

# Effects of Genomic Selection on Traditional Animal Model Evaluations

L. R. Schaeffer

March 24, 2009

## 1 Introduction

### 1.1 Random Sampling

A traditional animal model has the implicit assumption that the progeny of an animal that have records in the data are a **random sample** of all possible progeny that an animal could have produced. The average relationship between the animal and its progeny would be 0.5, across all parents and progeny. The additive genetic relationship matrix, ( $\mathbf{A}$ ), is a reasonable approximation to the correct covariance structure between animals IF the assumption of random sampling is valid.

### 1.2 Genomic Sampling

Today, genomic estimated breeding values, (GEBV), can be calculated at birth of an animal. Amongst several progeny of the same animal, then the animals with the better GEBV can be selected to be in the herd or to be progeny tested or to be mated to other animals. The progeny with low GEBV would be discarded. An animal's true breeding value (TBV) is

$$\text{TBV} = 0.5(\text{Sire TBV} + \text{Dam TBV}) + \text{Mendelian Sampling},$$

and, therefore, the GEBV allows the selection of progeny with the better Mendelian Sampling contribution. The additive relationship of the selected progeny with their parents is likely greater than the expected average of 0.5. The actual value can be approximated from the genotypes for 50,000 SNPs between parent and offspring. The resulting relationship matrix is better than  $\mathbf{A}$  in approximating the correct covariance structure, but for the case of millions of animals, the inverse of this matrix is not possible to calculate.

### 1.3 Selection of Sires

In dairy cattle, vastly improved genetic gains are predicted for selecting young bulls at birth based on their GEBV. Depending on the country this may mean the young bull is selected for progeny testing, or is utilized immediately as a proven sire in the population. The progeny of the young bull would be a **random sample** of all possible progeny, but the genetic relationship between the young bull and its sire and dam would be different from the average values assumed in the  $\mathbf{A}$  matrix. Thus, an EBV computed from progeny data of this bull would be unbiased as long as the bull's sire and dam information were not included. Bias would be introduced if the usual relationship matrix were used in the analysis, containing the incorrect relationship between the bull and its parents. This bias would then flow to the bull's progeny and the mates that produced those progeny.

## 1.4 Selection of Cows

Dairy producers also want to genotype their animals. A newly born female calf could be genotyped and, based on its GEBV, could be selected for entry into the herd. If so, then this female would not be a **randomly sampled** progeny of its sire and dam. This would violate the implicit assumption of **random sampling** for the animal model. In addition, the relationship in the **A** matrix would not be correct between that female and its parents.

## 1.5 Summary of Introduction

The successful application of GEBV to dairy cattle, will cause bias to genetic evaluations from traditional animal models through violation of the implicit assumption of **random sampling** of progeny needed for the animal model.

Combining GEBV with traditional EBV from an animal model may not be the best strategy to implement, given that EBV will likely be biased. The violation of **random sampling** of progeny of sires and dams will cause biases whether or not genotypes are included in the animal model. Such biases have been demonstrated by Patry and Ducrocq (2009). GEBV should be unbiased, and should not be combined with EBV from data and animal models. Patry and Ducrocq (2009) suggest finding methods to account for the selection intensity in the animal model.

The purpose of this paper is to explore the extent of the bias that might exist as a result of a successful GEBV selection strategy.

# 2 Effects on the Animal Model

## 2.1 Simulation Population

A starting population of 50 males and 1000 females were randomly generated for a trait with heritability of 0.4. The males and females were randomly mated to produce 5 progeny each. A GEBV was produced for each animal assuming an accuracy of 0.70. Then 50 males and 1000 females were chosen from amongst all progeny and the parents based on their GEBV. Records were created for every selected female to be included in genetic evaluation, and only one record per female. Selections and matings were continued through 5 overlapping generations, to provide a total of 6300 animals. Half the sires and half the dams were replaced each generation. Fifty replicates were run for each scenario.

Scenario 1 was where all animals were randomly chosen to be parents of the next generation. That is, there was no selection based on GEBV or other genetic basis. This was the control population in that there was no violation of the assumptions for the animal model.

Scenario 2 was where progeny were selected on the basis of their GEBV (one out of 5 progeny created), and parents of the next generation were also selected based on their GEBV. GEBV

were assumed to have a correlation with true breeding values of 0.70. The implicit assumptions of the animal model were violated to the extreme.

## 2.2 Animal Model Analysis

A traditional animal model with animal additive genetic effects was applied to the records. The purpose was to demonstrate the impact of genomic selection on the accuracy of evaluations. The model was

$$y_i = \mu + a_i + e_i,$$

where  $\mu$  is the overall mean,  $a_i$  is the additive genetic effect of the  $i^{th}$  female, and  $e_i$  was the residual effect from a Normal distribution with mean 0, and variance  $\sigma_e^2$ . Only animals with records and their ancestors were used to create the usual additive genetic relationship matrix, **A**. No information on SNP genotypes was included in the analysis.

## 2.3 Comparison Criteria

- Correlations among animal solutions with TBV for sires, dams, and animals with records.
- Rank correlations of animal solutions with true rankings for sires, dams, and animals with records.
- For animals in generation 5,
  - Correlation of solutions with TBV.
  - Rank correlation.
  - Correlation of estimated Mendelian Sampling, EMS, and true Mendelian Sampling, TMS, effects.
  - Correlation of phenotypic record,  $\mathbf{y}$ , and  $\hat{\mathbf{y}}$ .
  - Correlation of estimated Parent Average, EPA, with true Parent Average, TPA.
  - Correlation of estimated Parent Average with estimated Mendelian Sampling effects.
  - Average TBV of top 200 animals based on EBV.
  - Average TBV of top 200 animals based on TBV.

## 2.4 Results and Discussion

**Table 1**

Group	Statistic	Selection	
		Random	Genomic
Sires	Cor(EBV,TBV)	.851	.905
	Rank Cor(EBV,TBV)	.779	.844
Dams	Cor(EBV,TBV)	.721	.880
	Rank Cor(EBV,TBV)	.641	.775
Animals with records	Cor(EBV,TBV)	.721	.910
	Rank Cor(EBV,TBV)	.640	.815
Animals, Gen. 5	Cor(EBV,TBV)	.733	.850
	Rank Cor(EBV,TBV)	.698	.819
	Cor(EMS,TMS)	.474	.392
	Cor( $\mathbf{y},\hat{\mathbf{y}}$ )	.864	.870
	Cor(EPA,TPA)	.832	.914
	Cor(EPA,EMS)	.199	.207
	Average TBV top 200	4.81	27.99
	Average TBV Maximum	6.52	29.25

For nearly all comparison criteria, the animal model results using data created by genomic sampling of animals were better for genomic selection than for random sampling. Correlations of EBVs with TBVs were actually much better. Rank correlations were much better. The only negative result was the correlation between estimated Mendelian Sampling and True Mendelian Sampling effects, which decreased with genomic selection, but not as much as the increase in all other correlations. The biases of genomic selection do seem to appear in the Mendelian sampling effects, as might be expected.

Possible biases due to genomic selection can be ignored because the solutions from the animal model allow breeders to select among animals with higher correlations to their true breeding values. Adjusting for bias will likely lower the correlations of EBV with TBV, and thereby result in less genetic gain. The animal model seems to be robust to extreme violations in assumptions about random samples of progeny. While evaluations may be biased, rankings of animals were more accurate and selection of animals on this basis should lead to greater genetic change. Estimation of genetic trends may suffer using biased evaluations from an animal model, but could also be estimated using GEBV. This work should be repeated to validate the conclusions.

## 3 References

2009 Patry, C., V. Ducrocq. 2009. Bias due to genomic selection. Proceedings of Interbull Workshop, January 26-29, Bulletin 39, in Uppsala, Sweden.