Session II

Validation
ITB Working group for validation tests of genetic and genomic evaluations

Working Group was assigned 2018

**Members**
Esa Mäntysaari
Zengting Liu
Peter Sullivan
Raphael Mrode
Paul VanRadem
Valentina Palucci
The need for EBV/GEBV validation tests is still important:

1) Inclusion of national genetic evaluations to MACE
   • the trend tests II and III and the MS trend test

2) Inclusion of genomic evaluations to GMACE (GEBVtest)
   • Unbiasedness, inflation

3) Certification of genomic evaluations (GEBVtest)
   • Unbiasedness, inflation, and accuracy
   • To be used in international trade; required also by EU
Trend validation tests II, III, and the MS variance test

Applicability of the tests has not changed in genetic evaluations of countries that are NOT using genomics in selection.

In countries applying genomic selection:

1) **tests for EBV:**

- All the trend tests apply as before, except that the power has been reduced because the sires are used only a short time.
- Especially Test III has lost its power.
- Therefore the test II, based on DYD, should be preferred, but the DYD for the test should be approximated rationally.

In the case of efficient GS - the EBV are likely to fail the tests!
Distribution of sire birth years in DFS Holsteins calving first time year 2014
Distribution of sire birth years in DFS Holsteins calving first time year 2018
Distribution of sire birth years in DFS Holsteins calving first time year 2020
Number of tests submitted to Interbull during last 3 years (TRAIT MILK)

<table>
<thead>
<tr>
<th>Tests applied</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test II</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Test III</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>GEBVtest</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Very difficult to conclude anything about the changes in results

Valentina Palucci, Interbull 2023
2) **tests to be applied on GEBV**

**Countries own interest:**
- Validation test II can be used for single-step evaluations
- Validation test III can be modified to use GEBVs and $R^2$ of GEBV

**International evaluations (for exporting semen, or for GMACE input)**
- Interbull GEBV test
  - Tests how well the early GEBV predicts (recent) progeny based EBVs
  - To pass: regression coefficient $b1 \sim 1.0$ and $R^2_{GEBV} > R^2_{PA}$
Number of test submitted to Interbull during last 3 years (TRAIT MILK)

<table>
<thead>
<tr>
<th>Tests applied</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test II</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Test III</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>GEBVtest</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
GEBVtest example:
$R^2$ results of countries having applied test >2 times (HOL protein)
Number of test submitted to Interbull during the last 3 years (TRAIT MILK)

<table>
<thead>
<tr>
<th>Tests applied</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test II</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Test III</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>GEBVtest</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Age of the last accepted GEBVtest in 13 countries:

- 4 countries have 8 years old results.
Weak points in the current GEBVtest

Can we assume that the DYD or DRP from the EBV evaluations are unbiased?

• Genomic evaluations appear biased if the recent EBV of the validation bulls is lower than the GEBV
• Genomic evaluations appear inflated if the recent EBV has less variance than the variance in GEBV
• Genomic evaluation accuracy ($R^2$) is under-estimated if EBV are biased
2) tests to be applied on GEBV:

GEBVtest should not assume EBV evaluations to be a "golden standard"

Better approaches:

• Testing using two consecutive GEBVs: \( \text{GEBV}_{\text{red}} \) and \( \text{GEBV}_{\text{full}} \)
• Test of changes in GEBV i.e. deregressed GEBV (dGEBV)

To standardize the testing, these new options were included into current GEBVtest program: GEBVtest2022.py
Thank you!

Ja kiitos hyvistä vuosista!

Acknowledgements:
Valentina Palucci compiled the statistics of current tests
Validation working group
Peter Sullivan for the new version of GEBVtest2022
NAV provided HOL calving data
theory & actual experiences with method LR for validation tests

Andrés Legarra, andres.legarra@uscdcb.com

Interbull Technical Workshop, Rome, 14-15 Feb 2023
• Intro & why
• how does it work (theory & practicalities)
• examples of output
Motivations

• Esa Mantysaari invited me to give this talk (before 😊 he knew that I now work for CDCB)

• I show theory and experiences that we got from the use of the LR method (mainly in dairy sheep from France)

• There is a tutorial:
  http://genoweb.toulouse.inra.fr/~alegarra/SMARTER_D5.3_Use_of_method_LR.pdf

• and in the course notes:
  http://genoweb.toulouse.inra.fr/~alegarra/GSIP.pdf
Comparing models, history

- In 1994 Toni Reverter (then in Colorado) and Didier Boichard (INRA) attacked the problem of checking bias in BLUP evaluations

- Reverter focused on “Hendersonian » properties of BLUP:
  - comparing successive evaluations
  - more theorems

- Boichard focused on dairy cattle problems and tools:
  - comparing first-lactation and all-lactation
  - check of DYDs
  - comparing successive evaluations
Why method LR

• Later we had the explosion of cross-validation due to Genomics
• My own work in species with small offspring size (pigs, sheep)
• In these cases DYDs do not exist or... getting DYDs right is very hard!!
  • are dams well estimated? are contemporary groups well estimated? ...etc
• Should we trust DRP?
  • I don’t trust Reliabilities (approximations)
  • Equations for DRPs are often also approximations (even if Reliabilities are correct)
  • I trust (G)EBVs
Why method LR

• We wanted something more general
• Toni and I made mutual visits
• We re-derived his 1994 paper to apply to “multiple individual” case (the 1994 one considered individual EBVs)
• In this manner we derived more general equations that explicitly show relationships, Prediction Error Covariances, and also effects of selection
• We also focused on WHICH properties we want to check and WHY
Metrics

• Theory of quantitative genetics suggest using Metrics from linear regression of $u$ (TBV) on $\hat{u}$ (EBV) for a *collection* of animals (and hence the vectors)

• Bias: $\Delta = \frac{1}{n}(1'\hat{u} - 1'u)$ (it is NOT the intercept of the regression of $u$ on $\hat{u}$)

• Slope: $b = \frac{Cov(u,\hat{u})}{Var(\hat{u})}$ (slope of the regression of $u$ on $\hat{u}$)

• Accuracy: $r = \frac{Cov(u,\hat{u})}{\sqrt{Var(u)Var(\hat{u})}}$

In fact: $MSE = \frac{1}{n} \sum_i (u_i - \hat{u}_i)^2 = \Delta^2 + Var(u)\left(1 + \frac{r^2}{b^2} - \frac{2r^2}{b}\right)$

• Why are these relevant? Genetic progress !!
True and estimated genetic progress

- When we select animals, we believe our $\Delta G = \frac{1}{n} \Sigma (GEBV) = \tilde{u}$ of selected animals
- This only holds if bias $\Delta = 0$, regression $b = 1$
- $\Delta > 0$ (bias) or $b < 1$ (overdispersion) lead to too high values of selected young animals
- Both need to be checked
  - $\Delta > 0$ implies that all animals were overevaluated (regardless of $b = 1$)
  - $b < 1$ implies that selected animals are overevaluated (regardless of $\Delta = 0$)
Genetic gain: $\Delta$

Consider a Genetic Evaluation

**Ideal situation**

- Young animals EBVs should lie around the true genetic mean of their generation.

**If bias**

If the EBVs are biased, all animals are now underestimated (for example).

**Selection rule**

- Young animals EBVs should lie around the true genetic mean of their generation.

**True Genetic gain**

- Old animals: Good accuracy.

**Year of birth**

- If bias:
  - You don't select as many young animals as you should.

**Old animals**

- Good accuracy.

**Young animals**

- Bad accuracy.

**EBV**

- Bias $\Delta$
Dispersion: b
Consider a Genetic Evaluation

Young animals EBVs should have the right dispersion
Correct mean after selection
True Genetic gain
Old animals
Good accuracy

Year of birth
EBV

Ideal situation

If we have too much dispersion...
The genetic gain after selection of young animals is overestimated
Good dispersion if
\[
\text{reg}(\text{TBV} \sim \text{EBV})
\]
Slope = 1

Selection rule

Dispersion b

If problems in dispersion

Good accuracy
Bad accuracy
Young animals

Correct mean after selection

The genetic gain after selection of young animals is overestimated

If we have too much dispersion...
Bias + overdispersion

Old animals
Good accuracy

Young animals
Bad accuracy

Selected animals:
Well placed. Not too dispersed.

Too high and too dispersed
We don’t have true EBVs

yet Legarra & Reverter (2018) proposed a new method based on comparisons of EBV from “partial” (old) data vs “whole” (old+new) data.

• Does not require “true” breeding values
• Does not require pre-corrected phenotypes
• Could be used for any kind of traits

Check of bias using successive evaluations

• We proved (analytically) that in successive genetic evaluations there are useful statistical properties of the JOINT distributions of “early” (partial) and “late” (whole) EBVs
• We use these properties to get estimators of biases and accuracies
• Some details in the paper referring to accuracies (not to biases) were later refined
  • series of papers by Macedo et al. and the Tutorial
  • don’t worry about them today
The proposed method LR
How does LR method work?

- Pedigree (markers)
- Records

- 2005
- 2010
- 2015

- (ssG)BLUP with Partial (old) data
- (ssG)BLUP with Whole (old+new) data

- (G)EBVs ($\hat{u}_p$) of “Focal group” e.g. Young males without daughters
- (G)EBVs ($\hat{u}_w$) of “Focal group” e.g. Same males with daughters

Estimators of Bias Slope Accuracies
Estimators of LR method: Bias and Slope

**Bias**
\[ \hat{\Delta}_p = \tilde{u}_p - \tilde{u}_w \]
Expected value of 0 in absence of bias

**Slope**
\[ \hat{b}_{w,p} = \frac{cov(\tilde{u}_p, \tilde{u}_w)}{var(\tilde{u}_p)} \]
Expected value of 1 in unbiased genetic evaluations
\[ \hat{\Delta}_p = \frac{999 + 849 + 831 + 953 + 764}{5} - \frac{973 + 833 + 904 + 963 + 807}{5} = -16.8 \]

\[ \text{EBV2018}=c(999,849,831,953,764) \]
\[ \text{EBV2019}=c(973,833,904,963,807) \]
\[ \text{delta}_p=\text{mean}($\text{EBV2018}$)-\text{mean}($\text{EBV2019}$) = -16.8 \]
\[ \hat{\Delta}_p \]
\[ \text{aa} = \text{lm}(\text{EBV2019} \sim \text{EBV2018}) \]
\[ \hat{b}_{wp} = \text{aa$\$coefficients}[2] \approx 0.71 \]
Some of the algebra

\[
\text{Var}\left( \begin{pmatrix} \hat{u}_p \\ \hat{u}_w \end{pmatrix} \right) = \begin{pmatrix} G - C_p^{uu} & G - C_p^{uu} & G - C_p^{uu} \\ G - C_p^{uu} & G - C_w^{uu} & G - C_w^{uu} \\ G - C_p^{uu} & G - C_w^{uu} & G \end{pmatrix}
\]

\[
E\left( \frac{1}{n} ( \hat{u}_p - \bar{u}_p )' ( \hat{u}_w - \bar{u}_w ) \right) = \frac{1}{n} \text{tr} \left( SG - SC_p^{uu} \right)
\]

\[
= \frac{1}{n} \text{diag}(G) - \bar{G} - \left( \text{diag}(C_p^{uu}) - (C_p^{uu}) \right)
\]

Regression of whole on partial

after algebra on expectation of quadratic forms, \( E(\hat{b}_{w,p}) = 1 \)

\[
\hat{b}_{w,p} = \frac{\text{cov}(\hat{u}_p, \hat{u}_w)}{\text{var}(\hat{u}_p)} = \frac{\frac{1}{n} ( \hat{u}_p - \bar{u}_p )' ( \hat{u}_w - \bar{u}_w )}{\frac{1}{n} ( \hat{u}_p - \bar{u}_p )' ( \hat{u}_p - \bar{u}_p )}.
\]

- expressing everything as quadratic forms allows to derive properties.
- e.g. the s.e. of \( \hat{b}_{w,p} \) (slope of “whole” on “partial”), will be a function of
  - (1) number of individuals
  - (2) their relationships
  - (3) their PEV and PEC.
Practicalities: defining focal groups

• The properties of the method hold and are useful for a group of animals that are contemporaries and have
  • similar selection pressure
  • similar information at "partial" (e.g. only Parent Average or PA+genomics)
  • and similar information at “whole” predictions (e.g. Parent Average +genomics + phenotype, or Parent Average + genomics + offspring, or...)

• we call this “focal group”

• e.g.
  • young genomic bulls vs. same bulls with daughters
  • 1st-calving cows vs. same cows at 2nd-calving
Practicalities: defining “whole” and “partial”

• You can do many “partials” and many “wholes”
• for instance you can do “partial” at 2010, 2011,...
• and compare each of them vs. “whole” at 2014, 2015...
• it is important to do several comparisons !!
Evaluations with data until 2005, until 2006 and so on until 2017.

We compare

- **EBVs at birth** (EBV<sub>P</sub>) of a set of Artificial Insemination males (2005 – 2014)
- **EBVs of the same males in later evaluations** (after having progeny) (EBV<sub>W</sub>) (until 2017).

For example for males born in 2005, 11 pairs of evaluations were analysed,

- 2005 vs 2007; 2005 vs 2008; ... and 2005 vs 2017

The same for males born in 2006, 2007 ... 2015

- 2006 vs 2008; ... and 2015 vs 2017

Total of 65 comparisons that we “averaged” using a linear model to account for unbalance
(details in the Macedo et al. paper)
Practicalities: referring to same genetic base

• in addition to the slope $\hat{b}_{w,p}$ of the regression $\hat{u}_w \sim \hat{u}_p$

• we do need to check bias in the genetic trends $\Delta_p = \bar{u}_p - \bar{u}_w$

• In genetic evaluations with Unknown Parent Groups, the EBVs are not estimable functions

• So you need to refer EBVs in “whole” and in “partial” to the same genetic base in order to infer “bias”

• Typically the genetic base is something like “average EBV of all females born in 2010” or something like that.
Some results

• Dairy sheep improvement is a French specialty!
• This is **not** meat or wool sheep
• Very well structured co-operative “mini – dairy cattle style” breeding program
• AI, performance recording, etc etc
  <2015: progeny-testing (first crop ~30 daughters)
  >2015: genomic selection

• All results here concern Milk Yield
The breeds

**Manech Tête Rousse**

Female population size: 274,000
Females in the breeding flocks: 80,260 (29%)
Tested rams per year: 150
Rams at AI Center: 600
Individuals in the pedigree: 540,999
Number of records Milk Yield: 1,842,295
Missing pedigree: ≈25%

Traits selected: Milk Yield and contents

**Lacaune**

Female population size: 890,000
Females in the breeding flocks: 174,472 (19%)
Tested rams per year: 440
Rams at AI Center: 1400
Individuals in the pedigree: 1,868,975
Number of records Milk Yield: 5,696,348
Missing pedigree: ≈9%

Traits: Milk Yield and contents, SCS, Udder traits
Main results Manech Tête Rousse

Each focal group has ~150 rams.


So we get 3 values of $\hat{b}_{w,p}$.

Rams born in 2013.
Main results Manech Tête Rousse

each focal group has ~150 rams

Important variation between truncation points!
Very small bias in Manech Tête Rousse!
slope $\hat{b}_{w,p} \sim 1$
small, positive bias $\hat{\Delta}_p$ (0.2 genetic s.d.)
lambs are over-predicted

### Across models
slope $\hat{b}_{w,p}$:
MF performs better

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\Delta}_p$</th>
<th>$\hat{b}_{w,p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLUP-MF</td>
<td>0.25</td>
<td>0.98</td>
</tr>
<tr>
<td>BLUP-UPGA</td>
<td>0.48</td>
<td>0.96</td>
</tr>
<tr>
<td>SSGBLUP-MF</td>
<td>0.23</td>
<td>0.97</td>
</tr>
<tr>
<td>SSGBLUP-UPGA</td>
<td>0.32</td>
<td>0.94</td>
</tr>
<tr>
<td>SSGBLUP-UPGH</td>
<td>0.48</td>
<td>0.88</td>
</tr>
</tbody>
</table>

### Some models (UPGH...) are really biased
Results in Lacaune

• Slopes $\hat{b}_{w,p}$ of several cohorts of ~400 AI rams born in 2000...2016 vs. several “whole”
Results in Lacaune

- Slopes $\hat{b}_{w,p}$ of several cohorts of ~200 AI rams born in 2015...2019 (partials at 2015...2019) vs. a single “whole” at 2021

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alone</td>
<td>0.82</td>
<td>0.77</td>
<td>0.90</td>
<td>0.66</td>
<td>0.81</td>
</tr>
<tr>
<td>TogetherSameMF</td>
<td>0.90</td>
<td>0.87</td>
<td>0.93</td>
<td>0.83</td>
<td>0.89</td>
</tr>
<tr>
<td>TogetherDifferentMF</td>
<td>0.89</td>
<td>0.87</td>
<td>0.93</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td>Indirect</td>
<td>0.08</td>
<td>0.28</td>
<td>0.09</td>
<td>0.06</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Marine Wicki
PhD student
INRAE, Toulouse
Variation in estimates

• There’s a LOT of variation
• Properties of LR method (and of Interbull methods) rely on asymptotics
• strictly speaking $\Delta_p \to 0$ and $\hat{b}_{w,p} \to 1$ for $n \to \infty$
• so $\hat{b}_{w,p} = 1$ either for one year on a large sample of bulls or averaging small series across years
• The property of unbiasedness is a property of the prediction method as a whole (across years and animals), not of every single batch of young bulls.
• Tampering the model to avoid point biases could create further biases down the road.
• The s.e. of $\hat{b}_{w,p}$ depends on more things than the number of bulls (we can derive that)
Sources of variation

• I see two sources
• The cohort of focal individuals may be not what we expected, i.e. they were themselves biased or not representative: $\bar{u}_p \neq \bar{u}$
• Their further evaluations were not correct, i.e. $\hat{u}_w(t)$ was more “correct” in some years ($t$) than another years
• e.g. a value of $\hat{b}_{w,p} < 1$ may indicate overdispersion of $\hat{u}_p$ but also underdispersion of $\hat{u}_w$
• For instance this can happen if a fixed effect interacts with genetic trend in a complicated manner (I’m thinking UPGs)
• My personal suggestion (in dairy sheep) is to check models using averages across years
What if my model is already wrong?

• The LR theory assumes that the model is correct!!
• Can we verify if a model is correct if the model is not correct?
• Fernando Macedo explored that
By simulation

- Software: QMsim, Blupf90 family and our own
- 20 replicates of a “dairy” population
- 10 generations
- Two heritabilities (0.1 and 0.3)
- Three scenarios

---

**Correct Model**

Genetic evaluations performed with correct heritabilities and effects

**Wrong Heritability**

Using higher (+0.05) and lower (-0.05) heritabilities in the evaluation model

**Environmental trend not (well) accounted for**

Simulate an environmental trend.
Fit contemporary groups either as fixed, or as random heavily shrunken to 0.
Main Results: the correct model

There was no surprise with the right model. Bias, slope, and accuracies were well estimated.
Main Results: the wrong heritabilities

The slope was estimated but with low precision

Accuracies were well estimated

True bias was generated
LR method could estimate the good direction but not the magnitude
Main Results: the environmental trend

It was impossible to estimate the Bias, neither fitting CG as fixed nor as random effect.

The slope was poorly estimated. The estimation was better when CG were fit as fixed effects.

In general, accuracies were well estimated.
Main conclusion

The LR method can estimate the Bias, Slope, and Accuracies when the genetic evaluation model is robust, even if not perfect.

When the model is really wrong, the estimates from LR method are unreliable.
How does all this compare to Interbull test?

• I believe that if DRP/DYDs are correctly computed, Interbull $b1$ is our “slope” $\hat{b}_{w,p}$

• Interbull doesn’t check pure averages of EBVs ($\Delta_p$)

• In my view the logic of Interbull tests is “early=PA”, “late=progeny-tested” so $Rel_{PA} \ll Rel_{progeny}$ but with genomics is this still true?
Take-home messages

• validation with EBVs (by method LR) allows
  • faster & automatic implementation and
  • (maybe) less approximations/incertitudes than DYDs / DRP
  • theory-based results
  • I find theory important now that we select based on early genomic predictions

• the variation in validation results needs to be considered
  • in principle we can derive theory to get e.g. the s.e. of $\hat{\Delta}_p$, $\hat{b}_{w,p}$, etc
  • I think that checking several years is a necessity for small breeds

• I (personally) think that difference in means(GEBV) should be checked
  i.e. $\hat{\Delta}_p = \hat{u}_p - \hat{u}_w$
• thank you
New GEBVtest Program

Pete Sullivan    (Lactanet, Canada)
• The Interbull software “gebvtest.py” is used for GEBV “Interbull validation tests”
• This Python program was modified to facilitate new and expanded validation tests of interest to the Interbull Validation Test Working Group
• the modified version is named “gebvtest_2022C.py”
• Backward compatibility was maintained
  - Input and output files are identical between new and old versions, if none of the optional new features are requested.
Test reliability and unbiasedness of GEBV to predict future daughter performance (dEBV)

1) Reliability: \( R(\text{dEBV}, \text{GEBV}_{\text{red}})^2 > R(\text{dEBV}, \text{EBV}_{\text{red}})^2 \) (genomics improves \( R^2 \))

2) Unbiasedness: \( b(\text{dEBV}, \text{GEBV}_{\text{red}}) = 1 \) (acceptable range ~ 0.90 – 1.20)

Working Group concerns with the current tests

1) De-regression methods used to generate the dEBV can be arbitrary

2) Different bases of expression between Reduced versus Full data can directly affect the regression test, and can easily cause “FALSE FAIL” regression test results

3) EBV and dEBV carry genomic preselection bias, which means that biased GEBV look better than unbiased GEBV, and this is the opposite of what we want

4) The significance testing could be improved (e.g. with boot strapping)
Usage and New Command-line Options

<table>
<thead>
<tr>
<th>gebvtest.py -h</th>
<th>gebvtest_2022C.py -h</th>
</tr>
</thead>
</table>

**Positional Arguments:**
- brd: evaluation breed code (BSW/GUE/JER/HOL/RDC/SIM)
- pop: population code (same as country code except for CHR/DEA/DFS/FRR/FRM)
- datadir: absolute or relative path to data files

**Optional Arguments:**
- -h, --help: show this help message and exit
- -v, --verbose: increase output verbosity
- -m, --mergefiles: write merged data files for independent data checks
- -M MERGEDIR, --mergedir MERGEDIR: absolute or relative path for merged data files (default=DATADIR/merged)
- -Z, --no-zip: do not create a zip file (e.g., for preliminary testing or usage at ITBC)
- -C, --cleanup: delete all files successfully added to the zip file

---

Validation target options are: [ DGEBV, DGPA, VFEBV, GEBV, EBV, DEBV ] (default=DEBV)

**Positional Arguments:**
- brd: evaluation breed code (BSW/GUE/JER/HOL/RDC/SIM)
- pop: population code (same as country code except for CHR/DEA/DFS/FRR/FRM)
- datadir: absolute or relative path to data files

**Optional Arguments:**
- -h, --help: show this help message and exit
- -v, --verbose: increase output verbosity
- -m, --mergefiles: write merged data files for independent data checks
- -M MERGEDIR, --mergedir MERGEDIR: absolute or relative path for merged data files (default=DATADIR/merged)
- -Z, --no-zip: do not create a zip file (e.g., for preliminary testing or usage at ITBC)
- -C, --cleanup: delete all files successfully added to the zip file
- --target [DGEBV, DGPA, VFEBV, GEBV, EBV, DEBV]: validation target options are: [ DGEBV, DGPA, VFEBV, GEBV, EBV, DEBV ] (default=DEBV)

**Options:**
- --weight [ITB,LR]: Options are: [ ITB or LR ], for the Interbull weighted-regression test or Legarra-Reverter un-weighted regression, respectively (default=ITB)
- --min_byear MIN_BYEAR: specify a minimum birth year to use instead of using the value specified in the traits file
- --baseadj [GEBV, EBV, NONE]: evaluation variable to use for base adjustments, options are: [ NONE, EBV, GEBV ] (default=NONE)
- --power POWER: specify a base for the power function weighting records in base adjustments, instead of optimizing the base from the data
- --baseincl BASEINCL: comma-separated lists of restrictions on bulls to include for base adjustment estimates, [ min,max byr : proof type list : proof status list : official Y/N ]
- --traitsincl TRAITSINCL: comma-separated list of traits to process
- --outdir OUTDIR: absolute or relative path to write output files (default=.)
1. Analyze and adjust the base of expression in Reduced data to match Full data
   • Uses a regression heavily weighted to bulls who do not add any recent data
     ➢ Same reliability should usually mean same evaluation, if bases of expression are the same
   • Regression weights are optimized for the distribution of bull reliability changes for the trait
     --baseadj = \{EBV\} or \{GEBV\}
   • Validation tests are thus based on “relative changes” for bulls adding versus not adding new data

2. Allow different validation targets
   • Allowing validation targets like GEBV that are not biased by GPS (e.g. Legarra-Reverter)
   • An EBV target can also still be used, for new tests that might replace Trend Validation test III
     --target = \{DEBV\} or \{EBV\} or \{DGEBV\} or \{GEBV\}
     \{DGEBV\} are derived by the software using method of Van Raden (2021 Interbull bulletin)
3. Allow filtering of data to use in **--baseadj**
   
   **--baseincl** `{min_byr, max_byr : proof types : bull status : official Y|N}`
   
   **--outdir** to save output files in a desired location
   
   ➢ Useful for comparing test results based on different requested options

4. Command-line options to over-ride data and values read from input files
   
   **--min_byear** YYMM
   
   ➢ Eliminates need to re-create several versions of input files for different tests

   **--traitsincl** mil,ocs,dlo
   
   ➢ Useful to focus extra checks on only subsets of traits with poor test results
   
   ➢ Limits **--mergefiles** output to only the trait subsets of interest
5. Apply unweighted regression and correlation tests, instead of the standard weighted regressions used for Interbull validation tests

    --weight LR

    Ø Useful to see the impacts of using weighted regression in the Interbull tests

    Ø In combination with --min_byear these options make it easy to see the impacts of autocorrelations on validation test results, if a weighted test is not used
Countries Reports

- CANADA
Applications with Canadian Data

- Full-data evaluations (EBV and GEBV) = August 2022
- Reduced-data evaluations = December 2018

- Tests applied to all traits in MACE (38)
- Summarizing results for the required test traits in GMACE (20)
  - Excludes all but 2 of the type traits
Regression tests with Canadian Data

Regression tests with Canadian Data

--target DEBV (tD) --target GEBV (tG)

--baseadj GEBV (bG)

--weight LR Legarra-Reverter Simple Regression
Conclusions

• Matching the Base of Expression is important: --baseadj

• Seeking good predictions of biased future results makes no sense
  --target GEBV or DGBV is better than the current test using --target DEBV

• We can improve on the “basic” Legarra-Reverter regression test in 2 ways:
  1. With the new --baseadj feature
  2. With weighted regressions that target animals adding data for the validation tests.
     ➢ Reduces auto-correlations in the regression tests and increases power to detect bias

• Several new options, in combination, were designed to help focus on problem trait(s)
  --baseadj --traitsincl --mergefiles and flexible formatting now allowed for file300
  ➢ The mergefiles include all variables for post-analyses, including extra regression variables
  ➢ All reduced-data evaluations are base-adjusted to match the full-data base of expression
     ➢ Average differences have expectation=0 for any group of animals
Countries Reports

- USA
Genomic validation software: USA results

Rodrigo Mota\textsuperscript{1}, Ezequiel Nicolazzi\textsuperscript{1}, and Paul VanRaden\textsuperscript{2}

\textsuperscript{1}CDCB - Council on Dairy Cattle Breeding, Bowie, MD, USA
\textsuperscript{2}Animal Genomics and Improvement Laboratory, ARS, USDA, Beltsville, MD, USA

Interbull Meeting - Rome, Italy
Feb 13, 2023
Apply new validation to 5 breeds and 7 traits

• Validation of USA genomic predictions (GPTA)
  • August 2022 official GPTA including MACE input
  • August 2018 truncated GPTA using official evaluations
  • Breeds tested were HOL, JER, BSW, RDC, and GUE
  • Traits tested were mil, fat, pro, scs, dlo(*), int, and mas (HO only)
New GEBV test software (Sullivan)

- gebvtest_2022C.py
- Minimum birth year: 2014
- Predicted deregressed dGPTA instead of dPTA
  - target DGEBV: uses the method of VanRaden, 2021 (Interbull Bulletin)
  - baseadj GEBV
### Validation results: Holstein

<table>
<thead>
<tr>
<th>Trait</th>
<th>Bulls</th>
<th>S.E. (Bulls)</th>
<th>R(^2) GEBV</th>
<th>R(^2) PA</th>
<th>Pass/Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>3,562</td>
<td>1.10</td>
<td>0.01</td>
<td>74</td>
<td>35</td>
</tr>
<tr>
<td>Fat</td>
<td>3,562</td>
<td>1.08</td>
<td>0.01</td>
<td>77</td>
<td>41</td>
</tr>
<tr>
<td>Protein</td>
<td>3,562</td>
<td>1.04</td>
<td>0.01</td>
<td>74</td>
<td>43</td>
</tr>
<tr>
<td>SCS</td>
<td>3,502</td>
<td>1.12</td>
<td>0.01</td>
<td>69</td>
<td>27</td>
</tr>
<tr>
<td>Longevity (dlo)</td>
<td>3,330</td>
<td>1.01</td>
<td>0.01</td>
<td>61</td>
<td>30</td>
</tr>
<tr>
<td>DPR (int)</td>
<td>3,425</td>
<td>0.93</td>
<td>0.01</td>
<td>54</td>
<td>21</td>
</tr>
<tr>
<td>Mastitis</td>
<td>2,379</td>
<td>1.30</td>
<td>0.03</td>
<td>40</td>
<td>17</td>
</tr>
</tbody>
</table>
### Validation results: Jersey

<table>
<thead>
<tr>
<th>Trait</th>
<th>Trait Code</th>
<th>R² GEBV</th>
<th>S.E. (R² GEBV)</th>
<th>R² PA</th>
<th>S.E. (R² PA)</th>
<th>Pass/Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>648</td>
<td>1.06</td>
<td>0.03</td>
<td>73</td>
<td>0.03</td>
<td>Pass</td>
</tr>
<tr>
<td>Fat</td>
<td>648</td>
<td>1.05</td>
<td>0.03</td>
<td>63</td>
<td>0.03</td>
<td>Pass</td>
</tr>
<tr>
<td>Protein</td>
<td>648</td>
<td>1.05</td>
<td>0.03</td>
<td>69</td>
<td>0.03</td>
<td>Pass</td>
</tr>
<tr>
<td>SCS</td>
<td>604</td>
<td>1.01</td>
<td>0.05</td>
<td>45</td>
<td>0.05</td>
<td>Pass</td>
</tr>
<tr>
<td>Longevity (dlo)</td>
<td>571</td>
<td>0.88</td>
<td>0.05</td>
<td>36</td>
<td>0.05</td>
<td>Fail</td>
</tr>
<tr>
<td>DPR (int)</td>
<td>588</td>
<td>0.79</td>
<td>0.03</td>
<td>47</td>
<td>0.03</td>
<td>Fail</td>
</tr>
</tbody>
</table>
## Validation results: Brown Swiss

<table>
<thead>
<tr>
<th>Trait</th>
<th>R² GEBV</th>
<th>R² PA</th>
<th>Pass / Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>94</td>
<td>1.29</td>
<td>0.11</td>
</tr>
<tr>
<td>Fat</td>
<td>94</td>
<td>0.89</td>
<td>0.11</td>
</tr>
<tr>
<td>Protein</td>
<td>94</td>
<td>1.03</td>
<td>0.11</td>
</tr>
<tr>
<td>SCS</td>
<td>93</td>
<td>0.60</td>
<td>0.09</td>
</tr>
<tr>
<td>Longevity (dlo)</td>
<td>65</td>
<td>0.53</td>
<td>0.12</td>
</tr>
<tr>
<td>DPR (int)</td>
<td>88</td>
<td>0.64</td>
<td>0.16</td>
</tr>
</tbody>
</table>

- **Milk** with GEBV R² of 59 and PA R² of 20 fails the test.
- **Fat** with GEBV R² of 41 and PA R² of 18 passes the test.
- **Protein** with GEBV R² of 47 and PA R² of 16 passes the test.
- **SCS** with GEBV R² of 31 and PA R² of 9 fails the test.
- **Longevity (dlo)** with GEBV R² of 24 and PA R² of 12 fails the test.
- **DPR (int)** with GEBV R² of 16 and PA R² of 27 fails the test.
### Validation results: **Ayrshire (RDC)**

<table>
<thead>
<tr>
<th>Trait</th>
<th>Bulls</th>
<th>( \text{R}^2 ) GEBV</th>
<th>( \text{R}^2 ) PA</th>
<th>Pass / Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>22</td>
<td>0.61</td>
<td>0.21</td>
<td>35</td>
</tr>
<tr>
<td>Fat</td>
<td>22</td>
<td>0.91</td>
<td>0.20</td>
<td>54</td>
</tr>
<tr>
<td>Protein</td>
<td>22</td>
<td>0.85</td>
<td>0.21</td>
<td>50</td>
</tr>
<tr>
<td>SCS</td>
<td>21</td>
<td>0.79</td>
<td>0.30</td>
<td>21</td>
</tr>
<tr>
<td>Longevity</td>
<td>7</td>
<td>1.36</td>
<td>1.26</td>
<td>1</td>
</tr>
<tr>
<td>DPR (int)</td>
<td>22</td>
<td>0.60</td>
<td>0.52</td>
<td>3</td>
</tr>
</tbody>
</table>
## Validation results: Guernsey

<table>
<thead>
<tr>
<th>Trait</th>
<th>GEBV</th>
<th>S.E. (GEBV)</th>
<th>R² GEBV</th>
<th>PA</th>
<th>S.E. (PA)</th>
<th>R² PA</th>
<th>Pass/Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>19</td>
<td>0.92</td>
<td>0.22</td>
<td>51</td>
<td></td>
<td>36</td>
<td>Pass</td>
</tr>
<tr>
<td>Fat</td>
<td>19</td>
<td>0.51</td>
<td>0.18</td>
<td>32</td>
<td></td>
<td>34</td>
<td>Fail</td>
</tr>
<tr>
<td>Protein</td>
<td>19</td>
<td>0.77</td>
<td>0.25</td>
<td>36</td>
<td></td>
<td>26</td>
<td>Pass</td>
</tr>
<tr>
<td>SCS</td>
<td>18</td>
<td>1.48</td>
<td>0.43</td>
<td>42</td>
<td></td>
<td>19</td>
<td>Fail</td>
</tr>
<tr>
<td>Longevity</td>
<td>7</td>
<td>0.52</td>
<td>0.52</td>
<td>37</td>
<td></td>
<td>5</td>
<td>Pass</td>
</tr>
<tr>
<td>DPR (int)</td>
<td>19</td>
<td>0.62</td>
<td>0.62</td>
<td>18</td>
<td></td>
<td>30</td>
<td>Fail</td>
</tr>
</tbody>
</table>
Summary

• Larger breeds and more heritable traits had more stable results

• Smaller breeds and less heritable traits are hard to validate. Tests often fail:
  • B1 more or less than expected from S.E., which may be underestimated.
  • Upper biological limit of 1.2 should allow for S.E. of B1
  • R2 of parent average may exceed GEBV with small sample sizes
  • Extra regressions could help test other biases (trend, parent average, etc.)
Acknowledgements and disclaimers

- **Pete Sullivan** of Lactanet for providing an advance copy of the validation software
- Participating **dairy producers** for supplying data
- **DHI** organizations and **DRPCs** for processing and relaying the information to the Council on Dairy Cattle Breeding (CDCB)
- **Purebred breed associations** for providing pedigree data
- 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
- **CDCB** is an equal opportunity provider and employer
Thank you!
• DENMARK-FINLAND-SWEDEN
Test of GEBV software
Ulrik Sander Nielsen and Trine Andersen
SEGES Innovation P/S
Denmark

Interbull Technical Workshop, Rome, 14-15 February 2023
DFS HOL type traits

- Official indexes published Nov 2022
  Mean:100, SD:10
  App. 100 Nordic AI bulls tested per year
- Genotype cut by birth class 2009
- Datasets created according to Interbull guidelines for GEBV test from 2010
- Reduced GEBV dataset deleting progeny of bulls born after 2014.01.01
GEBV tests

1. Official Interbull program
2. Test program with same options as official
3. Legarra-Reverter regression
4. Legarra-Reverter regression with option – min_byear=2010
5. --target DGEBV, method of VanRaden
6. --weight=ITB –baseincl 2014,2015:11,12,10:Y
7. Official but only DFS bulls included (stature)
## Results: Stature

<table>
<thead>
<tr>
<th>Test</th>
<th>ntest</th>
<th>b0</th>
<th>b1</th>
<th>ncand</th>
<th>Exp_b1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 intb test</td>
<td>431</td>
<td>-1.74</td>
<td>1.00</td>
<td>602</td>
<td>0.93</td>
<td>80.3</td>
</tr>
<tr>
<td>2 test</td>
<td>431</td>
<td>-1.74</td>
<td>1.00</td>
<td>602</td>
<td>0.93</td>
<td>80.3</td>
</tr>
<tr>
<td>3 LR test</td>
<td>607</td>
<td>-2.74</td>
<td>1.01</td>
<td>1183</td>
<td>0.95</td>
<td>84.1</td>
</tr>
<tr>
<td>4 LR 2010</td>
<td>607</td>
<td>-2.73</td>
<td>1.01</td>
<td>2790</td>
<td>1.00</td>
<td>84.1</td>
</tr>
<tr>
<td>5 DGEBV</td>
<td>531</td>
<td>-3.65</td>
<td>1.02</td>
<td>602</td>
<td>0.95</td>
<td>84.4</td>
</tr>
<tr>
<td>6 b,2014,15</td>
<td>431</td>
<td>-1.74</td>
<td>1.00</td>
<td>602</td>
<td>0.94</td>
<td>84.1</td>
</tr>
<tr>
<td>7 DFS bulls</td>
<td>328</td>
<td>1.20</td>
<td>0.97</td>
<td>602</td>
<td>0.89</td>
<td>77.7</td>
</tr>
</tbody>
</table>

All tests pass: N-Y-Y-Y-Y PASS
## Results: Udder support

<table>
<thead>
<tr>
<th>Test</th>
<th>ntest</th>
<th>b0</th>
<th>b1</th>
<th>ncand</th>
<th>Exp_b1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 intb test</td>
<td>395</td>
<td>-2.74</td>
<td>1.02</td>
<td>520</td>
<td>0.96</td>
<td>80.5</td>
</tr>
<tr>
<td>2 test</td>
<td>395</td>
<td>-2.74</td>
<td>1.02</td>
<td>520</td>
<td>0.96</td>
<td>80.5</td>
</tr>
<tr>
<td>3 LR test</td>
<td>607</td>
<td>-2.74</td>
<td>1.02</td>
<td>1183</td>
<td>0.95</td>
<td>84.1</td>
</tr>
<tr>
<td>4 LR 2010</td>
<td>607</td>
<td>-2.73</td>
<td>1.02</td>
<td>2779</td>
<td>0.95</td>
<td>84.1</td>
</tr>
<tr>
<td>5 DGEBV</td>
<td>395</td>
<td>90.40</td>
<td>0.08</td>
<td>520</td>
<td>0.80</td>
<td>0.6</td>
</tr>
<tr>
<td>6 b,2014,15</td>
<td>395</td>
<td>-2.74</td>
<td>1.02</td>
<td>520</td>
<td>0.96</td>
<td>80.5</td>
</tr>
</tbody>
</table>

Test 1,2,3,4,6 pass: N-Y-Y-Y-Y PASS
Test 5 fails: N-N-N-N-Y FAIL
# Udder support - DGEBV test

Summary statistics on candidate bulls (CB) and test bulls (TB)

<table>
<thead>
<tr>
<th>Trait Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>usu CB EBV</td>
<td>520</td>
<td>100.468</td>
<td>10.029</td>
<td>74.56</td>
<td>126.37</td>
</tr>
<tr>
<td>usu TB EBV</td>
<td>395</td>
<td>99.315</td>
<td>10.122</td>
<td>74.56</td>
<td>126.37</td>
</tr>
<tr>
<td>usu TB VAL(y)</td>
<td>395</td>
<td>99.317</td>
<td>11.233</td>
<td>71.82</td>
<td>130.94</td>
</tr>
<tr>
<td>usu TB GEBV(x1)</td>
<td>395</td>
<td>104.208</td>
<td>10.468</td>
<td>71.62</td>
<td>132.84</td>
</tr>
<tr>
<td>usu TB EBVr(x2)</td>
<td>395</td>
<td>100.782</td>
<td>8.102</td>
<td>78.31</td>
<td>126.08</td>
</tr>
</tbody>
</table>

Details of GEBVtest calculations
usu \( i_{est} = (99.315 - 100.468) / 10.029 = -0.115 \)

After Base adjustments:
usu \( p=0.947 \) \( x=-1.612 \) \( i=0.115 \) \( k=0.198 \) \( R2_b=0.008 \) \( E(b1)=0.803 \)
usu \( b1=0.085 \) \( se=0.054 \) \( E(b1)=0.803 \) \( t=-13.40 \) \( R2_1=0.6 \) \( R2_2=30.5 \)
usu passes \( t\)-test=N \( \text{bio-test}=N \) \( b1>1=N \) \( R2\)-test=N \( b1<1.2=Y \) overall=FAIL
Udder support – weight=LR

Summary statistics on candidate bulls (CB) and test bulls (TB)

<table>
<thead>
<tr>
<th>Trait Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>usu CB EBV</td>
<td>1183</td>
<td>102.592</td>
<td>8.817</td>
<td>74.56</td>
<td>128.51</td>
</tr>
<tr>
<td>usu TB EBV</td>
<td>607</td>
<td>100.825</td>
<td>9.419</td>
<td>74.56</td>
<td>126.37</td>
</tr>
<tr>
<td>usu TB VAL(y)</td>
<td>607</td>
<td>105.402</td>
<td>11.950</td>
<td>63.93</td>
<td>137.01</td>
</tr>
<tr>
<td>usu TB GEBV(x1)</td>
<td>607</td>
<td>106.293</td>
<td>10.770</td>
<td>71.62</td>
<td>138.38</td>
</tr>
<tr>
<td>usu TB EBVr(x2)</td>
<td>607</td>
<td>102.112</td>
<td>8.330</td>
<td>78.31</td>
<td>126.08</td>
</tr>
</tbody>
</table>

Details of GEBVtest calculations
usu i_est = (100.825 - 102.592) / 8.817 = -0.200

After Base adjustments:
usu p=0.897 x=-1.263 i=0.200 k=0.293 R2b=0.882 E(b1)=0.953
usu b1=1.017 se=0.018 E(b1)=0.953 t=3.56 R2_1=84.1 R2_2=4.1
usu passes t-test=N bio-test=Y b1>1=Y R2-test=Y b1<1.2=Y overall=PASS
Summary

• Without use of options, test version gives same results as official test
• Correction for base year has little effect
• LR test results in line with official test results
• Using –target DGEBV gave unexpected results for the trait, udder support
• THE NETHERLANDS
Results from various validation methods
Ibrahim Jibrila, Herwin Eding
Interbull Technical Workshop
Rome, 15 Feb 2023
Introduction

- General validation procedure
- Results
- Conclusions and discussion
- Experiences with Interbull validation software.
General validation procedure

Routine AEU validation; for each trait in evaluation
- Full data (deregressed) GEBV used as dependent variable (GEBV or DRP)
- Reduced data (- 4 years) GEBV and PA

- Bull validations:
  - Sires with at least 20 daughters in Full, but no daughters in Reduced

- Cow validations:
  - Cows with at least one record in Full, but no observations in Reduced
General validation procedure

Analysis
• Regress reduced run GEBV and PA on full DRP or GEBV
  – $\text{DRP/GEBV}_{\text{ful}} = a + b \text{ PA}_{\text{red}}$
  – $\text{DRP/GEBV}_{\text{ful}} = c + d \text{ GEBV}_{\text{red}}$ ( = $\text{PA}_{\text{red}} + \text{genotype}$)

• Main statistics of interest:
  – Regression factor, added EDC, DGV reliability
Testing of validation methods

• Using deregressed proofs (DRP) versus unregressed proofs (GEBV)

• Using bull validation (classical) versus cow validation.
  – Could possibly alleviate lack of data issues

• Results shown for illustrative traits
  – Udderhealth
  – Calving ease
  – Ketosis
  – Reproductive disorders
<table>
<thead>
<tr>
<th>Trait name</th>
<th>h2</th>
<th>N bulls</th>
<th>N cows</th>
<th>Regression coeff</th>
<th>bul drp</th>
<th>bull gebv</th>
<th>cow drp</th>
<th>cow gebv</th>
</tr>
</thead>
<tbody>
<tr>
<td>subclinical mastitis</td>
<td>0,056</td>
<td>763</td>
<td>89776</td>
<td>1,03</td>
<td>0,99</td>
<td>1,12</td>
<td>1,02</td>
<td></td>
</tr>
<tr>
<td>clinical mastitis</td>
<td>0,062</td>
<td>775</td>
<td>14698</td>
<td>0,89</td>
<td>0,86</td>
<td>1,08</td>
<td>0,96</td>
<td></td>
</tr>
<tr>
<td>udder health index</td>
<td>0,090</td>
<td>775</td>
<td>90196</td>
<td>0,96</td>
<td>0,92</td>
<td>1,11</td>
<td>0,99</td>
<td></td>
</tr>
<tr>
<td>maternal stillbirth heifers</td>
<td>0,084</td>
<td>571</td>
<td>81396</td>
<td>1,06</td>
<td>1,03</td>
<td>1,21</td>
<td>1,04</td>
<td></td>
</tr>
<tr>
<td>maternal stillbirth cow</td>
<td>0,005</td>
<td>485</td>
<td>71515</td>
<td>1,12</td>
<td>1,03</td>
<td>1,33</td>
<td>1,06</td>
<td></td>
</tr>
<tr>
<td>direct stillbirth heifers</td>
<td>0,041</td>
<td>444</td>
<td>38741</td>
<td>0,94</td>
<td>0,90</td>
<td>1,20</td>
<td>0,93</td>
<td></td>
</tr>
<tr>
<td>direct stillbirth cows</td>
<td>0,006</td>
<td>571</td>
<td>66131</td>
<td>0,96</td>
<td>0,88</td>
<td>1,22</td>
<td>0,94</td>
<td></td>
</tr>
<tr>
<td>milk fever parity</td>
<td>0,035</td>
<td>51</td>
<td>3565</td>
<td>0,59</td>
<td>0,61</td>
<td>1,09</td>
<td>0,94</td>
<td></td>
</tr>
<tr>
<td>clinical ketosis overall</td>
<td>0,096</td>
<td>115</td>
<td>8619</td>
<td>0,62</td>
<td>0,60</td>
<td>0,87</td>
<td>0,75</td>
<td></td>
</tr>
<tr>
<td>retained placenta</td>
<td>0,064</td>
<td>272</td>
<td>17672</td>
<td>0,69</td>
<td>0,65</td>
<td>0,90</td>
<td>0,78</td>
<td></td>
</tr>
<tr>
<td>endometritis</td>
<td>0,060</td>
<td>256</td>
<td>17519</td>
<td>0,67</td>
<td>0,65</td>
<td>0,93</td>
<td>0,79</td>
<td></td>
</tr>
<tr>
<td>metritis</td>
<td>0,049</td>
<td>224</td>
<td>14543</td>
<td>0,61</td>
<td>0,60</td>
<td>1,01</td>
<td>0,81</td>
<td></td>
</tr>
<tr>
<td>cystic ovaries</td>
<td>0,029</td>
<td>210</td>
<td>12186</td>
<td>0,85</td>
<td>0,79</td>
<td>1,00</td>
<td>0,85</td>
<td></td>
</tr>
<tr>
<td>anoestrus</td>
<td>0,034</td>
<td>293</td>
<td>18325</td>
<td>0,70</td>
<td>0,67</td>
<td>0,89</td>
<td>0,77</td>
<td></td>
</tr>
<tr>
<td>index reproduction disorders</td>
<td>0,123</td>
<td>337</td>
<td>22646</td>
<td>0,71</td>
<td>0,65</td>
<td>0,93</td>
<td>0,76</td>
<td></td>
</tr>
</tbody>
</table>
### Results: Added EDC

<table>
<thead>
<tr>
<th>Trait name</th>
<th>h2</th>
<th>N bulls</th>
<th>N cows</th>
<th>EDC added</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>bul drp</td>
</tr>
<tr>
<td>subclinical mastitis</td>
<td>0.056</td>
<td>763</td>
<td>89776</td>
<td>145.4</td>
</tr>
<tr>
<td>clinical mastitis</td>
<td>0.062</td>
<td>775</td>
<td>14698</td>
<td>115.4</td>
</tr>
<tr>
<td>udder health index</td>
<td>0.090</td>
<td>775</td>
<td>90196</td>
<td>87.2</td>
</tr>
<tr>
<td>maternal stillbirth heifers</td>
<td>0.084</td>
<td>571</td>
<td>81396</td>
<td>81.1</td>
</tr>
<tr>
<td>maternal stillbirth cow</td>
<td>0.005</td>
<td>485</td>
<td>71515</td>
<td>803.7</td>
</tr>
<tr>
<td>direct stillbirth heifers</td>
<td>0.041</td>
<td>444</td>
<td>38741</td>
<td>48.9</td>
</tr>
<tr>
<td>direct stillbirth cows</td>
<td>0.006</td>
<td>571</td>
<td>66131</td>
<td>318.2</td>
</tr>
<tr>
<td>milk fever parity</td>
<td>0.035</td>
<td>51</td>
<td>3565</td>
<td>14.4</td>
</tr>
<tr>
<td>clinical ketosis overall</td>
<td>0.096</td>
<td>115</td>
<td>8619</td>
<td>6.3</td>
</tr>
<tr>
<td>retained placenta</td>
<td>0.064</td>
<td>272</td>
<td>17672</td>
<td>7.1</td>
</tr>
<tr>
<td>endometritis</td>
<td>0.060</td>
<td>256</td>
<td>17519</td>
<td>8.5</td>
</tr>
<tr>
<td>metritis</td>
<td>0.049</td>
<td>224</td>
<td>14543</td>
<td>8.3</td>
</tr>
<tr>
<td>cystic ovaries</td>
<td>0.029</td>
<td>210</td>
<td>12186</td>
<td>27.5</td>
</tr>
<tr>
<td>anoestrus</td>
<td>0.034</td>
<td>293</td>
<td>18325</td>
<td>16.6</td>
</tr>
<tr>
<td>index reproduction disorders</td>
<td>0.123</td>
<td>337</td>
<td>22646</td>
<td>4.5</td>
</tr>
</tbody>
</table>
Conclusions

• DRP validation seems more consistent than GEBV validation

• Bull validation seems more stable than cow validation
  – Unless lack of data is an issue. Cow validations can be used when lack of validation bulls occurs.

• Added EDC (diff in regression R^2) higher in GEBV validation than in DRP validation

• Added EDC in cow validations higher than in bull validation

• Leads to inflated estimates of mean DGV reliability
  – Validations of GEBV on bull DRP seems preferable.
  – In a pinch cow DRP validations can be done.
AEU versus ITB validations

New validation software made available
Limited testing on a few number of traits
• Compare validation on
  – EBV (Full conventional)
  – GEBV (Full genomic)
### AEU vs ITB validation: EBV validation

<table>
<thead>
<tr>
<th>Trait name</th>
<th>B1</th>
<th>Diff _R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aeu_debv</td>
<td>aeu_ebv</td>
</tr>
<tr>
<td>Milking speed</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>Temperament</td>
<td>0.61</td>
<td>0.62</td>
</tr>
<tr>
<td>Direct stillbirth</td>
<td>0.55</td>
<td>0.51</td>
</tr>
<tr>
<td>Maternal stillbirth</td>
<td>0.64</td>
<td>0.65</td>
</tr>
<tr>
<td>Clinical mastitis</td>
<td>0.56</td>
<td>0.56</td>
</tr>
</tbody>
</table>
# AEU vs ITB validation: GEBV validation

<table>
<thead>
<tr>
<th>Trait name</th>
<th>B1</th>
<th>Diff _R2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aeu_dgebv</td>
<td>aeu_gebv</td>
</tr>
<tr>
<td>Milking speed</td>
<td>1,07</td>
<td>1,03</td>
</tr>
<tr>
<td>Temperament</td>
<td>0,92</td>
<td>0,89</td>
</tr>
<tr>
<td>Direct stillbirth</td>
<td>0,93</td>
<td>0,89</td>
</tr>
<tr>
<td>Maternal stillbirth</td>
<td>0,99</td>
<td>0,96</td>
</tr>
<tr>
<td>Clinical mastitis</td>
<td>0,85</td>
<td>0,82</td>
</tr>
</tbody>
</table>
**AEU vs ITB validation**

- GEBV validation shows better agreement (AEU, ITB)
  - Higher success rate
  - Conventional EBV biased?
    - Does not account for genomics in PA => less predictability
    - False FAILs

- Validation on GEBV seems to be preferable
  - For both AEU and ITB validations

- Testing of new Interbull software continues...
Thank you...
• GERMANY
Validating German Holstein single-step evaluations for test-day traits using Interbull’s new GEBVtest software

Zengting Liu and Hatem Alkhoder
IT Solutions for Animal Production (vit), Germany
Scenarios for testing the new software for GEBV test

- **Dependent variable**
  - Full evaluation GEBV
  - Deregressed GEBV (VanRaden, 2021)
  - Deregressed EBV (DRP) from the current conventional evaluations
    - Deregressed MACE EBV for bulls (MACE_DRP)
    - Deregressed national EBV for cows with national phenotype data only (NAT_DRP)

- **Validation animals**
  - Genotyped bulls having daughters in DEU and EDC ≥ 20 (1,655 Holstein bulls)
  - Genotyped domestic cows with test-day records (180,389 Holstein cows)
    - Own test modification of the software for validation cows: removal of the minimum EDC 20

- Full and truncated evaluations both adjusted for the base population average
  - Using the same cow base population
  - Option of *no base adjustment* for the GEBVtest software
Validation results using full evaluation \( \text{GEBV} \) as dependent variable (I)

- **Validation bulls**
- **Model 1**: \( \text{GEBV}_{\text{full}} = b_0 + b_1 \times \text{GEBV}_{\text{trunc}} \)
  - Weighted LR regression
- **Model 2**: \( \text{GEBV}_{\text{full}} = b_0 + b_1 \times \text{EBV}_{\text{trunc}} \)

### Table

<table>
<thead>
<tr>
<th>Trait</th>
<th>Model 1</th>
<th>M1-M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b_1 )</td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Milk yield</td>
<td>1.01</td>
<td>0.80</td>
</tr>
<tr>
<td>Fat yield</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>Protein yield</td>
<td>0.95</td>
<td>0.71</td>
</tr>
<tr>
<td>SCS</td>
<td>0.99</td>
<td>0.78</td>
</tr>
</tbody>
</table>

---

**\( b_1 \) of Model 1**

**\( R^2 \) of Model 1**

**\( \Delta R^2 \): Model 1-2**
Validation results using full evaluation GEBV as dependent variable (II)

- **Validation cows**

- **Model 1:** GEBV\textsubscript{full} = \( b_0 + b_1 \times \text{GEBV}_{\text{trunc}} \)
  - Weighted LR Regression

- **Model 2:** GEBV\textsubscript{full} = \( b_0 + b_1 \times \text{EBV}_{\text{trunc}} \)

### Model 1: GEBV\textsubscript{full} = \( b_0 + b_1 \times \text{GEBV}_{\text{trunc}} \)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Model 1</th>
<th>M1-M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b_1 )</td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Milk yield</td>
<td>1.03</td>
<td>1.01</td>
</tr>
<tr>
<td>Fat yield</td>
<td>1.03</td>
<td>1.00</td>
</tr>
<tr>
<td>Protein yield</td>
<td>1.01</td>
<td>0.95</td>
</tr>
<tr>
<td>SCS</td>
<td>1.02</td>
<td>0.99</td>
</tr>
</tbody>
</table>

### Validation of Trait using Model 1

- **\( b_1 \) of Model 1**
  - Milk yield: 1.03
  - Fat yield: 1.03
  - Protein yield: 1.01
  - SCS: 1.02

- **\( R^2 \) of Model 1**
  - Milk yield: 0.89
  - Fat yield: 0.91
  - Protein yield: 0.88
  - SCS: 0.91

- **\( \Delta R^2 \): Model 1-2**
  - Milk yield: 0.59
  - Fat yield: 0.55
  - Protein yield: 0.48
  - SCS: 0.66
Validation results using VanRaden deregressed GEBV as dependent variable (I)

- Validation bulls

- Model 1: DGEBV\textsubscript{full} = b_0 + b_1 \cdot \text{GEBV}_{\text{trunc}}
- Model 2: DGEBV\textsubscript{full} = b_0 + b_1 \cdot \text{EBV}_{\text{trunc}}

### Trait | Model 1 | M1-M2
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>b_1</td>
<td>GEBV</td>
<td>R^2</td>
<td>GEBV</td>
<td>ΔR^2</td>
<td>GEBV</td>
</tr>
<tr>
<td>Milk yield</td>
<td>1.01</td>
<td>1.01</td>
<td>0.77</td>
<td>0.80</td>
<td>0.60</td>
</tr>
<tr>
<td>Fat yield</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>0.80</td>
<td>0.48</td>
</tr>
<tr>
<td>Protein yield</td>
<td>0.94</td>
<td>0.95</td>
<td>0.64</td>
<td>0.71</td>
<td>0.43</td>
</tr>
<tr>
<td>SCS</td>
<td>0.98</td>
<td>0.99</td>
<td>0.71</td>
<td>0.78</td>
<td>0.50</td>
</tr>
</tbody>
</table>

### Parameters

- \( b_1 \) of Model 1
- \( R^2 \) of Model 1
- \( ΔR^2 \): Model 1-2
Validation results using VanRaden deregressed GEBV as dependent variable (II)

- Validation cows
  - Model 1: $\text{DGEBV}_{\text{full}} = b_0 + b_1 \cdot \text{GEBV}_{\text{trunc}}$
  - Model 2: $\text{DGEBV}_{\text{full}} = b_0 + b_1 \cdot \text{EBV}_{\text{trunc}}$

<table>
<thead>
<tr>
<th>Trait</th>
<th>$b_1$ GEBV</th>
<th>$R^2$ GEBV</th>
<th>$\Delta R^2$ GEBV</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk yield</td>
<td>1.07</td>
<td>1.03</td>
<td>0.66</td>
<td>0.46</td>
</tr>
<tr>
<td>Fat yield</td>
<td>1.06</td>
<td>1.03</td>
<td>0.68</td>
<td>0.45</td>
</tr>
<tr>
<td>Protein yield</td>
<td>1.02</td>
<td>1.01</td>
<td>0.54</td>
<td>0.30</td>
</tr>
<tr>
<td>SCS</td>
<td>1.05</td>
<td>1.02</td>
<td>0.69</td>
<td>0.52</td>
</tr>
</tbody>
</table>

- $b_1$ of Model 1
- $R^2$ of Model 1
- $\Delta R^2$: Model 1-2
Summary and conclusions (I)

- The new GEBVtest software successfully tested with DEU single-step model for test-day traits using
  - Three types of dependent variables, and
  - Two groups of validation animals
  - And also for the current 2-step genomic model (DEU conformation traits)

- For all tested scenarios, \( b_1, R^2 \) and \( \Delta R^2 \) seem to meet expectations
  - Validation cows (low reliability animals) and bulls (high reliability) behaved differently
  - Validation results vary across subgroups of validation animals, e.g. born in different years

- Deregressed GEBV / EBV resulted in lower model \( R^2 \) values than GEBV
  - Not directly comparable between the two dependent variables  DGEBV and GEBV
  - Regression slope \( b_1 \) deviated slightly more from 1

- For high reliability validation animals, e.g. national bulls, dependent variables GEBV and DGEBV led to more similar \( R^2 \) values than for low reliability validation animals, e.g. cows
  - Depending on the contribution of own phenotype data to the total reliability of animal
Issues concerning the genomic validation method

- GEBV as dependent variable for low-reliability validation animals \( \rightarrow \) extremely high R\(^2\) value
  - Own phenotype data contributing less to own GEBV than bulls with many daughters

- Validated genomic reliabilities as input data for VanRaden’s DGBEV, e.g. using Interbull GREL method

- Deregressed GEBV as dependent variable are more desirable than GEBV
  - Alternative way of computing deregressed GEBV for single-step model (Liu and Masuda, 2021)
    - Using genomic and pedigree relationship matrices
    - Iterative procedure of solving deregressed GEBV
    - Independent of the truncated genomic evaluation

- 4-year truncation of phenotype data may be revised for more realistic forward prediction
  - A short history of large-scale cow genotyping in most countries

- Requirement of special conventional evaluations (Cf and Cr) in the era of single-step evaluation
  - Testing \( \Delta R^2 > 0 \) still necessary after ~15 years of genomic selection?
Thanks for your attention!

Acknowledgement: Dr. S. Rensing and Dr. J. Heise for discussion