

## Evaluation of Dormer sires for litter size and lamb mortality using a threshold model

K.V. Konstantinov\*, G.J. Erasmus and J.B. van Wyk

Department of Animal Science, University of the Orange Free State, P.O. Box 339, Bloemfontein, 9300 Republic of South Africa

Received 22 March 1993; revised 16 August 1994; accepted 14 September 1994

Heritability estimates and breeding value predictions of Dormer sires for litter size and lamb mortality were obtained using a threshold model. Heritability estimates on the underlying scale were 0.24 and 0.12 for litter size and mortality, respectively. Breeding value predictions varied between  $0.41\sigma_c^2$  and  $-0.43\sigma_c^2$  for litter size and  $0.30\sigma_c^2$  and  $-0.14\sigma_c^2$  for lamb mortality. Probability distributions for progeny of sires across 'fixed effects' were also calculated. It is concluded that a threshold model can effectively be applied for routine evaluation of Dormer sires for categorical traits.

Oorerflikheidsberamings en teeltwaardevoorspellings van Dormer vaders vir werpselgrootte en lam mortaliteit is verkry deur die gebruik van 'n drumpelwaarde-model. Oorerflikheidsberamings op die onderliggende skaal was 0.24 en 0.12 vir fekunditeit en mortaliteit, onderskeidelik. Teeltwaardevoorspellings het gewissel tussen  $0.41\sigma_c^2$  en  $-0.43\sigma_c^2$  vir fekunditeit en  $0.30\sigma_c^2$  en  $-0.14\sigma_c^2$  vir mortaliteit. Waarskynlikheidsverdelings vir nageslag van vaars oor 'vaste effekte' is ook beraam. Dit blyk 'n drumpelwaarde-model aangewend kan word vir roetine evaluasie van vaars vir kategorieë kenmerke.

**Keywords:** Categorical traits, Dormer sheep, sire evaluation, threshold model.

\* To whom correspondence should be addressed.

### Introduction

The estimation of genetic parameters and predictions of breeding values of the candidates for selection from observations of relatives play an important role in animal breeding. If the data and the genetic values to be predicted follow joint normal distributions, Best Linear Unbiased Prediction (BLUP) yields the maximum likelihood estimator of the best predictor and maximizes the probability of the correct ranking of individuals (Henderson, 1973). However, many traits of economic importance in animal breeding (e.g. calving ease, disease resistance, mortality, type scores, pelt traits in Karakul sheep, etc.) are measured as a response in a small number of mutually exclusive and exhaustive categories. These variables are not normally distributed and, in this case, Portnoy (1982) showed that linear predictors may behave poorly for ranking purposes.

Nel (1967) used intra-sire regression of progeny on dam and Henderson's method 3 for heritability estimates of pelt traits in Karakul sheep and suggested a selection index for genetic evaluation of animals. Greeff *et al.* (1991) also used Henderson's method 3 to estimate sire variances for pelt traits, arguing that least-squares analysis is valid if the values scored are an indication of quantitative differences among classes.

Schaeffer & Wilton (1976) used a linear model, developed by Grizzle *et al.* (1969), for the analysis of categorical data. They suggested that the use of BLUP methodology in sire evaluation for categorical responses might be justified, given certain conditions, which unfortunately are inconsistent with the assumptions required by their model. In spite of its statistical drawbacks, this work gave an impetus for the widespread use of BLUP in the evaluation of sires for categorical traits (Berger & Freeman, 1978; Van Vleck & Karner, 1979; Cady & Burnside, 1982; Westell *et al.*, 1982).

In recent years efforts have been made to develop procedures of analysis specifically in the area of prediction of breeding values for categorical traits (Gianola & Foulley, 1983a; 1983b; Harville & Mee, 1984; Gilmour *et al.*, 1985; Foulley & Gianola,

1984; Foulley *et al.*, 1983; Hoeschele *et al.*, 1986). The main theoretical reason for not using BLUP with categorical data is that breeding values and residuals are not independent of each other and their marginal distributions are difficult to specify (Gianola, 1982). A single heritability does not exist on the discontinuous scale. To account for this heterogeneity, BLUP would require a large number of parameters to be estimated, probably very imprecisely. Heterogeneity of heritability is also affected by the number of effects in the model and by differences between levels (Hoeschele, 1988).

A general approach to the prediction of genetic merit from categorical data has been proposed by Gianola & Foulley (1983a; 1983b). This method is based on the threshold concept (Wright, 1934; Dempster & Lerner, 1950; Bulmer, 1980; Gianola, 1982) and employs a Bayesian procedure for statistical inference which allows a large range of data structures and models to be treated. This method can be regarded as an extension of BLUP to a type of non-linear problem.

The purpose of this study was to estimate heritabilities for litter size and lamb mortality and to utilize these estimates in the genetic evaluation of Dormer sires from the Elsenburg stud, employing a threshold model.

### Material and Methods

The data used in this study consisted of 5 214 records of the progeny of 112 sires for litter size and 4 125 records of the progeny of 96 sires for lamb mortality, respectively. A detailed description of the data structure is given by Van Wyk *et al.* (1993). The traits analysed were prolificacy (number of lambs born; 0, 1, 2 or 3) and lamb mortality (born live or born dead). The threshold model was assumed to be the same as that in Gianola & Foulley (1983a). Under this model litter size and lamb mortality occur as a result of an underlying unobserved phenotype exceeding a given threshold (three thresholds for litter size and one for mortality). The unobserved continuous phenotypes are assumed to be normally distributed. For each trait a vector,  $\mu$ ,

of means corresponding to subpopulations determined by combinations of levels of 'fixed'  $\beta$  and random  $s$  factors, is modelled as:

$$\mu = X\beta + Zs$$

where

$\mu$  is a vector of underlying means,

$\beta$  is a vector associated with the effects of year-season (48) and age of dam (7) for litter size and year-season (18), age of dam (7) and birth status (3) for mortality,

$s$  is a vector of sire effects and

$X$  and  $Z$  are design matrices.

The  $s$  effects are assumed to be normally distributed with  $E(s) = 0$  and  $\text{Var}(s) = A\sigma_s^2$ , where  $A$  is a numerator relationship matrix. The two traits were analysed separately. The solutions for  $t$  (thresholds),  $\beta$  and  $s$  can be computed by iterating from round ( $k$ ) to ( $k + 1$ ) on the following system of equations (Gianola & Foulley, 1983a):

$$\begin{bmatrix} L^{(k)'} M^{(k)} L^{(k)} + H^{(k)} & L^{(k)'} M^{(k)} X & L^{(k)'} M^{(k)} Z \\ X' M^{(k)} L^{(k)} & X' M^{(k)} X & X' M^{(k)} Z \\ Z' M^{(k)} L^{(k)} & Z' M^{(k)} X & Z' M^{(k)} Z + G^{-1} \end{bmatrix} \times \begin{bmatrix} t \\ \beta \\ s \end{bmatrix}^{(k+1)} = \begin{bmatrix} L^{(k)'} M^{(k)} y^{(k)} + q^{(k)} \\ X' M^{(k)} y^{(k)} \\ Z' M^{(k)} y^{(k)} \end{bmatrix} \quad (1)$$

where  $G^{-1} = A^{-1}\sigma_s^2$  and matrices  $M^{(k)}$ ,  $L^{(k)}$ ,  $H^{(k)}$ , and the vectors  $y^{(k)}$  and  $q^{(k)}$  are defined in Misztal *et al.* (1989).

Since this procedure requires the variance of  $s$  to be known, and no literature estimates were available for the traits under consideration, it was decided to estimate the sire variance from the data by a **REML-type** procedure, proposed by Harville & Mee (1984):

$$\sigma_s^{2(k+1)} = (s'A^{-1}s + \text{tr} A^{-1}C^{-1})^{(k)}/q$$

where  $C$  is a part of the coefficient matrix in (1), corresponding to the  $s$  effects, and  $q$  is the number of levels in  $s$ .

For each sire the probability distributions for progeny across 'fixed' effects were calculated, using the formula presented by Gianola & Foulley (1983a).

### Numerical aspects

Computing solutions to threshold models poses several problems not encountered in linear model analyses. Many of these problems and computing strategies were discussed by Misztal *et al.* (1989). The computing strategy used in this study was as follows:

First the data were arranged in contingency tables. The system (1) was created in memory, using the formulae presented by Gianola & Foulley (1983a) and Misztal *et al.* (1989). The first threshold was set to zero, as well as the last level of each of the  $m - 1$  'fixed' effects. The inverse of the numerator relationship matrix was computed using the method of Quaas (1976). The parents were absorbed into the sire part of the matrix and only this part was written onto the disk. The system (1) was solved by obtaining a generalized inverse of the coefficient matrix and multiplying it by the right-hand side vector. In theory the complete analysis requires two levels of nested iterations, one for variance component estimation and the other for estimation of the effects

in the model. Convergence criteria used for both iterations were those suggested by Misztal *et al.* (1989). Since the formulae for certain matrices and vectors in (1) involve the probability

$$P_{jk} = \Phi(t_k - \mu_j) - \Phi(t_{k-1} - \mu_j)$$

in denominators, where  $\mu_j = x_j\beta + z_j s$  is a location parameter for observations  $j$ , and  $\Phi(\cdot)$  is a standard normal distribution function, it was calculated using the **s15abf** function from the **NAG** library. In order to increase the speed for constructing the system (1), the probabilities were first calculated for a wide range of  $t_k - \mu_j$  and stored in a large vector. The values of  $P_{jk}$  were not allowed to drop below a certain value ( $P_{min} = 10^{-10}$ ) in order to avoid division by zero, as suggested by Misztal *et al.* (1989).

### Results and Discussion

The total number of rounds required to reach convergence of  $10^{-5}$  for Newton-Raphson iterations was 10 and 12 for litter size and mortality, respectively. After five rounds of Newton-Raphson iteration, two rounds of **REML-type** iterations were performed. Convergence was reached after 12 and 14 rounds for litter size and mortality, respectively.

Solutions for thresholds, age of dam and birth status are presented in Table 1. They are expressed in units of residual standard deviations of the underlying variable.

The results showed that prolificacy increased with age, although ages 6 and 7 seemed to escape from this pattern. There was no trend in age effect for lamb mortality. It should be pointed out that more data per age group are needed to assess the effect of dam's age for lamb mortality.

For this trait birth status was added as an additional 'fixed' factor. It is interesting to note that the elimination of birth status from the model caused an increase in the prediction error variance for sires from 0.2114 to 0.2169. A slight change in sire ranking was observed as was the case when birth status was excluded from the model. Hence, it was decided to include birth status in the model for analysis and the results are shown in

**Table 1** Solutions for age of dam and birth status

	Litter size	Mortality
Thresholds		
1	0.0000000	0.0000000
2	0.8221543	-
3	2.7843010	-
Age of dam		
2	1.2802117	0.0779802
3	1.5464498	0.1529798
4	1.7578906	0.0441869
5	1.7583555	-0.0585607
6	1.6700321	-0.0696860
7	1.6418813	0.0702199
8	1.9350504	0.0000000
Birth status		
1	-	-1.4933747
2	-	-1.5253850
3	-	-0.9141237

Table 1. As one would expect, there were more deaths with triplets than with twins and single-born animals. All estimates were, however, negative signs, which is probably due to the data structure used for analysis.

Table 2 contains the estimated sire variances and heritabilities for the traits under consideration. Heritability estimates for litter size and lamb mortality were 0.24 and 0.12, respectively. No references in which a threshold model was applied to estimate heritabilities for these traits have been found. The results from this study suggest that improvement by means of selection for the traits analysed is possible.

**Table 2** Estimates of sire variances and heritabilities on underlying scale

Trait	Sire variance	Heritability
Litter size	0.064576	0.243
Lamb mortality	0.029928	0.116

Table 3 shows features of the sire evaluation for litter size and lamb mortality. In addition to the breeding values, probability distributions for progeny of sires across 'fixed' effects were calculated. The best sire had a breeding value estimate of  $0.41\sigma^2_e$  for litter size. The corresponding value for mortality was  $-0.15\sigma^2_e$ .

**Table 3** Features of the sire genetic evaluation for litter size and lamb mortality

	Litter size		Mortality	
	Best sire	Worst sire	Best sire	Worst sire
No. of progeny	74	35	72	30
Sire solution	0.4097	-0.4289	-0.1465	0.3004
Probability	14.06	22.94	62.75	61.37

Extreme evaluations for sires are not consistent with extreme observed percentages per sire. For example, the best evaluated sire for litter size had progeny with 6.4, 16.1, 63.4 and 14.1% in categories 1, 2, 3 and 4. The corresponding observed percentages for this sire were 31.0, 9.7, 7.0 and 52.3%, respectively. Crude percentages are not good indicators of litter size and lamb mortality and they should not be used for selection purposes.

This study shows that the threshold model can effectively be applied for routine evaluation of sires for categorical traits in Dormer sheep. Although the current version of the programme uses matrix inversion, it is not difficult to make the necessary modifications and employ some of the iterative procedures available to obtain solutions for the effects in the model. To estimate large data sets with several variance components along with the breeding values, the 'tilde-hat' approach of Van Raden & Young (1988) could be used, as shown by Manfredi *et al.* (1991).

**Acknowledgements**

The authors are grateful to an anonymous reviewer for his helpful comments and criticism.

**References**

BERGER, P.J. & FREEMAN, A.E., 1978. Prediction of sire merit for calving difficulty. *J. Dairy Sci.* 61, 1146.

BULMER, M.G., 1980. The mathematical theory of quantitative genetics. Clarendon Press, Oxford.

CADY, R.A. & BURNSIDE, E.B., 1982. Evaluation of dairy bulls in Ontario for calving ease of offspring. *J. Dairy Sci.* 65, 2150.

DEMPSTER, E.R. & LERNER, I.M., 1950. Heritability of threshold characters. *Genetics* 35, 212.

FOULLEY, J.L. & GIANOLA, D., 1984. Estimation of genetic merit from bivariate 'all-or-non' responses. *Genet. Sel. Evol.* 16, 285.

FOULLEY, J.L., GIANOLA, D. & THOMPSON, R., 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty. *Genet. Sel. Evol.* 15, 401.

GIANOLA, D., 1982. Theory and analysis of threshold characters. *J. Anim. Sci.* 54, 1079.

GIANOLA, D. & FOULLEY, J.L., 1983a. Sire evaluation for ordered data with a threshold model. *Genet. Sel. Evol.* 15, 201.

GIANOLA, D. & FOULLEY, J.L., 1983b. New techniques of prediction of breeding values for discontinuous traits. 32nd Ann. Breed. Roundtable, St. Louis, Missouri.

GILMOUR, A.R., ANDERSON, R.D. & RAE, A.C., 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 593.

GREEFF, J.C., FAURE, A.S., MINNAAR, G.J. & SCHOEMAN, S.J., 1991. Genetic and phenotypic parameters of pelt traits in a Karakul control flock. *S. Afr. J. Anim. Sci.* 21, 156.

GRIZZLE, D.A., STARMER, C.F. & KOCH, G.C., 1969. Analysis of categorical data by linear models. *Biometrics* 25, 489.

HARVILLE, D.A. & MEE, R.W., 1984. A mixed model procedure for analysing ordered categorical data. *Biometrics* 40, 393.

HENDERSON, C.R., 1973. Sire evaluation and genetic trends. Proc. of Anim. Breed. Genet. Symp. in honour of Dr J.L. Lush, Amer. Soc. of Anim. Sci. and Amer. Dairy sci. Assoc., Champaign, IL.

HOESCHELE, INA, FOULLEY, J.L., COLLEAU, J.J. & GIANOLA, D., 1986. Genetic evaluation for multiple binary responses. *Genet. Sel. Evol.* 18, 299.

HOESCHELE, INA., 1988. Comparison of 'maximum a-posteriori estimation' and 'quasi best linear unbiased prediction' with threshold characters. *J. Anim. Breed. Genet.* 105, 337.

MANFREDI, E., DUCROCQ, V. & FOULLEY, J.L., 1991. Genetic analysis of distocia in dairy cattle. *J. Dairy Sci.* 74, 1715.

MISZTAL, I., GIANOLA, D. & FOULLEY, J.L., 1989. Computing aspects of non-linear method of sire evaluation for categorical data. *J. Dairy Sci.* 72, 1557.

NEL, J.A., 1967. Genetic studies in Karakul sheep. Annale Universiteit van Stellenbosch.

PORTNOY, S., 1982. Maximizing the probability of correctly ordering random variables using linear predictors. *J. Mult. Anal.* 12, 256.

SCHAEFFER, L.R. & WILTON, J.W., 1976. Methods of sire evaluation for calving ease. *J. Dairy Sci.* 59, 544.

VAN RADEN, P.M. & YOUNG, Y.C., 1988. A general purpose approximation to restricted maximum likelihood: the tilde-hat approach. *J. Dairy Sci.* 71, 187.

VAN VLECK, L.D. & KARNER, D.J., 1980. Sire evaluation by best linear unbiased prediction for categorical scored type traits. *J. Dairy Sci.* 63, 1328.

VAN WYK, J.B., ERASMUS, G.J. & KONSTANTINOV, K.V., 1993. Non-genetic factors influencing growth traits in the Elsenburg Dormer sheep stud. *S. Afr. J. Anim. Sci.* 23, 67.

WESTELL, R.A., BURNSIDE, E.B. & SCHAEFFER, L.R., 1982. Evaluation of Canadian Holstein Friesian sires on disposal reasons of their daughters. *J. Dairy Sci.* 65, 23.

WRIGHT, S., 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19, 506.