NATURAL RESOURCES
INSTITUTE FINLAND

Usability of different genetic evaluation validation tests in a population subjected to a strong genomic selection and in testing the single-step genomic evaluations

Esa A. Mäntysaari & Andrei A. Kudinov
Natural Resources Institute Finland (Luke)
Animal Genetics and Breeding

Background

Genomic selection has changed the applicability of (Interbull) validation tests

- the assumptions of the tests might not be valid and thereafter the value of tests are questionable

Interbull Technical Committee working group ”Validation tests”

- Esa Mäntysaari, Zengting Liu, Peter Sullivan, Raphael Mrode, Paul VanRaden, Valentina Palucci

In the present study

We applied Interbull validation tests on Nordic RDC data
in which the genomic selection has been now used >10 years

Applied tests were:

- Interbull validation tests II and III
- Validation using prediction of deRegressed GEBVs (VanRaden, 2021)

Data and Models

Unofficial
Data and Model !

Data and Traits:

- 9 trait model for 305d lactation yields
- Multi-trait (milk, protein, fat), Multi-lactation (1,2,3)
- All Nordic RDC and Finnish HOL animals,

Recorded cows:	4.3 M
Pedigree:	5.9 M
Genotypes:	169 t

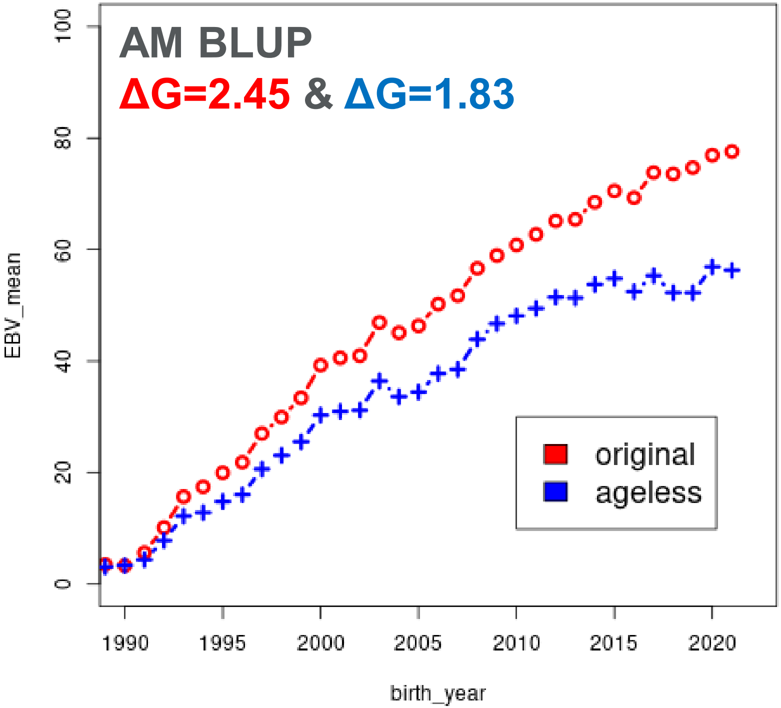
TWO MODELS:

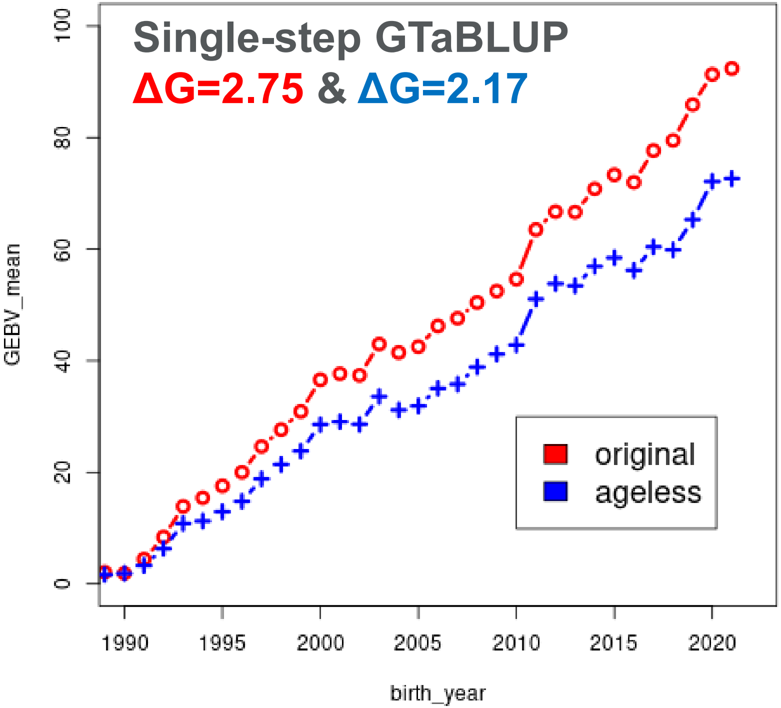
Control model vs. **Model without calving age**

Genetic trends Nordic RDC AI bulls (Protein 305d yield)

Two models:

- **Control** vs. a **model without calving age** : **(ΔG bias ~ -0.6)**





Validation Test II (Boichard, et al. 1995.)

Rationale:

Daughters of a same bull should perform equally during the data time span bull is used – year after year

The performance is measured by Daughter Yield Deviations (DYD)

Action: Bull yearly DYD are tested with a fixed effect model:

$$Y_{ij} = BULL_i + b*\mathbf{j} + e_{ij} \qquad [1]$$

where Y_{ij} is the DYD in the j^{th} year of the i^{th} bull; **j**=0 for the first year 10 daughters **born**;

Test II results for protein

Model	DYDij=BULLi+**b***j	Testval = abs(b)/σ_{bv}
AM BLUP	-0.11	0.11/26.53 = 0.004
AM BLUP_ageless	-0.43	**0.43/26.53 = 0.016**
ssGTBLUP	-0.09	0.09/26.53 = 0.003
ssGTBLUP_ageless	-0.39	**0.39/26.53 = 0.015**

Number of bulls = 2768

Control models PASS the test
Models without calving age FAIL the test !

Same conclusion for
Genetic and Genomic Evaluations

Conclusions: Validation test II

- Test detected the reduction on ΔG caused by model modification fairly well (70%)
- Validation test is usable also for the single-step
    - However, estimation of DYD requires suitable program

Estimation of DYD for the basic models

Usually DYD are calculated from the observations and fixed effect solutions

Alternatively, they could be calculated using the pedigree file and EBV solutions

As

sire means of $(\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{A}^{-1}\lambda)\widehat{\boldsymbol{EBV}} - 0.5 * \boldsymbol{Z}_{dam}\widehat{EBV}$ over daughters w. records

This can be easily accumulated by reading the pedigre file and the solution file

Validation Test III (Boichard, et al. 1995.)

Rationale:

The bull EBV should remain stable over time (t), both when

- calculated from the first crop daughters (EBV_{red}),
- or from daughters included at latter (EBV_{full})

Action:

- Perform two evaluations:

Current (=**full**) and **red**uced (current – 4 years of data)

Analyze:

$$EBVfull_i = a + b * EBVredi + \delta \boldsymbol{t} + ei$$

If the regression coefficient δ >0, the EBVs have on average increased and vice verse

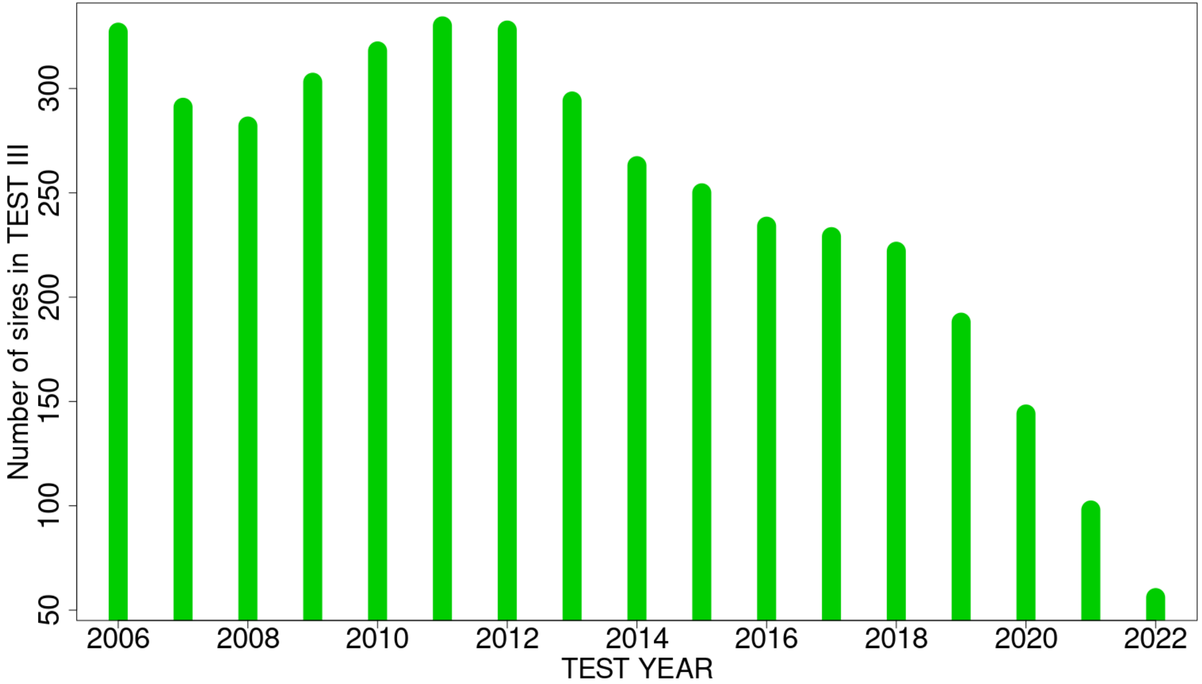
Validation Test III

Note: Test III is based on change in bull EBV from the first crop to second crop

QUESTION:

- Can Test III be used if the breeding program is based on only young genomic selected sires?

Number of progeny tested sires in validation test III in each year



For example:
Year 2022:

- Only 54 bulls have second crop daughters during 2016-2019

Test III results for protein

Model	**N bulls**	**t** **Mean (SD)**	**b*t (SD)**	**T**	**b/σ_{bv}**
AM BLUP	54	1.16 (0.18)	-0.367 (0.438)	0.8	-0.014
AM_ageless	same	same	-0.582 (0.452)	1.3	-0.022
ssGTBLUP			-0.593 (0.481)	1.2	-0.022
ssGTBLUP_ageless			-0.875 (0.483)	1.8	-0.033

All models are PASSing the test!
- all pass because the changes are not concluded to be statistically significant !

GEBV validation
using prediction of deregressed GEBV (dGEBV)

- VanRaden, P. M. 2021. Improved genomic validation including extra regressions Interbull Bulletin No. 56 (2021): https://journal.interbull.org/index.php/ib/article/view/82

Tests the evaluation stability (and relative accuracy)
by modeling "changes" in predictions.

Similar to the Interbull Validation test III, the test requires 2 evaluations:

- Current evaluations (*full*) and evaluations 4 years ago (reduced, *red*)

- The reduced data evaluations can require *base adjustment* to assure the comparability of the two evaluations

Define:

$$\mathrm{d}GEBV = GEBV_{red} + (GEBV_{full} - GEBV_{red})/Rdiff$$

where the *Rdiff* is the increase in reliability from R^2_{red} to R^2_{full}

For EBVs:

$$dEBV = PA + (EBV_{full} - PA)/Rdiff$$

Motivation:

- For the pedigree BLUP models, the dEBV resembles approximated DRP from the simple de-regression

- For the genomic evaluations, the dGEBV reflects changes in
  - the information due to animals' own records,
  - records of relatives,
  - increased information in reference population

Action

De-regressed predictions (dGEBV) are modelled using three models:

Bias test: $dGEBV_i = a_0$

Inflation test: $dGEBV_i = a_0 + b_1 (G)\widehat{EBV}_{red, i}$

Trend test: $\mathrm{d}GEBV_i = a_0 + b_1 (G)\widehat{EBV}_{red, i} + b_2 X_i$

Bias and Trend tests are fitted using the weight w = Rdiff.

In the Trend test the regression $b_2 X_i$ examines the effect of the birth year of the bull, or birth years of the daughters

Expectations: $E[a_0] = 0.0$, $E[b_1]=1.0$ and $E[b_2] = 0.0$

Validation cohorts:

- Young bulls
  with no daughters in (G)EBV$_{red}$
  but >20 daughters in >10 herds in (G)EBV$_{full}$

- Progeny tested bulls
  with first crop daughters in reduced data
  and receiving the second crop daughters in full data

Single-step evaluation (protein) dGEBV validation test

Significance Indicated by color
T > 3, **T > 5**

Test $b_1=1$

Test $b_2=0$

	b_1 (EBV_{red})	**std(b_1)**	**T**	**b_2 (year)**	**std(b_2)**	**T**	**R^2**
			$b_1=1$			$b_2=0$	
SS model							
Inflation test	**0.60**	0.04	-11.0				0.48
Trend test	**0.75**	0.04	-6.1	**-4.56**	0.6	-6.0	0.53
SS model without age effect in model							
Inflation test	**0.56**	0.04	-11.2				0.39
Trend test	**0.75**	0.04	-5.8	**-6.23**	0.8	-8.0	0.50

320 validation bulls have EDC== 0. in reduced data, and EDC>20 in full data

Animal model (protein validation)
dEBV validation test

Test $b_1=1$

Test $b_2=0$

	b_1 (EBV_{red})	std(b_1)	T	b_2 (year)	std(b_2)	T	R^2
			$b_1=1$			$b_2=0$	
AM model							
Model (2)	**0.66**	0.06	-5.4				0.27
Model (3)	0.70	0.07	-4.6	-1.13	0.84	-1.3	0.28
AM model without age effect in model							
Model (2)	**0.61**	0.07	-5.8				0.22
Model (3)	0.71	0.07	-4.2	-3.56	0.85	-4.2	0.26

Conclusions: dGEBV validation

- dGEBV validation indicated over dispersion in all models
- All single-step models indicated over prediction of genetic trend (in reduced data)
- dGEBV prediction tests the prediction accuracy
  - The predictive power (R^2) values of GEBV was almost double to classical PA (0.27 vs. 0.48)

Final conclusions

Interbull tests II and III are

- originally designed for genetic evaluations, but they do work with genomic evaluations
- **Test II** was able to detect the model problems
    - Usable test as long as the DYD are can be obtained
- **Test III** did not yield statistically significant statistics (because of only 54 bulls in test)

- The **dGEBV prediction test** could replace the current ITB GEBV validation test.

Thank you!

Acknowledgement:

We thank NAV for making
the RDC data available for the tests.

NATURAL RESOURCES
INSTITUTE FINLAND