

A note on increasing the limit of selection through selection within families

By LEO DEMPFLER*

Department of Animal Science, Cornell University, Ithaca, N.Y. 14850†

(Received 10 January 1974)

SUMMARY

It is shown that, in a finite population, by ensuring an equal contribution of offspring from each family we achieve a higher selection limit than by using mass selection, given that the correlation between selection criterion and additive genetic value is sufficiently high. The difference between the selection limits of the two schemes increases with more intense selection. The theoretical results were verified by Monte Carlo Simulation, and the influences of several factors were investigated. It is shown that such a breeding scheme might be useful in dairy cattle.

1. INTRODUCTION

When we select a trait with low heritability whose genetic variation is caused by many independent loci each with a very small effect and without dominance, it is well known that the limit of selection is twice the effective population size times the genetic gain in the first generation (Robertson, 1960). If we select within full sib families in such a way as to have each family contribute exactly one male and one female, we double the effective population size, but we use only half the additive genetic variance and therefore the limit would not be changed as noted by Robertson (1960). However, under the circumstances of very intense selection, large families, and a high correlation between the selection criterion and the additive genetic value, it is possible to have a higher limit of selection through selection within families. In dairy cattle breeding, we have a situation where it may be useful to select within half-sib families. Only the case without dominance is considered here.

2. THEORY

When considering a trait in a large random breeding population which is influenced by many loci each with a very small effect and without dominance, we assume that the additive genetic values and the environmental effects are uncorrelated and follow normal distributions with variances V_A and V_E , respectively. It is further assumed that the nongenetic differences between full sib families are not

* A grant from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

† Present address: Institut für Tierzucht der Technischen Universität München, 8050 Freising-Weißenstephan, Germany.

important. By using a breeding scheme in which we test n offspring of each sex from each of the M full sib families every generation, we can compare the accumulated selection response of two procedures:

- (i) selecting the best male and the best female out of each full sib family (subsequently referred to as selection within families), and
- (ii) selecting among all tested animals disregarding any family relationships (subsequently referred to as mass selection).

In both schemes, the selected animals are then randomly mated and the number of families, M , is assumed to be large.

(i) *Selection of the best male and the best female out of each full sib family*

Here we do not have any variation in the number of offspring per individual, and therefore the effective population size approaches twice the number of breeding individuals when $N = 2M$ gets large. The expected selection gain per generation can be written as

$$\Delta G_W = \frac{i_W V_{AW}}{\sqrt{(V_E + V_{AW})}}$$

where i_W is the expected value of the highest order statistic from a sample of n standard normal variates and V_{AW} is the usable additive genetic variance within full sib families. In a random breeding population, V_{AW} is equal to $\frac{1}{2}V_A$, but with selection this value can change. It is clear that, without dominance and with random mating of the selected animals, V_{AW} is a function of the frequency of heterozygotes among the selected animals. It can be shown that the frequency of heterozygotes is generally changed very little by selection, as noted by Fisher (1918). At a locus that has no dominance, even with intense selection and large proportional effect, the change in the within family variance can be neglected. Therefore, the calculations are done with $V_{AW} = \frac{1}{2}V_A$. Under our assumptions, V_A changes each generation by approximately $(1 - (1/4N))$, and at generation t , V_{AW} will have a value of

$$V_{AW}(t) = \frac{1}{2}V_A \left(1 - \frac{1}{4N}\right)^t,$$

and the accumulated selection response for the two extreme cases will approach

$$\sum_{i=1}^{\infty} \Delta G(i) = \begin{cases} 2Ni_W V_A / \sqrt{V_E} & \text{as } V_A/V_E \text{ tends to zero.} \\ (\sqrt{2})4Ni_W / \sqrt{V_A} & \text{as } V_E/V_A \text{ tends to zero.} \end{cases}$$

(ii) *Selection disregarding pedigree relationships*

In this scheme, we select the top M of the females (males) out of the Mn females (males) tested. If the selection of the breeding animals were made at random, the distribution of the number of retained offspring would be hypergeometric with variance $(1 - 1/M)(1 - (M - 1)/(Mn - 1))$ in all four paths. This variance is lower than that assuming a binomial or Poisson distribution for the number of offspring

retained. Thus, the effective population size would be slightly greater than the number of breeding animals although the difference tends to zero when M and n are large. However, the individuals are selected according to their yield, and because members of a family are correlated the variance of the offspring number is usually increased. (For a more detailed discussion, see Robertson (1961) and Nei & Murata (1966).) Therefore, the effective population size calculated by using the variance of the hypergeometric distribution is an upper limit which may be considerably larger than the true effective population size when heritability is high and selection is intense. However, for the following calculation, N_e is taken equal to N .

The expected selection gain per generation can be written as

$$\Delta G_M = \frac{i_M V_{AM}}{\sqrt{(V_E + V_{AM})}},$$

where i_M is the expected mean of the top M order statistics from a sample of Mn standard normal variates and V_{AM} is the usable additive genetic variance, which is V_A in generation zero and for generation $t + 1$ can be expressed as

$$V_{AM}(t + 1) = [\frac{1}{2}V_A(t) + \frac{1}{4}V_D(t) + \frac{1}{4}V_S(t)]\left(1 - \frac{1}{2N_e}\right).$$

The first term reflects the within family variance (assuming that the frequency of heterozygotes is not changed much by selection), and V_D (V_S) is a measure of the additive genetic variation among the selected dams (sires). The exact value of $V_D(t)$ for small values of M and n will be given elsewhere, but it can be shown that

$$V_{AM}(t) [1 - r^2(t) i(i - x)] \left(\frac{M - 1}{M}\right)$$

is a good approximation to $V_D(t)$. Here i is the mean of the standardized normal distribution after truncation, x is the truncation point, and $r(t)$ is the correlation between selection criterion and additive genetic value in generation t . Random drift causes the genetic variance to change each generation by $(1 - (1/2N_e))$, which is taken to be equal to $(1 - (1/2N))$. Thus, we have

$$V_A(t) = V_A \left(1 - \frac{1}{2N}\right)^t$$

and
$$V_{AM}(1) = \frac{1}{2}V_A \left(1 - \frac{1}{2N}\right) \left\{1 + [1 - r^2(0) i(i - x)] \left(\frac{M - 1}{M}\right)\right\}.$$

Because $r(t)$ is changing, it is difficult to give a general expression for $V_{AM}(t)$ although numerical values can be easily calculated. It should be noted that in later generations the reduction of the usable additive genetic variance due to selection is larger than in the first generation. If $(M - 1)/M$ is close to one and if the reduction due to random drift can be neglected, V_{AM} goes to an equilibrium value of

$$\tilde{V}_{AM} = \{V_A - V_E + [(V_A + V_E)^2 + 4V_A V_E i(i - x)]^{\frac{1}{2}}\} / \{2[1 + i(i - x)]\}.$$

This result was also obtained by Bulmer (1971) using a different approach.

The accumulated selection response for the two extreme cases is approximately

$$\sum_{i=1}^{\infty} \Delta G(i) = \begin{cases} 2Ni_M V_A / \sqrt{V_E} & \text{as } V_A/V_E \text{ tends to zero,} \\ \sqrt{\left(\frac{1}{1+i(i-x)}\right)} 4Ni_M \sqrt{V_A} & \text{as } V_E/V_A \text{ tends to zero.} \end{cases}$$

Comparing the accumulated selection responses of the two schemes, disregarding the difference between i_W and i_M shows that the same total response is achieved if V_A/V_E is very small. However, if V_E/V_A tends to zero, the total response by selection within families is $\sqrt{2}$ times as high as the total response by mass selection when the selection intensity is low and twice as high if the selection intensity is very high. In the latter case, selection within families is always better than mass selection except in the first generation.

The reason for this behaviour is that with a very low heritability the increased effective population size by selection within families exactly balances the reduced genetic variance on which selection is acting and so the same limit as with mass selection is reached. However, when the heritability is very high, the selection response per generation becomes proportional to the square root of the usable additive genetic variance which gives the advantage to selection within families. In addition, with a high heritability the usable genetic variance is considerably reduced by a high selection intensity, but this affects only the between family component. Thus, the rate of response within families is the same in both schemes and only the response of mass selection is reduced giving a further advantage to selection within families. Also, if the family size is large, i_W tends to i_M and since the effective size by mass selection under the condition of high heritability and very intense selection may become considerably smaller than the actual number the advantage of selection within families is then even higher than calculated above.

3. MONTE CARLO SIMULATION AND RESULTS

In order to investigate the relative difference of the two accumulated selection responses for other situations, Monte Carlo Simulation was used for populations with $N = 4, 8$ and 12 individuals. Pairs were formed randomly and each pair had $n = 2, 5, 10, 20$ and 40 'female' offspring and the same number of 'male' offspring. Each individual had only one mate so there were no half-sibs. The loci were independent, had equal effect, and there was no dominance. Regardless of the number of loci, the difference between the two homozygotes at a locus was constant. The number of loci was $L = 10, 60, 150$ or 240 . At generation zero the frequency of the desired alleles was $q = 0.125, 0.5$ or 0.875 , and the correlation between additive genetic value and selection criterion was taken as $r^2 = 0.4, 0.65$ or 0.9 . The genotypes were in random mating proportions for a finite population at every locus in generation zero. During all generations, V_E was kept constant; as a consequence, V_P changed. With each generated population, both methods of selection were practised for 33 generations.

The results are shown in Figs. 1-4. The ratio of the accumulated selection response by selection within families to the accumulated selection response by mass selection is given for 33 generations.

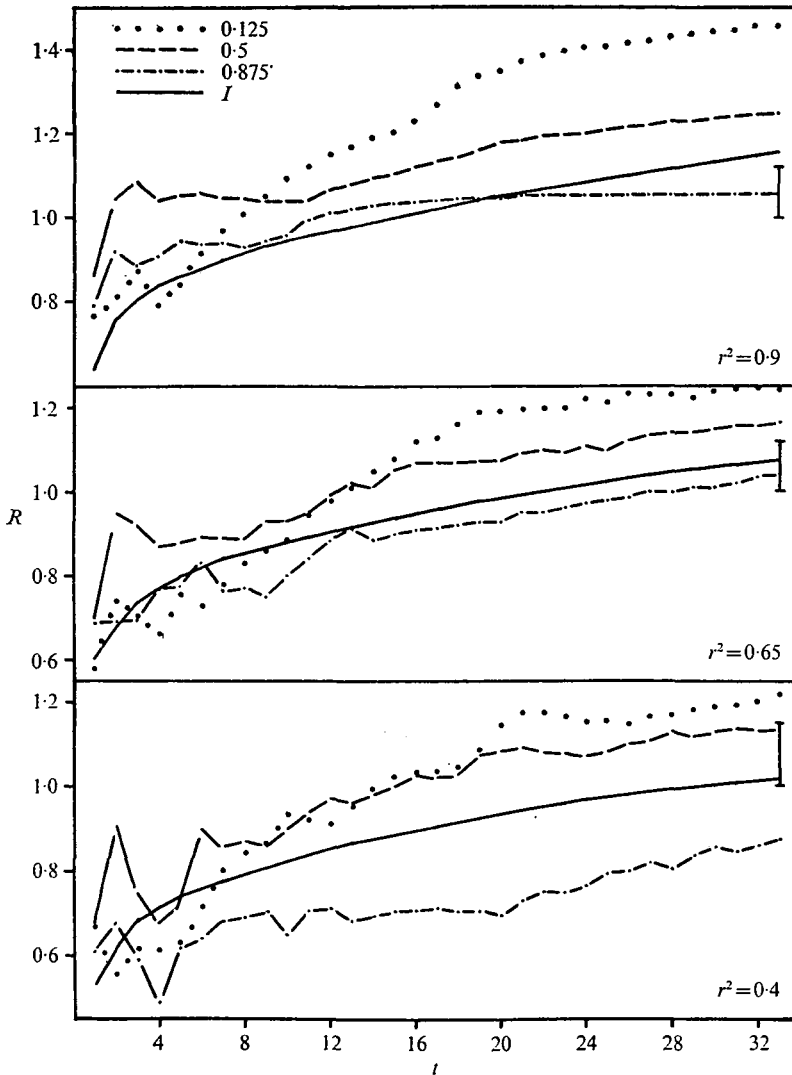


Fig. 1. Ratio of the accumulated selection response resulting from selection within families to the accumulated selection response resulting from mass selection

$$\left[R = \frac{\sum_{i=1}^t \Delta G_W(i)}{\sum_{i=1}^t \Delta G_M(i)} \right] \text{ with } N = 4, n = 20 \text{ and } L = 240, \text{ for three corre-}$$

lations at generation zero ($r^2 = 0.4, 0.65, 0.9$) and three initial allele frequencies ($q = 0.125, 0.5, 0.875$). Also shown are the results for a population with L going to infinity (I) and the approximate length of the largest standard error of the ratio in generation 33.

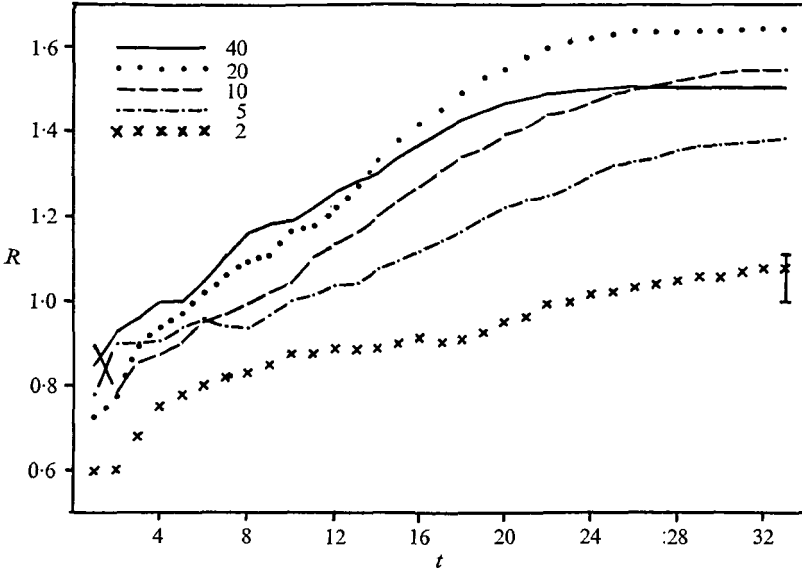


Fig. 2. Ratio of the accumulated selection response resulting from selection within families to the accumulated selection response resulting from mass selection for different family sizes ($n = 2, 5, 10, 20, 40$) with $N = 4, r^2 = 0.9, L = 150$ and $q = 0.125$.

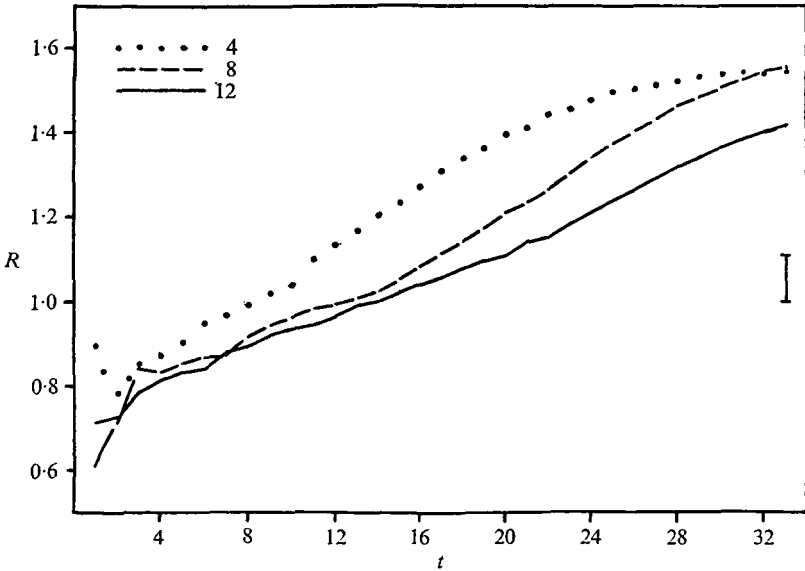


Fig. 3. Ratio of the accumulated selection response resulting from selection within families to the accumulated selection response resulting from mass selection for different population sizes ($N = 4, 8, 12$) with $n = 10, r^2 = 0.9, L = 150$ and $q = 0.125$.

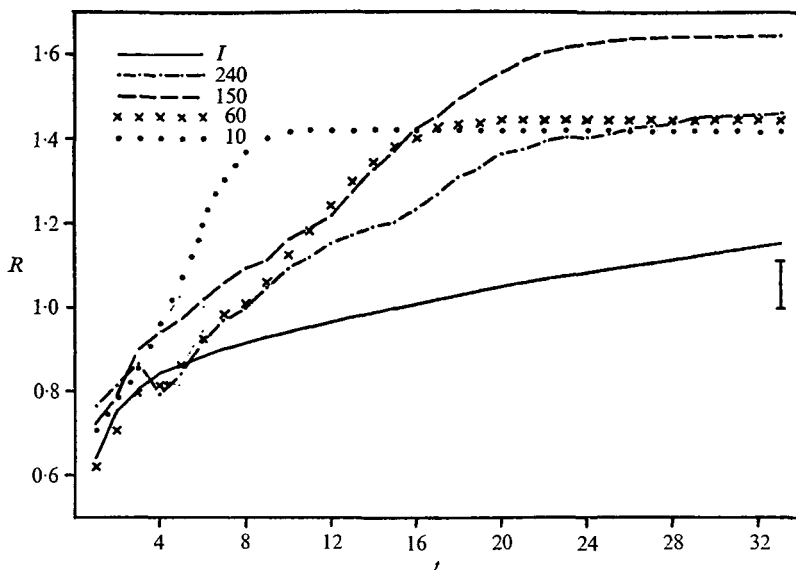


Fig. 4. Ratio of the accumulated selection response resulting from selection within families to the accumulated selection response resulting from mass selection for different numbers of loci ($L = 10, 60, 150, 240, \rightarrow \infty$) with $N = 4, n = 20, r^2 = 0.9$ and $q = 0.125$.

4. DISCUSSION

In good agreement with the theory, Fig. 1 shows clearly that the higher the correlation between selection criterion and additive genetic value the greater is the relative efficiency of selection within families compared to mass selection. The smaller the initial allele frequency of the desired allele the better is selection within families. This mainly reflects the fact that if we start with a high allele frequency the desired allele is almost always fixed in both selection schemes and therefore the ratio must go to one. For this reason, only the results for an allele frequency of 0.125 are shown in Figs. 2, 3 and 4.

Fig. 2 shows the influence of selection intensity and family size. The intensity of selection and thus the reduction of the additive genetic variance increases with the family size. With large families, the difference between $i_{\overline{w}}$ and i_M gets smaller and with more intense selection the effective population size under mass selection is likely to decrease. All three factors increase the advantage of selection within families and the results shown in Fig. 2 confirm this. The smaller ratio in later generations for $n = 40$ may be due to chance.

Fig. 3 illustrates the effect of population size. Selection within families increases the effective population size, but this increase is not independent of the population size. With the two breeding schemes used for computing Fig. 3, the ratio of the effective population sizes is approximately 1.5 if $N = 4$, 1.7 if $N = 8$, and 1.8 if $N = 12$. Therefore, the effect of selecting within families should be greater with $N = 8$ or $N = 12$. The reason why that superiority is not shown is mainly due to the time factor. The smaller the population the sooner are the limits reached in both

schemes. With the parameters of Fig. 3 but with $L \rightarrow \infty$, the ratio of $N = 4$ is highest and of $N = 12$ still lower at generation 33 according to the theoretical expectation and is reversed in later generations. It should be noted that the increase of the ratio in generations 30–33 is largest with $N = 12$ and smallest with $N = 4$ in agreement with the expectation.

The simulation results of Fig. 4 indicate that the number of loci or more generally the heritability of the locus effect in the range studied does not have a significant influence on the ultimate ratio of the two accumulated selection responses, but does have a marked influence on the time taken to reach it. The greater the heritability of the locus effect the sooner will the limit be reached. The simulation results are all higher than the expected curve for L going to infinity. That can be attributed to three factors. The smaller the loci effect the later is the ultimate ratio achieved. A comparison at generation 33 is therefore biased against $L \rightarrow \infty$. With an infinite number of loci, the additive genetic variance is always decreasing, whereas in the simulation it was at least partly compensated for by the increase in allele frequency from its initial value of 0.125. With the parameters of Fig. 4, this situation favours selection within families after the first few generations if L is large. The third and probably most important factor is the effective population size. For calculating the curve for $L \rightarrow \infty$, it was assumed that with mass selection the offspring number follows the hypergeometric distribution. Under the conditions of Fig. 4, that overestimates the true size and so the ratio is expected to be higher.

It should be noted that the reduction in the additive genetic variance due to selection which is not caused by a change in allele frequency is reversible. In 2(ii) we saw that by using mass selection the additive genetic variation among the selected animals in generation zero is approximately $V_A[1 - r^2(0) i(i-x)](1 - 1/M)$. If the effect of every locus is very small, then on discontinuing selection, increasing the population size, and letting the animals mate at random, the usable additive genetic variance would asymptotically return to $V_A(1 - (1/2N_e))$. At the gene level, the reduction in the variance of the selected animals is explained by an induced negative covariance between any two allele effects. With no dominance or epistasy, we can write the genotype of an animal as the sum of all allele effects,

$$G = \sum_{i=1}^L \sum_{j=1}^2 a_{ij}.$$

One generation of random mating is sufficient to nullify $\text{Cov}(a_{i1}, a_{i2})$, the covariance between allele effects at the same locus and to halve $\text{Cov}(a_{ij}, a_{hj})$, the covariance between allele effects of different loci. The covariance is halved by each further generation of random mating if there is no linkage and it is reduced less than half otherwise, as noted by Bulmer (1971).

From these results, it can be concluded that selection within families has its greatest effect if the allele frequency of the desired allele is low, the correlation between selection criterion and additive genetic value is high, and the family size is large. From the derivation, it is clear that these results apply similarly to selec-

tion within half-sib families. In a well-designed dairy cattle breeding scheme using AI, we have a high correlation between the selection criterion (daughter mean) and the additive genetic value of the test sire, and we can have large half-sib families by testing many sons from bulls with proven high evaluations. By selecting the best male of each half-sib family, we make the variance in the number of offspring in the male to male path equal to zero. If the numbers of males and females are equal, this increases the effective population size by $1/7$ over that of the random mating population. However, if the number of males is very much smaller than the number of females, the effective population size is increased by one-third when the variance in the male to male path is changed from one to zero. The increase is even higher if there is also a positive covariance between the numbers of male and female offspring of a sire. It can be shown that under these circumstances the decrease in immediate selection gain using within half-sib family selection is very small because the lower genetic gain in the first few generations is partly due to the fact that in the theoretical derivation and the simulation study a random breeding population was assumed in generation zero. If we change the selection scheme in an already existing breeding project, the reduction would be smaller. In addition, other factors such as inbreeding depression and nongenetic differences between families also make selection within families more advantageous.

I wish to thank the referee for helpful comments and suggestions.

REFERENCES

- BULMER, M. G. (1971). The effect of selection on genetic variability. *The American Naturalist* **105**, 201–211.
- FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.
- NEI, M. & MURATA, M. (1966). Effective population size when fertility is inherited. *Genetical Research* **8**, 257–260.
- ROBERTSON, A. (1960). A theory of limits in artificial selection. *Proceedings of the Royal Society of London B* **153**, 234–249.
- ROBERTSON, A. (1961). Inbreeding in artificial selection programmes. *Genetical Research* **2**, 189–194.