3168/jds.2024-24767.

Kadarmideen, H. N., R. Thompson, and G. Simm. 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. Anim. Sci. 71:411–419. https://doi.org/10.1017/S1357729800055338.

Koeck, A., B. Heringstad, C. Egger-Danner, C. Fuerst, and B. Fuerst-Waltl. 2010. Comparison of different models for genetic analysis of clinical mastitis in Austrian Fleckvieh dual-purpose cows. J. Dairy Sci. 93:4351–4358. https://doi.org/10.3168/jds.2009-2955.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012. Health recording in Canadian Holsteins: Data and genetic parameters. J. Dairy Sci. 95:4099–4108. https://doi.org/10.3168/jds.2011-5127.

Krishnamoorthy, P., A. L. Goudar, K. P. Suresh, and P. Roy. 2021. Global and countrywide prevalence of subclinical and clinical mastitis in dairy cattle and buffaloes by systematic review and meta-analysis. Res. Vet. Sci. 136:561–586. https://doi.org/10.1016/j.rvsc.2021.04.021.

Kwak, S. G., and J. H. Kim. 2017. Central limit theorem: The cornerstone of modern statistics. Korean J. Anesthesiol. 70:144. https://doi.org/10.4097/kjae.2017.70.2.144.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061.

Malchiodi, F., A. Koeck, S. Mason, A. M. Christen, D. F. Kelton, F. S. Schenkel, and F. Miglior. 2017. Genetic parameters for hoof health traits estimated with linear and threshold models using alternative cohorts. J. Dairy Sci. 100:2828–2836. https://doi.org/10.3168/jds.2016-11558.

McWhorter, T. M., M. Bermann, A. L. S. Garcia, A. Legarra, I. Aguilar, I. Misztal, and D. Lourenco. 2023. Implication of the order of blending and tuning when computing the genomic relationship matrix in single-step GBLUP. J. Anim. Breed. Genet. 140:60–78. https://doi.org/10.1111/jbg.12734.

Miglior, F., A. Fleming, F. Malchiodi, L. F. Brito, P. Martin, and C. F. Baes. 2017. A 100-Year Review: Identification and genetic selec-

tion of economically important traits in dairy cattle. J. Dairy Sci. 100:10251–10271. https://doi.org/10.3168/jds.2017-12968.
Misztal, I., D. Gianola, and J. L. Foulley. 1989. Computing aspects of a nonlinear method of sire evaluation for categorical data. J. Dairy Sci. 72:1557–1568. https://doi.org/10.3168/jds.S0022-0302(89)79267-5.
Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014a. Manual for BLUPF90 family of programs. Accessed Aug. 26, 2025. https://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all.pdf.
Misztal, I., A. Legarra, and I. Aguilar. 2014b. 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.
Misztal, I., and D. Lourenco. 2024. Potential negative effects of genomic selection. J. Anim. Sci. 102:skae155. https://doi.org/10.1093/jas/skae155.
Misztal, I., D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. Manual for BLUPF90 family of programs University of Georgia, Athens, USA 2022.
Misztal, I., S. Tsuruta, I. Pocrnic, and D. Lourenco. 2020. Core-dependent changes in genomic predictions using the algorithm for proven and young in single-step genomic best linear unbiased prediction. J. Anim. Sci. 98:skaa374. https://doi.org/10.1093/jas/skaa374.
Ojavee, S. E., Z. Kutalik, and M. R. Robinson. 2022. Liability-scale heritability estimation for biobank studies of low-prevalence disease. Am. J. Hum. Genet. 109:2009–2017. https://doi.org/10.1016/j.ajhg.2022.09.011.
Pocrnic, I., D. A. L. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016a. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203:573–581. https://doi.org/10.1534/genetics.116.187013.
Pocrnic, I., D. A. L. Lourenco, Y. Masuda, and I. Misztal. 2016b. Dimensionality of genomic information and performance of the algorithm for proven and young for different livestock species. Genet. Sel. Evol. 48:82. https://doi.org/10.1186/s12711-016-0261-6.
R Core Team. (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.
Sorensen, D. 2023. Statistical Learning in Genetics: An Introduction Using R. Springer Nature Switzerland AG, Cham, Switzerland.
Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health. Springer New York, New York, NY.
VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.
Viazzi, S., C. Bahr, A. Schlageter-Tello, T. Van Hertem, C. E. B. Romanini, A. Pluk, I. Halachmi, C. Lokhorst, and D. Berckmans. 2013. Analysis of individual classification of lameness using automatic measurement of back posture in dairy cattle. J. Dairy Sci. 96:257–266. https://doi.org/10.3168/jds.2012-5806.
Vukasinovic, N., N. Bacciu, C. A. Przybyla, P. Boddhireddy, and S. K. DeNise. 2017. Development of genetic and genomic evaluation for wellness traits in US Holstein cows. J. Dairy Sci. 100:428–438. https://doi.org/10.3168/jds.2016-11520.
H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.