Using Interbull validation test results to improve national genomic evaluations

P.G. Sullivan
Lactanet, 660 Speedvale Ave, Guelph, Ontario, Canada, N1K 1E5
Corresponding author: psullivan@Lactanet.ca

Abstract

The Interbull GEBV test software was recently updated, with new methods to account for genomic pre-selection of young AI bulls, and to add diagnostic features that help national evaluation units isolate and resolve issues if sub-optimal validation test results are observed. Applications of the updated Interbull tests to recent Canadian data confirmed genomic evaluations published by Lactanet are accurate and unbiased, particularly for traits with highest emphasis in current selection programs. Average GEBV validation slopes for Holstein were 1.06 across the three production traits, 1.06 for twenty conformation traits validated for internal use in Canada, and 1.07 for the five production and conformation traits requiring validation by Interbull. Validation slopes were notably lower for other traits, and especially for longevity, fertility, and calving traits when the test was based on a mixed population of foreign and local bulls and while ignoring expected bias in the early progeny-based evaluations of validation test bulls with a limited amount of progeny data. New features in the Interbull validation software were used to explore modeling improvements for traits with sub-optimal validation test results, and with direct longevity as an example trait to demonstrate proper uses of new diagnostic features now available in the updated Interbull validation software. Relative to initial validation test slopes for the mixed population of local plus foreign bulls, which averaged 0.73, 0.71, and 0.77 for the direct longevity, fertility, and calving traits evaluated by Interbull in GMACE, validation slopes for these same three groups of MACE traits were 0.98, 0.99 and 0.89 respectively, for local validation bulls born in Canada, when using an unbiased GEBV validation target. Low validation test slopes for GEBV were almost exclusively attributable to biases in the parent average contributions in GEBV, rather than genomic modeling of the young bull genotype effects, secondarily to the inclusion of foreign bulls born and pre-selected outside of Canada and then progeny-tested within Canada, and thirdly due to genomic pre-selection bias that is expected in a validation bull's first GEBV that are based on only one year of progeny data. This research is ongoing, with plans to further explore validation of foreign bulls in Single Step evaluation systems that will include an integration of MACE information.

Key words: genomic validation, software, GEBV, GMACE

Introduction

The Interbull Centre recently released new standards and software, to validate the accuracy and unbiasedness of national GEBV used for international selection and marketing of young genomic bulls, and for input to the Interbull GMACE evaluation service. The newly added software features were described in detail by Sullivan (2023). The purpose of the present paper was to demonstrate proper use of new diagnostic features, with applications of the new Interbull software to Canadian data.

Materials and Methods

Official GEBV tests were applied to the traits in GMACE that included national GEBV from Lactanet Canada in August 2025. The full data sets were national GEBV published in August

2025, and the truncated data sets were national GEBV generated with current evaluation systems, current and full data sets for pedigree and genotypes, and truncated phenotypic data sets, reduced in size by excluding the four most recent years of phenotypes recorded after August 2021.

For foreign bulls with progeny four years ago in countries outside of Canada only, truncated-data MACE proofs from the Interbull TMACE evaluation service were used. The TMACE results from October 2025 were based on truncated-data national EBV from all countries participating in TMACE, including the truncated-data EBV for local bulls in Canada that were used in the present study. The TMACE results for foreign bulls in Canada are thus on the same scale as truncated-data EBV of local bulls that were submitted to TMACE by Lactanet. The integration of TMACE evaluations into national genomic evaluation of truncated data is preferred for validation testing, over the integration of MACE evaluations published four years ago, based on currently outdated methods for any countries making changes to evaluation systems for any of the traits evaluated in MACE over the previous four years. It would be extremely rare that no country participating in MACE would make changes for any trait, during a period of four years. The theoretical advantages of using TMACE for GEBV validation testing are clear, but multiple practical benefits have also been demonstrated by Mota et al. (2023; 2024; 2026).

The following strategy is recommended to identify and make improvements to national GEBV systems, while making best use of new features in the updated Interbull validation software:

1. Apply the official GEBV validation test as described in the Interbull code of practice (https://interbull.org/ib/cop_appendix8), by using only the default software options.
2. Identify traits with the greatest potential for improvement (i.e. with largest detected biases that could be reduced).
3. Use exploratory features of the Interbull validation software to diagnose the biases. Modified validation tests are encouraged in this step, to most efficiently find and isolate underlying causes of bias.
4. Consider improvements to the evaluation model and/or data pipelines to address the identified causes of bias, in a suitably revised national evaluation system.
5. Confirm better diagnostic results from step 3, when testing GEBV from the modified national system.
6. Repeat step 1 to confirm official Interbull test results are also improved by using the modified national system.

Traits with potential for improvement

For Canadian Holstein traits, genomic reference populations are large, and the Interbull validation test results for GEBV accuracy are consequently always acceptable by Interbull standards. The accuracy test is to confirm ranking of young bulls is improved by genotyping, a condition that is usually met if the genomic reference population is large.

The validation slope test is to confirm no bias in the scale, or variance, of GEBV used to pre-select genomic young bulls. The variance of GEBV for young bulls should be consistent with that of progeny-proven bulls. When the variance of GEBV for young genomic bulls is too high or too low, and it is therefore misaligned with the variance of full-data GEBV after progeny testing, this is detected as a bias with the Interbull validation slope test. A GEBV scale bias is detected when the slope of regression from the validation test model deviates significantly from its expected value, which is equal to 1 for any trait (Legarra and Reverter, 2019), if the trait definition is not changing over time for the validation bulls. Traits with largest deviation from an expected validation slope of 1 have largest relative degrees of bias in the scale of variance for GEBV of young genomic bulls. Traits with largest magnitudes of bias have the greatest potential for improvement, with changes to the

national genomic evaluation system, and the greatest potential for successful diagnosis of underlying causes of the bias.

New diagnostics

One of the important updates in the latest version of Interbull software was a feature to guarantee genetic bases of expression are the same between national GEBV from full versus truncated data sets. Methods and the general need to align genetic bases of expression between the two different data sets used for validation testing were fully described by Sullivan (2023). After aligning the scales of expression for bulls already progeny-proven in reduced data, unexpected changes in scale that remain for young genomic bulls who were not yet progeny-tested four years ago, but who are progeny proven in the full data, can be more easily analyzed, and with a better and faster diagnosis of the reasons for bias.

One approach to find underlying causes of bias is to examine either individuals or groups of bulls with largest differences between GEBV from full versus reduced data sets, or to explore plots of base-aligned GEBV from full versus reduced data to identify outliers with strong impacts on the validation slope. Another approach is to add extra variables to the validation test regression model, as proposed by VanRaden, (2021). The use and benefit of these different approaches will be demonstrated by example below, for a Canadian trait with relatively large magnitude of detected bias.

Results & Discussion

Validation test results for accuracy of Lactanet GEBV were acceptable to Interbull for all traits. The accuracy test results are not reported here because our main interest was in diagnosing biases for Lactanet GEBV that are used in Interbull GMACE calculations.

The validation slope test results after three iterations applying the new Interbull software, with official bias tests using the default software options, are shown in Table 1. Details of the investigations between each iteration of testing are described more fully below.

Table 1: Average validation test slope values, by MACE trait group.

Validation test		#1	#1	#2	#3
With foreign bulls		Y	Y	N	N
Years of progeny*		1-4	1-4	1-4	2-4
Trait Group	Trait tests	PA**	GPA**		
Production	3	0.97	1.06	1.10	1.06
Conf'n	2	1.10	1.09	1.17	1.16
Udder	1	0.84	1.14	1.16	1.20
Longevity	1	0.73	0.74	0.88	0.98
Calving	4	0.77	0.77	0.83	0.89
Fertility	5	0.71	0.84	0.95	0.99
Work'y	1	0.63	0.85	0.83	0.80

*Genomic preselection bias is largest in a bull's first year of progeny proving. Excluding the youngest group of validation bulls limits the effects of GPS-bias on validation slope estimates.

**Regression estimates for GEBV from full data on PA and GPA, denoting the EBV and GEBV from truncated data, respectively.

Traits with highest potential for improvement

For production and conformation traits, all validation test slopes (b_1) for Canadian data were within the acceptable levels of tolerance set by Interbull (i.e. $0.90 < b_1 < 1.20$), for all three iterations of validation testing. In the first set of tests, average validation slopes were 1.06 for the three production traits, and 1.09 for the two conformation traits, stature and median udder support, designated by Interbull as reference traits to validate national evaluation systems for all legacy conformation traits as a group. Validation test slopes for all other Canadian conformation traits in MACE were consistent with slope tests of the reference traits. The average validation slope was 1.06 across twenty conformation traits. For the remaining trait groups, validation test slopes for GPA were low due to slopes that were equally low or lower from regressions of full-data GEBV on the young bull's truncated-data PA. The genotype of a young genomic bull has no expected impact on the evaluations of the bull's parents, and much of the bias detected in GPA (the GEBV test) is due to an over-estimated variance of the differences between parent

averages, rather than due to the bull's own genotype.

The direct longevity trait submitted to MACE by Lactanet was among those with largest detected biases. Investigations of underlying causes of bias for direct longevity were examples of efficient ways to use new diagnostics now provided by the updated Interbull validation software.

New diagnostics

Firstly, from a view of the simple plot of base-aligned GEBV from full versus reduced data sets (Figure 1), we observed an especially high number of outlier points (circled) were for a group of foreign bulls who had very high GEBV before progeny testing (G_{red}=105) followed by average GEBV after progeny testing (G_F=100). A majority of these were foreign young bulls born in the USA, who in truncated data had parental information (PA) heavily influenced by ancestors with MACE proofs, rather than Canadian EBV based on local daughters in Canada. In USA, the direct longevity trait includes the effects of high production on survival (USA productive life), while in Canada the direct longevity trait is independent of production (Canadian functional herd life). By effectively changing the trait definition for validation bulls born in USA, as they transition from information on USA ancestry to Canadian progeny, it should be expected that production effects will be reduced in the Canadian GEBV, and thus the validation slope will have an expected value lower than 1 for high-production USA validation bulls who are progeny-proven in Canada.

Excluding foreign bulls from the validation test group had a dramatic effect on the estimated validation slopes, not only for direct longevity, but also for fertility and calving traits. Validation slopes increased notably with the second iteration of validation testing, based on only local bulls, from 0.74 to 0.88 for direct longevity, from 0.84 to 0.95 for the average of five fertility traits, and from 0.77 to 0.83 for the average for four calving traits (Table 1). As with

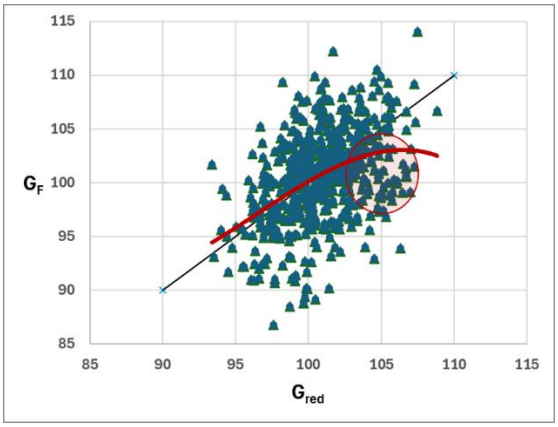


Figure 1. Non-linear polynomial regression of GEBV for direct longevity from full data (G_F) on base-aligned GEBV from data sets reduced by truncation of phenotypes (G_{red}). The highlighted group of points circled, for bulls with high G_{red} and average G_F, caused significant downward pressure on the linear validation test slope.

longevity, the fertility group includes traits that are defined differently in USA versus Canada. The harmonization of calving trait definitions has also been complicated internationally, by different models being used to separate direct and maternal effects of calf and dam respectively, and with categorical data for calving ease and calf survival.

The MACE system accounts for trait definition and modeling differences for bull re-rankings among countries at the population level, but not necessarily at the individually selected bull level. Validation test bulls and their sires are routinely being pre-selected, based on genomics, and based on intense levels of genomic pre-selection (GPS). If the GPS rule has a different correlation with the national definitions of a MACE validation trait defined differently among countries, then the level of correlated GPS effect will change as a validation bull's GEBV evolves, with respect to the relative contributions of information from each country. A pre-selected foreign young bull will transition from having GEBV information that is heavily based on the correlated GPS effects of ancestors in the selecting country, for the truncated-data PA and GPA, to a GEBV based on progeny in the local country in full data. The expectation of regression slopes for

full on truncated data can therefore be lower than 1 for foreign bulls in these cases. In Canadian Holstein genomic evaluation systems, approximately 2/3 of the genomic reference populations are foreign bulls with a MACE proof in Canada versus 1/3 being local bulls with a domestic EBV from progeny data recorded in Canada. The issue of foreign bulls is therefore especially important for validation tests by Lactanet, and potentially for other countries with similar mixtures of foreign versus local genomic reference bulls as Canada.

From a theoretical point of view, it is easily argued that expected values of validation slopes are lower than 1 for foreign bulls, due to changes in the respective trait definitions that drive GEBV information in full versus truncated data. This is the case for direct longevity, fertility, and calving traits in Canada. Interbull has never required the inclusion of foreign bulls in validation test groups, but it has been an optional way to increase validation group sizes that are otherwise too small based on only local bulls.

For Canadian data, Lactanet has historically included foreign bulls for GEBV validation testing, with a continuing interest to validate all progeny-tested bulls in Canada, and because there are many young genomic bulls born in USA who are used for breeding in both USA and Canada. Validation test results for the foreign bulls, however, can be very misleading for traits with expected slopes lower than 1 due to changes in the trait definitions among countries. The inclusion of foreign Holstein bulls is unnecessary for Lactanet because local bulls alone are sufficient for reliable validation tests with the Canadian data.

Further investigations made use of the VanRaden suggestion to add extra regressions in the Interbull validation test model. Of particular interest were the effects of local versus foreign bulls, and the number of years with progeny data in a bull's full-data GEBV. The number of years with accumulated progeny data was considered important for two reasons. Firstly, GPS biases in traditional EBV are highest in a bull's first year of progeny testing, and these biases quickly decrease to small levels as additional progeny data accumulate. Secondly, if a trait definition includes measurement of the trait across multiple parities, the data recorded after first parity are not available until the bull's second and later years with an official progeny-based GEBV. This is the case for direct longevity at Lactanet, which includes daughter survival rates to first, second, and third parity.

A variety of different validation models with extra regression variables were studied, and a good example for direct longevity, fitted separately for local versus foreign bulls was:

$$Y = b_0 + B_1 * (X_1 - \bar{X}_1),$$
$$\text{where } B_1 = b_1 * (1 + b_2 * Z(X_2))$$

and where Y and X_1 are the base-aligned national GEBV from full and reduced data sets, respectively, b_1 is the linear regression used for the Interbull slope test, b_2 is an extra regressions on sire's relative year of birth (Z ranging from 0 to 3), fitted by category of X_2 denoting low versus high levels of expected GPS bias in Y. Variable Z was used as a proxy for the number of years with progeny data in the bull's full-data GEBV. A combined data set was also analysis, fitting separate estimates by bull type, local versus foreign. Estimates were practically the same from combined versus separate analyses by bull type.

The solutions for B_1, the sum of estimated regression effects fitted as interactions with X_1, are in Table 2. Although validation slopes for direct longevity were initially very far from 1 (first test result in Table 1), the slopes were nearly equal to 1 for local bulls with a validation test based on a low-bias validation target (X_2=1, and the third test in Table 1). For both local and foreign bulls, the validation slopes were closer to 1 with a low-bias validation target, but always much lower than 1 for foreign bulls, and this was most likely due to the trait definition in

Table 2: Validation test slopes (B_1) by type of bull, from a validation test model with extra regressions, for the trait direct longevity.

Validation			Variable*		
Bulls	Born	n	X_2	Z	B_1
Foreign	2016	89	1	0	0.67
Foreign	2017	209	1	1	0.68
Foreign	2018	99	1	2	0.70
Foreign	2019	57	0	3	0.58
Local	2016	29	1	0	0.99
Local	2017	86	1	1	0.98
Local	2018	61	1	2	0.97
Local	2019	34	0	3	0.69

*As defined in the text.

change, from including in USA to excluding Canada, the impacts of high production effects on survival.

Although results are not shown here, another factor causing problems with our initial set of validation test results for direct longevity was the inclusion of multiple copies of a cloned USA donor bull's GEBV in the validation test data, once for the donor bull and again for each of the clones. In Canada, the GEBV generated by Lactanet for identical twins, which can include clones, split embryos or naturally occurring identical twins, are a copy for all twins of the single predicted GEBV from pooled data across each set of identical twins in the same identical family. It is correct to include the pooled-data GEBV as a single observation in the validation test data set, but incorrect to include replicated copies of the pooled-data observation, as if they were differently predicted GEBV for each identical twin. The correction to prevent using multiple copies of the same pooled-data GEBV had a remarkably large effect on the slope test results for direct longevity. This discovery came only after reviewing the individual outlier bulls identified in Figure 1.

Across all trait groups, validation slopes summarized in Table 1 were closer to expected values of 1 when based on local bulls only (the second set of tests), and when excluding GEBVs from full data with high expected levels of GPS bias (the third set of tests). In the third set of tests, validation slopes were close to 1 for all trait groups except Udder (Somatic Cell Score) and Workability (Milking Speed). The reasons for bias in national GEBV can differ among traits, and among the national genomic evaluation systems and data sets used for GEBV predictions in different countries.

The GEBV predicted differences for production, conformation, and Udder (SCS) traits are conservative for young, genomic, local bulls born in Canada, as indicated by validation test slopes that are higher than 1 for these trait groups. These results are consistent with historical efforts at Lactanet to limit GEBV bias and to avoid over-prediction of young bulls across all traits, and for all progeny-tested bulls in a mixed population of both local and foreign (mainly USA) young genomic bulls being used for breeding in Canada (e.g. Sullivan et al, 2015). The present study improved our understanding and expectation of different validation test results for local versus foreign bulls. While focusing on local bulls for official validation tests required by Interbull, we also continue efforts to optimize national GEBV for foreign bulls.

Single Step evaluation systems are used by Lactanet to evaluate newer traits not yet evaluated in MACE, while a multi-step genomic evaluation system is used for MACE traits. Methods to integrate foreign progeny information from MACE can differ between multi-step and Single Step genomic evaluation systems, and it has been unclear which of these two systems works best with MACE integration. Better validation methods and improved understanding of validation test results for both foreign and local bulls will help determine if MACE integration into Single Step systems can be equally or more effective than in multi-step genomic evaluation systems. Concerns about GPS bias in a bull's first year progeny proof will be lower with Single Step than with multi-step genomic evaluation systems. Development of Single Step systems that include integration of MACE information is currently underway at Lactanet.

Conclusions

Detecting underlying causes of bias in scale was previously complicated by differing scales of expression between national GEBV from full versus reduced data sets, especially for GEBV expressed on a variance-standardized scale, like the relative breeding value scale of GEBV published in Canada. The complexity of different scales of expression was eliminated by a new feature in the updated Interbull validation software, which was designed to detect and account for scale mis-alignments that can exist in validation test data sets, and to thereby ensure full-data and truncated-data GEBV are adjusted when necessary, and always directly comparable after scale alignment, for all traits and countries. The advantages of base-aligned scales of expression, combined with new software tools and methods to diagnose bias, are more efficient paths to both validating and improving national genomic evaluation models.

The underlying factors shown to cause bias in GEBV tests with Lactanet data, in the present study, were previously difficult to isolate and identify. The new Interbull validation software was invaluable for these new discoveries, and with the software being freely available to all, others can take similar advantage of the new diagnostic tools.

Acknowledgments

Contributions from past and present members of the Interbull validations working group are gratefully acknowledged, in the development and review of new validation methods and software, including Esa Mäntysaari, Paul VanRaden, Zengting Liu, Raphael Mrode, Andres Legarra, Ignacio Aguilar, and Valentina Palucci.

References

Legarra, A. and Reverter, A. 2019. Correction to: Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Gen. Sel. Evol. 51:69. https://doi.org/10.1186/s12711-019-0511-5.

Mota, R.R., Nicolazzi, E, McWhorter, T. and VanRaden, P.M. 2023. Genomic validation software: USA results. Interbull bulletin 58, 27-32.

Mota, R.R., Sullivan, P.G., Nicolazzi, E, McWhorter, T.M., Legarra, A. and VanRaden, P.M. 2024. Genomic validation software: USA update including truncated MACE. Interbull bulletin 60, 27-32.

Mota, R.R., Savoia, S., Legarra, A., McWhorter, T.M. and Nocolazzi, E.L. 2026. A comprehensive GEBV validation framework using Truncated MACE and new traits in U.S. dairy cattle. Proceedings 13th WCGALP 23414.

Sullivan, P.G, Jamrozik, J. and Kistemaker, G.J. 2015. De-regressing MACE versus domestic EBV for genomics. Interbull bulletin 49: 97-102.

Sullivan, 2023. Updated Interbull software for genomic validation tests. Interbull bulletin 58: 7-16.

VanRaden, P.M. 2021. Improved genomic validation including extra regressions. Interbull bulletin 56: 65-69.