

# Bias of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups.

- Evaluations are unbiased on average but with random variations across years: don't trust a single check of your evaluations
- With Unknown Parent Groups:
  - BLUP unbiased
  - Some SSGBLUP methods are biased

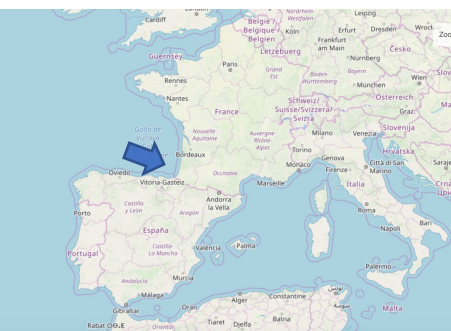


Fernando L. Macedo,  
Ole Christensen, Jean-Michel Astruc,  
Ignacio Aguilar, Yutaka Masuda,  
Andrés Legarra  
(your speaker)



# Why should I look at dairy sheep?

- Dairy sheep milk is a drop in an ocean of cow milk
- But we, dairy sheep geneticists, can do things that dairy cattle geneticists can not 😊
  - Simpler selection schemes
  - No import/export of animals, no Interbull
  - Human-size data sets (in the few millions of records)
  - Can try things quickly
- Some of our results might be of interest for the whole community of dairy geneticists



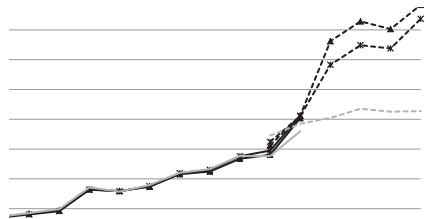
# Manech Tete Rousse breeding program

- Progeny testing , Artificial Insemination (AI)
- Steady genetic progress  $\sim 0.2\sigma_g/year$
- 35 years of pedigree and records
  - 60,000 records / year
  - 15,000 females / year
  - 200 elite (AI) males / year
- Genomic selection started in 2018
- We don't want overestimation of young animals' GEBVs (bias)
- All rams have very complete pedigrees, but
- 20% ewes have missing sires in pedigree due to Natural Mating
  - These unknown Natural Mating sires are actually offspring of AI rams
  - Use of Unknown Parent Groups in BLUP to model genetic trend
  - unclear how to include missing pedigrees in SSGBLUP



# Objectives

- Do we have bias in genetic (genomic) evaluations?
- How to best model missing pedigrees?
- The time-honored method to check bias is time-point truncation (Mantysaari et al., 2010, Olson et al., 2011)
- We want to check bias across multiple truncation time points
  - we don't want to use DYDs as we have small progeny groups: method LR
- Should we use...
  - 13 Unknown Parent Groups (fixed)
- or
  - 13 Metafounders (random with relationships)



```

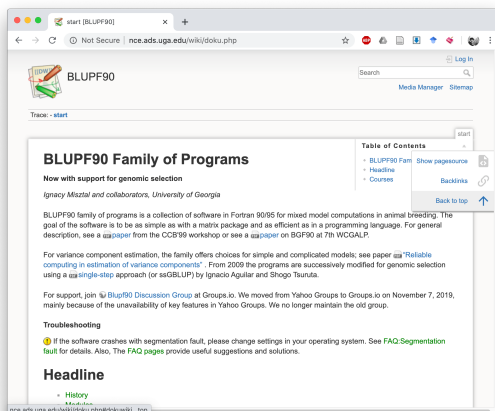
LONESPFBRES20000040030 0 LONESPFBRES14000030596
LONESPFBRES20000040037 LONESPFBRES14000031504 LONESPFBRES14000036619
LONESPFBRES20000040038 LONESPFBRES14000031494 LONESPFBRES14000036619
LONESPFBRES20000040056 LONESPFBRES14000032904 LONESPFBRES14000038530
LONESPFBRES20000040122 LONESPFBRES14000032904 LONESPFBRES14000038530
LONESPFBRES20000040124 LONESPFBRES14000034800 LONESPFBRES14000038529
LONESPFBRES20000033425 0 0
LONESPFBRES200000556572 0 0
LONESPFBRES200000556573 0 0
LONESPFBRES200000556575 0 0
LONESPFBRES200000556576 0 0
LONESPFBRES200000556577 0 0
LONESPFBRES200000556578 0 0
LONESPFBRES200000556579 0 0
LONESPFBRES200000556581 0 0
LONESPFBRES200000570001 0 0
LONESPFBRES200000570003 0 0
LONESPFBRES200000570008 0 0
LONESPFBRES200000570009 0 0
LONESPFBRES200000570010 0 0
LONESPFBRES200000570012 0 0
LONESPFBRES200000570013 0 0
LONESPFBRES200000570014 0 0
LONESPFBRES200000570015 0 0
LONESPFBRES200000570016 0 0
LONESPFBRES200000570017 0 0
LONESPFBRES200000570018 0 0
LONESPFBRES200000570019 0 0
LONESPFBRES200000570020 0 0
LONESPFBRES200000570021 0 0
LONESPFBRES200000570022 0 0
LONESPFBRES200000570023 0 0
LONESPFBRES200000570024 0 0
LONESPFBRES200000570025 0 0
LONESPFBRES200000570026 0 0
LONESPFBRES200000570027 0 0
LONESPFBRES200000570028 0 0
LONESPFBRES200000570029 0 0
LONESPFBRES200000570030 0 0
LONESPFBRES200000570031 0 0
LONESPFBRES200000570032 0 0
LONESPFBRES200000570033 0 0
LONESPFBRES200000570034 0 0
LONESPFBRES200000570035 0 0
LONESPFBRES200000570036 0 0
LONESPFBRES200000570037 0 0
LONESPFBRES200000570038 0 0
LONESPFBRES200000570039 0 0
LONESPFBRES200000570040 0 0
LONESPFBRES200000570041 0 0
LONESPFBRES200000570042 0 0
LONESPFBRES200000570043 0 0
LONESPFBRES200000570044 0 0
LONESPFBRES200000570045 0 0
LONESPFBRES200000570046 0 0
LONESPFBRES200000570047 0 0
LONESPFBRES200000570048 0 0
LONESPFBRES200000570049 0 0
LONESPFBRES200000570050 0 0
LONESPFBRES200000570051 0 0
LONESPFBRES200000570052 0 0
LONESPFBRES200000570053 0 0
LONESPFBRES200000570054 0 0
LONESPFBRES200000570055 0 0
LONESPFBRES200000570056 0 0
LONESPFBRES200000570057 0 0
LONESPFBRES200000570058 0 0
LONESPFBRES200000570059 0 0
LONESPFBRES200000570060 0 0
LONESPFBRES200000570061 0 0
LONESPFBRES200000570062 0 0
LONESPFBRES200000570063 0 0
LONESPFBRES200000570064 0 0
LONESPFBRES200000570065 0 0
LONESPFBRES200000570066 0 0
LONESPFBRES200000570067 0 0
LONESPFBRES200000570068 0 0
LONESPFBRES200000570069 0 0
LONESPFBRES200000570070 0 0
LONESPFBRES200000570071 0 0
LONESPFBRES200000570072 0 0
LONESPFBRES200000570073 0 0
LONESPFBRES200000570074 0 0
LONESPFBRES200000570075 0 0
LONESPFBRES200000570076 0 0
LONESPFBRES200000570077 0 0
LONESPFBRES200000570078 0 0
LONESPFBRES200000570079 0 0
LONESPFBRES200000570080 0 0
LONESPFBRES200000570081 0 0
LONESPFBRES200000570082 0 0
LONESPFBRES200000570083 0 0
LONESPFBRES200000570084 0 0
LONESPFBRES200000570085 0 0
LONESPFBRES200000570086 0 0
LONESPFBRES200000570087 0 0
LONESPFBRES200000570088 0 0
LONESPFBRES200000570089 0 0
LONESPFBRES200000570090 0 0
LONESPFBRES200000570091 0 0
LONESPFBRES200000570092 0 0
LONESPFBRES200000570093 0 0
LONESPFBRES200000570094 0 0
LONESPFBRES200000570095 0 0

```



# Data & records

- 2 M records of milk yield in the complete data set
- 500,000 animals in pedigree
- 3,000 genotyped AI males (all progeny-tested males since 2000) with the 50K Illumina ovine chip
- (SSG)BLUP evaluations for milk yield
  - animal model with repeated records and correction for heterogeneity of variances (Meuwissen et al., 1996)
- Software `heterf90`, `blup90iod2` (Misztal et al.)



# Method LR in a nutshell

- Define a “focal set” of “comparable” animals of interest
  - e.g. contemporary young males born in 2010 or young females born in 2008
- Define evaluations “partial data” (early) and “whole data” (late). From “partial” to “whole”:
- Averages of EBVs should not change (i.e. some young animals will go up and some down)
- Dispersion of EBVs should increase by the right amount (more information -> more dispersion)
  - Slope (EBVs<sub>whole</sub> ~ EBVs<sub>partial</sub>) should be 1
- Corr(EBVs<sub>whole</sub>, EBVs<sub>partial</sub>) is a function of respective accuracies

Legarra and Reverter *Genet Sel Evol* (2018) 50:53  
<https://doi.org/10.1186/s12711-018-0426-6>



RESEARCH ARTICLE

Open Access



Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method

Andres Legarra<sup>1\*</sup> and Antonio Reverter<sup>2</sup>



J. Dairy Sci. 103

<https://doi.org/10.3168/jds.2019-16603>

© 2020, The Authors. Published by FASS Inc. and Elsevier Inc. on behalf of the American Dairy Science Association®.  
This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models**

F. L. Macedo,<sup>1,2\*</sup> A. Reverter,<sup>3</sup> and A. Legarra<sup>1</sup>

# Application of method LR

Evaluations with data until 2005, until 2006 and so on until 2017.

We compare \*within-model\*

- (G)EBVs at birth of a set (~200) of future AI males
- (G)EBVs of the same males in later evaluations (after having progeny).

For example for males born in 2005, 11 pairs of evaluations were compared

- 2005 vs 2007
- 2005 vs 2008
- ...
- 2005 vs 2017

The same for males born in 2006, 2007 ... 2015

- 2006 vs 2008
- ...
- 2015 vs 2017

Total of 66 comparisons that were then “averaged” using a pseudo-model

## Bias

$$\hat{\Delta}_p = \bar{\hat{u}}_{partial} - \bar{\hat{u}}_{whole}$$

Expected value of 0 in absence of bias.

## Slope of the regression $EBV_w$ on $EBV_p$

$$\hat{b}_p = \frac{cov(\hat{u}_{partial}, \hat{u}_{whole})}{var(\hat{u}_{partial})}$$

With a value of 1 in unbiased procedure.

# Models for UPG/metafounders

## 1. BLUP\_MF (metafounders)

$$A^{[\Gamma]-1} = \begin{pmatrix} A_{anim,anim}^{[\Gamma]} & A_{anim,mf}^{[\Gamma]} \\ A_{mf,anim}^{[\Gamma]} & \Gamma \end{pmatrix}^{-1}$$

Legarra et al. 2015

## 2. BLUP\_UPG

$$A^* = \begin{pmatrix} A_{anim,anim}^{-1} & -A_{anim,anim}^{-1}Q \\ -Q'A_{anim,anim}^{-1} & Q'A_{anim,anim}^{-1}Q \end{pmatrix}$$

Thompson 1979

Quaas 1988

## 3. ssGBLUP\_EUPG ("exact"\_UPG)

$$H^{(exact)-1} = A^* + \begin{pmatrix} 0 & 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} & -(G^{-1} - A_{22}^{-1})Q_2 \\ 0 & -Q_2(G^{-1} - A_{22}^{-1}) & Q_2(G^{-1} - A_{22}^{-1})Q_2 \end{pmatrix}$$

Misztal et al. 2013

Matilainen et al 2018

## 4. ssGBLUP\_MF (metafounders)

$$H^{[\Gamma]-1} = A^{[\Gamma]-1} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_{05}^{-1} - A_{22}^{[\Gamma]-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Legarra et al. 2015

## 5. ssGBLUP\_UPG

$$H^* = A^* + \begin{pmatrix} 0 & 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

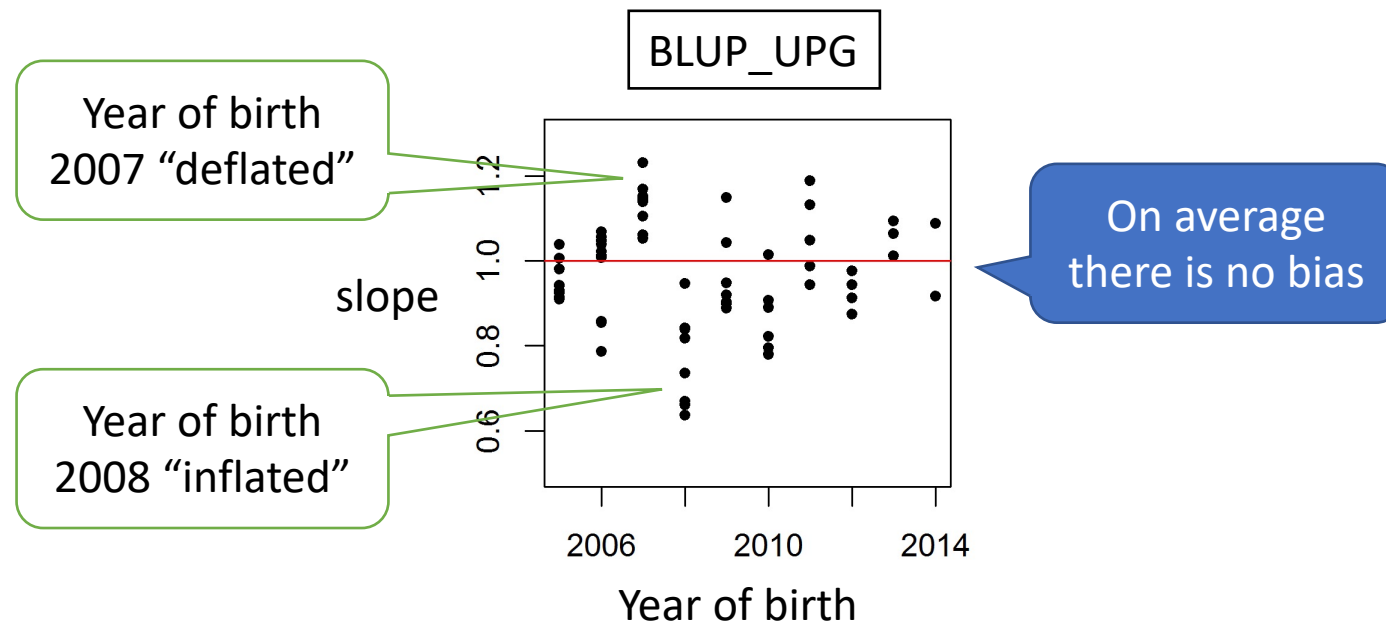
Default blupf90

\*  $\Gamma$  was computed by GLS from all genotypes

# Results 1: more than one check

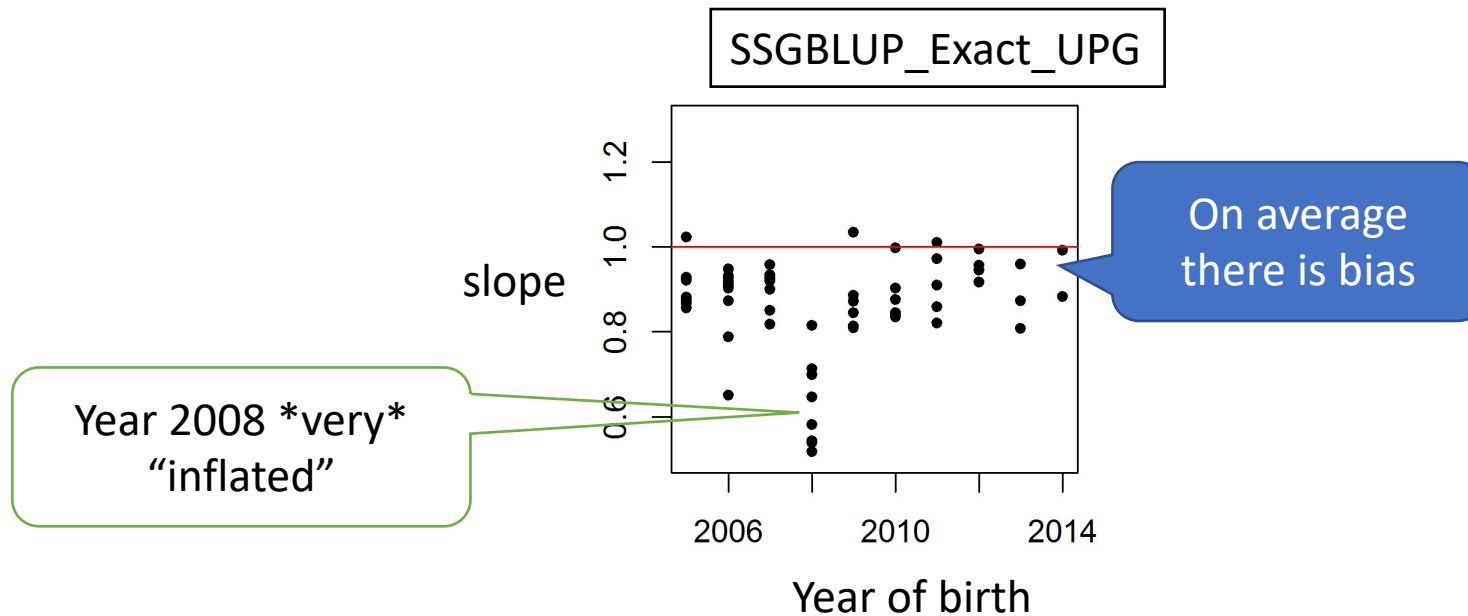
Use of method LR allows many, easy, systematic checks of bias

- High variability due to year of birth (“partial”) and year of “whole” evaluation
- We should use several comparisons to decide the rightness of the model



# Results 2: some methods are better

Some methods (SSGBLUP\_Exact\_UPG) look wrong even compared to themselves (they are incoherent from one year to the next ones)



# Results 3: all methods together

Model	$\bar{u}_p - \bar{u}_w$ Bias	$\hat{u}_w \sim \hat{u}_p$ Slope	Accuracies		
			$\left(\frac{\widehat{acc_p}}{acc_w}\right)$	$\widehat{acc_p^2}$	$\left(\frac{\widehat{acc_p^2}}{acc_w^2}\right)$
BLUP-MF	0.21	0.99	0.56	0.22	0.33
BLUP-UPG	0.36	0.97	0.55	0.24	0.31
SSGBLUP-EUPG	0.39	0.87	0.62	NA	0.44
SSGBLUP-MF	0.20	0.98	0.67	0.33	0.46
SSGBLUP-UPG	0.27	0.95	0.65	NA	0.44

- There is some small bias in the genetic trend of  $\sim 0.25\sigma_g$
- BLUPs are generally unbiased
- SSGBLUP more accurate than BLUP
- SSGBLUP is better with metafounders (less bias, more accurate)

\* All estimators have s.e. <0.02

# Conclusions

- Run more than one test !!!
- Method LR allows quick and automatic checks
- \*In our data set\*, no bias if we do things correctly
  - BLUP\_UPG and BLUP\_MF were unbiased
  - SSGBLUP\_MF was unbiased
  - Good compatibility of marker and pedigree information while avoiding double counting

Fernando L. Macedo<sup>12\*</sup>, Ole Christensen<sup>3</sup>, Jean-Michel Astruc<sup>4</sup>,  
Ignacio Aguilar<sup>5</sup>, Yutaka Masuda<sup>6</sup>, Andrés Legarra<sup>1</sup>



INRAE



<sup>1</sup> GenPhySE, INRAE, 31326 Castanet Tolosan, France.

<sup>2</sup> Facultad de Veterinaria. UdelaR. A. Lasplaces 1620. Montevideo. Uruguay.

<sup>3</sup> Center for Quantitative Genetics and Genomics. Blichers Allé 20. 8830 Tjele. Denmark.

<sup>4</sup> Institut de l'Elevage, CS52627, 31326 Castanet Tolosan, France

<sup>5</sup> Instituto Nacional de Investigación Agropecuaria. Las Brujas, Uruguay

<sup>6</sup> Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA

This work has received funding from the project ARDI funded by  
INTERREG POCTEFA, the European Unions' Horizon 2020 Research &  
Innovation program under grant agreement N°772787



# Conclusions

*Thank you for your attention*

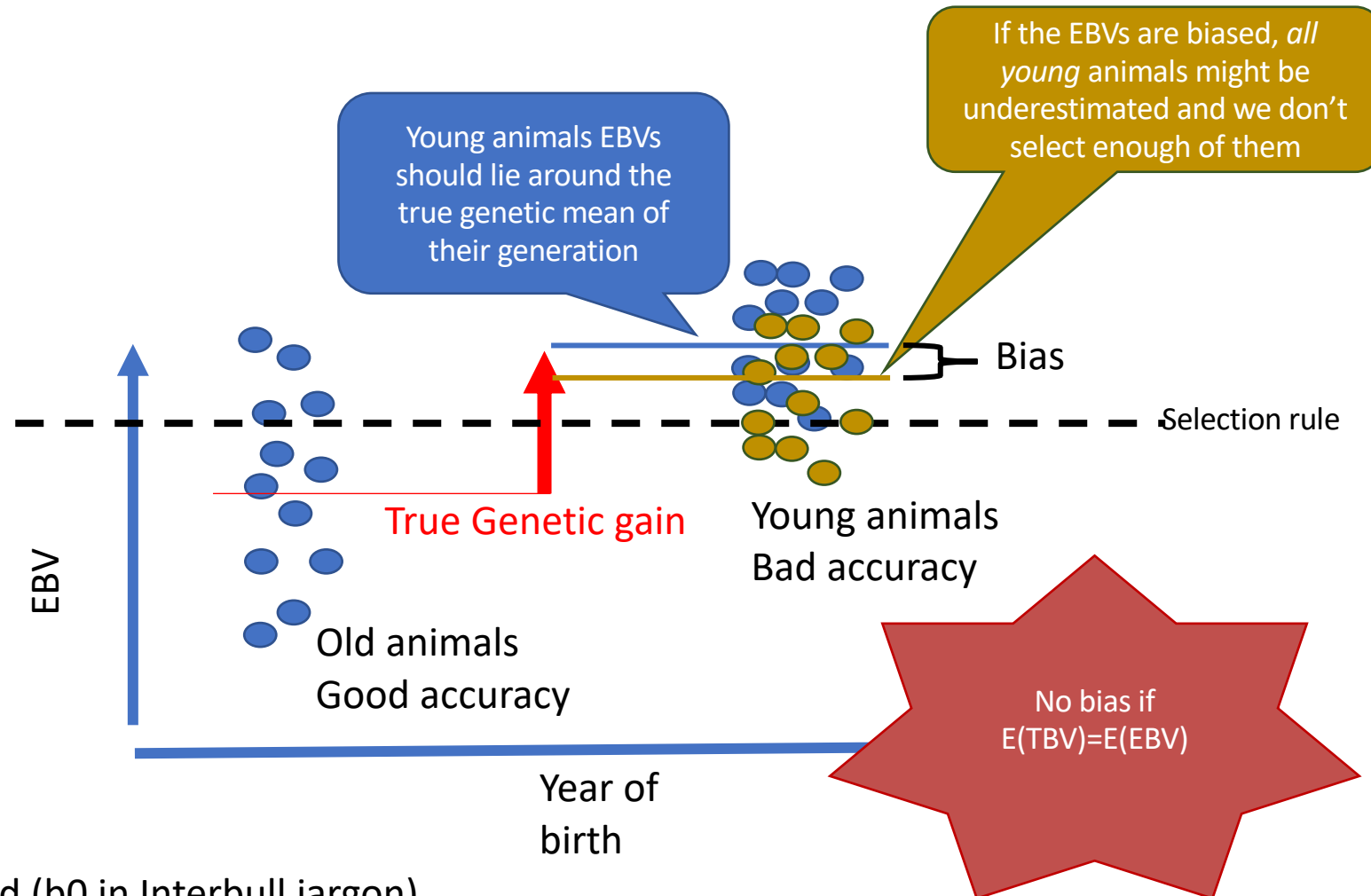
- Run more than one test !!!
- Method LR allows quick and automatic checks
- \*In our data set\*, no bias if we do things correctly
  - BLUP\_UPG and BLUP\_MF were unbiased
  - SSGBLUP\_MF was unbiased
  - Good compatibility of marker and pedigree information while avoiding double counting



# Leftovers

# Bias due to genetic trend

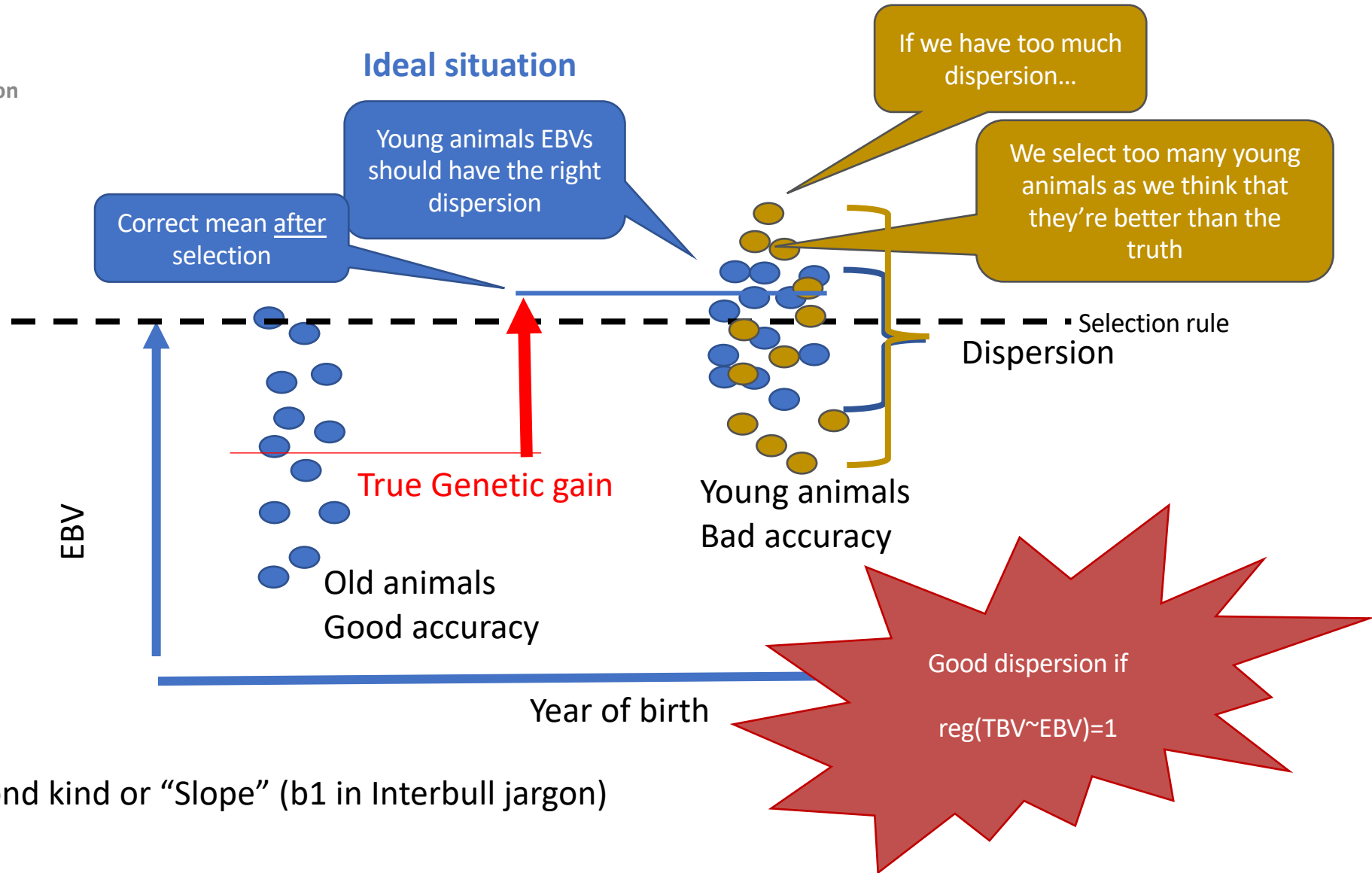
Consider a Genetic Evaluation



Bias of the first kind (b0 in Interbull jargon)

# Bias due to dispersion (slope)

Consider a Genetic Evaluation



Bias of the second kind or "Slope" ( $b_1$  in Interbull jargon)

# Checks for bias

- Traditional method
  - Compare DYDs after daughters with (G)EBVs before daughters
  - Not always possible:
    - Small data sets: RFI, methane emissions, small populations (sheep & goat but also small breeds)
    - Traits such as maternal ability in beef, carcass yield in pigs
    - Hard to make automatic
- Hence method LR:
  - comparison of old EBVs ( $\hat{u}_p$ ) with new EBVs ( $\hat{u}_w$ ) for a group of animals of interest

Legarra and Reverter *Genet Sel Evol* (2018) 50:53  
<https://doi.org/10.1186/s12711-018-0426-6>



RESEARCH ARTICLE

Open Access



Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method

Andrés Legarra<sup>1</sup> and Antonio Reverter<sup>2</sup>



J. Dairy Sci. 103

<https://doi.org/10.3168/jds.2019-16603>

© 2020, The Authors. Published by FASS Inc. and Elsevier Inc. on behalf of the American Dairy Science Association®. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models**

F. L. Macedo,<sup>1,2\*</sup> A. Reverter,<sup>3</sup> and A. Legarra<sup>1</sup>

# Method LR in a nutshell

- We can use these ideas in reverse:
  - If the mean(EBV) does change there is a problem
  - If the dispersion does not hold to the theory there is a problem
  - either Henderson's theory was wrong (maybe) or our genetic evaluations are wrong (likely)
- Also, average change indicates reliabilities