DESIGNING ARTIFICIAL SELECTION
EXPERIMENTS FOR SPECIFIC OBJECTIVES

B. B. BOHREN

Purdue University, Agricultural Experiment Station, Lafayette, Indiana 47907

Manuscript received February 13, 1974
Revised copy received December 23, 1974

ABSTRACT

The observed genetic gain (ΔP) from selection in a finite population is the possible expected genetic gain E(ΔG) minus the difference in inbreeding depression effects in the selected and control lines. The inbreeding depression can be avoided by crossing the control and selected parents to unrelated mates and summing the observed gains. The possible expected gain will be reduced by an amount D from the predicted gain because of the effects of the genetic limit and random genetic drift, the magnitude of which is a function of effective population size, N. The expected value of D is zero in unselected control populations and in the first generation for selected populations. Therefore, this source of bias can be reduced by increasing N in the selected populations and can be avoided by selecting for a single generation. To obtain observed responses which are unbiased estimates of the predicted response from which to estimate the realized heritability (or regression) in the zero generation, or to test genetic theory based on infinite population size, single-generation selection with many replications would be most efficient. To measure the “total” effect or genetic efficiency of a selection criterion or method, including the effect of different selection intensities, effective population sizes, and space requirements, more than one generation of selection is required to estimate the expected response in breeding values. The efficiency, in the sense of minimum variance, of estimating the expected breeding values at any generation t will decline as the number of generations t increases. The variance of either the estimated mean gain or the regression of gain on selection differential can be reduced more by increasing the number of replicates K than by increasing the number of generations t. Also the general pattern of the response over t can be estimated if the N’s are known. Therefore, two- or not more than three-generation selection experiments with many replications would be most efficient.

EXPERIMENTS involving artificial selection for many traits in many species have been reported. Frequently the results have been puzzling to the authors because they did not closely follow predictions based on selection theory. In such a case it is necessary to reexamine the theory involved. In a series of excellent papers, Hill (1971, 1972a,b,c,d) reviewed problems in interpreting selection responses for single replications, but frequently made simplifying assumptions applying only to selection in populations of infinite size. Since artificial selection

1 This investigation was conducted as a portion of the cooperative research of the N.C. 89 Regional Poultry Breeding Project. Journal Paper No. 5408 of the Purdue University Agricultural Experiment Station.

is conducted only in finite populations, it is desirable to attempt to further summarize what is known that would be important in the design, analysis and interpretation of artificial selection experiments.

GENERAL CONSIDERATIONS

Consider the expected phenotypic means for any trait $T$ in an unselected or control population $(c)$ of size $NK$ where $K$ is the number of subpopulations, each of effective size $N$. The expected value $(E)$ of the mean is defined in the statistical sense as the mean obtained from an infinite number of samples $(K=\infty)$ of effective size $N$. If $N=\infty$, the value of $K$ is immaterial, and at generation $t$ the expectation of the mean $\bar{P}_{ct}$ is $E(\bar{P}_{ct}) = \bar{P}_{co}$ and has no variance. The only variation would be between generations, and would be due to environmental changes. If $N<\infty$, there would be variance among the $K$ sublines around the population mean due to genetic drift which would accumulate over generations. If only additive genetic variance is present, the $E(\bar{P}_{ct})$ would still be $\bar{P}_{co}$ and would also be unaffected by $K$. But, if $K$ is less than an infinite number, there is an additional variance of the population mean around the expected value. The expected values of the means over generations for such a control line are illustrated as curve 5 in Figure 1.

More serious consequences of finite population size are evident in selected populations. If $N=\infty$ (K again immaterial) the population mean would eventually reach the unknown genetic limit (curve 2, Figure 1). In most cases the mean at any time $t>1$ would be below the linear prediction on the basis of parameters estimated in the first generation by an unknown amount $D^c$ (curve 1, Figure 1) because of the asymptotic approach to the selection limit, brought about by restrictions on the selection differentials and the genetic variance imposed by the upper genetic limit. An exception would occur in the unlikely situation where many desirable alleles are at frequencies much below 0.5 in the zero generation. Contrary to the expectations in the unselected populations, when $N<\infty$, $K=1\rightarrow \infty$, the expected mean at any generation after the first and at the limit (curve 3, Figure 1) is reduced by a further unknown amount $D^d$ from that expected for $N=\infty$. This reduction in response, $D^d$, occurs as a result of random drift causing the $K$ sublines to vary around the expected mean, and some loci in some sublines will become fixed at values less than the upper genetic limit. This reduction in the expected value of the mean due to drift, $D^d$, does not occur in the first selected generation of parents because they are chosen from the zero generation in which the expected mean of the parents chosen in each subpopulation would be the population mean plus the genetic gain.

If non-additive genetic variance is present in the trait under selection, any study in which $N<\infty$ would show the phenotypic effect of inbreeding depression, resulting in an additional reduction $(F_i)$ in the expected phenotypic mean. The magnitude of this reduction would depend not only upon the amount of inbreeding, but upon the number of undesirable recessive alleles segregating and the magnitude of their effects. The depression would be less severe in lines under
Figure 1.—The relationship between the predicted $P(\Delta G)$ and possible $E(\Delta G)$ expected responses in the breeding values and the observed response ($\Delta P$) to selection in a finite population of effective numbers $N$ with $K$ replications at generation $t$.

These ideas are basically those of Wright (1931) who expressed them in terms of gene frequency distributions. The effect of population size on both unselected and selected populations has been summarized and quantified in terms of the probability of fixation for single loci by Kimura (1964) and in terms of the means by Robertson (1960), and the extension of the results to quantitative traits was discussed by the latter. In general, the reduction of the expected mean for a quantitative trait selected in a finite population from that expected in a selected population of infinite size is a function of the effective population size $N$ and the selection intensity $i$, or $Ni$. As pointed out by Robertson (1961), selection has its effects by preventing all pairs of parents from having an equal opportunity to have their progeny chosen as parents of the next generation, resulting in an
additional reduction in effective population size. It is clear that random genetic drift, a function of effective population size \( N \), has an effect on the expected gain at any generation \( t > 1 \). Experimental biological verification of this effect has been provided by Frankham, Jones and Barker (1968), Jones, Frankham and Barker (1968), Hanrahan, Eisen and Legates (1973) and Eisen, Hanrahan and Legates (1973), and by computer simulation studies by Gill (1965).

Another source of variation in selected or control populations is genotype by environment interactions. These interactions may be between lines and generation environments, or they may exist between lines and the environments of replications within any one generation. The effect of such interactions would be similar to that of random drift, except that the variance of these effects would not be cumulative over generations. If interaction is present, changing environments over generations would result in lower genetic gains than would be expected in a constant environment, because selection would be for the average effect over the set of environments. When selection and testing occur in the same constant environment, the genotype by environment interaction variance can be thought of as additional genetic variance. However, if selection is in one environment and testing in another, the genetic response could be negative. Within any one generation, a genotype by replication interaction would simply increase the variance of the means about the expected values in the same manner as does random drift.

While we have considered the curves in Figure 1 to represent the phenotypic means over generations of a single population, say \( A \), as affected by different selection limits, effective population sizes and types of gene action, they can also be considered as the means of the progeny in the \( t \) generation of the parents chosen in the \( t-1 \) generation. In this case these curves represent the average breeding values of the parents chosen each generation (Falconer 1960), and genetic gains each generation are measured as changes in breeding values estimated as the difference in the progeny means between generations \( t \) and \( t + 1 \).

If males and females from population \( A \), selected on their performance in population \( A \), are mated only to selected individuals from population \( A \), the situation is as previously discussed, and it is seen that the breeding value of population \( A \) when mated \( A \times A \) includes the effect of inbreeding on the phenotype (curve 4) in addition to the genetic gain. The controls would be mated within line also and their mean would also be affected by inbreeding (curve 6). It would be a rare occurrence if the inbreeding in the control and selected lines, and therefore the inbreeding effects, were identical.

The selected male and female parents in population \( A \) could be reciprocally mated to individuals in an unrelated population \( B \). The mates from population \( B \) for the selected parents from population \( A \) could be random or unselected, in which case the mean of the reciprocal crosses would represent the mean of the breeding values of the reciprocal crosses between the unselected control populations \( A \) and \( B \) (curve 5) plus half of the gain in breeding value due to selection in population \( A \) (curve 3). The predicted responses (curve 1) and the upper genetic limit (curve 2) would be based on the cross performance so would also represent only half of the change in breeding value in population \( A \). If the mates
in population $B$ were also selected on their performance in population $B$, then the mean of the reciprocal crosses of the selected parents would represent the mean of the reciprocal crosses of the controls plus the average gain in breeding values of populations $A$ and $B$. In such crosses, there would be no inbreeding effect in either the control or selected progeny.

Individuals in population $A$ could also be selected and their breeding value measured on the basis of the performance of their progeny when crossed to another population, say $C$, as in reciprocal recurrent selection. This could involve a change in the selection criterion and thus a change in the predicted response, even if only additive genetic variance is present. But if only additive or dominance variance is present the upper genetic limit in breeding values would be the same in $A \times B$ and $A \times C$ populations, and would be reached when all desirable genes had been fixed at frequencies of zero or one. If overdominant gene action was involved, however, the expected upper genetic limit at such loci for populations $A$ and $B$ selected within line would occur when the gene frequencies were at their equilibrium values, intermediate between zero and one. The expected upper limit for such loci in $K$ pairs of subpopulations $A$ and $C$ selected on crossbred performance would be when the average $q$ values over the $K$ subpopulations of $A$, or $q_A = 1 - q_C$ or when $q_A + q_C = 1.0$ or the average of $q_A$ and $q_C$ equals one-half. This is true whether selection occurs in the subpopulations of $C$ or not. However, the value of $q_A$ at the upper genetic limit in breeding value of population $A$ will depend upon whether or not selection is conducted in population $C$ and the relative initial gene frequencies of populations $A$ and $C$.

The fact that the maximum breeding values for population $A$ have different expectations, depending upon both the criterion of selection and the population with which it is mated for measurement, suggests that the breeding value measured by matings of $A \times A$, $A \times B$ and $A \times C$ should be considered as different traits.

**PREDICTED, POSSIBLE AND OBSERVED RESPONSES**

The theoretical genetic gain ($\Delta G_{T, C}$) or increase in breeding value for trait $T$ from selection on any criterion $C$ is,

$$\Delta G_{T, C} = I_0 \beta_{T,P_C} = I_0 \sigma_{T,P_C} \sigma_T / \sigma_P = I_0 \sigma_{T,P_C} \phi_{T,P_C} / \sigma_P,$$

(1)

where $I_0$ is the phenotypic selection differential in the criterion of selection, and $\beta_{T,P_C}$ is the regression of the breeding values for the trait $T$ to be improved on the phenotypes for the criterion of selection, $C$. The breeding value to be improved, $T$, can be for a single trait or any combination of traits or sources of information as a compound genotype. The criterion of selection may be the phenotype of a single trait or any combination of sources of genetic information and/or traits in a phenotypic index, and may or may not include the trait to be improved by the selection. The breeding value for the trait $T$ is estimated for an individual in a pure line, and could be quite different if the mates come from the same line rather than from an unrelated line. The criterion is also estimated for an individual and may be based on the individual's performance within its own
population or when crossed to another specific population. In the simplest case, when the criterion of selection and the traits to be improved are the same \( (C = T) \), \( \beta_{\alpha \gamma p_t} = h^2 \). If \( C \neq T \), then a correlated response is involved. Predictions of response in breeding value to selection \( P(\Delta G) \) are made by substituting parameters from the zero generation into the right-hand side of equation (1). On the other hand the observed (or realized) genetic gains or responses to selection can be estimated only as the difference \( (\Delta \alpha \gamma p_t) \) between two phenotypic means. The difference may be between the mean of a selected line after \( t \) generations of selection, and the mean of the original population (in which case a necessary assumption is that the environment does not change between generations), or between the mean of the selected line and that of a control population in the same generation, the only situation to be considered further.

The model for the individual phenotypic observation in a given selection method in any generation \( t \) would be:

\[
P_{ijn} = \mu + L_i + R_j + LR_{ij} + e_{ijn},
\]

where \( L_i \) represents the line effect of the \( i^{th} \) selection method \( (i = 1, \ldots, S) \), \( R_j \) is the replicate effect \( (j = 1, \ldots, K) \), and \( LR_{ij} \) is the interaction of the breeding value of the \( i^{th} \) method with the \( j^{th} \) replicate, and \( e_{ijn} \) is the random deviation of the \( n^{th} \) observation \( (n = 1, \ldots, M) \) from the mean of the \( ij^{th} \) cell. The term \( L_i \) is the only estimable line effect and is based on the observed means as seen in curves 3, 4, 5 or 6 in Figure 1, and can be considered as the sum of three parts, or \( L_i = G_i + D_i + F_i \). Here \( G_i \) is the expected breeding value predicted from the base population parameters and the accumulated selection differentials (curve 1). The \( D_i \) effect is made up of two parts. The first part, \( D^4 \), is the decrease in the expected breeding value because of the existence of, and the proximity to, a limiting genetic value which cannot be estimated (curve 2), and a second part, \( D^6 \), the reduction in the expected values of the mean due to the fixation of loci resulting from the fact that the effective population size \( N \) is finite (curve 3) and is not dependent upon the type of gene action involved. This curve could be referred to as the "possible" expected breeding values after \( t \) generations of selection. The last term, \( F_i \), is the reduction of the phenotypic mean due to inbreeding depression (curve 4) and would occur only if non-additive genetic variance is present. Since the expected value of any observed phenotypic mean is \( E(\bar{P}_i) = \mu + L_i \), the expected difference between any two different observed phenotypic line means \( \bar{P}_{i..} \) and \( \bar{P}_{i..'} \) would be

\[
E(\bar{P}_{i..} - \bar{P}_{i..'}) = (L_{i..} - L_{i..'}) = (G_i - G_{i'}) + (D_i - D_{i'}) + (F_i - F_{i'}).
\]

When one of the means is that of an unselected control, the difference is the response to selection or the genetic gain. The observed response is \( \Delta P = (\bar{P}_{i..} - \bar{P}_{i..'}) \) with expectation as in (3). Then the "possible" expected gain in breeding value is \( E(\Delta G) = (G_i - G_{i'}) + (D_i - D_{i'}) \), while the "predicted" gain in breeding value would be \( P(\Delta G) = (G_i - G_{i'}) \). To summarize,

\[
(\Delta P) = (\bar{P}_{i..} - \bar{P}_{i..'}) = P(\Delta G) + (D_i - D_{i'}) + (F_i - F_{i'}) = E(\Delta G) + (F_i - F_{i'}).
\]
The relationship between these values is illustrated in Figure 1.

In the selected population the values of $D$ and $F$ are cumulative over generations, and are inverse functions of the effective population size ($N$) but are directly increasing functions of the number of generations ($t$). Because the first selected parents are chosen in the zero generation, the expectation of $D$ is zero regardless of the size of $N$ for the progeny means of the first generation. But inbreeding depression could exist in the first progeny generation since the selected parents, some of which could be related, are mated together so that their breeding value can be estimated from the phenotypic mean of their progeny (Kojima 1961). In the unselected control population, the expected value of $D$ is zero regardless of the effective population size, but the amount of inbreeding and the inbreeding effect is a function of $N$ and $t$ just as in the selected populations.

An idea of the magnitude of $D$ and $F$ can be obtained from the difference between or the ratio of the predicted mean or gain and that observed. The size of the $F$ term can be estimated by comparing the means of the selected pure lines with their crosses.

**REGRESSION OF GENOTYPE ON PHENOTYPE**

Of primary interest in many selection experiments is the estimation of the regression of the breeding value of the trait on the phenotype of the selection criterion, the so-called "realized" regression, or heritability in the simple case. This has been estimated as the least-squares estimate of the regression of cumulative response on cumulative selection differential by Falconer (1960) and others. For four generations of selection the expectations of the plotted values (as deviations from control) would be

<table>
<thead>
<tr>
<th>Generation</th>
<th>$X$ = cum. sel. diff.</th>
<th>$Y$ = cum. response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$I_1$</td>
<td>$I_1\beta_1$</td>
</tr>
<tr>
<td>2</td>
<td>$I_1 + I_2$</td>
<td>$I_2\beta_1 + I_2\beta_2$</td>
</tr>
<tr>
<td>3</td>
<td>$I_1 + I_2 + I_3$</td>
<td>$I_3\beta_1 + I_3\beta_2 + I_3\beta_3$</td>
</tr>
<tr>
<td>4</td>
<td>$I_1 + I_2 + I_3 + I_4$</td>
<td>$I_4\beta_1 + I_4\beta_2 + I_4\beta_3 + I_4\beta_4$</td>
</tr>
</tbody>
</table>

and the least-squares estimate of the regression is

$$b_c = \frac{\sum xy}{\sum x^2} = \frac{(\frac{3}{4}I_1^2 + \frac{1}{2}I_1I_3 + \frac{1}{4}I_3I_4)\beta_2 + (\frac{1}{2}I_1I_3 + I_3^2 + \frac{1}{2}I_3I_4)\beta_3 + (\frac{1}{4}I_3I_4 + \frac{1}{2}I_3^2I_4 + \frac{3}{4}I_3^2)\beta_4}{\frac{3}{4}I_1^2 + I_1^2 + \frac{3}{4}I_3^2 + I_3I_4 + \frac{1}{2}I_3^2I_4 + I_3I_4}.$$  

If the selection differentials are all equal ($I_1 = I_3 = I_4$), then

$$b_c = \frac{3\beta_2 + 4\beta_3 + 3\beta_4}{10},$$

and only if the regressions are equal ($\beta_2 = \beta_3 = \beta_4$ or $D_2 = D_3 = D_4 = 0$), does the expectation of $b_c = \beta$. 


Cumulative data such as these also do not comply with the assumption of independent errors underlying least-squares regression since the errors due to drift are additive over generations. This can lead to serious underestimation of the standard error of the estimated regression coefficient (see Mandel 1957; Painter 1968) unless adjusted as proposed by Hill (1972a, b). This adjustment is still an approximation. Hill (1972a) considers \( b_t \) as well as the ratio of total response to the total selection differential at generation \( t \),

\[
b_t = \frac{I_1\beta_1 + I_2\beta_2 + \ldots + I_t\beta_t}{I_1 + I_2 + \ldots + I_t}
\]  

(4)
to be unbiased estimators of \( \beta_1 \). This would be correct in an infinite population with no selection limit where \( \beta \) would be constant, but if \( \beta \) is not constant over generations as is always the case in finite populations, they would be unbiased estimates of different things. Since the generation \( \beta \)'s are unequal due to increasing values of \( D \), the genetic interpretation of \( b_t \) is not clear, but \( b_t \) is interpretable as the weighted average over the generations of the estimated \( \beta \)'s, the weights being the selection differentials. So if one is interested in predicting gain to generation \( t \), \( b_t \) would be the preferred estimator, but unless \( t = 1 \), even this estimate will predict the result only for selection of the same intensities each generation and for populations having the same effective size. On the other hand, if an estimate of \( \beta_1 \) is desired, this should be based on only the first generation of selection, when the bias due to \( D \) is zero for the selected as well as for the control line.

**STATISTICAL ANALYSES**

As shown by Hill (1971), the variance of the gains due to genetic drift is not different if selection is conducted in a single population of effective size \( NK \) from what it would be if selection was practiced in \( K \) subpopulations of size \( N \). However, if \( t > 1 \), the expected response in the selected subgroups would be reduced because of the smaller \( N \). Experimental verification of this effect is evident in the results of Jones, Frankham and Barker (1968). In a selection experiment of size \( NK \), a large value \( K \) is preferred. Reducing the size of \( N \) is not critical because inferences can be made only to populations of the same effective size as those studied, unless \( t = 1 \) when \( E(D) = 0 \). The advantage of the increased \( K \) in a replicated experiment is that the experimental error can be better estimated from the data, rather than being based upon an approximation which may not take into account all sources of error.

It has been shown that interest should be centered on the responses to selection to any generation \( t \), in which case analysis of variance based on the phenotypic model in equation (2) is a powerful analytical tool. To illustrate, consider an analysis of up and down selection for \( t \) generations with separate controls with \( M \) observations per genetic group and \( K \) random replications of effective size \( N \) (Model A) or an analysis of a one-way selection experiment with a control (Model B) (see Table 1).
### Table 1

<table>
<thead>
<tr>
<th>Sources of variation</th>
<th>Model A Up and down selection and control</th>
<th>Model B One-way selection and control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>E.M.S.</td>
</tr>
<tr>
<td>Genetic groups (S)</td>
<td>3</td>
<td>$\sigma^2_w + M\sigma^2_{RL} + K\Phi_L$</td>
</tr>
<tr>
<td>Selection V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control (S)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Selection up V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control down (S)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Control up V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control down (S)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Replications (R)</td>
<td>$K-1$</td>
<td>irrelevant</td>
</tr>
<tr>
<td>$R \times S$</td>
<td>$3(K-1)$</td>
<td>$\sigma^2_w + M\sigma^2_{RL}$</td>
</tr>
<tr>
<td>Within</td>
<td>$4K(M-1)$</td>
<td>$\sigma^2_w$</td>
</tr>
</tbody>
</table>

If the response in breeding values is estimated in a crossbred population, no cross matings to eliminate the effect of inbreeding are necessary, because the expected average phenotypic value of the cross-progeny is equal to the average of the breeding values of the two parent pure lines. But if measured in a pure line in any generation, inbreeding effects must be removed if the true genetic gain is to be estimated. This is accomplished by mating the selected and control males and the selected and control females to an unrelated line as shown previously. The resulting four populations can be analyzed as in Model A by replacing up and down selection by male and female, in which case contrast $S$, would be an estimate of gain in breeding value unbiased by inbreeding effect. If there is no interest in the differences between sexes, the crosses could be summed over sex and compared as in Model B. If the objective is to estimate the significance of the difference between responses resulting from two different selection criteria or methods, analysis as in Model B just as for one-way selection, but without any controls, would be preferred.

No method has yet been devised to keep a finite control population from being affected to some extent by random drift. Therefore, if controls are needed to meet the objectives of the experiment, each population and replication selected for a given criterion should have its own control, so that the estimates of gains from different criteria will be independent and that drift in the controls can be measured. But it is not necessary that the control test-populations be as large as the selected populations or that they have the same effective size. This is because the expected reduction of the expected response due to any genetic limit or drift, $E(D)$, is zero in the unselected population. The resulting unequal subclasses can be accommodated in the ANOVA by use of an unweighted means or a proportional subclass analysis. Reducing the number of controls tested would increase the within-cell variance. But if the facilities made available by reducing the number of controls tested were available to increase the number of replications, the standard error of the gains observed could be reduced, especially if the variance due to drift is large.
The significance of the genetic differences in either model would be tested by the interaction term, which would include the variance due to genetic drift and any interactions of groups by replications. For Model A, if the up and down selected lines have separate controls, the contrast \( S_i \) is an unbiased estimate of asymmetry, while \( S_i \) is an unbiased estimate of \( \Delta P \) up + \( \Delta P \) down. Additional orthogonal contrasts of the gains from up and down selection would not be possible. If this is the desired objective, three independent experiments following Model B would be required.

The values \( P_{i,j,n} \) are observed only one time, at generation \( t \). They can be expressed in a generation model (\( g = 1 \ldots t \)) as deviations from \( P_o \), which is constant for all lines and replicates, in which case they represent the sum of the genetic gains to a terminal generation \( t \). The gain to the given generation \( t \), but for the \( j \)th replicate within the \( i \)th line, can be expressed in a model as,

\[
(P_i - P_o)_{i,t} = \Delta \hat{G}_{j,t} = \sum_{g=1}^{t} (\Delta \hat{G})_{j,t} + C_{j,t} + e_{j,t},
\]

\[
= \sum_{g=1}^{t} \{\Delta G + (\bar{C}_{j,t})_g\} + C_{j,t} + e_{j,t}.
\]  

Here (\( \Delta \hat{G}_{j,t} \)) \( g \) is the mean over generations within a single replicate at terminal generation \( t \). It is the estimator of the mean gain within any one replication. The term \( C_{j,t} \) is the accumulated drift variance within one replication and equals \( t \sigma_d^2 \).

It has an expected value of zero since it represents the deviations of the individual generation gains from the mean gain over generations but within replicates. The \( e_{j,t} \) term is the measurement error obtained in each replicate but only at terminal generation \( t \). Then (\( \bar{C}_{j,t} \)) \( g \) is the deviation of the mean gain over generations within replicates from the expected mean over replicates. (\( \bar{C}_{j,t} \)) \( g \) is a constant over \( g \), as shown in (6), that is created when the gain in breeding value estimated at generation \( t \) is divided by \( t \) to obtain the mean gain per generation. It will have a variance among the replicates, however, which is defined so that

\[
V(\bar{C}_{j,t})_g = E(\bar{C}_{j,t}^2)_g = \frac{1}{t^2} \sum_{g=1}^{t} E(C_{j,t}^2) = \sigma_d^2.
\]

When selection is conducted in a single replicate over \( t \) generations and the regression is estimated as \( (P_i - P_o)/t = (\Delta \hat{G}_{j,t})/t = \Delta \hat{G}_{j,t} \), it is seen from (5) that one is estimating \( \Delta \hat{G}_{j,t} \), and not the expected value of \( \Delta G \), since for a single value of \( t \), (\( \bar{C}_{j,t} \)) \( g \) is constant over \( g \), and has a variance only when \( t \) varies. On the other hand (\( \bar{C}_{j,t} \)) \( g \) will always vary between replicate means; but if the gain is calculated within replicates, this deviation is not included in the variance, and appears as a constant bias of the estimator as seen in (5) and (6).

From (5) it is seen that the variance at generation \( t \) within a single replicate would be

\[
V(\Delta \hat{G}_{j,t}) = E(C_{j,t}^2) + \frac{1}{M^2} \sum_{n=1}^{M} \sigma_e^2 = t \sigma_d^2 + \sigma_e^2/M,
\]
and the variance of the regression estimated over the $t$ generations in a single replicate would be

$$V\left(\frac{\Delta \hat{G}_{jt}}{t^2}\right) = \frac{1}{t^2} \left[ \frac{\sigma^2_d}{t} + \frac{\sigma^2_e}{Mt^2} \right].$$  (9)

If estimation is based on $j = 1 \ldots K$ replications to generation $t$, however, the variance could also be derived from (6) as

$$V\left(\frac{\Delta \hat{G}_{jt}}{Kt^2}\right) = \frac{1}{t^2} \frac{1}{K} \frac{1}{\bar{t}^2} \sum_{i=1}^{K} \left[ \frac{t}{\sum_{g=1}^{t} E(\hat{C}_{igt})} + E(C_{igt}) + \frac{1}{M} \sum_{w=1}^{K} E(e^W) \right],$$

$$= \frac{1}{t^2} \left[ \frac{\sigma^2_d}{K} + \frac{\sigma^2_d}{\bar{t}^2} + \frac{\sigma^2_e}{MKt^2} \right] + \frac{\sigma^2_e}{MKt^2}, \quad t > 1.$$  (10)

Since the second term results from variation of the individual generation regression estimates around the mean regression estimated within each replicate, this formula applies only when $t > 1$ because when $t = 1$, there are no drift deviations within replicates. The variance of the difference between the regression estimates for any two lines would be obtained by replacing $1/\bar{t}^2$ in (10) with $(1/\bar{t}_1^2) + (1/\bar{t}_2^2)$.

The interpretation of the variance components over generations derived from analysis of variance of $K$ replicates at one terminal generation $t$ can also be obtained from equation (6). We see from equation (8) that the within-replicate-line variance component would contain $\sigma^2_w = Mt\sigma^2_d + \sigma^2_e$. The interaction component would be interpreted to contain $t\sigma^2_d$ from the first term in (10), while the line component would contain $t^2\Delta G^2$ from the first term in (6). If the replicates were nested within lines, then the replicate component would contain the $t^2\sigma^2_d$ portion of the drift. But if no replicates were run or if their components were not taken into consideration, the $t^2\sigma^2_d$ term would be completely confounded with the line component.

If it is desired to determine the change in breeding values after some $t$ generations of selection in any line, it is clear that an increase in the number of replicates has the effect of increasing the precision of the estimates. Any increase in $t$ would result in increased variation around the true breeding value due to the accumulation of drift errors over generations. Increasing $t$, therefore, will not cause convergence to the expected value. To illustrate with an extreme case, consider a trait controlled by a single locus, having only additive effects, an initial value of $q = .5$, with no selection. A single replicate will deviate from $q = .5$ over generations, and at $t = \infty$, each replicate will have reached either $q = 1.0$ or $q = 0$. The expected value at $t = \infty$ would still be $q = .5$, but convergence to the expected value would only be attained by averaging over the $K$ replicates, since no replicates would have the expected $q$ value. Selection for an infinite number of generations would only change the average gene frequency by changing the proportion of replicate lines having $q = 1.0$ or $q = 0$. 
For estimation of the regression over replications it is seen from (10) that increasing \( t \) will reduce the variation only within replicates, while increasing \( K \) will decrease the variance both within and between replicates. The measurement error is reduced most by increasing \( t \), but the drift variance is reduced more rapidly by increasing \( K \). The effect on the total variance of increasing \( K \) versus increasing the number of generations, \( t \), depends upon the relative magnitudes of \( \sigma^2_d \) and \( \sigma^2_e/M \). As \( \sigma^2_d \) increases relative to \( \sigma^2_e/M \), then the benefit of increasing \( K \) relative to \( t \) also increases. The magnitude of \( \sigma^2_d \) is increased by an increase in \( h^2 \) or a decrease in the effective population size. The advantage of increasing the number of replicates rather than the number of generations exists even when the ratio of \( \sigma^2_d/\sigma^2_e \) is quite small. For instance, for the variance of the regression when \( K = 1, t = 10 \) to be the same as when \( K = 10, t = 1 \), \( \sigma^2_d \) needs to be only 1/10 as large as \( \sigma^2_e/M \). In the case of limited resources where only \( Kt \) generation-replicate estimates of the regression for a line can be made, the best estimate in terms of minimum variance will usually be obtained by increasing the number of replications and keeping \( Kt \) constant by reducing the number of generations.

**SINGLE-GENERATION SELECTION EXPERIMENTS**

Only the line mean values \((L_i)\) can be estimated from the analysis of variance, and these consist of three completely confounded effects as shown in equation (3). Experimental designs must be used which ensure that the line means reflect only the effects desired according to the objectives of the experiment.

If the objective of a selection experiment is to obtain unbiased estimates of population parameters in the base population or to test genetic theory based on infinite population sizes, this can only be done by single-generation experiments, because only in the first generation is the expected drift effect equal to zero. If pure lines are involved they should be outcrossed as shown previously, unless the objective of the experiment requires including the inbreeding depression effect. One-generation experiments would be necessary to obtain comparisons unbiased by drift effects, of the effectiveness of selection based on two criteria of selection—for example, individual phenotype and family average phenotype. It could also be used to compare the ability of selection on crossbred progeny performance, as in reciprocal recurrent selection, with selection on pure line performance to improve the breeding value of a line either when mated to a matched unrelated line or to its own line.

If no restrictions are placed on selection for the different criteria, the observed responses may differ due to differences in the predicted responses resulting from differences either in the predicted regressions or in the selection intensities as seen in equation (1). If the same proportions of the individuals and of the families are chosen as breeders for two different criteria, the section intensities would be equal. It is very difficult to keep the selection intensities equal in two different populations even when it is desired, and the selection intensities may be permitted to differ, which in itself could be an advantage of one method over another. Dividing the genetic response by the selection intensity \((i)\) would yield
\( \rho_{p \sigma} \), which is proportional to the efficiency of selection based on the trait itself, or \( \rho_{p \sigma} \). So the genetic efficiency relative to individual selection would be

\[
RE = \frac{\rho_{p \sigma}}{\rho_{p \sigma}} = \frac{\rho_{p \sigma} c}{h^2 c}. 
\]

One could also compare the regressions, \( \beta_{p \sigma} \), by dividing the responses by their selection differentials. For the individual selection case \( (T = C) \), this is the realized heritability, while for the case where the criterion of selection is different \( (T \neq C) \), this is the realized correlated response in actual units in \( T \) from selecting one standard deviation in \( C \), or \( h^2 c \beta_{p \sigma} \). The relative efficiency of two criteria of selection can be compared on these bases only if the trait being improved is the same for both criteria. However, ratios of the \( \rho_{p \sigma} c = \Delta \sigma_{T} / \sigma_{T} \) values provide relative efficiency of two different criteria in improving either the same or two different traits or for the same criterion to improve two different traits. Estimates of the standard errors for these values are obtainable from the analysis of variance.

The "total" value of a selection criterion in a single generation would include both its effect on genetic efficiency and the selection intensity that can be used. Thus, the comparative size of the gains themselves would be important for comparing the criteria of selection. One could examine the ratios of the observed gains to predicted gains and should find these ratios to be near unity for all criteria since the predicted, possible and observed gains for the additive genetic model should all be the same.

**MULTIGENERATION EXPERIMENTS**

These same estimates and comparisons could be made after \( t \) generations of selection but would clearly be different from the first generation because of the presence of \( (D_i - D_{i'}), \) and would be specific values for any given generation and for any pair of selection criteria, because the \( D \)'s, and therefore the gains, are functions of \( N \) and \( t \). If the objective of a selection experiment is to compare the response in a selected line to that in a control, the maximum responses will occur in the early generations, and the rate of response will decline through an increase in \( D \) as \( t \) increases due to the cumulative loss of desirable alleles resulting from genetic drift (Robertson 1960). Any predictive value of the response for \( t > 1 \) generations would apply only to another population of the same effective size. Furthermore, the pattern of the deviation of the possible from the predicted breeding value due to the finite \( N \) is established in the second generation.

If the objective of an experiment is to compare the response in the same trait \( T \) from two different selection criteria \( C_1 \) and \( C_2 \), the predicted gains and the genetic drift effects on \( T \) for the two criteria could be different, but the upper genetic limit would be the same. When the same trait is selected, the difference in predicted gains is completely determined by the criteria of selection chosen and the distance from the upper genetic limit. If the predicted gains are the same for \( C_1 \) and \( C_2 \), then the reduction in gain due to the upper genetic limit, \( D^2 \), would be the same for both criteria. As the two predicted gains diverge, the values of \( D^2 \) at any generation \( t \) would also diverge proportionally to the predicted
gains. Assuming infinite population size, the difference between the possible and predicted gains would increase over generations and the divergence in the population having the highest predicted gain would increase faster than that in the population with smaller predicted values, due to the approach toward the same upper limit. Conversely, the ratio of possible gains \( E(\Delta G) \) and predicted gains \( P(\Delta G) \) would be the smallest for the criterion having the largest predicted value and would decrease at a faster rate over generations than would the ratio for the criterion with the lower predicted gain. The curves for the means of the two populations over generations would never cross, but would meet at \( t=\infty \).

It often occurs that the selection criterion having the largest predicted response will have the smallest effective population size, as for example, when comparing family with individual selection. In this case, the means of the populations selected on the basis of criteria \( C_1 \) and \( C_2 \) will cross at some generation less than infinity. The greater the difference in effective population sizes, the earlier the generation when the two populations will have equal observed gains. This fact emphasizes the importance of effective population size in maximizing long-term genetic gains. Curves for the ratios of the observed to predicted values will not cross, but will diverge as \( t \) increases. If, on the other hand, the effective population size is smaller for the criterion having the smaller predicted gain, the curves for the population means or gains will not cross; but the curves of the ratios of the observed to expected values may cross if the difference in predicted gains is small.

When the objective is to compare the responses in two traits \( T_1 \) and \( T_2 \) from selection on a single criterion \( C \), different predicted gains and upper genetic limits would be involved, but the effective population size would be constant. Since the variance of the two traits may be different, it is necessary to standardize them before comparisons of the gains can be made. The genetic gains are then expressed as \( i \rho_{G,T} \). The genetic limits will determine the predicted gains since the criteria are otherwise the same; therefore the trait with the higher upper genetic limit will be expected to have the greater predicted response since it would have a higher heritability. As a result the response curves should not cross as they approach their respective limits.

The standardized response in trait \( T_1 \) from selection on the criterion of selection \( C_1 \), can be compared with that for response in trait \( T_2 \) from selection on criterion \( C_2 \) in single-generation experiments since they would differ only in the predicted responses. In multiple-generation experiments, differences in upper genetic limits and effective population sizes could exist as well as differences in predicted values. The expected values of the standardized responses in the traits \( T_1 \) and \( T_2 \) over generations could be parallel, could cross, meet or diverge under selection for \( C_1 \) and \( C_2 \) depending on the combination of upper limits, population sizes and predicted gains existing for the particular populations selected and the traits and criteria chosen. In this case the effects are confounded, and accurate prediction of genetic responses to selection are not possible without precise information on the distance the trait to be improved in the specific population is from the upper genetic limit, the effect of this distance on the predicted gains, and the
effect of finite population size on the responses to selection. Conversely, the causes of different responses to selection based on different criteria of selections over $t>1$ generations are not subject to accurate interpretation. Any results over $t>1$ generations would have predictive value only to a population the same distance from the upper genetic limit and for criteria having the same effective population size. The magnitude of the reduction in selection responses due to $D$ and $F$ after only a few generations is clearly seen in the results of Jones, Frankham and Barker (1968), of Kinney et al. (1970), and of Calhoon and Bohren (1974). Unconfounded experiments with only one variable can be designed. For example, the upper limits of selection can be estimated for two criteria of selection by keeping the effective population size equal for the two criteria. But even in such a case only the relative ranking would have predictive value, since the absolute values of the estimated upper limits would depend upon the effective population size chosen.

The unbiased information on genetic parameters for the zero generation and for testing genetic theory relevant to infinite populations with no upper genetic limit can only be obtained in single-generation selection experiments. But for practical evaluation of selection criteria or methods in a given population, their “total” effect, including differences in predicted values and the influence of different effective population sizes over $t$ generations, determines their expected value and their true efficiency for long-range improvement of the means. To compare two criteria or methods on the basis of their “total” effect, equalization should be on a fixed amount of capacity, which can be utilized for stock storage, testing or mating facilities. Within this limitation, genetic gains should be maximized for each criterion or method. The responses measured in cross progeny would be unbiased estimates of the expected gains to be obtained by each criterion in the given amount of capacity.

Even for evaluating the “total” effect of one or more criteria of selection for the same trait, only a few generations would be required, perhaps no more than two or three. This is especially true if information on effective population size is available. While precise prediction is not possible, the general nature of the trend of the expected values of the responses over generations is known. Therefore, little additional information would be obtained from additional generations. The number of generations in an experiment should then depend upon the desired objectives and the statistically most efficient method of achieving the objectives.

The ideas presented were developed primarily in discussions with Dr. T. B. Kinney while preparing the manuscript for Kinney et al. (1970) and numerous refinements resulted from discussions with Dr. V. A. Garwood, Dr. R. B. Harrington and Dr. V. L. Anderson.

LITERATURE CITED


Corresponding editor: R. C. Lewontin