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

How to validate genomic predictions?

Comparing genomic and non-genomic models

Daniela Lourenco
BLUPF90 TEAM – 08/2024

INIAV Course 2024

Daniela Lourenco
Fernando Bussiman

How to validate genomic predictions?

- We should look at accuracy!!!

Prediction (validation) Accuracy

Accuracy

Reliability

Prediction Accuracy

Prediction Reliability

Predictive Ability (predictability)

Theoretical Accuracy

Accuracy

Reliability

BIF Accuracy

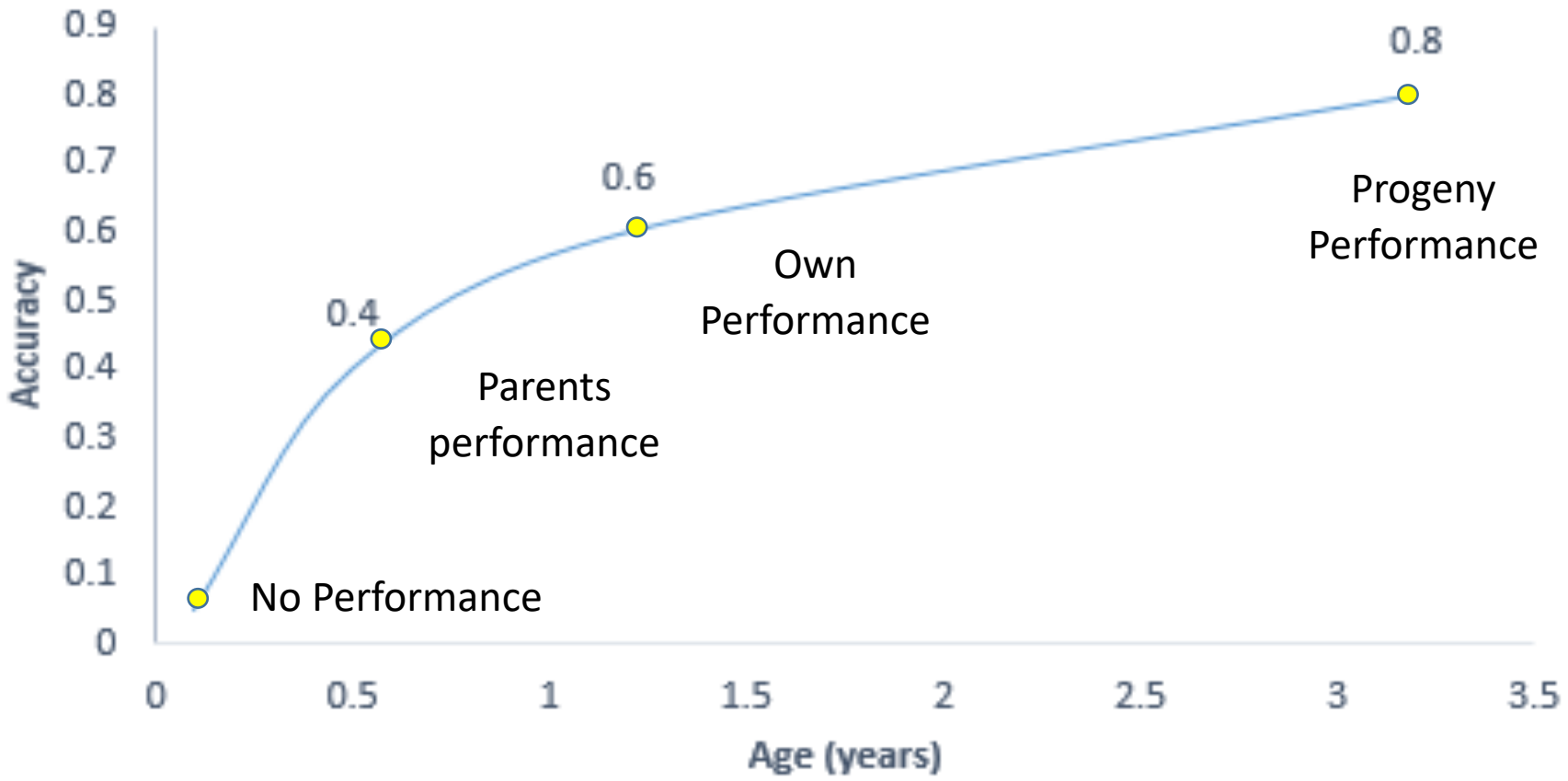
Understanding accuracy

- **Theoretical Accuracy**
  - Precision or stability of EBV
  - How much EBV changes when more data is added (risk)
  - SE of EBV
  - $Acc = \sqrt{1 - \frac{PEV}{\sigma_u^2}}$
  - $Acc_inb = \sqrt{1 - \frac{PEV}{\sigma_u^2(1+F)}}$
  - Individual
  - Model-based

- **Prediction Accuracy**
  - $\rho = COR(u, \hat{u})$
  - $\Delta G = i\ \rho\sigma_u/L$ (response to selection)
  - Potential ΔG of a breeding scheme
  - Very popular after genomics
  - Cross-validation
  - Population

Theoretical Accuracy of EBV

Theoretical Accuracy Increases as an more information is added



UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Accuracy and possible EBV change

BW EBV of 1.8	ACC	Possible Change
Bull A	.05	2.49
Bull B	.80	.53

Possible change = interval of 95% = $\text{EBV} \pm 1.96 \times \text{SE}$

Modified from: Dan Moser's GA talk - 2016

Accuracy and possible EPD change

Bull A, Acc. = .05

1.8 BW EBV ± 2.49

-0.69 to 4.29

Bull B, Acc. = .80

1.8 BW EBV ± .53

1.27 to 2.33

Which EBV will change most?

Which Bull has more reliable EBV?

Modified from: Dan Moser's GA talk - 2016

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

Variations of Theoretical Accuracy

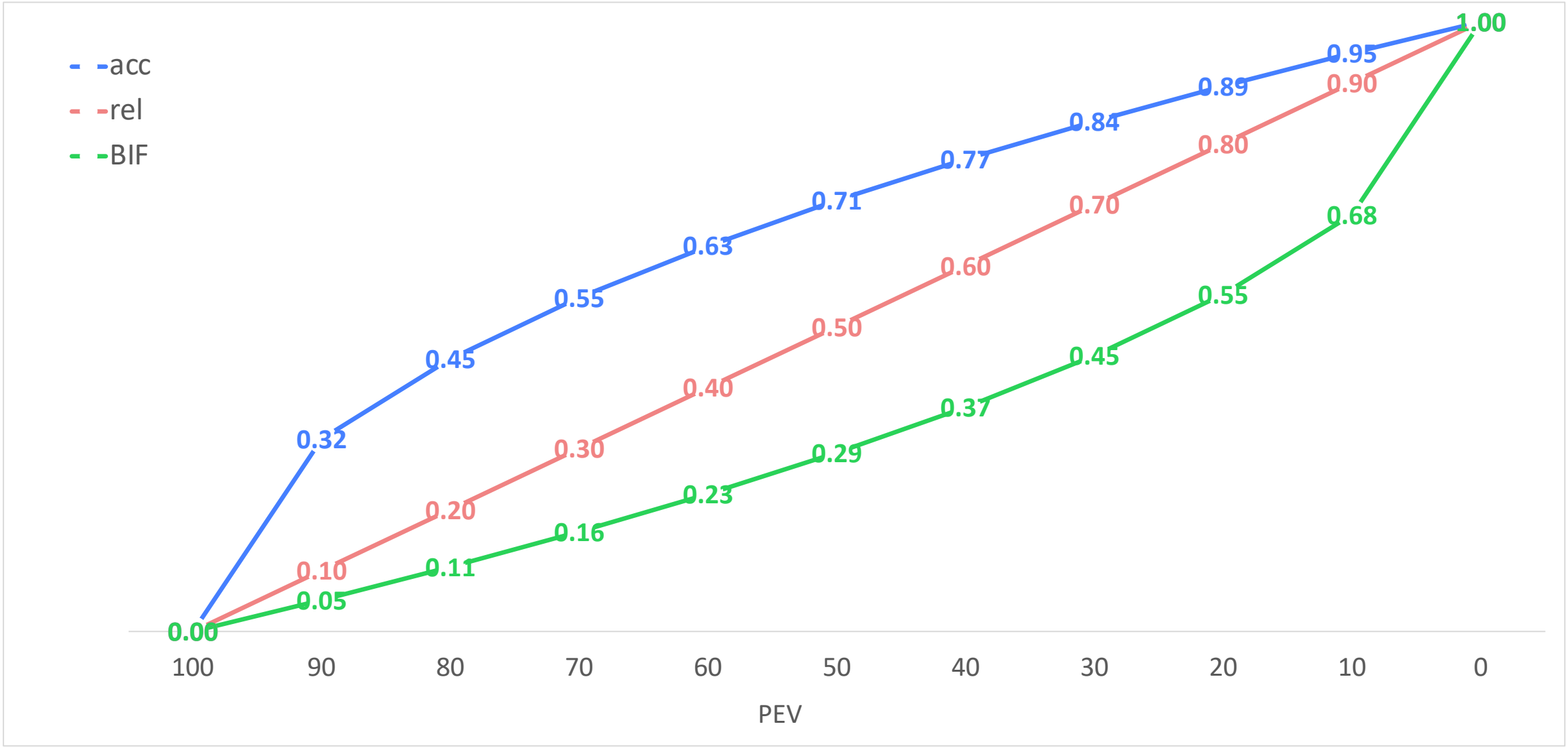
- Several: $Accuracy = \sqrt{1 - \frac{PEV}{\sigma_u^2(1+F)}}$ → Henderson (1975)
  Derivations under selection or not

- Beef cattle: $BIF\ Accuracy = 1 - \sqrt{\frac{PEV}{\sigma_u^2(1+F)}}$ → Lower values
  Approaches 1 very slowly
  Willham (~1985)

- Dairy cattle: $reliability = 1 - \frac{PEV}{\sigma_u^2(1+F)}$ → Lower values
  Approaches 1 more slowly
  Fraction of σ_u^2 accounted for by EBV
  VanRaden et al. (~1989)

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences

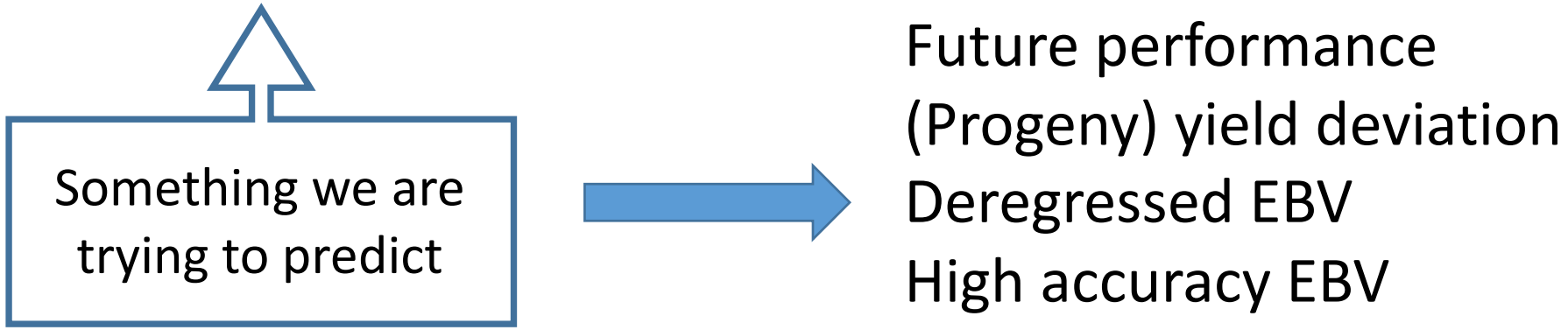
Variations of Theoretical Accuracy



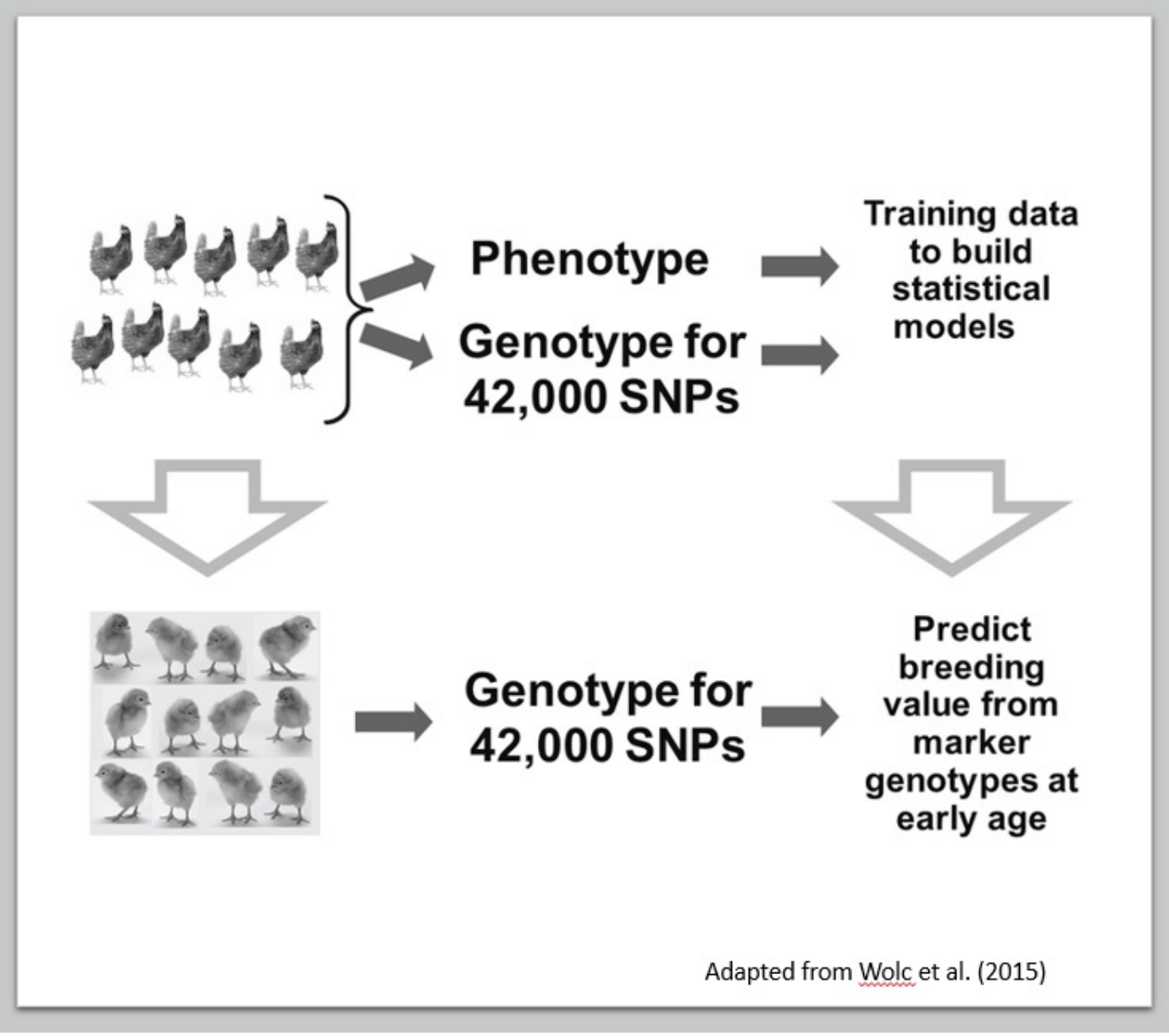
How to validate: Prediction accuracy

- Reflects the correlation between true and estimated breeding value
  - $accuracy = COR(u, \hat{u})$

- Do we have true breeding values in real populations?
  - $accuracy = COR(benchmark, \hat{u})$



Prediction accuracy

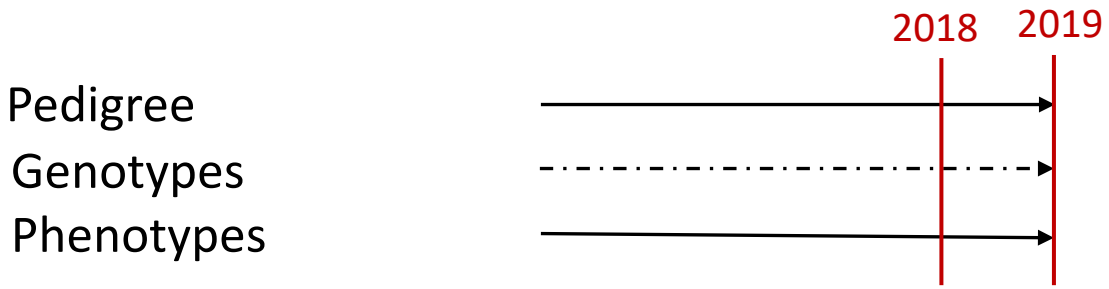


Training

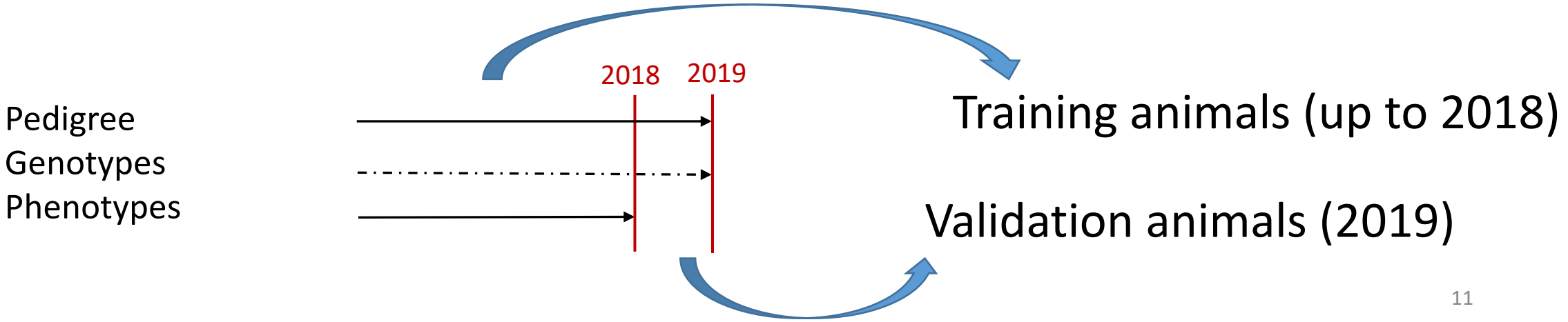
Validation

Prediction Accuracy

- Complete data (used to compute the benchmark)



- Reduced data (used to compute GEBV and EBV)



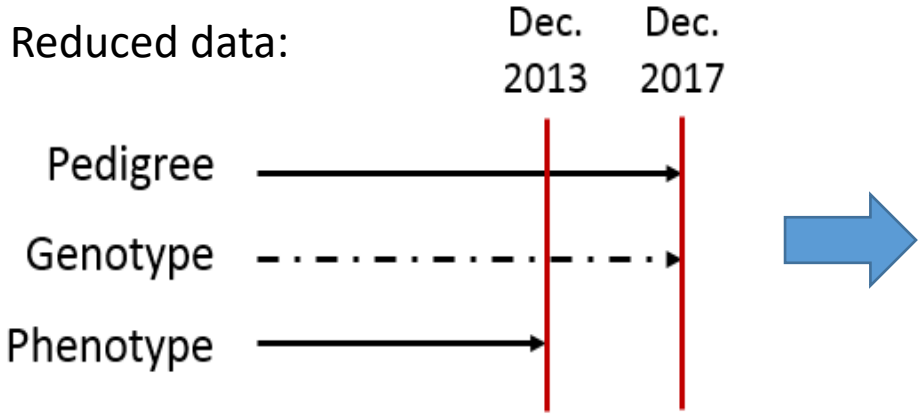
Which benchmark to use?

- $accuracy = COR(benchmark, \hat{u})$

Validation animals	Trait measured on	Benchmark
Dairy bulls	progeny	daughter yield deviation / deregressed proof

DYD or DEBV as benchmark

Reduced data:



- Remove 4 to 5 years of data
- Bulls with no daughter records in the reduced data
- Bulls have at least 10 daughters in the complete data
- At least 100 bulls

- Benchmark: Deregressed EBV (based on VanRaden et al., 2009)

$$\mathrm{DEBV}_{\mathrm{complete}} = \frac{\mathrm{EBV}_{\mathrm{complete}} - \mathrm{PA}_{\mathrm{complete}}}{\mathrm{R}_{\mathrm{complete}}} + \mathrm{PA}_{\mathrm{complete}}$$

$$\mathrm{DEBV}_{\mathrm{complete}} = b_0 + b_1(\mathrm{G})\mathrm{EBV}_{\mathrm{reduced}}$$

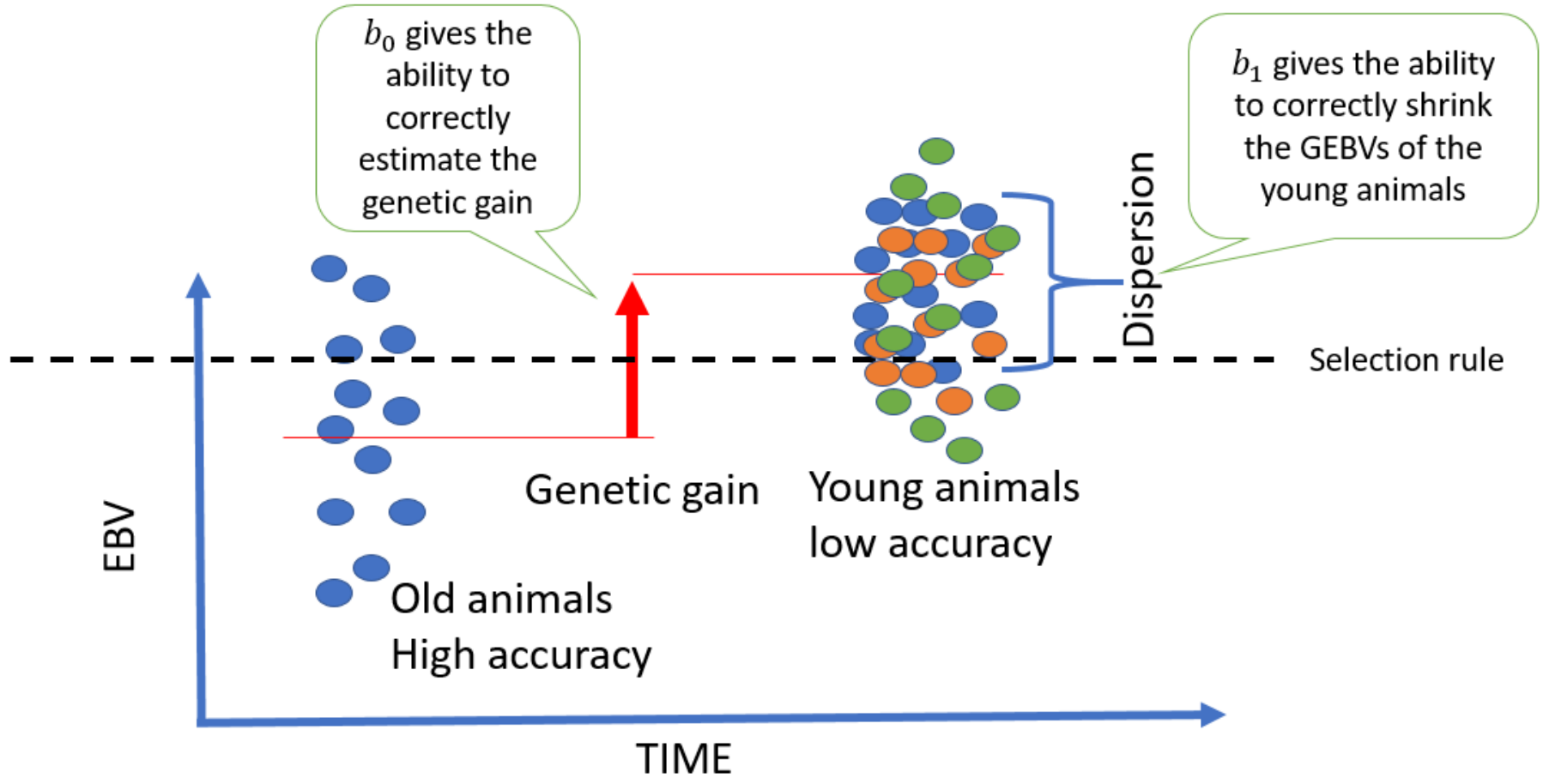
R^2 = prediction reliability

b_0 = bias

b_1 = dispersion

DYD or DEBV as benchmark

- Are bias and dispersion also important?

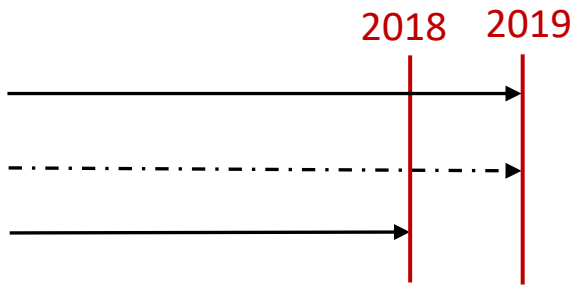


Adapted from Legarra and Reverter (2017)

Adjusted phenotypes as benchmark

Reduced data:

Pedigree
Genotypes
Phenotypes



- Remove 1 or 2 years of data
- Validation animals with own phenotypes in the complete data
- Phenotypes adjusted for fixed effects (complete data)

Predictivity or predictive ability of (G)EBV = Cor(Y_{adj}, $(G)EBV_{reduced}$)

$$\text{Accuracy} = \frac{\text{Predictivity}}{\text{sqrt}(h^2)}$$

$$Y_{adj} = b_0 + b_1(G)EBV_{reduced}$$

b_0 = bias

b_1 = dispersion

A new validation method

- LR Method
  - Linear Regression metrics
  - Legarra & Reverter (2018; GSE)

- Reduced (partial) and Complete (whole) data
- Validation animals have phenotypes in the *complete data* but not in the *reduced data*

- Benchmark: complete (G)EBV

- Compares EBV with EBV and GEBV with GEBV
  - Same scale

LR validation

- Accuracy

$$Accuracy = \sqrt{\frac{cov(\hat{\boldsymbol{u}}_c, \hat{\boldsymbol{u}}_r)}{\left(1 + \bar{F} - 2\bar{f}\right)\sigma^2_{u,\infty}}} \quad \text{or} \quad Accuracy = \sqrt{\frac{cov(\hat{\boldsymbol{u}}_c, \hat{\boldsymbol{u}}_r)}{(1 - \bar{F})\sigma^2_u}}$$

- Dispersion $\hat{\boldsymbol{u}}_c = \mathrm{b}_0 + \mathrm{b}_1\hat{\boldsymbol{u}}_r$
- Bias $\mu_{cr} = \overline{\hat{\boldsymbol{u}}_r} - \overline{\hat{\boldsymbol{u}}_c}$
- Consistency between subsequent evaluations $cor_{c,r} = cor(\hat{\boldsymbol{u}}_c, \hat{\boldsymbol{u}}_r)$
  - Estimator of the ratio of accuracies using the “reduced” or the “complete” data