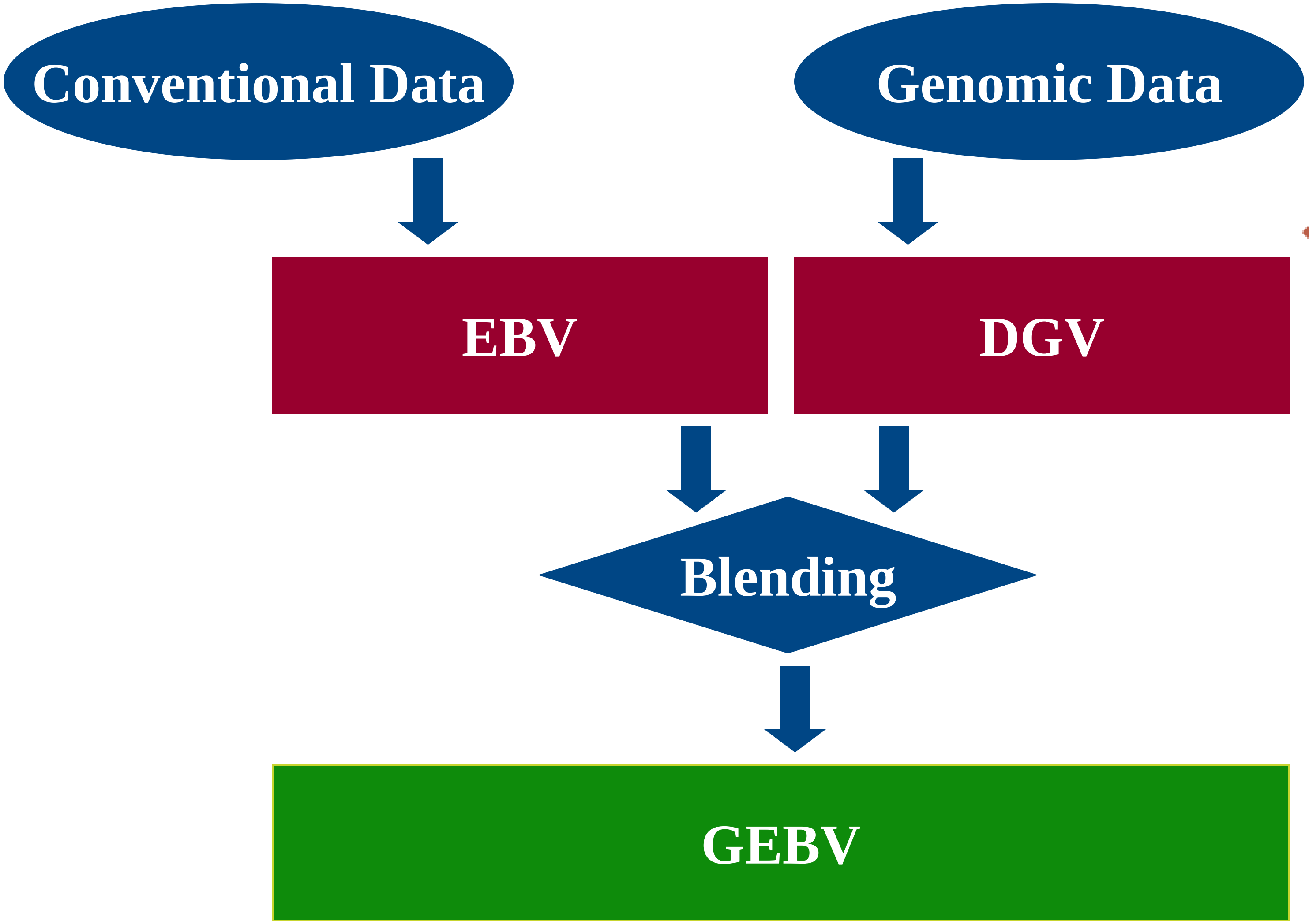
Using Pseudo-observations to combine Genomic and Conventional Data in the Dutch-Flemish National Evaluation

Herwin Eding
W.M. Stoop, M.L. van Pelt, L.C.M. de Haer, G. de Jong
CRV – The Netherlands

Content

- GEBV in the Netherlands
- Pseudo-Record Procedure (PSR)
- Results:
  - Compare EBV vs. GEBV_{PSR}
  - Compare $\text{GEBV}_{\text{BLEND}}$ vs. GEBV_{PSR}
  - MT GEBV reliability validation
- Conclusions

GEBV in NLD: Post-processing



Conventional Data

Genomic Data

EBV

DGV

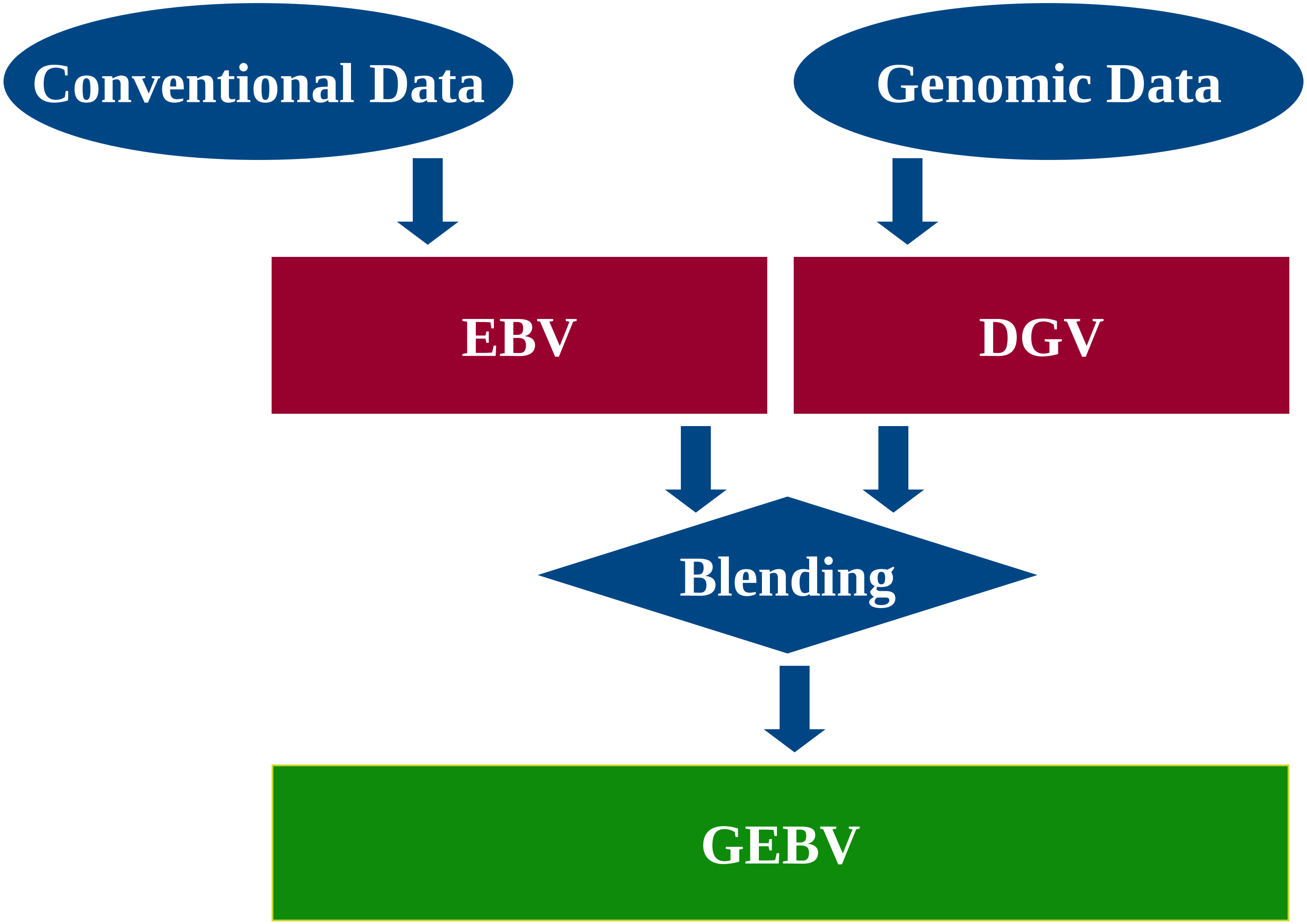
Blending

GEBV

Reasons to introduce PSR

- Simplification of the integration process
- More efficient use of genomic data
- Reduction of bias as a result of genomic pre-selection

Simplification: Post-processing



Conventional Data

Genomic Data

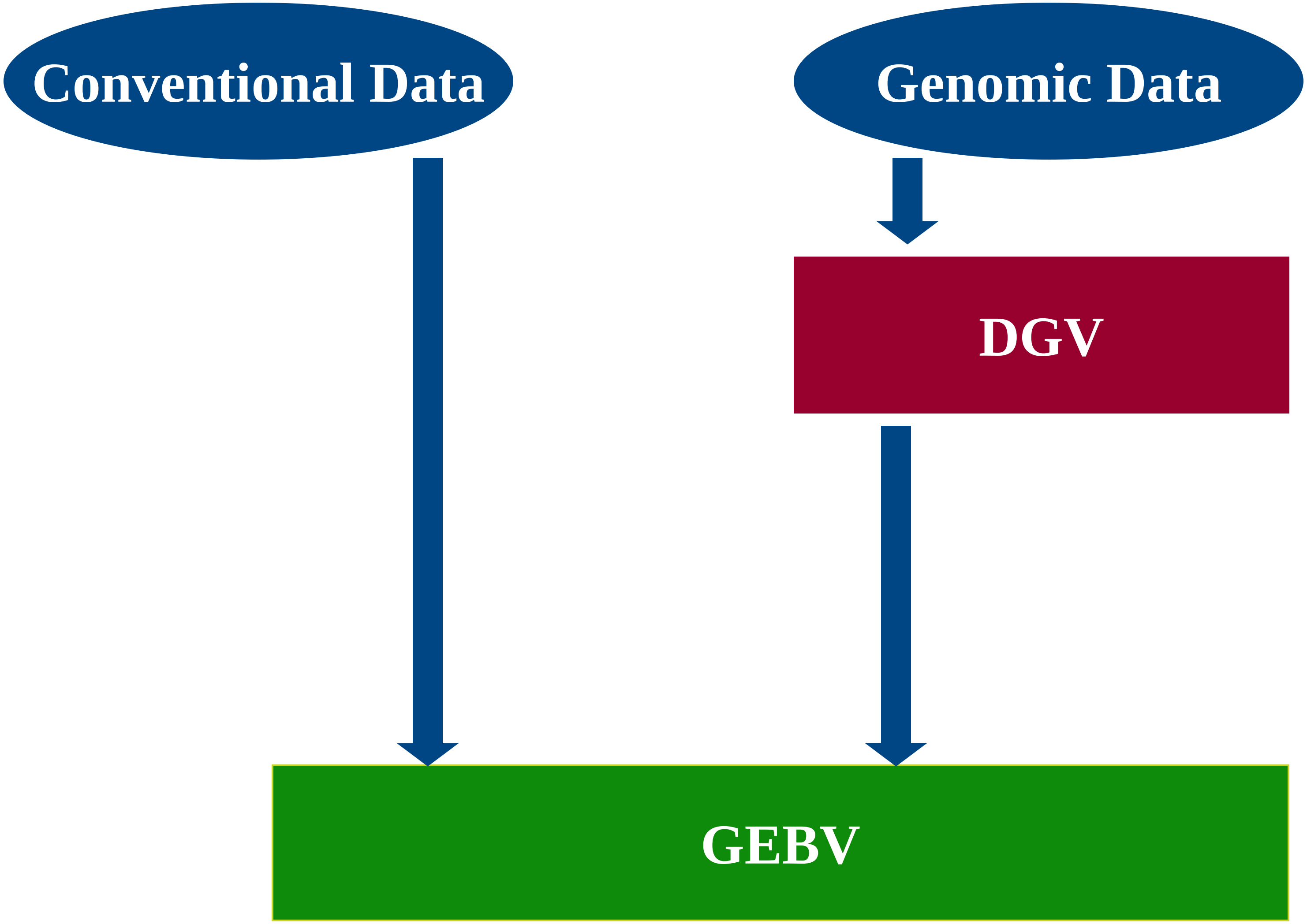
EBV

DGV

Blending

GEBV

Simplification: Pseudo-records



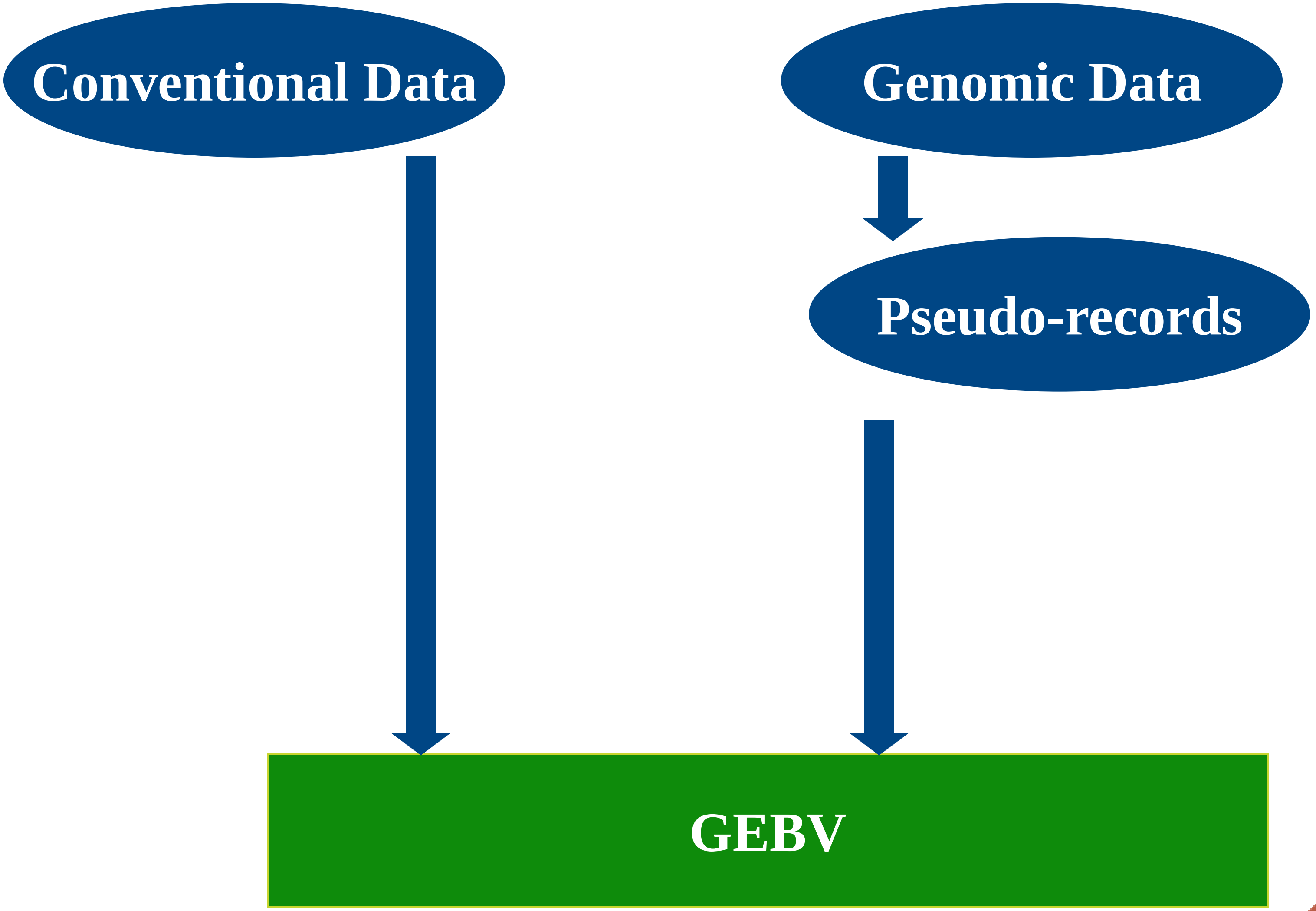
Conventional Data

Genomic Data

DGV

GEBV

Simplification: Pseudo-records

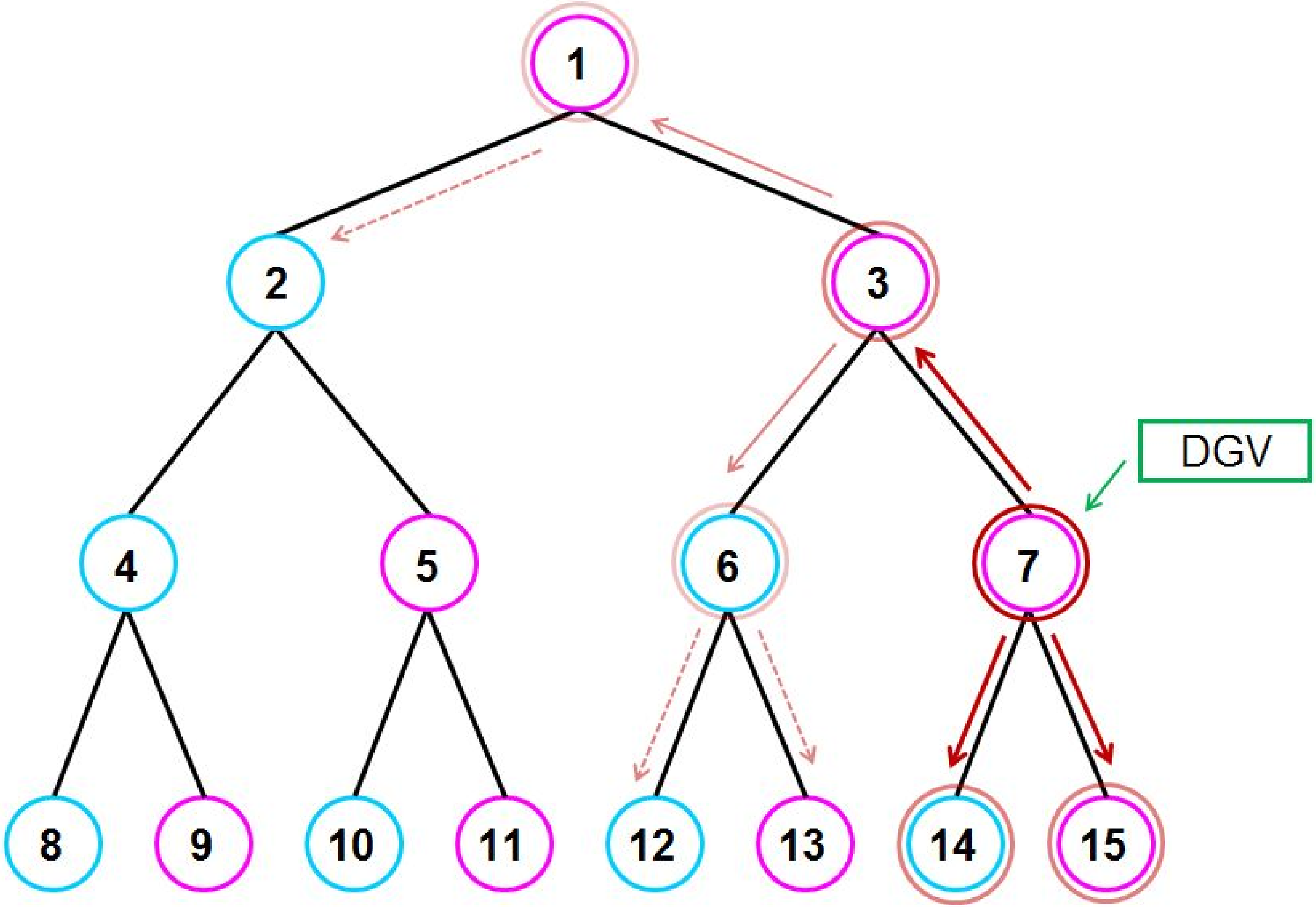

Conventional Data

Genomic Data

Pseudo-records

GEBV

Efficient use of genomic data



Sources of bias in GEBV estimation

- Bias due to genomic preselection
  - Bias due to selection on information outside BVE
    - E.g. foreign breeding values
    - GS makes things worse
  - Incorporating all DGV of all genotyped animals
    - Genomic information in BVE => No/less bias

- Bias due to dam of sire
  - Second source of bias in GEBV
  - Known issue in conventional BVE
    - Causes overestimation of PA => bias in BV
    - Blending uses SIx => no bull dam bias
    - PSR system == AM => Use of PA in GEBV

PSR procedure: Mantysaari and Strandén (2010)

- DGV => pseudo-observation (PSR) on absolute scale
  - h^2 of 0.999
  - $\sigma_{g,psr} = \sigma_{g,original}$ ➔ $\mathbf{G}_{\mathbf{psr}} = \mathbf{G}_{\mathbf{original}}$
- Genomic part of the DGV ~ summation of SNP effects
  - Non-SNP part of DGV is negligible
- Observed variation strictly explained by genetic makeup
  - No residual covariance, $\mathbf{E}_{\mathbf{psr}}$ is diagonal

$$y_{ij} = mu_{PSR} + animal_i + e_{ij}$$

- Multitrait setting with $r_{g\ (pseudo,original)}$ from DGV validation study
  - $r_g = sqrt(R^2_{add})$

PSR procedure: Extension to MT

- Define **R** = diag(**r**), where $r_i = r_{g\text{ (pseudo,original)}}$ for *i*-th trait
- Structure of genetic (**G**) and error (**E**) covariance matrix :

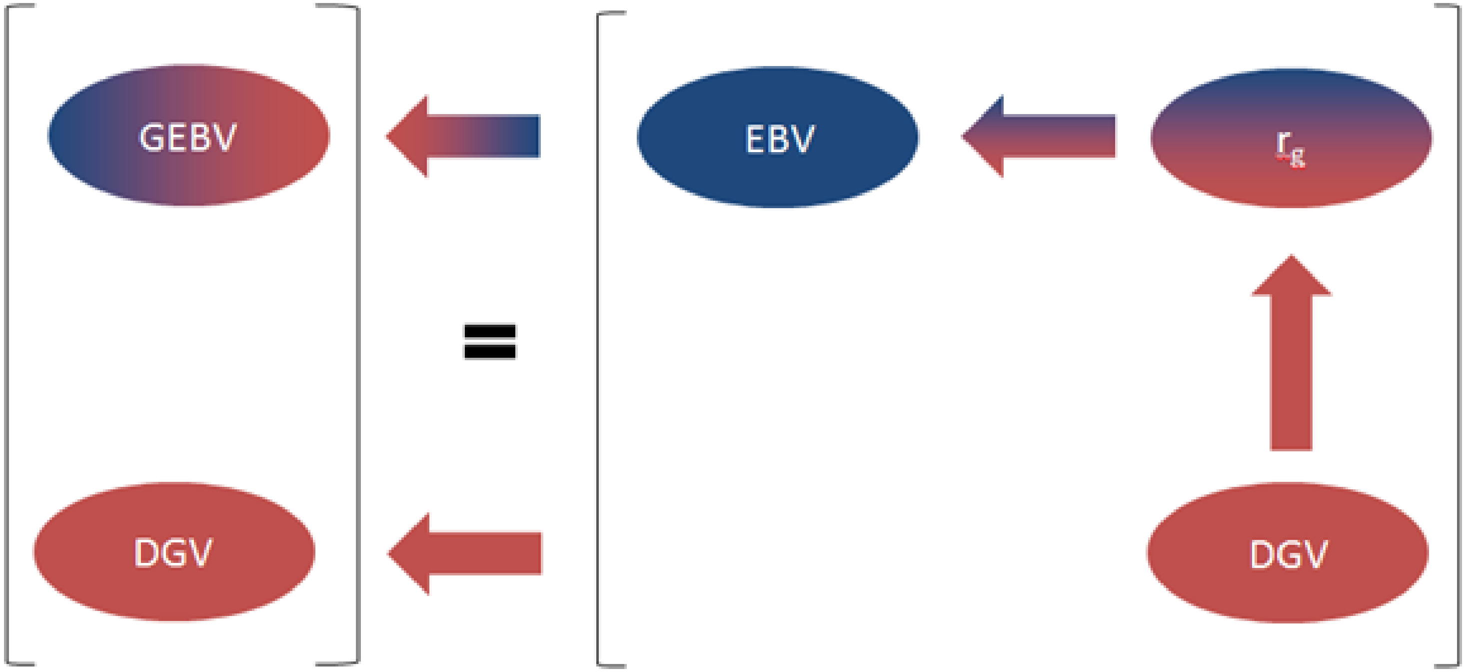
$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & & \\ \mathbf{G}_{12} & \mathbf{G}_{22} & \\ \mathbf{R}\mathbf{G}_{11} & \mathbf{R}\mathbf{G}_{12} & \mathbf{G}_{11} \end{bmatrix}$$

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_{11} & & \\ \mathbf{E}_{12} & \mathbf{E}_{22} & \\ \mathbf{0} & \mathbf{0} & \frac{(1-h^2)}{h^2} dg(\mathbf{G}_{11}) \end{bmatrix} \Rightarrow h^2 \overset{\text{def}}{=} 0.999$$

$\mathbf{G}_{11}$ for traits for which a pseudo-trait exists; $\mathbf{G}_{22}$ for traits without PSR

PSR procedure: Summary

- Automatic integration of DGV using the PSR system



Results: Conformation

Conventional EBV vs. GEBV$_{PSR}$: Expectations

- Only changes in Holstein & Holstein pedigrees

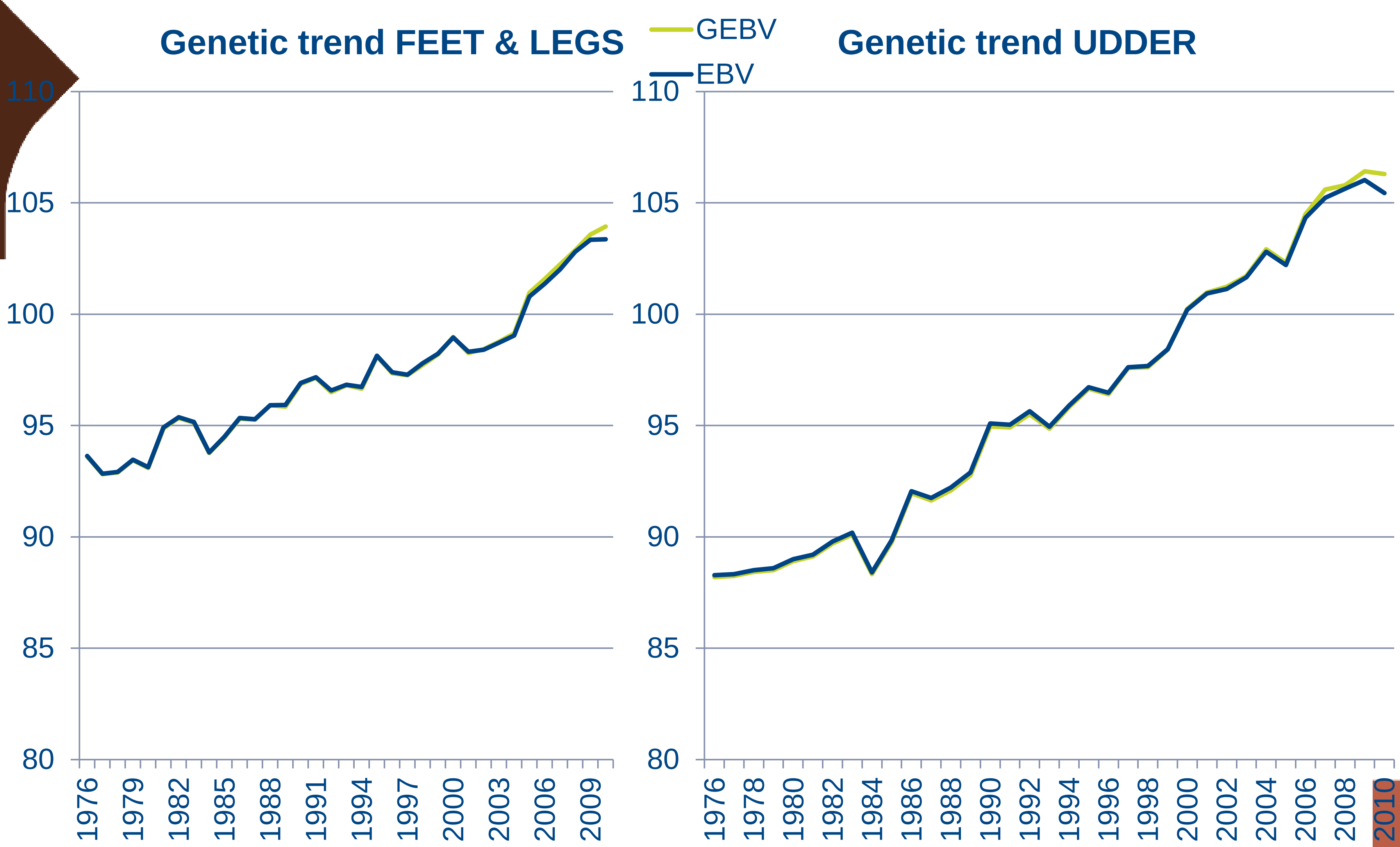
- No PSR & with dau : very small changes (due to pedigree)
- Own PSR & with dau : small changes (some extra data)
- Own PSR, no dau : some changes (from PA/PI -> GEBV)

Conventional EBV vs. GEBV_{PSR}

Difference in Reliability (GEBV – EBV)

Trait	No PSR, > 10 dau (N=15.614)	PSR, >10 dau (N=4.720)	PSR, no dau (N=2.692)
Correlation BV	~ 1.00	~ 0.99	~ 0.90
Frame	0	8	28
Dairy strength	0	9	23
Udder	0	6	29
Feet&Legs	0	6	20

Conventional EBV vs. GEBV_{PSR}



Blending vs. PSR: Expectations

- Only changes in Holstein & Holstein pedigrees

- Own PSR & with dau : very small changes (due to pedigree)
- Own PSR, no dau : small changes (due to pedigree)

Blending vs. PSR

GEBV ~N (100, 4)

Trait	PSR	BLEND	diff	corr
Frame	104.0	103.6	0.4	0.86
Dairy strength	103.8	103.1	0.7	0.92
Udder	105.7	105.3	0.4	0.95
Feet&Legs	103.3	102.1	1.2	0.92

GEBV reliability

Trait	PSR	BLEND	diff
Frame	64.5	53.8	10.7
Dairy strength	57.7	53.8	3.9
Udder	66.2	63.3	2.9
Feet&Legs	55.3	52.1	3.2

Blending vs. PSR

GEBV ~N (100, 4)

Trait	PSR	BLEND	diff	corr
Frame	103.3	102.9	0.4	0.86
Dairy strength	103.7	103.2	0.5	0.92
Udder	106.4	105.8	0.6	0.95
Feet & Legs	103.0	102.5	0.5	0.92

GEBV reliability

Trait	PSR	BLEND	diff
Frame	64.5	53.8	10.7
Dairy strength	57.7	53.8	3.9
Udder	66.2	63.3	2.9
Feet & Legs	55.3	52.1	3.2

Return of the Bias

Mean difference of GEBV from PSR and blending vs.
mean difference GEBV from blending with PA or SIx

Trait	Diff GEBV (psr - blend)	Bias cause by PA *
Milk	198,5	190,0
Fat	8,1	8,4
Protein	6,8	6,7
Somatic cell count	-0,2	n/a
Frame	0,4	n/a
Dairy strenght	0,5	n/a
Udder	0,6	0,7
Feet & Legs	0,5	0,7

Bias = (GEBV from PA) – (GEBV from SIx) for young bulls without daughters

(memo ‘Bias in integrated breeding values’, 2009, R&D/09.0085/HE/GvO)

Blending vs. PSR

GEBV ~N (100, 4)

Trait	PSR	BLEND	diff	corr
Frame	104.0	103.6	0.4	0.86
Dairy strength	103.8	103.1	0.7	0.92
Udder	105.7	105.3	0.4	0.95
Feet&Legs	103.3	102.1	1.2	0.92

GEBV reliability

Trait	PSR	BLEND	diff
Frame	64.5	53.8	10.7
Dairy strength	57.7	53.8	3.9
Udder	66.2	63.3	2.9
Feet&Legs	55.3	52.1	3.2

Blending vs. PSR in short

- For EXT (and PRD) we see a re-introduction of bias in GEBV
  - GEBV based on PA, where it was a PI
  - Correction needed

- The GEBV from PSR have greater reliability then GEBV from blend
  - Possibly an effect of multi-trait BVE
  - Can it be confirmed?

➔ New validation study predicted vs. realized reliability

  - Using Fertility as a case study

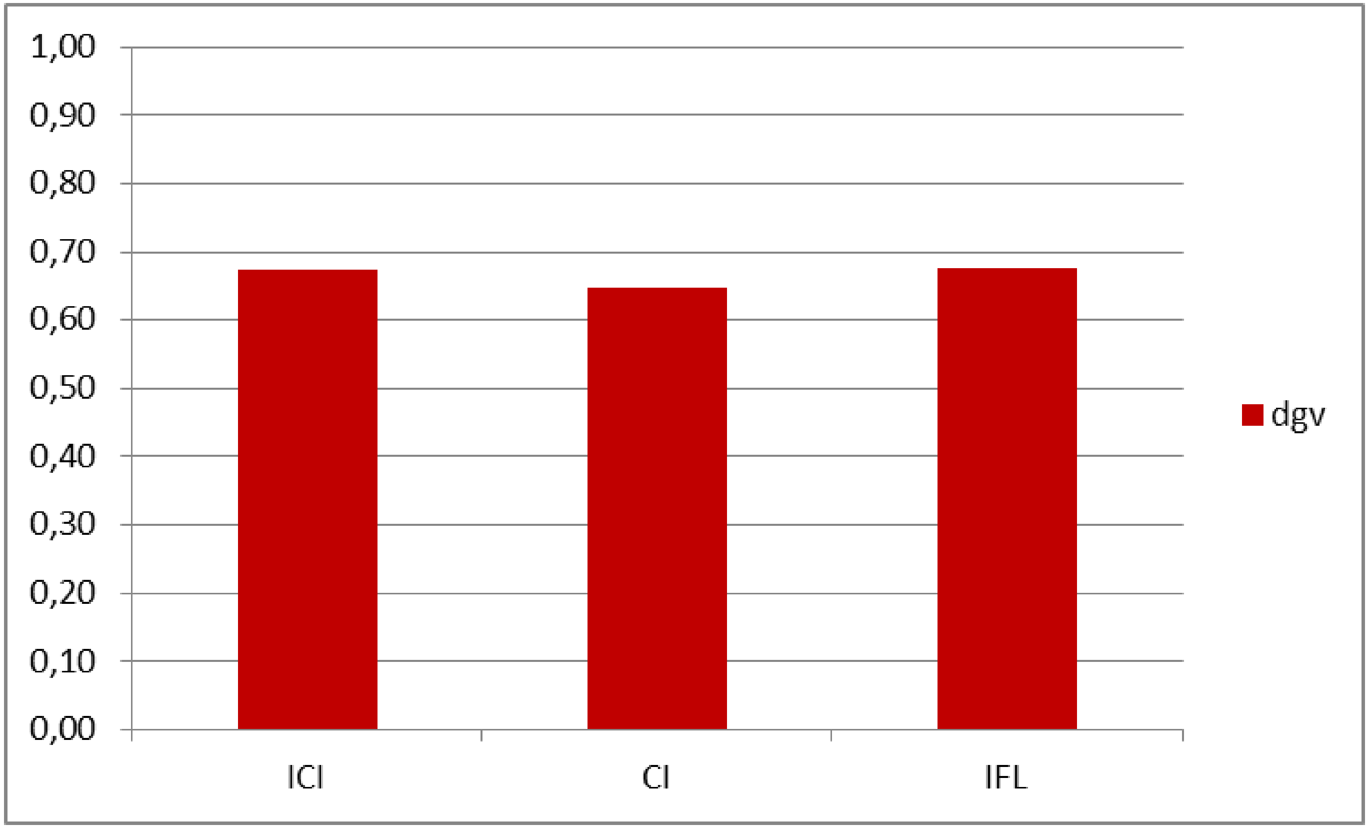
GEBV predicted and realized reliabilities

- Predicted reliability is R^2_{gebv} from PSR BVE
- Compare predicted to realized reliability, based on R^2_{ebv}

$$R^2_{real} = (r_{ebv,gebv} / \sqrt{R^2_{ebv}})^2$$

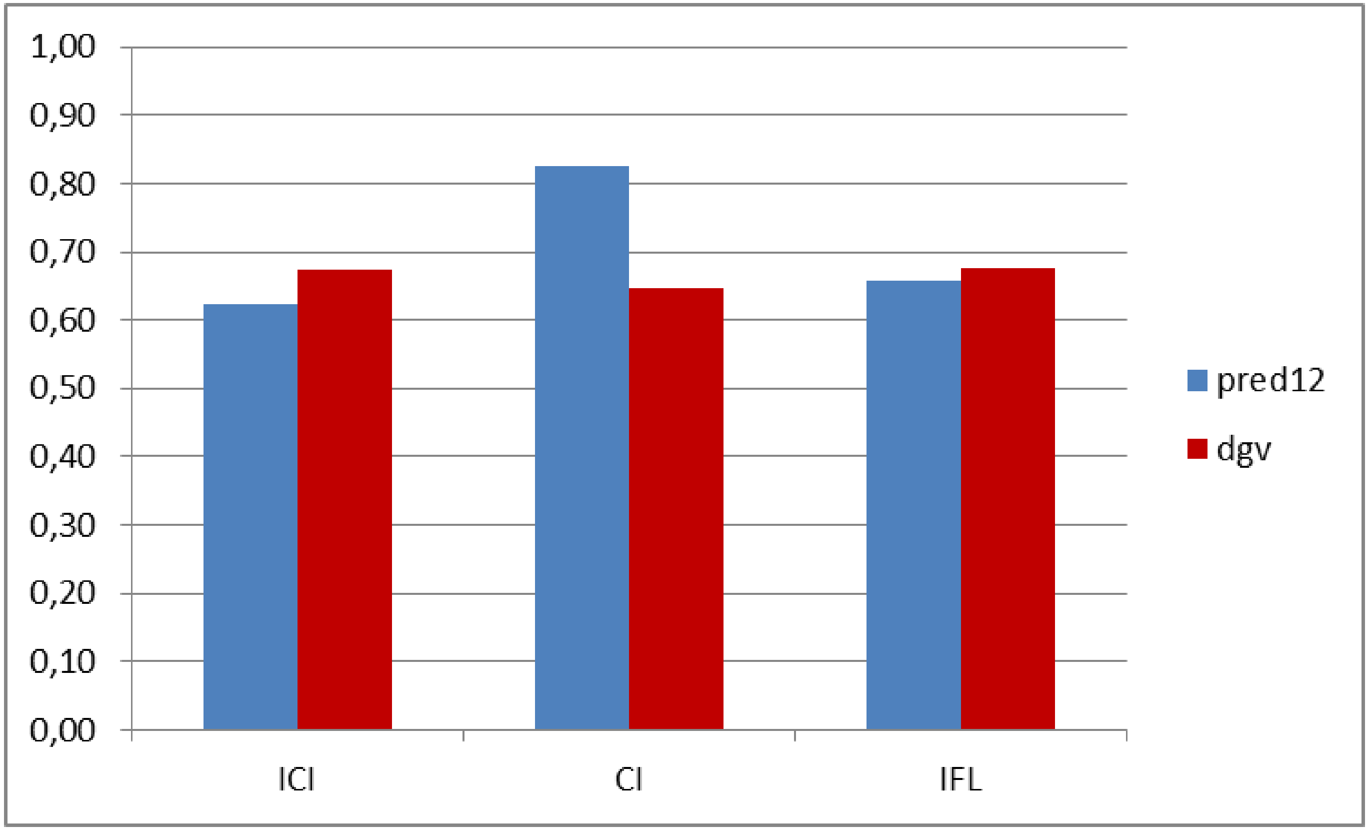
- Example: Three fertility traits
  - – ICI : Interv. Calving - 1st ins.
  - – CI : Calving interv. ~ 57% of FERT
  - – IFL : Interv. 1st – last insemination ~ 43% of FERT

GEBV predicted and realized reliabilities



Realized reliability of raw DGV ~ 0,65

GEBV predicted and realized reliabilities



Validation 2012: Reference = bulls with phenotype EDC > 10

➔ predicted rel. CI > 0.80

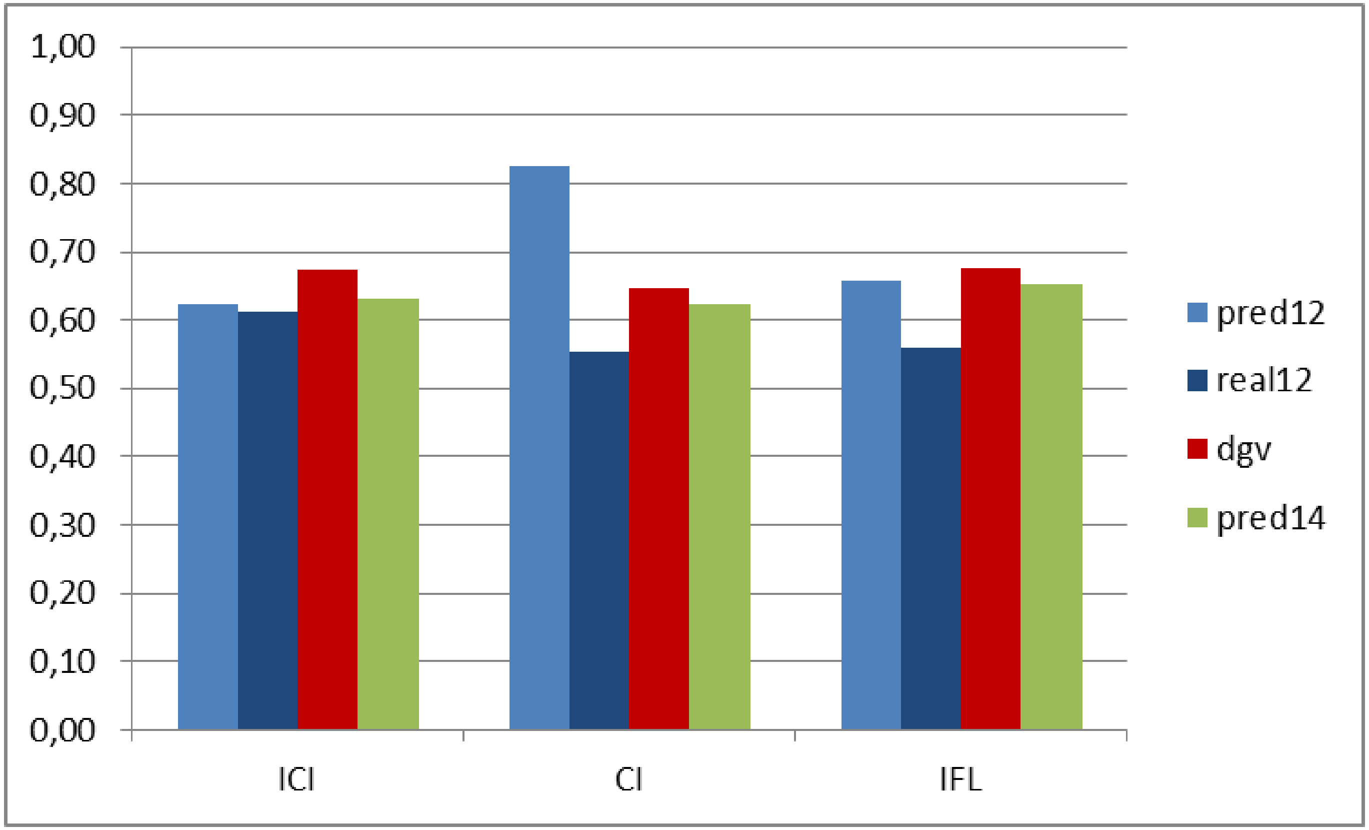
(?) Effect of multi-trait estimation?

GEBV predicted and realized reliabilities

Validation 2012: Reference = bulls with phenotype EDC > 10

➔ predicted rel. CI > 0.80, realized = 0.55 << DGV

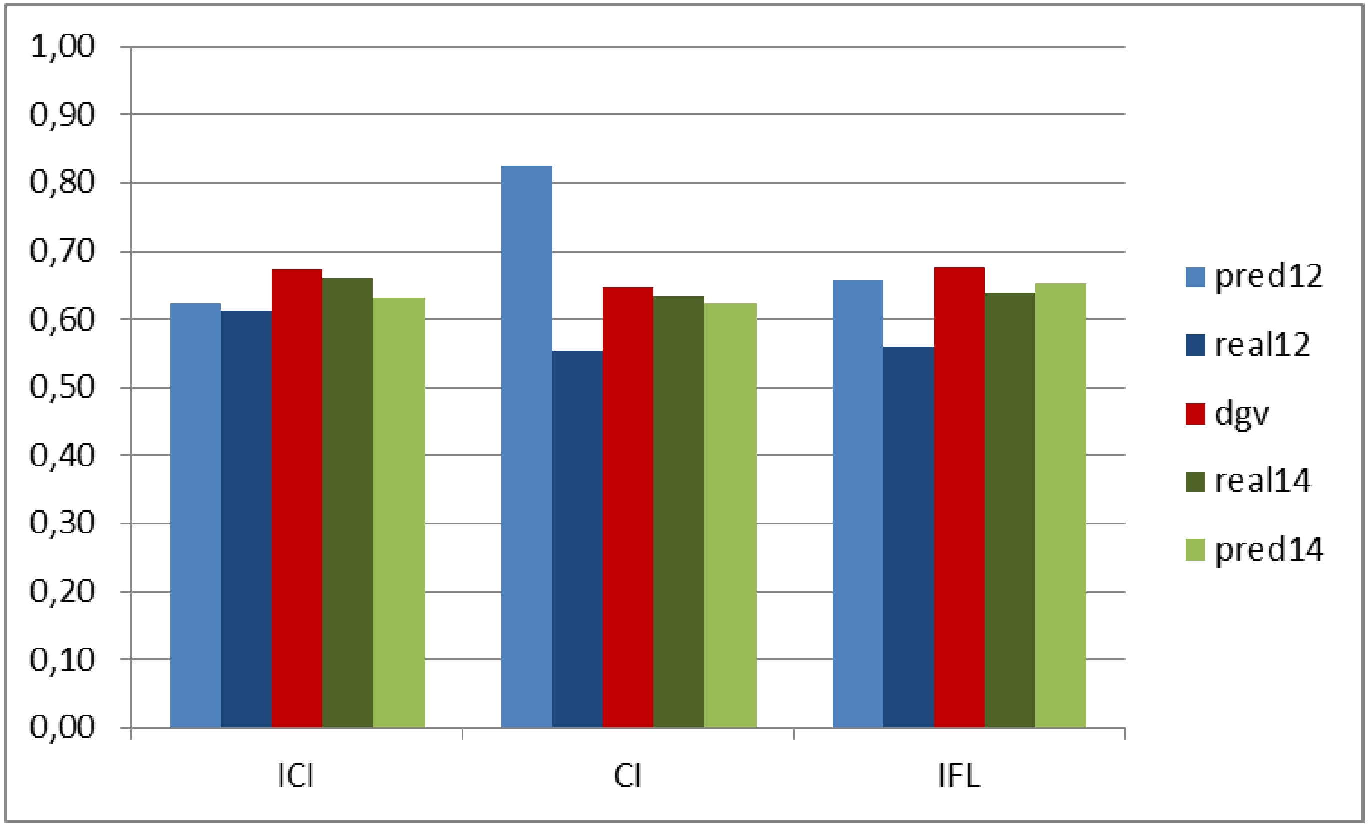
GEBV predicted and realized reliabilities



Validation 2014: Reference = bulls with phenotype rel. > 50%

➔ predicted rel. CI = 0.62

GEBV predicted and realized reliabilities



Validation 2014: Reference = bulls with phenotype rel. > 50%

➔ predicted rel. CI = 0.62, realized = 0.63

➔ In line with realized rel. DGV

GEBV predicted and realized reliabilities

- Change in definition of reference population increased accuracy of DGV validation (R^2_{dgv} and R^2_{pi})

  - Better validation results ➔ More accurate $r_{g,psr}$
    - predicted rel. ≈ realized rel.

- Predicted and realized reliability of PSR GEBV ≈ R^2_{add}
    - Hence reliability_{psr} ~ $\text{reliability}_{blend}$
    - MT-effect on reliability is not real

General conclusions

- Method of Mantysaari and Strandén (2010) seems to work well, results are according to expectations
- GEBV validation shows that increase in reliability (MT setting) is comparable to old blending method.
- Some BIAS left in CONF (and PROD) when using PA + PSR instead of PI + PSR
- First implementation in December 2014

General conclusions: Bias

- Two sources of bias
  - Genomic pre-selection
  - Bull dams
- Genomic pre-selection: PSR system takes care of this
- Bull dams: Not solved
  - Possibly issue even in Single Step Genomics
  - Young bulls without daughters do not receive PSR GEBV
    - Receive integrated GEBV (blending).
- Next step: Single Step Genomics

THANK YOU