

Measuring genomic pre-selection in theory and in practice

Paul VanRaden and Jan Wright

Animal Improvement Programs Lab, Beltsville, MD

Paul.VanRaden@ars.usda.gov

In Theory

- **Traditional mixed models do not account for genomic selection**
 - Phenotypes only for animals with highest Mendelian sampling
 - **GBV** differ from **EBV** for progeny, mates, parents, or herdmates
- **Multi-step methods may be biased**
- **Single-step methods reduce bias**



1-Step Relationship Inverse

Aguilar et al. (2010)

$$H^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} + G^{-1} - A_{22}^{-1} \end{bmatrix}$$

1 = non-genotyped animals (60 million)

2 = genotyped animals (400,000)

Traditional and 1-Step Models

➤ Animal model

- $EBV = w_1 PA + w_2 YD + w_3 PC$
- (parent average, yield deviation, progeny contribution)

➤ 1-step genomic info (GI) model

- $GBV = w_1 PA_g + w_2 YD_g + w_3 PC_g + w_4 GI$
- $GI = \frac{\sum \text{off-diagonal}_j \text{ of } G^{-1} - A_{22}^{-1} (GBV_j)}{\text{diagonal}_i \text{ of } G^{-1} - A_{22}^{-1}}$
- Numerator of w_4 in denominator of w

Diagonals of $G^{-1}-A_{22}^{-1}$

- Computed for 8,300 Brown Swiss
- Diagonals of G and A_{22} are similar
 - G : **mean** $F_G = 3.98\%$ and **SD** = 4.15%
 - A_{22} : **mean** $F_A = 3.95\%$ and **SD** = 2.97%
- Diagonals of G^{-1} larger than A_{22}^{-1}
 - **Mean** = 5.83 for G , 2.18 for A_{22}
 - G^{-1} , A_{22}^{-1} and difference all highly correlated

Correlations of Diagonals

Genotyped Brown Swiss

	G	A	G⁻¹	A⁻¹	G⁻¹ - A⁻¹
G	1.0	.70	.05	.03	.06
A		1.0	.02	-.02	.04
G⁻¹			1.0	.98	.99
A⁻¹				1.0	.94
G⁻¹ - A⁻¹					1.0

Simulations of Pre-Selection Bias

- **Phenotyped only animals with good Mendelian sampling genotypes**
- **Discrete or overlapping generations**
 - **Bias if discrete (Patry, Ducrocq 2011)**
 - **OK with overlap (Nielsen et al, 2012)**
 - **Large bias for dams (Liu et al, 2009)**
- **Actual studies of pre-selection and genomic assortative mating needed**

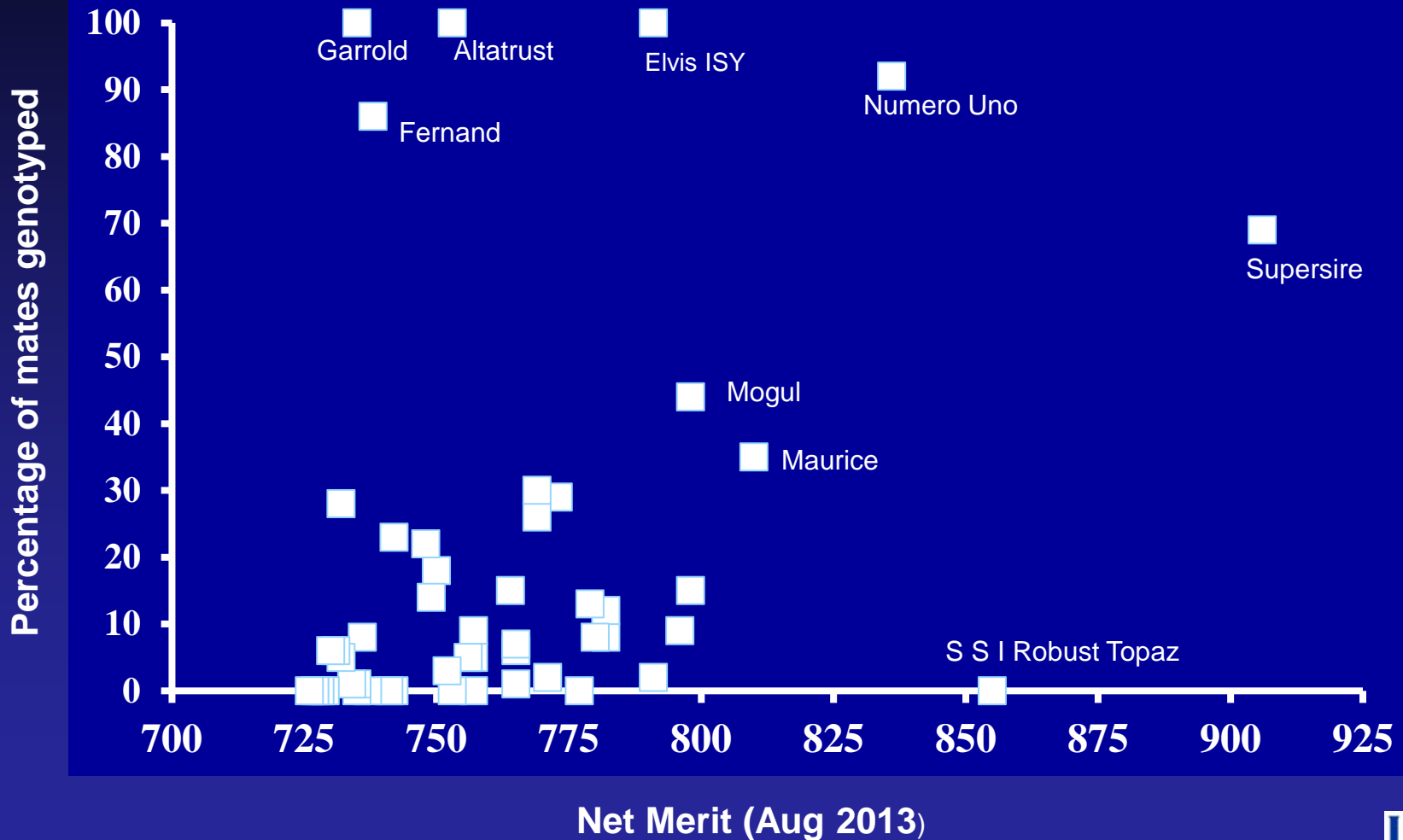


In Practice

- **Test actual selection and mating**
- **Quantify genomic pre-selection in:**
 - **Mates of proven bulls (group 1)**
 - **Mates of young bulls (group 2)**
 - **Dams of young selected sons**
- **Measure future bias because pre-selection has already occurred**

Percentage of Genotyped Mates

Group 2 bulls ranked by NM\$



Mates of Group 1 and 2 Bulls

- **Group 1 (proven bulls)**
 - Daughters with records
 - Top 50, no dtrs in April 2010
 - Were mates pre-selected?
- **Group 2 (young bulls)**
 - Top 50, born 2009 and 2010
 - Study calves born in USA
 - Will pre-selected mates cause bias?



Group 1 Realized Mate Bias

Trait	Mean	SD	Min	Max
NM\$	2	4	-5	13
Protein	0	0	-1	1
Prod Life	.0	.0	.0	.1
Dtr Preg Rate	.0	.0	.0	.1
SCS	.00	.00	-.01	.01
Final Score	.00	.01	-.04	.02
Udder Depth	.00	.01	-.02	.02

Group 2 Future Bias from Mates

Trait	Mean	SD	Min	Max
NM\$	8	9	0	33
Protein	0	0	0	1
Prod Life	.1	.1	.0	.5
Dtr Preg Rate	.1	.1	.0	.3
SCS	-.01	.01	-.03	.00
Final Score	.02	.03	-.01	.10
Udder Depth	.03	.04	-.01	.13

Future Bias – Dams of Young Bulls

- Dams with ≥ 1 sampled son born 2008 to 2012
- Son selection differential = $\sum(\text{GPTA} - \text{PA}) / \# \text{ of sons sampled}$
- Dam's bias = $2 * \text{sons' selection differential} * (\text{DE from sampled sons}) / (\text{total conventional DE})$
 - DE = daughter equivalents or EDC

Example Dam

HOUSA000065597532

- 29 sons genotyped, 6 selected, each will provide 5.4 DE
- Son selection differential for milk = $\sum(\text{GPTA} - \text{PA}) / 6 = 583$ pounds
- 30 daughters, each provide 1.5 DE
- 8.3 DE from PA, 7.8 from records
- Dam's future bias = $2 * 583 * 6 * 5.4 / [8.3 + 7.8 + 6 * 5.4 + 30 * 1.5] = 808$

Expected Future Bias – Bull Dams

Trait	Mean	SD	Min	Max
NM\$	29	33	-124	156
Protein	1	3	-10	14
PL	.3	.5	-1.7	2.0
DPR	.1	.2	-.9	.9
SCS	-.01	.04	-.22	.14

Potential Biases

- **Also from preferential treatment**
 - High-priced early daughters
 - Lack of random sampling
- **Deregression removes some bias**
 - Example: dam gets credit only for own records and non-genotyped progeny, not genotyped sons
 - Use matrix instead of simple one at a time deregression

Conclusions

- Evaluations should adjust for **GBV** instead of **EBV** of:
 - Progeny, mates, contemporaries, and parents
- Biases from pre-selection:
 - Very small for recently proven bulls
 - Moderate from mates top young bulls
 - Will be large for dams of several highly selected sons, but deregression can remove some of the bias

Acknowledgments

- Phenotypes, genotypes, and pedigrees were provided by the Council on Dairy Cattle Breeding