Effects of weighting schemes and base adjustment on genomic validation outcomes in the Interbull GEBVtest framework

R.R. Mota[1], S. Savoia[1], A. Legarra[1], T.M. McWhorter[1], E. Nicolazzi[1]

[1] Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

Abstract

Genomic validation is required to ensure the reliability and international acceptance of genomic estimated breeding values (GEBV). Recent updates to the Interbull GEBVtest software introduced alternative options for checking, including different regression equations, weights, and base-adjustment procedures. The objective of this study was to evaluate the effects of these options on genomic validation outcomes in U.S. dairy cattle populations. Genomic validations were performed using the Interbull GEBVtest software and August 2025 U.S. genomic evaluations. A reduced dataset was generated by truncating domestic information after 2021 and combining the resulting national evaluations with foreign evaluations from Truncated Multiple Across Country Evaluation (TMACE) information from Interbull. Five dairy breeds and nine production, health, longevity, and fertility traits were evaluated. All validation scenarios used predicted deregressed genomic estimated breeding values (dGEBV) as the validation target. Four scenarios were compared, differing in regression methodology [Interbull weighted regression (ITB) or unweighted Legarra-Reverter (LR) regression] and application [yes/no] of base adjustment. A total of 42 breed-trait combinations were evaluated under four validation scenarios, resulting in 168 scenario-specific validation analyses. Across all analyses, 91 resulted in PASS outcomes, 18 in FAIL outcomes, 12 in WARN outcomes, and 47 in HiSE outcomes. WARN indicates validation statistics close to predefined acceptance thresholds, whereas HiSE indicates inconclusive validation results due to high slope standard errors. Validation results were generally consistent across scenarios, indicating that the genomic validation framework was robust for most breed-trait combinations. However, a limited number of analyses produced different validation outcomes depending on the methodological configuration applied. These differences were associated with interactions among regression methodology, base adjustment, and validation-group definition rather than any single factor alone. However, for traits that failed validation in some scenarios, validation statistics classified as WARN remained close to the threshold for FAIL. The results demonstrate that genomic validation outcomes are sensitive to details when validation statistics are close to predefined acceptance thresholds. A WARN result close to a FAIL threshold should prompt further investigation of the factors contributing to potential bias. The observed FAIL outcomes may be associated with limitations of the current multiple-step evaluation framework. CDCB is transitioning to single-step evaluation, which should impact validation performance.

Key words: base adjustment, candidate-bull selection, GEBVtest, validation target, weighting schemes

Introduction

Genomic estimated breeding values (GEBV) are widely used in dairy cattle breeding programs for the early selection and international marketing of young bulls. To ensure confidence in genomic predictions, countries participating in international evaluations are required to validate their genomic evaluation systems before GEBV can be used as input for Interbull services and accepted for routine publication. Since its introduction, the Interbull GEBV test has become the standard framework for genomic validation and has undergone several modifications to accommodate changes in genomic evaluation methodologies and data

structures (Mäntysaari et al., 2011; Interbull, 2021).

The rapid adoption of genomic selection has substantially altered dairy cattle populations and evaluation systems. Genomic preselection, increasing use of foreign information, and the implementation of single-step genomic evaluation models have created challenges for traditional validation procedures. In response, several studies have proposed methodological improvements to better account for these changes and to provide validation procedures that remain robust under modern genomic evaluation systems (Legarra and Reverter, 2018; Legarra and Reverter, 2019; VanRaden, 2021).

Recent updates to the Interbull validation software introduced new validation targets, alternative weighting schemes, and procedures for aligning evaluation scales between full and reduced datasets (Sullivan, 2023). Among these updates, the use of deregressed genomic estimated breeding values (dGEBV) as validation targets and the implementation of base-adjustment procedures were intended to reduce potential biases associated with genomic preselection and scale differences between datasets and evaluation time points. The software also allows, for exploration, the application of either the Interbull weighted regression approach or an unweighted regression framework similar to the Legarra-Reverter (LR) methodology described by Legarra and Reverter (2018, 2019).

Previous studies conducted in the United States demonstrated that genomic validation results are generally more stable for large dairy populations and highly heritable traits, whereas smaller populations and more complex traits remain challenging due to small numbers of validation animals yielding increased variability in validation statistics (Mota et al., 2023). More recently, the use of Truncated Multiple Across Country Evaluation (TMACE) evaluations as reference information improved validation consistency by reducing the effects of model and data changes accumulated over time and by allowing validation of traits implemented after the traditional four-year validation window (Mota et al., 2024).

Although the updated validation software provides several methodological options, limited information is available regarding the sensitivity of validation outcomes to different combinations of weighting schemes and base-adjustment procedures. Understanding the magnitude of these effects is important because validation results can influence the interpretation of genomic evaluation performance and, in some cases, may alter validation conclusions when statistics are close to predefined acceptance thresholds.

Therefore, the objective of this study was to evaluate the effects of weighting schemes and base adjustment on genomic validation outcomes in U.S. dairy cattle populations using the current Interbull GEBV validation framework. Emphasis was placed on comparing alternative validation scenarios and assessing the consistency of validation outcomes across methodological configurations.

Materials and Methods

Genomic validations were conducted using the Interbull GEBVtest software (gebvtest.py), corresponding to the version available from the Interbull Centre (Interbull, 2026). The software incorporates the most recent updates to validation statistics, candidate-bull definitions, and validation procedures described by Sullivan (2023).

The genomic prediction datasets in US dairy cattle populations were GEBV extracted from the August 2025 genomic evaluation (full dataset), which included MACE input, and truncated GEBV to the year of 2021 plus TMACE input (truncated dataset). TMACE follows the same principles as conventional MACE evaluations but uses truncated national information as input, thereby providing international information consistent with the reduced-data scenario (Jorjani and Dürr, 2011).

Five dairy breeds were included in the analyses: Holstein (HOL), Jersey (JER), Brown Swiss (BSW), Red Dairy Cattle (RDC), and

Guernsey (GUE). Nine traits representing production, health, longevity, and fertility were evaluated: milk yield (MIL), fat yield (FAT), protein yield (PRO), longevity (DLO), somatic cell score (SCS), clinical mastitis (MAS), heifer conception rate (HCO), cow conception rate (CC1), and daughter pregnancy rate (INT). Heifer conception rate was evaluated for all breeds except GUE, whereas clinical mastitis was evaluated for BSW, HOL, and JER, reflecting the breeds for which routine national evaluations were available.

All validation scenarios used predicted deregressed genomic estimated breeding values (dGEBV) as the validation target. Validation using GEBV-derived targets is easier to communicate to stakeholders and allows straightforward application in national evaluation systems that have adopted single-step methodologies (VanRaden, 2021). In addition, the Council on Dairy Cattle Breeding (CDCB) is currently transitioning toward single-step genomic evaluations with direct GEBV output, making dGEBV a more appropriate validation target for future implementation. Therefore, the validation target was kept constant across all scenarios so that the effects of weighting schemes and base adjustment could be evaluated independently.

Four validation scenarios were investigated (Table 1). The scenarios differed according to regression methodology and application of base adjustment. Scenarios 1 and 2 used the Interbull weighted regression approach (ITB), whereas Scenarios 3 and 4 used the unweighted LR regression approach described by Legarra and Reverter (2018, 2019). Base adjustment was not applied in Scenarios 1 and 3, while Scenarios 2 and 4 included base adjustment to align evaluation scales between full and reduced datasets as described by Sullivan (2023). Scenario 2 corresponds to the current official Interbull weighting and base-adjustment configuration, although dGEBV rather than EBV was used as the validation target in the present study.

Foreign bulls were included as candidate animals for the smaller breeds to increase the size of the validation population. Candidate and tested bull definitions followed the criteria implemented within the GEBVtest software. Validation outcomes were evaluated using the statistics provided by the software, including regression slope (b1), prediction accuracy (R^2), and overall PASS or FAIL status based on the official Interbull validation criteria. To focus on methodological effects rather than individual evaluation outcomes, breed-trait combinations are presented in aggregate form throughout the Results and Discussion section.

Table 1: Validation scenarios evaluated

S	*Target*	*Reg.*	*BA*	*Wt.*
S1	dGEBV	ITB	No	Yes
S2	dGEBV	ITB	Yes	Yes
S3	dGEBV	LR	No	No
S4	dGEBV	LR	Yes	No

S = validation scenario; **Reg.** = regression method; **BA** = base adjustment; **Wt.** = weighting applied; **ITB** = Interbull weighted regression; **LR** = unweighted regression approach; **dGEBV** = deregressed genomic estimated breeding value.

Results & Discussion

A total of 42 breed-trait combinations were evaluated under four validation scenarios, resulting in 168 scenario-specific validation analyses. Across all analyses, 91 (54.2%) resulted in a PASS outcome, whereas 18 (10.7%) resulted in FAIL, 12 (7.1%) resulted in WARN, and 47 (28.0%) resulted in HiSE (Table 2). WARN indicates validation statistics close to predefined acceptance thresholds, whereas HiSE indicates inconclusive validation results due to high slope standard errors. Most validation outcomes were therefore consistent with acceptable genomic validation performance, although a substantial proportion of analyses were classified as HiSE due to limited validation information available for some breed-trait combinations.

Overall, validation outcomes were generally consistent across scenarios, supporting

conclusions from previous studies that genomic validation performance is largely stable across traits and populations when sufficient information is available (Mota et al., 2023; Mota et al., 2024). Nevertheless, differences among scenarios were observed for a limited number of breed-trait combinations, indicating that choices can influence validation statistics and, in some cases, alter validation conclusions.

Table 2 : Validation outcomes across scenarios

Outcome	*Number of analyses*	*Percentage (%)*
PASS	91	54.2
FAIL	18	10.7
WARN	12	7.1
HiSE	47	28.0
TOTAL	168	100.0

Warn: validation statistics close to predefined acceptance thresholds; HiSE: high standard error that indicates an inconclusive statistical PASS due to high SE(b1) while FAILING the practical test: b1<0.8 or b1>1.2.

To illustrate the effect of alternative validation settings, Table 3 presents an example of a breed-trait combination for which validation outcomes differed among scenarios. Breed and trait identities are intentionally omitted to focus on methodological effects rather than individual evaluation outcomes.

Table 3: Example of validation outcomes under four validation scenarios

S	*Reg.*	*BA*	*Wt.*	*b1*	*Pass*
S1	ITB	No	Yes	1.17	PASS
S2	ITB	Yes	Yes	1.22	FAIL
S3	LR	No	No	1.15	PASS
S4	LR	Yes	No	1.20	PASS

S = validation scenario; **Reg.** = regression method; **BA** = base adjustment; **Wt.** = weighting applied; **ITB** = Interbull weighted regression; **LR** = unweighted regression approach; **dGEBV** = deregressed genomic estimated breeding value; b1 = regression slope.

The example demonstrates that alternative validation settings can produce different validation conclusions despite using the same underlying data and validation target. The only failing scenario corresponded to the ITB regression with base adjustment, whereas the same breed-trait combination passed under the remaining three scenarios. Although the differences in regression slopes were relatively small, they were sufficient to alter the final validation outcome because the statistics were close to the predefined acceptance thresholds. This observation highlights the sensitivity of genomic validation results to methodological choices when validation statistics are near decision boundaries. However, even scenarios classified as PASS but with statistics close to FAIL thresholds warrant further investigation.

The observed differences were not attributable solely to weighting schemes or base adjustment. Instead, the results suggest interactions among regression methodology, scale alignment, and validation-group definition. Sullivan (2023) reported that modifications to validation targets, candidate-bull selection, and base-adjustment procedures can affect the scale of reduced genomic evaluations and consequently influence regression statistics. The present results support those observations and demonstrate that alternative validation configurations may yield different outcomes even when the same genomic evaluations are analyzed.

A key observation from this study is that the effect of base adjustment depended on the regression framework being applied. Base adjustment altered the regression slope under both ITB and LR approaches; however, the resulting validation outcome was not identical across methodologies. This finding suggests that base adjustment should not be interpreted as an isolated factor but rather as one component of a broader validation framework. The combined effects of regression methodology, weighting strategy, and candidate-bull definition ultimately determine the validation statistics produced by the software.

These findings have practical implications for the interpretation of genomic validation results. PASS and FAIL outcomes are often viewed as definitive indicators of evaluation quality; however, the present study demonstrates that validation conclusions can be influenced by slight differences in configuration. Consequently, interpretation of GEBVtest results should consider the validation scenario employed, particularly when regression statistics are close to the acceptance thresholds. The results also emphasize the importance of transparency and consistency when comparing genomic validation outcomes across countries and evaluation systems.

The use of predicted dGEBV as the validation target aligns with recent recommendations for modern genomic evaluation systems and is particularly relevant given the ongoing transition of CDCB toward single-step genomic evaluations. As additional countries adopt single-step methodologies, validation procedures based on genomic predictions rather than traditional daughter-based measures may provide a more consistent framework for evaluating genomic performance. Continued evaluation of validation methodologies will therefore remain important to ensure that genomic validation procedures evolve alongside advances in genetic evaluation systems.

Conclusions

Validation outcomes were generally consistent across scenarios, although a limited number of analyses produced different conclusions depending on weighting scheme and base adjustment. The results indicate that genomic validation outcomes are influenced by interactions among methodological choices rather than individual components alone. Therefore, interpretation of exploratory GEBVtest results should consider the validation scenario employed. The use of dGEBV as the validation target provides a framework compatible with future single-step genomic evaluations. CDCB is currently transitioning toward single-step genomic evaluations with direct GEBV output, and future validation studies will assess whether this framework impacts scale consistency and genomic validation performance.

Acknowledgements and disclaimers

The authors acknowledge the support of Pete Sullivan of Lactanet (https://lactanet.ca/en/auteur/dr-pete-sullivan/) for providing an advance copy of the GEBV validation software, of participating dairy producers for data supply, Dairy Herd Improvement (**DHI**) organizations and Dairy Records Processing Centers (**DRPCs**) for processing and relaying the information to the CDCB. Purebred breed associations for providing pedigree data.

The mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by CDCB. The CDCB is an equal opportunity provider and employer.

References

Interbull Centre. 2021. Interbull CoP – Appendix VIII – Interbull validation test for genomic evaluations – GEBV test. https://wiki.interbull.org/public/CoPAppendixVIII?action=print&rev=48

Interbull. 2026. GEBVtest software. Interbull Centre, Uppsala, Sweden. Available at: https://interbull.org/ib/gebvtest_software

Jorjani, H., and Dürr, J. 2011. Truncated MACE for validation purposes. Interbull Technical Workshop, March 2011, Guelph, Canada.

Legarra, A., and Reverter, A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* 50:53.

Legarra, A., and Reverter, A. 2019. Predicting the accuracy of genomic breeding values. *Genet. Sel. Evol.* 51:55.

Mäntysaari, E., Liu, Z., and VanRaden, P. 2011. Interbull Validation Test for Genomic Evaluations. *Interbull Bulletin* 41:17-21

Mota, R.R., Nicolazzi, E.N., McWhorter, T.M. and VanRaden, P.M. 2023. Genomic validation software: USA results. *Interbull Bulletin*. 58:27–32

Mota, R.R., Sullivan, P., Nicolazzi, E., McWhorter, T.M., Legarra, A., and VanRaden, P. 2024. Genomic validation software: USA update including truncated MACE. *Interbull Bulletin* 60:20–21.

Sullivan, P.G. 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