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*[1] *and Andrei A. Kudinov*[1]**

[1] *Genetics and Breeding, Natural Resources Institute Finland (Luke)*

Abstract

Interbull trend validation tests II and III and the dGEBV prediction test were applied on genetic and genomic evaluations of production traits in Nordic Red Dairy cattle evaluations. For all tests, both the animal model and the single-step GTBLUP evaluations were applied to a control model and to a faulty model where the calving age effect was removed. The trend test II was able to detect faulty models, while the test III had insufficient power because of only few bulls returning to service. Both trend validation tests seemed to work equally well for genetic and genomic evaluations. The dGEBV prediction test indicated problems in both genomic evaluation models, although results for inflation and bias were worse with the faulty model. Given the good properties of test II, an alternative approach was proposed for the calculation of daughter yield deviations. The approach is based implicitly on MME equations and requires only the estimates of breeding values and pedigree.

Key words: Genetic evaluation validation, Genetic trend, deregressed GEBV, single step GBLUP

Introduction

Genetic evaluation validation tests have multiple purposes. They are used in development of national evaluations, they are certificate of trustworthy indices, and an accepted validation result is required for national data to be submitted to the international evaluations offered by Interbull (MACE and GMACE). The pedigree based EBVs (AM BLUP) have to pass two out of three Interbull trend validation tests (ITB tests I-III) and the genomic evaluations (GBLUP) have to pass ITB GEBV validation test. In addition, for AM BLUP evaluations, countries are asked to run the Mendelian Sampling variance test.

The genomic selection has changed the usability of validation tests. For example, the ITB test II addresses differences in daughter yield deviations (DYD) of bulls first and second daughter crops, and the test III on changes in EBVs of bulls due to the second crop (Boichard et al 1995). In countries that use only genomic selected bulls with high turnover, all the daughters of bulls can be born in one single year. In the GEBV validation tests the DYD or deregressed genomic predictions (DRP) from the recent AM BLUP are predicted using the GEBV from 4 years reduced data (Mäntysaari et al. 2010). In the countries with efficient genomic selection the DYD and DRP from the AM BLUP are known to be biased alike their source EBV.

To avoid validation of GBLUP using AM BLUP results, VanRaden (2021) presented a regression method where the current GEBV and GEBV from four-year reduced data are used to predict so-called deregressed genomic evaluations (dGEBV).

In the presented study we tested the efficiency of ITB trend validation tests II and III and the dGEBV regression test on AM BLUP and single step genomic evaluations. The data was the full Nordic RDC population in which the genomic selection has been applied intensively since 2010.

Materials and Methods

Test Data and Model

The phenotypic data were extracted from the official DFS (Denmark, Finland and Sweden) test-day evaluations May 2021. The test day observations were used to compile 305d lactation yields, and the corresponding model and genetic parameters (NAV 2022) were reduced to 9 trait model for milk, protein and fat in 3 first lactations. Unlike in official test-day model, the same heritabilities were assumed in each country, and no heterogenous variance adjustments were applied. Records were from 3,534 thousand RDC, 852k Finnish Holstein, and 28k Finncattle cows. The pedigree included 4,911k RDC, 1,102k HOL, and 42k Finncattle animals.

Genomic evaluations were run using single-step GTBLUP with 30% residual polygenic proportion. Genotypes of 46,914 SNPs were from 39k RDC bulls, and 167k RDC cows. There were 148 unknown parent groups (UPG) formed using breed by country by time classes. In the ssGTBLUP evaluations the UPG were replaced by metafounders (Legarra et al. 2015). See the details of the data and analyses (Kudinov et al 2022).

The validation methods are testing whether the evaluations can model the genetic trend in data. Therefore, the data was analyzed using two different models. The first model (CONTROL) was trying to mimic the official DFS RDC evaluation. From the second model we removed the calving age effect (AGELESS). The genetic trends for the AM BLUP were 2.45 kg/y for the CONTROL and 1.83 kg/y for the AGELESS model (Figure 1). The single-step evaluations gave higher estimates, 2.75 kg/y in CONTROL and 2.17 kg/y in AGELESS, and was thus more robust against the model change (Figure 2).

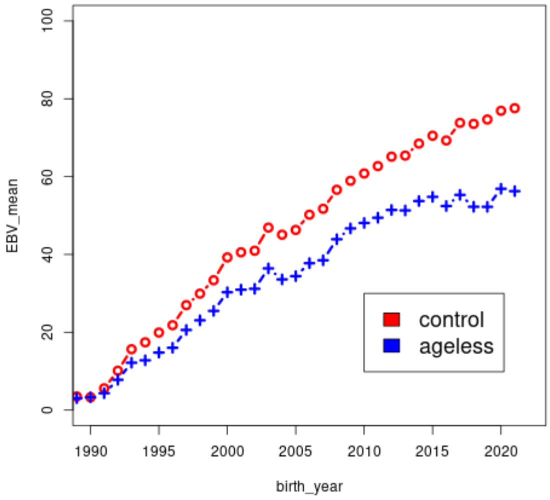

Figure 1. Trends of 305d protein EBV in Nordic (DFS) RDC bulls. Control is a model adapted from TD evaluations; from ageless the calving age effect has been removed.

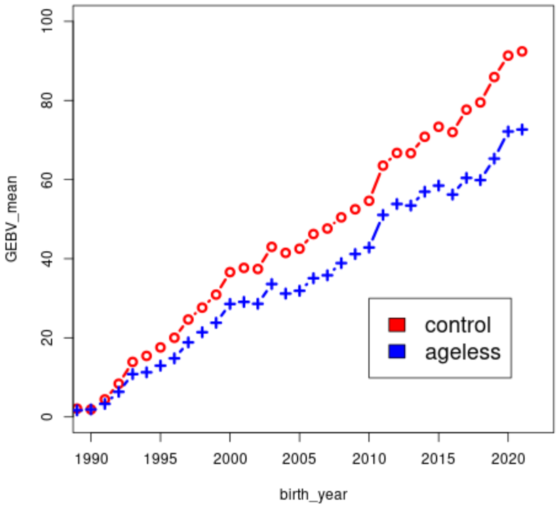

Figure 2. Trends of 305d protein GEBV in Nordic (DFS) RDC bulls. Control is a single-step GTBLUP model adapted from TD evaluations; from the ageless the calving age effect has been removed.

Validation methods

The validation tests II and III were performed using the python programs provided by Interbull (Palucci, pers comm. 2022). The bulls born after 1990 were used in test II, and bulls born 2007-2011 were used in test III. The test II estimates the year trend in DYDs within sires. In our multi-lactation model, the yield deviations (YD) were first estimated separately for each lactation observation y_i as y_i-x_ib; next these were adjusted by half of the breeding value of their dams; and finally, averaged over sire x first calving year. For ssGTBLUP DYD the solutions for environmental effects **b** and the breeding value estimates for the dams were from the single-step run. Final combined DYD were formed as weighted means of lactation wise DYD using the official NAV lactation weights 0.25, 0.35 and 0.45 for the 1st, 2nd and 3rd lactation, respectively.

The assumption in dGEBV regression test (VanRaden, 2021) is that the full data and reduced data evaluations have the same base and same variance. This was assured by adjusting the base of (G)EBV of the reduced run to agree with the full run as described by Sullivan (2020). The validation cohort in dGEBV regression test were 320 bulls that had no daughters in 4 years reduced evaluation but had more than 20 daughters in the current. Inflation was tested with only the reduced (G)EBV in the regression, and bias with including the birth year of the validation bull into the validation regression model.

In official ITB validation tests II and III the genetic trend is considered under/over -estimated if the absolute value of regression coefficient exceeds the biological acceptance level. This acceptance level is 0.01 and 0.02 of the genetic standard deviation for the tests II and III, respectively. In addition, in test III, the test statistic has to be statistically significant. In dGEBV test we used approximate statistical significance, i.e. the t-statistic of regression coefficient had to be between -2.0 and +2.0.

Results & Discussion

Results of the ITB validation test II are given in Table 1. In both runs, AM BLUP and ssGTBLUP, the control model passed the test, but the ageless model failed the test. There were 2,768 bulls used for the test. The average value of j (year of DYD from the first DYD) was 1.64 indicating the variation among birth years of the daughters was relatively little.

Table 1. ITB validation test II results for the control model and for the model without calving age effect (ageless). EBV are from a multi-trait model with milk, protein and fat, but only the protein is presented.

Model	b[1]	Testval[2]
AM BLUP	-0.11	0.004
AM BLUP_ageless	-0.43	0.016
ssGTBLUP	-0.09	0.003
ssGTBLUP_ageless	-0.39	0.015

[1] Regression coefficient b from the test model DYDij=BULLi+b*j

[2] Testval =abs(b)/σ_{bv}

Results of the test III are in Table 2. All the models passed the statistical significance test, although for every model the trend estimate was below the biological significance limit. ITB applies the statistical test based on bootstrap confidence limits. In case of ageless ssGTBLUP this did not include zero, thus that model would have failed in true situation.

Table 2. ITB validation test III results for the control model and for the model without calving age effect (ageless). The trait is protein and number of test bulls 54. Statistically non-significant coefficients are marked with superscript NS.

Model	b*t (SD)	b/σ_{bv}
AM BLUP	-0.544^{NS} (0.561)	-0.021
AM ageless	-0.699^{NS} (0.583)	-0.026
ssGTBLUP	-0.715^{NS} (0.637)	-0.027
ssGTBLUPageless	-0.875^{NS} (0.483)	-0.033

The problem in test III is the lack of statistical power to detect the non-zero bias. In our test with data from 2021, only 54 bulls had attained usable second crop progeny during the last 4 years. For illustration, we counted the number of bulls that received more than 20 daughters after the age of 4 years. During 2017-2020 only 57 “old” bulls received ≥20 daughters, while during 2006-2009 such number was 328.

Table 3 shows the results from the dGEBV regression validation. The test was done also to AM BLUP but only the ssGTBLUP results are given here. The b_1 coefficients for the $GEBV_{red}$ are expected to be 1.00 and b_2 for the birth year of bulls are expected to be zero. A high R^2 value in the inflation test indicates that the $GEBV_{red}$ is predicting the dGEBV well. A significant coefficient for b_2 and corresponding increase in R^2 indicate that the average bias is associated with the age of the bulls. A negative -4.56 for b_2 in control model, indicates to decrease in overestimation for the youngest validation bulls. The absolute value of coefficient was even higher with the ageless model. Note that this estimate is a partial regression, i.e. after the $GEBV_{red}$ has been included in the model.

Table 3. dGEBV regression test (VanRaden 2021) for the single-step control model and for the ageless model. The trait is protein. In the inflation test only the GEBV from 4-year data and in the trend test both the GEBV and the birth year of the bull in the model.

	b_1 ($GEBV_{red}$)	T (b_1=1)	b_2 (year)	T (b_2=1)	R^2
Control model					
Inflation test	0.60	-11.0			0.48
Trend test	0.75	-6.1	-4.56	-6.0	0.53
Ageless model					
Inflation test	0.56	-11.2			0.39
Trend test	0.75	-5.8	-6.23	-8.0	0.50

The R^2 of inflation test model was 0.48. This is higher than the 0.27 for the AM BLUP (not in Table 3), but lower than 0.52, the direct regression of GEBV of the validation bulls from the full run on the GEBV from the reduced run (Legarra and Reverter 2018). An ITB GEBV validation test from the same data was computed by regressing the dEBV from AM BLUP run to the GEBV from the ssGTBLUP reduced data run. This was 0.43.

Comparison of the three test results is difficult because the tests II and III and dGEBV regression test have all different cohorts. The time period in test II is the bulls born 1990 to 2017 while bulls in the test III were born 2007-2011 and received a second daughter crop 2016-2019. This is again different to bulls in dGEBV regression test, where the validation cohort was mostly born in 2013-2017.

MME approach to compute DYD

The test II seemed to detect the faulty model most reliably. It has not been used much because of difficulties computing the DYD. These are not well defined for multi-lactation models when the tested (G)EBV is a combined trait. In test-day models, the daughters that have less observations in lactation than the degrees of freedom in corresponding non-genetic lactation curve do not contribute to their sire DYD (Mrode 2014).

We tested also an alternative approach to calculate the DYD. The DYD are essentially sire x daughter year averages of the YD minus half of the breeding values of mates. In a simple single trait model, the elements of the matrix product $(\boldsymbol{Z}'\boldsymbol{Z}+\boldsymbol{A}^{-1}\lambda)\widehat{EBV}$ corresponding to the cows with records are equivalent to YD. In the formula above, **Z** is the design matrix linking observations to breeding values, **A** is the numerator relationship matrix, and λ is a ratio of residual over genetic variance. Note that the multiplication is easy to do by simply reading the sires and dams and EBV solutions into a data table. With multi-trait models the above requires summing of vector quantities, but we hypothesize that for many models the single-trait MME approach could be used as simple approximation.

In our test we:

1) Formed the combined protein yield with the official NAV weights for 6.2M animals.
2) Computed MME based YD assuming combined index to have a heritability of 0.44.
3) Subtracted $-0.5*EBV_{dam}$, from each YD.
4) Summed results over sires x daughter birth years to generate 69,204 DYD.

Our R implementation steps 2) to 4) took 9 seconds including the data input and output. The MME based approximation and the DYD formed as described in Material and Methods had a correlation of 0.98 (0.99) for DYD including more than 20 (50) daughter observations.

Conclusions

The validation tests II and III can be used for single-step genomic evaluations as well as for the AM BLUP. In our tests the ITB validation test II was able to detect the artificially generated faulty estimates of genetic trend. The test III indicated proper functioning, but because of too few proven bulls returning to service, the test results were not statistically significant. The dGEBV prediction test indicated problems in both genomic evaluation models, although results with faulty model were worse.

Acknowledgements

The authors thank Paul VanRaden, Pete Sullivan, Raphael Mrode, and Zengting Liu for their helpful suggestions as members of Interbull's Validation Working Group with coordination by Valentina Palucci. We thank Nordic Cattle Genetic Evaluation NAV for providing the data for the study.

References

Boichard, D., Bonaiti, B., Barbat, A. and Mattalia, S. 1995. Three methods to validate the estimation of genetic trend for dairy cattle. Journal of dairy science, 78(2), 431-437.

Kudinov, A.A., Koivula M., Strandén, I., Aamand, G.P. and Mäntysaari, E.A. 2022. Single-step genomic prediction in small breeds: Finncattle case. WCGALP 2022 Rotterdam, Netherlands.

Legarra, A., and Reverter, A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genetics Selection Evolution, 50(1), 1-18.

Legarra, A., Christensen, O. F., Vitezica, Z. G., Aguilar, I. and Misztal, I. 2015. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. Genetics, 200(2), 455-468.

Mrode, R. A. 2014. Linear models for the prediction of animal breeding values. Cabi.

NAV 2022. NAV routine genetic evaluation of Dairy Cattle – data and genetic models. https://nordicebv.info/wp-content/uploads/2022/06/NAV-routine-genetic-evaluation_Heiferfertility-71012022NEW_gap.pdf

Mäntysaari, E. A., Liu, Z. and VanRaden, P. 2010. Interbull validation test for genomic evaluations. Interbull bulletin, (41), 17-26.

Sullivan, P.M. 2020. Base adjustments to improve validation tests. Interbull Technical committee report 17.6.2020.

VanRaden, P. (2021). Improved genomic validation including extra regressions. Interbull bulletin, (56), p 65-69.