GEBV Validation Test Revisited

E. A. Mäntysaari and M. Koivula
Genetic Research, Agrifood Research Finland,
31600 Jokioinen.
E-mail: esa.mantysaari@mtt.fi; minna.koivula@mtt.fi

Introduction

Interbull GEBV validation test has been presented to assure that the genomic evaluations (GEBVs) in the member countries can be reliably used as an input for international GEBVs. Moreover, GEBV validation tested genomic indices of young bulls or bull dams are trustworthy references in international trade of breeding animals.

The test protocol was presented in the Interbull genomic workshop February 2010 (Mäntysaari *et al.*, 2010). Eight countries participated on preliminary test for the protein GEBVs in July 2010. The protocol was further defined 2011 (Mäntysaari, 2011). The countries were instructed to provide test data to Interbull for all the traits the country will submit GEBVs for international evaluation. Thus, the protocol would be the same as for the conventional evaluations. Moreover, the requirements for consistence of GEBV evaluations were extended from statistical significance to also consider biological significance. By this amendment, any country or trait that fails statistical test $\hat{b}_1$=E[b_1], can still be accepted as long as the absolute value of $\hat{b}_1$-E[b_1] is less than 0.1.

Inclusion of biological acceptance region to the test resulted from discussion of the validity of the test statistic $(\hat{b}_1 - E[b_1])/SE(\hat{b}_1)$. Firstly, the estimation of the E[b_1] has distracted many, because it can be less than one for the traits where selection is not expected. Secondly, it has been suggested that $SE(\hat{b}_1)$ is too low, because the validation bulls are related, or, alternatively that $SE(\hat{b}_1)$ does not treat low heritability traits fairly (Loberg *et al.*, 2011).

The objective of this presentation is to examine the properties of the GEBV test in practice. A test run of GEBVs for the Nordic Red Cattle were used as an example. In addition, results from the January 2012 GEBV test from the Interbull Center are presented.

Material

GEBV test results were illustrated with genomic evaluations for milk production in Nordic Red Dairy Cattle. The evaluations were from a single step approach and based on deregressed proofs (DRP) of all cows having records in TD data. The evaluations are described in Mäntysaari *et al.* (2011). The Interbull GEBV test was adapted to single step approach by first truncating the data from year 2003 onwards, and, in addition, assuring that all daughters of validation test bulls were removed. The validation bulls were born 2000 or later. Date was chosen to provide more bulls than in Mäntysaari *et al.* (2011). Final reduced evaluation included 3211 genotyped bulls with daughters and 1509 validation bulls. Next, all the 3.401 million DRP were used to recalculate animal model EBVs for all the 54,175 bulls, and the EBVs were deregressed for bulls having non-zero EDC. The DRP for the validation bulls were then solved and used in GEBV validation.

January 2011 GEBV validation test results were received from the Interbull. The data included 211 tests of 38 different traits from 14 different countries and 6 breeds. The GEBV test results from the Intergenomics project were not included in the data set.

Methods

Nordic GEBV data were used to illustrate the sampling variation in the test and the effect of selective genotyping of the validation bulls.

Sampling variation in $\widehat{b_1}$ was first estimated from 10 independent random bull subsets. The empirical standard deviation of $\widehat{b_1}$ was estimated over the samples, and was compared to model based SE($\widehat{b_1}$). Next, the effect of size of test data was illustrated by increasing the number of test bulls from 132 step-by-step to all 1509 bulls. In both the analyses the model based SE was also estimated from 1000 bootstrap samples ($SE_b(\widehat{b_1})$). In each sample a corresponding number of bulls (e.g. 132,…, 1509) were drawn from the subset with replacement. Effect of selective genotyping on the b_1 estimates were tested by culling the test bulls batch-by-batch until only 723 (48%) selected bulls were used. Culling was done either by full data DRP or by reduced data EBV. First resembles the usual selection where the low EBV bulls in history file are not genotyped and the latter scenario represents practice in real breeding program where only the high parent average (PA) bulls are progeny tested.

In addition to linear regression estimation of b_1, the estimate was derived from maximum likelihood (ML) estimates of variance components. This was done by fitting a simple random model in SAS PROC MIXED with the option repeated. In ML all the animals are expected to have DRPs and for the non-genotyped animals the GEBVs are declared missing.

January 2011 GEBV validation tests were carried out in the Interbull Center as described in GEBV test documentation (Interbull, 2012). First the expected value of b_1 was estimated using the selection differential obtained from the difference of means of validation bulls (genotyped) and all the bulls available on the same birth years. Then, the bias in GEBVs was tested using the linear regression model that estimates the mean of GEBVs (b_0) and the consistency of differences among GEBVs (b_1). The value of genomic information was accessed by comparing the R^2 of the GEBV regression model to the R^2 of PA regression model on predicting the DRPs of the test bulls.

Results and Discussion

Validation tests with Nordic single step GEBV as an example

The GEBV validation test $\widehat{b_1}$ regression coefficient from the full data was 0.74. Although 700 bulls were moved from the reference to validation the estimate was the same as in Mäntysaari et al. 2011. As the $E[b_1] = 1.0$, the result indicates that variation in GEBV can be considered biased. Figure 1 shows how the $\widehat{b_1}$ converged to final value when more validation bulls were added to test. However, all the estimates were within one standard deviation from the final estimate. Although the estimate showed no trend, the SE($\widehat{b_1}$) decreased steadily when more bulls were added to test. The difference between bootstrap estimate of SE and model based estimate was small. The SE($\widehat{b_1}$) was 0.049 when the number of bulls was 443. This is close to 0.051 that has been agreed to be the smallest SE($\widehat{b_1}$) to be used in official GEBV test.

The standard deviation of E($\widehat{b_1}$)estimates on 10 independent validation samples was 0.076. This is slightly lower than the model based SE($\widehat{b_1}$), which on average was 0.088. The range of estimates was 0.57 to 0.85 (Figure 2), but the values were well within 2* SE($\widehat{b_1}$) from the mean value 0.73.

Non-random selection of validation bulls caused clear reduction in GEBV validation test $\widehat{b_1}$ (Figure 3) when the selection was based on DRPs. Already a 9% culling led to $\widehat{b_1} = 0.58$, i.e. reduction in value by 0.16. According to the Interbull GEBV test instructions, a 9% culling will reduce the E($\widehat{b_1}$) from 1.00 to 0.80 when the R^2 of the GEBV is 0.34, as here. In the real selection situation also the R^2 has to be estimated from the selected sample. Therefore, the estimation of E($\widehat{b_1}$) has to be done first with the original R^2 and then redone with R^2/E($\widehat{b_1}$). To avoid this, the correlation, and thereafter also the $\widehat{b_1}$, could be estimated with ML methods. Figure 3 includes the $\widehat{b_1}$ derived from ML variance component estimates. This

method seems to be fairly robust against selection as the reduction in b_1 is only 0.03 in the fourth sample (25% culling). However, figure 3 shows that even when all the data are used, the ML based estimate is not the same as $\widehat{b_1}$ in the official GEBV test. This is caused by different definition of weights in regression analysis and in ML variance component estimation.

In the past breeding programs the selection of bulls for the progeny testing, and thereafter to genotyping, has been based on PA. We imitated progeny testing selection by culling bulls based on reduced data EBVs. In the extreme case this reduced $\widehat{b_1}$ from 0.74 to 0.66 (culling of 52%). The selection on PA has a small effect because it reduces almost equally the numerator and denominator of the b_1 estimation equation. In the future genomic selection program the bull calves are selected based on GEBVs. This, again, should not have a notable effect on b_1 regression coefficient.

January 2012 GEBV tests

All 14 countries participating in the test had submitted the protein GEBV test results. The conformation traits were received from 7, udder health from 6 and fertility traits from 3 countries. The rest of the traits or trait groups were from one or two countries.

Figure 4 summarizes the results from the protein tests. From the tests 15 were from Holstein populations, 4 Brown Swiss, 2 Jersey, 2 Simmental, and 1 Red Holstein. Data included 2 repeated tests from one HOL and BSW populations. The error bars in the graph will cover 1.96*SE($\widehat{b_1}$) around the estimated b_1. The bars with straight end are model based and the beveled end bars are based on biological acceptance region (corresponding SE($\widehat{b_1}$)=0.1/1.96). Thus, the trait fails, if the E[b_1] (i.e. the red X) is not in the confidence interval (CI).

None of the protein b_1 tests failed. There were few $\widehat{b_1}$ not in the model based SE($\widehat{b_1}$) interval, but also those fitted into biological interval. Generally it was observed that when the number of test bulls was higher than 500, the model based CI was narrower than biological. The variation in E[b_1] was surprisingly large with smallest value being 0.54. The smallest values were associated with very low R^2.

For the reproduction traits (Figure 5) three of the tests indicated $\widehat{b_1}$ was lower than the E[b_1], and one being higher. Interestingly, the first and the 12th test are submission and resubmission of the same breed and population. In the resubmission the GEBV have been reworked so that CI of $\widehat{b_1}$ includes 1.0, but if test is applied strictly, the value differs significantly from E[b_1]. The test number 11 shows out a very large SE($\widehat{b_1}$). This is caused by small number of test bulls (79). While this does not allow to judge the evaluations being biased, the small number of test bulls is obviously accompanied by small number of reference bulls. In this case, the R^2 of the GEBVs was 0.41 while the R^2 of a PA was 0.47.

On the average the R^2 of genomic evaluations for protein was 37% (13 %-units) higher than corresponding R^2 of PAs (Figure 6). Corresponding increase was 46% (16 %-units) and 25% (13) for fat and somatic cell score GEBVs. The gain in accuracy is associated with size of the population. For the protein GEBVs, the increase in R^2 was visible when the number of test bulls was larger. A correlation of number of validation bulls and gain from genomic evaluations (i.e. $R^2_{GEBV} - R^2_{PA}$) was 0.60.

Conclusions

The Nordic GEBV data were used to illustrate the sampling variation in the test. The variation in $\widehat{b_1}$ among independent random samples corresponded well the model based SE($\widehat{b_1}$). In addition, the SE estimated using boot strap approach did not differ much from model based SE($\widehat{b_1}$).

The effect of selective genotyping of the validation bulls was shown clearly to affect the estimate of b_1. Although in the selection examples considered, the Interbull approach to estimate E[b_1] seems to agree well with

observed reduction in b_1, the problem could be studied more. It was illustrated that an estimation of variance-covariance components with a simple bi-variate model with missing GEBVs for non-genotyped animals is robust against culling with respect to DRPs.

The GEBV tests submitted to the Interbull by national evaluation centers showed that the test for b_1 can be passed by large and small populations. In cases with a small number of test bulls, the test $b_1=E[b_1]$ has a low power on detection of biasedness. Therefore the improvement of R^2 by GEBV over PA should also be required.

Acknowledgements

Genotypes analyzed were from the Nordic project *Genomic selection* by Aarhus University, Nordic Cattle breeding industries, MTT and SLU. Phenotypes (EBVs) were obtained from *NAV Nordic Cattle Genetic Evaluation.*

References

Interbull, 2012. Interbull validation test for genomic evaluations – GEBV test. http://www.interbull.org/images/stories/GEBV_validationtest_June2010.pdf. Accessed 19.3.2012.

Loberg, A., Jorjani, H. & Dürr, J. 2011. Validation of genomic national evaluations. Proceedings Interbull Meeting August 27-28, 2011, Stavanger, Norway. *Interbull Bulletin 44*, 62-66.

Mäntysaari, E.A., Liu, Z. & Van Raden, P. 2010. Interbull validation test for genomic evaluations. Proceedings of the Interbull International Workshop March 4-5, 2010, Paris, France. *Interbull Bulletin 41*, 17-21.

Mäntysaari, E.A. 2011. GEBV Test for all traits. Proceedings of the Interbull international workshop, Guelph, Ontario, Canada. February 27-28, 2011. *Interbull Bulletin 43*, 29-31.

Mäntysaari, E.A, Koivula, M., Strandén, I., Pösö, J. & Aamand, G.P. 2011. Estimation of GEBVs using deregressed individual cow breeding values. Proceedings of the Interbull international workshop, Stavanger, Norway. August 27-29, 2011. *Interbull Bulletin 44*, 19-24.

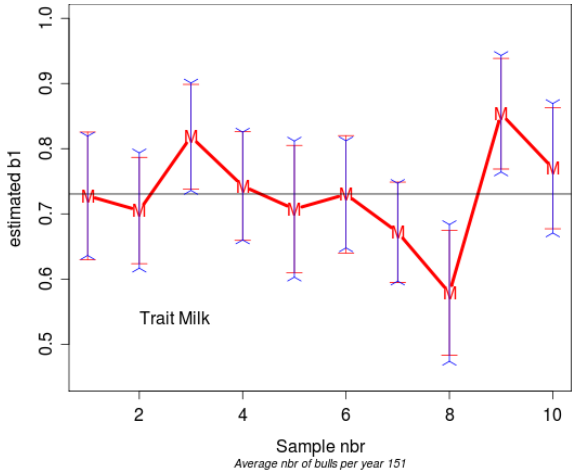

Figure 1. Estimates for b_1 from 10 independent GEBV tests with Nordic Red single step genomic evaluation. Each test has different validation bulls. Average number of bulls in tests 151. Straight error bars are model based, and beveled are from bootstrap samples.

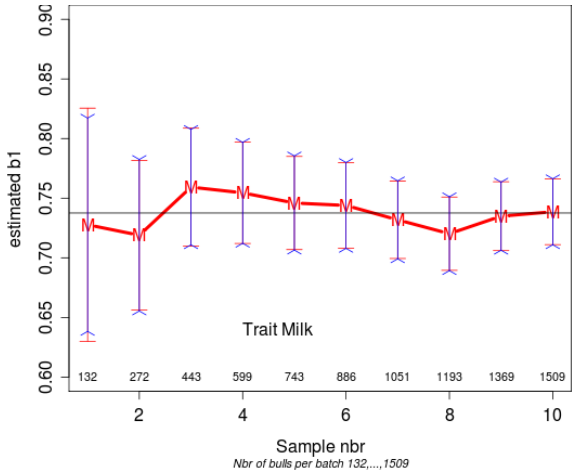

Figure 2. Estimates for b_1 from 10 GEBV tests with Nordic Red single step genomic evaluation. Number of bulls in test is incremented stepwise until all 1509 are in test. Straight error bars are model based standard deviations, and beveled are from bootstrap samples.

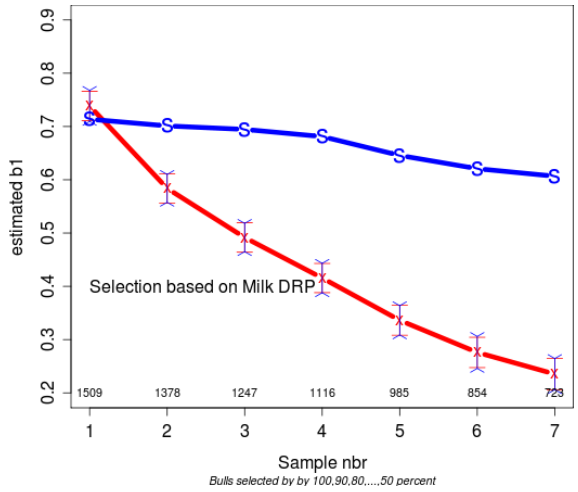

Figure 3. Estimates for b_1 from 7 GEBV tests with Nordic Red single step genomic evaluation. Test 1 includes all validation bulls, from test 2 onwards the worst 132 bulls are removed, until only 723 bulls are kept. Points with X are from regression analysis, S is from ML analysis. Straight error bars are model based and beveled are from boot strap samples.

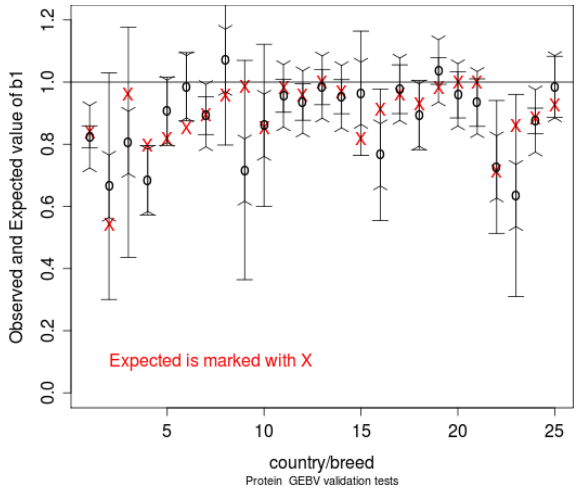

Figure 4. GEBV test results for the protein GEBV from 25 country/populations. Red X marks the E[b_1], O are the $\hat{b}_1$, straight error bars are model based 2*SE($\hat{b}_1$) and beveled are biological acceptance region.

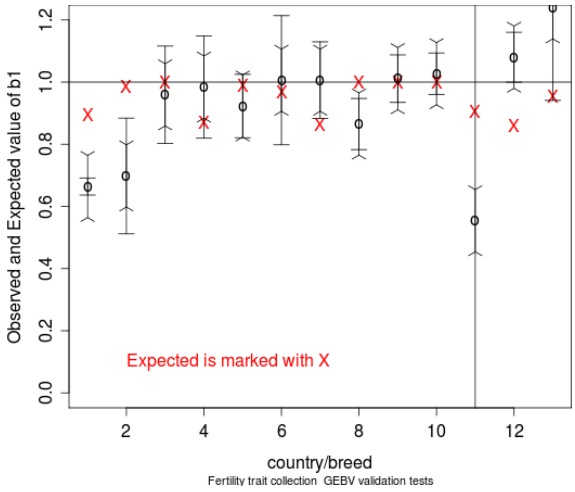

Figure 5. GEBV test results for the reproduction traits on 13 countries/traits. Red X marks the E[b_1], O are the $\hat{b}_1$, and straight error bars are model based 2*SE($\hat{b}_1$) and beveled are biological acceptance region.

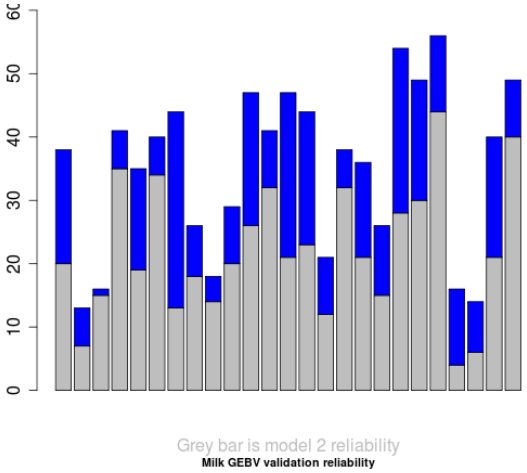

Figure 6. GEBV test R^2 for the protein traits on 25 country/populations. Grey bar represents the R^2 of parent average and blue is the added genomic information.