

Family Losses Following Truncation Selection in Populations of Half-sib Families¹⁾

By J. H. ROBERDS, G. NAMKOONG and H. KANG²⁾

(Received July 1979 / June 1980)

Summary

Family losses during truncation selection may be sizable in populations of half-sib families. Substantial losses may occur even in populations containing little or no variation among families. Heavier losses will occur, however, under conditions of high heritability where there is considerable family variation. Standard deviations and therefore variances of family loss, however, change little over the range of heritability. Increasing the intensity of selection causes the mean loss to increase but has little effect on the standard deviation of loss. As the number of families in the base population rises, both the mean loss and variance of loss increase, but the proportion of families lost remains almost constant. Some families will be lost in almost all selection operations, even those undertaken under conditions of low expected family losses, since variance of family loss is at moderate levels over a wide range of conditions.

Under conditions in which family losses are likely to be high, losses can be reduced by increasing the number of individuals selected.

Key words: Heritability, mean loss, loss variance, selection proportion, loss probability

Zusammenfassung

In Populationen von Halbgeschwister-Familien können, bei Selektionsmethoden, bei denen bei einem festen Schwellenwert abgeschnitten wird, die Verluste von Familien beträchtlich sein. Sogar in Populationen mit kleiner oder gar keiner Variation zwischen Familienmitteln können reale Verluste vorkommen. Größere Verluste wird es allerdings unter Bedingungen mit hoher Heritabilität, wo beträchtliche Familienvariabilität zu finden ist, geben. Die Standardabweichungen und daher auch die Varianzen des Familienverlustes ändern sich wenig über den Heritabilitätsbereich.

Eine Steigerung der Selektionsintensität verursacht eine Erhöhung des mittleren Verlustes, hat aber eine geringe Auswirkung auf die Standardabweichung des Verlustes. Mit der Zunahme der Anzahl der Familien in der Basispopulation steigt sowohl der mittlere Verlust als auch die Varianz des Verlustes, aber die Proportion an verloren gegangenen Familien bleibt fast konstant. Einige Familien gehen in nahezu allen Selektionsverfahren verloren, sogar in denjenigen die unter Bedingungen unternommen werden, bei denen erwartet wird, daß Familienverluste gering sind. Der Grund hierfür ist, daß die Varianz des Familienverlustes über einen weiten Bereich von Bedingungen auf mäßigem Niveaus ist. Unter Bedingungen, bei denen Familienverluste wahrscheinlich hoch sein werden, können Verluste durch Erhöhung der Anzahl selektierter Individuen vermindert werden.

Introduction

Simple recurrent selection almost inevitably excludes the progeny of some parents from breeding populations.

¹⁾ Joint contribution of the Department of Genetics, North Carolina Agricultural Experiment Station, Raleigh, North Carolina, and the Southeastern Forest Experiment Station, Paper No. 5873 of the Journal Series.

²⁾ Geneticist and Research Geneticist, Forest Service, USDA and North Carolina State University, Raleigh, North Carolina and Research Geneticist, Forest Service, USDA, Rhinelander, Wisconsin.

Any unique genes possessed by these parents are lost to breeding in future generations. Since alleles of potential future value are not easily recognized in the early cycles of forest tree selection, loss of a large number of families can severely limit breeding progress in future generations. This problem is particularly acute if the number of families in the breeding populations is small. Evaluation of breeding systems should therefore include consideration of expected number of families lost, variance in the number of families lost, and chances of losing specific numbers of families. We report here results for losses of half-sib families in one cycle of truncation selection of a simple recurrent selection system. We consider only selection on individual performance, ignoring information on family affiliation and family performance. This type of selection is simple to practice and is commonly used in tree breeding programs. Using this selection method, probabilities of family loss and means and variances of family loss were determined for two intensities of selection and for values of heritability extending from zero to one.

The Model

We consider populations made up of unrelated half-sib families; each family member has one common parent and one parent not shared by any other member of the family. The parents of any one family are assumed to be unique to that family and are unrelated to the parents of other families. Each such family is considered to be countably infinite in size.

If x represents the value of a trait under selection, we assume that x is normally distributed within each family and that the within family variance (σ^2) for x is constant for all families. We primarily consider spatial patterns for the family means in which the means are symmetrically spread at equal intervals about the population mean. The value of x at the point of truncation (x_t) for a given selection proportion will remain constant over repetitions of identical selection processes because the populations considered are countably infinite in size.

The variable x may be expressed as:

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad i = 1, 2, \dots, f \quad j = 1, 2, \dots, \infty$$

where μ is the population mean, α_i is the deviation due to the i^{th} half-sib family, ε_{ij} is the effect for the j^{th} individual in the i^{th} family, and f is the number of families. The α_i

have mean equal to zero and variance equal to $\frac{\sum_{i=1}^f \alpha_i^2}{f}$

and the ε_{ij} are normally distributed with means and variance of zero and σ^2 . For our model

$$\frac{\sum_{i=1}^f \alpha_i^2}{f} = \frac{\sigma_T^2 h^2}{4}, \quad \text{the } \alpha_i = \sigma_T \frac{(f - 2i + 1)}{4} \sqrt{3h^2/(f^2 - 1)}$$

and $\sigma^2 = \sigma_T^2 (4 - h^2)/4$,

where h^2 is the narrow sense heritability and σ_T^2 is the total variance among the x_{ij} .

Application of the model. This model can be used for analysis of family losses resulting from individual-tree selection in half-sib family performance tests when family information is ignored. For example, during the second stage of selection to produce seedling seed orchards (NAMKOONG 1978), some families arising from selections made

during the initial stage will be lost unless special measures are undertaken to ensure the survival of each family. The model describes selection in infinite or large populations in which a finite number of individuals is selected at random from the portion of the population with $x_{ij} \geq x_t$. It is thus appropriate for selection involving traits in which improvement through breeding is not sought beyond a certain level as, for example, is true for stem straightness in some breeding programs (SHELBOURNE, 1970). The model gives approximate results for selection programs involving large populations from which Nb individuals with the greatest values for x are selected, where N is the population size and b is the proportion of the population selected. The population size must, however, be large enough so that there is little variation in x_t over repetitions of selection if the model is to be adequate for analysis of these selection procedures.

Probability of family losses. Let p_i represent the probability that an individual chosen during truncation selection is from the i^{th} family.

Each p_i is given by

$$P_i = \frac{1}{fb} \int_{x_t}^{\infty} g_i(x) dx$$

where $g_i(x)$ represents the $N(\mu + \alpha_i, \sigma^2)$ probability density function, x_t is the truncation point for selection, μ is the population mean, the α_i are family effects, σ^2 is the within family variance, f is the number of families, and b is the proportion of the population selected. The proportion selected can be shown to be

$$b = \frac{1}{f} \sum_{i=1}^f \int_{x_t}^{\infty} g_i(x) dx.$$

Since α_i and σ^2 can be expressed in terms of σ^2_T and h^2 , values of x_t that satisfy this equation for specified values of b , h^2 and f can be determined by numerical integration. With these determinations for x_t , values of p_i can be calculated and used to compute probabilities for family losses.

Consider the probability distribution for the vector-valued random variable (n_1, n_2, \dots, n_f) where n_i is the number selected from the i^{th} family and $\sum_{i=1}^f n_i = n$.

Since the families are countably infinite in size, the probability function for (n_1, n_2, \dots, n_f) is multinomial and can be written as

$$\Pr(n_1, n_2, \dots, n_f) = n! \prod_{i=1}^f p_i^{n_i} / n_i!$$

The random variable (n_1, n_2, \dots, n_f) is an ordered partition (HALL, 1967) of n individuals into f families. When families are lost — do not have a member chosen during selection — one or more of the $n_i = 0$. Let the random variable m represent the number of families lost during selection.

Then m is defined on the sample space of the $\binom{n+f-1}{f}$ ordered partitions of n individuals into f families, and each

of the $\binom{f}{f-m} \binom{n-1}{f-m-1}$ ordered partitions of n individuals into $f-m$ families map into the same value of m . Thus, the probability of losing m families is the sum of the probabilities for obtaining each of these ordered partitions. Hence, the probability function for m can be written

$$P(m) = \begin{cases} n! \sum_{i=1}^f \prod_{i=1}^f p_i^{n_i} / n_i! & m \geq f-n \\ 0 & m < f-n \end{cases}$$

where the summation extends over all $\binom{f}{f-m} \binom{n-1}{f-m-1}$ ordered partitions of n individuals into $f-m$ families and $m = 0, 1, \dots, f-1$.

Populations without variance among families. Selection in populations in which the variance among families is zero — $h^2 = 0$ is of special interest. In these populations, each family has an equal chance of contributing individuals to the selected group. Each family then also has an equal chance of being eliminated from the selected portion. Conversely, in populations with $h^2 > 0$, some families always have higher probabilities of being rejected from the selected portion than others. This unequal distribution of selection probabilities causes the mean family loss to be higher in populations with $h^2 > 0$ than in those with $h^2 = 0$. Therefore, mean family loss at $h^2 = 0$ serves as a lower bound for mean loss in populations with $h^2 > 0$.

In populations with $h^2 = 0$, $p_1 = p_2 = p_3 = \dots = p_f = f^{-1}$, the probability function for m can be expressed as

$$P(m) = \begin{cases} \binom{f}{m} f^{-n} \sum_{k=0}^f (-1)^k \binom{f-m}{k} (f-m-k)^n & m \geq f-n \\ 0 & m < f-n \end{cases}$$

which is shown by FELLER (1962) to be the probability of m empty cells occurring when n objects are randomly distributed into f cells. The mean and variance of m are found to be:

$$E(m) = f(1-f^{-1})^n$$

$$\text{Var}(m) = f\{(f-1)(1-2f^{-1})^n + (1-f^{-2})^n - f(1-f^{-1})^{2n}\}.$$

The expected proportion of families lost is seen to be $E(m)/f$, which is equal to $(1-f^{-1})^n$. If the number of individuals selected is proportional to the number of families — then $n = kf$ — the expected proportion of families lost approaches $e^{-n/f}$ as the number of families increases:

$$\lim_{f \rightarrow \infty} (1-f^{-1})^n = e^{-n/f}.$$

Computational Procedures

Probabilities for family losses were calculated for heritabilities ranging from zero to one by increments of 0.05. For cases where heritability is greater than zero, it was necessary to calculate these probabilities numerically. As the number of families and the number of individuals selected increases, the number of ordered partitions that must be considered in computing probabilities for medium range losses increases drastically. Difficulty of computing the probabilities is therefore increased, and as a result we were unable to compute probabilities for populations with more than a moderate number of families. Loss probabilities were determined for populations of 5 and 10 families for selection proportions of 0.1 and 0.01, but for populations of 12 families calculations were done only for the 0.1 selection proportion for heritabilities of 0, 0.1, 0.25, and 1.0. Main emphasis was placed on selection procedures in which the number of individuals selected equals the number of families in the populations, and this condition is assumed throughout the discussion of results unless it is otherwise stated. However, probabilities were additionally calculated for selection procedures involving populations of 5 families in which the number of individuals selected equals twice the number of families. Mean family loss and the standard deviation for the numbers of families lost were computed using the probabilities calculated.

Although the results we obtained for populations with $h^2 > 0$ were for small numbers of families, they are descriptive of family losses in subdivided populations containing large numbers of families. If breeding populations are divided into small subpopulations as suggested by NAMKOONG (1976), the distribution of family losses among the subpopulations is of interest. Mean loss and standard deviation of loss in this context describe how family losses are distributed among the subpopulations.

For cases involving $h^2 = 0$, computation of family loss probabilities is not difficult and means and variances of family loss can be obtained directly from the formulas introduced in the previous section. Results therefore can be

readily computed for selection procedures involving large numbers of families. Since mean family loss is at a minimum when $h^2 = 0$, family loss means were calculated at $h^2 = 0$ and can be used as lower bounds for means in populations with $h^2 > 0$. Family loss variances at $h^2 = 0$ were calculated and can be used as estimates for the loss variances in populations with $h^2 > 0$. We determined probabilities of family losses, mean family loss and variance of family loss for populations containing numbers of families ranging from 5 to 50. Results were obtained for selection procedures in which the number of selected individuals equals both the number of families and twice the number of families.

Results

A large number of half-sib families can be unintentionally eliminated from future breeding consideration following a cycle of truncation selection. Approximately one-third of the families is expected to be lost when the number of individuals selected equals the number of families, even though the families have identical means ($h^2 = 0$) (Table 1). The expected loss is greater when there is genetic variation among the families ($h^2 > 0$). In the populations we studied, approximately one-half of the families is expected to be lost when $h^2 = 0.5$ and the selection rate is $b = 0.01$. If selection takes place under these conditions in a population having 10 families, the probability of losing 6 or more families equals 0.25 (Table 2).

Family loss is subject to sampling variation among repetitions of selection. The level of loss experienced in a particular selection experiment may differ considerably from that expected. For example, in a population of 10 families having $h^2 = 0.5$ from which 10 individuals are selected using a selection proportion of $b = 0.01$, the mean loss is 4.3. However, the probability of losing 6 or more families is 0.11, and the probability of losing 2 or fewer families is smaller but still a significant 0.04 (Table 2).

Effect of heritability. As might be expected, the mean family loss rises as heritability, and thus variation among family means, increases. In populations of 5 and 10 families the expected family loss increases curvilinearly with heritability (Figure 1). Concomitant with the change in mean loss as heritability increases is the increase in probability

of loss of a large number of families and the decrease in the probability of loss of small numbers of families. This result is demonstrated in Table 2 for populations of 10 families.

Change in heritability has little effect on the standard deviation of family loss under the conditions studied. That the magnitude of variation among these standard deviations is small across the range of heritability is illustrated for populations of 5 and 10 families in Figure 1. These results suggest that the standard deviations for populations with

Table 2. — Family loss data for populations of 10 families.

	$h^2 = 0$		$h^2 = .25$		$h^2 = .5$	
	$b > 0$	$b = .1$	$b = .01$	$b = .1$	$b = .01$	
Expected family loss	3.5	3.9	4.2	4.2	4.8	
Standard deviation of family loss	0.9964	1.0298	1.0440	1.0383	1.0246	
Probability that loss ≤ 2 families	.15	.09	.04	.05	.01	
Probability that loss ≥ 6 families	.02	.05	.11	.10	.25	

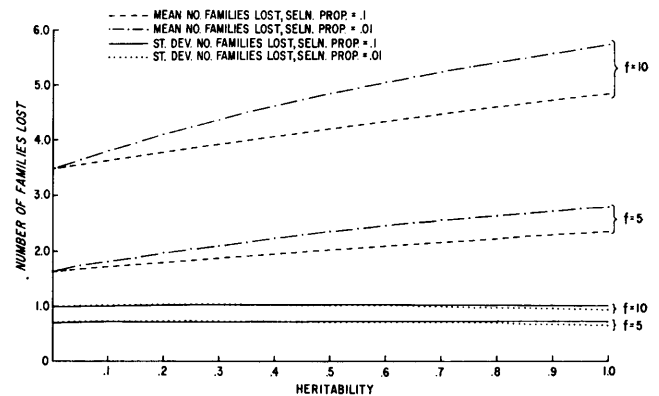


Figure 1. — Heritability effect on the mean and standard deviation of number of families lost at two selection proportions in populations of five and ten families with equally spaced means.

Table 1. — Expected values and variance for number of families lost when there is no variance among families ($h^2 = 0$).

Number of Families (f)	Number of Individuals Selected Equals			Twice the Number of Families		
	Expected Family Loss	Expected Proportion of Families Lost	Variance of Family Loss	Expected Family Loss	Expected Proportion of Families Lost	Variance of Family Loss
5	1.6	0.328	0.5092	0.54	0.107	0.3696
8	2.7	.344	.7989	0.94	.118	.6137
10	3.5	.349	.9928	1.2	.122	.7753
12	4.2	.352	1.1869	1.5	.124	.9366
15	5.3	.355	1.4782	1.9	.126	1.1784
20	7.2	.359	1.9640	2.6	.129	1.5808
25	9.0	.360	2.4499	3.2	.130	1.9831
30	10.9	.362	2.9358	3.9	.131	2.3852
35	12.7	.363	3.4218	4.6	.131	2.7873
40	14.5	.363	3.9078	5.3	.132	3.1894
45	16.4	.364	4.3938	6.0	.132	3.5914
50	18.2	.364	4.8798	6.6	.133	3.9934
∞		.368			.135	

$h^2 > 0$, can be estimated without large bias by computing standard deviations for $h^2 = 0$.

Number of families. As the number of families in the base population increases, the expected family loss increases but the expected proportion of families lost remains fairly stable. In populations with $h^2 = 0$, this expected proportion varies from 32.8 % to 36.4 % as the number of families increases from 5 to 50 and approaches 36.8 % ($100 e^{-1}$) as the number of families becomes large (Table 1). Similar results are obtained when there is genetic variation among families. At $h^2 = 0.25$ and a b of 0.1, the expected proportion of families lost is 36.6 % for 5 family populations and increases to 38.8 % for populations of 12 families.

The standard deviation of family loss increases as the number of families in the base population increases. However, the increase in standard deviation per unit family increment (hereafter referred to as $\Delta s/\Delta f$) is much less than one and declines as the number of families increases. The near parallelism of the standard deviation curves in Figure 1 indicates that $\Delta s/\Delta f$ is nearly uniform over changes in heritability and selection intensity when a small number of families is involved. If it is assumed that this uniformity also holds for large numbers of families, the extent to which $\Delta s/\Delta f$ decreases with increasing family numbers can be determined by studying populations with $h^2 = 0$. In these populations, as the number of families increases from 45 to 50, $\Delta s/\Delta f$ is less than half the $\Delta s/\Delta f$ value obtained as families are increased from 5 to 10, $\Delta s/\Delta f = 0.023$ and 0.057 respectively. These results indicated that although the standard deviation increases as the number of families in the base population rises, the rate of increase on a per family basis is small and becomes smaller as the number of families increases.

Selection intensity effect. The effect of selection intensity on mean family loss and variation in family loss was studied in populations of 5 and 10 families. An increase in the severity of selection results in an increase in the mean family loss when $h^2 > 0$ (Figure 1). Increasing the selection pressure tends to increase the probability that progeny are selected from families with a large proportion of high values for the trait under selection. At the same time, it diminishes the probability that individuals are chosen from families with a large proportion of low values. The intensity of selection has little effect on the standard deviation of family loss (Figure 1).

Nonequally spaced means. Several models with non-equally spaced family means were investigated to determine if they produced family loss data substantially different from that for the model with equally spaced means. These models involved patterns in which either extreme means or central means are farther apart or closer together than all other means. Numerical comparisons in 5-family

populations were used to evaluate differences among the models. Values for expected family loss and standard deviation of family loss were similar for the two type models except under conditions of high heritability, large departures from equal spacing, and intense selection ($b = 0.01$).

Alternative Selection Strategies

Under conditions of high heritability where family losses may be high, breeders may opt for alternative selection strategies to reduce losses. One method to reduce losses is to place an upper bound on the number of individuals that may be selected from each family. An extreme form of such a strategy is to have each family equally represented in the selected population, thus reducing selection to a within-family basis.

The proportion of families preserved can also be increased by increasing the number of individuals selected. This increase can be interpreted as a decrease in selection intensity. However, if the selection process involves taking a random sample from that part of the base population at or above the truncation point, the increase can be viewed as elevating the sample size rather than as incrementing the proportion selected. If, on the other hand, selection involves choosing only the best individuals from a large finite base population, the increased number of selected individuals can be viewed as coming from a base population, enlarged in the same proportion as the number of selected individuals. Thus, in this case, the intensity of selection also remains constant.

We studied the effects of such an increase by comparing results when the number selected is twice the family number to results obtained when the number selected equals the number of families. When $h^2 = 0$ the expected family loss for 5-family populations when $n = 2f$ is 32.8 % of the loss expected when $n = f$. For 50-family populations this percentage increases to 36.4% (Table 1). Results for 5-family populations under selection with $b = 0.1$ vary from 32.8% for $h^2 = 0$ to 69.4% for $h^2 = 1$. Values for $h^2 = 0.25$ and 0.5 are 54.1% and 55%, respectively. It is clear from these data that increasing the census of the selected population in this way results in a marked decrease in the expected family loss.

References

- FELLER, W.: An introduction to probability theory and its applications. Volume 1, 2nd Edition. John Wiley & Sons, Inc. New York, N.Y. (1962). — HALL, M.: Combinatorial theory. Blaisdell Publishing Co., Waltham, Mass. (1967). — NAMKOONG, G.: Introduction to quantitative genetics in forestry. Technical Bulletin No. 1588. USDA Forest Service, 342 p. (1979). — NAMKOONG, G.: A multiple-index selection strategy. *Silvae Genetica* 25: 199–201 (1976). — SHELBOURNE, C. J. A.: Breeding for stem straightness in conifers. *Second World Consultation on Forest Tree Breeding*. 1: 293–302 (1970).