Genomic validation software: USA update including truncated MACE

R.R. Mota[1], Peter Sullivan[2], E. Nicolazzi[1], T.M. McWhorter[1], Andres Legarra[1], P.M. VanRaden[3]

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

[2] Lactanet Canada, Guelph, Ontario, N1K 1E5, Canada

[3] U.S. Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350, USA

Abstract

With the need to establish a standardized method to validate genomic estimated breeding values (**GEBV**) to meet the requirements for marketing the semen of young bulls in Europe, the Interbull Centre has routinely added new features to the GEBVtest software. In 2023, the United States (**US**) conducted a GEBV validation and reported that large population breeds and traits with high heritability were more stable, whereas smaller populations and complex traits often failed due to various reasons. In addition, the use of Truncated MACE (**TMACE**)-based genomic evaluations was recommended to verify if this model would outperform 4-year-old official results. A new version of the GEBVtest software will become the standard for GEBV validation in 2024. The new version includes bootstrapping to improve and expand significance, with better tests for slopes, validation accuracy, and bias and does not allow bulls with GEBV foreign proof to be included as candidates. In this study, GEBV validation was performed using the newest version of the GEBVtest software while validating truncated domestic plus TMACE instead of using official US predictions from 4 years ago and applied to US dairy cattle populations. Nine traits were tested and GEBV from August 2023 were used as the full dataset, whereas TMACE-based GEBV were used as the truncated dataset. A TMACE-based model can accommodate model or data changes over time as well as a validation on traits that were not even implemented four years ago such as mastitis for more breeds implemented in 2020 and 2022 in the US. In general, the inclusion of TMACE improved results for all breeds. For Holstein, all traits passed validation except for one trait that failed with a slope ($\mathbf{b_1}$) of 1.31 (>1.2). The b_1 standard error was 0.02, which confirms an underestimation of this trait. In smaller breeds, a few other traits failed validation due to a $b_1 < 0.8$, but showed clear improvement of the b_1 by including TMACE. Finally, the smallest breeds showed several inconclusive passes and fewer failures compared to a previous study. The results may be due to the complexity of traits and the small number of candidate bulls. The use of TMACE-based genomic evaluations improves the validation test and is a tool to be considered as standard when performing GEBV validation, especially for smaller breeds.

Key words: bootstrapping, foreign information, model change, reference population

Introduction

With the need to establish a standardized method to validate genomic estimated breeding values (**GEBV**), the Interbull Centre in Uppsala, Sweden, https://interbull.org/index, has developed and routinely improved the GEBVtest software (Interbull, 2021; Mäntysaari et al., 2011). Recognizing that standardized software may be the easiest and most practical way to validate GEBV, the United States of America (**US**) conducted a genomic validation in 2023 using the software proposed by the Interbull working group. The software incorporated new features such as different validation targets, base adjustments, and a larger birth year window for candidate bulls (Mota et al., 2023a).

It was reported by Mota et al. (2023a) that large population breeds and traits with high

heritability were more stable, whereas smaller populations and complex traits often failed due to various reasons, such as the small number of candidate bulls, the slope (**b_1**) being more or less than expected from the standard error (**S.E.**), the b_1 upper biological limit being higher than 1.20, and the prediction accuracy (**R^2**) of parent average (**PA**) exceeding the GEBV with small sample sizes.

In addition, the use of extra regressions to assist other tests (VanRaden, 2021) and Truncated MACE (**TMACE**)-based genomic evaluations was recommended to verify if this model would outperform validation results using 4-year-old official breeding values. TMACE is a relatively new voluntary base service introduced by Interbull Centre, scheduled annually in October. This service aims to supply validation inputs for countries that include foreign bulls without domestic daughters in their reference population. It follows the same logic as conventional MACE but requires countries to use the current conventional model with the most recent four years of data truncated. The truncated EBV are then submitted to Interbull for a TMACE evaluation (Jorjani and Dürr, 2011). The use of a TMACE-based model in countries that blend their domestic with international evaluations, allows for accommodating model or data changes over time, in any and all countries participating in MACE and TMACE, as well as validating traits that were not implemented four years ago where neither domestic estimated breeding values (**EBV**) nor GEBV were available. In the US, this is the case for traits such as clinical mastitis in the Jersey and Brown Swiss breeds, which were implemented in 2020 and 2022, respectively (Norman et al., 2020; Mota et al., 2021; CDCB Connection, 2022; CDCB Connection, 2023; Mota et al., 2023b).

A year later, Interbull developed a new version of the GEBVtest software (**gebvtest.py**) that will become the official validation standard in December 2024. This new version adds bootstrapping to improve and expand significance testing, with better tests for slopes, validation accuracy, and bias. In addition, the software automatically excludes bulls from the validation group if they have type of proof = 24 in the truncated data, which denotes GEBV that included performance records of foreign daughters. This new edit prevents countries from using foreign bulls that are already progeny-proven in the truncated data, with no domestic daughters within the country but having daughters worldwide. If the genomic evaluation includes MACE data, as was the case of this study, the GEBV will include all daughters, both foreign and domestic. Moreover, if a wider birth year window is used to add more validation bulls with smaller data sets, it could become a more significant issue that the bulls with foreign daughters only from four years ago should be excluded from the validation test group, as will be the case with this new edit.

Therefore, a genomic validation was performed using the newest version of the GEBVtest software, validating truncated domestic plus TMACE instead of using official US predictions from 4 years ago, and applied to US dairy cattle populations.

Materials and Methods

To provide updated results on the US dairy cattle populations, a genomic validation was conducted using the gebvtest.py software version *gebvtest_2023C2.py*. A new version *gebvtest_2024A.py* is under testing and will likely become the official version.

The genomic prediction datasets in US dairy cattle populations were GEBV extracted from the August 2023 genomic evaluation (full dataset), which included MACE input, and truncated GEBV to the year of 2019 plus TMACE input (truncated dataset).

In this study, five breeds were evaluated: Holstein (**HOL**), Jersey (**JER**), Brown Swiss (**BSW**), Red Dairy Cattle (**RDC**), and Guernsey (**GUE**). Nine traits were tested: milk yield (**MIL**), fat yield (**FAT**), protein yield (**PRO**), longevity (**DLO**), somatic cell score (**SCS**), clinical mastitis (**MAS**), heifer conception rate (**HCO**), cow conception (**CC1**), and calving

interval (**INT**). All breeds were evaluated for all nine traits with one exception: MAS was only tested in HOL, JER, and BSW since the US had no evaluations for this trait for the RDC and GUE breeds.

In addition to using bootstrapping and the exclusion of bulls with GEBV foreign proof, the following parameters were applied as in Mota et al. (2023a): (1) Predicted deregressed GEBV (**dGEBV**) were used rather than the conventional deregressed EBV (**dEBV**) or daughter yield deviations (**DYD**). Validating using later GEBV is easier for the public to understand and allows national evaluations that have adopted single-step methods to apply the validation straightforwardly (VanRaden, 2021). This was done by using the option "*--target DGEBV*" from the software; (2) Base adjustments were applied to the GEBV and not EBV as conventionally done by using the option "*--baseadj GEBV*". The minimum birth year used was 2015, which reflects the current year of data (2023) minus eight years, as recommended by Interbull; (3) Foreign bulls were included as candidate bulls to increase the validation group size for the small breeds only: BSW, GUE, and RDC.

The criteria for candidate bulls are reported in Interbull (2021) and Mota et al. (2023a). The number of candidates bulls ranged from 9 to 3,277 depending upon the trait and breed evaluated (Table 1).

Results & Discussion

In general, the inclusion of TMACE improved results for all breeds (Table 1) compared to the results reported by Mota et al. (2023a) when official GEBV from August 2018 were used as truncated data. One of the main reasons for better US genomic validation results in the present study is that both the US and Canada participated in TMACE simultaneously.

For the HOL breed, a PASS was observed for all traits except HCO. This trait failed due to a b_1 of 1.31, higher than the upper biological limit (1.20). The b_1 standard error was 0.02, which confirms an underestimation of this trait. An important point to highlight is the trait MAS. As seen in Table 1, with the use of the TMACE methodology, a PASS was observed, whereas Mota et al. (2023a) reported a FAIL using official GEBV from 4 years prior due to the b_1 being higher than the biological limit. This is because model and data ingestion differences (Gaddis et al., 2020) between full and truncated GEBV used as input by Mota et al. (2023a) were overcome by the use of the TMACE methodology. In addition, the use of TMACE provided a less biased PA prediction (10% vs. 17%) and a b_1 within the biological interval of 0.80 and 1.20 (1.08 vs. 1.30). Another example of the benefits of TMACE is if the same data is used in this study as full but replaced the truncated data with official GEBV from 4 years ago (i.e., 2019), a FAIL for MAS continues to be observed (Table 2). After the implementation of MAS, the model was changed, and there was a significant effort to include much more data for this trait, which clearly impacted GEBV predictions over time. Therefore, current GEBV and those from four years ago are not directly comparable (Mota et al., 2023a).

The JER breed passed for most traits except the fertility traits of CC1 and INT (Table 1). There was a clear improvement in b_1 with the inclusion of TMACE compared to results reported by Mota et al. (2023a), as shown in Table 2. However, it was still not enough to pass the validation test due to high standard error. Fertility traits are under significant work to improve their predictions. As with HOL, MAS is again a noteworthy trait. This is because MAS evaluations for JER were implemented after 2018 (Norman et al., 2020; Mota et al., 2021), the year of the GEBV predictions used by Mota et al. (2023a) as truncated data input. Therefore, while Mota et al. (2023a) were not able to genetically validate this trait, the use of TMACE methodology allowed us to do so for this breed and trait combination, accounting for the current model in use and all data and model changes over time.

Table 1. GEBV Validation results for the five breeds evaluated in this study: Holstein, Jersey, Brown Swiss, Red Dairy Cattle and Guernsey

Holstein

Trait	*Bulls*	b_1 ±*S.E.*	R^2 *GEBV*	R^2 *EBV*	Pass
MIL	2,767	1.08±0.01	68	36	Yes
FAT	2,767	1.07±0.01	74	48	Yes
PRO	2,767	1.03±0.01	70	44	Yes
DLO	2,509	1.18±0.02	65	43	Yes
SCS	2,731	1.09±0.01	75	36	Yes
MAS	1,738	1.08±0.03	50	10	Yes
HCO	3,277	1.32±0.02	53	20	No
CC1	3,277	1.10±0.02	68	31	Yes
INT	3,277	1.03±0.01	65	27	Yes

Jersey

Trait	*Bulls*	b_1 ±*S.E.*	R^2 *GEBV*	R^2 *EBV*	Pass
MIL	486	1.07±0.03	79	51	Yes
FAT	486	1.05±0.03	75	44	Yes
PRO	486	1.02±0.03	76	51	Yes
DLO	435	0.86±0.05	41	36	Yes
SCS	481	1.09±0.04	63	37	Yes
MAS	222	0.81±0.15	13	12	Yes
HCO	516	0.98±0.08	27	10	Yes
CC1	445	0.83±0.05	40	28	No
INT	480	0.81±0.04	45	32	No

Brown Swiss

Trait	*Bulls*	b_1 ±*S.E.*	R^2 *GEBV*	R^2 *EBV*	Pass
MIL	71	0.86±0.07	66	46	Yes
FAT	71	0.77±0.08	54	31	No
PRO	71	0.82±0.08	60	45	Yes
DLO	63	0.73±0.15	33	18	hSE
SCS	69	0.74±0.10	39	32	No
MAS	NA				
HCO	75	1.04±0.20	22	6	Yes
CC1	63	0.93±0.13	35	27	Yes
INT	71	0.84±0.11	43	31	Yes

Red Dairy Cattle

Trait	*Bulls*	b_1 ±*S.E.*	R^2 *GEBV*	R^2 *EBV*	Pass
MIL	18	0.68±0.15	43	43	hSE
FAT	18	0.83±0.22	55	57	Yes
PRO	18	0.75±0.16	52	53	hSE
DLO	9	0.59±1.19	5	13	hSE
SCS	18	0.90±0.51	16	30	Yes
MAS	NA				
HCO	16	2.18±0.84	30	5	hSE
CC1	16	-0.21±0.8	1	4	hSE
INT	17	-0.04±0.5	4	0.1	No

Guernsey

Trait	*Bulls*	b_1 ±*S.E.*	R^2 *GEBV*	R^2 *EBV*	Pass
MIL	16	0.87±0.25	35	8	Yes
FAT	16	0.31±0.33	5	4	hSE
PRO	16	0.08±0.54	0.3	0.1	hSE
DLO	NA				
SCS	16	1.79±0.59	41	22	hSE
MAS	NA				
HCO	NA				
CC1	12	2.19±0.58	68	77	No
INT	16	1.87±0.31	70	52	No

MIL: milk yield; FAT: fat yield; PRO: protein yield; DLO: longevity; SCS: somatic cell score; MAS: clinical mastitis; HCO: heifer conception rate; CCR: cow recycling; INT: calving interval; hSE: high standard error; EBV/GEBV R^2 are expressed in %.

Table 2. Comparison of three genomic validation scenarios (S) using different truncated data as input

S	*Bulls*	b_1 ±*S.E.*	R^2 *GEBV*	R^2 *EBV*	Pass
Brown Swiss - INT					
S1	88	0.64±0.16	16	27	No
S2	77	1.30±0.19	40	40	hSE
S3	71	0.84±0.11	43	31	Yes
Jersey - INT					
S1	588	0.79±0.03	47	31	No
S2	500	0.71±0.04	32	32	No
S3	480	0.81±0.04	45	32	No
Holstein - MAS					
S1	2,379	1.30±0.03	40	17	No
S2	1,548	0.60±0.05	9	10	No
S3	1,738	1.08±0.03	50	10	Yes
Jersey - MAS					
S1		NA			
S2		NA			
S3	222	0.81±0.15	13	12	Yes

S1: 2022-2018 official GEBV; S2: 2023-2019 official GEBV; S3: 2023 official GEBV and 2019 truncated MACE; INT: calving interval; MAS: clinical mastitis; hSE: high standard error.

The BSW breed had a FAIL for FAT and SCS due to a $b_1 < 0.8$, an inconclusive PASS test for PRO, but also with a $b_1 < 0.8$, and a PASS was observed for the other traits (Table 1). The S.E. of b_1 were much larger than those observed for the aforementioned larger breeds HOL and JER. This is likely linked to the much smaller number of candidate bulls and the fact that more than 50% of the BSW reference population in the US is composed of foreign bulls, primarily from Switzerland and France (Mota et al., 2023b). These foreign bulls likely have daughters outside the US earlier than within the country. So, even if the effective daughter contribution (**EDC**) in the US is equal to zero when truncating the data, this is not true worldwide, making these bulls ineligible as candidate bulls.

As seen in Table 1, there are no results for MAS for the BSW breed. This is because there were no bulls to validate MAS at this time. However, if there were candidate bulls, this trait would have been validated using the TMACE methodology, even if it is a trait recently implemented by the Council on Dairy Cattle Breeding (**CDCB**), i.e., in 2022 (CDCB Connection, 2022; CDCB Connection, 2023; Mota et al., 2023b).

Finally, the smaller breeds RDC and GUE showed several inconclusive passes and far fewer failures compared to the results reported by Mota et al. (2023a), as shown in Table 1. These results may be due to the complexity of traits and the small number of candidate bulls.

For RDC, a PASS was observed for FAT and SCS, even though the R-squared for the GEBV is smaller than for PA. This indicates a high S.E. for the R^2 test and it cannot be concluded that the R^2 is significantly lower with the GEBV ($P > .05$). Numerical differences are often due to sampling bias, making the R^2 too high for the PA, and the (lower) R^2 for GEBV is actually more reasonable. A FAIL in this case is likely an unreliable decision because the small sample inflation of R^2 for PA are ignored and combined with a reasonable alignment of model R^2 with truncated GEBV and the corresponding genomic reliabilities. With the new software, the R^2 test is still applied as additional information, but it causes an overall FAIL in the GEBV test if the R^2 is significantly lower ($P < .05$) for the GEBV than it is for the PA.

The GUE results here were very similar to RDC with several inconclusive passes, mostly due to the small number of candidate bulls (Table 1). DLO and HCO had no bulls to apply a genomic validation for GUE, whereas MAS is not implemented for either of these breeds, RDC and GUE.

Significance testing of the validation slope parameter was theoretically improved with the implementation of bootstrapping in the updated validation test software. In practice, the

bootstrap tests were very similar to t-test results applied in earlier versions of the software (e.g. Mantysaari et al, 2011), as verified by both Canada (Table 3) and similarly by USA in the present study. Several new tests were also added which make use of the full posterior probability distributions now available from bootstrap samples. For example, the software now includes new significance tests for bias in top young genomic bulls specifically, for the average bias across all young genomic bulls, and additionally the new tests described earlier for significance of R^2 improvements due to genomics.

In summary, the use of TMACE-based genomic evaluations improves the validation test and is a tool to be considered as standard when performing genomic validation, especially for smaller breed populations.

Table 3. Estimated t-values based on bootstrap samples (Boot.) versus the standard t-test, for *n* traits by breed in Canadian research data described by Sullivan (2023)

Breed	*Value*	*Mean*	*SD*	*Min*	*Max*
RDC	Boot.	-0.47	1.43	-3.7	2.4
n=32	t-test	-0.56	1.54	-4.1	2.7
JER	Boot.	-0.58	1.89	-5.6	3.4
n=30	t-test	-0.63	1.93	-5.3	3.3
HOL	Boot.	-1.31	6.55	-18.0	9.1
n=36	t-test	-1.58	7.05	-18.1	9.5

Conclusions

Countries must ensure they use candidate bulls with no daughters, domestic or foreign, from four years ago. The tests continued to fail for smaller breeds and less heritable traits due to b1 underestimation, the biological interval being between 0.80 and 1.20, and/or not enough bulls to validate. As in CAN, bootstrapping provided trivial differences to the results. TMACE resulted in a fairer test, and countries are strongly encouraged to use Truncated MACE service when performing GEBV validation, especially those with small populations. Finally, for maximum benefit, it is recommended that groups of countries share their genotypes to all participate in TMACE simultaneously.

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 association 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

CDCB Connection. 2022. https://mailchi.mp/uscdcb.com/cdcb-connection-february-13509112?e=d1116e694c

CDCB Connection. 2023. https://mailchi.mp/uscdcb.com/cdcb-connection-february-13647400?e=790a193632#mctoc1

Gaddis, K.P., VanRaden, P.M., Cole, J.B., Norman, H.D., Nicolazzi, E. and Dürr, J.W. 2020. Symposium review: Development, implementation, and perspectives of health evaluations in the United States. *J. Dairy Sci.* 103: 5354-5365.

GEBV Test, https://interbull.org/ib/cop_appendix8

Interbull Centre. 2021. Interbull CoP – Appendix VIII – Interbull validation test for genomic evaluations – GEBV test. https://interbull.org/ib/cop_appendix8

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

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

Mota, R.R., Nicolazzi, E., Megonigal, J.H., Gaddis, K.P., Dürr, J.W. and VanRaden, P.M. 2021. Use of international clinical mastitis data as independent trait in the US evaluation system. *Interbull Bulletin*. 56:11-16

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

Mota, R.R., Parker Gaddis, K., VanRaden, P.M., McWhorter, T.M. and Nicolazzi, E.N. 2023b. International evaluations for clinical mastitis in Brown Swiss. *Interbull Bulletin*. 59:182–186

Norman. H.D., Gaddis, K.P. and Jensen, L. 2020. CDCB connection. https://uscdcb.com/wp-content/uploads/2020/01/CDCB-Jersey-Health-Traits.pdf

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