



Novel approach to incorporate information about recessive lethal genes increases the accuracy of genomic prediction for mortality traits

Grum Gebreyesus¹ · Goutam Sahana¹ · A. Christian Sørensen¹ · Mogens S. Lund¹ · Guosheng Su¹

Received: 21 January 2020 / Revised: 2 June 2020 / Accepted: 2 June 2020 / Published online: 12 June 2020
© The Author(s) 2020. This article is published with open access

Abstract

The genetic underpinnings of calf mortality can be partly polygenic and partly due to deleterious effects of recessive lethal alleles. Prediction of the genetic merits of selection candidates should thus take into account both genetic components contributing to calf mortality. However, simultaneously modeling polygenic risk and recessive lethal allele effects in genomic prediction is challenging due to effects that behave differently. In this study, we present a novel approach where mortality risk probabilities from polygenic and lethal allele components are predicted separately to compute the total risk probability of an individual for its future offspring as a basis for selection. We present methods for transforming genomic estimated breeding values of polygenic effect into risk probabilities using normal density and cumulative distribution functions and show computations of risk probability from recessive lethal alleles given sire genotypes and population recessive allele frequencies. Simulated data were used to test the novel approach as implemented in probit, logit, and linear models. In the simulation study, the accuracy of predicted risk probabilities was computed as the correlation between predicted mortality probabilities and observed calf mortality for validation sires. The results indicate that our novel approach can greatly increase the accuracy of selection for mortality traits compared with the accuracy of predictions obtained without distinguishing polygenic and lethal gene effects.

Introduction

Important fractions of deleterious mutations segregating in diploid organisms are recessive alleles that cause fatal effects when present in a homozygous state. In dairy cattle breeds, intensive use of a limited number of elite breeding sires through artificial insemination has led to the spread of recessive lethal alleles in populations (e.g., Shuster et al. 1992; Agerholm et al. 2001). Recessive lethal alleles are of considerable economic consequence in the dairy industry because they cause calf and young stock mortality as well as reproductive inefficiency (e.g., Cole et al. 2016). Calf mortality represents a major economic loss for farmers,

poses great animal welfare issues, and threatens public perceptions of the dairy industry. Therefore, efficient strategies for limiting the impact of recessive lethal alleles on calf mortality are critical.

Several strategies have been proposed that involve the use of genotype information to limit the harmful effects of recessive lethal alleles (e.g., Charlier et al. 2008; Pryce et al. 2012; Cole 2015), including the possibility of their removal through genome editing (Johnsson et al. 2019). Currently, the most practiced approach in dairy breeding is to limit carrier-to-carrier mating (Charlier et al. 2008). While such an approach might be feasible with few lethal alleles segregating in the population, it is becoming considerably more difficult with the increasing number of detected recessive lethal alleles (VanRaden et al. 2011; Fritz et al. 2013; Sahana et al. 2013, 2016; Kadri et al. 2014; Hoff et al. 2017). Preselection based on carrier status reduces selection intensity, consequently affecting genetic gains in economically important traits. In addition, the genetic causes of calf mortality are also partly polygenic. Therefore, genomic prediction of breeding values for calf mortality traits should

Associate editor: Sara Knott

✉ Guosheng Su
guosheng.su@qgg.au.dk

¹ Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Alle, 8830 Tjele, Denmark

take into account both polygenic and recessive lethal components.

Several approaches have been proposed for incorporating genotypic information of major genes into genomic prediction for traits controlled by both “major” genes and polygenic inheritance (e.g., Hoeschele 1988; Fernando and Grossman 1989). However, the proposed models traditionally rely on the assumption of additive effects for major quantitative trait loci (QTLs), as in the remaining genome-wide markers (Cole 2015), and different weights, with the a priori assumption of different effect sizes. Adopting such approaches to model polygenic effects and recessive lethal alleles simultaneously is problematic, as not only the effect sizes but also the modes of gene action are different. The effect of the heterozygous genotype for a recessive lethal locus on an individual is the same as that of the homozygous genotype of the wild type, but the effects on the future mortality risk of offspring differ. Additional sources of complexity in modeling the two effects simultaneously are that different recessive lethal alleles might have different penetrance levels and the fact that the mortality risk transferred to offspring depends on both penetrance and recessive allele frequency. Therefore, fitting polygenic and recessive lethal effects simultaneously in genomic prediction models has been a challenge.

In this study, we hypothesize that modeling risk probabilities from polygenic QTL effects and recessive lethal alleles separately can improve the accuracy of selection for mortality traits. The objective of this study is therefore to present a strategy where a breeding animal’s risk with regard to the mortality of its future offspring can be predicted in terms of probabilities for the recessive lethal allele and polygenic risk components. In this approach, risk probabilities from recessive lethal loci are computed as a function of carrier status and the recessive allele frequency in the population. We develop methodologies for transforming genomic estimated breeding values (GEBVs) into polygenic risk probabilities using normal density and cumulative distribution functions. Simulated data are used to test the advantages of the presented approach by comparing its prediction accuracies with those of an alternative approach where risk probabilities are predicted without distinguishing polygenic and recessive lethal allele effects.

Materials and methods

In this section, we present an approach that allows the prediction of genetic risk posed to the survival of future offspring by both the recessive lethal locus and polygenic components of a breeding animal as the basis for selection to improve mortality traits.

Risk probabilities from recessive lethal alleles

The computation of risk probabilities from recessive lethal alleles is straightforward when using the carrier status of an animal as inferred from its genotype and the recessive allele frequencies in the population. Given the genotype of a breeding animal, e.g., a sire, for a particular recessive lethal locus, an offspring’s risk probability of succumbing to a lethal gene is the probability that it receives two copies of the recessive lethal allele. Assuming a locus i with recessive lethal allele a and wild-type allele A , the probability of the offspring’s mortality due to the lethal allele (hereafter referred to as p_{leth_i}) is the probability of the offspring having an aa genotype. This can be computed given the sire’s genotype and the frequency of the recessive allele in the population (p_a), which accounts for the probability of receiving a copy of the lethal allele from a dam under the assumption of random mating:

$$p_{\text{leth}_i} = p(\text{a from sire} \mid \text{sire_geno}) \times p_a. \quad (1)$$

Assuming complete penetrance and given three possible outcomes of the sire’s genotype, i.e., AA , Aa , or aa , the lethal risk probability at locus i for an offspring can be given as:

$$\text{Sire } AA : p_{\text{leth}_i} = 0,$$

$$\text{Sire } Aa : p_{\text{leth}_i} = 0.5 \times p_a,$$

$\text{Sire } aa : p_{\text{leth}_i} = p_a$ (however, the affected individual may not be a breeding animal under complete penetrance).

In cases of recessive lethal alleles with a penetrance other than 100%, the risk probability in Eq. (1) can be calculated by multiplying by the penetrance level of a given lethal allele:

$$p_{\text{leth}_i} = p(\text{a from sire} \mid \text{sire_geno}) \times p_a \times \text{penetrance}. \quad (2)$$

An individual might carry more than one recessive lethal gene. Thus, the risk probability across all n loci can be computed as:

$$p(\text{leth}) = 1 - \prod_{i=1}^n (1 - p_{\text{leth}_i}). \quad (3)$$

Prediction of polygenic risk probabilities

Obtaining risk probabilities from the polygenic component requires a transformation from GEBVs, which are predicted using different genomic prediction models. Mortality-related traits are often recorded as categorical outcomes,

which are usually binary and hence non-normally distributed. Commonly used approaches for modeling categorical traits include threshold-liability (probit) and logistic regression (logit) models. Despite their violations of normality assumptions, linear models (LM) have also been employed in many studies for genetic analysis of categorical traits (e.g., Rao and Xia 2000; Peñagaricano et al. 2011). To accommodate the specific computations in the prediction of GEBVs in these different models, we present strategies for transforming GEBVs into polygenic risk probabilities as implemented in probit model, logit model, and LM.

Liability threshold (probit) model

In genetic modeling of categorical traits, one of the most commonly used approaches is the threshold model (Wright 1934). In the threshold model, the observed categorical responses are assumed to be the outcome of an underlying, normally distributed latent variable, often termed the liability (l), in relation to a fixed threshold (τ). In the context of mortality traits, the categories of response are usually “survival,” denoted as event 0, or “death,” denoted as event 1, i.e., a binary response, during a given monitored period. Accordingly, the observed categorical outcomes (y) are linked to the underlying liability (l) such that:

$$y = \begin{cases} 1, & \text{if } l > \tau \\ 0, & \text{if } l < \tau \end{cases} \quad (4)$$

Thus, the expected liability (η) is assumed to be a function of the predictors such that:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a}, \quad (5)$$

where $\boldsymbol{\eta}$ is the vector of all expectations η_i , $\boldsymbol{\beta}$ is the vector of fixed effects, \mathbf{a} is the vector of random additive genetic effects, and \mathbf{X} and \mathbf{Z} are the design matrices for the fixed and random effects, respectively. Thus, the true underlying liability (l) will be the expectation plus the residuals such that:

$$\mathbf{l} = \boldsymbol{\eta} + \mathbf{e}, \quad (6)$$

where \mathbf{l} is the vector of all l_i and \mathbf{e} is the vector of random residuals. The underlying liability for each individual as a linear function of the linear predictors can thus be rewritten as:

$$l_i = \eta_i + e_i.$$

Combining Eqs. (4) and (6):

$$y = \begin{cases} 1, & \text{if } \eta_i + e_i > \tau \\ 0, & \text{if } \eta_i + e_i < \tau \end{cases} \quad (7)$$

Given η_i estimated from the data and an assumed value for the fixed threshold τ , the observed outcomes are conditional on the residuals as follows:

$$y = \begin{cases} 1, & \text{if } e_i > \tau - \eta_i \\ 0, & \text{if } e_i < \tau - \eta_i \end{cases} \quad (8)$$

The probability of observing event 1 (mortality) given η_i and τ can then be estimated as:

$$p(y_i = 1 | \eta_i, \tau) = p(e_i > \tau - \eta_i) = 1 - \Phi_e(\tau - \eta_i), \quad (9)$$

where $\Phi_e(\cdot)$ is the cumulative density function with $e_i \sim N(0, \sigma_e^2)$, where σ_e^2 is residual variance. During implementation of the probit model, the threshold τ is commonly set to 0 as a convenient origin. Since the liability cannot be observed, the variation in the liability is scaled to be $\sigma_e^2 = 1$. Thus:

$$p(y_i = 1 | \eta_i, \tau) = 1 - \Phi_e(-\eta_i), \quad (10)$$

with $e_i \sim N(0, 1)$.

Considering the simplest case, where the population mean is the only fixed effect in the threshold model, i.e., $\mathbf{X}_i\boldsymbol{\beta} = 1\boldsymbol{\mu}$, the probability of observing event 1 can be estimated as:

$$p(y_i = 1 | \hat{\boldsymbol{\mu}}, \hat{a}_i) = \hat{\pi}_i = 1 - \Phi_e(-(\hat{\boldsymbol{\mu}} + \hat{a}_i)), \quad (11)$$

where $\hat{\boldsymbol{\mu}}$ is the predicted population mean and \hat{a}_i is the EBV of individual i .

Logit model

Alternatively, a logistic distribution can be assigned to the residuals, resulting in a model known as the logit model. In this model, it is assumed that the logit of an underlying probability (π_i) is a function of the linear predictors:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i\mathbf{b}_i + \mathbf{z}_i\mathbf{a}_i. \quad (12)$$

Again, assuming a simple scenario, where the only fixed effect in the model is the population mean, Eq. (12) can be rewritten as:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \hat{\boldsymbol{\mu}} + \hat{a}_i.$$

The underlying probability (π_i) can then be estimated by the inverse logit transformation:

$$\pi_i = \text{logit}^{-1}(\hat{\boldsymbol{\mu}} + \hat{a}_i) = \frac{\text{Exp}(\hat{\boldsymbol{\mu}} + \hat{a}_i)}{1 + \text{Exp}(\hat{\boldsymbol{\mu}} + \hat{a}_i)}. \quad (13)$$

Linear models

In addition to the probit and logistic regression models, it is a common practice in routine genetic evaluation to fit categorical trait data using LM, treating the traits as normally distributed. It has been shown that the loss of power when using a linear Gaussian model for categorical traits is negligible compared with implementing logit or probit regression, despite the violations of assumptions of normality (e.g., Meijering and Gianola 1985). Here, we present an approach for transforming EBVs obtained from a LM into risk probabilities in a way comparable to the probit approach. An advantage of the probit approach is that the threshold is usually set to 0. However, in a LM, the threshold (τ) must be estimated. Here, we estimate an approximate τ based on the cumulative distribution function of the normal distribution.

Given a particular threshold (τ), the proportion of an event ($y_i = 1$) (hereafter π) is the proportion of the underlying liability (l) above the threshold, i.e.:

$$1 - \pi = \Phi(\tau), \quad (14)$$

where $\Phi(\tau)$ is the cumulative probability of the normal distribution $N(\mu, \sigma^2)$. τ is unknown, but π can be estimated as the proportion of observed $y_i = 1$ events in the data. Thus, the unknown threshold (τ) can be estimated by inverse probability transformation as:

$$\tau = \Phi^{-1}(1 - \pi). \quad (15)$$

As mentioned above, the LM treats the binary observations as variables with a normal distribution. To be consistent with the mean and variance of the distribution for binary observations, the normal distribution used to derive τ is assumed to be:

$$N(\pi, \pi \times (1 - \pi)).$$

Furthermore, risk probabilities can then be estimated by Eqs. (9)–(11), but with an approximated threshold τ , which is not set to 0 as in the liability threshold model. Thus:

$$p(y_i = 1 | \hat{\mu}, \hat{a}, \tau) = \hat{\pi}_i = 1 - \Phi_e(\tau - (\hat{\mu} + \hat{a}_i)). \quad (16)$$

Finally, since we are interested in the polygenic risk transmitted from a breeding animal, e.g., a sire, which on average passes half of its breeding value to its future offspring, the polygenic risk probability transmitted is half the probability of the sire estimated using the models presented above. Thus:

$$p_{\text{poly_offspring}} = 0.5 \times p_{\text{poly_sire}}. \quad (17)$$

Combining risk probabilities from the lethal allele and polygenic components

The risk probabilities computed separately from the recessive lethal alleles and the polygenic effect can finally be combined to give the total risk probability of an individual's future offspring. Assuming that the probabilities of survival from the lethal allele and polygenic components are independent, the total risk probability (p_{total}) can be given as:

$$p_{\text{total}} = 1 - ((1 - p_{\text{leth}}) \times (1 - p_{\text{poly}})), \quad (18)$$

where p_{leth} is the risk probability from the lethal component and p_{poly} is the risk probability from the polygenic component.

Assessing the accuracy of predicted risk probabilities

It is possible to assess the accuracy of predicted total risk probabilities with a validation procedure, where EBVs for the polygenic effect of validation individuals are predicted without using observations from their offspring. The predicted EBVs are subsequently transformed into risk probabilities, which together with the risk from the lethal genes are used to compute the total risk probability. The accuracy of the predicted risk probabilities can then be computed as the correlation between the predicted total risk probabilities of the test individual's offspring and the observed proportion of offspring mortality.

Simulation experiments

We tested the proposed approach with a dataset of 15 replicates simulated using the stochastic simulation program ADAM (Pedersen et al. 2009). In each replicate, a population of animals was simulated for 4 years with overlapping generations and assuming no selection. In each year, 50 males were randomly selected to mate with 10,000 females of different parities, with each male mated to 200 females. Offspring's sex was assigned randomly with a probability of 50% for males and females. The simulations resulted in a total of 40,000 individuals in four generations with approximately equal proportions of males and females. Animals born in generations 1–3 were used as the reference population, while animals in generation 4 (G4) were used as the test population.

Genotype data were simulated mimicking the real linkage disequilibrium profile in the Danish Holstein as described in detail by Thomasen et al. (2019). The simulated genotype data included 40K markers, 1980 QTLs with polygenic effects, and 20 lethal genes with recessive allele frequencies between 0.04 and 0.05. A large number of QTLs with polygenic effects were assumed in order to

mimic a trait of mixed polygenic-major gene inheritance, where a large number of QTLs with small effects each and a few lethal genes composed the genetic architecture. Other simulation studies mimicking the bovine genome for genomic prediction have assumed similar number of QTLs (Lourenco et al. 2013), or slightly more QTLs (Hayes et al. 2009; Thomasen et al. 2019), underlying various polygenic traits. The assumed recessive allele frequencies were chosen based on the range reported for recessive lethal haplotypes detected in the Danish and Nordic cattle breeds (Wu et al. 2020). SNPs were distributed across 30 chromosomes, each 100 cM in length. The QTLs were assumed to be evenly distributed across the genome, such that on each chromosome, 66 SNPs were randomly sampled to be QTLs. The effect of each QTL was sampled from a normal distribution, and the effects collectively explained all the variation in the simulated polygenic true breeding values (TBV).

Thus, the TBV for each animal i was defined as the sum of all QTL genotypic values:

$$TBV_i = \sum g_j Q_{ij},$$

where g_j is the allele substitution effect of j th QTL, Q_{ij} is the QTL genotype at locus j in individual i , coded as 0, 1, or 2 representing the number of copies for a particular allele in the genotype. The TBVs were finally scaled to have a variance of 1 in base population through dividing TBV by the standard deviation of TBVs in the base population, i.e., setting the additive genetic variance as: $\sigma_a^2 = 1$.

Simulation of the TBVs was performed at the liability scale, with a target heritability of 0.02 at the observed scale, according to heritability estimates reported in the literature for the Holstein breed (e.g., Hansen et al. 2003; Fuerst-Waltl and Sørensen 2010; Henderson et al. 2011). The target observed-scale heritability was transformed to the underlying scale using the formula proposed by Dempster and Lerner (1950):

$$h_l^2 = \frac{h_x^2 \times \pi(1 - \pi)}{z^2}, \quad (19)$$

where h_l^2 is the heritability at the underlying scale, z is the height of the normal distribution curve at the threshold, h_x^2 is the heritability at the observed scale, which is 0.02 in this study, and π is the proportion for $y = 1$ ($\pi = 0.068$ in this study). Thus, heritability at the underlying scale was 0.075.

Simulated QTLs were not included in the construction of the genomic relationship matrices (GRMs) used for prediction. Recessive lethal loci were assigned by randomly sampling from SNPs with minor allele frequencies (MAFs) between 0.04 and 0.05 on 20 randomly selected chromosomes. Each lethal locus was located on a different chromosome, and the loci were thus assumed to be independent of one another.

Four different scenarios were simulated with regard to the penetrance of recessive lethal alleles. These included three scenarios where all 20 lethal alleles were assumed to have an equal penetrance of 60, 80, or 100%. The fourth scenario considered a mixture of four penetrance groups (with an equal number of lethal alleles): 60, 70, 80, and 100% penetrance.

Liability for death was generated by adding a residual effect to TBV. The residual effect was sampled from $e \sim N(0, \sigma_e^2)$, where the residual variance (σ_e^2) was $\frac{(1-h_l^2)\sigma_a^2}{h_l^2} = (1 - 0.075)/0.075$. Simulated phenotypic values in observed scale were either 0 if the animal survived or 1 if the animal died. Individuals were assigned phenotypic values in a stepwise manner, considering both polygenic and recessive lethal allele components. First, a threshold was calculated as the inverse cumulative distribution function of a target mortality incidence ($y = 1$) of 6.8% from a normal distribution with mean 0 and variance $\sigma_a^2 + \sigma_e^2$. Individuals with a liability greater than the threshold were subsequently assigned a phenotypic value of 1. Second, an individual's phenotype was assigned as 1, regardless of the assigned phenotype due to polygenic component, if its genotype for at least one of the recessive lethal alleles was in a homozygous state. When recessive lethal alleles were assigned a penetrance other than 100%, a proportion of the homozygous individuals for each allele was assigned a phenotype of 1 in accordance with the penetrance value assumed (60, 80%, or a mixture with an average of 75%). The average observed mortality over 15 replicates was 9.43%. On average, 29.73% of the total mortality was caused by recessive lethal allele effects, while 68.21% was due to polygenic risk and 2.06% was due to both.

Statistical analysis

Breeding values and risk probabilities were estimated using the above approach. To predict risks from polygenic and recessive lethal allele effects separately, the data used to predict polygenic effects excluded records of death caused by recessive lethal alleles (Data_poly). Similarly, the GRM used for the prediction of GEBVs was constructed without recessive lethal alleles. All GRMs used for the different scenarios were calculated using the first method presented by VanRaden (2008), and SNP allele frequencies for building GRMs were calculated directly from the SNP data. Risk probabilities from the recessive lethal alleles were computed separately and then used to compute the total risk using Eq. (18).

To test the proposed approach's superiority over the conventional approach, breeding values and risk probabilities were also estimated using a conventional approach that did not distinguish the polygenic and lethal allele

effects. Thus, the conventional approach predicted total breeding values using phenotypic data without excluding observations of death due to recessive lethal alleles (Data_all) and GRMs that did not include recessive lethal genotypes. Subsequently, the predicted GEBVs were transformed into total risk probabilities.

In addition, we also assessed the accuracy of the polygenic GEBVs predicted without distinguishing the two effects, i.e., with the data that included death due to recessive lethal alleles, but taking into account the effects of the recessive lethal alleles, either by including recessive lethal alleles in the construction of the GRM or using models that included fixed regression on recessive allelic genotype code. Below, we present the methods used to predict the GEBVs that were subsequently transformed into risk probabilities for each of the models (probit, logit, and linear).

These approaches were implemented using three statistical models, namely, generalized linear mixed models with logit and probit link functions and a linear mixed model, using DMU software (Madsen and Jensen 2013).

The probit models for analysis of the simulated data

Three probit models were used to estimate breeding values and risk probabilities due to polygenic effects. The first probit model was:

$$\text{Probit1} : \boldsymbol{\eta} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g}, \quad (20)$$

where element i of $\boldsymbol{\eta}$ is $\eta_i = \Phi^{-1}(\pi_i)$, μ is the overall mean, \mathbf{g} is the random additive genetic effects with distribution $\mathbf{g} \sim N(0, \mathbf{G}\sigma_a^2)$, and \mathbf{G} is the GRM constructed using only markers, i.e., excluding the recessive lethal loci. Both Data_poly and Data_all were analyzed using this model.

The second probit model includes fixed regression on lethal genotype:

$$\text{Probit2} : \boldsymbol{\eta} = \mathbf{1}\mu + \mathbf{x}d + \mathbf{Z}\mathbf{g}, \quad (21)$$

where d is the fixed regression coefficient for lethal genotype score and \mathbf{x} is the vector of recessive lethal statuses. Since a homozygous recessive genotype at a given locus can cause mortality regardless of the genotype at another recessive lethal locus, the element of x is 1 as long as the recessive lethal allele is in a homozygous state at any locus and “0” otherwise. The G matrix in this model also excluded genotypes of recessive lethal loci, and this model was used to analyze Data_all.

The third probit model was implemented to test the impact of including recessive lethal alleles in the GRM on genomic prediction:

$$\text{Probit3} : \boldsymbol{\eta} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g}^*, \quad (22)$$

where \mathbf{g}^* are the vectors of random additive genetic effects with distribution $N(0, \mathbf{G}^*\sigma_a^2)$, where \mathbf{G}^* is the GRM based on the markers including lethal alleles and σ_a^2 is the corresponding genetic variance. This model was used to analyze Data_all.

The logit models for analysis of the simulated data

The logit models used in the analysis were:

$$\text{Logit1} : \boldsymbol{\eta} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g}, \quad (23)$$

$$\text{Logit2} : \boldsymbol{\eta} = \mathbf{1}\mu + \mathbf{x}d + \mathbf{Z}\mathbf{g}, \quad (24)$$

$$\text{Logit3} : \boldsymbol{\eta} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g}^*, \quad (25)$$

where element i of $\boldsymbol{\eta}$ is $\eta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ and the rest of the model components are the same as those in Models (20)–(22).

The linear models for analysis of the simulated data

The implemented LM included three genomic best linear unbiased prediction models:

$$\text{LM1} : \mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}, \quad (26)$$

$$\text{LM2} : \mathbf{y} = \mathbf{1}\mu + \mathbf{x}d + \mathbf{Z}\mathbf{g} + \mathbf{e}, \quad (27)$$

$$\text{LM3} : \mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g}^* + \mathbf{e}, \quad (28)$$

where \mathbf{y} is the vector of observations (0, 1) and \mathbf{e} is the vector of residuals with distribution $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where \mathbf{I} is an identity matrix and σ_e^2 is the residual variance. The rest of the model components in (26)–(28) are as described in Models (20)–(22).

Computation of prediction accuracy and bias

To assess the accuracy of predicted risk probabilities, the predicted total transmitted risk probabilities of the sires of the G4 animals were compared with the observed proportion of deaths among their G4 offspring, whose phenotypes were masked during the prediction. This is to be consistent with the situation where the candidate animals do not have offspring at the time of selection. For the approach that distinguished polygenic and recessive lethal allele effects, the predicted total risk probability was calculated using Eq. (18) to combine the risk probabilities from the recessive lethal allele and polygenic components obtained from GEBVs that were predicted using LM1, Logit1, or Probit1 and Data_poly. For the analysis that did not consider lethal

genes, the predicted total risk probabilities were calculated from the GEBVs predicted using LM1, Logit1, or Probit1 but based on Data_all. Prediction bias was measured as the coefficient of regression of the observed rate of calf mortality against the predicted probabilities for the validation sires.

In addition, the advantages of incorporating recessive lethal alleles, by either including them in the GRM (LM3, Logit3, and Probit3) or considering lethal genotype as a fixed effect in regression (LM2, Logit2, and Probit2) to predict polygenic breeding values, were assessed by the accuracy of GEBVs, which was measured as the correlation between GEBVs and simulated polygenic TBVs.

The statistical significance of differences in prediction accuracies between scenarios (approaches) and models, i.e., probit, logit, and linear, was tested using a pairwise *t*-test across the replicates.

Results

Accuracy and bias of predicted risk probabilities

Figure 1 presents the accuracy of the total risk probability predicted with the two approaches in comparison: (1) the novel approach in which risk probabilities from the polygenic and recessive lethal components were estimated separately, with the polygenic component predicted using LM1, Logit1, or Probit1 based on Data_poly, and (2) the

conventional approach where risk probability was estimated without distinguishing the polygenic and recessive lethal effects and obtained from GEBVs predicted using LM1, Logit1, or Probit1 but based on Data_all. Across all penetrance scenarios, the accuracies obtained with the novel approach were significantly higher ($P < 0.001$) than those obtained with the conventional approach. The difference in prediction accuracy between the two approaches ranged between 20 and 29.1 percentage points, depending on the penetrance scenario assumed.

In all three statistical models, i.e., LM1, Probit1, and Logit1, the highest accuracy was observed when all lethal alleles had 100% penetrance (Pen100), while the lowest accuracy was observed when all alleles had the lowest penetrance (Pen60). Among the three models, Probit1 resulted in the highest prediction accuracies, followed by Logit1, in both approaches. However, these differences in prediction accuracy between the three models were not statistically significant.

Table 1 presents the regression coefficients for observed calf mortality against predicted total risk probability for the novel approach which distinguished recessive lethal allele and polygenic effects and used Data_poly, and the conventional approach, which did not distinguish the two effects and used Data_all. For the analysis distinguishing the two effects, the regression coefficients were close to 1 for all three models and penetrance scenarios. For the analysis with conventional approach, however, the regression coefficients deviated from 1 for all models and

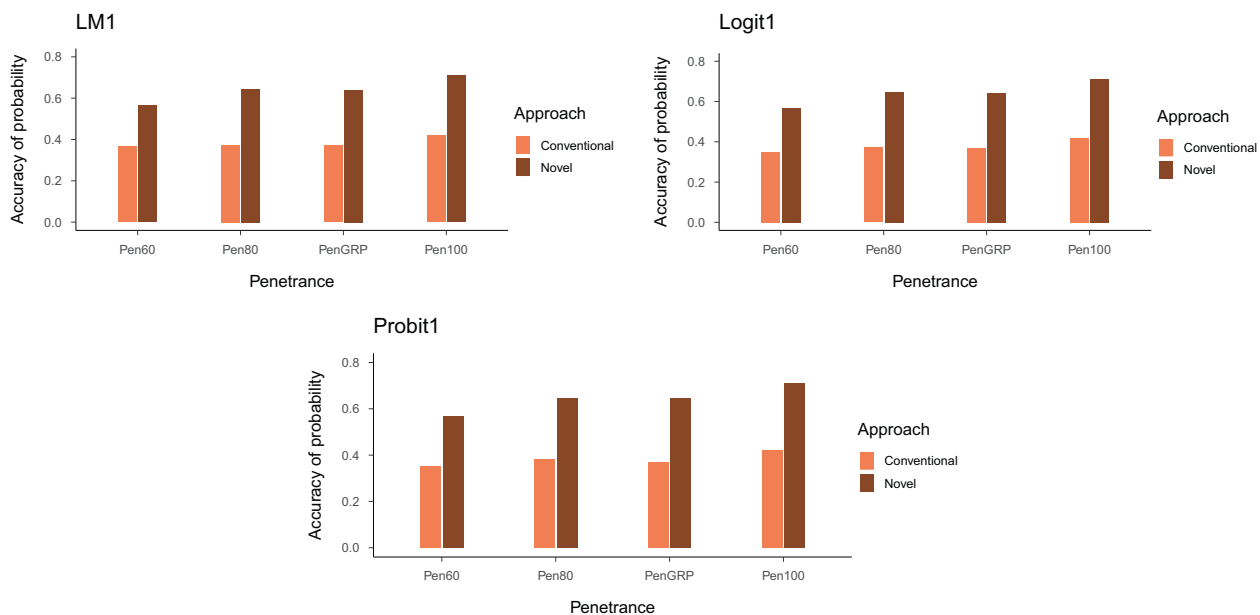


Fig. 1 Risk probability prediction accuracies in the two approaches. Accuracies of predicted total risk probabilities obtained by the novel approach, distinguishing polygenic and recessive lethal allele effects (based on Data_poly), and conventional approach, not

distinguishing polygenic and recessive lethal allele effects (based on Data_all), across the penetrance scenarios, plotted for each model (LM1, Probit1, and Logit1).

Table 1 Regression coefficients for observed mortality against predicted total risk probability, obtained by the approach distinguishing polygenic and recessive lethal allele effects (based on Data_poly) and the approach not distinguishing polygenic and recessive lethal allele effects (based on Data_all).

Model	Data_poly				Data_all			
	Pen100	PenGRP	Pen80	Pen60	Pen100	PenGRP	Pen80	Pen60
LM1	1.086	1.086	1.100	1.125	2.073	1.989	2.061	2.195
Logit1	1.066	1.061	1.070	1.072	1.344	1.252	1.303	1.275
Probit1	1.072	1.069	1.080	1.089	1.412	1.332	1.428	1.362

Pen60, 80, 100, and GRP = penetrance level of 60, 80, 100 and a mixture of 60%, 70%, 80%, and 100% penetrance levels, respectively. Data_poly = phenotype data that excluded the records of death due to lethal alleles. Data_all = phenotype data that included the records of death due to lethal alleles.

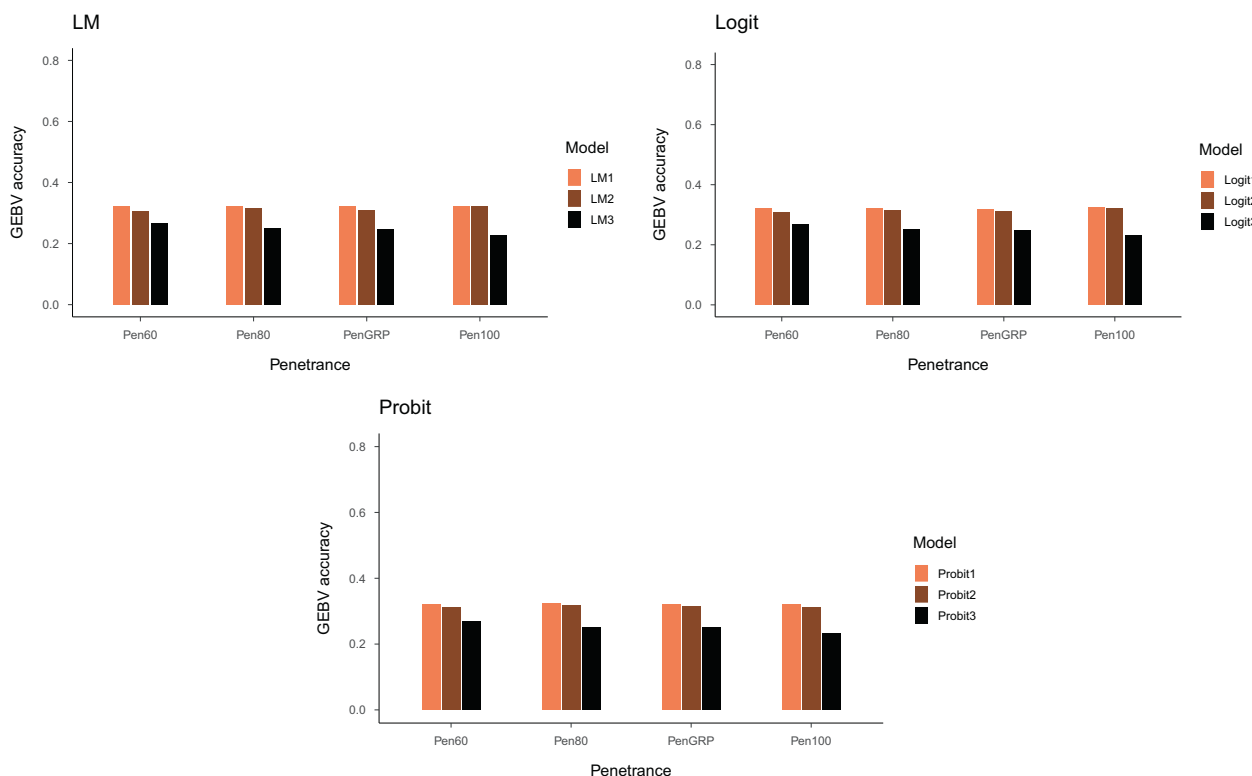


Fig. 2 GEBV accuracies across the different approaches. Accuracy of GEBVs predicted using phenotypic data that excluded the records of mortality due to lethal alleles (Data_poly) with the LM1, Probit1, and Logit1 models; data that included records of mortality due to lethal

alleles (Data_all) and models considering regression on lethal genotype (LM2, Probit2, and Logit2); and data that included records of mortality due to lethal alleles (Data_all) and models with a GRM including recessive lethal alleles (LM3, Probit3, and Logit3).

penetrance scenarios, with the largest deviation observed for the LM.

Accuracy of prediction of GEBVs

Figure 2 shows the accuracy of GEBVs predicted with the different approaches: (1) using Data_poly and a model with a GRM that did not include recessive lethal loci (LM1, Probit1, and Logit1), (2) using Data_all and a model with regression on lethal genotype (LM2, Probit2, and Logit2), and (3) using Data_all and a model with a GRM including genotypes of recessive lethal loci. In general, approach (1)

resulted in the highest accuracy of predicted GEBVs, ranging from 0.319 to 0.323 according to penetrance class. Approach (2) resulted in slightly lower accuracies compared with those in approach (1), ranging from 0.307 to 0.322. However, the differences in GEBV prediction accuracies between the two approaches were only statistically significant in scenarios Pen60 ($P < 0.01$) and PenGRP ($P < 0.05$). Among the three approaches, approach (3) produced the lowest GEBV prediction accuracies. The prediction accuracies obtained using this approach were significantly lower ($P < 0.001$) than those obtained using approaches (1) and (2).

Discussion

Prediction of risk probabilities

Genomic prediction for traits with mixed “major” genes and polygenic inheritance has been shown to benefit from models that account for differences in marker effects compared with models with “infinitesimal” assumptions (e.g., Cole et al. 2009; Hayes et al. 2010; Legarra et al. 2011). Recessive lethal loci might be considered special cases of “major” genes. While a single recessive lethal locus might have a large effect on an observed phenotype, the effect on the individual itself might be different from the effect on its future offspring. The carrier status of a single recessive lethal allele alone does not affect the carrier’s mortality but can determine the categorical outcome (death or survival) of an offspring. In this study, we present an efficient approach for predicting the total risk probabilities for future offspring of selection candidates by predicting risk probabilities from polygenic and recessive lethal components separately. By using simulated data, the prediction accuracies of this approach were compared with those of a conventional approach that did not distinguish polygenic and recessive lethal allele effects (Data_all vs. Data_poly in Fig. 1). The results show that the prediction of risk probabilities with the proposed approach leads to high accuracy in predicting mortality for the future offspring of selection candidates, with a gain in accuracy up to 29.1 percentage points. By blending the risk probabilities estimated separately for the two components, the novel approach allows efficient utilization of information from the recessive lethal component, which otherwise tends to be difficult to untangle with simultaneous modeling due to differing modes of gene action. The gain in accuracy achieved by distinguishing the polygenic and recessive lethal allele effects, compared with the approach that does not distinguish the two effects, is dependent on the rate of mortality caused by the two effects. In our analysis, the gain in accuracy (29.1%) was comparable to the rate of mortality caused by recessive lethal alleles in the simulation (29.73%).

A potential challenge for the separate prediction of risk probabilities caused by polygenic and lethal allele effects in real data scenarios might be the difficulty of conclusively distinguishing mortality caused by recessive lethal alleles and that caused by polygenic components. This is because genotypic information may not be available for dead animals. In this study, we further investigated if taking into account the recessive lethal genotypes through, either including regression on lethal genotypes in the models or by accounting for genotypes of the recessive lethal loci in the GRM, could improve the accuracy of predicted GEBVs when using data that included mortality due to lethal alleles. The results showed that when the data included records of

mortality caused by recessive lethal alleles, including the recessive lethal alleles in the GRM did not improve GEBV prediction accuracy. In contrast, when using data that included records of mortality due to lethal alleles, models with regression on lethal genotype resulted in prediction accuracy being comparable to the approach that distinguished the two effects using Data_poly. Compared with including genotypes of recessive lethal loci in the GRM, using models with fixed regression on lethal genotypes improved the accuracy of polygenic GEBVs by 4–9.3 percentage points, based on data that included records of mortality due to lethal alleles. These results demonstrate that in the situations where excluding mortality caused by recessive lethal alleles is difficult, using models with regression on lethal genotypes can improve prediction accuracies. Potential challenges in the approach considering fixed regressions on lethal genotype is the definition of a lethal covariable and the uncertain relationship between a lethal covariable and observations in the case of incomplete penetrance and unequal penetrances among lethal loci. Consequently, prediction accuracies might be affected by penetrance. This was demonstrated in our simulation, where the prediction accuracy of the model considering fixed regression on lethal genotype was significantly lower than that of the approach using Data_poly for scenarios having a lower penetrance (Pen60) and a mixture of penetrance levels (PenGRP).

In our simulation study, the accuracy of the predicted GEBVs was generally low across the compared models. This is expected given the low heritability considered in the simulation. In dairy cattle breeding, definitions of different calf and young stock mortality traits are dependent on monitoring period. In general, studies across several dairy cattle breeds have shown very low heritability estimates for calf and young stock mortality traits (e.g., Hansen et al. 2003; Fuerst-Waltl and Sørensen 2010; Henderson et al. 2011), which limit the expected genomic prediction accuracy. However, given the major deleterious effects of individual recessive lethal genes, the prediction accuracy for mortality traits can be improved with the efficient incorporation of genotypic information on such genes. The results of this study indicate that the presented novel approach is quite advantageous in integrating information on recessive lethal and polygenic components for the prediction of mortality traits. By bringing the two components to a comparable scale, i.e., risk probability, the approach allows utilizing information from both effects to predict the mortality status of future offspring of a breeding animal. A somewhat comparable approach to the prediction of risk probabilities presented in this study is the polygenic risk score (PRS) approach, which is commonly used in human genetics to predict an individual’s risk of succumbing to a particular disease (Wray et al. 2007, 2019; Evans et al.

2009). However, the PRS used in human genetics is predicted based on SNP effects estimated from genome-wide association studies that are often based on fitting one SNP at a time, thus ignoring all other SNPs (Wray et al. 2007). Moreover, PRSs in human genetics are used to predict the future phenotypes of an individual, while the primary objective in our approach, and in animal breeding more generally, is to predict a selection candidate's transmission ability to its future offspring.

There are several assumptions in our simulation study that might not be fully consistent with the features of real data and thus might affect prediction accuracies to some extent. We have shown that the gain in accuracy achieved by distinguishing polygenic and recessive lethal allele effects is dependent on the rate of mortality caused by recessive lethal alleles. This rate, in turn, depends on the number of recessive lethal loci and recessive allele frequencies. In the simulation, 20 loci with lethal allele frequencies between 0.04 and 0.05 were assumed. These might be considered high-frequency lethal alleles compared with what one would expect for a lethal allele under mutation-selection balance or drift. Therefore, in cases with smaller numbers of recessive lethal loci with lower MAFs, the mortality caused by lethal alleles will be lower, subsequently resulting in a smaller gain achieved by distinguishing the two effects. However, several recessive lethal mutations have been identified in cattle breeds, and the numbers continue to increase (Cole 2015), with some reaching high recessive allele frequency (e.g., Kadri et al. 2014; Sahana et al. 2016; Hoff et al. 2017).

An additional assumption potentially prone to violation in real scenarios is the independence of recessive lethal loci and the independence of recessive lethal loci and non-lethal loci across the genome. In reality, recessive lethal alleles might be in LD with each other as well as with other loci. However, this is expected to have negligible consequences when using the novel approach that predicts risk probability due to lethal allele and polygenic effects separately, where polygenic GEBVs are estimated using Data_poly, but may cause confounding between the two effects when using the approach that does not distinguish the two effects, based on Data_all. An additional issue that was not taken into account in our simulation is the possibility of synergistic epistasis between the recessive lethal loci and other loci with polygenic effects. Under such interaction, the lethality, or penetrance, of recessive lethal loci may depend on polygenic effects, thus risking double counting of lethal effects when GEBVs are estimated in the presence of the lethal alleles. Such epistatic interactions were not considered in this study due to the complexity and lack of prior information for the simulation.

Comparison of models

Categorical traits are not normally distributed, and thus linear mixed models are believed to behave poorly in modeling such traits (Portnoy 1982). Despite such violations of normality assumptions, the use of linear mixed models in the genetic analysis of categorical traits is gaining popularity due to their straightforward implementation. Meijering and Gianola (1985) demonstrated that LM can be applied without much loss of statistical power. In our study, slight differences in prediction accuracy were observed between the three models implemented, i.e., logit model, probit model, and LM, but the differences were not statistically significant. These results indicate that our approach can be implemented in LM with negligible loss of accuracy.

The regression coefficients of observed proportions of calf mortality against predicted risk probabilities were different from 1 when using the models that did not distinguish polygenic and lethal allele effects. The deviation from 1 was much larger for the LM than for the logit and probit models. This could partly be because in the LM, the threshold is approximated by direct calculation of mortality from the data, as opposed to the probit model, where the threshold is set to 0 for convenience and the underlying liability moves the origin accordingly. For the LM, the deviation from 1 was even larger for the approach based on Data_all that did not distinguish recessive lethal allele and polygenic effects. This could be explained by the fact that the threshold was approximated from observed mortality in the data, including mortality due to recessive lethal alleles, and hence the approximate threshold could be far from the threshold for the polygenic model. Moreover, the relationship between sire risk probability due to polygenic effects and offspring mortality is not necessarily linear. Consequently, the regression coefficient of observed mortality against the predicted risk probability might not necessarily be 1.

Management of recessive lethal alleles in breeding programs

To date, commonly proposed methods for managing recessive lethal alleles have focused on the optimization of mate selection to avoid carrier-to-carrier matings. Van Eenennaam and Kinghorn (2014) proposed methods and programs that allow selection against the total number of lethal alleles and recessive lethal genotypes. Cole (2015) extended the parent-average penalizing method for controlling inbreeding proposed by Pryce et al. (2012), allowing it to consider information on recessive lethal alleles. Some studies also suggested the complete removal of carriers from the breeding population to eradicate recessive lethal mutations (e.g., Thompson et al. 2006).

Managing recessive lethal alleles requires a trade-off between controlling recessive lethal alleles in the long run and maintaining genetic gains in production and functional traits (Segelke et al. 2014). Previously proposed methods aimed at optimizing mate selection as well as culling carriers might allow the control of recessive lethal allele frequencies and avoid lethal homozygous genotypes. However, these methods represent classic tandem selection, where breeding animals are excluded from mating due to low merit for one trait (recessive lethal alleles in this case), regardless of their superiority in other traits. The consequence of this approach is a reduction in selection intensity and a subsequent reduction in genetic gain. The slightly different approach proposed by Segelke et al. (2014) recommends a selection index that weights the carrier status of recessive lethal haplotypes based on economic consequences and population allele frequencies when selecting females for mating. A drawback of this approach, and many other mate-allocation-based approaches, is the inability to handle many recessive lethal alleles. For instance, Cole (2015) pointed out the difficulty of assigning proper weights and costs for each recessive lethal allele as the number of identified alleles increases.

The approach proposed in this study enables blending polygenic breeding values for a given trait with risk probabilities from recessive lethal alleles. Thus, the method is beneficial for a balance between controlling recessive lethal frequencies in the population and maintaining genetic gains in economically important traits. In contrast to the methods where carrier status for each recessive lethal allele is a selection criterion, the proposed method integrates the effect of each recessive lethal allele into the breeding value for a particular trait (mortality or survival), which can be used for selection decisions. Therefore, an overall weight for the trait of interest can be used to integrate the breeding values, which account for both the polygenic and recessive lethal allele components, into a selection index with no need to assign weights for each recessive lethal allele.

Conclusions

This study proposed an approach for predicting the probability of mortality of future offspring by predicting the risk probabilities from polygenic and recessive lethal components separately. The approach was tested using simulated data and found to be superior to approaches that do not distinguish polygenic and lethal allele effects. No statistically significant differences in prediction accuracy were observed between the probit model, logit model, and LM, suggesting that the novel approach can be implemented using different models, with comparable power.

Data availability

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.xd2547ddv>.

Acknowledgements This work was supported by the GUDP project “LiveCalf” (No. 34009-16-1101) from the Ministry of Environment and Food of Denmark (Copenhagen).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agerholm JS, Bendixen C, Andersen O, Arnbjerg J (2001) Complex vertebral malformation in holstein calves. *J Vet Diagn Invest* 13(4):283–289
- Charlier C, Coppiepieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N et al. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* 40(4):449–454. <https://doi.org/10.1038/ng.96>
- Cole JB (2015) A simple strategy for managing many recessive disorders in a dairy cattle breeding program. *Genet Sel Evol* 30(47):94. <https://doi.org/10.1186/s12711-015-0174-9>
- Cole JB, Null DJ, VanRaden PM (2016) Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. *J Dairy Sci* 99(9):7274–7288. <https://doi.org/10.3168/jds.2015-10777>
- Cole JB, VanRaden PM, O’Connell JR, Van Tassell CP, Sonstegard TS, Schnabel RD et al. (2009) Distribution and location of genetic effects for dairy traits. *J Dairy Sci* 92(6):2931–2946. <https://doi.org/10.3168/jds.2008-1762>
- Dempster ER, Lerner IM (1950) Heritability of threshold characters. *Genetics* 35:212–235
- Evans DM, Visscher PM, Wray NR (2009) Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet* 18:3525–3531. <https://doi.org/10.1093/hmg/ddp295>
- Fernando RL (1989) Grossman M. Marker assisted selection using best linear unbiased prediction. *Genet Sel Evol* 21:467–477
- Fritz S, Capitan A, Djari A, Rodriguez SC, Barbat A, Baur A et al. (2013) Detection of haplotypes associated with prenatal death in

- dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS One* 8(6):e65550
- Fuerst-Waltl B, Sørensen MK (2010) Genetic analysis of calf and heifer losses in Danish Holstein. *J Dairy Sci* 93(11):5436–5442. <https://doi.org/10.3168/jds.2010-3227>
- Hansen M, Madsen P, Jensen J, Pedersen J, Christensen LG (2003) Genetic parameters of postnatal mortality in Danish Holstein calves. *J Dairy Sci* 86(5):1807–1817
- Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res* 91(1):47–60. <https://doi.org/10.1017/S0016672308009981>
- Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet* 6:e1001139
- Henderson L, Miglior F, Sewalem A, Kelton D, Robinson A, Leslie KE (2011) Estimation of genetic parameters for measures of calf survival in a population of Holstein heifer calves from a heifer-raising facility in New York State. *J Dairy Sci* 94(1):461–470. <https://doi.org/10.3168/jds.2010-3243>
- Hoeschele I (1988) Genetic evaluation with data presenting evidence of mixed major gene and polygenic inheritance. *Theor Appl Genet* 76(1):81–92. <https://doi.org/10.1007/BF00288836>
- Hoff JL, Decker JE, Schnabel RD, Taylor JF (2017) Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics* 18(1):799
- Johnsson M, Gaynor RC, Jenko J, Gorjanc G, de Koning DJ, Hickey JM (2019) Removal of alleles by genome editing (RAGE) against deleterious load. *Genet Sel Evol* 51(1):14. <https://doi.org/10.1186/s12711-019-0456-8>
- Kadri NK, Sahana G, Charlier C, Iso-Touru T, Gulbrandsen B, Karim L et al (2014) A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock *PLoS Genet* 10(1):e1004049
- Legarra A, Robert-Granié C, Croiseau P, Guillaume F, Fritz S (2011) Improved Lasso for genomic selection. *Genet Res* 93:77–87
- Lourenco DA, Misztal I, Wang H, Aguilar I, Tsuruta S, Bertrand JK (2013) Prediction accuracy for a simulated maternally affected trait of beef cattle using different genomic evaluation models. *J Anim Sci* 91(9):4090–4098. <https://doi.org/10.2527/jas.2012-5826>
- Madsen P, Jensen J (2013) A user's guide to DMU. version 6, release 5.2. Aarhus University Foulum, Denmark
- Meijering A, Gianola D (1985) Linear versus nonlinear methods of sire evaluation for categorical traits: a simulation study. *Genet Sel Evol* 17(1):115–132. <https://doi.org/10.1186/1297-9686-17-1-115>
- Pedersen LD, Sørensen AC, Henryon M, Ansari-Mahyari S, Berg P (2009) ADAM: a computer program to simulate selectivebreeding schemes for animals. *Livest Sci* 121:343–344. <https://doi.org/10.1016/j.livsci.2008.06.028>
- Peñagaricano F, Urioste JI, Naya H, de los Campos G, Gianola D (2011) Assessment of poisson, probit and linear models for genetic analysis of presence and number of black spots in Corriedale sheep. *J Anim Breed Genet* 128(2):105–113. <https://doi.org/10.1111/j.1439-0388.2010.00893.x>
- Portnoy S (1982) Maximizing the probability of correctly ordering random variables using linear predictors. *J Mult Anal* 12:256–269
- Pryce JE, Hayes BJ, Goddard ME (2012) Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *J Dairy Sci* 95:377–388
- Rao S, Xia L (2000) Strategies for genetic mapping of categorical traits. *Genetica* 109(3):183–197
- Sahana G, Nielsen US, Aamand GP, Lund MS, Gulbrandsen B (2013) Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. *PLoS One* 20(12):e82909
- Sahana G, Iso-Touru T, Wu X, Nielsen US, de Koning DJ, Lund MS et al. (2016) A 0.5-Mbp deletion on bovine chromosome 23 is a strong candidate for stillbirth in Nordic Red cattle. *Genet Sel Evol* 48:35
- Segelke D, Täubert H, Jansen S, Pausch H, Reinhardt F, Thaller G (2014) Management of genetic characteristics. *Interbull Bull* 48:85–88
- Shuster DE, Kehrl Jr ME, Ackermann MR, Gilbert RO (1992) Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proc Nat Acad Sci USA* 89(19):9225–9229
- Thomassen JR, Liu H, Sørensen AC (2019) Genotyping more cows increases genetic gain and reduces rate of true inbreeding in a dairy cattle breeding scheme using female reproductive technologies. *J Dairy Sci* 13. <https://doi.org/10.3168/jds.2019-16974>
- Thompson PN, Heesterbeek JA, van Arendonk JA (2006) Changes in disease gene frequency over time with differential genotypic fitness and various control strategies. *J Anim Sci* 84(10):2629–2635
- Van Eenennaam AL, Kinghorn BP (2014) Use of mate selection software to manage lethal recessive conditions in livestock populations. In: *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*, Vancouver, 17–22 Aug 2014
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423
- VanRaden PM, Olson KM, Null DJ, Hutchison JL (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J Dairy Sci* 94:6153–6161
- Wray NR, Goddard ME, Visscher PM (2007) Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 17:1520–1528. <https://doi.org/10.1101/gr.6665407>
- Wray NR, Kemper KE, Hayes BJ, Goddard ME, Visscher PM (2019) Complex trait prediction from genome data: contrasting EBV in livestock to PRS in humans: genomic prediction. *Genetics* 211(4):1131–1141. <https://doi.org/10.1534/genetics.119.301859>
- Wright S (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19:506
- Wu X, Mesbah-Uddin M, Gulbrandsen B, Lund MS, Sahana G (2020) Novel haplotypes responsible for prenatal death in Nordic Red and Danish Jersey cattle. *J Dairy Sci* 103(5):4570–4578. <https://doi.org/10.3168/jds.2019-17831>