Improved genomic validation with extra regressions

Paul M. VanRaden

USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD, USA.

paul.vanraden@usda.gov

Current and future validation options

- **Tests I, II, and III** check for bias in proven bull genetic **trend**

- **Test IV** checks for stability of EBV **variance** (MS)

- **Genomic** validation checks if young bull GEBV match later DYD

- **Future options for genomic validation:**

  - **Use later GEBV or deregressed GEBV instead of later DYD or deregressed EBV as dependent variable (better for ssGBLUP)**

  - **Add regression on age to test if genetic trend in young bulls changes when they later add daughter records**

  - **Add other regressions such as for PA or inbreeding**

Dependent variables and weighting

- **Prediction of new data should account for selection in old data**
  - **Conventional DYD excludes genomic info (could be biased)**
  - **New GEBV contains info from old GEBV (not independent)**
  - **Deregressed GEBV = GEBVold + (GEBVnew – GEBVold) / Rdif**
  - **Similar to deregressed EBV = PA + (EBV – PA) / Rdif**
- **Weighting based on difference in reliability**
  - **Compute newEDC – oldEDC and convert difference to Rdif**
  - **Weight bull dGEBV by Rdif (no weight if no reliability gain)**

USDA

Information added after truncation

- **Gains in reliability come from added:**

  – **Parent** info and higher REL(PA)

  – **Genomic** info from larger reference population

  – **Own** records for cows

  – **Daughter** records as summarized in DYD

  – **Granddaughter** records (may be > daughters but not in DYD)

- **Validation could test the sum of all changes in GEBV**

- **Requires EDC difference from all sources as a function of REL**

Data and methods used

- **Official GEBVs for 3,504 HO bulls with daughters in >10 herds in Dec 2020 but none in Dec 2016**

- **Earlier GEBVs were adjusted to the later genetic base**

- **Predict later dGEBV to get model R^2, or predict the changes (later dGEBV – earlier GEBV) to get residual R^2**

  - **High model R^2 values are desired**

  - **Low residual R^2 values are desired if the extra regressions do not contribute significantly (GEBV changes not predictable)**

  - **Large t test values (> ±3) indicate significant regressions**

USDA

Which dependent variable to validate?

Dependent variable	B1	T tests (significance)		R^2 (%)	
		B1[1]	B0	Residual	Model
DYD	1.03	1.9	-14.4	0.10%	59%
dGEBV	0.99	-0.9	-16.5	0.02%	69%
GEBV	0.99	-0.8	-16.3	0.02%	72%

[1]Test of B1 difference from expected 1.0

USDA

Genomic validation with just 1 or extra regressions

B1	T tests (significance)					R^2 (%)	
	B1[1]	Age(yr)	PA	EFI[2]	B0	Residual	Model
0.99	-0.9				-14.4	0.02	69
1.08	7.1	-19.1			-25.7	9	72
1.43	19.4		-22.6		-14.7	13	73
1.00	-0.2			-4.6	3.4	1	69
1.47	20.0	-14.8	-18.5	-3.1	1.0	18	77

[1]Test of B1 difference from expected 1.0
[2]Expected Future Inbreeding for bull in 2016

USDA

B1	T tests (significance)				R² (%)	
	B1	Birth yr	PA	B0	Residual	Model
0.99	-0.7			-2.6	0.03	60
1.06	2.8	-10.0		10.0	6	62
1.29	8.5		-10.9	1.7	7	63
1.39	11.1	-11.1	-12.0	11.1	14	65

Provided by Pete Sullivan, LactaNet

Model R^2 for CAN and USA differ due to edits and bulls included,
Dependent variable was later GEBV rather than dGEBV.

USDA

Results for USA milk

- **Regression B1 was near 1.0 in standard genomic validation**

- **Three extra regressions were each significant:**

  - **Later GEBV declined for the youngest bulls, those with highest PA, and those with higher EFI**

  - **B1 became much >1.0 when those were added**

  - **Model R^2 increased to 77% from 69% with all 3 terms added**

- **Blending DGV with PA reduced B0 bias but also reduced the R^2**

- **The EFI decline was more than expected from F depression**

USDA

Genomic validations for other USA traits

Trait	B1	T tests (significance)					R² (%)	
		B1	Age(yr)	PA	EFI	B0	Residual	Model
Protein	1.47	17.5	-14.4	-16.3	-3.1	1.8	13	73
Fat	1.40	18.7	-12.9	-18.6	-0.8	-0.7	16	75
SCS	1.22	11.6	1.4	-3.7	-4.1	-4.6	6	73
PL	0.89	-3.8	-6.9	4.1	-6.1	4.0	3	53
DPR	0.91	-3.9	-14.1	3.8	-0.8	-2.4	6	58
CCR	1.15	7.0	-11.3	-3.9	-1.5	-2.0	4	60
HCR	1.10	3.3	-6.0	6.0	-6.5	4.4	5	48

USDA

Conclusions

- **Validation could use published GEBV or dGEBV and ssGBLUP**

- **GEBV predictions are simple to explain but not independent**

- **Later deregressed GEBVs are independent of earlier GEBVs**

- **Extra regressions can show which bull groups change and why**

- **Trend differences may reflect inflation of the youngest bull GEBV or preselection bias in proven bull GEBV**

- **Models may need revision to balance accuracy and bias**

USDA

Acknowledgments

- **Interbull working group on genomic validation:**

  - **Esa Mäntysaari, Raphael Mrode, Zengting Liu, Pete Sullivan, Paul VanRaden**

- **Interbull staff for support**

- **CDCB staff and industry cooperators for data**

- **AGIL staff and USDA funding of project 8042-31000-002-00-D, "Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals"**