# Validation of Genomic National Evaluations

***A. Loberg, H. Jorjani and J. Dürr***
*Interbull Centre, Box 7023, SE-750 07 Uppsala, Sweden*

## Abstract

Validation of national genomic data is now implemented as an official Interbull service, the GEBV-test. The test has grown from initially only validating production trait protein yield to be able to handle all Interbull traits. In addition to the statistical significance test (t-test) for passing the test a new biological significance criteria has been introduced. Among the countries that have validated for all three production traits, all pass for all traits. For the other traits the results are diverse, indicating that the test works best for the traits with medium to high heritability.

## Introduction

The introduction of genomic data and the development of new national evaluation methods to estimate marker effects, international trade within the dairy business met some new challenges, such as the need to extend the current international evaluation to accommodate genomic data. Among other things a validation procedure of national GEBVs was necessary to ensure unbiased international genomic evaluation.

The *GEBV test* is offered as a service to the Interbull customers since August 2010, but it is still limited to protein yield, although the plan is to extend its application to all traits. During the Interbull international workshop in Guelph (February 2011) a new set of rules for the *GEBV test* was established by the Steering Committee: a new validation is needed when major changes in national evaluation or in the reference population are carried out or every two years; two types of significance level were defined for the test, a set of estimates will pass the test either if reach statistical significance (±2 SE), or if lies within a biological significance (Confidence Interval: ±0.1). The motivation for adding the biological criteria was the perception that large populations were penalized by the t test when they reached very low standard errors, and small populations would likely pass the test even with much larger deviations from the expected value due to equally large standard errors. The Steering Committee also stimulated countries to submit validation data for milk yield, fat yield and key traits other than production (stature, somatic cell score, direct longevity and fertility trait cc1) to help further evaluation of the method (Dürr, 2011; Mäntysaari, 2011).

The objective of this paper is to give an overview of the results from countries that have validated all production traits so far, as well as the first results for other key traits. The impact of the biological significance test on the validation is also discussed.

## Material and Methods

From the populations that have validated for protein, 17 populations are included in this study. These include the populations validated after the Interbull workshop in Guelph February 2011, and the populations that had validated for all production traits before the workshop. For the other traits results are included if three or more populations validating after the workshop.

**Table.** Number of populations per trait.

| Trait[1] | no. populations |
|---|---|
| **pro** | 17 |
| **mil** | 14 |
| **fat** | 14 |
| **sta** | 9 |
| **scs** | 8 |
| **cc2** | 4 |
| **int** | 3 |
| **msp** | 4 |
| **dlo** | 6 |

[1]pro= protein yield, mil=milk yield, fat=fat yield, sta= stature, scs= milk somatic cell, cc2= abilility to conceive, interval, int=calving-conception interval, msp= milking speed, dlo= direct longevity

The validation was performed by the method described by Mäntysaari *et al.* (2010), Nilforooshan et al (2010) and Interbull (2011).

## Results and Discussion

The results are shown in Figure 1-9. The absolute t-value should be equal or below the threshold of 1.96 to pass the t-test. To pass the biological significance test the expected $b_1$ value should be within the range of $b_1 \pm 0.1$, labeled as the lower and the upper bound in the figures.

All populations that have participated in the validation test for a production trait have passed (Figure 1-3). For protein yield all population but three pass the t-test, but those populations did pass the biological significance test. For the populations that also validated for milk and fat yield, there is only one population in each trait that did not pass the t-test, and in both incidences the populations passed the biological significance test. The conclusion is that it seems like if a genomic evaluation pass the GEBV-test for protein yield, it will also pass for the other production traits.

For the other traits results are diverse. The conformation trait stature show good results, with all populations but one passing both significance tests (Figure 4). For milk somatic cell, three out of 8 populations pass the validation test. For most of the populations that did not pass the test, the value of the slope ($b_1$) is low. This is also seen for direct longevity (Figure 9), with three out of six passing the test. For the fertility and workability traits there are so far very few countries validating, but we can see that milking speed (Figure 8) has a better result than the fertility traits (Figure 6-7).

It seems like the GEBV-test works better for the traits with medium to high heritability than for the ones with low heritability.

## References

Dürr, J. 2011. Executive summary, Interbull workshop, Guelph, Canada. http://www.interbull.org/images/stories/Ex_summary_final_corrected2.pdf. Accessed 22.08.2011

Interbull, 2010. Interbull validation test for genomic evaluations - GEBV test. http://www.interbull.org/images/stories/GEBV_validationtest_June2010.pdf. Accessed 16.08.2011

Mäntysaari, E.A. 2011. GEBV Test for all traits. *Interbull Bulletin 43*, 27-29.

Mäntysaari, E.A., Liu, Z. & VanRaden, P.M. 2010. Interbull validation test for genomic evaluations. *Interbull Bulletin 41*, 17-21.

Nilforooshan, M.A., Zumbach, B., Jakobsen, J.H., Loberg, A., Jorjani, H. & Dürr, J.W. 2010. Validation of national genomic evaluations. *Interbull Bulletin 42*, 56-61.

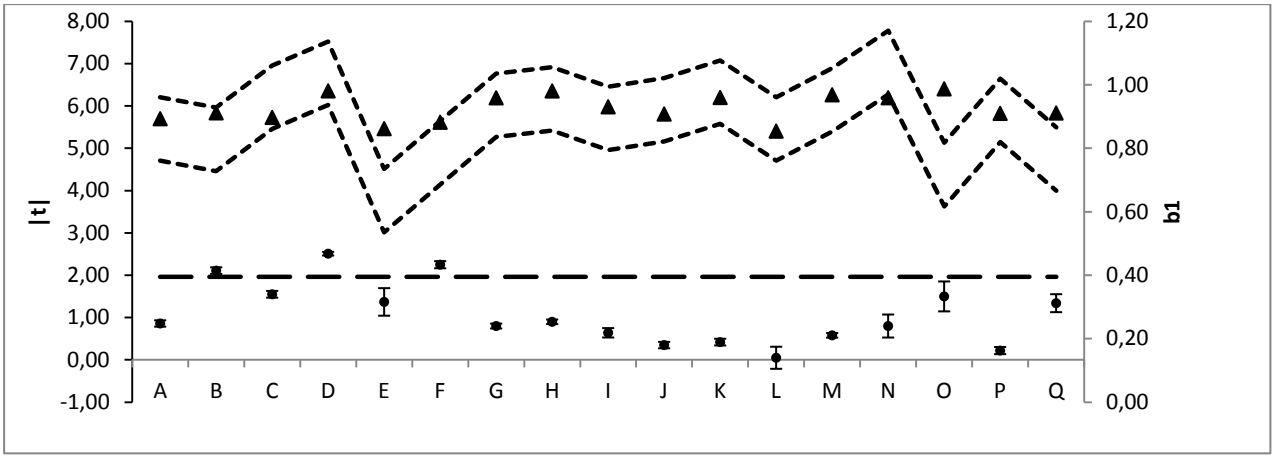


**Figure 1.** Results from the GEBV test for protein yield showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

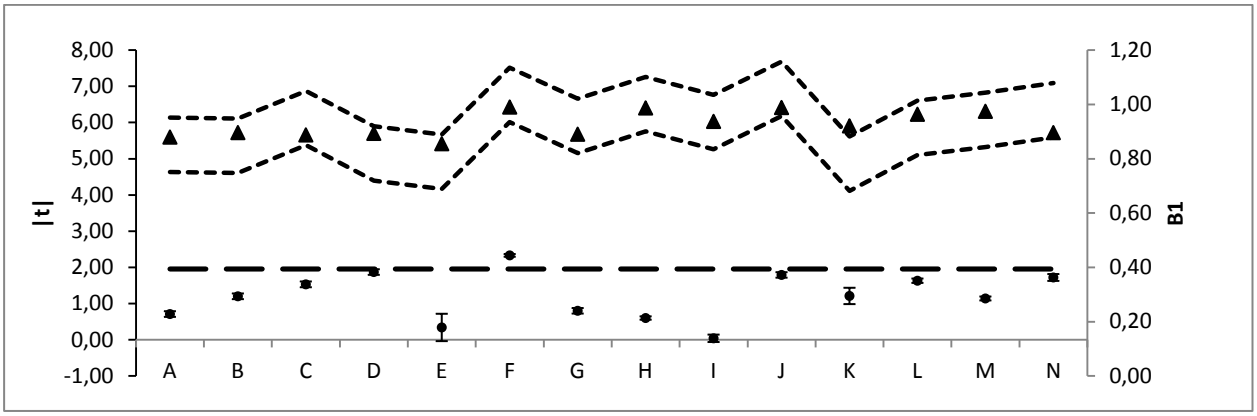


**Figure 2.** Results from the GEBV test for milk yield showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

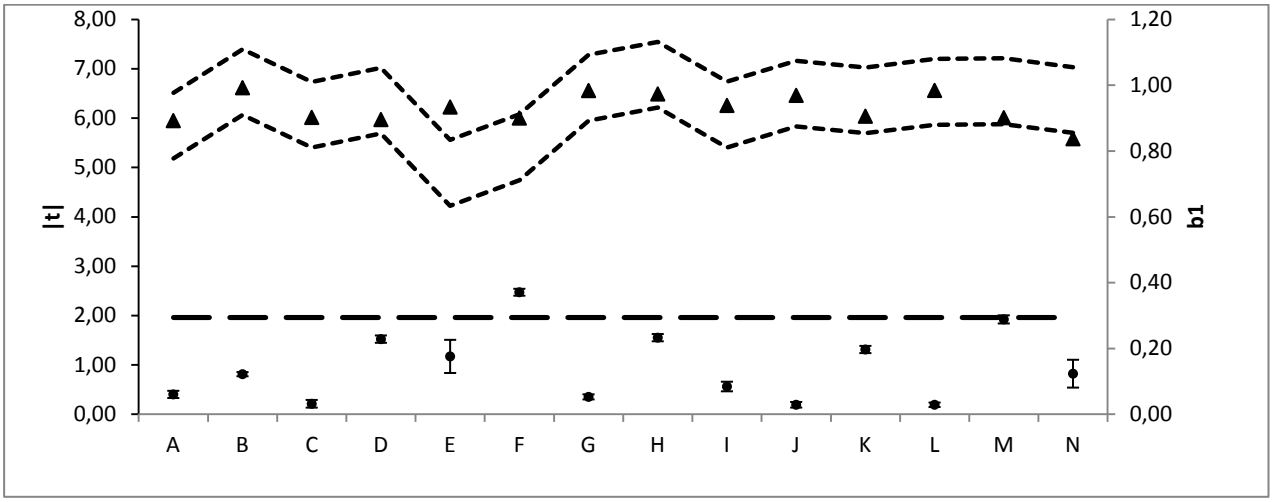


**Figure 3.** Results from the GEBV test for fat yield showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

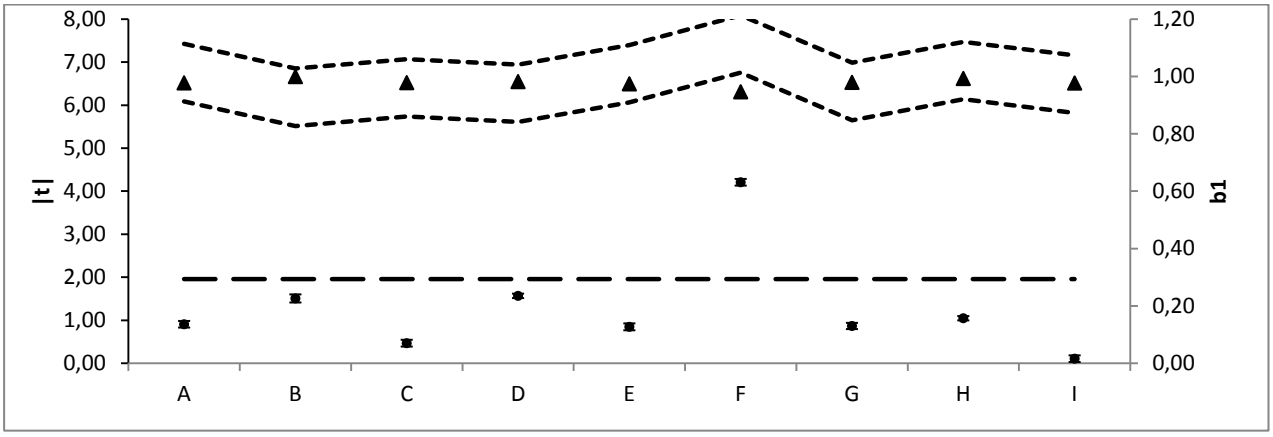


**Figure 4.** Results from the GEBV test for stature showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

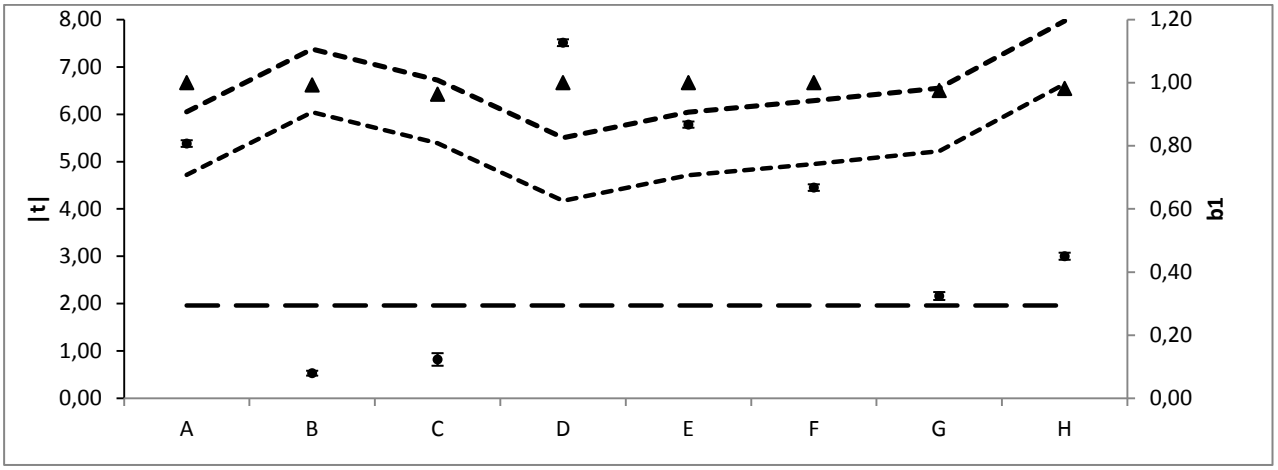


**Figure 5.** Results from the GEBV test for milk somatic cell showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

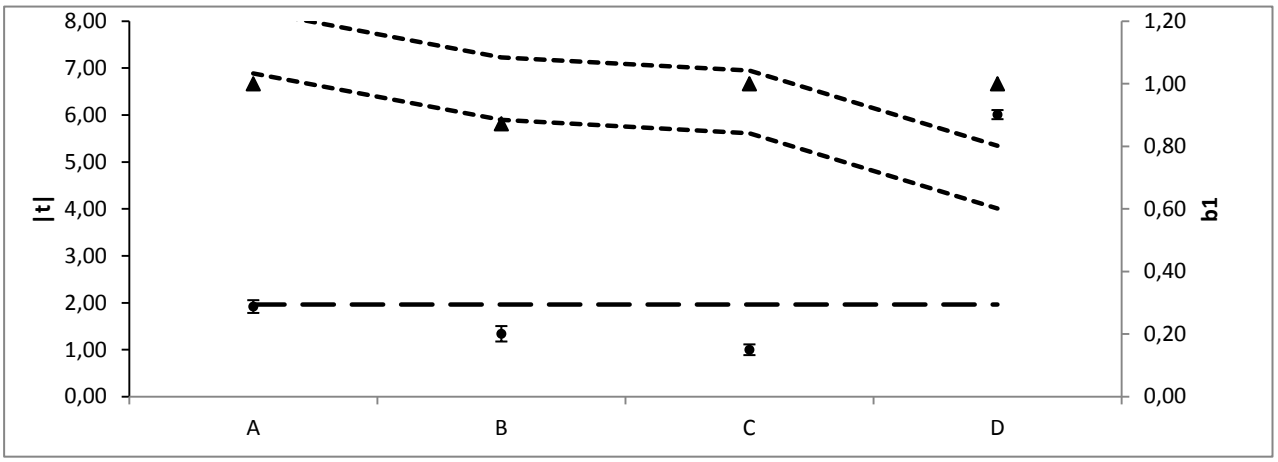


**Figure 6.** Results from the GEBV test for ability to conceive, interval showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

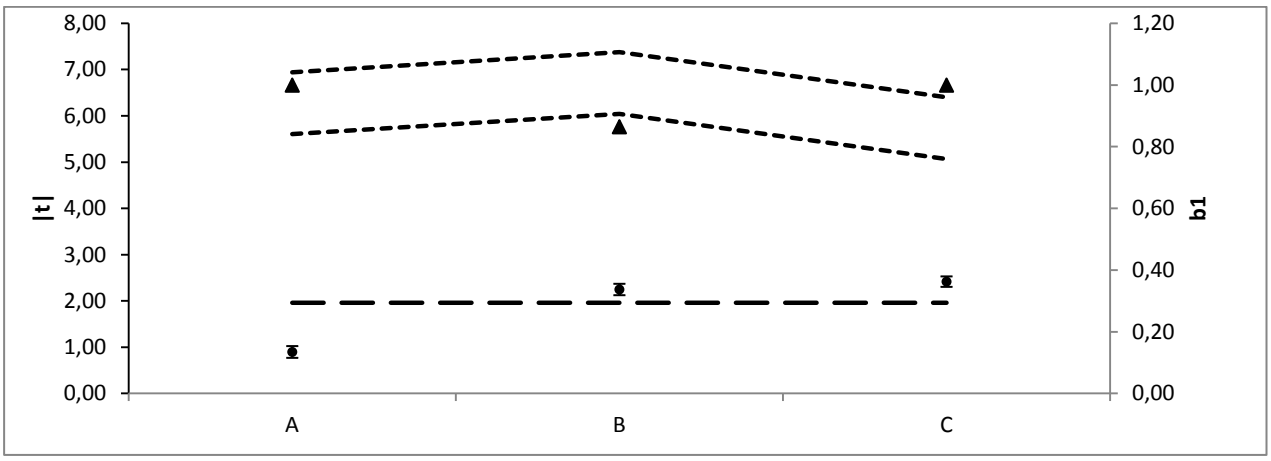


**Figure 7.** Results from the GEBV test for calving-conception interval showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

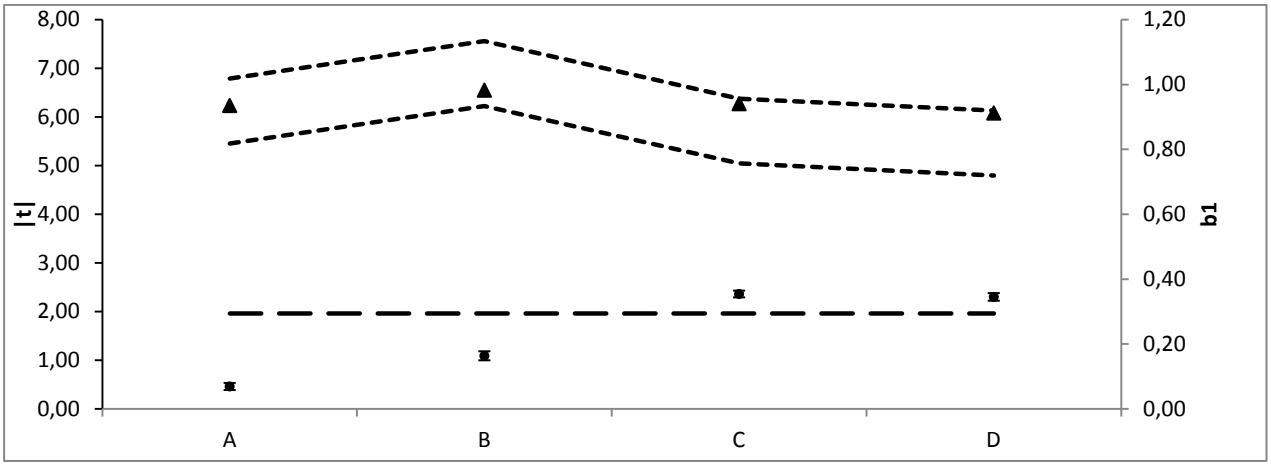


**Figure 8.** Results from the GEBV test for milking speed showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).

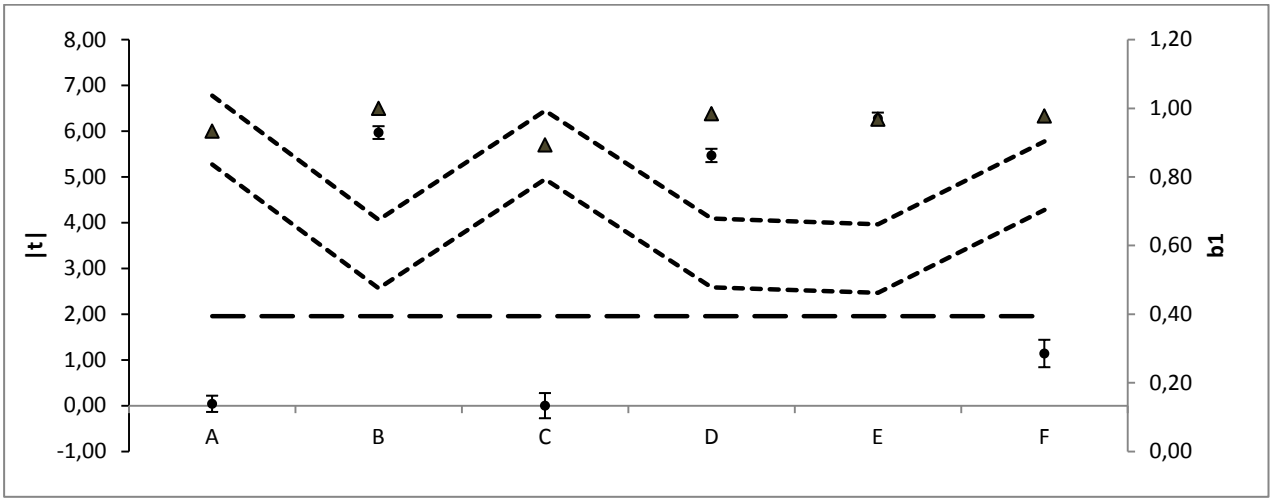


**Figure 9.** Results from the GEBV test for direct longevity showing both the statistical and the biological significance tests. Each letter in the X axis represents a different population (● = |t|; ▲= E(b1); ▬ ▬ = t test threshold of 1.96; --- = biological significance boundaries).