

# Procedure for Validation of Trends in Genetic Variance

**W.F. Fikse<sup>1</sup>, L. Klei<sup>2</sup>, Z. Liu<sup>3</sup>, P.G. Sullivan<sup>4</sup>**

<sup>1</sup> Interbull Centre, Uppsala, Sweden

<sup>2</sup> Holstein Association, Brattleboro, Vermont, USA

<sup>3</sup> VIT, Verden, Germany

<sup>4</sup> CDN, Guelph, Ontario, Canada

## Introduction

Several reports mention the existence of systematic trends in genetic variances estimated from bull breeding values (e.g., VanDoormaal *et al.*, 1999). Such trends make bull rankings sensitive to the time period of data that is used for sire variance estimation (Miglior *et al.*, 2001), as conversions depend on ratio of sire variances in each country (e.g., Klei *et al.*, 1999). Moreover, trends in genetic variance will lead to suboptimal selection decisions. A greater proportion of animals will be selected from time periods with higher genetic variance (e.g., Vinson, 1987).

The plan of the Technical Committee to address heterogeneous variances in relation to international evaluations has three components:

- 1) Develop a test to validate trends in estimated genetic variances, this includes determination if input data conform to the assumptions of the model for international evaluations;
- 2) Determine whether the trend in the genetic variance computed at national level (often with animal model) and at international level (assuming a sire model) agree;
- 3) Modify the international genetic evaluation model if heterogeneity of genetic variances occurs only at international level.

The aim of this paper is to present the general outline of a procedure to validate Mendelian sampling variances, and to illustrate parts of this procedure with field data.

## Outline of the procedure

The whole procedure can be divided into three different steps:

1. Computation of Mendelian sampling deviation and its prediction error variance for each animal
2. Estimation of genetic variance within year ( $\sigma_{u_i}^2$ )
3. Testing of the hypothesis whether the regression coefficient of ( $\sigma_{u_i}^2$ ) on year differs from zero.

### **Step 1: Computation of Mendelian sampling deviation and its prediction error variance**

For each animal, the Mendelian sampling deviation is computed as:

$$\hat{m}_a = \hat{u}_a - 0.5 \cdot (\hat{u}_s + \hat{u}_d),$$

where  $u_a$ ,  $u_s$  and  $u_d$  are the predicted genetic merit for the animal, and its sire and dam, respectively. In case of missing parents, solutions for the unknown parent group are to be used instead.

If the reliability for animal, sire and dam are known, the information on the same animals if their relationships are not considered ( $q$ ) can be calculated by solving the following non-linear equation:

$$\begin{bmatrix} 1.5\alpha + q_s & 0.5\alpha & -\alpha \\ 0.5\alpha & 1.5\alpha + q_d & -\alpha \\ -\alpha & -\alpha & 2\alpha + q_a \end{bmatrix}^{-1} = \begin{bmatrix} 1/(\alpha + b_s) & \bullet & \bullet \\ \bullet & 1/(\alpha + b_d) & \bullet \\ \bullet & \bullet & 1/(\alpha + b_a) \end{bmatrix}$$

where  $q$  is the residual to genetic variance ratio, and

$$b_i = \alpha \cdot REL_i / (1 - REL_i),$$

with  $REL_i$  being the reliability for animal  $i$ .

In case of missing parents,  $q$  is constrained to zero.

The prediction error variance of the Mendelian sampling deviation for animal  $i$  is then approximated as:

$$PEV(\hat{m}_i) = \begin{bmatrix} -0.5 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} 1.5\alpha + q_s & 0.5\alpha & -\alpha \\ 0.5\alpha & 1.5 + q_d & -\alpha \\ -\alpha & -\alpha & 2 + q_a \end{bmatrix}^{-1} \begin{bmatrix} -0.5 \\ -0.5 \\ 1 \end{bmatrix} \sigma_e^2,$$

where  $q_s$ ,  $q_d$  and  $q_a$  are the solutions obtained above, and  $\sigma_e^2$  the residual variance. Elements in the row and column vector that correspond to missing parents need to be zeroed.

### Step 2: Estimation of the genetic variance within year ( $\hat{\sigma}_{u_i}^2$ )

The genetic variance can be estimated for a group of animals as follows (Sullivan, 1999):

$$\hat{\sigma}_{u_i}^2 = \frac{1}{q_i} \sum_{k=1}^{q_i} d_k [\hat{m}_k^2 + PEV(\hat{m}_k)],$$

where  $q_i$  is the number of animals in year  $i$ , and  $d_k$  the inverse of the genetic variance (proportional to  $\sigma_u^2$ ) that not has been explained by known parents.

### Step 3: Hypothesis testing

Analysis of  $\hat{\sigma}_{u_i}^2$  with a fixed linear model:

$$\hat{\sigma}_{u_i}^2 = a + b_1 \cdot i + b_2 \cdot i^2 + e,$$

where  $a$ ,  $b_1$ , and  $b_2$  are regression coefficients.

## Illustration

### Data

First lactation records on milk yield for 32,306 Guernsey cows produced between January 1990 and December 1999 from United States were used. Pedigree records on an additional 2861 bulls and 33,710 cows were added. There were 4568 herd-year-season contemporary groups. The mean and median contemporary group size was 7.1 and 6, respectively. Description of the data is in Table 1.

### Method

The data was analysed with a animal model containing the fixed effects of herd-year-season, age at calving and percent Guernsey genes. Genetic and residual variances were estimated with an estimated with an AI-REML algorithm (Misztal *et al.*, 2002). Solutions for the animal effects were saved at convergence. In addition, the block of the inverse of the mixed model pertaining to animal, sire and dam ( $C^{aa}$ ) was saved for each animal.

The true prediction error variance for the Mendelian sampling deviation ( $tPEV_{MS}$ ) was computed as:

$$PEV(\hat{m}_i) = I'C^{aa}I,$$

where **l** is a vector containing -0.5 and 1.0 at the position of parents and animal, respectively.

The prediction error variance of the Mendelian sampling deviations was also approximated as outlined in step 1 above, using either true or approximated reliabilities as input (referred to as  $aPEV_{MS}(tREL)$  and  $aPEV_{MS}(aREL)$ , respectively). Approximate reliabilities were computed according to the procedure described by Misztal and Wiggans (1988).

**Table 1.** Characteristics of the data set.

	Number
Animals	68877
Males	2861
Females	66016
Animals	
with parents unknown	14728
with both parents known	53600
Males	
with 1-5 progeny	2087
with 6-10 progeny	215
with 11-20 progeny	196
with 21-50 progeny	224
with 51-100 progeny	53
with > 100 progeny	86
Females	
with record, with progeny	8156
with record, without progeny	24150
without record, with progeny	33710
Records per CG	
<6 records per CG	11257
6-10 records per CG	9709
11-20 records per CG	9396
>21 records per CG	1944

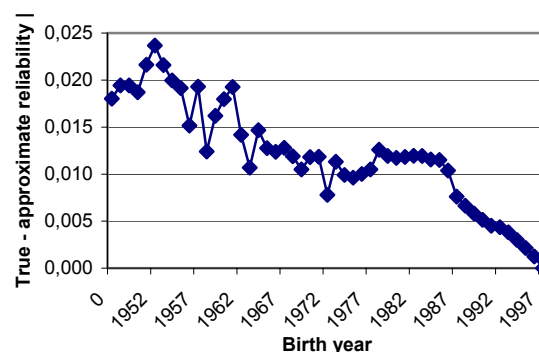
Genetic variance was estimated for the whole population applying the formula in step 2, using  $tPEV_{MS}$ ,  $aPEV_{MS}(tREL)$  and  $aPEV_{MS}(aREL)$ . These three methods were also used to compute the genetic variance within year and gender.

## Results

### *Approximation of reliabilities*

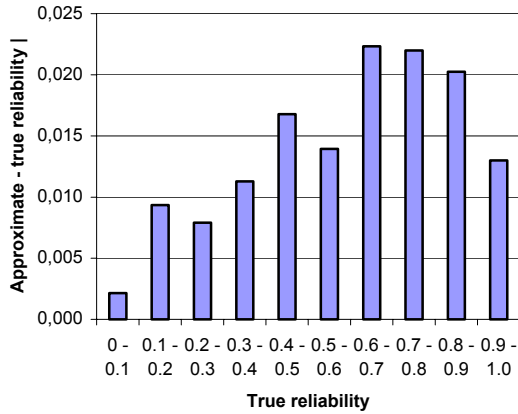
Actual and approximated reliability averaged for all animals was 0.32 and 0.33, respectively. The absolute bias in approximation was on average 0.011. The correlation between true and approximated reliability was 0.999, and the regression of true on approximated reliability was 0.972. These results indicate that there was generally good agreement between true and approximated reliabilities.

The approximation of reliabilities performed worst for females with several daughters and males with 10-50 daughters, especially when parent information was incomplete (Table 2). This was also observed by Meyer (1989), who attributed it to accumulation of errors with the number of progeny. These become less important, however, when the number of offspring is sufficiently large so that the marginal contribution by additional offspring is small. Bias increased with increasing true reliability, but levelled off at the highest level of true reliability (Figure 1).



**Figure 1.** Time trend in bias in approximate reliabilities.

There was a decreasing trend in bias for recent years (Figure 2). The year effect was at least in part due to bias in approximated reliability of the parents, as the effect of birth year was reduced after adjustment for bias in approximated reliability of the parent.



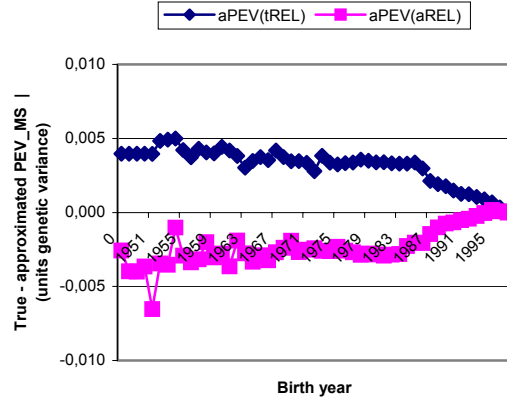
**Figure 2.** Bias in approximated reliability.

#### *Approximation of $PEV_{MS}$*

Actual  $PEV_{MS}$  was on average 0.544 units genetic variance, while approximate  $PEV_{MS}$  based on true and approximate reliabilities was 0.546 and 0.542 units genetic variance, respectively. Regression of  $PEV_{MS}$  on approximate  $PEV_{MS}$  using true and approximate reliabilities was 1.004 and 0.997, respectively.

The number of known parents and the error in approximation of the animal's reliability had the largest effect on bias in approximated  $PEV_{MS}$ . The bias was almost twice as large for cows with records and only one parent known compared to cows records and both parents known (Table 2). The largest bias (0.05 units genetic variance) was found for bulls without parents and more than 10-50 daughters.

The time trend in bias of  $aPEV_{MS}(aREL)$  was nearly flat (Figure 3).

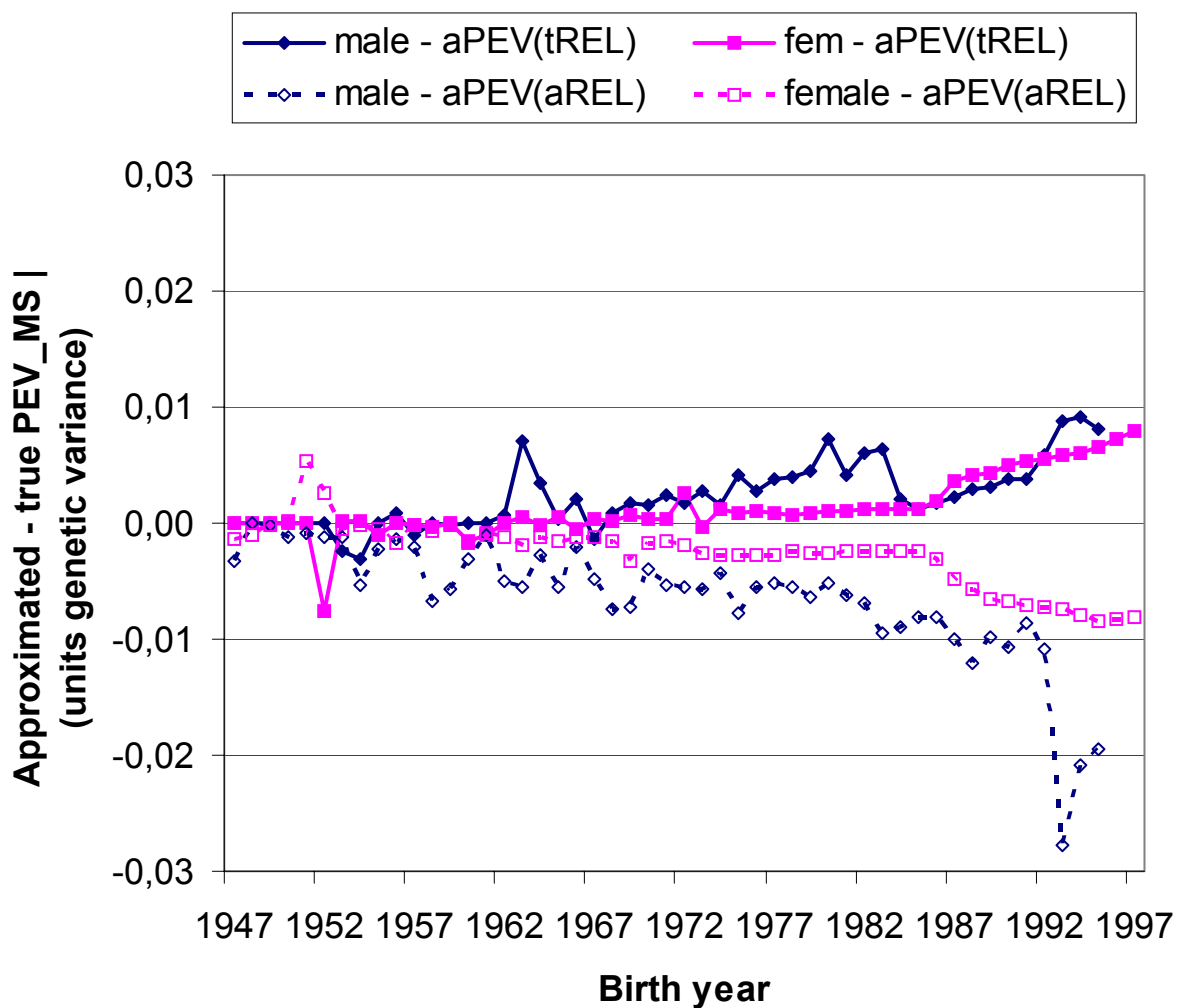


**Figure 3.** Time trend in bias in approximated prediction error variance in Mendelian sampling deviations.

#### *Estimation of $VarA_{ij}$*

The genetic variance calculated with the formula in step 2 using  $tPEV_{MS}$  was 343,930  $kg^2$ , exactly equal to the variance estimated with AI-REML. When  $aPEV_{MS}(tREL)$  and  $aPEV_{MS}(aREL)$  were used instead of  $tPEV_{MS}$ , the estimates were 345,078  $kg^2$  and 342,204  $kg^2$ , respectively.

Difference in genetic variance computed within year and gender using true and either of the approximated  $PEV_{MS}$  was at most 9,540  $kg^2$  (~3% units genetic variance), for males born in 1993 (Figure 4). The number of animals in that group was only 29. There was a tendency for the genetic variance computed with  $aPEV_{MS}(aREL)$  to differ more in recent years, though the magnitude of the difference was relatively small.



**Figure 4.** Error in estimation of genetic variance.

## Discussion

It was feasible to estimate genetic variance within year and gender when good approximate reliabilities were available. Given the wide range of methods that are being applied to approximate reliabilities, it is necessary to investigate the sensitivity of the outlined procedure to poor reliability approximations.

A coefficient of regression of within-year estimates on birth year that is not significantly different from zero would indicate the absence of a time trend in genetic variance. Thus, the test outlined in step 3 would fulfil the objective of the validation procedure. However, an important aspect, tolerance values for the regression coefficient, have not been addressed yet. This will be done during a pilot study.

## References

- Klei, L., Weigel, K.A. & Lawlor, T. 1999. Mace, the relative importance of information sources. *Interbull Bulletin* 22, 9-13.
- Meyer, K. 1987. Approximate accuracy of genetic evaluation under an animal model. *Livest. Prod. Sci.* 21, 87-100.
- Miglior, F., Sullivan, P.G. & VanDoormaal, B.J. 2002. Alternative strategies for estimation of country sire variance in international evaluations of dairy bulls. *Interbull Bulletin* 27, 84-88.
- Misztal, I. & Wiggans, G.R. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* 71, 27-32.

- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T. & Lee, D.H. 2002. BLUPF90 and related programs. *Proc. 7<sup>th</sup> WCGALP*, Communication n° 28-07.
- Sullivan, P.G. 1999. REML estimation of heterogeneous sire (co)variances for MACE. *Interbull Bulletin* 22, 146-148.
- VanDoormaal, B.J., Kistemaker, G.J. & Sullivan, P.G. 1999. Heterogeneous variances of bull EBVs over time. *Interbull Bulletin* 22, 141-145.
- Vinson, W.E. 1987. Potential bias in genetic evaluations from differences in variation within herds. *J. Dairy Sci.* 70, 2450-2455.

**Table 2.** True reliability and bias in approximation of reliabilities and prediction error variance of Mendelian sampling deviations (as percentage of the genetic variance).

D/P	No	tREL	Difference (tREL-aREL)			Difference (tPEV <sub>MS</sub> -(aREL <sub>MS</sub> ))			
			Mean	SD	Mean Abs	aPEV <sub>MS</sub> (tREL)		aPEV <sub>MS</sub> (aREL)	
						Mean	SD	Mean	SD
Dam without record; unknown parent									
0	28	0.019	-0.001	0.001	0.001	0	0	0.06	0.13
1	11045	0.028	-0.001	0.002	0.001	0	0	0.13	0.17
2	1039	0.060	-0.004	0.005	0.004	0	0	0.04	0.50
3	112	0.089	-0.007	0.006	0.007	0	0	0.72	0.61
4	10	0.188	-0.011	0.004	0.011	0	0	1.08	0.39
5	2	0.202	-0.012	0.002	0.012	0	0	1.24	0.23
Dam without records; one known parent									
0	3	0.167	-0.004	0.002	0.004	-0.01	0.02	0.03	0.09
1	280	0.205	-0.007	0.005	0.007	-0.07	0.16	0.20	0.13
2	49	0.251	-0.011	0.009	0.011	-0.11	0.22	0.48	0.55
3	15	0.266	-0.016	0.014	0.016	-0.32	0.60	0.57	0.44
4	4	0.229	-0.023	0.012	0.023	-0.03	0.03	1.31	0.67
Dam without records; both parents known									
0	354	0.228	-0.005	0.006	0.006	-0.02	0.16	0.10	0.11
1	15239	0.244	-0.007	0.005	0.008	-0.07	0.11	0.08	0.06
2	4099	0.280	-0.010	0.007	0.010	-0.09	0.14	0.18	0.17
3	1079	0.316	-0.013	0.008	0.013	-0.11	0.15	0.27	0.21
4	255	0.369	-0.016	0.009	0.017	-0.14	0.21	0.40	0.28
5	97	0.422	-0.023	0.014	0.023	-0.19	0.21	0.63	0.51
Cows with records									
0/1	596	0.434	-0.015	0.009	0.015	-0.26	0.41	0.93	0.45
0/2	23567	0.475	-0.016	0.007	0.164	-0.30	0.22	0.39	0.19
1/1	39	0.464	-0.014	0.010	0.014	-0.26	0.48	0.90	0.44
1/2	6110	0.501	-0.016	0.006	0.016	-0.31	0.21	0.38	0.18
2/1	9	0.501	-0.012	0.005	0.012	-0.19	0.14	0.88	0.46
2/2	1567	0.524	-0.016	0.006	0.016	-0.32	0.21	0.41	0.20
Sires; unknown parents									
0-5	1881	0.036	-0.003	0.005	0.003	0	0	0.29	0.48
6-10	117	0.167	-0.018	0.012	0.018	0	0	1.82	1.22
11-20	33	0.309	-0.051	0.034	0.051	0	0	5.06	6.45
21-50	11	0.458	-0.088	0.098	0.088	0	0	8.78	9.85
Sires; both parents known									
0-5	212	0.202	-0.006	0.007	0.007	-0.05	0.15	0.10	0.19
6-10	95	0.420	-0.016	0.014	0.017	-0.17	0.26	0.49	0.63
11-20	161	0.567	-0.020	0.008	0.020	-0.32	0.24	0.72	0.42
21-50	211	0.710	-0.023	0.007	0.023	-0.46	0.19	0.90	0.44
51-100	53	0.808	-0.021	0.007	0.021	-0.49	0.34	0.81	0.37
101-500	65	0.912	-0.015	0.003	0.015	-0.59	0.33	0.47	0.28
> 500	19	0.973	-0.011	0.002	0.011	-0.75	0.71	0.14	0.43