Theor Appl Genet (1995) 91:553–556

S. E. Aggrey · C. Y. Lin · K. M. Cheng

Size of breeding populations required for selection programs

Received: 2 April 1994 / Accepted: 17 January 1995

Abstract The minimum population size required for selection in order to reduce the effect of genetic drift to a particular level has been considered. The model of Nicholas was extended to include the measurement-error variance in the response variance. Situations where the sex ratios among scored and breeding individuals are unequal are also considered. When the duration of a selection experiment is relatively long, Nicholas' approximation (i.e., assuming that measurement error is negligible relative to drift) is useful in determining the minimum effective population size required. However, the measurement-error variance becomes an important source of variation in short-term (≤ 5 generations) selection experiments, and should not be ignored.

Key words Coefficient of variation of response · Genetic drift · Selection · Effective-population size · Short- and long-term

Introduction

Response to selection for a single trait depends on five factors: genetic variability, generation interval, intensity of selection, effective population size and accuracy of selection. Of these several factors, population size has the widest range of consequences. In the short term, it influences the selection differential, inbreeding depression, and the reduction in genetic variance due to genetic drift. In the long term, it affects selection limits and the utilization of new variation arising from mutation (Robertson 1960; Hill 1986). In small populations, genetic drift is an important source of variation among selected lines, causing variation in within-line additive genetic variance (Avery and Hill 1977) as well as variation in response (Bohren 1975; Hill 1980, 1986).

The variance of the estimate of genetic change in a selection population comprises of drift and measurement-error variances. Genetic drift is the random change in gene frequency due to the sampling of gametes (Wright 1949). Measurement-error variance is due to the sampling of a limited number of individuals for testing, which depends on variation within families about their genetic mean, and on common environmental effects (Hill 1978, 1980). The drift term accumulates across generations, whilst the measurement term does not. Measurement-error variance cannot be completely ignored as a component of response variance except in long-term selection experiments (Hill 1972a). Measurement-error variance also becomes increasingly important as the number of individuals recorded per family, and the heritability of the trait under selection, decrease.

Nicholas (1980) considered the minimum effective population size required for selection experiments using the coefficient of variation of response as a criterion. To obtain a simple formula, he assumed that measurement-error variance is negligible as compared to drift variance. In the present study, Nicholas' model was extended to include measurement error. The implications in both short-term and long-term selection programs are discussed. The case of unequal sex numbers in both the scored and selected groups were also incorporated directly into the model to examine the effect on the minimum effective population size required for selection experiments.

Model

Consider a selection line derived from a large base population and undergoing artifical directional selection with discrete generations for a trait determined only by additive genes and having additive genetic variance σ_a^2, phenotypic variance σ_p^2, and heritability h^2. Consider that M males and F females ($F = \alpha M$) are scored every generation

---

Communicated by D. Van Vleck

S. E. Aggrey · K. M. Cheng (✉)
Avian Genetics Laboratory, Department of Animal Science, University of British Columbia, 248-2357 Main Mall Vancouver, B. C., V6T 1Z4, Canada

C. Y. Lin
Centre for Food and Animal Research, Agriculture and Agri-food Canada, Ottawa, Ontario, K1A 0C6, Canada

($M + F = T$ is constant) and that, from these, m males and f females ($f = \beta m$) are selected. The proportions selected in each sex are $p_m = m/M$ and $p_f = f/F$, giving rise to a selection intensity i_m and i_f respectively. For such a population, the usual approximation for effective population size, N_e (the number of individuals in an indealised population that would give the same drift), is given by $N_e = 4mf/(m + f)$. The drift variance in a population with two sexes over t generations, σ_d^2, is approximated by Hill (1972) as

$$\sigma_d^2 = \frac{t\sigma_a^2}{N_e}. \tag{1}$$

The measurement-error variance of mean response to selection is given by Hill (1980) as

$$\sigma_e^2 = \sigma_p^2 \left[\frac{c^2}{f} + \frac{(1 - c^2 - h^2/2}{T} \right],$$

where c^2 is the environmental correlation between full sibs. This expression can be written as

$$\sigma_e^2 = \frac{4\sigma_p^2}{N_e(1+\beta)} \left[c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2) \right]. \tag{2}$$

For $\alpha = \beta = 1$, and $p_m = p_f = p$, then $\sigma_e^2 = \sigma_p^2/N_e[2c^2 + p(1 - c^2 - h^2/2)]$ (Hill 1980).

Criteria for analysis

The coefficient of variation of response, $CV(R_t)$, is the ratio of the standard deviation of the response to its expectation, $CV(R_t) = SD(R_t)/E(R_t)$. This parameter can be used as a criterion for determining the size of a breeding population required for a selection program, and to compare the relative results achieved from various selection schemes.

Case 1: selection with a control population

The same symbols as in Nicholas' model will be used whenever applicable. If m males and f females are selected as parents from M males and F females scored respectively in the selection line, then drift variance in the selection line is given by equation (1) and measurement-errror variance is given by equation (2). Similarly, over t generations, drift variance and measurement-error variance in a contemporaneous control line can be written as $t\sigma_a^2/J_e$ and $[4\sigma_p/J_e(1 + \beta')][c^2 + \beta p'_m/(1 + \alpha'))(1 - c^2 - h^2/2)]$ respectively, where α' and β' are the sex ratios of the scored and selected groups of the control line respectively, p'_m is the proportion of males selected in the control line, and J_e is the effective population size of the control line.

Let $\bar{X}_t$ and $\bar{C}_t$ be the phenotypic means of the selection and control lines, respectively, at generation t. Response to t generations of selection is estimated as

$$\hat{R} = \bar{X}_t - \bar{C}_t. \tag{3}$$

Assuming no covariance between the selection line and its control, the variance of the estimate of response, $\sigma_{\hat{R}}^2$, is given by

$$\sigma_{\hat{R}}^2 = \sigma_{dx}^2 + \sigma_{dc}^2 + \sigma_{ex}^2 + \sigma_{ec}^2, \tag{4}$$

where σ_{dx}^2 and σ_{dc}^2 are drift variances of the selection and control lines respectively, and σ_{ex}^2 and σ_{ec}^2 are the measurement-error variances in the selection and control lines respectively. If $\alpha = \alpha'$, $\beta = \beta'$ and $p_m = p'_m$, then substituting equations (1) and (2) and appropriate drift and measurement-error variances of the control line into equation (4) gives

$$\sigma_{\hat{R}}^2 = t\sigma_a^2 \left(\frac{1}{N_e} + \frac{1}{J_e} \right) + \frac{4\sigma_p^2}{1+\beta} \left[c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2) \right] \times \left(\frac{1}{N_e} + \frac{1}{J_e} \right). \tag{5}$$

If selection and control lines are the same size in terms of the numbers of males and females selected as parents, then equation (5) reduces to

$$\sigma_{\hat{R}}^2 = \frac{2\sigma_a^2}{N_e} \left\{ t + \frac{4}{(1+\beta)h^2} \left[c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2) \right] \right\}. \tag{6}$$

Disregarding any linkage disequilirium and changes in gene frequency which affect linearity of response then, in mass selection experiments of t generations, the expected response to selection, $E(R_t)$, is given by

$$E(R_t) = tih\sigma_a, \tag{7}$$

where i is the average selection intensity. From equations (6) and (7), the coefficient of variation of response is,

$$CV(R_t) = \frac{\sqrt{\frac{2\sigma_a^2}{N_e} \left[t + \frac{4}{(1+\beta)h^2} \left(c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2 \right) \right]}}{tih\sigma_a}. \tag{8}$$

After rearrangement, equation (8) can be written as

$$CV(R_t) = \frac{\sqrt{2}}{ih\sqrt{N_e t}} \sqrt{1 + \left[\frac{4}{t(1+\beta)h^2} \left(c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2) \right) \right]}. \tag{9}$$

If $\alpha = \beta = 1$ and $p_m = p_f = p$, equation (9) reduces to $\sqrt{2}$ times the equation (13) of Hill (1980). The difference in the factor of $\sqrt{2}$ is because Hill considered σ_d^2 and σ_e^2 of the selected line only. For $\alpha = 1$, $t = 5$ and $c^2 = 0.1$, the corresponding N_e required under different combinations of $CV(R_t)$, h^2, β and i values were arithmetically determined, and are shown in Table 1.

Table 1 Effects of heritability, h^2, sex ratio of breeding individuals, β, and selection intensity, i, on the effective population size, N_e required for two values of coefficient of variation of response, $CV(R_t)$ (when $\alpha = 1$, $t = 5$ and $c^2 = 0.1$)[a]

$CV(R_t)$	h^2	β	i	N_e required
5%	0.25	1	1.400	399
	0.25	1	1.159	582
	0.25	5	1.400	377
	0.25	5	1.159	551
	0.40	1	1.400	232
	0.40	1	1.159	338
	0.40	5	1.400	206
	0.40	5	1.159	300
20%	0.25	1	1.400	25
	0.25	1	1.159	36
	0.25	5	1.400	24
	0.25	5	1.159	34
	0.40	1	1.400	14
	0.40	1	1.159	21
	0.40	5	1.400	13
	0.40	5	1.159	19

[a] α = sex ratio of scored individuals; t = number of generations; and c^2 = environmental correlation between full sibs

Case 2: selection without a control population

When the objective is to compare alternative selection schemes for improving the same trait, selection programs are often conducted without a control line. However, the importance of variance in common environmental effects among geneations must be considered. It is customary to assume that effects common to individuals within a generation are randomly and independently distributed with mean zero and variance σ_c^2, which becomes a part of error variance in the selected line. Consequently, measurement-error variance of the mean response to selection is

$$\sigma_e^2 = \frac{4\sigma_p^2}{N_e(1+\beta)}\left[c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2)\right] + \sigma_c^2. \quad (10)$$

In the absence of a control line, the observed response is

$$\hat{R} = \bar{X}_t - \bar{X}_0 \quad (11)$$

with expected variance

$$\sigma_{\hat{R}}^2 = \sigma_{dx}^2 + 2\sigma_{ex}^2. \quad (12)$$

Substituting equations (1) and (10) into (12) leads to

$$\sigma_{\hat{R}}^2 = \frac{\sigma_a^2}{N_e}\left[t + \frac{8}{(1+\beta)h^2}\left(c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2)\right)\right] + 2\sigma_c^2. \quad (13)$$

Unlike drift variance, σ_c^2 does not accumulate, and may not be a major source of error in the long term, but certainly is of concern in short-term experiments. The magnitude of σ_c^2 cannot be predicted *a priori*. However, it can be estimated from selection experiments (Hill 1972b).

Case 3: divergent selection

Divergent selection has been used as an alternative to running a selection line with a control. One line is selected for a high value (H), and another contemporaneous line is selected for a low value (L), of the trait. Let $\bar{X}_H$ and $\bar{X}_L$ represent the phenotypic means for the high- and low-selection lines, respectively, after t generations of selection. Response to divergent selection is

$$\hat{R} = \bar{X}_H - \bar{X}_L \quad (14)$$

and the variance of response is

$$\sigma_{\hat{R}}^2 = 2\sigma_{dx}^2 + 2\sigma_{ex}^2. \quad (15)$$

If symmetry of response in divergent selection can be assumed for a short-term experiment:

$$E(R) = 2tih\sigma_a. \quad (16)$$

Then, the coefficient of variation of response in case 3 is

$$CV(R_t) = \frac{1}{ih\sqrt{2N_e t}}\sqrt{1 + \left[\frac{4}{t(1+\beta)h^2}\left(c^2 + \frac{\beta p_m}{1+\alpha}(1 - c^2 - h^2/2)\right)\right]} \quad (17)$$

which is one-half of the coefficient of variation for selection with a control, as noted by Nicholas (1980).

Discussion

Assumptions for derivation of expressions for $CV(R_t)$ are that the number of individuals scored and the proportion selected are the same for both sexes. This situation is rare in practical animal breeding although common with laboratory experiments. In deriving equation (9), the number of parents in both selection and control lines was assumed to be the same, but $CV(R_t)$ in a more general case could be evaluated by substituting the most general formulae for σ_d^2 and σ_e^2 directly into equation (4).

Effective population size can be expressed as $N_e = 4\beta m/(1+\beta)$ (Nicholas, 1980); then, the required number of breeding males in the selected line is $m = N_e(1+\beta)/4\beta$ and the required number of females is $f = \beta m$. To determine the required size of a selection experiment the maximum value for $CV(R_t)$ must be set. For example, $CV(R_t)$ can be set to be no greater than 20% (i.e., expected selection response five times as great as its standard deviation). If a specific case is considered where $p_m = 10\%$, $\alpha = 1$, $\beta = 3$, $h^2 = 0.25$, $p_f = 30\%$ and $c^2 = 0.1$, and all of these terms are incorporated into equation (9), it turns out that effective population sizes of 43 for three generations and 24 for five generations are required. If the measurement-error variance was ignored, e.g., using Nicholas' (1980) approximation, the effective population sizes required would have been determined to be 34 and 20 for three and five generations, respectively. In this case, Nicholas' approximations are 20% and 17% less than the more accurate figure, for three generations and five generations, respectively. The decision to incorporate the measurement-error variance into the model, or to use Nicholas' (1980) approximation, should be based on the number of generations. As (t) increases, measurement error becomes insignificant as compared to drift variance. Under such conditions, Nicholas' (1980) approximation is useful. However, not all selection experiments span large numbers of generations. In selection experiments where t is small, the contribution of the variance of measurement error cannot be ignored. Although there is no fixed definition for short or long term, up to five generations might be taken as appropriate for short term. The degree to which Nicholas' approximation underestimates the required population size for short-term selection programs is proportional to the value of c^2. This underestimation is greater when the number of individuals recorded in each family is small and h^2 is small (Hill 1972a). Under such conditions, the more complete equations, rather than Nicholas' simple approximation, should be used. When t is small, variation in α and β also influence the required effective population size.

As shown in Table 1, the N_e required for a selection program is much larger when $CV(R_t)$ is set at 5% than it is at 20%. The required value of $CV(R_t)$ is more important in deciding the size of the breeding population required than is h^2, β or i. Equations (9) and (17) show that, as with Nicholas' approximation, $CV(R_t)$ is inversely related to selection intensity (i), heritability (h^2), effective population size (N_e), and generation number (t). For a pre-defined value of $CV(R_t)$, the number of breeding animals required is smaller for high selection intensity, a highly heritable trait, or long-term selection, than for low selection intensity, a lowly heritable trait or short-term selection. These relationships are reflected in Table 1 which was derived from equation (9).

In the present study the effect of selection on drift variance was not accounted for (i.e., N_e was assumed to be the same in every generation). Drift variance is reduced during selection (Meuwissen 1991). Therefore, the effective population size determined from equations (9) and (17) would over-predict the drift variance, which leads to conservative estimates for required population sizes.

Directional selection acts on genetic variance directly first by changing gene frequencies and secondly by inducing linkage disequilibrium which also causes changes in genetic variance (Bulmer 1980). Selection also acts on genetic variance indirectly by changing the family structure and therefore enhancing loss of variation through the inbreeding effect of selection (Robertson 1961). Heritability of a trait was also assumed to be known *a priori* without error, and i, the standardised selection differential, was assumed not to change over generations. However, with a finite population size, selection differentials do change over generations (Hill 1985).

In the short term, genetic parameters estimated in the base population were assumed not to change. This assumption does not hold true for small populations (Aggrey 1994) and long-term selection experiments. Linkage disequilibrium, inbreeding depression, and changing family structure should all be accounted for in a more complete model. Nevertheless, the model used in this paper can provide useful indications of how large selection programs need to be.

Acknowledgements This paper was submitted by S. E. Aggrey as part of his thesis for the partial fulfillment of the requirements for a PhD degree at the University of British Columbia, Vancouver. We thank Prof. W. G. Hill and Dr. F. W. Nicholas for their constructive comments and suggestions.

References

Aggrey, SE (1994) Studies on genetic drift and inbreeding in small populations under artificial selection. PhD thesis, University of British Columbia, Vancouver

Avery PJ, Hill WG (1977) Variability in genetic parameters among small populations. Genet Res 29:193–213

Bohren BB (1975) Designing arfitical experiments for specific objectives. Genetics 80:205–220

Hill WG (1972 a) Estimation of genetic change. I. General theory and design of control populations. Anim Breed Abstr 40:1–15

Hill WG (1972 b) Estimation of realized heritabilities from selection experiments. II. Selection in one direction. Biometrics 28:767–780

Hill WG (1978) Design of selection experiments for comparing alternative testing regimes. Heredity 41:371–376

Hill WG (1980) Design of quantitative genetic selection experiments. In: Robertson A (ed) Selection experiments in laboratory and domestic animals. Proc Symp, Harrogate, UK, pp 1–13

Hill WG (1985) Effects of population size on rates of variability of response to short- or long-term selection. Z Tierzuchtungsbiol 102:161–173

Hill WG (1986) Population size and design of breeding programmes. In: Proc 3rd World Cong on Genetics Applied to Livestock Production, Lincoln, Nebraska, vol 12, pp 245–256

Meuwissen THE (1991) Expectation and variance of genetic gain in open and closed nucleus and progeny testing schemes. Anim Prod 53:133–141

Nicholas FW (1980) Size of population required for artificial selection. Genet Res 35:85–105

Robertson A (1960) A theory of limits in artificial selection. Proc Roy Soc Lond B153:234–249

Robertson A (1961) Inbreeding in selection programmes. Genet Res 2:189–194

Wright S (1949) Population structure in evolution. Proc Am Phil Soc 93:471–478