Converting linear breeding values to probabilities for health traits in dairy cattle

J. Hidalgo, S. Tsuruta, D. Gonzalez, G. de Oliveira, M. Sanchez, A. Kulkarni, C. Przybyla, G. Vargas, N. Vukasinovic, I. Misztal, D. Lourenco

zoetis

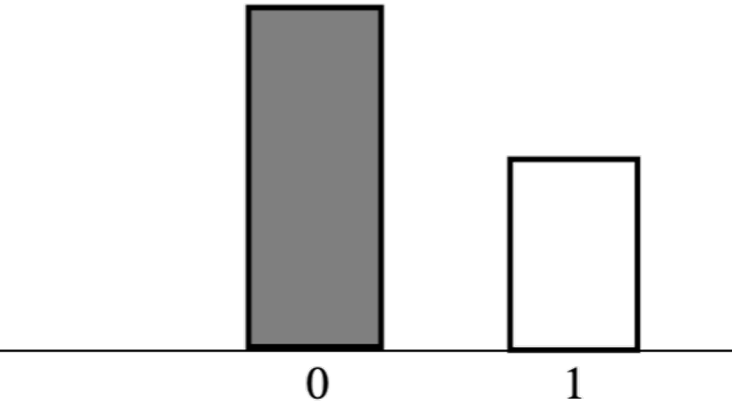**Animal Breeding and Genetics Group**
College of Agricultural & Environmental Sciences
UNIVERSITY OF GEORGIA

1785

About threshold models

Breeding values as probabilities

"A neat feature"

$$P_i = 1 - \Phi\left(\frac{t - x_i'\beta - z_i'u}{\sigma_e}\right)$$
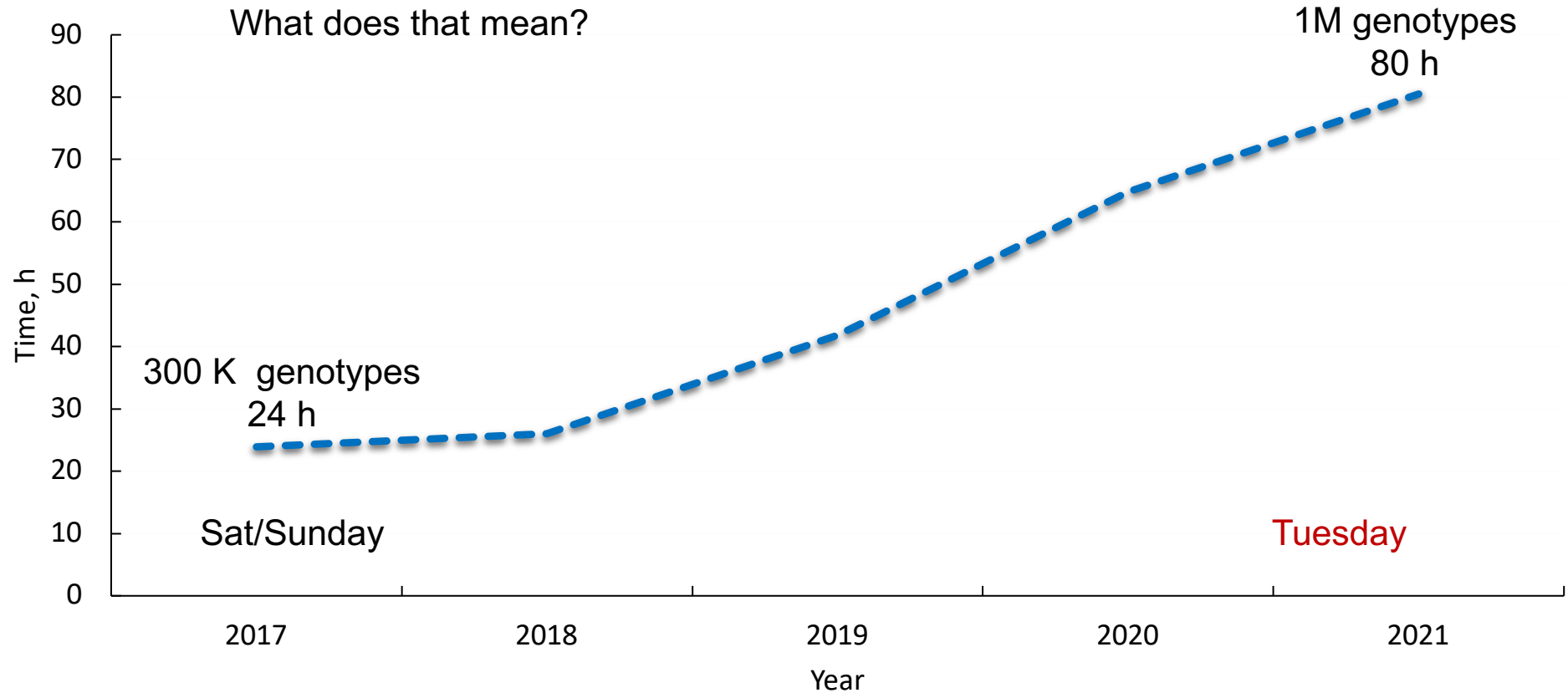
$$\mathcal{L} = \mathbf{1}v + \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

$$\begin{bmatrix} \mathbf{T}^{[i-1]} & \mathbf{L}'^{[i-1]}\mathbf{X} & \mathbf{L}'^{[i-1]}\mathbf{Z} \\ \mathbf{X}'\mathbf{L}^{[i-1]} & \mathbf{X}'\mathbf{W}^{[i-1]}\mathbf{X} & \mathbf{X}'\mathbf{W}^{[i-1]}\mathbf{Z} \\ \mathbf{Z}'\mathbf{L}^{[i-1]} & \mathbf{Z}'\mathbf{W}^{[i-1]}\mathbf{X} & \mathbf{Z}'\mathbf{W}^{[i-1]}\mathbf{Z} + \mathbf{H}^{-1}\lambda \end{bmatrix}\begin{bmatrix} \Delta_t^{[i]} \\ \Delta_b^{[i]} \\ \Delta_u^{[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{p}^{[i-1]} \\ \mathbf{X}'\mathbf{v}^{[i-1]} \\ \mathbf{Z}'\mathbf{v}^{[i-1]} - \mathbf{H}^{-1}\lambda\mathbf{u}^{[i-1]} \end{bmatrix}$$

"The mathematics resists simple computations"

Dempster and Lerner, 1949
Gianola and Foulley, 1983

Increased computational demands

Example: weekly evaluations for calving ease

$r_{GEBV_{lia,lin}} \geq 0.95$

Different scale

Derive a formula to convert linear breeding values to probabilities

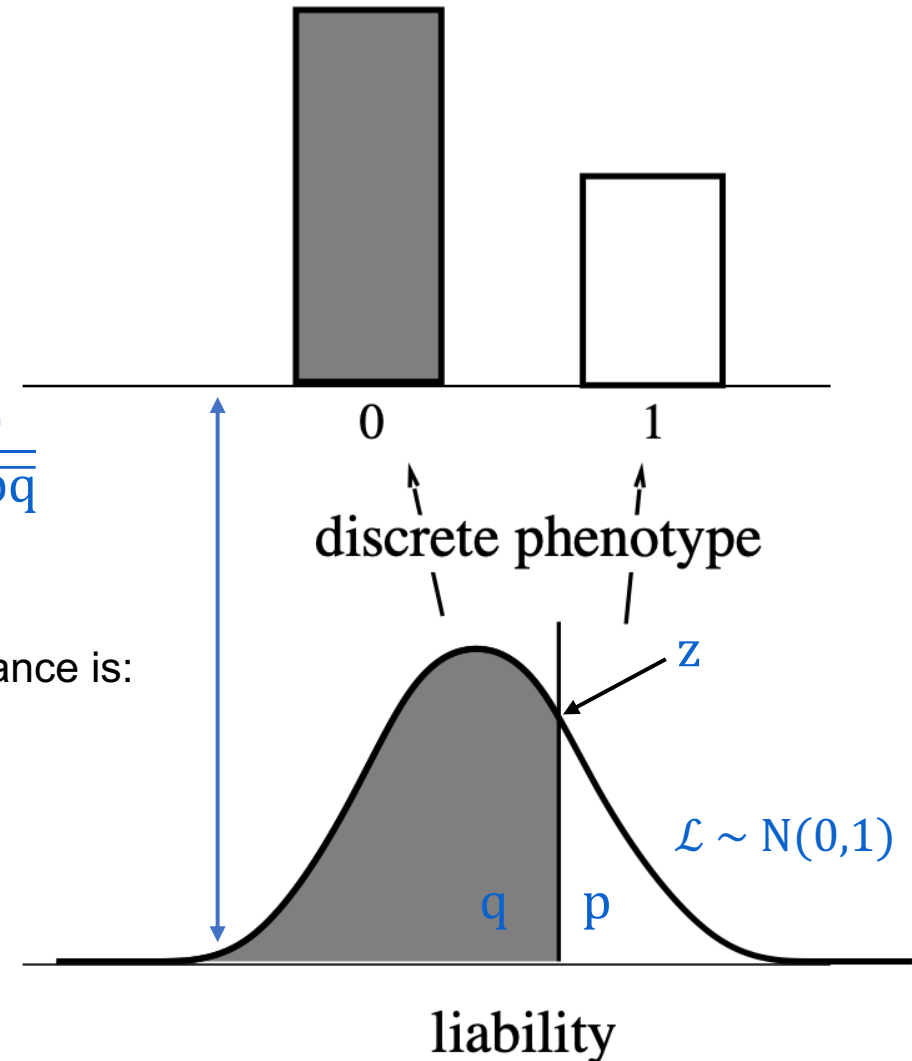# Derivation

$$h^2_{lin} = \frac{z^2 h^2_{lia}}{pq}$$

Genetic variance is fully additive in the liability scale, but this property does not hold on the linear scale

There is some non-additive genetic variance

$$r_{lia-lin} = \frac{\sigma_{lia-lin}}{\sqrt{\sigma^2_{lia}\sigma^2_{lin}}} = \frac{z}{\sqrt{pq}}$$

The proportion of non-additive variance is:
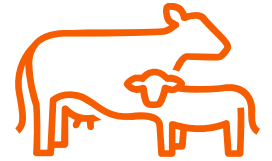
$$1 - r^2_{lia-lin} = 1 - \frac{z^2}{pq}$$

discrete phenotype

$\mathcal{L} \sim N(0,1)$

liability

Robertson, 1949
Dempster and Lerner, 1949

Derivation

The difference in heritabilities is determined by the proportion of phenotypic variance that is non-additive in the linear scale

$$\frac{\sigma^2_{non-a_{lin}}}{\sigma^2_{phen_{lin}}} = 1 - r^2_{lia-lin} = 1 - \frac{z^2}{pq} = 1 - \frac{h^2_{lin}}{h^2_{lia}}$$

Therefore, to approximate $GEBV_{lin}$ to $GEBV_{lia}$

$$GEBV_{lia} \approx \frac{GEBV_{lin}}{\sqrt{\sigma^2_{e_{lin}} * \left(1 - r^2_{lia-lin}\right)}} \approx \frac{GEBV_{lin}}{\sqrt{\sigma^2_{e_{lin}} * \left(1 - \frac{h^2_{lin}}{h^2_{lia}}\right)}}$$

Data and statistical model

Trait	Records (M)	Incidence	h^2_{lia}	h^2_{lin}
Mastitis	6.3	0.27	0.09	0.05
Lameness	5.0	0.14	0.08	0.03
Ketosis	3.6	0.06	0.08	0.02
Displaced abomasum	5.1	0.02	0.09	0.01

Pedigree with 6M animals

Genotypes for 1.7M

Threshold   vs.   Linear models

$$\begin{bmatrix} \mathbf{T}^{[i-1]} & \mathbf{L}'^{[i-1]}\mathbf{X} & \mathbf{L}'^{[i-1]}\mathbf{Z} \\ \mathbf{X}'\mathbf{L}^{[i-1]} & \mathbf{X}'\mathbf{W}^{[i-1]}\mathbf{X} & \mathbf{X}'\mathbf{W}^{[i-1]}\mathbf{Z} \\ \mathbf{Z}'\mathbf{L}^{[i-1]} & \mathbf{Z}'\mathbf{W}^{[i-1]}\mathbf{X} & \mathbf{Z}'\mathbf{W}^{[i-1]}\mathbf{Z} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \Delta_t^{[i]} \\ \Delta_b^{[i]} \\ \Delta_u^{[i]} \end{bmatrix} = \begin{bmatrix} \mathbf{p}^{[i-1]} \\ \mathbf{X}'\mathbf{v}^{[i-1]} \\ \mathbf{Z}'\mathbf{v}^{[i-1]} - \mathbf{H}^{-1}\lambda\mathbf{u}^{[i-1]} \end{bmatrix}$$
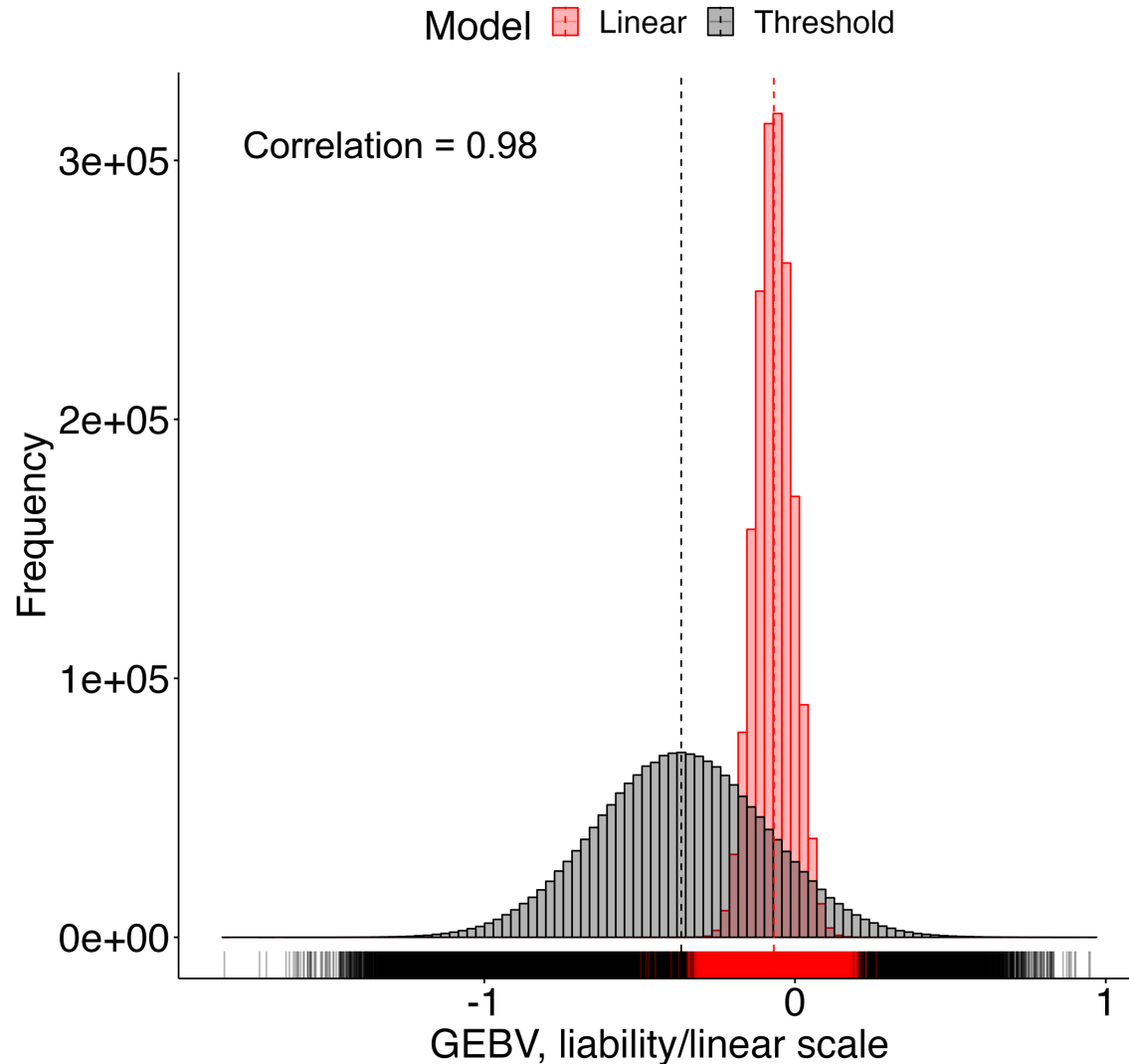
$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$
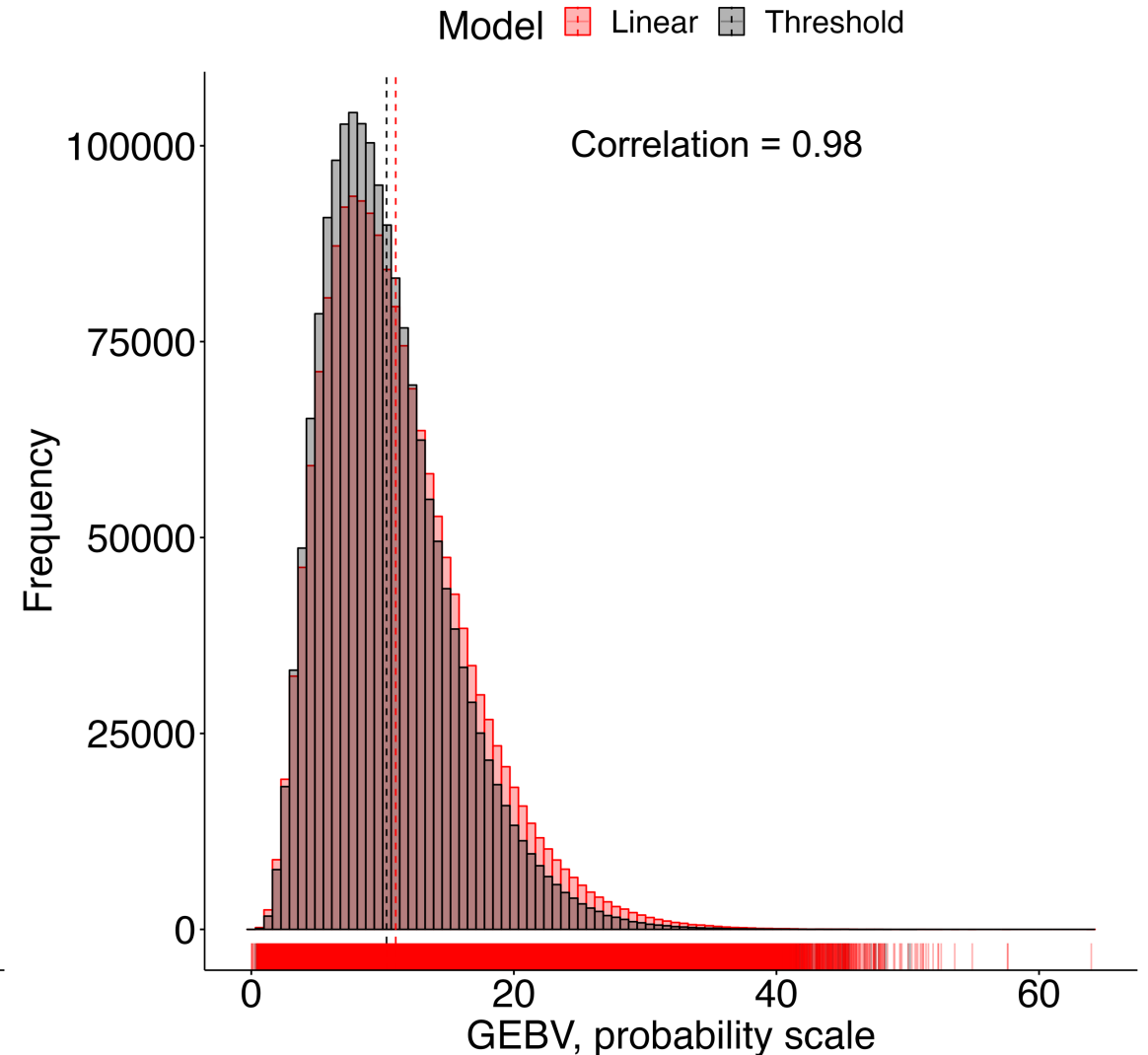
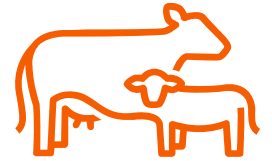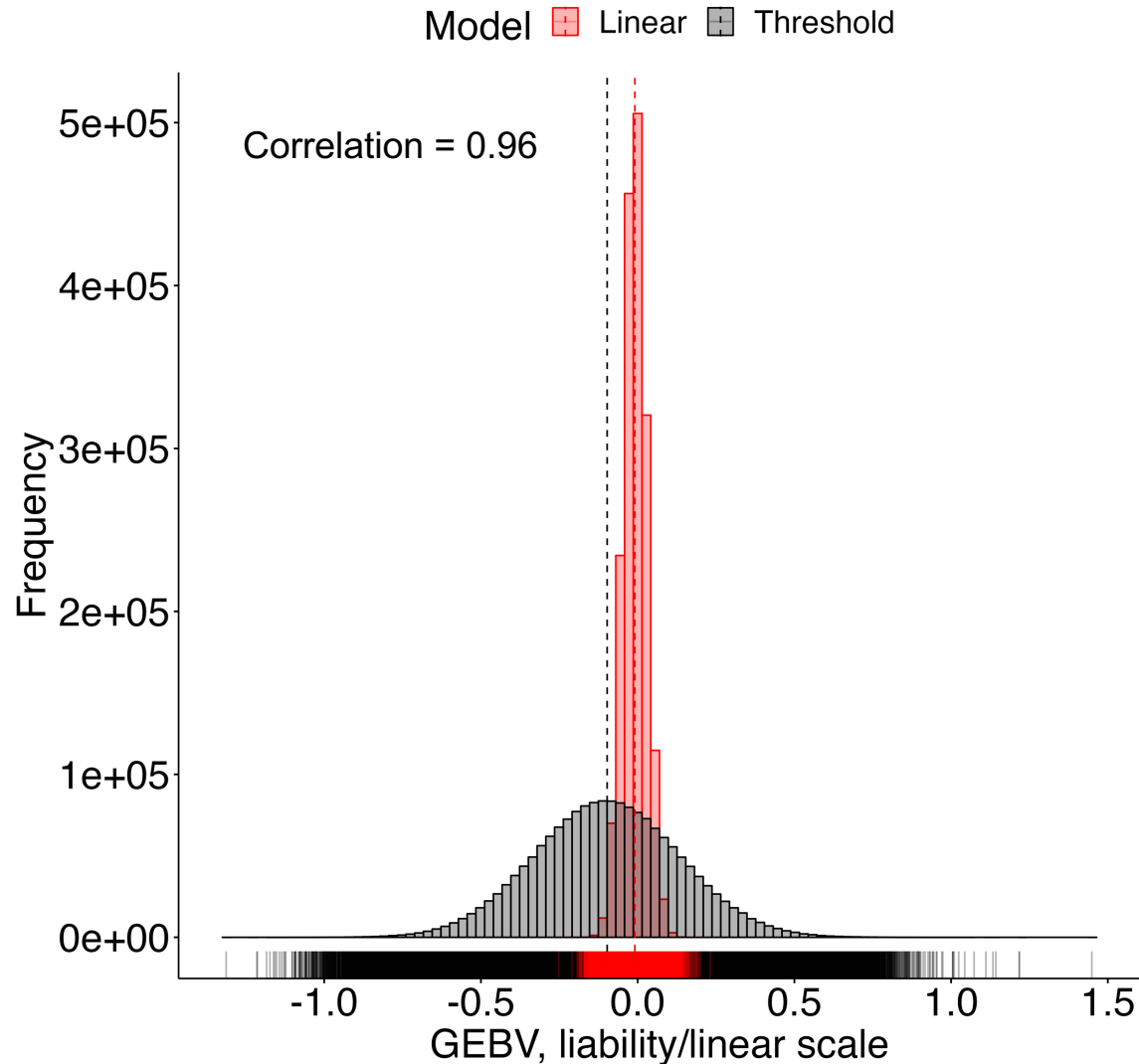# From linear to probability scale

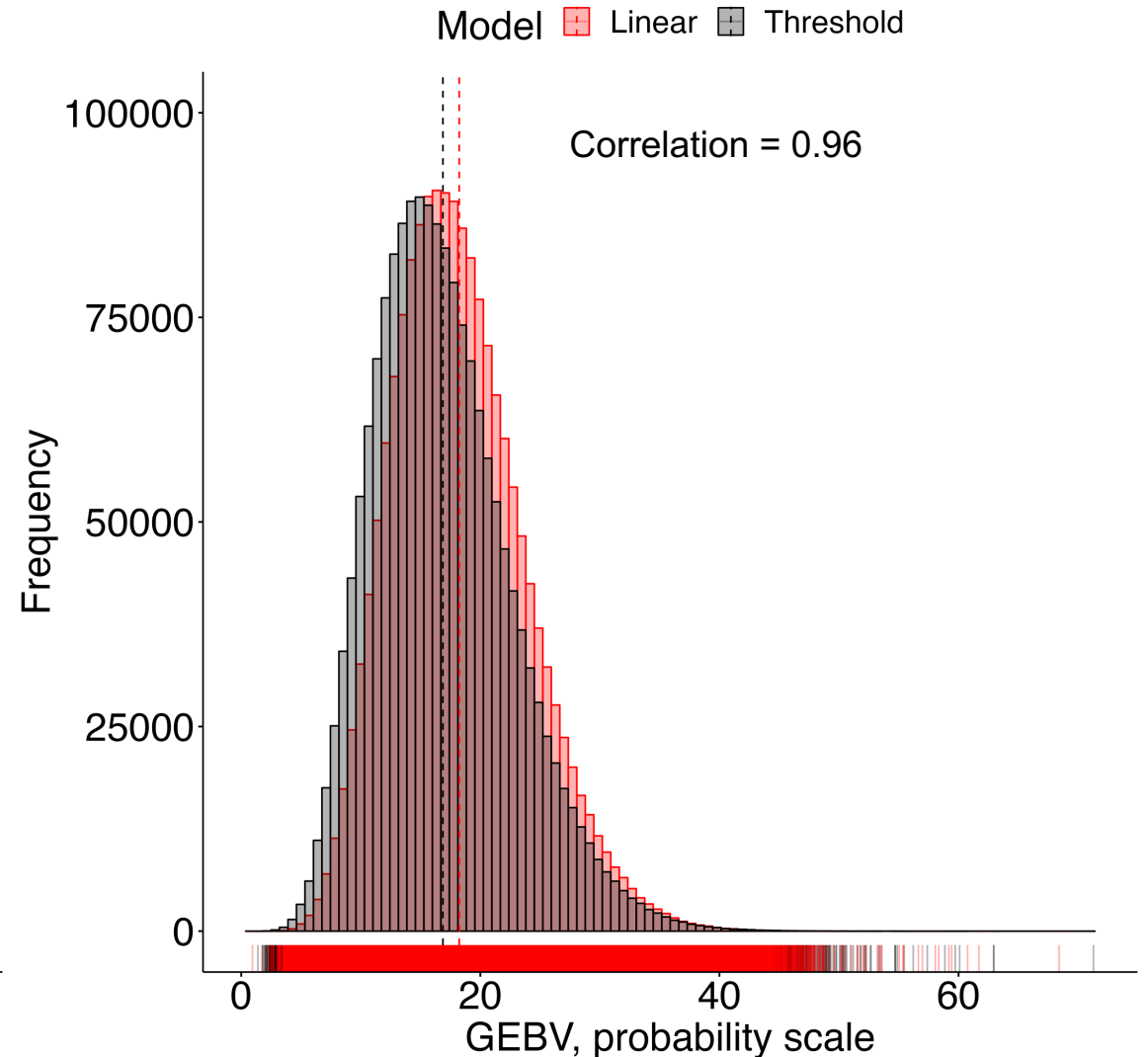Mastitis – incidence = 0.27
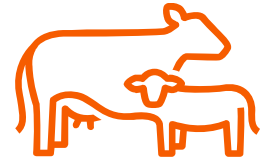
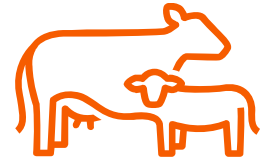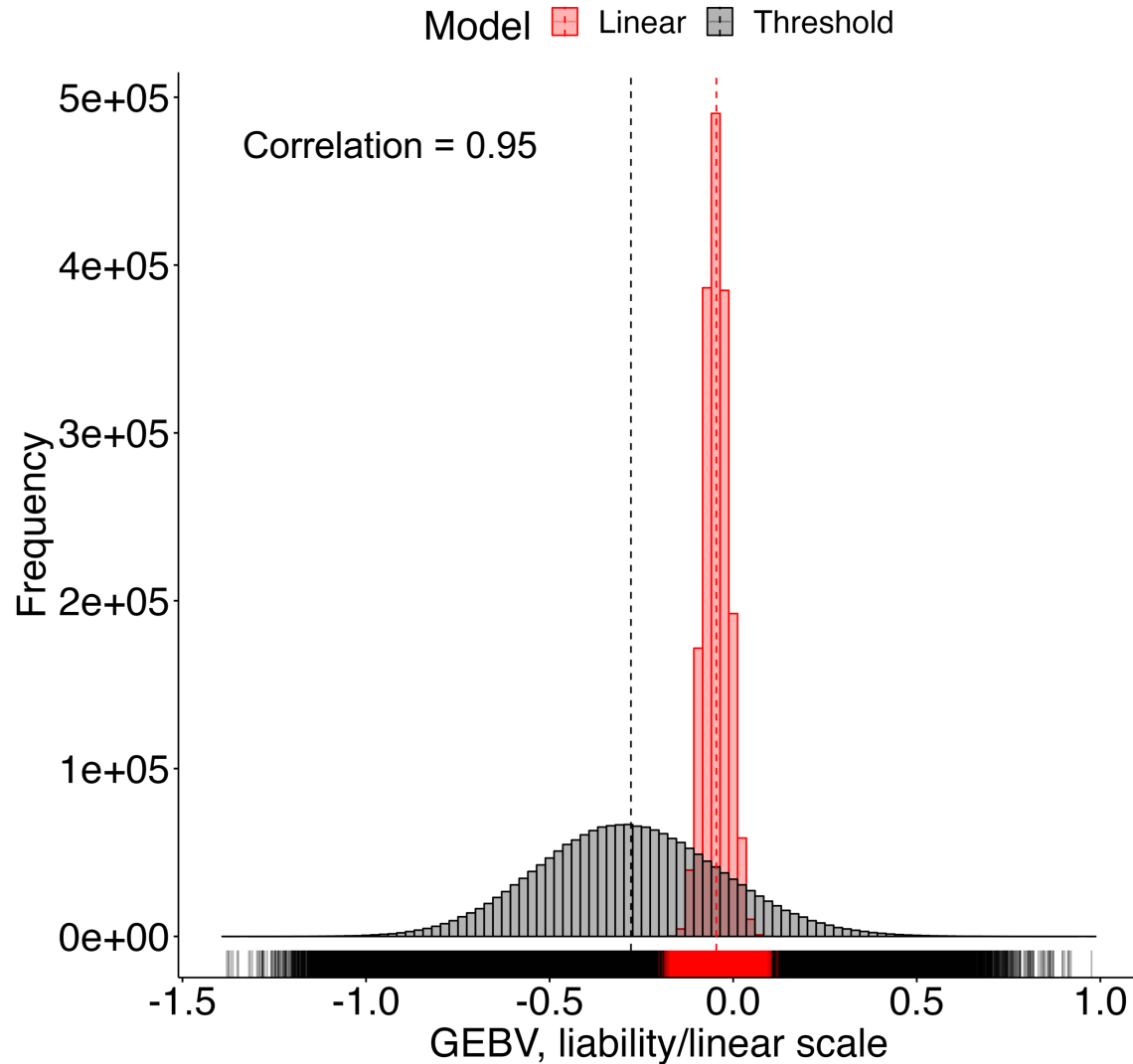# From linear to probability scale
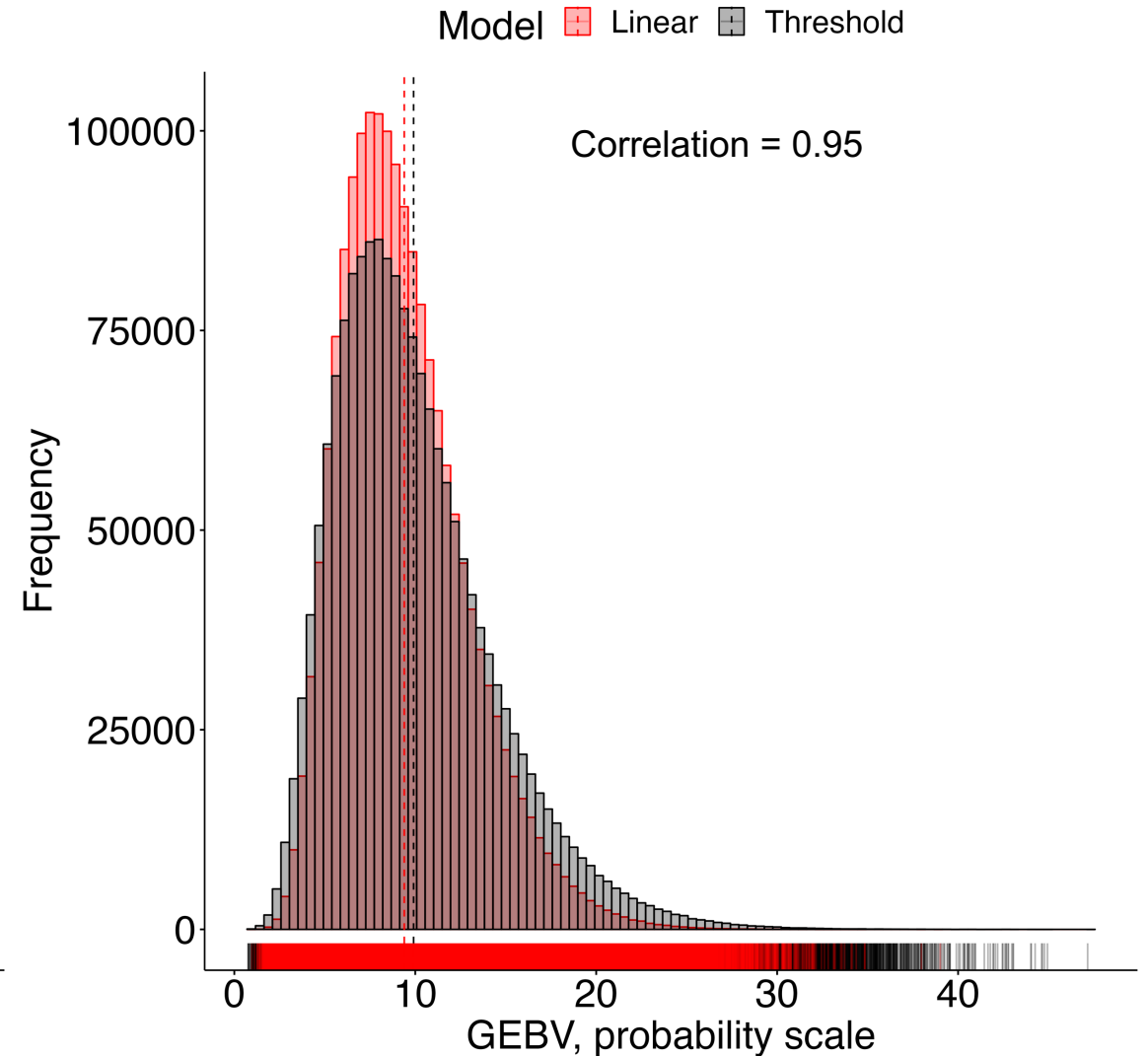
Lameness – incidence = 0.14

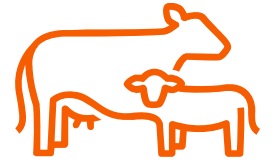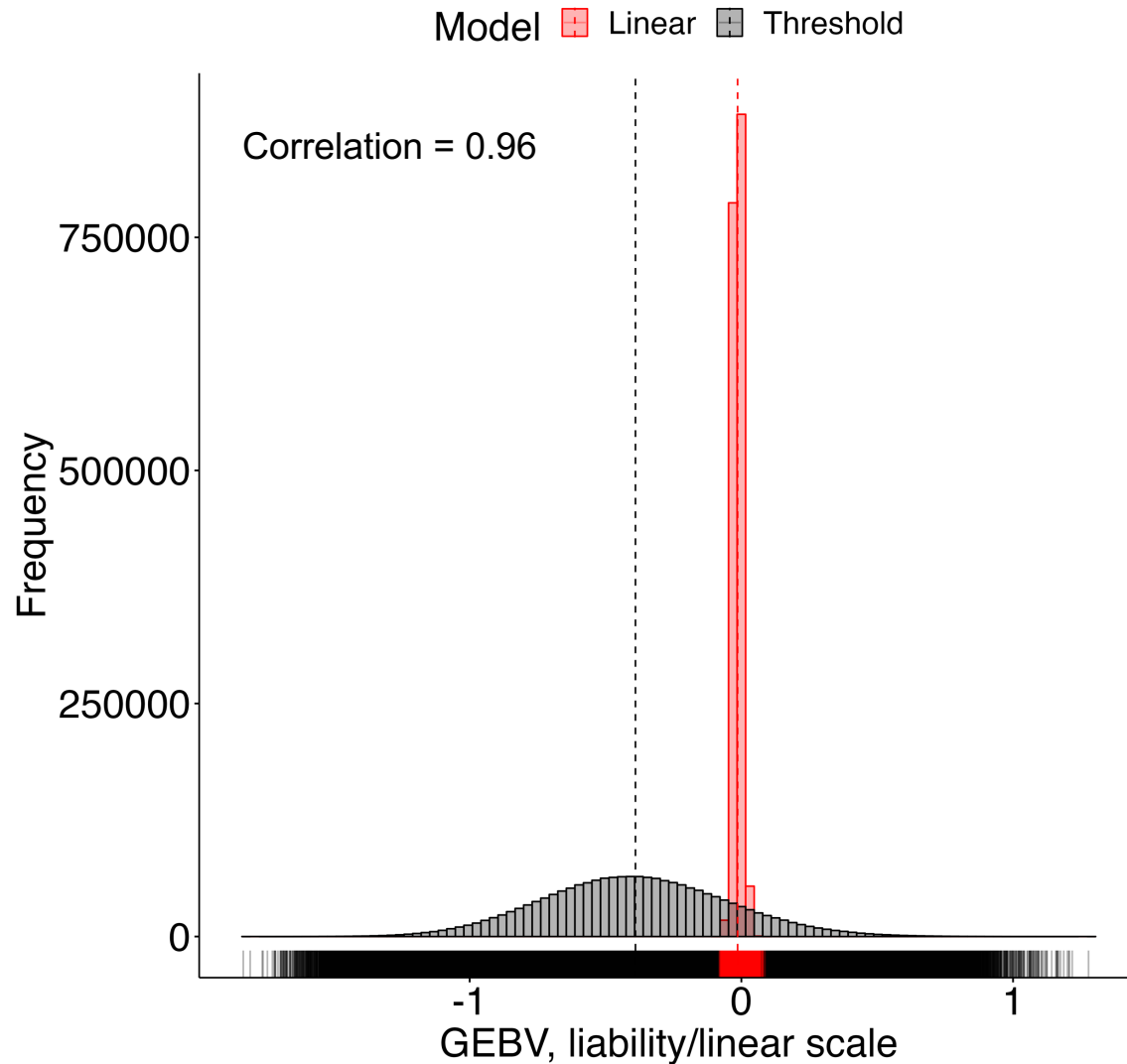# From linear to probability scale

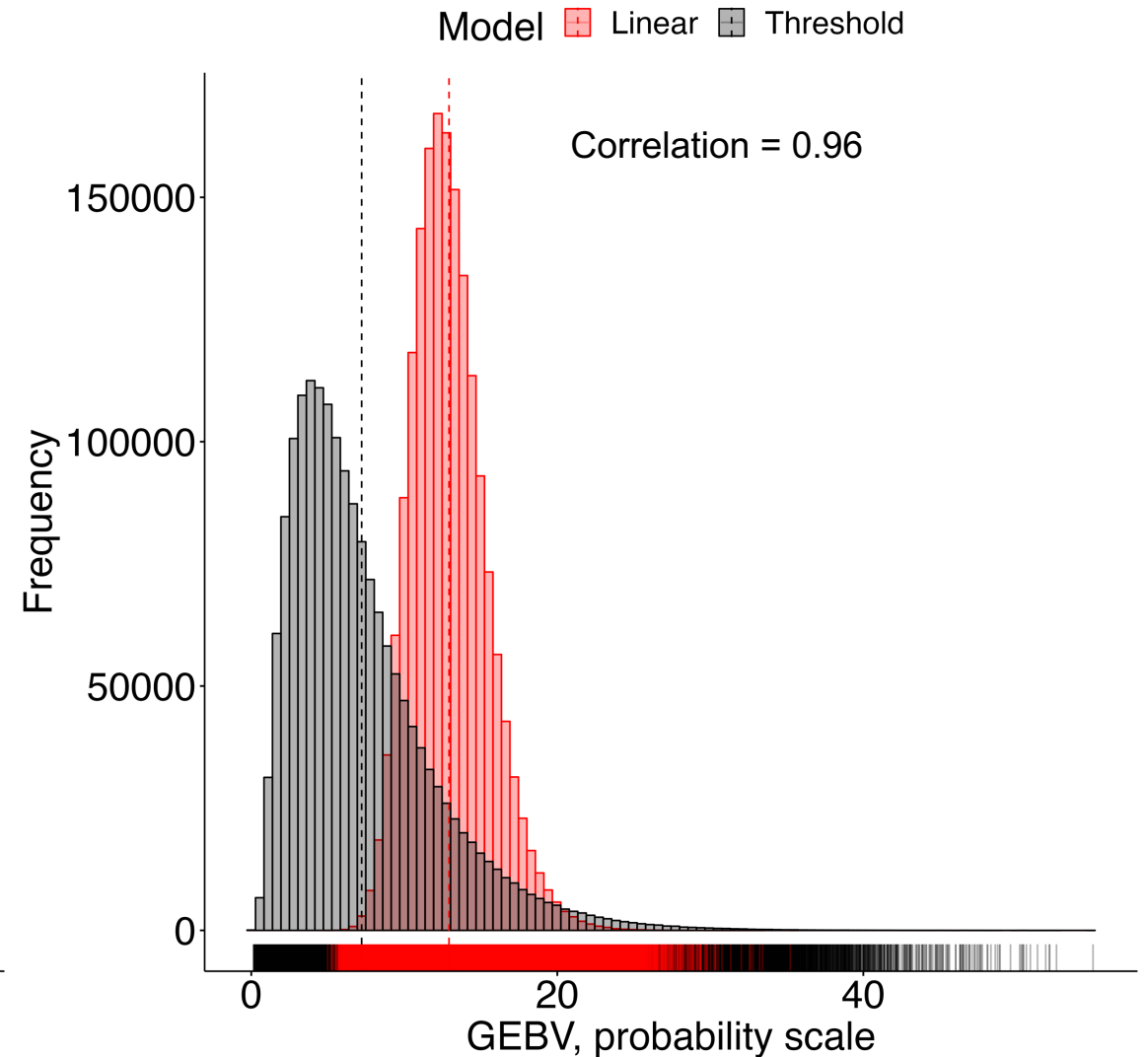Ketosis – incidence = 0.06

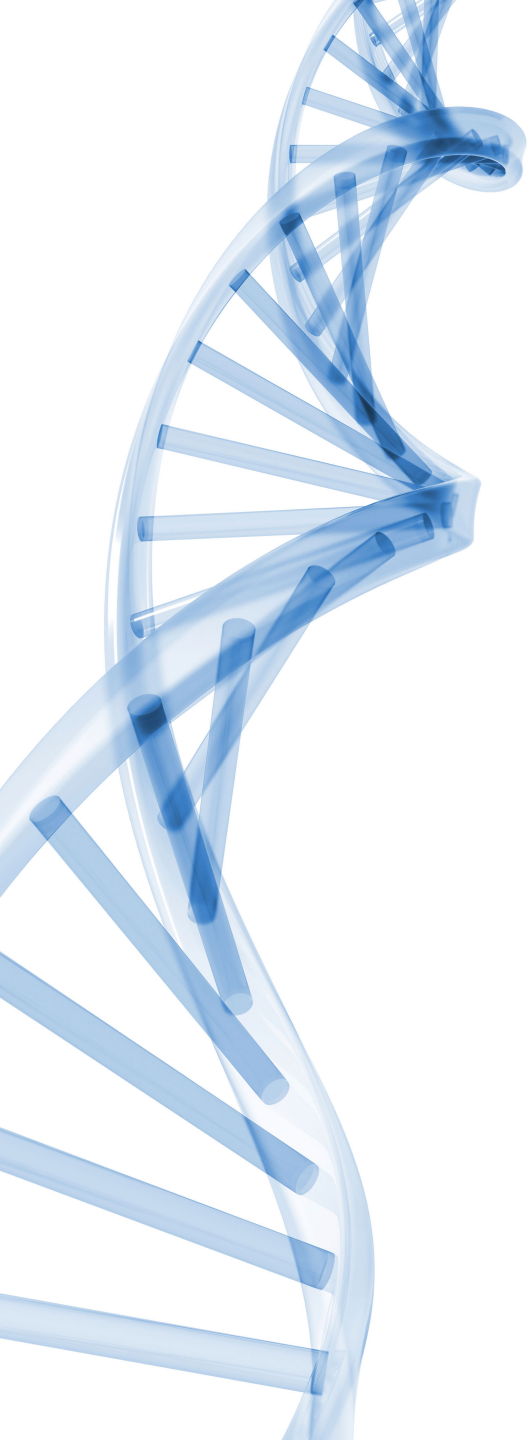# From linear to probability scale

Displaced abomasum – incidence = 0.02

Conclusions

- Linear models provide GEBV highly correlated with GEBV from threshold models

- Computing time can be reduced 3 – 5x

- The derived formula enables the use of linear models to obtain GEBV as probabilities

- Linear models may facilitate large-scale evaluations with multiple binary traits

- For traits with extreme incidence, the approximation provided a suboptimal match

  - Can it be improved by multi-trait models? Research in progress

Thank you!