Validation of genomic and genetic evaluations in 305d production traits of Nordic Holstein cattle

M. Koivula, I. Strandén, G. P. Aamand and E. A. Mäntysaari*

Biometrical Genetics, Luke, Finland
NAV, Denmark
GEBV validation test

Model I

\[ \text{DYD} = b_0 + b_1 \text{GEBV} + e \]

Model II

\[ \text{DYD} = b_0 + b_1 \text{EBV} + e \] \% i.e. EBV is parent average

- EBV and GEBV are estimated using truncated (-4 years) data
- DRP (or DYD) are estimated using full data

✓ Regression \( b_1 = 1.00 \)
  if the evaluations and DYD are consistent

✓ GEBV should explain more of DYD than the EBV

GEBV validation test problems (ssGBLUP)

• Generally poorly suited for testing genomic animal models (single-step GBLUP)
  – Validation bulls are by definition young, and have no daughters. But their genotyped daughters might be essential part of genomic reference population

• Generally young bulls are (heavily) selected using GEBVs
  – Eventually the bull based validation $R^2$ starts to decrease
• With our single-step test day model the GEBV validation results are often unsatisfactory (see e.g. Koivula et al. 2016 EAAP, Belfast)
  – Too low $b_1$ and $R^2$
  – Bit, also the parent average validations are poor

**Is this a problem of model or test?**

The aim in this presentation is to take a closer look on accuracy and stability of our evaluations (traditional and genomic)
Model validation alternatives

1. Interbull GEBV validation test

   • Problem 1  Validity of ”golden standard”
     - After years of Genomic Selection
       the EBV model accuracy starts to deteriorate:
       WE CAN NOT FULLY TRUST THE DRPs (Deregressed genetic predictions)
     - Solution:  Start using DYDs and YDs from the ssGBLUP

   • Problem 2  Reduced reliability due selection
     - The validation bulls in 4 year reduced data are selected with GEBV
       values \(\rightarrow\) correlation between GEBV and BV are reduced
       Assumption:  the correlation of GEBV and DY will be closer to expected one
       in (unselected) genotyped cows.
Model validation alternatives

1. Interbull GEBV validation test
   • **Problem 1  Validity of ”golden standard”**
     – After years of Genomic Selection
       the EBV model accuracy starts to deteriorate:
       WE CAN NOT FULLY TRUST THE DRPs (Deregressed genetic predictions)
     – Solution?    Start using DYDs and YDs from the ssGBLUP

   • **Problem 2  Reduced reliability due selection**
     – The validation bulls in 4 year reduced data are selected with GEBV
       → correlation between GEBV and BV are reduced

     Solution?    In (unselected) genotyped cows, the correlation
                   of GEBV and YD will be closer to what we expect
Model validation alternatives

2. **Reliability**
   Correlation \((EBV1, EBV2)\) and regression \(EBV2\) on \(EBV1\)

3. **(G)EBV Stability**

\[
(EBV2 - EBV1) = C\text{-Year} + \text{Parity} + \text{SireType} + \text{Parity} \times \text{SireType}
\]

\(EBV1\) is always the evaluation with less information, and \(EBV2\) the following, next evaluation
Evaluations tested

Data set

• Nordic Holstein 305 d production data
  – 305d lactation records compiled from the data used in official nordic TD evaluations

Analysis models

• Multitrait (lactations 1-3) model for protein
  • Variance parameters derived from national evaluations

\[
\text{Model:} \quad \text{Protein} = \text{herd\_year} + \text{calving\_year\_season} + \text{calving\_age} + \text{animal} + \text{residual}
\]

\[
\text{Weights} = (\text{Number of TD})/10
\]

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3th</th>
</tr>
</thead>
<tbody>
<tr>
<td>(h^2)</td>
<td>0.36</td>
<td>0.29</td>
<td>0.26</td>
</tr>
</tbody>
</table>

• MT -animal model (EBVs)
• ssGBLUP (GEBVs)
Nordic HOL Data & reduced data sets

• Full data Data0
  • Calvings up to March 2016
  • 7.3 million cows with 15.6 million observations
  • Pedigree 9.9 million animals, 30056 genotyped

• Reduced data sets:

<table>
<thead>
<tr>
<th>Year</th>
<th>Data-3</th>
<th>Data-2</th>
<th>Data-1</th>
<th>Data0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- March 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- March 2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- March 2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

27.8.2017
Interbull Open, Tallinn, Estonia
Genetic Trends with different evaluations
Protein (G)EBV; bulls with ≥50 daughters

(G)EBV trend, combined and weighted over lactations 1-3

--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | ---
N    | 723  | 624  | 616  | 585  | 605  | 633  | 508  | 439  | 386  | 351  | 246  | 40

27.8.2017
Interbull Open, Tallinn, Estonia
GEBV validation test results for protein

Regression of $\text{DYD}_{\text{data0}}$ on $\text{PA}_{\text{data-3}}$ or $\text{GEBV}_{\text{data-3}}$

723 Validation bulls

<table>
<thead>
<tr>
<th></th>
<th>PA</th>
<th>GEBV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combined (G)EBV</strong></td>
<td>$b_1$</td>
<td>$R^2$</td>
</tr>
<tr>
<td><strong>EBV_DYD</strong></td>
<td>0.67</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>GEBV_DYD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1. Lactation (G)EBV (783 bulls)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EBV1_DYD1</strong></td>
<td>0.71</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>GEBV1_DYD1</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$R_{\text{GEBV}}^2 = \frac{R_{\text{model}}^2}{\bar{w}}$$

27.8.2017

Interbull Open, Tallinn, Estonia
GEBV validation test results for protein

Regression of $DYD_{\text{data0}}$ on $PA_{\text{data-3}}$ or $GEBV_{\text{data-3}}$

723 Validation bulls

<table>
<thead>
<tr>
<th></th>
<th>PA</th>
<th></th>
<th>GEBV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined (G)EBV</td>
<td>$b_1$</td>
<td>$R^2$</td>
<td>$b_1$</td>
<td>$R^2$</td>
</tr>
<tr>
<td><strong>EBV_DYD</strong></td>
<td>0.67</td>
<td>0.14</td>
<td>0.75</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>GEBV_DYD</strong></td>
<td></td>
<td></td>
<td>0.77</td>
<td>0.39</td>
</tr>
<tr>
<td>1. Lactation (G)EBV (783 bulls)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EBV1_DYD1</strong></td>
<td>0.71</td>
<td>0.17</td>
<td>0.74</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>GEBV1_DYD1</strong></td>
<td></td>
<td></td>
<td>0.78</td>
<td>0.44</td>
</tr>
</tbody>
</table>

$$R^2_{GEBV} = R^2_{model_{1}} / \bar{w}$$
GEBV validation test results

**genotyped cows**

Regression of YD to GEBV or EBV (PA), 1. Lactation (G)EBV only

<table>
<thead>
<tr>
<th>EBV_YD</th>
<th>PA1&lt;sup&gt;st&lt;/sup&gt;</th>
<th>GEBV1&lt;sup&gt;st&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b&lt;sub&gt;1&lt;/sub&gt;</td>
<td>R&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Prod.year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012 (n=2967)</td>
<td>1.19</td>
<td>0.36</td>
</tr>
<tr>
<td>2013 (n=4446)</td>
<td>1.01</td>
<td>0.29</td>
</tr>
<tr>
<td>2014 (n=8556)</td>
<td>1.11</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Note: Female reference population:

<table>
<thead>
<tr>
<th>EBV_YD</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2012 (n=2967)</td>
<td>2012 n=4376;</td>
<td>1.02</td>
<td>0.60</td>
</tr>
<tr>
<td>2013 (n=4446)</td>
<td>2013 n=7342;</td>
<td>1.03</td>
<td>0.59</td>
</tr>
<tr>
<td>2014 (n=8556)</td>
<td>2014 n=11788</td>
<td>1.06</td>
<td>0.58</td>
</tr>
</tbody>
</table>

\[ R^2_{GEBV} = \frac{R^2_{model}}{\bar{W}} \]

27.8.2017
Interbull Open, Tallinn, Estonia
LS means for difference EBV2-EBV1 by lactation, siretype and year of EBV1 evaluation (± 95%CL)

EBV difference= EBV2 – EBV1, change from evaluation to another

Siretype 1= no daughters, Siretype 2 = progeny tested bull

27.8.2017 Interbull Open, Tallinn, Estonia
LS means for change GEBV2-GEBV1 by lactation, siretype and year of GEBV1 evaluation (± 95% CL)

GEBV difference = GEBV2 − GEBV1, change from evaluation to another

Siretype 1 = no daughters, Siretype 2 = progeny tested bull

PA = from PA --> 1. Lactation
Lact1 = from 1. lactation --> 2. lactation

bull EBV SD ~100
Conclusions

• Use of DYDs from animal model run will give lower validation reliability (0.36) than using DYD from ssGBLUP (0.39)

• Estimate of validation reliability using bull DYDs is lower (0.44) than using cow YD (0.60)

• In the stability test both the evaluations were equally good: No obvious pattern was found
Thank you!