

Comparisons between genetic variances estimated from different types of relatives in dairy cattle

P. M. Visscher¹† and R. Thompson²

¹*Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT*

²*AFRC Institute of Animal Physiology and Genetics Research, Roslin, Midlothian EH25 9PS*

Abstract

Genetic variances were estimated from different types of relatives for a data set of 19038 first lactation fat yield (FY) and fat concentration (FC) records. For FY, h^2 estimates were 0.377, 0.322 and 0.377 from an animal model, a sire model and a model using only female relationships. The corresponding estimates for FC were 0.632, 0.628 and 0.610. Regarding sires with most progeny in the data set as levels of a fixed effect slightly increased h^2 estimates, most likely through accounting for selection of proven sires. Statistical models were presented to account for heterogeneous genetic variances for males and females in the prediction of breeding values or in estimating variance components. Simultaneously estimating genetic variances for FY from paternal half-sibs, maternal half-sibs, and daughter-dam comparisons, resulted in heterogeneous genetic variance estimates from male v. female comparisons and similar variance estimates from mhs and dd comparisons. For FC, h^2 estimates from different types of relatives were similar.

Dam-daughter comparisons provide 0.75 to 0.80 of the information on heritability from female relationships and provide 0.42 to 0.55 of the total information on heritability.

Keywords: *dairy cattle, genetic variance, heritability, mathematical models, milk fat.*

Introduction

In dairy cattle, prediction of breeding values using an animal model (AM) is replacing separate genetic evaluations for cows and bulls. Some countries have implemented a national AM evaluation, e.g. the USA (Wiggans, Misztal, and Van Vleck, 1988), France (Ducrocq, Boichard, Bonaiti, Barbat, and Briend, 1990), and Australia (Jones and Goddard, 1990), and many others are in the process of doing so. Genetic parameters required for best linear unbiased prediction (BLUP) evaluations are usually estimated using residual maximum likelihood (REML; Patterson and Thompson, 1971) with the same model as used to estimate breeding values. However, AM-REML parameter estimation is computationally

demanding, and few AM estimates have been reported (Swalve and Van Vleck, 1987; Van Vleck and Dong, 1988; Visscher, Thompson and Hill, 1991; Visscher and Thompson, 1992).

An advantage of using an AM, both for prediction of breeding values and for estimating parameters, is that all relationships between animals with records are taken into account. Different types of relatives contributing to parameter estimates in dairy cattle are, for example, paternal half-sibs (phs), daughter-dam (dd) pairs, and maternal half-sibs (mhs). Usually, AM estimates are treated as 'black-box estimates', but a relevant question is how the different animal comparisons contribute to variance component estimates. For example, there is ample evidence of higher heritability estimates from dd regression compared with estimates from phs covariances (see, for example, Majjala and Hanna, 1974; Van Vleck, 1986). Heritability estimates for

† Present address: Department of Food and Agriculture, Livestock Improvement Unit, 166-176 Wellington Parade, East Melbourne, Victoria 3002, Australia.

milk production traits using an animal model (Swalve and Van Vleck, 1987; Van Vleck and Dong, 1988; Visscher and Thompson, 1992) are often higher than previously reported estimates using sire models.

Previously, we quantified contributions from comparisons between bulls and cows by reporting simultaneous estimates of 'male' and 'female' heritabilities (Visscher and Thompson, 1990). The justification for allowing different genetic variances for males and females was the observed difference between *dd* and *p*hs heritabilities from a number of different studies (see above). An additional reason is that animal model REML estimation uses smaller data sets due to limited computer resources, and this may result in selection not being accounted for properly, in particular on the male side. The model proposed by Visscher and Thompson (1990) is, therefore, based on statistical rather than genetical arguments.

The aim of this study was to show how to assess the contributions from relationships between male ancestors (including *p*hs) and female ancestors (including *dd* and *m*hs comparisons), and to quantify such contributions to AM estimates for fat yield and fat concentration. The two traits were chosen because they are known to have different heritabilities, so that the relative (weighted) contribution from various animal comparisons may differ for these traits. A second related aim was to investigate the effect of treating some sires as levels of a fixed, rather than a random effect, as suggested by Meyer (1983), Hill, Edwards, Ahmed and Thompson (1983), and Van Vleck (1985) for sire models, in order to measure genetic variation in the unselected sire population. More recently Graser, Smith and Tier (1987) advocated using some fixed

animals in animal models, when there is insufficient information about the ancestry of certain animals.

Material and methods

First lactation fat yield (FY) and fat concentration (FC) records for the period 1979 to 1987 from 50 large pedigree Holstein-Friesian (HF) herds were extracted from the production files of the Milk Marketing Board. The data set was a subset of that used by Visscher and Thompson (1992). Details of the data structure and the traits are presented in Table 1.

Data were analysed using REML. For all analyses, the fixed part of the fitted model was the same; herd-year-seasons (HYS), with seasons defined in 4-month periods, and month of calving were treated as fixed effects, and proportion North American HF, age at calving and lactation length were fitted as covariables. The following analyses were carried out.

AM standard animal model analysis; AM100 AM analysis: sires with ≥ 100 progeny fixed; AM25 AM analysis: sires with ≥ 25 progeny fixed; AM01 AM analysis: all sires fixed. For the animal models, $y = Xb + Zu + e$, and $v(y) = A\sigma_a^2 + I\sigma_e^2$, with the usual definitions: y , b , u , and e are vectors of data, fixed effects, animal effects and residual effects respectively, and X and Z are known incidence matrices. The matrix A is the numerator relationship matrix, and may be written as TDT' (Thompson, 1977), where T is a lower triangular matrix and D is diagonal matrix containing Mendelian sampling terms. For fixed animals the corresponding elements of D are infinite (from Graser *et al.*, 1987).

SIRE standard sire model; SIRE100 sire model: sires with ≥ 100 progeny fixed; SIRE75 sire model: sires with ≥ 75 progeny fixed; SIRE50 sire model: sires with ≥ 50 progeny fixed; SIRE25 sire model: sires with ≥ 25 progeny fixed. For the sire models, $y = Xb + Zs + e$, and $v(y) = I\sigma_s^2 + I\sigma_e^2$, where s is a vector of sire effects (with identity covariance matrix). The additive genetic variance is estimated as four times the estimate of the between-sire variance.

DAM dam model: dams with records are not recognized as such — their records are only used as progeny records of their dams and sire information is ignored. For the dam model, $y = Xb + Zd + e$, with $v(y) = I\sigma_d^2 + I\sigma_e^2$, where d is a vector of dam effects (with identity covariance matrix). The additive genetic variance is estimated as four times the estimate of the between dam variance.

FEM female relationship model: all sire information is ignored; *dd*, *m*hs and daughter-granddam comparisons are used, cows with records but no

Table 1 Data structure and summary statistics of traits

No. of records	19038	
No. of animals in analyses	28435	
Mean fat yield (s.d.) (kg)	213.1	(45.2)
Mean fat concentration (s.d.) (g/kg)	39.6	(3.95)
No. of males	1280	
with ≥ 100 progeny	34	
with ≥ 75 progeny	47	
with ≥ 50 progeny	89	
with ≥ 25 progeny	199	
No. of females	27155	
No. of dams with progeny	13383	
with ≥ 4 progeny	305	
with ≥ 3 progeny	1150	
with ≥ 2 progeny	4094	
No. of daughter-dam comparisons	7151	

female relatives measured provide information on phenotypic variation.

AM ♀ ♂ AM with heterogeneity of additive genetic variance for male and female comparisons (see Visscher and Thompson, 1990). The model is the same as described above for animal models, with the relationship matrix, A , partitioned as:

$$A = T \begin{bmatrix} D_m & 0 \\ 0 & 0 \end{bmatrix} T' + T \begin{bmatrix} 0 & 0 \\ 0 & D_f \end{bmatrix} T'$$

with D_m and D_f the Mendelian sampling terms for males and females respectively. Now

$$v(\mathbf{u}) = TD_m' T \sigma_{a\sigma}^2 + TD_f' T \sigma_{a\phi}^2$$

where D_m' and D_f' are diagonal matrices with the appropriate number of zero diagonal elements for females and males respectively. A limiting case is with $\sigma_{a\sigma}^2 = \sigma_{a\phi}^2$ then this model reduces to the standard animal model. The phenotypic variance is estimated as the sum of the genetic and error variance, and the genetic variance is partitioned as the sum of three-quarters the genetic variance between females and one-quarter the genetic variance between males:

$$\sigma_p^2 = \frac{3}{4} \sigma_{a\phi}^2 + \frac{1}{4} \sigma_{a\sigma}^2 + \sigma_e^2$$

The corresponding heritabilities are:

$$h_{\phi}^2 = \text{female heritability} = \sigma_{a\phi}^2 / \sigma_p^2$$

$$h_{\sigma}^2 = \text{male heritability} = \sigma_{a\sigma}^2 / \sigma_p^2$$

The model assumes that the Mendelian sampling variance is equal to half the female genetic variance. This assumption was made because records were only on females.

AM ♀ ♂ ♀ AM with possible heterogeneity of genetic variance with a maternal half-sib effect as an extra (uncorrelated) random effect (with identity covariance matrix). For the parameterization used, the covariance between maternal half-sibs (σ_c^2) is the variance of the additional random effect plus one quarter of the female genetic variance. Thus the female genetic variance should correspond to an estimate of the variance from dd comparisons, while four times the covariance between maternal half-sibs should correspond to an estimate of genetic variance from mhs. For this model the phenotypic variance is estimated as:

$$\sigma_p^2 = \frac{1}{2} \sigma_{a\phi}^2 + \frac{1}{4} \sigma_{a\sigma}^2 + \sigma_c^2 + \sigma_e^2$$

with σ_c^2 = covariance between half-sibs, assumed to be $\frac{1}{4} \sigma_{a\text{mhs}}^2$, and

$$h_{dd}^2 = \sigma_{a\phi}^2 / \sigma_p^2 = \text{heritability from dd covariance}$$

$$h_{mhs}^2 = 4\sigma_c^2 / \sigma_p^2 = \text{heritability from mhs covariance.}$$

The parents of all sires were assumed to be unknown, hence all sires were treated as base animals. Approximate standard errors and (Fisher's) information were calculated using a quadratic approximation of the likelihood surface, as suggested by Smith and Graser (1986) and discussed by Visscher *et al.* (1991). All analyses were univariate analyses for either FY or FC.

Table 2 Parameter estimates from models with homogeneous genetic variance†

Model	Fat yield (FY)					Fat concentration (FC)				
	$\hat{\sigma}_a^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_p^2$	h^2	s.e.	$\hat{\sigma}_a^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_p^2$	h^2	s.e.
AM	342	566	908	0.377	0.021	6.8	4.0	10.8	0.632	0.019
AM100	350	561	911	0.384	0.021	7.0	3.9	10.9	0.640	0.019
AM25	368	549	917	0.402	0.022	6.9	3.9	10.8	0.639	0.019
AM01	393	530	923	0.426	0.024	7.0	3.9	10.9	0.645	0.021
SIRE	288	606	894	0.322	0.033	6.7	4.0	10.7	0.628	0.044
SIRE100	303	594	897	0.338	0.034	7.1	3.7	10.8	0.656	0.044
SIRE75	309	589	898	0.344	0.034	7.1	3.7	10.8	0.657	0.045
SIRE50	344	562	906	0.379	0.040	7.1	3.7	10.8	0.657	0.048
SIRE25	338	567	905	0.374	0.051	7.1	3.7	10.8	0.657	0.058
DAM	321	549	871	0.369	0.051	6.0	4.2	10.2	0.592	0.051
FEM	329	543	871	0.377	0.024	6.2	4.0	10.2	0.610	0.022

† All variance components in units kg^2 for FY and $(\text{g}/\text{kg})^2$ for FC.

Results

Models with homogeneous genetic variances

The main results for models for which one genetic component was estimated are shown in Table 2. Heritabilities were similar to previously reported estimates (Visscher and Thompson, 1992). Estimates of heritabilities and phenotypic variances differed slightly for models, although difference between highest and lowest estimates were substantial.

Table 3 Parameter estimates and relative likelihoods from AM and models with heterogeneous genetic variance†

Parameter estimate	Model		
	AM♀♂♀	AM♀♂‡	AM§
For trait FY			
\hat{h}_{phs}^2 (s.e.)	0.307 (0.031)	0.307 (0.031)	0.377 (0.021)
\hat{h}_{mhs}^2 (s.e.)	0.385 (0.050)	0.426 (0.028)	
\hat{h}_{dd}^2 (s.e.)	0.437 (0.030)		
$\hat{\sigma}_{\text{phs}}^2$	276	276	342
$\hat{\sigma}_{\text{mhs}}^2$	346	383	
$\hat{\sigma}_{\text{dd}}^2$	393		
$\hat{\sigma}_e^2$	547	542	566
$\hat{\sigma}_p^2$	899	899	908
ML	0.0	3.2	3.9
For trait FC			
\hat{h}_{phs}^2 (s.e.)	0.611 (0.042)	0.610 (0.042)	0.632 (0.019)
\hat{h}_{mhs}^2 (s.e.)	0.610 (0.044)	0.646 (0.026)	
\hat{h}_{dd}^2 (s.e.)	0.654 (0.027)		
$\hat{\sigma}_{\text{phs}}^2$	6.6	6.6	6.8
$\hat{\sigma}_{\text{mhs}}^2$	6.5	7.0	
$\hat{\sigma}_{\text{dd}}^2$	7.0		
$\hat{\sigma}_e^2$	3.9	3.9	4.0
$\hat{\sigma}_p^2$	10.8	10.8	10.8
ML	0.0	0.2	0.7

† Variances in units kg^2 for FY and $(\text{g}/\text{kg})^2$ for FC.

‡ No distinction between dd and mhs h^2 and variance estimates.

§ No distinction between phs, dd and mhs h^2 and variance estimates.

|| For each trait compared with the (log) maximum likelihood of AM.

Largest differences for heritability estimates between models were 0.104 for FY (AM01 – SIRE) and 0.065 for FC (SIRE75 – DAM). For both traits the heritability increased if more sires were treated as fixed, and this increase was largest for FY. For FY the estimate of the heritability increased if relatively more female comparisons were used to estimate genetic variances, while the opposite was found for FC. For example, for FY nearly all heritability estimates from sire models were lower than estimates from animal models. Fitting animal models, the estimate of the phenotypic variance slightly increased when (more) sires were treated as fixed. Compared with AM01, phenotypic variances were proportionately about 0.06 lower for both traits when sires' information was ignored (models DAM and FEM), and this reduction in variance was due to a reduction in the genetic variance. Standard errors increased for models which used less information, and were largest when only sires with up to 24 progeny contributed to the parameter estimates (model SIRE25). For FY, the dam model (DAM) gave a similar standard error as SIRE25.

Models with heterogeneous genetic variances

In Table 3 the results from AM♀♂♂ and AM♀♂♀ are presented. For the model with two genetic variance estimates (AM♀♂♂), male and female heritability estimates for FY were similar to estimates from the standard sire model (SIRE) and the animal model with all sires being fixed (AM01). For FC the corresponding heritability estimates from homogeneous genetic variance models was less clear, although the lower estimate (0.610) was similar to that for models using only female comparisons (FEM and DAM). The difference between male and female heritability estimates, 0.119 for FY (and 0.036 for FC), was similar to that previously reported (Visscher and Thompson, 1990). Estimates of the correlation between male and female heritability estimates, obtained from the information matrix, were -0.15 for FY and -0.36 for FC. Differences in the logarithm of the maximum likelihood (ML) for AM and AM♀♂♂ indicate significant heterogeneous genetic variances for FY (twice difference in ML, which is approximately distributed as a χ^2 with 1 d.f., is 6.4), but no difference in male and female genetic variances for FC.

Heritability estimates from daughter-dam regression and maternal half-sib covariance (from model AM♀♂♀) were similar for each of the two traits, although the heritability estimate from dd comparisons was consistently higher for both traits. Similar heritabilities indicate that female genetic variances were homogeneous, and this is confirmed by the outcome of a likelihood ratio test on (twice) the difference of the ML for AM♀♂♂ and AM♀♂♀.

As expected, the estimate of the phs heritability was nearly identical to the estimate from AM♀♂.

Discussion

Heterogeneity of genetic variance was demonstrated, in particular for FY. The explanation for this observation is not straightforward. The sires present in the pedigree may be a selected group for the traits under consideration, but unfortunately there was insufficient information about the country of origin and the year of birth of the sires to investigate this. If sires represented in the data were related, this was ignored; this may bias the genetic variance downward. Since sires in the pedigree were a mixture of proven and unproven bulls, and were from European and North American origin, it is not obvious that the genetic variance among them should be reduced. However, analyses with sire models (SIRE, SIRE25, SIRE50, SIRE75 and SIRE100) suggest that the variance between sires with large numbers of progeny, presumably the proven sires, was less than the variance between sires with few progeny. By treating the proven sires as fixed, some selection effect seems to be taken into account. The increase in phenotypic variance for models where (more) sires were treated as fixed, for example for AM01, may be explained using the results from model AM♀♂: for AM01, the phenotypic variance is approximately the sum of the female genetic variance ($\approx 383 \text{ kg}^2$ for FY, from AM♀♂ in Table 3) and the environmental variance ($\approx 542 \text{ kg}^2$, from Table 3). Therefore, although the analyses for AM01 was in effect within sires, the estimate of the phenotypic variance was increased. Van der Werf (1992) shows that by treating (selected) base animals as fixed does not necessarily lead to unbiased variance components estimates.

Ignoring sires (models FEM and DAM) led to a decrease in the phenotypic variance (see Table 2). If we approximate the pedigree structure by a (balanced) hierarchical design with dams within sires and observations only on progeny and sires selected, then performing an ANOVA ignoring sires gives a similar result. For a given number of sires the largest decrease corresponds to the most intense selection of sires. The extreme case giving the largest reduction in phenotypic variance is if all dams were mated to one sire, since then the estimate of the phenotypic variance would be three-quarters of the female genetic variance plus the environmental variance. Hence models FEM and DAM are not as satisfactory as models including all relationships.

Combining the male and female heritability estimates to find the minimum variance estimate, resulted in combined estimates for each trait which

were close to the estimated heritabilities from the standard animal model. The weights for the male and female heritability were 1263 and 1450 for FY, and 1027 and 2085 for FC. Therefore, the relative information from males and females was similar for FY, while for FC females provided more information (because of the higher heritability).

Similarly by converting the standard errors in Table 3 on female comparisons to measures of information (1/variance) then this suggests that the weights given to dd and mhs comparisons were 1012 and 262 for FY and 1228 and 250 for FC, showing that dam-daughter comparisons were more valuable in providing information on heritability.

There was little evidence of heterogeneous female genetic variances, although for both traits the daughter-dam heritability was slightly higher than the heritability from maternal half-sibs. Differences in ML for AM♀♂ and AM♀♂♀ indicated no difference between genetic variances from dd and mhs comparisons. If a genotype by herd interaction exists, this would still give homogeneous female variances (provided dam families have records in the same herds), but may explain the observed differences in male and female genetic variances. A cytoplasmic effect (e.g. Bell, McDaniel, and Robison, 1985) would give a larger genetic variance from mhs compared with dd comparisons, and would result in the genetic variance calculated from phs to be smaller than either estimates from mhs and dd. Although the latter was observed, for both traits the dd variance was consistently higher than the mhs variance. Epistatic genetic effects, for example additive by additive gene effects, could explain a difference between dd and phs genetic variance, but the effects have to be extremely large to explain the observed differences (the expected difference between phs and dd is only one-quarter of the additive by additive genetic variance). Furthermore, epistatic effects would give the same variances for phs and mhs, which was not observed.

The modelling used to estimate male and female genetic variances by allowing different Mendelian sampling terms for males and females (Visscher and Thompson, 1990) is general, and can, for example, be used to estimate heterogeneous genetic variances from selection experiment data (Beniwal, Thompson, and Hill, 1991). However, when using the male and female grouping as presented in this study over multiple generations, the genetic model and the parameterization of the phenotypic variance become less clear. Even if the genetic variance among males is consistently lower in all generations, allowing for different Mendelian sampling terms would not account for this properly.

Acknowledgement

We thank W. G. Hill for making us formalize our intuitive (handwaving) arguments. The Milk Marketing Boards in the UK are thanked for providing data and for financial support for PV.

References

- Bell, B. R., McDaniel, B. T. and Robison, O. W. 1985. Effect of cytoplasmic inheritance on production traits of dairy cattle. *Journal of Dairy Science* **68**: 2038-2051.
- Beniwal, B. K., Thompson, R. and Hill, W. G. 1991. Estimates of genetic parameters for lean mass in long-term selection with mice. *Animal Production* **52**: 562 (abstr.).
- Ducrocq, V., Boichard, D., Bonaiti, B., Barbat, A. and Briend, M. 1990. A pseudo-absorption strategy for solving animal model equations for large data files. *Journal of Dairy Science* **73**: 1945-1955.
- Graser, H-U, Smith, S. P. and Tier, B. 1987. A derivative-free approach for estimating variance components in animal models by Restricted Maximum Likelihood. *Journal of Animal Science* **64**: 1362-1370.
- Hill, W. G., Edwards, M. R., Ahmed, M.-K. A. and Thompson, R. 1983. Heritability of milk yield and composition at different levels and variability of production. *Animal Production* **36**: 59-68.
- Jones, L. P. and Goddard, M. E. 1990. Five years experience with the animal model for dairy cattle evaluations in Australia. *Proceedings of the fourth world congress on genetics applied to livestock production, Edinburgh, Vol. XIII*, pp. 382-385.
- Maijala, K. and Hanna, M. 1974. Reliable phenotypic and genetic parameters in dairy cattle. *Proceedings of the first world congress on genetics applied to livestock production, Madrid, Vol. I*, pp. 541-563.
- Meyer, K. 1983. Scope for evaluating dairy sires using first and second lactation records. *Livestock Production Science* **10**: 531-553.
- Patterson, H. D. and Thompson, R. 1971. Recovery of interblock-information when block sizes are unequal. *Biometrika* **58**: 545-554.
- Smith, S. P. and Graser, H.-U. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *Journal of Dairy Science* **69**: 1156-1165.
- Swalve, H. and Van Vleck, L. D. 1987. Estimation of genetic (co)variances for milk yield in first three lactations using an animal model and restricted maximum likelihood. *Journal of Dairy Science* **70**: 842-849.
- Thompson, R. 1977. The estimation of heritability with unbalanced data. II. Data available on more than two generations. *Biometrics* **33**: 497-504.
- Van der Werf, J. H. J. 1992. Restricted maximum likelihood estimation of additive genetic variance when selected base animals are considered fixed. *Journal of Animal Science* **70**: 1068-1076.
- Van Vleck, L. D. 1985. Including records of daughters of selected bulls in estimation of sire components of variance. *Journal of Dairy Science* **68**: 2396-2407.
- Van Vleck, L. D. 1986. Evaluation of dairy cattle breeding programs: specialised milk production. *Proceedings of the third world congress on genetics applied to livestock production, Lincoln, Nebraska, Vol. IX*, pp. 141-152.
- Van Vleck, L. D. and Dong, M. C. 1988. Genetic (co)variances for milk, fat, and protein yield in Holsteins using an animal model. *Journal of Dairy Science* **71**: 3040-3046.
- Visscher, P. M. and Thompson, R. 1990. REML estimates of parameters for fat yield in pedigree herds in the UK using an individual animal model; male and female heritability estimates. *Proceedings of the fourth world congress on genetics applied to livestock production, Edinburgh, Vol. XIV*, pp. 233-236.
- Visscher, P. M. and Thompson, R. 1992. Univariate and multivariate parameter estimates for milk production traits in lactations 1-3 using an animal model. I. Description of analyses and presentation of REML estimates. *Genetics, Selection and Evolution* In press.
- Visscher, P. M., Thompson, R. and Hill, W. G. 1991. Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livestock Production Science* **28**: 273-290.
- Wiggans, G. R., Misztal, I. and Van Vleck, L. D. 1988. Implementation of an Animal Model for genetic evaluation of dairy cattle in the United States. *Journal of Dairy Science* **71**: suppl. 2, pp. 54-69.

(Received 11 November 1991—Accepted 1 June 1992)