

## Prediction of response to selection within families

WG Hill<sup>1</sup>, A Caballero<sup>1</sup>, L Dempfle<sup>2</sup>

<sup>1</sup> *Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JT, UK;*

<sup>2</sup> *International Trypanotolerance Centre, PMB 14, Banjul, the Gambia*

(Received 19 January 1996; accepted 24 May 1996)

**Summary** – The distinction is clarified between two different uses of the term ‘within-family selection’, either to imply that one individual of each sex is selected from each family or to imply that individuals are selected on their deviation from family mean, in which case families may not be equally represented. In the short term, selection on within-family deviation is expected to give higher responses, but in the long term, selection within families is expected to give higher responses as the effective population size is larger. The two schemes are the same, however, if only two individuals of each sex are recorded in each family.

**artificial selection / selection index / accuracy of selection / response / effective population size**

**Résumé** – **Prédiction de la réponse à la sélection intrafamiliale.** *Une distinction est faite entre deux emplois du terme « sélection intrafamiliale », qui implique tantôt une sélection d'un nombre égal d'individus de chaque sexe dans chaque famille, tantôt une sélection des individus sur leur écart à la moyenne de famille, auquel cas toutes les familles ne sont pas nécessairement également représentées. À court terme, l'espérance de la réponse à une sélection sur l'écart intrafamiliale est plus grande ; mais à long terme, l'espérance de la réponse à une sélection intrafamiliale stricte est supérieure, car une telle sélection accroît l'effectif génétique. Les deux schémas de sélection sont cependant identiques si seulement deux individus de chaque sexe sont contrôlés dans chaque famille.*

**sélection artificielle / indice de sélection / précision de la sélection / réponse / effectif génétique**

Only one-half of the additive genetic variance is expressed within families, so selection within families is not usually predicted to lead to higher rates of response than other schemes in which variance between families is also utilised (Lush, 1947). It can, however, be efficient for short-term selection if there is a very high

environmental correlation of sibs, and for long-term selection because the effective population size is double that for random sampling among families, and may be many times larger if selection leads to very unequal family representation. Further, under the infinitesimal model, selection leads to a reduction of variance between but not within families (Bulmer, 1971), so the relative efficiency of selection within versus across families is higher than if the 'Bulmer effect' is ignored. Selection within full-sib families is often practised in selection experiments for traits such as juvenile body weight in mice, where the environmental correlation is high, long-term response is required, and a straightforward management programme is needed (eg, Falconer, 1973). Selection within half-sib families may be practised in livestock selection programmes and experiments so as to maintain a high effective population size and increase selection limits (Dempfle, 1975). There is, however, some confusion in the literature as to predicted rates of response to selection within families, which this note is intended to clarify.

There are actually two alternative selection schemes which can be considered (Hill, 1985; Dempfle, 1990; Toro and Pérez-Enciso, 1990), and these are illustrated for the simple case where there are pair matings and therefore only full-sib families. The first is selection within families (SWF), in which the best (ie, highest scoring on whatever trait or index of traits is used) male and the best female are selected on the basis only of their own performance ( $X$ ) from each family. The second is where individuals are selected on deviation from family mean (SDM), ie on  $X - X_F$ , where  $X_F$  is family mean, with family mean computed either for each sex separately or pooled, after correction for a sex effect if necessary. For SDM, it is unlikely that all families will contribute one male and one female to the next generation. This case is being considered both because it is an index of individual and family mean performance,  $b_1X + b_2X_F$  in which  $b_1 = 1$  and  $b_2 = -1$ , so that calculations of predicted response can be computed and compared directly with other such indices, and because, in his classic text, Falconer (most recently in Falconer and Mackay, 1996) does not distinguish clearly between the two schemes, and uses formulae to describe SWF which actually relate to SDM. Correct results for both cases are given by Dempfle (1990) and applied by Toro and Pérez-Enciso (1990), but seem not to have been generally noticed.

Following Falconer's notation as far as possible, let  $n$  = family size (usually in these examples  $n$  is the number of individuals recorded of each sex, so  $X_F$  refers to the mean of one sex),  $r$  = the relationship of family members ( $r = 1/2$  for full-sib families),  $h^2$  = heritability,  $t$  = intra-class correlation of full-sibs,  $\sigma_P^2$  = phenotypic variance (both  $h^2$  and  $\sigma_P^2$  refer to the population before selection, ie, before the 'Bulmer effect' applies). The additive genetic variance within families is  $(1-r)h^2\sigma_P^2$  and the total variance within families is  $(1-t)\sigma_P^2$ . As selection on SDM applies over the total population, let the corresponding selection intensity be  $i_T$ , and as SWF applies separately within each family of size  $n$ , let the corresponding selection intensity be  $i_n$ . For example, if  $n = 4$  in all families and there are very many families, 25% of individuals of each sex are selected giving  $i_T = 1.271$  and  $i_n = 1.029$  (from tables A and B, respectively, in the appendix of Falconer and Mackay, 1996). The predicted selection response is  $R = icov(I, A)/\sigma_I$ , where the terms are the selection intensity, the covariance of the selection criterion and breeding value, and the standard deviation of the selection criterion, respectively.

**SDM**

This relates to the population as a whole, so

$$\begin{aligned}\sigma_I^2 &= \text{var}(X - X_F) = (1 - t)(1 - 1/n)\sigma_P^2 \\ \text{cov}(I, A) &= \text{cov}(X - X_F, A) = h^2(1 - r)(1 - 1/n)\sigma_P^2 \\ R(\text{SDM}) &= i_T h^2 \sigma_P (1 - r) [(1 - 1/n)/(1 - t)]^{1/2}\end{aligned}$$

which agrees with the response given by Falconer and Mackay (1966, table 13.4) for 'within-family selection'.

**SWF**

As selection is practised independently within each family, each can be regarded as a subpopulation with phenotypes and breeding values distributed about true family means, say  $\mu_F$  of phenotypes and  $A_F$  of breeding values ( $A_F$  is the mean breeding value of the parents, or of the sire for a half-sib family;  $\mu_F$  includes common environmental and dominance effects). Hence

$$\begin{aligned}\sigma_I^2 &= \text{var}(X - \mu_F) = (1 - t)\sigma_P^2 \\ \text{cov}(I, A) &= \text{cov}(X - A_F, A) = h^2(1 - r)\sigma_P^2 \\ R(\text{SWF}) &= i_n h^2 \sigma_P (1 - r)/(1 - t)^{1/2}\end{aligned}$$

as given by Dempfle (1975) for selection within families. For both SDM and SWF, the regression of breeding value on selection criterion, termed 'heritability of within-family deviations' by Falconer and Mackay (1996, eq 13.5), is given by:

$$\text{cov}(I, A)/\sigma_I^2 = h^2(1 - r)/(1 - t)$$

**COMPARISONS**

If two individuals of each sex are recorded in each family and SDM is practised as deviations from sex mean, the schemes are identical. This is because in SDM only one member of each family has a positive deviation and is selected; in terms of the formulae,  $R(\text{SDM})/R(\text{SWF}) = [i_T \sqrt{(1 - 1/n)}] / i_n = [0.798 \sqrt{(1/2)}] / 0.564 = 1$ . If family sizes are larger, higher responses are predicted for SDM because no constraints are made on selecting the best individuals. Examples are given in table I, assuming the population size is large (so infinite population values can be used for  $i_T$ ). The maximum relative difference is seen to be for families of about ten of each sex. Alternatively, if SDM is practised after correction for sex, using a common mean, its relative efficiency rises further. Thus, with two males and two females recorded in each family,  $i_n = 0.564$  and  $i_T \sqrt{(1 - 1/4)} = 0.691$ . The selection intensity for SDM has been computed assuming that the total population

size is very large. If not,  $i_T$  has to be calculated accordingly, taking account of the negative correlation,  $-1/(n-1)$ , of values of  $X - X_F$  of family members. Rawlings (1976) and Hill (1976, used by Toro and Pérez-Enciso, 1990) give formulae to correct selection intensity for correlation of family members which, although intended to allow for positive correlations as with mass selection, have been found by simulation (not shown) to give good predictions for the present case of negative correlations of deviations, that of Rawlings being simpler and fitting somewhat better. For  $m$  families, Rawlings' formula becomes  $i_{Tm} = i_T[1 + 1/(nm-1)]^{1/2}$ . For example, with  $m = 5, 10$  and  $20$  families each of size  $n = 4$ ,  $i_T = 1.214, 1.242$  and  $1.257$ , respectively, so it predicts  $i_{Tm} = 1.246, 1.258$  and  $1.265$ , respectively; and as  $m \rightarrow \infty$ ,  $i_T = i_{Tm} = 1.271$ .

If families are not of equal size, variation in size has a different impact on the two alternatives (and it may not be possible to apply SWF strictly if any families fail to rear at least one of each sex). For example, assume that on average four progeny are reared per family, but these are in relative frequencies 10% for 2 and 6, 20% for 3 and 5, and 40% for 4. Then, on average,  $i_T \sqrt{(1-1/n)} = 1.081$  and  $i_n = 0.996$ , so  $R(\text{SDM})/R(\text{SWF}) = 1.085$ , a little higher than given in table I. Use of SWF, however, removes any need to correct for differences in variation between families in within-family environmental variance, which would lead to loss of efficiency for SDM.

**Table I.** Predictions<sup>a</sup> for different family sizes ( $n$ ) of relative responses ( $R$ ) in the first generation and relative effective population sizes ( $N_e$ ) for selection within families (SWF) and selection on within-family deviation (SDM).

	n						
	3	4	5	8	10	20	50
$R(\text{SDM})/R(\text{SWF})$	1.053	1.070	1.077	1.082	1.082	1.077	1.066
$N_e(\text{SDM})/N_e(\text{SWF})^b$	0.78	0.73	0.69	0.64	0.62	0.57	0.54

<sup>a</sup> Assuming a large number of families and the same number recorded in each family. <sup>b</sup> By simulation,  $SE < 0.01$ .

The effective population size ( $N_e$ ) for SWF equals  $2m - 1$ , ie, almost twice the number of parents (Falconer and Mackay, 1996, p 69). Depending on the extent of unequal family representation,  $N_e$  for SDM is smaller, but in contrast to other selection schemes (Caballero, 1994),  $N_e$  is independent of  $h^2$  and  $t$  because these parameters do not affect the distribution of numbers of selected individuals per family and there is no correlation of family size over generations (assuming variance in family size is not inherited). Using simulation, relative values of  $N_e$  are given for the two schemes in table I assuming the same number are recorded in each family; if this varies,  $N_e$  will be reduced further for SDM. As the effective population size is smaller for SDM than SWF, long-term responses would be less; as in other situations, there is a conflict between short- and long-term responses. Toro and Pérez-Enciso (1990) discuss alternative structures further.

**REFERENCES**

- Bulmer MG (1971) The effect of selection on genetic variability. *Am Nat* 105, 201-211
- Caballero A (1994) Developments in the prediction of effective population size. *Heredity* 73, 657-679
- Dempfle L (1975) A note on increasing the limit of selection through selection within families. *Genet Res* 24, 127-135
- Dempfle L (1990) Statistical aspects of design of animal breeding programs: a comparison among various selection strategies. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, Berlin, 98-117
- Falconer DS (1973) Replicated selection for body weight in mice. *Genet Res* 22, 291-321
- Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. 4th ed, Longman, Harlow
- Hill WG (1985) Fixation probabilities of mutant genes with artificial selection. *Genet Sel Evol* 17, 351-358
- Hill WG (1976) Order statistics of correlated variables and implication in genetic selection programmes. *Biometrics* 32, 889-902
- Lush JL (1947) Family merit and individual merit as bases for selection. *Am Nat* 81, 241-261, 293-301
- Rawlings JO (1976) Order statistics for a special class of unequally correlated multinormal variates. *Biometrics* 32, 875-887
- Toro M, Pérez-Enciso M (1990) Optimization of selection response under restricted inbreeding. *Genet Sel Evol* 22, 93-107