

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

(Interbull validation test for genomic evaluations.
Mäntysaari, Liu, and VanRaden, 2010; Interbull Bull, 41).

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 → 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

$$\boxed{(EBV2-EBV1)} = \boxed{C-Year} + \boxed{Parity} + \boxed{SireType} + \boxed{\begin{array}{c} C-year \\ * \\ Parity \\ * \\ SireType \end{array}}$$

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

	1st	2nd	3th
h^2	0.36	0.29	0.26

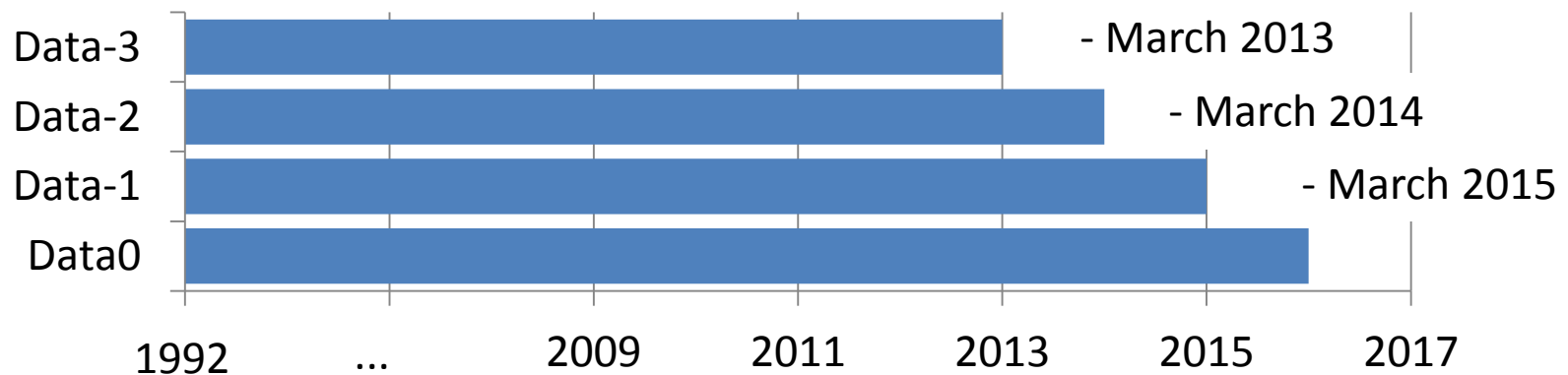
Model:

Protein=herd_year+calving_year_season
+calving_age+animal+residual
! Weights= (Number of TD)/10

- MT -animal model (**EBVs**)
- ssGBLUP (**GEBVs**)

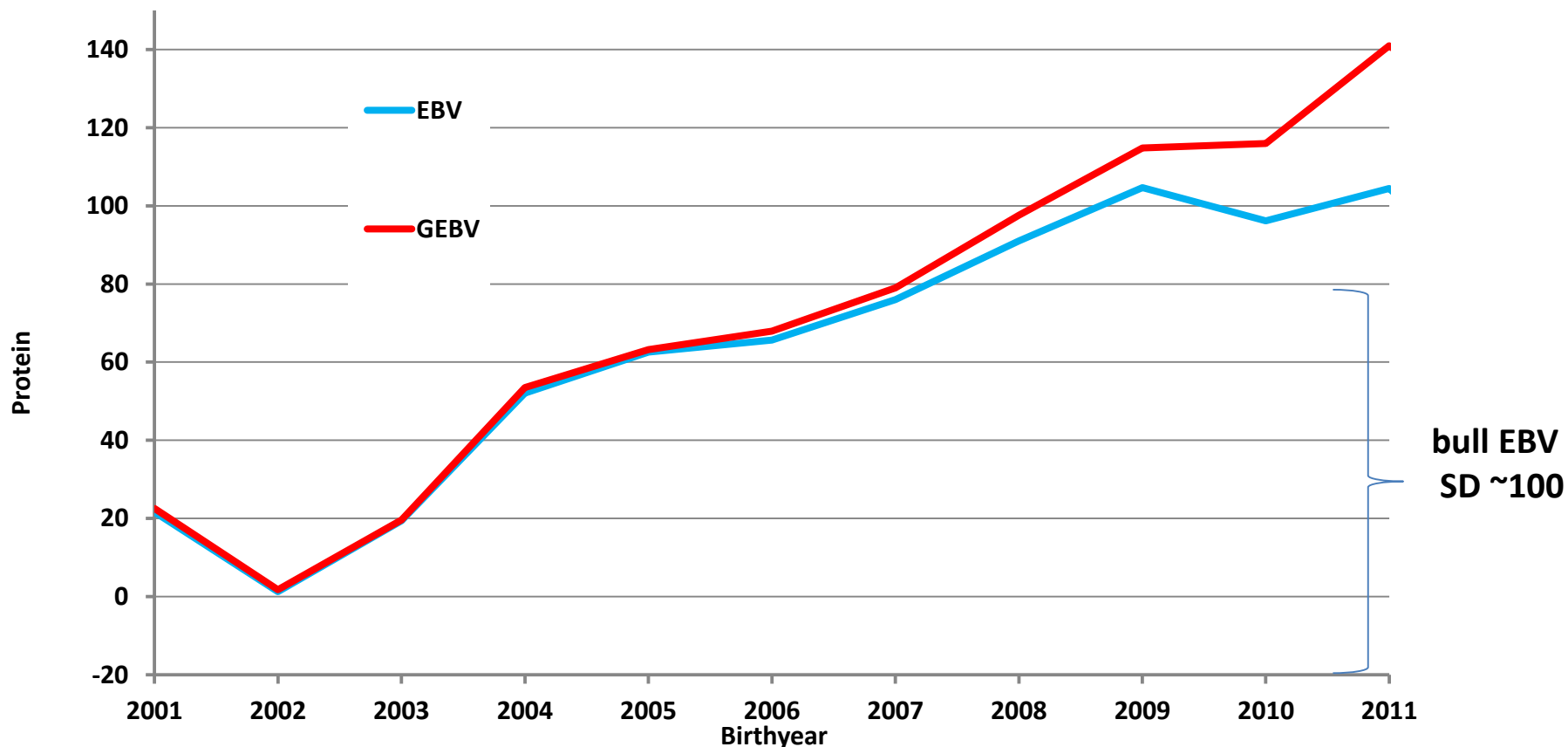
Nordic HOL Data & reduced data sets

- 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:



Genetic Trends with different evaluations

Protein (G)EBV; bulls with ≥ 50 daughters



Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
N	723	624	616	585	605	633	508	439	386	351	246	40

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

GEBV validation test results for protein

Regression of DYD_{data0} on PA_{data-3} or $GEBV_{data-3}$
723 Validation bulls

	PA		GEBV	
Combined (G)EBV	b_1	R^2	b_1	R^2
EBV_DYD	0.67	0.14	0.75	0.36
GEBV_DYD			0.77	0.39
1. Lactation (G)EBV (783 bulls)				
EBV1_DYD1	0.71	0.17	0.74	0.40
GEBV1_DYD1			0.78	0.44

$$R_{GEBV}^2 = R_{model\ 1}^2 / \bar{w}$$

GEBV validation test results for protein

Regression of DYD_{data0} on PA_{data-3} or $GEBV_{data-3}$
723 Validation bulls

	PA		GEBV	
Combined (G)EBV	b_1	R^2	b_1	R^2
EBV_DYD	0.67	0.14	0.75	0.36
GEBV_DYD			0.77	0.39
1. Lactation (G)EBV (783 bulls)				
EBV1_DYD1	0.71	0.17	0.74	0.40
GEBV1_DYD1			0.78	0.44

$$R_{GEBV}^2 = R_{model\ 1}^2 / \bar{w}$$

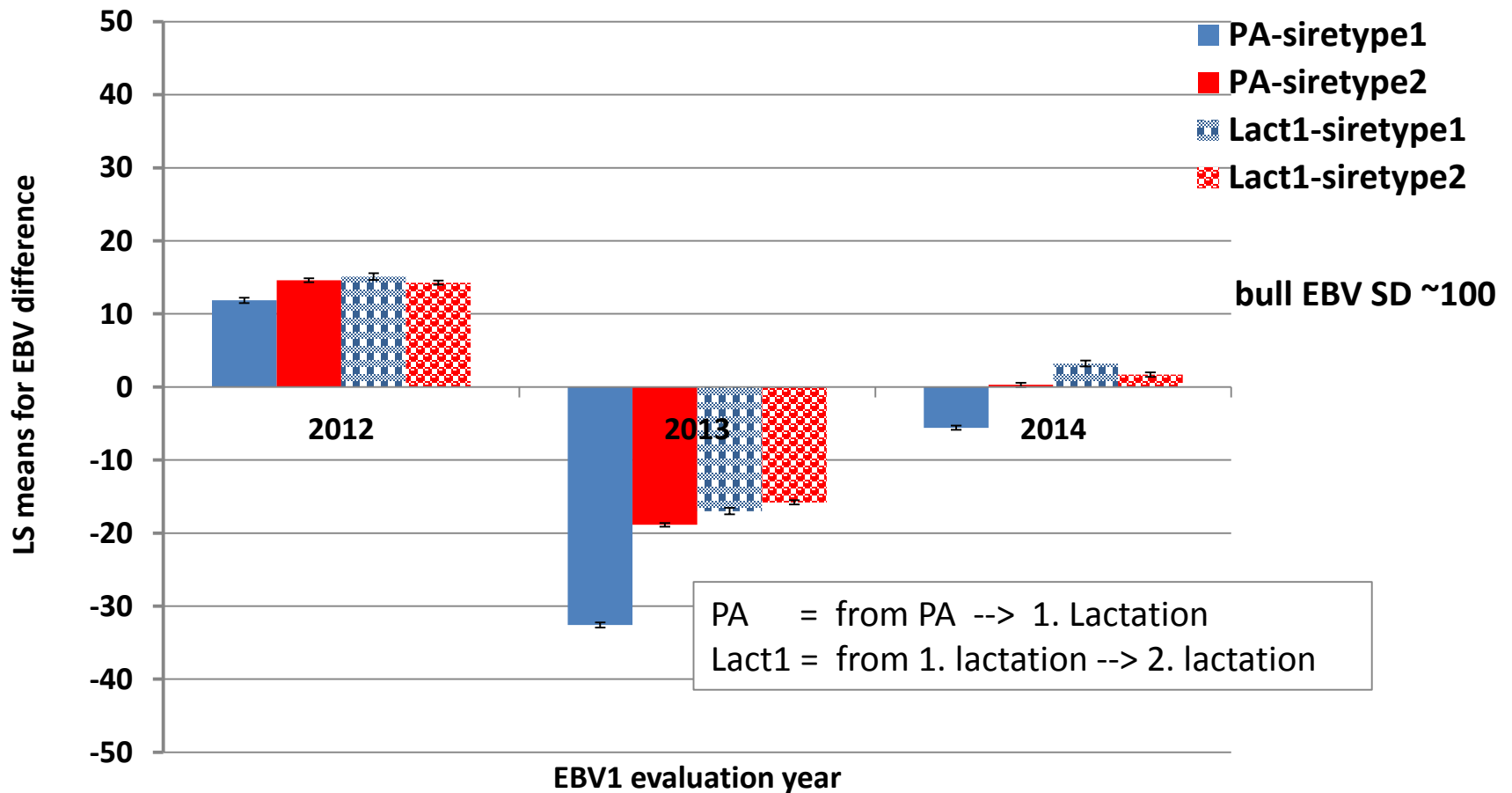
GEBV validation test results genotyped cows

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

EBV_YD	PA1 st		GEBV1 st	
Prod.year	b ₁	R ²	b ₁	R ²
2012 (n=2967)	1.19	0.36	1.00	0.57
2013 (n=4446)	1.01	0.29	1.01	0.57
2014 (n=8556)	1.11	0.28	1.04	0.56
GEBV_YD	Note: Female reference population: 2012 n=4376; 2013 n=7342; 2014 n=11788			
2012 (n=2967)			1.02	0.60
2013 (n=4446)			1.03	0.59
2014 (n=8556)			1.06	0.58

$$R_{GEBV}^2 = R_{model\ 1}^2 / \bar{w}$$

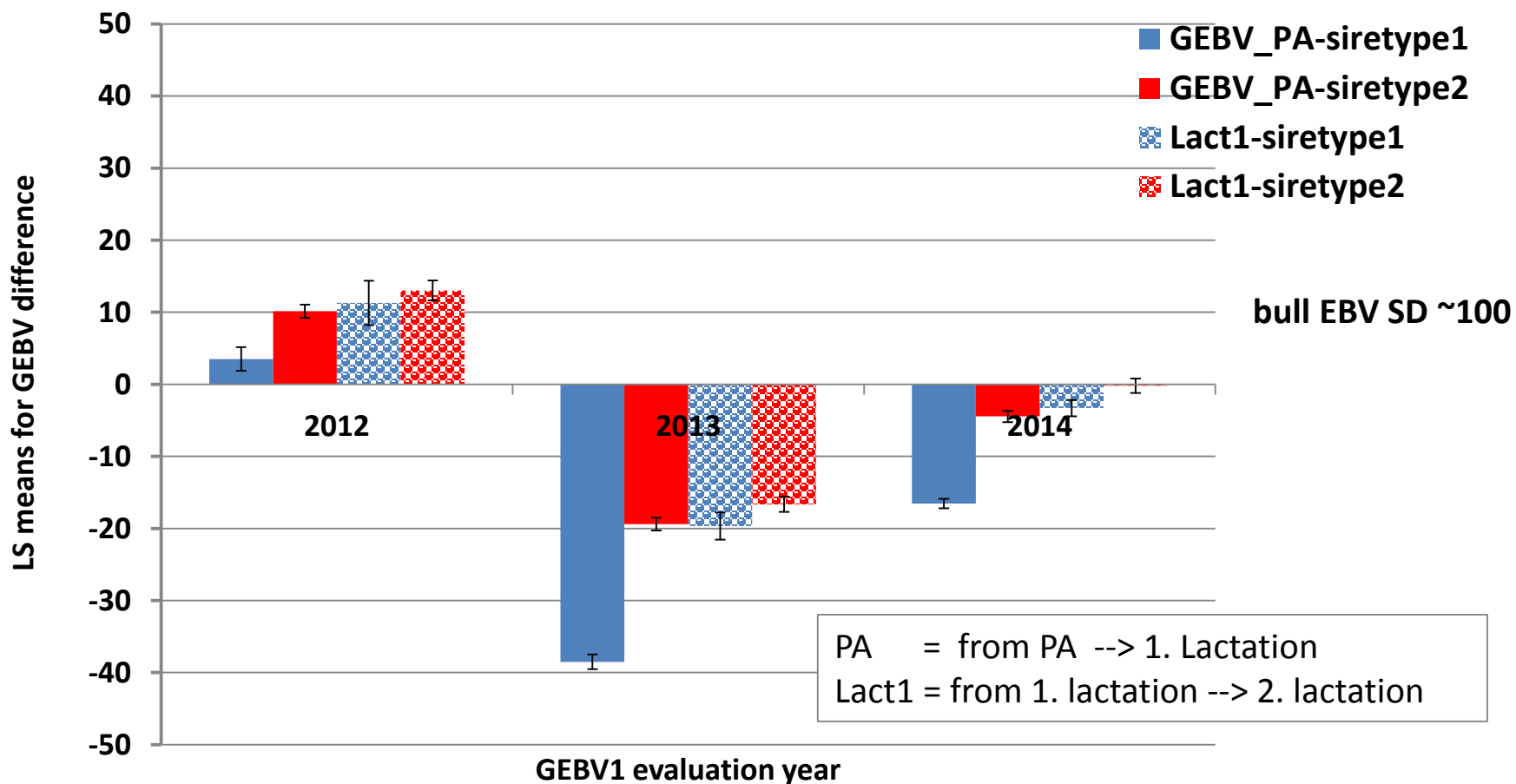
LS means for difference EBV2-EBV1 by lactation, siretype and year of EBV1 evaluation ($\pm 95\%CL$)



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

LS means for change GEBV2-GEBV1

by lactation, siretype and year of GEBV1 evaluation (\pm 95% CL)



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

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

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!

