

EFFICIENT DESIGN AND ANALYSIS OF SELECTION EXPERIMENTS*

WILLIAM M. MUIR, USA

SUMMARY

Problems associated with estimating response to selection by researchers examining 1. unidirectional selection, 2. bidirectional selection, or 3. alternative selection procedures were examined. Theoretical results showed that if genotype by environment interactions exist, and there is an inter-generation environmental trend, then the usual methods of analysis, i.e., by differences between either the experimental and control or between experimental lines, will result in biased estimates. Covariance analysis, with the control used as a covariate, was examined as a method of removing environmental trends when such interactions exist. The procedure was shown to give relatively unbiased estimates in this situation. Some bias results from errors associated with estimating the control's performance. Procedures to correct for this bias were discussed.

A statistic was also developed that measures the relative proportion of the environmental variance which a control can remove by covariance techniques. This statistic was used to experimentally compare alternative controls from data previously published by Bray *et al.* (1962). The statistic proved to be very useful, and, contrary to popular belief, identified an inbred as being consistently the best single control for more than one experimental population. The benefits of inbreds as controls were further discussed. From theoretical considerations and practical utility, an inbred or a single cross among inbreds may be the best universal control.

INTRODUCTION

Geneticists have recognized that in order to separate environmental trends from unidirectional selection response, a control population is necessary. However, maintaining a control is considered a waste of facilities when comparing alternative methods of selection or when estimating realized heritabilities by divergent selection (Hill, 1972c). In each of the three situations described, environmental trends are assumed to be eliminated by taking differences between either the experimental and control lines or between the experimentals. These procedures have become classic (Dickerson, 1969) but ignore one of the basic assumptions necessary to obtain unbiased results; it is assumed that the common environmental factor will have the same effect on all populations, i.e., no genotype by environment (GxE) interaction. From theoretical considerations, Hill (1972c) suggests that the degree of interaction will likely increase in proportion to the genetic difference between individuals from the same or different populations. If such is the case, then the basic assumption of no GxE is expected to become invalid as selection proceeds.

There are three basic questions which need to be addressed regarding the design and analysis of the types of experiments mentioned. 1. What statistical procedure is used to remove environmental trends in the presence of GxE interactions? 2. How are such interactions detected? And 3. How does

*Journal Paper 10452 of the Purdue Agricultural Experiment Station.

**Department of Animal Sciences, Purdue University, West Lafayette, IN 47907.

one choose and maintain a control population to remove these effects? Each of these questions will be addressed by the research presented in the paper.

STATISTICAL METHODS TO INCORPORATE A CONTROL

Theoretical considerations relative to statistical procedures are presented elsewhere (Muir, 1986). For the following discussion, the control population is assumed to be such that drift is negligible. This assumption may be satisfied if either the control's population size is large, or a highly inbred line is chosen as the control.

1. Unidirectional Selection with a Control Population

Assume a selection experiment in which the response (Y_{ijk}) is measured on 'n' individuals for 'a' generations in each of 'r' replicate lines, where the replicate lines are initially established by the subdivision of a base population before selection takes place. A control population is also maintained in the same environment as the experimentals with 'q' individuals measured in each of the 'a' generations (C_{jp}). Assuming that environmental effects are common to all replicate experimental lines and the control population, the following models may be used. For the experimental lines,

$$Y_{ijk} = \mu_1 + R_i + G_j + E_j + RG_{ij} + \varepsilon_{(ij)k} \quad (1)$$

where:

- μ_1 is the overall mean of the experimental populations
- R_i is the effect of the i^{th} replicate line (random), $i = 1, \dots, r$
- G_j is the effect of the j^{th} generation (fixed), $j = 1, \dots, a$
- E_j is the environmental effect in the j^{th} generation on the experimental population (random)
- RG_{ij} is the effect of the interaction of the i^{th} replicate with the j^{th} generation (random)
- $\varepsilon_{(ij)k}$ is the within error assumed NID ($0, \sigma_\varepsilon^2$), $k = 1, \dots, n$.

The generation effect (G_j) is actually the cumulative gains per generation,

$$G_j = \sum_{m=1}^j g_m, \text{ where } g_m \text{ is the gain in the } m^{\text{th}} \text{ generation. Drift is also}$$

cumulative, but within a replicate line. Thus, The RG_{ij} interaction is due to the cumulative drift of the i^{th} replicate line through the j^{th} generation, therefore, $RG_{ij} = \sum_{m=1}^j d_{im}$, where d_{im} is the drift effect which occurred in the m^{th} generation for the i^{th} replicate line. The expected mean squares and degrees of freedom for this model are given in Table 1. The drift effects (d_{im}) are assumed to be sampled from a population with homogeneous drift variances (σ_d^2) across generations. Since selection will change genetic and phenotypic variances to some extent, the expectations are only approximate.

For the control population, the model is,

$$C_{jp} = \mu_2 + E_j + \rho_{(j)p}$$

where:

- μ_2 is the overall mean of the control population
- E'_j is the environmental effect in the j^{th} generation on the control (random), $j = 1, \dots, a$
- $\rho_{(j)p}$ is random measurement error assumed NID $(0, \sigma_\rho^2)$, $p = 1, \dots, q$

Table 1. Expected mean squares for the model given by Equation 1

Source of variation	Degrees of freedom	Mean square	Expected means squares
Replication	$r-1$	MS(R)	$\sigma_\epsilon^2 + [n(a+1)(2a+1)/6] \sigma_d^2 + a n \sigma_R^2$
Generations	$a-1$	MS(G)	$\sigma_\epsilon^2 + [n(a+1)/6] \sigma_d^2 + r n \sigma_E^2 + r n \sigma_G^2$
RXG	$(r-1)(a-1)$	MS(RXG)	$\sigma_\epsilon^2 + [n(a+1)/6] \sigma_d^2$
Within	$ra(n-1)$	MS(E)	σ_ϵ^2

a. Analysis of Deviations From Control

The usual method of estimating genetic gain is to analyze the difference between the experimental values and the mean control value,

$$Y_{ijk}^* = Y_{ijk} - \bar{C}_j.$$

$$Y_{ijk}^* = \mu_1 - \mu_2 + R_i + E_j - E'_j + G_j + D_{ij} + \epsilon_{(ij)k} - \bar{\rho}_{(j)} \dots \quad (3)$$

This procedure has the advantage of simplicity and will remove environmental effects if the experimental and control populations respond the same to the environmental effect ($E_j = E'_j$). Because E and E' are a function of the genotype, this assumption may not be valid due to $G \times E$ interactions. In an extreme form of $G \times E$ interaction, the control may actually respond in the opposite direction to that of the experimental line.

The variance of the regression of the experimental values deviated from the control (Y_{ijk}^*) on any independent variable (X) is

$$\sigma_{\beta_{Y^*.X}}^2 = V(\sum_j Z_j (\bar{Y}_{.j} - \bar{C}_j) / \sum_j Z_j^2)$$

$$\sigma_{\beta_{Y^*.X}}^2 = \sigma_{\beta_{Y.X}}^2 + \frac{1}{\sum_j Z_j^2} (\sigma_E^2 + \sigma_{E'}^2 - 2 \sigma_{EE'} + \sigma_\rho^2/q), \quad (4)$$

where $Z_j = X_j - \bar{X}$,

$$\sigma_{\beta_{Y.X}}^2 = \frac{1}{rn \sum_{j=1}^a Z_j^2} (\sigma_\epsilon^2 + n c \sigma_d^2), \quad (5)$$

$$c = \frac{\sum_{j=1}^a (\sum_{j'=j}^a Z_j)^2}{\sum_{j=1}^a Z_j^2} \quad \text{and} \quad (6)$$

σ_{EE} , is the covariance between the environmental effect on the experimental and control populations. The independent variable 'X' is usually cumulative selection differential but generation number may be used if selection differentials do not vary greatly (Hill, 1972c). The variance presented by Equation 5 was previously given by Hill (1972b).

As seen from Equation 4, environmental effects are removed if

$$\sigma_E^2 + \sigma_{E'}^2 - 2 \sigma_{EE'} = 0. \quad (7)$$

Yamada (1962) has shown that the GxE interaction variance (σ_I^2) can be formulated as:

$$2\sigma_I^2 = \sigma_{b_1}^2 + \sigma_{b_2}^2 - 2\sigma_{b_{12}}, \quad (8)$$

where $\sigma_{b_1}^2$ and $\sigma_{b_2}^2$ are the among group, within environment, variance components for environment 1 and 2 respectively and $\sigma_{b_{12}}$ is the among group covariance component. Thus, Equality 7 is true only if $\sigma_I^2 = 0$. If $\sigma_{EE'}$ is less than $\sigma_E^2/2$ then $\sigma_{b_{12}}$ will be greater than $\sigma_{b_{12}}$, the unadjusted regression. Hill (1972c) recognized that GxE interactions would increase the variance of the differences and used this criteria to evaluate control population (Hill, 1972d). However, examination of Equation 5 shows that an increase in variance will not always result when GxE interaction is present. In addition, the estimates of genetic gain ($\beta_{G,X}$) will be biased since

$$\begin{aligned} E(\beta_{Y,X}^*) &= E\left(\frac{\sum_j Z_j (\bar{Y}_{.j} - \bar{C}_{j..})}{\sum_j Z_j^2}\right) \\ &= \frac{\sum_j Z_j G_j}{\sum_j Z_j^2} + E\left[\frac{\sum_j Z_j (E_j - E'_j)}{\sum_j Z_j^2}\right] \\ &= \beta_{G,X} + \text{bias}. \end{aligned} \quad (9)$$

If $E_j = E'_j$ for all j, then bias = 0 and $\sigma_I^2 = 0$.

b. Analysis by Covariance Techniques

A procedure not previously used in genetic studies is to adjust experimental values by using the control as a covariate, as follows:

$$Y'_{ijk} = Y_{ijk} - \beta_{Y,C} (\bar{C}_{j..} - \bar{C}_{...}), \quad (10)$$

where $\beta_{Y,C}$ is a regression parameter expressing the true relationship between the experimental and control populations, and

$$E[\beta_{Y,C} (\bar{C}_{j..} - \bar{C}_{...})] = E_j. \quad (11)$$

The adjusted values of the experimental line (Y''_{ijk}) are:

$$Y''_{ijk} = \mu + R_i + G_j + E_j - \beta_{Y.C} (E'_j - \bar{E}'.) - \beta_{Y.C} (\bar{\rho}_{j.} - \bar{\rho}_{..}) + D_{ij} + \varepsilon_{(ij)k} \quad (12)$$

The estimate of genetic gain $\beta_{G.X}$ is unbiased even if interaction is present because

$$\begin{aligned} E[\beta_{Y.C}] &= E\left[\frac{\sum_j Z_j (\bar{Y}''_{.j.})}{\sum_j Z_j^2}\right] \\ &= E\left[\frac{\sum_j Z_j (\mu + \bar{R}. + G_j + E_j - \beta_{Y.C} (E'_j - \bar{E}'.) - \beta_{Y.C} (\bar{\rho}_{j.} - \bar{\rho}_{..}) + \bar{D}_{.j} + \bar{\varepsilon}_{.j.})}{\sum_j Z_j^2}\right] \\ &= \beta_{G.X} \end{aligned}$$

and $E[\beta_{Y.C} (E'_j - \bar{E}'.)] = E_j$. Therefore, analysis by covariance techniques is an effective statistical procedure to use in the presence of GxE interactions and provides an answer to the first primary question addressed.

The expected mean square for generations, after adjustment,

$$\begin{aligned} E(\text{MS (Generation)}) &= \sigma_e^2 + [n(a+1)/6] \sigma_d^2 + rn [\sigma_E^2 - \beta_{Y.C} \sigma_{EE'}] \\ &\quad + rn \beta_{Y.C}^2 \sigma_\rho^2 / q + rn \theta_G^2 \end{aligned} \quad (13)$$

The relative proportion (RP) of environmental variation in the experimental line accounted for by the covariate is

$$\begin{aligned} \text{RP} &= \beta_{Y.C}^2 \sigma_{EE'} / \sigma_E^2 \\ &= r_{EE'}^2 \end{aligned} \quad (14)$$

which is always positive and is the square of the correlation between the environmental effects on the experimental and control populations. Note that RP presents an effective statistic for comparing various populations to use as a control and thereby provides a means to answer the third question addressed by this research. This result shows that it is not important that a control and experimental population respond with similar magnitude or direction to an environmental effect. What is important is that the environmental effects produce consistent responses from each population. The magnitude of correlation ($r_{EE'}$) will increase as the environmental effects become more consistent. The accuracy of prediction correspondingly increases, which allows for the effective removal of environmental effects. In this regard, an inbred population may serve as the most effective control, since their increased sensitivity to environmental effects allows for precise estimation of the environmental effect.

The estimate of genetic gain, adjusted for the control ($\beta_{Y.C}$), may be computed by several alternative procedures (see Muir, 1986, for details), the simplest of which is as a partial regression coefficient based on the following model:

$$\bar{Y}_{.j} = \beta_0 + \beta_{YX.C} X_j + \beta_{YC.X} \bar{C}_j + \bar{\omega}_{.j}. \quad (15)$$

This procedure has the advantage of providing information on the presence of GxE interactions. The estimate of $\beta_{YC.X}$ provides information on the relative response of the two lines to common intra-generation environmental factors (GxE interaction), and thereby addresses the final question. The partial regression coefficient $\beta_{YC.X}$ measures both direction and magnitude with which the experimental and control populations respond to common intra-generation environmental fluctuations. If $\beta_{YC.X} = 1$ then there is no interaction, and the fluctuations of the experimental line, on the average, match those of the control. If $\beta_{YC.X} > 1$, then the fluctuations of the experimental line are an amplification of the control's. An estimate of $0 < \beta_{YC.X} < 1$ indicates that the environment has a greater effect on the control than on the experimental population. Any value less than zero ($\beta_{YC.X} < 0$) is indicative of extreme interaction. That is, a common environmental fluctuation causes the experimental and control lines to respond in opposite directions. The residual variance from the model given by Equation 15 should not be used to construct tests. Instead, the replicate by effect interaction should be used for testing due to the correlation among generations (Muir, 1986).

c. Analysis When G x E Interaction Develops With Generation

By including an interaction term in the following model, the procedure can be expanded to test for and adjust for the effect of GxE interactions which develop as a result of selection:

$$\bar{Y}_{.j} = \beta_0 + \beta_1(X_j - \bar{X}_{.}) + \beta_2(\bar{C}_{j.} - \bar{C}_{..}) + \beta_3(X_j - \bar{X}_{.})(\bar{C}_{j.} - \bar{C}_{..}) + \bar{\omega}_{.j}. \quad (16)$$

where corresponding coefficients are as defined by the model given in Equation 15, and β_3 estimates the magnitude of interactions which develop with generations. This procedure will remove that part of selection response which is due to GxE interaction.

2. Experiments Involving Divergent Selection or Comparison of Alternative Selection Procedures

As previously mentioned, a control is usually not used in these situations. However, if GxE interactions are expected as populations become genetically distant, then divergent selection is most likely to induce such an effect. Analysis by differences will either overestimate or underestimate the true divergence if there is an environmental trend depending on the direction of the interaction.

When comparing alternative selection procedures, response to selection is compared for traits selected in the same direction but based on different criteria. Because the experimental populations under selection remain phenotypically similar, the conclusion might be made that GxE interactions are not expected in this situation. However, Griffing (1954) found that heterosis may be exhibited by a specific cross in a specific environment. He concluded

that changes in genetic parameters may occur in varying degrees in different environments. Bray *et al.* (1962) also found evidence of conditioning of the expression of dominance effects by the environment. Thus, if one of the methods of selection tends to take advantage of dominance effects, such as reciprocal recurrent selection (RRS), while another may not, such as within line selection (WLS), then a GxE interaction may be induced through the differential dominance effects exhibited by the experimental populations. Rich and Bell (1980) demonstrated this result in two experimental lines of *Tribolium* selected for 70 generations by either RRS or WLS. In such a situation, the comparison of response between selection procedures will either overestimate or underestimate the difference if there is an environmental trend. A control population is thus needed to (1) verify the presence or absence of a GxE interaction, (2) measure the environmental trend, and (3) adjust the experimental population for any such trend. The statistical procedures to perform these tasks are accomplished as an extension of the technique already described.

Assume two replicated experimental populations are each selected for the same trait but either in the opposite direction or based on different criteria. Response (Y_{ijkl}) is measured on the l^{th} individual in the k^{th} generation of the j^{th} replicate of the i^{th} experimental line. A control population is also maintained in the same environment as the experimentals. The measure of the p^{th} individual of this control population ($p = 1, \dots, q$) in the k^{th} generation ($k = 1, \dots, a$) is represented by C_{kp} . Estimates of effects are found as partial regression coefficients in the following model:

$$\begin{aligned}
 Y_{ijkl} = & \beta_0 + \beta_1 L_i + \beta_2 (X_k - \bar{X}) + \beta_3 (\bar{C}_{k.} - \bar{C}_{..}) \\
 & + \beta_4 L_i (X_k - \bar{X}) + \beta_5 L_i (\bar{C}_{k.} - \bar{C}_{..}) \\
 & + \beta_6 L_i (X_k - \bar{X}) (\bar{C}_{k.} - \bar{C}_{..}) + \omega_{ijkl}
 \end{aligned} \tag{17}$$

where:

- L_i = dummy variable for the effect of the i^{th} line ($\sum_i L_i = 0$)
- X_k = generation number or cumulative selection differential
- $\bar{C}_{k.}$ = mean value of the control in the k^{th} generation
- ω_{ijkl} = residual error.

Other effects are defined as interactions with the main effects. The β 's are the partial regression coefficients.

The residual error is not to be used for tests of hypothesis. Those tests must be computed separately for each effect due to the correlated error structure of response. Tests of hypothesis in such cases are made by comparing the overall regression to the variation among such regression coefficients computed within each replicate. The same test is accomplished in the analysis of variance (ANOVA) by comparing the mean square due to that effect with the mean square for the effect's interaction with replicates. An increase in power may be possible by constructing the proper error term from variance components which are more accurately estimated by other sources of variance in the ANOVA (Muir, 1986).

Interpretation of the partial regression coefficients is as follows: $\tilde{\beta}_1$ estimates the average difference between lines and $\tilde{\beta}_2$ the average regression on generations or cumulative selection differential. The test of $H_0: \beta_2 = 0$ tests for asymmetry of response in divergently selected lines. $\tilde{\beta}_3$ measures the average GxE interaction between the experimental and control lines over generations. $\tilde{\beta}_4$ estimates the rate of divergence between experimental lines after adjustment for GxE interactions. The test of $H_0: \beta_4 = 0$ tests for the equality of alternative selection procedures. The $\tilde{\beta}_5$ effect determines if the relationship between the experimentals and control is the same for both experimental populations, and $\tilde{\beta}_6$ estimates the amount of interaction which develops during the selection process. Thus, with this model, the relationship of the experimentals with the controls may be the same or different and may change with generations independently for each experimental line.

3. Assumptions

The expected mean squares in Table 1 and Equation 5 for the variance of response assume that within population variance (σ_e^2) and the increment in drift variance (σ_d^2) remain constant. Tests of hypothesis assume that the phenotypic values are normally distributed. Since selection may change gene frequencies and is likely to change variances, these results are only approximate. Hill (1972a) also acknowledges these assumptions and further discusses their validity. He concludes the model assumptions hold more closely when heritabilities are low and the experiment is of short duration.

Unfortunately, the use of a control as a covariate violates the assumption that the independent variable should be measured without error. Two potential sources of error may be associated with the control, that of drift and measurement. The assumption regarding drift may be satisfied if the control is highly inbred or is maintained as a large population. The desired model for regression of experimental means on control means is:

$$Y = \beta_0 + \beta_{Y.C} C + \omega, \quad (18)$$

where C is the unknown true value of the control. The observed measure of the control is $C' = C + \theta$, where θ is measurement error. If the assumption is made that C is NID (μ_C, σ_C^2) and θ is NID ($0, \sigma_\theta^2$), then Y and C' follow a bivariate normal distribution. The linear regression of Y on C' is $\beta_{Y.C'} = \beta_{Y.C} / (1 + \lambda)$ where $\lambda = \sigma_\theta^2 / \sigma_C^2$ (Snedecor and Cochran, 1967). If the mean value of the control is used in the model given by Equation 18, then

$$\sigma_\theta^2 = \sigma_p^2 / q \quad (19)$$

and

$$\sigma_C^2 = \sigma_{C'}^2 - \sigma_\theta^2. \quad (20)$$

An unbiased estimate of the regression of experimental on control means is

$$\beta_{Y.C} = \beta_{Y.C'} [1 + (\tilde{\sigma}_\theta^2 / (\tilde{\sigma}_{C'}^2 - \tilde{\sigma}_\theta^2))]. \quad (21)$$

The true regression ($\beta_{Y.C}$) is therefore underestimated by $\tilde{\beta}_{Y.C'}$, but as the number of samples of the control increases, the bias decreases.

The influence of measurement error on the estimated response to selection, corrected for the control, can be observed from the following equation which gives an unbiased estimate of the partial regression coefficient from simple regression coefficients.

$$\beta_{YX.C} = \frac{\tilde{\beta}_{Y.X} - \tilde{\beta}_{Y.C'} (1+\lambda) \beta_{C'.X}}{1 - \tilde{\beta}_{X.C'} (1+\lambda) \beta_{C'.X}} \quad (22)$$

Thus, if λ is ignored, $\beta_{YX.C}$ will either be overestimated or underestimated, depending on the sign of $\tilde{\beta}_{Y.C'}$ and $\tilde{\beta}_{C'.X}$. The bias induced by this effect is usually minimal (Muir, 1986).

Corrections for errors due to drift can be made by similar techniques if replicated control populations are maintained so that drift variance can be estimated. Because the theory given by Snedecor and Cochran (1967), assumes that the errors are independently distributed, while drift effects are cumulative, the necessary correction factor for drift variance is not obvious. However, some idea as to the effect of drift on the estimates may be obtained by letting the drift error of each generation take on the average value expected for drift, i.e., assume the drift errors are NID(0, $(a+1)\sigma_d^2/2$). Equation 19 then becomes

$$\sigma_{\emptyset}^2 = \sigma_p^2/q + (a+1)\sigma_d^2/2 \quad (23)$$

EXPERIMENTAL EVALUATION OF CONTROL POPULATIONS AND EXAMPLES

Hill (1972d) presented an extensive review of control populations, their importance and errors associated when using a control. He compared variance about regression of performance on generations using both deviations and unadjusted results. The conclusion was drawn that if the control successfully removes variability in response between generations, then the two strains react in the same way to the changing environment. As noted earlier, it is possible for a GxE interaction to exist even though the variance of differences is reduced. When covariance procedures are used to incorporate controls, the presence of GxE interactions is not important. The relevant statistic is the RP value presented in Equation 14. With this new procedure, a reexamination of controls is appropriate.

1. Experimental Comparison of Controls

Since the work of Bray *et al.* (1962) with Tribolium castaneum is the only major experiment undertaken in any species to compare the efficiency of alternative types of control populations, the comparisons will be limited to their results. The experiment included eight generations of selection plus the initial estimates at generation 0. Relative humidity was alternated between high and low every other generation to impose a known environmental effect. In each of two replicates, two experimental lines were divergently selected for large (L) and small (S) pupal weight and maintained along with fifteen types of controls. The controls included two masters (C_1 and C_2), two inbreds (I_1 and I_2), and its cross (I_{12}), five controls with different mating structures limited to fifty mating pairs (M, R_1^D, R, R, R), a stabilized selection line (Z), a repeat mating for large (TL) and small (TS) lines, and finally, a relaxed line formed at the 4th generation for the large (XL) and small (XS) lines.

The variation for each experimental line was first partitioned by usual analysis of variance techniques into the sources given by Table 1. Variation due to between generation differences was further partitioned into the linear effect of generation (G_j) and effects due to the covariate (C) by the use of the model given in Equation 15. The mean value of the control was used as the covariate in each case except for the repeat matings. In that case, the covariate was the difference between the mean of the repeat mated line in the j^{th} generation and the corresponding experimental line in the $j-1^{\text{th}}$ generation.

The variance components σ_E^2 and $\beta_{Y,C} \sigma_{EE'}$, were estimated by setting the mean squares equal to their expectation and solving. The term $r\beta_{Y,C}^2 \sigma_p^2 / qn$, due to measurement error associated with the control, was assumed to be trivial in comparison with the other variance components and ignored. The residual mean square for generations, reduced by the variation due to the covariate and the linear effect of generation, was set equal to the expectation:

$$E[\text{MS (gen. residual}_1)] = \sigma_e^2/n + (a+1)\sigma_d^2/6 + r(\sigma_E^2 - \beta_{Y,C} \sigma_{EE'}). \quad (24)$$

After adding back the sums of squares due to the covariate to the residual sums of squares due to generations, the mean square was set equal to the expectation:

$$E[\text{MS (gen. residual}_2)] = \sigma_e^2/n + (a+1)\sigma_d^2/6 + r\sigma_E^2. \quad (25)$$

Thus, $\sigma_E^2 = [\text{MS (gen. residual}_2) - \text{MS (RXG)}]/r$ and $\tilde{\beta}_{Y,C} \sigma_{EE'} = [\text{MS (gen. residual}_1) - \text{MS (gen. residual}_2)]/r$. The results for each control are shown in Table 2.

The only control with an RP greater than 90 percent for both experimental lines was I_1 , an inbred line. The relaxed lines, corresponding to each experimental (XL and XS), were slightly better than I_1 , when used only for the experimental line from which it was derived. Surprisingly, the repeat mated control was only effective for the small line, not the large. Examination of the partial regression coefficients show several extreme types of GxE interaction. The sign of the interaction with each experimental line was opposite each other for all controls except C_1 , IL, RS, R and I_2 . In addition, when the opposite experimental line was used as the covariate, the partials were $-.78$ and $-.49$ for large and small, respectively. These results support Hill's (1972c) hypothesis that the degree of interaction will increase in proportion to the genetic difference between individuals from the same population or different populations. Also, note that few of the partials were close to 1.0, the situation where analysis by differences will give unbiased results. This conclusion is particularly true with respect to divergent selection since the reaction of the two lines to environmental trends is in the opposite direction for each line.

The results with respect to inbred controls are contrary to previous beliefs. Objections to inbreds as controls have been made by Gowe *et al.* (1959), King *et al.* (1959) and Hill (1972d). Bray *et al.* (1962) discusses these objections and concludes that the "major difficulty is that inbreds and inbred crosses represent at best only a narrow range of genotypes so that their response to environmental shifts is likely to be specific to the genotypes which type represent." They also conclude that "individual inbred lines cannot be truly representative of any original outbred population from which

Table 2. Experimental comparison of control populations

	Experimental Line			
	Small (S)		Large (L)	
	RP*	$\beta_{YC.X}$	RP*	$\beta_{YC.X}$
<u>Control</u>				
C ₁	28.9	1.10	-17.5	.60
C ₂	97.6	1.14	44.2	-.36
TL	-20.2	.01	-16.8	.30
XL	78.0	-1.49	98.4	1.32
TS	93.1	.60	78.2	-.30
XS	103.0	1.11	74.0	-.78
Z	97.1	1.31	70.5	-.21
M	88.9	2.04	91.3	-.84
R _o	100.1	1.53	40.2	-.61
R _s	49.3	.75	80.6	.93
R _n	33.2	1.31	62.2	.08
R _{sn}	88.4	1.76	-17.8	-.34
I ₁	99.2	-1.38	97.2	1.01
I ₂	85.4	-1.56	83.1	-.82
I _{1,2}	101.9	-.91	86.5	.40

*RP = percent of the environmental variation removed from the experimental line by the control.

selected line may be drawn and should not be expected to be good indicators of how environmental shifts affect other genotypes." The results presented in this paper clearly refute these objections if the proper statistical procedure is used. Bray *et al.* further conclude that inbred lines should give highly repeatable results with respect to how environmental changes affect their own genotype. The results presented here support this conclusion and it is, in fact, this property of inbreds which make them excellent controls. Dickerson (1969) suggests using crosses of homozygous lines to provide a means of producing genetically constant populations in successive generations. He notes some potential problems of using inbreds as controls in farm animals since only partially inbred lines are available and there may be problems related to fitness.

Inbreds, on the other hand, have several advantages. The maintenance of large random-bred controls is very expensive in terms of facilities. In contrast, inbred controls should cost much less to maintain as only a few mating pairs are needed. Another advantage of inbreds is that samples of the control could be distributed without problems of sampling (drift). Questions have been raised as to the suitability of random-bred controls over long periods of time due to accumulated drift.

This research suggests a different approach to choosing a control population. Rather than trying to maintain a population which is genetically close to the experimental, such as repeat mated or relaxed line, use a genetically constant line, such as an inbred, and use statistical procedures to correct for interactions which develop with time. Since all inbreds and their crosses are not equally effective as controls, as seen by comparing RP of I_1 with I_2 and its cross, the best procedure may be to produce several inbred lines and first eliminate those lines which are difficult to maintain due to reproduction problems. The remaining lines would be compared with several experimental lines as to their RP values. The inbred or cross with the consistently best RP value would be chosen as the permanent control.

2. Estimation of Response with Unidirectional Selection

As previously shown, the estimate of $\beta_{YX.C}$ is adjusted for inter-generation trends regardless of GxE interactions. This result can also be demonstrated by expressing $\beta_{YX.C}$ as an adjustment of the simple regression coefficient $\beta_{Y.X}$

$$\beta_{YX.C} = \beta_{Y.X} - \beta_{YC.X} \beta_{C.X} \quad (26)$$

The estimate of $\beta_{C.X}$ measures the effect of the inter-generation environmental trend on the control. Therefore, the simple regression $\beta_{Y.X}$ is adjusted by the product of the functional relationship between the experimental and control populations ($\beta_{YC.X}$) and the measured change in the environment ($\beta_{C.X}$) as it influences the control. For the case of extreme interaction, where the experimental and control populations respond in the opposite direction to common environmental effects, $\beta_{YC.X}$ would be negative, and the adjustment to the experimental population would be in the opposite direction to that which the environmental trend influences the control.

Data from the previous example for the experimental S line and the inbred control I_1 will be used to illustrate the procedures. The estimated genetic gain, ignoring environmental effects, is $\tilde{\beta}_{Y.X} = -10.12$. The same regression for control values shows a long-term environmental trend of $\tilde{\beta}_{C.X} = 2.35$. Thus, analyzing deviations from control gives a value of $\tilde{\beta}_{Y.X}^* = -12.48$. However, the functional relationship between the experimental and the control population is $\tilde{\beta}_{YC.X} = -1.38$. An extreme form of GxE interaction is thereby indicated, in which the control responds in a reverse phase to that of the experimental population due to common intra-generation environmental fluctuations. The estimate of response should therefore be adjusted in the reverse direction. The adjusted estimate of genetic gain can be found directly from fitting the model given by Equation 15 or from Equation 26 as

$$\begin{aligned} \tilde{\beta}_{YX.C} &= -10.12 - (-1.38)(2.38) \\ &= -6.87 \end{aligned}$$

3. Bidirectional Selection

If the assumption were made that environmental effects cancel each other with bidirectional selection and no control was maintained, the data would either be analyzed by taking differences between the two experimental lines and regressing differences on generations or by analyzing the data by the

model given in Equation 17 with only β_0 , β_1 , β_2 , and β_4 in the model. The two analyses would give the same result for the rate of divergence between the lines. As an example, the data from Bray *et al.* (1962) was used, with the inbred I₁ used as the control since it give the highest RP for both lines. The data was then fit to the model given in Equation 17. The results are seen in Table 3.

Table 3. Analysis of bidirectional selection by the model given in Equation 17, with and without a control.

Parameter estimated	Without a control	With a control*
β_0	219.1	219.1
β_1	40.8	40.2
β_2	-.03	.42
β_3	—	-.18
β_4	10.09	7.38
β_5	—	1.14
β_6	—	.04

*Inbred I₁ used as a control.

Without a control, the rate of divergence is 10.09 $\mu\text{g/gen}$, while with a control the rate is reduced to 7.38 $\mu\text{g/gen}$. The reason for the reduction is that the experimental lines were adjusted in opposite directions. As can be seen, without a control, realized heritability would have been overestimated by approximately 18%. If the environmental trend had been in the opposite direction, the estimate would have been underestimated.

The estimate of β_2 indicates that there was a small amount of assymetry in the direction of the large line. The small value of $\tilde{\beta}_6$ shows that the increase in the interaction variance with generations is small. The large value of $\tilde{\beta}_5$ was expected from previous examination of the data.

REFERENCES

- Bray, D.E., Bell, A.E. and King, S.C. (1962). The importance of genotype by environment interaction with reference to control populations. *Genet. Res. Camb.* 3, 282-302.
- Dickerson, G.E. (1969). Techniques for researchers in quantitative annual genetics. In *Techniques and Procedures in Animal Production Research*. Amer. Soc. Anim. Sci., New York.
- Gowe, R.S., Robertson, Alan and Latter, B.D.H. (1959). Environment and poultry breeding problems. 5. The design of poultry control strains. *Poult. Sci.* 38, 462-471.

- Griffing, B. (1954). Contributions of genotypic and environmental effects to earliness of tomato fruiting. Statistics and Mathematics in Biology. Chapter 41, 523-536. Iowa State College Press: Kempthorne, Bancroft, Gowen and Lush.
- Hill, W.G. (1972a). Estimation of realized heritabilities from selection experiments. I. Divergent selection. *Biometrics* 28, 747-765.
- Hill, W.G. (1972b). Estimation of realized heritabilities from selection experiments. II. Selection in one direction. *Biometrics* 28, 767-780.
- Hill, W.G. (1972c). Estimation of Genetic Change. I. General theory and design of control populations. *Animal Breeding Abstracts* 40, 1-15.
- Hill, W.G. (1972d). Estimation of Genetic Change. II. Experimental evaluation of control populations. *Animal Breeding Abstracts* 40, 193-213.
- King, S.C., Carson, J.R. and Doolittle, D.P. (1959). The Connecticut and Cornell randombred populations of chickens. *World's Poultr. Sci. Jour.* 15, 139-159.
- Muir, W.M. (1986). Estimation of response to selection and utilization of control populations for additional information and accuracy. *Biometrics* (In press).
- Rich, S.S. and Bell, A.E. (1980). Genotype-environment interaction in long term selected populations of *Tribolium*. *Jour. Heredity* 71, 319-322.
- Snedecor, G.W. and Cochran, W.G. (1967). Statistical Methods. University Press, Ames, Iowa.
- Yamada, Y. (1962). Genotype by environment interaction and genetic correlation of the same trait under different environments. *Jap. Jour. Genet.* 37, 498-509.