

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Faculty Papers and Publications in Animal  
Science

Animal Science Department

---

January 1993

## Effect of Genetic Groups on Estimates of Additive Genetic Variance

C. Pieramati  
*Italy*

L. Dale Van Vleck  
*University of Nebraska-Lincoln, dvan-vleck1@unl.edu*

Follow this and additional works at: <https://digitalcommons.unl.edu/animalscifacpub>



Part of the [Animal Sciences Commons](#)

---

Pieramati, C. and Van Vleck, L. Dale, "Effect of Genetic Groups on Estimates of Additive Genetic Variance" (1993). *Faculty Papers and Publications in Animal Science*. 250.  
<https://digitalcommons.unl.edu/animalscifacpub/250>

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# Effect of Genetic Groups on Estimates of Additive Genetic Variance<sup>1</sup>

C. Pieramati<sup>2</sup> and L. D. Van Vleck

Department of Animal Science, University of Nebraska, Lincoln and  
Roman L. Hruska U.S. Meat Animal Research Center, ARS, USDA, Clay Center, NE 68933-0166

**ABSTRACT:** This study examined the effect of genetic grouping on REML estimates of additive genetic variance with an animal model with selected base populations. A simulated population of 40 animals (20 males and 20 females) was followed under selection or random mating conditions for 10 generations. Each population was replicated 20 or 50 times. Genetic grouping reduced estimates of additive genetic variance in populations with selected base animals, whereas grouping had little effect on the estimate in unselected populations. The effect of

genetic grouping varied according to the quantity and kind of information that was missing (percentage of deleted data and pattern of deletion). When genetic grouping was completely random, the estimates were unaffected. Because including genetic groups in the model for analysis affected the estimates of additive genetic variance, the question should be considered of what value or values for heritability should be used for genetic evaluation when grouping is used to account for prior selection.

Key Words: REML, Genetic Variance, Selection, Genetic Groups

J. Anim. Sci. 1993. 71:66–70

## Introduction

Prediction of additive genetic values for production traits is usually based on data from selected animals. Selection seems to be accounted for by the model if all numerator relationships and all records used for selection decisions are included (Henderson, 1975; Sorensen and Kennedy, 1984b; Gianola and Fernando, 1986). In practice this condition is not met because pedigree information and(or) production records usually do not date back to a single base population (e.g., a distinct generation of non-inbred, unrelated, and unselected animals with records).

The genetic group effect has been used in genetic prediction to account for selection when pedigree information is missing. Quaas and Pollak (1981) suggested a transformation of the mixed-model equa-

tions to incorporate the effect of genetic groups into the breeding values and, therefore, allow solving directly for estimated breeding values. Westell et al. (1988) demonstrated that a simple modification of rules of Henderson (1976) and Quaas (1976) can be used to set-up directly the transformed mixed-model equations developed by Quaas and Pollak (1981). Graser et al. (1987) suggested a method to estimate variance components with selected records. The mixed-model equations they proposed seem to be summarized by Westell's rules.

The aim of this study was to examine how REML estimates of additive genetic variance are affected by genetic grouping to account for prior selection in populations undergoing selection when base animals have been selected at different times. A second goal was to compare those estimates with results of van der Werf and de Boer (1990), who did not include genetic groups in their model for analysis. The estimates can also be compared with those reported by van der Werf (1992), while this paper was being reviewed, for a model when selected base animals are fixed.

## Materials and Methods

Data sets were simulated following a Monte Carlo method (e.g., Sorensen and Kennedy, 1984a). Five males and 20 females were randomly sampled from a

<sup>1</sup>Published as Paper no. 9871, Journal Ser., Nebraska Agric. Res. Div., Univ. of Nebraska, Lincoln 68583-0908. The authors wish to thank Karin Meyer for the use of her DF-REML program and the National Research Council of Italy for the fellowship of the senior author during the period of this study.

<sup>2</sup>Istituto Produzioni Animali, Facoltà di Medicina Veterinaria, Università degli Studi di Perugia, via S. Costanzo, 5-06100 Perugia, Italy.

Received March 12, 1992.

Accepted August 20, 1992.

pseudo-infinite base population of unrelated and unselected animals. Each male was randomly mated with four females. One male and one female progeny were generated from each mating (i.e., the 20 males and 20 females that make up Generation 1). In the selected population the five phenotypically best males were selected for the next generation and each was randomly mated with four females, to produce Generation 2 in the same way as described for Generation 1. This procedure was repeated through Generation 10. The generations did not overlap and full-sib mating was permitted. This procedure follows that for small population size in the simulation studies of van der Werf and de Boer (1990).

Records for Generation 0 were simulated as follows:  $y_i = \mu + a_i + e_i$  where  $y_i$  is the record of the  $i^{\text{th}}$  animal (i.e., phenotypic value),  $\mu$  is a phenotypic constant assigned to all records,  $a_i$  is the additive genetic value, and  $e_i$  the random environmental effect on the record of the  $i^{\text{th}}$  animal.

For Generation 0, values for the  $a_i$  were drawn from a pseudo-normal distribution with mean 0 and variance  $\sigma_a^2 = 10$  (Kinderman and Ramage, 1976). In subsequent generations, additive genetic values were simulated as follows:  $a_i = (\frac{1}{2})a_{si} + (\frac{1}{2})a_{di} + \Phi_i$ , where  $a_{si}$  and  $a_{di}$  are additive genetic values of sire and dam of  $i^{\text{th}}$  animal and  $\Phi_i$  is a value corresponding to Mendelian sampling, with mean = 0 and variance,  $\sigma_{\Phi_i}^2 = [\frac{1}{2} - (\frac{1}{4})(F_{si} + F_{di})]\sigma_a^2$  where  $F_{si}$  and  $F_{di}$  are the inbreeding coefficients for the sire and dam of the  $i^{\text{th}}$  animal (Bulmer, 1971). Values for  $e_i$  were drawn from a normal distribution with mean = 0 and variance  $\sigma_e^2 = 10$ , so that the initial heritability,  $h^2$ , was .5. A second population with no selection (random mating) was simulated with the same criteria previously described. In this population the five males were randomly chosen for each generation.

In an attempt to simulate a "real" situation, several incomplete data sets were prepared omitting different data from these two populations. When data were deleted both records and relationship ties from the deleted data were not used.

For the analyses with group effects in the model, animals were assigned to genetic groups according to their sex and generation. In fact, when selection was applied, only males were directly selected, and animals could always be assigned to their true generation using the progeny information. Two proxy groups were used for each generation, one for males and one for females: in fact, animals could always be assigned to their true generation using progeny information. When all the animals assumed to be the "base" are from the same generation, this method of assigning of genetic groups is useless: male and female genetic groups are confounded and all animals will receive the same coefficients for group effects. Such a situation is unusual in real populations because when data recording starts animals are from different

generations or some animals enter the recording system at later generations. In some situations, the available information is not enough to assign genetic groups to account for selection: an extreme possibility is that, with respect to selection, animals are assigned to groups randomly, a situation described here as "random group" effects.

The REML estimates were obtained with a derivative-free algorithm (Smith and Graser, 1986; Graser et al., 1987) using the DFREML programs of Meyer (1988). The genetic group effects were treated as covariates (i.e., without performing the Quaas-Pollak, 1981 transformation). The starting value for  $h^2$  was .5 for all analyses. Required convergence in the DFREML program was  $10^{-9}$ .

## Results and Discussion

Inbreeding coefficients and empirical additive genetic variances were calculated during the generation of data sets (Table 1). These populations differed slightly from those of van der Werf and de Boer (1990), which is assumed to be because they did not allow full sib matings in the first generation. Results similar to their results were found for inbreeding coefficients only when full sib mating was not allowed in the first generation. The restricted size of the population could explain the similarities in inbreeding coefficients and additive genetic variances for selected and random mating populations. Genetic drift produced some negative means in the selected populations and many mean values quite different from zero in the random mating population.

Required convergence of the solutions was always reached in very few iterates using the polytope method (e.g., Meyer, 1989). The starting value for estimation of heritability had little effect on the number of iterates required for convergence. In every data set very high or low starting values usually required only one iterate more than starting with the true parameter value.

Estimates of additive genetic variances using all information were similar on average to the original genetic variance in both populations (Table 2). There was a direct correlation between the quantity of data omitted from the oldest generations and the smaller estimate and larger empirical standard deviation of the variance estimate. As more early generations were discarded, a smaller estimate of the additive genetic variance was evident, especially in the selected population. The standard deviation of the estimate of the genetic variance increased in both populations. The standard deviations of the estimated genetic variances were larger in the random mating population. The correlation between the quantity of omitted data and the bias in the estimates confirms the results of van der Werf and de Boer (1990), who also

Table 1. Averages of inbreeding coefficients (F) and empirical variances of additive genetic values ( $\sigma_a^2$ ) for 50 replicates for each of 10 generations of selection or random mating

Generation	Selected population		Random mating	
	F	$\sigma_a^2$	F	$\sigma_a^2$
0	—	9.71	—	9.71
1	—	9.62	—	9.62
2	.030	8.75	.033	9.23
3	.060	8.20	.054	8.71
4	.081	7.33	.079	8.37
5	.107	7.52	.103	8.22
6	.129	7.88	.125	8.51
7	.157	7.24	.149	7.92
8	.176	7.07	.171	7.52
9	.201	6.85	.189	7.82
10	.224	6.56	.215	7.60

demonstrated that when the base population is selected, the use of relationships among base animals can reduce the bias of the estimates. However, this information usually would not be available, as was the situation for the simulations shown in Table 2. If relationships among base animals are not available, a fixed effect for the generation might be thought to be a simple way to reduce the bias due to selection in the estimates. The present results (line 2, Table 2) show that including in the model a fixed effect for generation, in an attempt to adjust for the missing data and relationships, gave only slightly different estimates of the additive genetic variance but considerably increased standard deviations of the estimated variance, especially with selection.

Situations with missing information were simulated by using less information as the number of generations increased. In a first data set records of animals of Generation 1 that were not descendants of

Table 2. Mean  $\pm$  standard deviation of estimates of additive genetic variance for 20 replicates of selected or random mating populations omitting data from an increasing number of the oldest generations

Generations included	Selected population	Random mating
1-10 <sup>a</sup>	9.92 $\pm$ 2.13	10.21 $\pm$ 2.72
1-10 <sup>ab</sup>	10.30 $\pm$ 2.97	10.11 $\pm$ 2.84
2-10	9.84 $\pm$ 1.84	9.96 $\pm$ 3.51
3-10	9.33 $\pm$ 1.80	9.83 $\pm$ 3.55
4-10	8.75 $\pm$ 2.09	9.46 $\pm$ 3.62
5-10	8.61 $\pm$ 2.13	9.29 $\pm$ 3.56
6-10	8.37 $\pm$ 2.53	9.33 $\pm$ 3.98

<sup>a</sup>Fifty replicates.

<sup>b</sup>Model for analysis also included a fixed effect for generation.

Table 3. Mean  $\pm$  standard deviation of estimates of additive genetic variance for 50 replicates of selected or random mating populations for models including in the model for analysis genetic groups and(or) generation fixed effects with data for a decreasing number of animals omitted from the five oldest of 10 generations depending on their descending from sires of Generation 0 for both selected and random mating populations

Model for analysis		Selected population	Random mating
Groups	Generations		
No	No	9.63 $\pm$ 2.13	9.48 $\pm$ 2.66
Yes	No	8.82 $\pm$ 2.02	9.15 $\pm$ 2.72
No	Yes	8.49 $\pm$ 2.85	9.26 $\pm$ 2.81
Yes	Yes	9.07 $\pm$ 2.86	9.05 $\pm$ 2.82

the first sire of Generation 0 were deleted. Then deletion was made in Generation 2 for records of animals that were not descendants of the first two sires of Generation 0; in Generation 3, records of animals that were not descendants of the first three sires were deleted; and in Generation 4, records of animals that were not descendants of the first four sires were deleted. The first complete generation was Generation 5. In a second data set all records of females in the first five generations were deleted. Data concerning males were deleted as for the first data set. In Generation 1, the only records were for male descendants through male lines from the first sire of Generation 0; in Generation 2, the only records were for males descending through male lines from the first two sires, and so on. Both data sets were analyzed in both selection or random mating situations, using genetic group effects and generation effects, in combination or alone in the model for analysis.

Models including genetic group and generation effects gave generally lower estimates of genetic variance but larger standard deviations of the estimates than models with neither effect (Tables 3 and 4). This bias was more evident when the population was selected and when only one of the two effects was included in the model. Variance estimates from models with either or both of the effects had greater standard deviations. The increase in the standard deviation is more evident in the selected population and when generation effect is used in the model. The generation effect was not used in generating the data sets but was added to the model of analysis to determine whether it was useful to account for selection effects. When very little data concerning males were used, some disconnected animals with relatively low genetic values in the oldest generations of the pedigree in the selected populations might have led to an increased estimate of additive genetic variance (Table 4).

Table 4. Mean  $\pm$  standard deviation of estimates of additive genetic variance for 50 replicates of selected or random mating populations including in the model for analysis genetic groups and(or) generation fixed effects: in the five oldest of 10 generations all data for females were omitted, whereas data for males were omitted in a decreasing amount by generation depending on their descent through male lines from sires of Generation 0

Model for analysis		Selected population	Random mating
Groups	Generations		
No	No	10.58 $\pm$ 2.60	9.29 $\pm$ 3.07
Yes	No	8.85 $\pm$ 2.55	9.06 $\pm$ 3.12
No	Yes	8.73 $\pm$ 3.67	9.03 $\pm$ 3.35
Yes	Yes	9.39 $\pm$ 3.85	9.04 $\pm$ 3.40

Two other data sets with a progressively decreasing percentage of omitted data according to generation were prepared for both the selection and random mating populations. In the first data set, 80% of the data of Generation 1 was randomly deleted, then 60% of Generation 2, 40% of Generation 3, and finally 20% of Generation 4. In the second data set, 50% of the data for the five oldest generations was randomly deleted. These two data sets were analyzed in three different ways: 1) without using genetic grouping, 2) using genetic grouping, and 3) randomly assigning animals to groups (e.g., there were the same number of groups as with "proxy" groups), but no logical pattern was followed to decide which animals belonged to which group (Tables 5 and 6). Under random mating conditions including either true genetic groups or randomly allotted proxy-like groups in the model had little effect on the estimates. When the population was selected, including random group effects in the model, as expected, did not have much effect on the estimates, whereas analyses with genetic groups assigned by generation gave marked reductions in the additive genetic variance estimates. As mentioned earlier, when very little data were used the

Table 5. Mean  $\pm$  standard deviation of estimates of additive genetic variance for 50 replicates of selected or random mating populations using or not using genetic groups and using randomly assigned genetic groups: data for a decreasing percentage of animals were omitted from the five oldest of 10 generations

Model included	Selected population	Random mating
No groups	9.46 $\pm$ 2.44	9.14 $\pm$ 3.02
True groups	8.26 $\pm$ 2.35	9.07 $\pm$ 3.03
Random groups	9.22 $\pm$ 2.50	9.06 $\pm$ 3.14

Table 6. Mean  $\pm$  standard deviation of estimates of additive genetic variance for 50 replicates of selected or random mating populations using or not using genetic groups and using randomly assigned genetic groups: 50% of the data was omitted from the five oldest of 10 generations

Model included	Selected population	Random mating
No groups	12.59 $\pm$ 3.43	9.59 $\pm$ 3.29
True groups	8.06 $\pm$ 2.53	9.32 $\pm$ 3.05
Random groups	12.07 $\pm$ 3.54	9.70 $\pm$ 3.25

unrelated animals in the pedigree with low genetic values might lead to a larger estimate of additive genetic variance when the population is selected. Overestimation was more evident when the amount of information was the same in each generation because in that case the isolated animals were equally distributed in the oldest generations.

In a real situation, data recording will start at different times according to sex and when parental origin is uncertain. Simulation of this kind of situation was attempted by deleting all data of males or females in the five oldest generations both under selection and random mating (Table 7). In the selected population the use of proxy groups gave estimates close to the expectation of the original genetic variance. For models without genetic groups, estimates for additive genetic variance were overestimated, particularly when data concerning females were deleted. The overestimation can be attributed to the fact that females were selected indirectly. The more that mean variance was overestimated for models without genetic groups, the more it was underestimated for models with groups. In the random mating population, including proxy groups in the model did not affect the estimates very much; the same effect of grouping as in the selected population when records were deleted by sex was found, but the range of differences in the estimates was smaller.

Table 7. Mean  $\pm$  standard deviation of estimates of additive genetic variance for 50 replicates of selected or random mating populations using or not using genetic groups: data of either males or females were omitted from the five oldest of 10 generations

Groups	Omitted sex	Selection	Random mating
No	Males	11.94 $\pm$ 3.04	9.46 $\pm$ 3.15
Yes	Males	9.08 $\pm$ 2.56	9.32 $\pm$ 2.96
No	Females	12.42 $\pm$ 2.90	9.61 $\pm$ 2.85
Yes	Females	8.38 $\pm$ 2.02	9.23 $\pm$ 2.69

## Conclusions

Genetic group effects are a way of accounting for selection when pedigree information is missing. Westell's rules make setting up the transformed mixed-model equations easy, so that genetic grouping is widely used in BLUP evaluations. When animals are selected, the complete omission of records on which selection is based leads to reduced estimates of additive genetic variance. Van der Werf and de Boer (1990) demonstrated that the use of relationships among base animals can remove most of the bias in the estimates of genetic variance, even though records on which selection decisions are based are not available. However, the relationships among base animals may not always be available. Models with genetic group effects might correct for part of the bias in estimates of genetic variance if base animals are not all in the same generation.

Analyses of the simulated data suggest that genetic grouping leads to underestimation of additive genetic variance in populations with selected base animals. This effect of genetic grouping changes according to the amount and the kind of information that is missing (e.g., percentage of deleted data and pattern of deletion). When there was no selection, adding group effects to the model did not affect estimates of additive genetic variance. Therefore, the magnitude of the difference in the estimate of additive genetic variance between a model with genetic groups and a model without groups might be useful in detecting whether selection had occurred before data were available. When genetic grouping was completely random, estimates of  $\sigma_a^2$  were unaffected.

Van der Werf and de Boer (1990) reported underestimation of additive genetic variance when some information was missing; their simulations, however, always deleted whole generations. A model in which selected base animals are considered fixed ("conditional model") has been reported to correct for bias due to selection of base animals but a new bias was introduced as selection continued (van der Werf, 1992), resulting in an underestimation of additive genetic variance. The "conditional" model has been demonstrated to be equivalent to Westell grouping, assuming each base animal is assigned to a separate group, a grouping strategy quite different from any now used. The simulations reported here show that missing information can also lead to an overestimation if omitted records follow some particular patterns; computational problems of REML often require exclusion of some available records and ancestry information, which may lead to similar biases.

Models with genetic groups are used to obtain estimates of breeding values less biased by prior selection, but group effects to account for selection have not been used in models to estimate parameters such as heritability and repeatability. Because genetic grouping seems to affect estimates of additive genetic

variance, the question of whether heritability for BLUP evaluations should be different when grouping is used needs to be considered.

## Implications

Group effects are used in mixed-model procedures to account for prior selection when records are not available on which selection decisions were made. Simulations with missing records from a population undergoing selection indicate that including group effects in the model results in different estimates of variances from those for an unselected base population. Genetic evaluations for populations undergoing selection are unbiased by selection when records used for selection are included. A question this simulation study raises is what variances should be used for genetic evaluations when group effects are included to account for prior selection.

## Literature Cited

- Bulmer, M. G. 1971. The effect of selection on genetic variability. *Am. Nat.* 105:210.
- Gianola, D., and R. L. Fernando. 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.* 63:217.
- Graser, H.-U., S. P. Smith, and B. Tier. 1987. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.* 64:1362.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423.
- Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69.
- Kinderman, A. J., and J. G. Ramage. 1976. Computer generation of normal random variables. *J. Am. Stat. Assoc.* 71:893.
- Meyer, K. 1988. DFREML—A set of programs to estimate variance components under an animal model. *J. Dairy Sci.* 71(Suppl. 1): 33.
- Meyer, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet. Sel. Evol.* 21:317.
- Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32:949.
- Quaas, R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. *J. Dairy Sci.* 64:1868.
- Smith, S. P., and H.-U. Graser. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.* 69:1156.
- Sorensen, D. A., and B. W. Kennedy. 1984a. Estimation of response to selection using least-squares and mixed model methodology. *J. Anim. Sci.* 58:1097.
- Sorensen, D. A., and B. W. Kennedy. 1984b. Estimation of genetic variances from unselected and selected populations. *J. Anim. Sci.* 59:1213.
- van der Werf, J.H.J. 1992. Restricted maximum likelihood estimation of additive genetic variance when selected base animals are considered fixed. *J. Anim. Sci.* 70:1068.
- van der Werf, J.H.J., and I.J.M. de Boer. 1990. Estimation of additive genetic variance when base populations are selected. *J. Anim. Sci.* 68:3124.
- Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. *J. Dairy Sci.* 71:1310.