A Review of the Validation of National Genomic Evaluations

M. A. Nilforooshan
Interbull Centre, Department of Animal Breeding and Genetics
Swedish University of Agricultural Sciences, Uppsala, Sweden
Mohammad.Nilforooshan@slu.se

Abstract

A descriptive review of the data and results of GEBV tests in a period of January 2013 to May 2014 was provided in this study. There were 357 GEBV tests, 259 of them for Holstein. 51 of 259 tests were repeated tests for the same traits. Currently, 4 tests are involved in the GEBV test. 283 cases passed the GEBV test, 206 of them passed the statistical test. These numbers were 216 and 164 cases for Holstein, respectively. Considering the 259 cases for Holstein, the minimum and the maximum number of test bulls were 11 and 4,892. Half of the GEBV tests had less than 330 test bulls, and 58 cases had less than 100 test bulls. The average proportion of genotyped candidate bulls (test bulls) was 0.82, with no clear difference between the statistically passed and the failed tests. (Genotyping) selection intensity was generally low, and expected regression slope ($E(b_1)$) was close to 1 for most of the cases. The range of the regression slope (b_1) was from 0.16 to 1.8, and the range of $|b_1–E(b_1)|$ was from 0.001 to 0.799. The reason for some cases with large $|b_1–E(b_1)|$ passing the statistical test was large $SE(b_1)$, which was ranged from 0.007 to 0.874. 150 of 259 GEBV tests for Holstein had $SE(b_1) < 0.1$. R^2 of the GEBV test model were intermediate to low, with only 46 of 259 cases having $R^2 > 50$. There might be concern for populations passing the GEBV test with very low R^2 and very low number of test bulls.

Key words: GEBV test, validation, genomic evaluation, test bulls

Introduction

For the countries to join the GMACE service, the national genomic evaluation model has to be validated. Since August 2010, the method proposed by Mäntysaari *et al.* (2010) has been in place for this validation. This method is testing the significance of the bias in approximation of future DYD or deregressed proof from the GEBV of young bulls. This bias is mainly introduced through selective genotyping of candidate bulls (Patry & Ducrocq, 2009a,b). With genomic selection, if information in the selection procedure is not accommodated in the model, the genetic/genomic evaluation model would be biased. Ignoring a (non-random) part of the data, on which selection is based, the assumptions of BLUP are violated and Mendelian sampling effect would no longer have a mean of 0 and a variance equal to half of the genetic variance (Patry & Ducrocq, 2009a,b).

Since January 2013, Interbull Centre has released a python program for the GEBV test (gebvtest.py©, 2013). This program is distributed to the countries, so the countries can perform the GEBV test themselves, before submitting the data to the Interbull Centre. The input data files, the created log file and the result file would be compressed, ready to be submitted by the countries. This tool provides a full transparency between the edits applied at the national level and the Interbull Centre level. This program also enables countries to make necessary changes in their applied edits. For example, considering a group of a more or less recent born candidate bulls, or possibly considering foreign bulls among candidate bulls for very small populations.

Materials and Methods

Data

The data used in this study was the GEBV test results since January 2013 to May 2014, from different country-breeds (populations, Table 1) for as many as 38 traits. In total, there were 357 population-traits, with 306 of them being unique. There were 51 repeated tests in different times, all for Holstein.

Because there were not many tests across countries for non-Holstein breeds, only the results for Holstein are illustrated.

Table 1. Frequency of the GEBV tests for different countries and breeds.

	BSW	HOL	JER	NOR	RDC	SIM
AUS	0	5	3	0	0	0
BEL	0	56	0	0	0	0
CAN	15	36	36	0	36	0
CHE[1]	0	7	0	0	0	0
CHR[2]	0	10	0	0	0	0
DEU	3	1	0	0	0	3
DFS	0	15	0	0	0	0
ESP	0	44	0	0	0	0
FRA	0	10	0	1	0	1
GBR	0	23	0	0	0	0
ITA	0	11	0	0	0	0
NLD	0	2	0	0	0	0
POL	0	9	0	0	0	0
USA	0	30	0	0	0	0

[1]Swiss-HOL, [2]Swiss-Red-HOL

Model

The model used for the validation of national genomic evaluations is a weighted linear regression model (Mäntysaari *et al.*, 2010):

$$Y = b_0 + b_1 \times GEBVr + e$$

where, Y is either DEBV (deregressed proof) or DYD, b_0 is the intercept, b_1 is the linear regression slope, $GEBVr$ is the genomically enhanced EBV from a few years ago (EBVr), and e is the random residual effect.

The default cut-off year was chosen to be 8 years difference between have the current data based on daughter information (full data), and previous data based on parent averages (reduced data). This is the average number of years that a bull gets information based on its first batch of daughters. Another value can be chosen instead of the default value (gebvtest.py©, 2013).

In the presence of non-random selection (selective genotyping) b_1 and $E(b_1)$ of the model deviate from 1. Significant deviation of b1 from $E(b_1)$ indicates significant bias. For $E(b_1)$, please see Nilforooshan *et al.*, 2011. A 2-tailed t-test is involved to access the significance of the bias.

There is different amount of information and accuracy involved for different bulls, therefore a weighted least squares regression model is used.

The method is further described in other literature (Nilforooshan *et al.*, 2010; Mäntysaari *et al.*, 2010). Previously, an EDC_i weight was used for $DEBV_i$, and $EDC_i/(EDC_i+\lambda)$ for DYD_i, which since 27 February 2015 has changed to EDC_i for both DEBV and DYD.

Tests

Since the first introduction of the GEBV test (August 2010), 3 other tests have been put alongside the statistical test (test1). One test is to check whether there has been a gain in accuracy using genomic information (test4). To do this, a parallel linear model (model2) is considered, in which the dependent variable of the GEBV test model is replaced with parent averages or EBV from some years ago (EBVr). The R^2 of the GEBV test model (model1) should be greater than the R^2 from model2 (R^2M_2). Otherwise, that population-trait cannot pass the GEBV test. Another added condition is whether b_1 is greater than $E(b_1)$ (test3), which would lead to a pass from the test. For large populations, passing the statistical test

can be difficult, because of a SE(b_1) close to 0. Another test, called the biological test (test2), eases passing large populations with very low SE(b_1). Whereas in the statistical test: b_1–2SE(b_1) < E(b_1) < b_1+2SE(b_1), in the biological test: b_1–0.1 < E(b_1) < b_1+0.1

Therefore, the biological test would favor SE(b_1) < 0.5. The final pass or fail of a population-trait is an outcome of the 4 tests. Putting the results of test1 to test4 (Y/N) in a row, all combinations except ???N and NNNY pass the test.

Results & Discussion

283 of 357 tests (all breeds) could pass the GEBV test. 52 population-traits failed due to R^2 value less than R^2M_2 (???N), and 22 population-traits failed because none of test1, test2 and test3 passed (NNNY). From the 283 cases that passed the GEBV test, 206 of them passed the statistical test (Y??Y), 127 of them passed the biological test (?Y?Y), 110 of them passed both test1 and test2 (YY?Y). There were 60 cases that could qualify, only because those had b_1 > E(b_1) (NNYY). b1 > E(b_1) is equivalent to $R^2V_Y > E(b_1)^2V_{GEBVr}$.

Considering Holstein results, Figure 1a shows the frequency of test bulls. The peak of the frequency falls in 0–100 (58 cases). A closer look to this span (Figure 1b) shows that many of those population-traits (47 cases) had only 20–50 number of test bulls.

Figure 2 plots the number of candidate bulls against the number of test bulls. The slopes were 0.846 and 0.798 for the passed and the failed tests, indicating the average proportion of genotyped bulls.

Whereas the proportion of the number of test bulls to candidate bulls did not show any clear difference between the (statistically) passed and failed tests, the more important is how those candidate animals are selectively genotyped. Figure 3 shows the frequency of selection intensity among the studied population-traits.

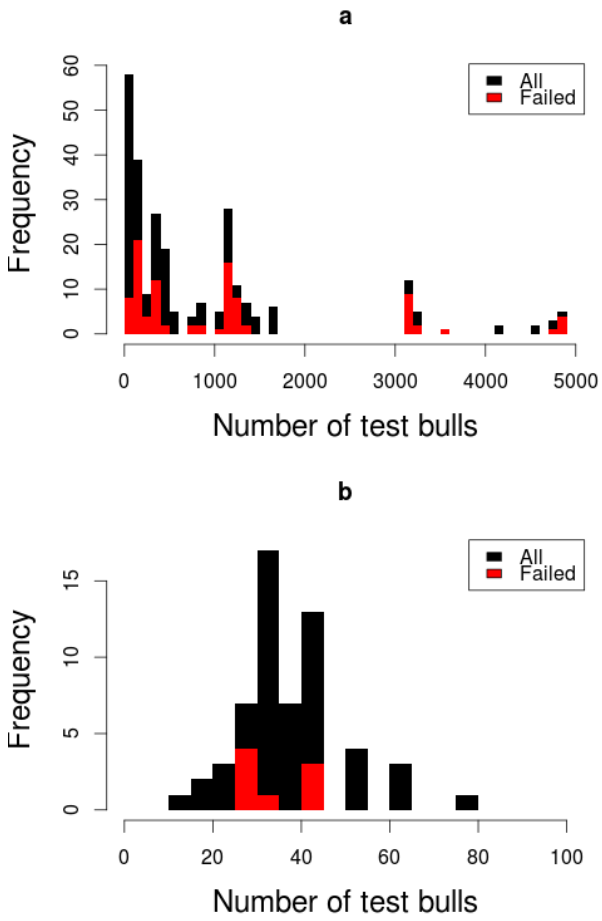

Figure 1. Frequency of test bulls across population-traits (red bars failed in the statistical test)

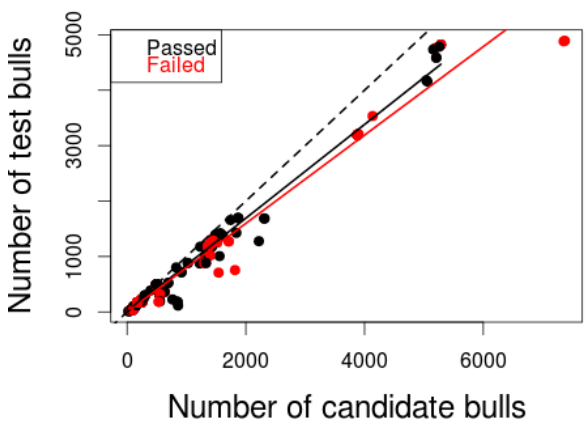

Figure 2. Number of test bulls and the number of candidate bulls in the studied GEBV tests (dashed line: number of test bulls = number of candidate bulls, red circles failed in the statistical tests)

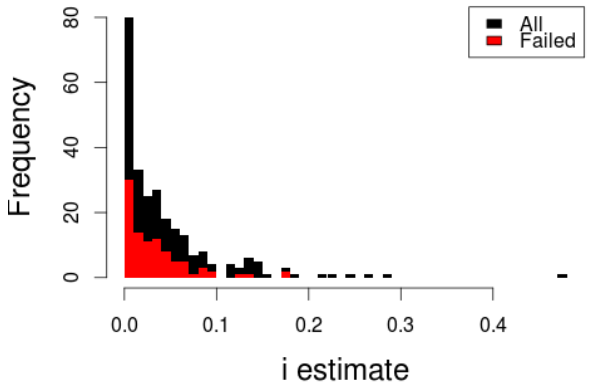

Figure 3. Frequency of selection intensity (*i*) across population-traits (red bars failed in the statistical test)

Selection intensity was small in most of the cases (i < 0.05, 183 of 259). Though, those were a few (9) i > 0.18 cases, all of them could pass the statistical test. The reason was that, though high selection intensity coincides with lower b_1, it reduced $E(b_1)$ too, which can help the population-trait to pass the test. A proof to this, is the frequency of $E(b_1)$ among the population-traits (Figure 4), which shows the opposite trend as in Figure 3.

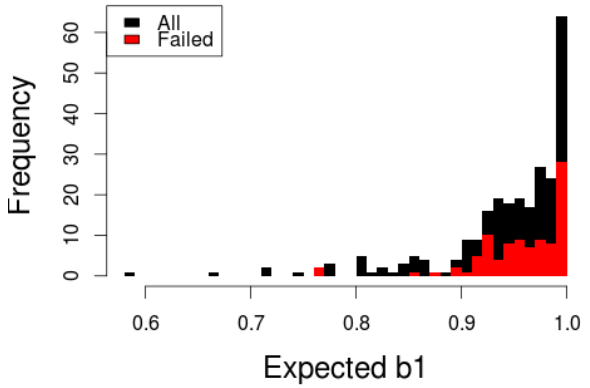

Figure 4. Frequency of Expected b_1 value across population-traits (red bars failed in the statistical test)

There are 3 influential factors in the GEBV test, b_1, $E(b_1)$ and $SE(b_1)$. The reason for some population-traits with low proportion of test bulls to candidate bulls passing the statistical test is either $E(b_1)$ was much lower than 1, b_1 was close to 1, or $SE(b_1)$ was high.

Whereas the test is currently relaxed for b_0, it is focused on b_1, and how it is deviated from $E(b_1)$. Figure 5 shows the frequency distribution of b_1 in the studied population-traits. There were more cases passed the statistical test for $b_1 < 1$ compared to $b_1 > 1$. However, according to test3, $b_1 > 1$ would finally pass the test (if $R^2 > R^2M_2$), because of $E(b_1)$ not being greater than 1. Some cases with b_1 largely deviated from 1 could pass the statistical test. The reason was large $SE(b_1)$ values for those cases. Though such cases get a pass report from the software (gebvtest.py©, 2013), the person in charge of the GEBV test at the Interbull Centre will report this issue to the corresponding country to check for possible problems in their data, before any pass decision is made.

The two main components of the F test are $|b_1–E(b_1)|$ and $SE(b_1)$. Figure 6 shows the distribution of $|b_1–E(b_1)|$, and Figure 7 shows the distribution of $SE(b_1)$. With $|b_1–E(b_1)|$ increasing and $SE(b_1)$ decreasing, the probability of passing the test decreases. There were many cases (109 of 259) with high $SE(b_1)$ (greater than 0.1). This can justify why a few large $|b_1–E(b_1)|$ could pass the statistical test (Figure 6).

The distribution of b_1 against $SE(b_1)$, limited to $SE(b_1) < 0.5$ (Figure 8) shows a greater probability of passing the statistical test with b_1 closer to 1, and greater $SE(b_1)$ values.
The studied GEBV tests showed low to intermediate model fitness (R^2, Figure 9). There were 46 cases (of 259) with $R^2 > 50$, from which 26 cases could statistically pass the GEBV test.

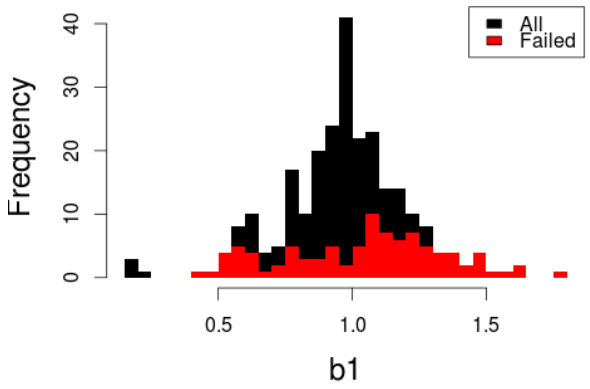

Figure 5. Frequency of b_1 value across population-traits (red bars failed in the statistical test)

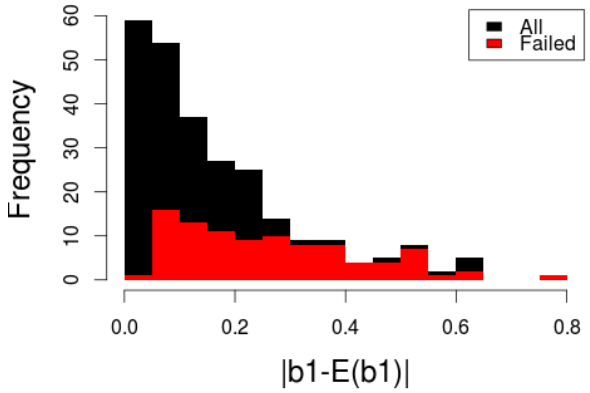

Figure 6. Frequency of $|b_1–E(b_1)|$ value across population-traits (red bars failed in the statistical test)

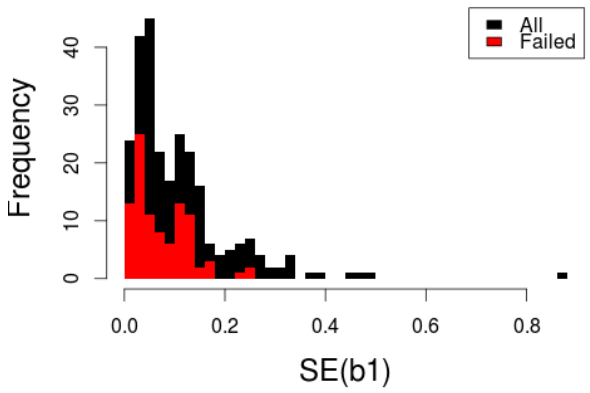

Figure 7. Frequency of $SE(b_1)$ value across population-traits (red bars failed in the statistical test)

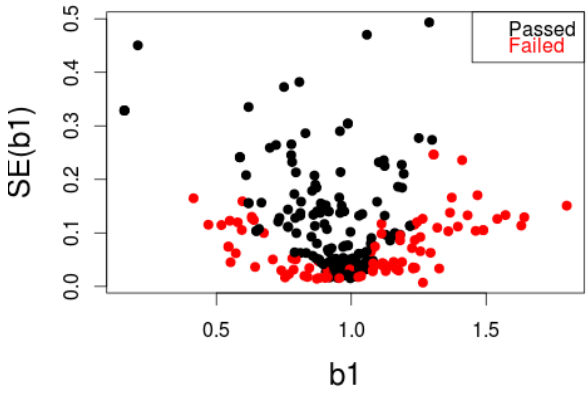

Figure 8. b_1 and $SE(b_1)$ values in the studied population-traits that had $SE(b_1) < 0.5$ (red circles failed in the statistical test)

Another important factor for national genomic evaluation systems is the gain in accuracy ($\Delta\sqrt{R^2}$) using information from DNA markers. Figure 10 plots R^2 against R^2M_2. The vertical distance of the circles from the dashed line shows the R^2 gain. The minor cases that could not gain R^2 (below the dashed line) were failed by test4. On average, in the cases that passed the statistical test, R^2 increased one unit by one unit increase of R^2M_2. For the cases that failed the statistical test, the R^2 gain was less than a unit by a unit increase in R^2M_2. Reliability gain was low for R^2 of model2 greater than 30, which can be an indication of higher reliability gain for lower heritable traits, using genomic information.

One alarming point was the low number of test bulls and the low value of R^2 among the tests. The small number of test bulls was mainly a matter of small number of candidate bulls rather than a low proportion of candidate bulls being selected (Figure 2). Figure 11 shows the distribution of R^2 against the number of test bulls multiplied by h^2, limited to 100 (167 of 259 cases). As it can be seen, there were numerous population-traits that could pass the test with very low R^2 and low number of test bulls. For example, there were 21 cases with $R^2 < 10$, and 21 cases with number of test bulls $\times$ $h^2 < 5$ (i.e., < 20 test bulls for an h^2 of 0.25) passing the test. There were 5 tests having both conditions.

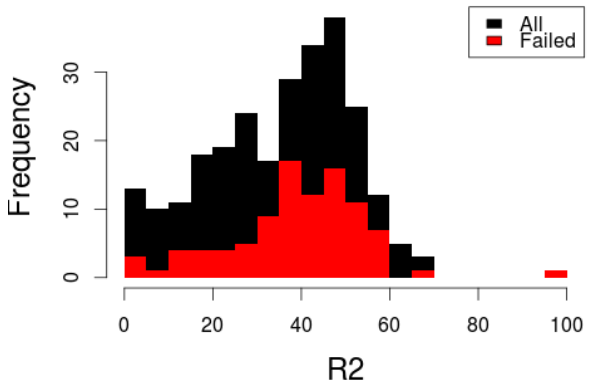

Figure 9. Frequency of R^2 value across population-traits (red bars failed in the statistical test)

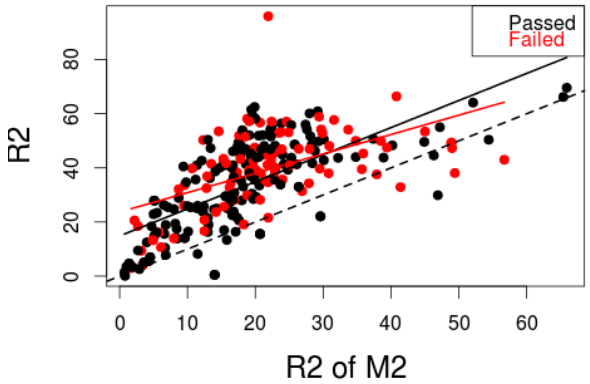

Figure 10. R^2 and R^2M_2 values in the studied population-traits (R^2M_2 is the R^2 of the GEBV test model with the GEBVr independent variable replaced by parent averages, dashed line: $R^2 = R^2M_2$, red circles failed in the statistical test)

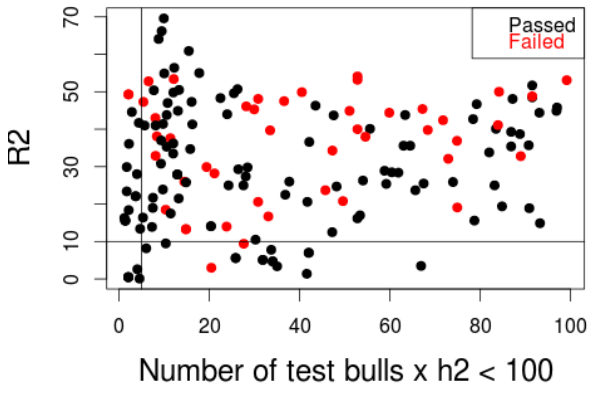

Figure 11. Number of test bulls × h^2 and R^2 in the studied population-traits that had number of test bulls × $h^2 < 100$ (red circles failed in the statistical test)

After all, with directional selection of candidate bulls for genotyping, national genomic evaluations would not be free of bias. Though it is important that the bias is insignificant, it is also important that the national genomic model is robust enough. Though a national genomic evaluation may pass the GEBV test with an $R^2 < 0.1$ and a few test bulls, the robustness of this model is under question with more data added in the future, especially if the reference population size is also small, and it is in a different phase than the testing population.

This study showed the data structure of GEBV test in several occasions of evaluation, and the role of some of the key factors in the GEBV test. Looking at one parameter at a time, may not show a clear distinction between the passed and the failed (statistical) tests. This lack of distinction was more evident in the last 3 graphs about the R^2 of the model, which indicates more underlying parameters that should come into the picture to make a trend. One of those parameters, as an example, is the size of the reference population for the estimation of marker effects.

References

gebvtest.py©. 2013. Program gebvtest.py - *User Manual. Version 1.* Interbull Centre, Uppsala, Sweden. Retrieved May 2014, from https://wiki.interbull.org/public/gebvtest_py?action=print

Mäntysaari, E., Liu, Z. & VanRaden, P. 2010. Interbull validation test for genomic evaluations. Proceedings of the Interbull Meeting, March 4–5, 2010, Paris, France. *Interbull Bulletin 41*, 17–21.

Nilforooshan, M.A., Zumbach, B., Jakobsen, J., Loberg, A., Jorjani, H. & Dürr, J. 2010. Validation of National Genomic Evaluations. Proceedings of the Interbull Meeting, May 31 - June 4, 2010, Riga, Latvia. *Interbull Bulletin 42*, 56–61.

Patry, P. & Ducrocq, V. 2009a. Bias due to genomic selection. Proceedings of the Interbull Meeting, January 27-28, 2009, Uppsala, Sweden. *Interbull Bulletin 39,* 77–82.

Patry, P. & Ducrocq, V. 2009b. Evidence of a bias in genetic evaluation due to genomic selection. Proceedings of the Interbull Meeting, Agust 21-24, 2009, Barcelona, Spain. *Interbull Bulletin 40,* 167–171.