



Development of Dairy Validation Tools

Last updated April 2026

Introduction

Over the years, Interbull Centre together with its collaborators, has developed a set of validation tools to ensure that data provided by the national genetic centers (NGEC) are as unbiased as possible and obtained by statistically sound models. The tools available have been developed for conventional EBVs (four different tests available: 3 trend tests and 1 Mendelian Sampling Variance test) as well as genomic breeding values (GEBVs) via the GEBV test.

On August 2024, during the course of the 2023-2024 EURC working programme, the GEBV test has been reviewed and improved to make it more robust to handle genomic pre-selection variance. The new version of the software was released in August 2024 (https://interbull.org/web/static/web/Dairy_Validation_EURC%20Report.pdf).

During the course of the 2025-2027 EURC working programme, Interbull Centre and its collaborators, is continuing working towards the review and improvements of its validation tools, delivering a clarification on how the new GEBVtest should be used (https://interbull.org/static/EURCreports2026/Clarification_on_GEBVtest_criteria.pdf), and focusing on identifying a more suitable replacement, within the genomic era, for the trend test validation model III, currently applied to conventional (EBVs) evaluation.

Background

The current trend test method III aims at investigating the random variation associated with the accumulation of new daughters within a set period of four years. The test was originally developed in the mid '90s where genetic evaluations were all based on progeny testing schemes where a good number of proven bulls was produced each year in every country. The advent of genomic evaluation had caused a dramatic reduction in the number of proven bulls available in certain countries, reducing, as a direct consequence, the statistical power of the trend test model III as fewer proven bulls would become eligible for the test. Moreover, with the fast-increasing number of organisations/countries moving towards a genomic evaluation, the genomic pre-selection bias is believed to accumulate and increase fast over the years (https://interbull.org/ib/2021_genomic_free_ebv_webinar). There was, therefore, the need to revise the validation method III to:

- Adapt it to the different data structure due to genomic evaluation





- Make it able to better detect genomic pre-selection
- Simplify the math underneath so to make it easier to understand by all users.

Work performed and achievements

The current method III is based on the following model:

$$C_f = C_r + t + e$$

where C_f and C_r are full and reduced (4-years truncated) datasets, respectively, and “ t ” a time trend variable. The test’s null hypothesis is that there will be no time trend detected ($H_0: t=0$).

The variable “ t ” is calculated from a bull’s daughter counts in the full and reduced (truncated) data sets. However, the values of “ t ” can also be predicted with very high accuracy using regression functions that include either all or most of this same information: for example by considering the bull’s birth year and the ratio n_x/n_y , which is the ratio of daughter counts in the reduced versus the full dataset (first crop / first + second crop). Considering that the value of “ t ” tends to 0 for bulls who did not add new daughters in C_f relative to C_r , i.e. the regression of C_f on (C_r and t), this appeared to be mathematically almost equivalent to the base adjustment approach used in the newly released GEBVtest software, which fits a regression of C_f on C_r that is weighted on bulls not adding new data (i.e. who would have values of “ t ” approaching 0). Furthermore, both method III and the GEBVtest software tests are contrasting scale-aligned differences between a bull’s EBV in full versus reduced data, for bulls adding versus not adding any new information in the most recent four years.

Based on these similarities, an alternative test to replace Method III could be developed by exploiting some of the functionalities already available within the new GEBVtest software. Making advantages of existing software would present the following benefits:

- 1) Reduce the time required to develop a new tool
- 2) Simplify the users’ data preparation as no additional data than what is already provided by the countries to apply the GEBV test, would be eventually required.

To identify a suitable new method, two possible target variables (models) are under evaluation: 1) based on de-regressed proofs (DRP) and 2) based on the change in EBVs between current and truncated (four years back) evaluations (EBVdiff). Both models will be fitted with the same number of regressions which will be:

- Base adjustment
- Rel_diff





- Year of birth.

Moving away from dealing with number of daughters and focusing on reliability (or EBV) changes seems to be a good improvement towards easing up the interpretation of the test's results. For example, reliability changes are considered a more accurate value for better identifying groups of bulls that would add data over the four-year period from those bulls who would not. Moreover, the current method III is also focusing on the time when the bulls are adding their daughters. This factor is no longer considered important: if bias is introduced it doesn't seem to be relevant to know if it was introduced in the first two years or last two years. It seems to be more important just to identify that such bias is present in the data.

The possible shift towards using countries' reliability as a new target variable could generate some concern regarding the possible impact that different methods/procedures applied by countries could have on the test result. On the other hand, considering that there has been more than 20 years of experience on developing and applying methods for approximating reliability, it is assumed that countries would have a good solid procedure in place for that and therefore the impact on the test results should be negligible.

Results and Impact

Preliminary results obtained testing the two models with real and simulated data, where a known bias was added to some traits, seemed to prefer the EBVdiff model as it was able to detect the bias in the simulated data for almost all traits. More investigation, however, is still required to fully understand the behavior and differences of the two tested models before being able to identify the most suitable replacement for model III.

