

## Optimising Two-Stage Independent Culling Selection in Tree and Animal Breeding

P.P. Cotterill and J.W. James

School of Wool and Pastoral Sciences, University of New South Wales, Kensington, N.S.W. (Australia)

**Summary.** Theory is given for a simple practical method of predicting gain from two-stage independent culling, where stage 1 of selection is for individual performance and stage 2 is for either progeny performance only, or an index combining individual and progeny performance. Expected gain is determined as a direct function of heritabilities, genetic correlations, selection intensities and progeny-testing capacity. Results show the effect these parameters can have on proportions selected at each stage and, if multiple selection criteria are used, traits selected for first. Methods are discussed in the context of tree and animal breeding, with an example taken from forestry.

**Key words:** Two-stage selection – Genetic response – Independent culling

### Introduction

When records on individuals within one generation accumulate over time, selection is often done in stages, usually two. This is particularly common where there is progeny testing. For example, a popular two-stage option in tree breeding is selection on individual performance (stage 1) followed by progeny testing (stage 2). Phenotypically superior trees are selected from wild populations, plantations, or previous-generation progeny trials. These selections are then progeny tested and at the same time vegetatively propagated in commercial seed orchards. The seed orchards are culled when results from the current round of testing become available. It is common in sheep breeding for sires to be individually selected as weaners and then progeny tested before use in the breeding population. In both cases important decisions must be made on what proportions to select at each stage and, if multiple selection criteria are used, what traits to select for first. Decisions would be made with the aim of maximizing gain, gain per-unit cost, or some similar criterion.

Previous determinations of expected gain from two-stage independent culling have mostly been applied to sheep breeding (Young and Weiler 1960; Williams and Weiler 1964; Young 1964). These authors used the moments of truncated bivariate normal distributions to calculate response to selection. Namkoong (1970) used equivalent methods to determine gain from two-stage independent culling in forestry, but also included a linear function based on relationships between cost and numbers retained for progeny testing. Each paper gives gains or gains per-unit cost for a range of parameter values, but extensive and tedious calculations are required to deal with circumstances not included in the tabulations. Cunningham (1975) studied alternative methods of selecting in successive stages, but gave no solutions for maximizing gain, while Rönningen (1970) and Eikje (1978) examined factors affecting the efficiency of two-stage selection compared with single-stage selection in particular animal breeding programs.

This paper attempts to provide a simple practical method for determining gain from two-stage independent culling selection. To do this, genetic gain is approximated by simply ignoring the fact that the population being considered at stage 2 is no longer normally distributed after selection at stage 1, and proceeding as if it were normal for all traits except that on which stage-1 selection is based. Calculations have been made for stage-1 selection on individual performance (may be an index value combining several traits) and stage-2 selection on either progeny test results alone, or on an index combining individual and progeny performance. It is assumed that progeny testing has not altered the generation interval.

### Theory

Consider a population of individuals that is to be selected at two stages to retain a final proportion  $p$ . Stage-1 selection is by truncation on a variable  $X$  at a point  $x_1$

standard deviations from the mean, so as to retain a proportion  $p_1$  having a standardised selection differential  $i_1$ . Stage-2 selection is by truncation on a variable Y to retain a proportion  $p_2$  of those selected at stage 1, where  $p_1 p_2 = p$ . Response from selection is measured as the change in the mean of a third variable, Z. Assume X, Y, Z are trinomally distributed in the base population with means  $\mu_x$ ,  $\mu_y$ , and  $\mu_z$ ; variances  $\sigma_x^2$ ,  $\sigma_y^2$  and  $\sigma_z^2$ ; and phenotypic correlations  $r_{xy}$ ,  $r_{xz}$  and  $r_{yz}$ .

The mean of Z in the population retained after stage-1 selection is

$$\bar{Z}_1 = \mu_z + i_1 r_{xz} \sigma_z \quad (1)$$

while the variance of X in this stage-1 selected population is

$$s_x^2 = \sigma_x^2 (1 - H) \quad (2)$$

where  $H = i_1 (i_1 - x_1)$ . The variance of Y in the selected population can be shown to be (Cochran 1951)

$$s_y^2 = \sigma_y^2 (1 - r_{xy}^2 H) \quad (3)$$

with a similar expression for  $s_z^2$ , the variance of Z. The covariance of Y and Z in the selected population is

$$s_{yz} = \sigma_y \sigma_z (r_{yz} - r_{xy} r_{xz} H). \quad (4)$$

Now we can approximate the mean of Z in the final population retained after stage-2 selection on Y. The approximation is obtained by simply assuming that Y and Z remain jointly normally distributed after stage-1 selection. This is not strictly correct but, as will be seen, is not very wrong. The mean of Z after stage-2 selection would then be

$$\bar{Z}_2 = \bar{Z}_1 + i_2 s_{yz}/s_y \quad (5)$$

where  $i_2$  is the standardised selection differential for a fraction  $p_2$  selected, and can be read from the usual tables. Using Eqs. (1, 3, 4 and 5)

$$\begin{aligned} \Delta Z &= (\bar{Z}_2 - \mu_z)/\sigma_z \\ &= i_1 r_{xz} + i_2 [(r_{yz} - r_{xy} r_{xz} H)/(1 - r_{xy}^2 H)^{1/2}] \end{aligned} \quad (6)$$

where response from selection is expressed in units of standard deviations of Z. The consequences of particular selection procedures can then be studied using appropriate  $i_1$ ,  $i_2$  and H values, together with correlations corresponding to the variables involved.

For example, suppose stage-1 selection is on individual performance for a trait U and stage-2 selection is on

progeny performance for a trait V, measured as the mean of n progeny per-individual tested at stage 2. The breeding objective is W, which may be a single trait or a combination of traits weighted for economic value. Then if  $h_u^2$  and  $h_v^2$  are the heritabilities,  $g_{uv}$ ,  $g_{uw}$  and  $g_{vw}$  the genetic correlations among traits, and t is the intra-class correlation between members of families in the progeny test; the phenotypic correlations can be written

$$r_{xy} = \frac{1}{2} g_{uv} h_u h_v / K^{1/2} \quad (7)$$

$$r_{xz} = g_{uw} h_u \quad (8)$$

$$r_{yz} = \frac{1}{2} g_{vw} h_v / K^{1/2} \quad (9)$$

where

$$K = (1 + (n - 1)t)/n. \quad (10)$$

If the progeny test is for the same trait as individual performance, Eqs. (7) to (9) simplify to

$$r_{xy} = \frac{1}{2} h^2 / K^{1/2} \quad (11)$$

$$r_{xz} = h \quad (12)$$

$$r_{yz} = \frac{1}{2} h / K^{1/2}. \quad (13)$$

In this paper values for n, the number of progeny per-individual tested at stage 2, have been determined as

$$n = N_t / (N_b p_1) \quad (14)$$

where  $N_t$  is the testing capacity of the breeding operation, measured as the total number of offspring which can be physically managed in a progeny test in a certain period of time,  $N_b$  is the size of the base population that is subjected to stage 1 of selection and  $N_b p_1$  is the number of individuals selected at stage 1 for testing at stage 2.

So far it has been assumed that stage-2 selection is on the progeny test alone. However, stage-2 selection may be on an index combining individual and progeny performance. In this case the appropriate index coefficients would be

$$(\sigma_z / \sigma_x) (r_{xz} - r_{xy} r_{yz}) / (1 - r_{xy}^2) \quad (15)$$

$$(\sigma_z / \sigma_y) (r_{yz} - r_{xy} r_{xz}) / (1 - r_{xy}^2) \quad (16)$$

where the correlations are as given by Eqs. (7) to (9). Then if I denotes the index it can be shown that in the base population

$$\begin{aligned} \sigma_i^2 &= \sigma_z^2 [(r_{xz} + r_{yz})^2 / (1 - r_{xy}^2) \\ &\quad - 2r_{xz} r_{yz} / (1 - r_{xy}^2)] \end{aligned} \quad (17)$$

$$r_{xi} = r_{xz} \sigma_z / \sigma_i \quad (18)$$

$$r_{zi} = \sigma_i / \sigma_z. \quad (19)$$

From the generalised formula of Cochran (1951) and Eqs. (18) and (19) it can be shown that in the stage-1 selected population

$$s_i^2 = \sigma_i^2 - r_{xz}^2 \sigma_z^2 H \quad (20)$$

$$s_{iz} = s_i^2. \quad (21)$$

Cunningham (1975) made an equivalent observation to Eq. (21), that index weighting factors are not affected by prior selection. Using this and previous relationships it was found that

$$\Delta Z = i_1 r_{xz} + i_2 [(r_{xz} + r_{yz})^2 / (1 - r_{xy}^2) - 2r_{xz}r_{yz} / (1 - r_{xy}^2) - r_{xz}^2 H]^{1/2}. \quad (22)$$

The gain calculations that follow are based on Eqs. (6) and (22), and their reliability has been judged by comparison with direct estimates determined using Pearson's Tables (1931) and gain formulations comparable to those in Young and Weiler (1960). Differences between the direct and indirect estimates were always small (less than 1%). This was true even for circumstances in which the approximation would be assumed most vulnerable, such as intense stage-1 selection on a highly heritable trait with strong genetic correlations with V and W.

If gains per-unit cost are required, then solutions to Eqs. (6) and (22) may be simply divided by a linear cost function of the type used by Namkoong (1970)

$$\text{cost} = N_b C_1 + p_1 N_b C_2 \quad (23)$$

where  $C_1$  and  $C_2$  are per-unit costs at stages 1 and 2.

### Estimates of Expected Gain

Expected gain has been plotted against  $p_1$  for a range of heritabilities, genetic correlations, final proportions selected, testing capacities, and stage-2 selection with and without the index. Calculations have been made for half-sib progeny ( $t = 1/4 h_V^2$  in equation 10), but it can be shown that trends apply equally for full-sibs. Actual gains from full-sib tests are just a little lower due to the higher coefficient of relationship between members of families ( $t = 1/2 h_V^2$ ).

Combining U and V into an index for stage-2 selection was found to have very little effect on either maximum

gains or the  $p_1$  at which these maximum were reached, hereafter referred to as maximum-gain  $p_1$  (Fig. 1a). Although the gain calculations presented are for a specific set of circumstances ( $h_U^2$  and  $h_V^2 = 0.2$ ,  $g_{UV} = -0.2$ ,  $g_{UW}$  and  $g_{VW} = 0.6$ ) the trend is general and, as a result, independent culling without the stage-2 index would usually be preferred. It is much simpler to operate and has only a minor gain disadvantage; under this option stage 1 records need not be retained or, for that matter, reassembled into an index at stage 2. Cunningham (1975) also found that gains from independent culling without the stage-2 index were almost as good as those with the index. This author's calculations were based on swine

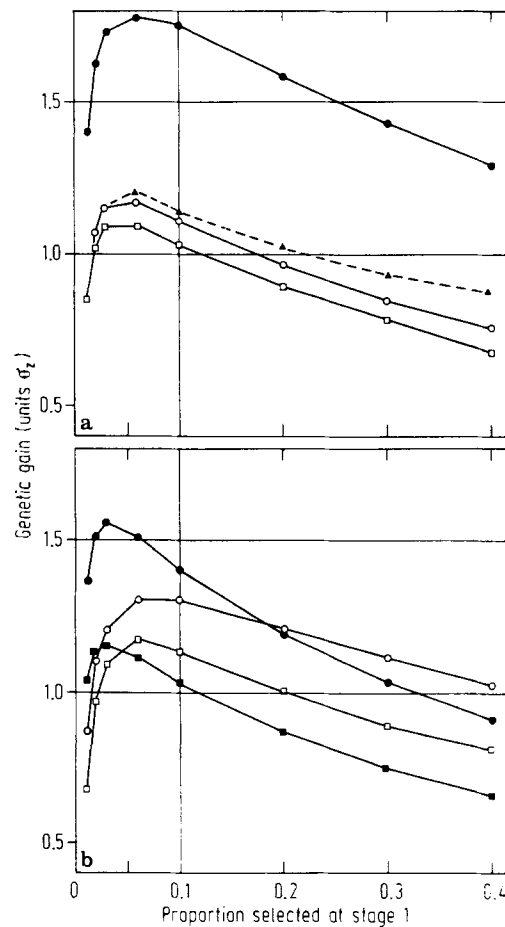


Fig. 1a and b. Two-stage independent culling on two different traits U and V:  $p$  is fixed at 0.01 and the testing capacity equals the base population. a Expected gains from selection with  $\blacktriangle$  and without  $\circ$  the stage-2 index; calculated for  $h_U^2$  and  $h_V^2 = 0.2$ ,  $g_{UV} = -0.2$ ,  $g_{UW}$  and  $g_{VW} = 0.6$ . Gains from selection without the stage-2 index showing the effect of general increases in  $h_U^2$  and  $h_V^2$  from 0.2 and 0.2  $\circ$  to 0.6 and 0.6  $\bullet$ ; or  $g_{UV}$  from  $-0.2$   $\circ$  to 0.6  $\square$ . b Expected gains from selection without the stage-2 index:  $g_{UV}$  is fixed at 0.1; with both  $g_{UW}$  and  $g_{VW} = 0.6$ , and  $h_U^2$  and  $h_V^2$  equal either 0.6 and 0.2  $\bullet$  or 0.2 and 0.6  $\circ$ ; or with both  $h_U^2$  and  $h_V^2 = 0.2$ , and  $g_{UW}$  and  $g_{VW}$  equal either 0.8 and 0.4  $\blacksquare$  or 0.4 and 0.8  $\square$ .

data with  $p$  fixed at 0.06 (or 6%) and  $p_1$  at 0.38. The remaining gain calculations presented in this paper will be for stage-2 selection on progeny test results only.

Equal increases in both  $h_u^2$  and  $h_v^2$  clearly caused substantial improvements in gain (Fig. 1a) and, although results are not presented, equal increases in  $g_{uw}$  and  $g_{vw}$  had a similar effect. Conversely, an increase in  $g_{uv}$  from  $-0.2$  to  $0.6$  caused a slight decrease in gain (Fig. 1a). However, none of these changes in genetic parameters significantly altered maximum-gain  $p_1$ .

In practice, heritabilities and genetic correlations can be potentially most important when either  $h_u^2$  and  $h_v^2$  or  $g_{uw}$  and  $g_{vw}$  differ. Under these circumstances the order in which traits are selected for can have a critical effect on the efficiency of application of two-stage selection. For example, Fig. 1b shows that when  $h_u^2$  and  $h_v^2$  differ, maximum gain can be considerably increased and maximum-gain  $p_1$  decreased by selecting for the most heritable trait first. The expense of progeny testing usually ensures that a considerable cost advantage is associated with any decrease in  $p_1$ . This is particularly true in forestry where estimates of the ratio of per-unit costs at stages 2 and 1 ( $C_2/C_1$ ; equation 23) are very high, ranging from around 800 to  $10^6$  (Namkoong 1970). When  $g_{uw}$  and  $g_{vw}$  differ, maximum-gain  $p_1$  are decreased by stage-1 selection for the trait which has the strongest genetic association with W (usually the trait of greatest economic importance), but maximum gain is not greatly affected (Fig. 1b).

When both  $h_u^2$ ,  $h_v^2$  and  $g_{uw}$ ,  $g_{vw}$  differ it can be shown that stage-1 selection should remain on the most heritable trait. If this trait also happens to have the strongest genetic association with W then maximum gain will be greatly increased and maximum-gain  $p_1$  greatly decreased. If the trait has the weakest genetic association with W then maximum-gain  $p_1$  will be slightly increased, but actual gains made will remain greater at every level of  $p_1$ . Of course the order in which traits are selected for is sometimes predetermined. For instance, traits which can only be measured by destructive sampling (biomass of trees, carcass composition of lambs) must obviously be selected for at stage 2.

So far all gain calculations have been made with  $p$  fixed at 0.01 and a testing capacity equal to the size of the base population (conveniently expressed by the ratio  $N_t/N_b = 1.0$ ; equation 14). Figure 2 illustrates the effect of independent changes in these parameters. A decrease in  $p$  from 0.05 to 0.001 clearly caused a considerable increase in maximum gain and a decrease in maximum-gain  $p_1$  (Fig. 2a). However, the intensity of final selection (lower limit to  $p$ ) is invariably predetermined by the size of the base population and tolerable levels of inbreeding. Tree breeding often involves relatively intense selection ( $p$  greater than say 0.001), particularly

in early generations when selection is from large 'wild' populations. Final selection intensities in sheep are usually somewhat lower.

An increase in testing capacity from 1/5 to five times the base population ( $N_t/N_b$  from 0.2 to 5.0) caused an increase in both maximum gain and maximum-gain  $p_1$  (Fig. 2b). If in advanced generations of tree breeding the progeny tests of one generation become the base population of the next, then  $N_t$  should at least equal  $N_b$  in order to perpetuate the breeding population and maintain final selection intensities. In circumstances where the progeny test and breeding populations are separate, the economics of large testing capacities deserve close scrutiny (Namkoong 1970), particularly since more individuals need to be selected at stage 1 to take full advantage of maximum gains (Fig. 2b). Sheep breeders usually operate with  $N_t/N_b$  ratios well below 1.0.

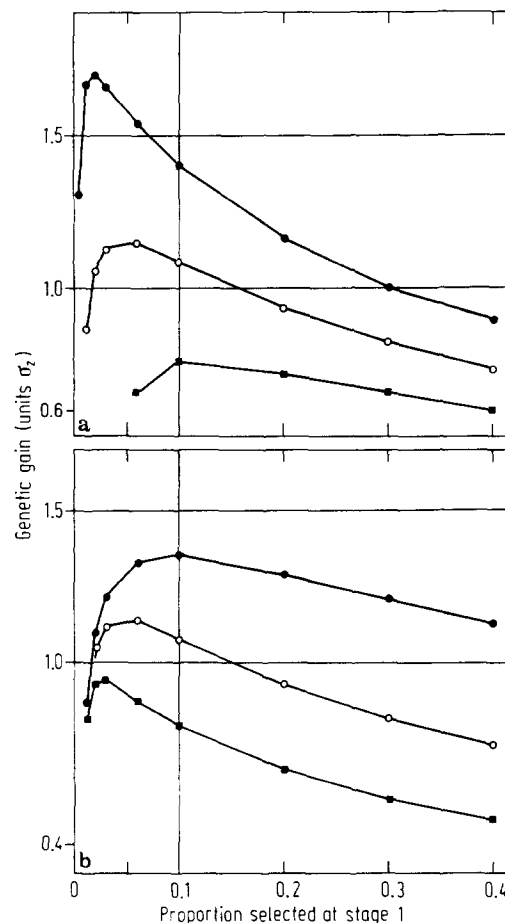


Fig. 2a and b. Two-stage independent culling on two different traits U and V without the stage-2 index:  $h_u^2$  and  $h_v^2$  are fixed at 0.2,  $g_{uv}$  at 0.1,  $g_{uw}$  and  $g_{vw}$  at 0.6. a Expected gains from selection with  $N_t/N_b = 1.0$  and showing the effect of an increase in  $p$  from 0.001 ● to 0.01 ○ to 0.05 ■. b Expected gains with  $p = 0.01$  and showing the effect of an increase in  $N_t/N_b$  from 0.2 ■ to 1.0 ○ to 5.0 ●

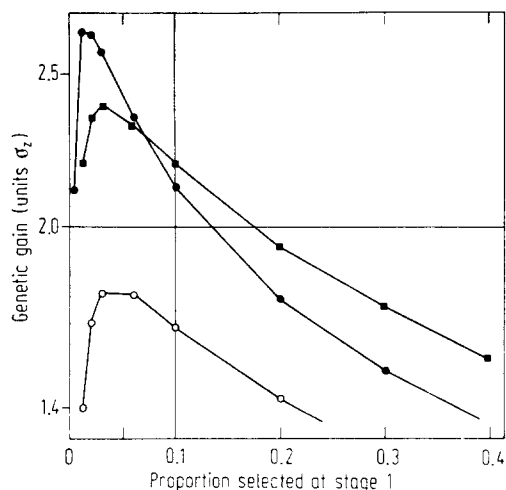


Fig. 3. Two-stage independent culling on one trait and without the stage-2 index:  $N_t/N_b$  is fixed at 1.0; with  $h^2 = 0.6$  and  $p = 0.01$  ■; or  $h^2 = 0.2$  and  $p$  equal either 0.001 ● or 0.01 ○

When selection is for the same trait at both stages (U and V are the same trait) gains are substantially greater and increase with heritability and final selection intensity, while the maximum-gain  $p_1$  tend to be lower (Fig. 3). Although results are not presented, changes in  $N_t/N_b$  and selection with and without the stage-2 index were found to have the same effects as they did on selection for two traits.

### Example from Tree Breeding

At the time of writing, two-stage independent culling without the stage-2 index was about to be used in a *Pinus radiata* D. Don breeding program in Australia run by the Woods and Forests Department of South Australia and CSIRO. Superior trees are to be selected within full-sib family-groups which occur in a previous-generation progeny test. Each family-group consists of approximately 100 progeny (or expressed another way,  $N_b = 100$ ); only one member from each group is to be finally retained ( $p = 0.01$ ); and testing capacity is equal to  $N_b$ .

Selection is for two indices U and V, which have been constructed using genetic and phenotypic parameters given in Cotterill and Zed (1980). Index U combines stem volume, stem straightness, branch diameter and branch angle; while index V combines stem volume and wood density. The breeding objective W is a combination of traits weighted for economic value (using a multivariate linear equation). The heritabilities of U and V, calculated from within-family deviations (Falconer 1960), are 0.23 and 0.15;  $g_{UV}$  is  $-0.29$ ;  $g_{UW}$  and  $g_{VW}$  are 0.77 and 0.63.

Index U has the highest heritability and strongest genetic association with W and will be selected for first. At this stage 2% of each family-group will be retained ( $p_1 = 0.02$ ). Table 1 shows that for the circumstances in this example, selecting 2% at stage 1 can be expected to give near-maximum gains in W. An increase in  $p_1$  from 0.01 to 0.02 caused a 36% increase in expected gain, while a further increase in  $p_1$  to 0.04 (the maximum-gain  $p_1$ ) increased response by only 5% (Table 1). The same conclusion

Table 1. Expected gains from two-stage independent culling on two different traits U and V without the stage-2 index:  $p = 0.01$ ,  $N_t/N_b = 1.0$ ,  $h_u^2 = 0.23$ ,  $h_v^2 = 0.15$ ,  $g_{UV} = -0.29$ ,  $g_{UW} = 0.77$ ,  $g_{VW} = 0.63$

	Proportion selected at stage 1 ( $p_1$ )			
	0.01	0.02	0.03	0.04
Expected gain <sup>a</sup>	0.984	1.355	1.394	1.401
Proportion selected at stage 2 ( $p_2$ )	zero	0.50	0.33	0.25
Progeny per individual tested (n)	100	50	33	25

<sup>a</sup> In units of  $\sigma_z$  or  $h_w\sigma_w$

could have been reached from previous gain calculations. Figure 1a shows that for circumstances similar to this example, that is  $h_u^2$  and  $h_v^2 = 0.2$ ,  $g_{UV} = -0.2$ , and  $g_{UW}$  and  $g_{VW} = 0.6$ , near-maximum gain is achieved by selecting about 3% at stage 1. The important difference is that in the worked example  $h_u^2$  and  $g_{UW}$  are greater than  $h_v^2$  and  $g_{VW}$ , and under these circumstances selecting 2% at stage 1 would reasonably be expected to approximate maximum gain (Fig. 1b).

To summarise the breeding operation: stage 1 requires that two individuals having the highest phenotypic values for index U be retained from each family-group for progeny testing. The testing capacity will permit about 50 progeny per-individual tested ( $n = 50$ ; Table 1). These progeny will probably be open-pollinated from seed collected in the base population. Stage 2, that individual whose progeny have the highest average performance for index V is finally retained from each family-pair. This would be expected to be the individual having the highest or near-highest breeding value for W of all members in its original family-group. If seed orchards are established after stage 1 they would be culled by 50% after stage 2 ( $p_2 = 0.50$ ; Table 1).

### Acknowledgement

Thanks to Christine Dean and Ludek Wolf for their assistance and Dr. A.R. Griffin for some useful suggestions.

### Literature

- Cochran, W.G. (1951): Improvement by means of selection. In: Proc. 2nd Berkeley Symp. on Math., Stat. and Probability (ed.: Neyman, J.), pp. 449-470
- Cotterill, P.P.; Zed, P.G. (1980): Estimates of genetic parameters for growth and form traits in four *Pinus radiata* D. Don progeny tests in South Australia. Austr. For. Res. 10, 155-167
- Cunningham, E.P. (1975): Multi-stage index selection. Theor. Appl. Genet. 46, 55-61
- Eikje, E.D. (1978): Genetic progress from performance and progeny test selection in Down sheep. Proc. N.Z. Soc. Anim. Prod. 38, 161-173
- Falconer, D.S. (1960): Introduction to Quantitative Genetics, pp. 231-237. London: Longman

- Namkoong, G. (1970): Optimum allocation of selection intensity in two stages of truncation selection. *Biometrics* 26, 465-476
- Pearson, K. (1931): *Tables for Statisticians and Biometricians. Part 2.* London: Cambridge Univers. Press
- Rönningen, K. (1970): Studies on selection in animal breeding. 2. The efficiency of selection based on the combination of performance testing and progeny testing compared with selection based only on performance testing when testing capacity is limited. *Acta Agric. Scand.* 20, 71-104
- Williams, J.M.; Weiler, H. (1964): Further charts for the means of truncated normal bivariate distributions. *Austr. J. Statistics* 6, 117-129
- Young, S.S.Y. (1964): Multi-stage selection for genetic gain. *Heredity* 19, 131-143
- Young, S.S.Y.; Weiler, H. (1960): Selection for two correlated traits by independent culling levels. *J. Genet.* 57, 329-338

Received August 14, 1980  
Communicated by W.J. Libby

Dr. Paul Cotterill  
CSIRO Division of Forest Research, P.O. Box 946  
Mount Gambier, South Australia 5290 (Australia)

Assoc. Prof. John James  
School of Wool and Pastoral Sciences  
University of New South Wales, P.O. Box 1  
Kensington, New South Wales 2033 (Australia)