

Prediction of long-term contributions and inbreeding in populations undergoing mass selection

J. A. WOOLLIAMS^{1,*}, N. R. WRAY² AND R. THOMPSON¹

¹AFRC Roslin Institute (Edinburgh), Roslin, Midlothian, EH25 9PS, UK

²Livestock Improvement Unit, Victoria Department of Agriculture, P.O. Box 500, East Melbourne, Victoria 3002, Australia

(Received 9 October 1992 and in revised form 26 February 1993)

Summary

For a population undergoing mass selection, derived from an unselected base population in generation zero, the expected long-term contribution to the population of an ancestor from generation 1 was shown to be equal to that expected during random selection multiplied by $1 + i(1 - c)^{-1}A'_i$ (where A'_i is one half the breeding value of the ancestor for the trait under selection standardized by the phenotypic standard deviation, i the intensity of selection, and $c = \frac{1}{2}(1 - kh_2^2)$ is the competitiveness which is defined by h_2^2 the heritability in generation 2 and k the variance reduction coefficient). It was shown that the rate of inbreeding (ΔF) could be partitioned into three components arising from expected contributions, sampling errors and sampling covariances respectively. Using this result ΔF was derived and shown to be dominated by terms that describe ΔF by variance of family size in a single generation plus a term that accounts for the expected proliferation of lines over generations from superior ancestors in generation 1. The basic prediction of ΔF was given by

$$(1 + i^2\rho_m)(8M)^{-1} + (1 + i^2\rho_m + 2i^2\rho_f)(8F)^{-1} + K[\rho_m(16M)^{-1} + \rho_f(16F)^{-1}] - (8T)^{-1}$$

where M and F are the numbers of breeding males and females, T the number of offspring of each sex, ρ_m and ρ_f are correlations among half-sibs in generation 2 for males and females respectively, and K is a function of the intensity and competitiveness.

1. Introduction

In a novel approach to the prediction of inbreeding, Wray & Thompson (1990) used the concept of the long-term contribution of an ancestor in the first generation of a population and showed that these contributions can be related to the rate of inbreeding. This concept can be described as follows: in a population maintained by the breeding of M males and F females each generation a total of $2^{t-1}(M + F)$ distinct genealogical pathways can be traced back from generation t to generation 1, the long-term contribution of a particular ancestor is $(M + F)$ times the proportion of these pathways that lead back to that ancestor. They developed implicit formulae to relate the long-term contribution of an ancestor to its breeding value when the population was undergoing mass selection, and using these advances they de-

veloped a recursive method for the computation of the inbreeding coefficient.

The advantage of such a method was that it took account of the dependence of one generation of selection on previous generations. In random selection the selection processes in each generation proceed independently of all previous generations, but when inheritance is involved then a selective advantage (or disadvantage) of a parent is passed, in part, to its offspring. Consequently the breeding value of the parent has some influence on the selection decisions of all subsequent generations. In mass selection, this influence is mediated entirely through the genes it passes to its offspring.

Only the method of Robertson (1961) had previously allowed for this interdependence of selection decisions. Whilst the method of prediction presented by Wray & Thompson (1990) was considerably better than previous methods (Burrows, 1984*a, b*; Verrier, 1989; Wray, Woolliams & Thompson, 1990; Robert-

* Corresponding author.

son, 1961) the method suffered from having no closed form that described the relationship of inbreeding to other predictable genetic parameters.

This paper will derive the explicit relationship between the breeding value of a selected individual in generation 1 and its expected long-term contribution in generation t . It will further derive terms for accumulation of the squared contributions involving both the expected values and chance deviations, together with some adjustments appropriate for small numbers of parents and for when rates of inbreeding are not small. From these an explicit formula for inbreeding is derived, and the origins of its constituent terms identified (i.e. from expected contributions or chance deviations) and their magnitudes evaluated.

Finally it will be shown that the rate of inbreeding is closely approximated by methods based on variances of family size developed by Latter (1959) and Hill (1972) but with a simple correction for the expected inequality of contributions of like-sexed ancestors that arises from the selection process.

2. Notation

Throughout conventions on notation will follow as closely as possible those of Wray & Thompson (1990). Thus the population is propagated through hierarchical random mating of F females with M males ($M \leq F$). Each female produces a family of full-sibs of n_f males and n_f females. Each male has $n_m = M^{-1}Fn_f$ offspring of each sex. T is used to denote the total number of offspring of each sex, thus $T = n_m M = n_f F$. X (or Y) or n_x (or n_y) or subscripts x and y are used to specify a single sex either male or female. The long-term genetic contribution from an ancestor i of sex x in generation 1 to descendants of sex y in generation t will be denoted by $r_{i(x)y,t}$ and its expected value by $\mu_{i(x)y,t}$.

The value of $\mu_{i(x)y,t}$ will be shown to be linearly related to $A_{i(x)}$, representing the breeding value of the ancestor for the trait undergoing selection and the slope of the relationship will be given by $\frac{1}{2}b_{xy,t}$. $A_{i(x)}$ has been adjusted so that $E(A_{i(x)}) = 0$. The rate of inbreeding (ΔF) was shown by Wray & Thompson (1990) to be predicted by

$$\frac{1}{4}(M + F)^{-2} \left(\sum_{i=1}^M r_{i(m)}^2 + \sum_{i=1}^F r_{i(f)}^2 \right),$$

where $r_{i(x)} = r_{i(x)m,\infty} + r_{i(x)f,\infty}$.

Generation 1 itself was assumed to have been produced by the mating structure described from an unselected base population in which the trait undergoing selection had heritability h_0^2 . Generation 1 was the first generation in which selection took place and generation 2 was the first generation produced from selected parents. The heritability in generation t will be denoted h_t^2 .

Various parameters relating to the normal distribution will be used throughout: i for the intensity of

selection, p for the upper tail probability after truncation at point v with ordinate z , and $k = i(i-v)$ for the variance reduction coefficient. These will be subscripted to refer to particular sexes.

3. Expected long-term contributions

In this section, $\mu_{i(x)y,t}$, the expected long-term genetic contribution from an ancestor i of sex x and with known breeding value, to descendants of sex y after selection in generation t is predicted. Of the $2^{t-1}(M + F)$ pathways, leading back from generation t to generation 1, $2^{t-1}M$ come from males and $2^{t-1}F$ come from females in generation t . Since male and female parents make equal genetic contributions to each individual in each generation exactly half these pathways lead back to males and half to females. Therefore for all descendants of sex y in generation t a total of $2^{t-2}Y$ pathways lead back to each sex in generation 1.

If selection were random each ancestor of sex x would be expected to contribute $2^{t-2}(Y/X)$ pathways from generation 1 to descendants of sex y in generation t . This represents a proportion $2^{t-2}YX^{-1}/2^{t-1}(M + F)$ of all the possible pathways and so the expected long-term contribution of ancestor i of sex x to descendants of sex y is $(M + F)$ times this proportion i.e. $\mu_{i(x)y,t} = \frac{1}{2}YX^{-1}$. With selection some ancestors are expected to contribute more descendants, and in doing so establish more pathways, than others due to the selective advantage that is a function of the superiority of their breeding value over the breeding value of their contemporaries. It follows that, in a linear model, $\mu_{i(x)y,t} = \frac{1}{2}(YX^{-1} + b_{xy,t}A_{i(x)})$. Denoting the slope by $\frac{1}{2}b_{xy,t}$ allows comparison with Wray & Thompson (1990). $\mu_{i(x)y,t}$ is strictly an expectation that is conditional on $A_{i(x)}$. It follows that the expected number of pathways from i of sex x to descendants of sex y in generation t is $2^{t-1}\mu_{i(x)y,t}$.

The number of pathways from ancestor i of sex x to descendants of sex y in generation t can also be expressed as the sum of pathways to sex y that pass through a male in generation $t-1$ and those that pass through a female in generation $t-1$. Each pathway arriving at an individual of sex w in generation $t-1$ has on average YW^{-1} extensions to sex y in generation t ; however, if it is known that the pathway originates from a particular ancestor then the expected number of extensions increases or decreases to some degree according to the breeding value of the ancestor i.e. the expected number of extensions is given by $YW^{-1} + b_{xwy,t}A_{i(x)}$. Therefore, the expected number of pathways to generation t is given by

$$2^{t-1}\mu_{i(x)y,t} = 2^{t-2}\mu_{i(x)m,t-1}(YM^{-1} + b_{xmy,t}A_{i(x)}) + 2^{t-2}\mu_{i(x)f,t-1}(YF^{-1} + b_{xfy,t}A_{i(x)}).$$

If

$$\mu_{i(x)wy,t} = \frac{1}{2}(YW^{-1} + b_{xwy,t}A_{i(x)}),$$

then

$$\mu_{i(x)y,t} = \mu_{i(x)m,t-1} \mu_{i(x)my,t} + \mu_{i(x)mf,t-1} \mu_{i(x)fy,t}$$

and this now expresses a recurrence relationship between generations $t-1$ and t .

Note that although linear relationships have been assumed this assumption was tested using simulation by Wray & Thompson (1990) and found to be appropriate.

Wray and Thompson (1990) showed (i) $b_{xy,2} = \frac{1}{2}n_x z_y \sigma_{P_2}^{-1}$ where σ_{P_2} is the phenotypic standard deviation in generation 2 and (ii) $b_{xwy,t}$ can be approximated by

$$n_w \left(\frac{1}{2}\right)^{t-1} (1 - k_w h_{t-1}^2) \left[\prod_{i=2}^{t-2} (1 - kh_i^2) \right] z_y \sigma_{P_2}^{-1}$$

In this paper it will be assumed that by generation 2 the values of σ_P^2 and h^2 will be close to their equilibrium values, and so σ_{P_2} will be denoted σ_P and $h_2^2 = h_0^2(1 - \frac{1}{2}kh_0^2)/(1 - \frac{1}{2}kh_0^4)$ will be used in place of h_t^2 for $t \geq 2$. It is possible to use the equilibrium values calculated from the equations of Bulmer (1971) but since the coefficients $b_{xwy,t}$ diminish rapidly to zero it is more important to estimate the early coefficients most accurately. If coefficients of competitiveness are defined for each sex by

$$c_x = \frac{1}{2}(1 - k_x h_x^2) \text{ (and } c = \frac{1}{2}(c_m + c_f))$$

then

$$b_{xwy,t} \approx \frac{1}{2}n_w c_w c^{t-3} z_y \sigma_P^{-1}$$

Also by noting $n_x z_y = YX^{-1}i_y$ and defining $A'_{i(x)} = \frac{1}{2}A_{i(x)}\sigma_P^{-1}$ it is observed that

$$\mu_{i(x)y,2} = \frac{1}{2}YX^{-1}(1 + i_y A'_{i(x)})$$

and

$$\mu_{i(x)wy,t} = \frac{1}{2}YW^{-1}(1 + i_y c_w c^{t-3} A'_{i(x)})$$

As t becomes large $\mu_{i(x)wy,t} \rightarrow \frac{1}{2}YW^{-1}$ indicating that the influence of ancestor's breeding value on the selection process in generation t decreases to zero.

Let $b_{x,t} = (b_{x_m,t}, b_{x_f,t})^T$ and $z = (z_m, z_f)^T$, then from collecting terms in the recurrence relationship that are linear in $A_{i(x)}$,

$$b_{x,t} = D b_{x,t-1} + \frac{1}{2}X^{-1}[M(b_{x_{mm},t}, b_{x_{mf},t})^T + F(b_{x_{fm},t}, b_{x_{ff},t})^T],$$

where D is the matrix

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2}MF^{-1} \\ \frac{1}{2}FM^{-1} & \frac{1}{2} \end{pmatrix}$$

D describes the expected dispersion of genes through the population in the absence of selection from generation to generation. For example, $\frac{1}{2}MF^{-1}$ is the expected number of copies of a gene sampled from selected females in generation $t-1$, among the selected males of generation t . D is idempotent, i.e. $D^2 = D$, and this property embodies the phenomenon that only

a single generation is required to disperse genes from one sex through a homogeneous, random mating, diploid population with discrete generations.

Using the property of idempotence Appendix 1 shows that the recurrence relationship can be solved to give

$$b_{x,t} = \frac{1}{2}\sigma_P^{-1}[S_{t-3}i(MX^{-1}, FX^{-1})^T + c^{t-2}(i_m MX^{-1}, i_f FX^{-1})^T],$$

and consequently

$$\mu_{i(x)y,t} = \frac{1}{2}YX^{-1}[1 + A'_{i(x)}(S_{t-3}i + c^{t-2}i_y)]$$

and

$$\mu_{i(x)y,\infty} = \frac{1}{2}YX^{-1}[1 + iS_\infty A'_{i(x)}]$$

where

$$S_t = \sum_{i=0}^t c^i,$$

and

$$S_\infty = \sum_{i=0}^{\infty} c^i = (1 - c)^{-1} = 2(1 + kh_2^2)^{-1}$$

Thus $b_{xy,\infty}/b_{xy,2} = i/i_y(1 + kh_2^2)$ independent of x . Application of this formula to the results of Wray & Thompson (1990) shows accurate prediction: some comparisons can be made with their simulation results present in their Table 4 which differ by ≤ 0.01 from the expectation given above.

4. Prediction of the rate of inbreeding

Wray & Thompson (1990) showed that ΔF is related to the expectation of the squared contributions and this involves not only the expectation of the square of the conditional expectations calculated in the previous section but also the expectation of the conditional variance. In this section the methods required to derive these expectations are described. From the definition,

$$\Delta F = \frac{1}{4}(M + F)^{-2} \left(\sum_{i=1}^M r_{i(m)}^2 + \sum_{i=1}^F r_{i(f)}^2 \right),$$

thus

$$E(\Delta F) = \frac{1}{4}(M + F)^{-2} \left(\sum_{i=1}^M E(r_{i(m)}^2) + \sum_{i=1}^F E(r_{i(f)}^2) \right) = \frac{1}{4}(M + F)^{-2} (ME(r_{i(m)}^2) + FE(r_{i(f)}^2)).$$

Suppose the squared contribution can be decomposed into s elements so that

$$r_{i(x)}^2 = \sum_{j=1}^s r_{i(x)}^2(j),$$

then

$$\Delta F = \frac{1}{4}(M + F)^{-2} \sum_{j=1}^s \sum_{x=m,f} XE(r_{i(x)}^2(j)).$$

The expression has been given in this form as it will be seen that simplification of terms is derived from it.

Wray & Thompson (1990) derive ΔF in a 'lateral' and recursive accumulation: terms involved in $r_{i(x),2}^2$ are derived and accumulated; the expected change in contribution in these moving from generations 2 to 3 is derived and $r_{i(x),3}^2$ is derived by adding further terms originating in generation 3, the recursion then proceeds through generations until contributions become negligibly small.

Following equations (24) and (27) of Wray & Thompson (1990)

$$E(r_{i(x),2}^2) = E\{\mathbf{1}^T[\boldsymbol{\mu}_{i(x),2} \boldsymbol{\mu}_{i(x),2}^T + V_{i(x),2} + C_{i(x),2}]\mathbf{1}\}$$

where $\mathbf{1}^T = (1, 1)$ and $V_{i(x),t}$ and $C_{i(x),t}$ are the matrices of new contributions arising in generation t from binomial sampling and from additional covariance through co-selection of sibs (i.e. the squared contribution is the sum of the mean contribution squared plus the variance of the contribution).

In generation 3

$$E(r_{i(x),3}^2) = \mathbf{1}^T[M_{i(x),3}(\boldsymbol{\mu}_{i(x),2} \boldsymbol{\mu}_{i(x),2}^T + V_{i(x),2} + C_{i(x),2}) + M_{i(x),3}^T + V_{i(x),3} + C_{i(x),3})\mathbf{1},$$

where

$$M_{i(x),t} = \begin{pmatrix} \mu_{i(x)mm,t} & \mu_{i(x)fm,t} \\ \mu_{i(x)mf,t} & \mu_{i(x)ff,t} \end{pmatrix}$$

and ultimately

$$E(r_{i(x)}^2) = \mathbf{1}^T P_{i(x),3} \boldsymbol{\mu}_{i(x),2} \boldsymbol{\mu}_{i(x),2}^T P_{i(x),3}^T \mathbf{1} + \sum_{t=2}^{\infty} \mathbf{1}^T P_{i(x),t+1} V_{i(x),t} P_{i(x),t+1}^T \mathbf{1} + \sum_{t=2}^{\infty} \mathbf{1}^T P_{i(x),t+1} C_{i(x),t} P_{i(x),t+1}^T \mathbf{1},$$

where

$$P_{i(x),t} = \prod_{j=t}^{\infty} M_{i(x),j}$$

Thus $E(r_{i(x)}^2)$ has three terms: due to the squared mean, to binomial sampling and to co-selection of sibs. By deriving these terms separately and accumulating them vertically over generations a closed expression for ΔF is obtained.

To achieve the vertical accumulation it is necessary to derive $P_{i(x),t+1}$ explicitly. Appendix 2 shows $P_{i(x),t+1}$ to be

$$D + A'_{i(x)} c^{t-2} i D \Theta + A'_{i(x)} (c^{t-2})^2 \cdot icq D \Lambda + \text{terms in } A'_{i(x)}.$$

Here Θ and Λ are diagonal matrices whose elements, along with q , depend on the intensity and competitiveness of the selection in each sex.

A general form for each of $V_{i(x),t}$ and $C_{i(x),t}$ is $u_{0,t} \boldsymbol{\delta} + u_{1,t} A'_{i(x)} \boldsymbol{\varepsilon} + u_{2,t} A'_{i(x)} \boldsymbol{\zeta}$ where $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}$ are matrices that depend on genetic parameters but are independent of t . $P_{i(x),t+1}$ is of the form $\boldsymbol{\alpha} + c^{t-2} A'_{i(x)} \boldsymbol{\beta} + c^{2t-4} A'_{i(x)} \boldsymbol{\gamma}$.

Since $E(A'_{i(x)}) = 0$ and contributions of $O(A'_{i(x)}^3)$ or higher are ignored the contributions to $\frac{1}{4}(M+F)^{-2} XE(r_{i(x)}^2)$ are of the form $v_0 + E(A'_{i(x)}^2) \sum_{j=1}^4 v_j$, where

$$v_0 = (\sum u_{0,t}) (\mathbf{1}^T \boldsymbol{\alpha} \boldsymbol{\delta} \boldsymbol{\alpha}^T \mathbf{1}), \quad v_1 = (\sum u_{2,t}) (\mathbf{1}^T \boldsymbol{\alpha} \boldsymbol{\zeta} \boldsymbol{\alpha}^T \mathbf{1}), \\ v_2 = (\sum c^{2t-4} u_{0,t}) (\mathbf{1}^T \boldsymbol{\beta} \boldsymbol{\delta} \boldsymbol{\beta}^T \mathbf{1}), \\ v_3 = (2 \sum c^{t-2} u_{1,t}) (\mathbf{1}^T \boldsymbol{\beta} \boldsymbol{\varepsilon} \boldsymbol{\alpha}^T \mathbf{1}), \\ v_4 = (2 \sum c^{2t-4} u_{0,t}) (\mathbf{1}^T \boldsymbol{\gamma} \boldsymbol{\delta} \boldsymbol{\alpha}^T \mathbf{1}).$$

The infinite sums in these expressions are convergent since $u_{0,t}$, $u_{1,t}$ and $u_{2,t}$ are $O(2^{-t})$. The symbols v_0 and v_j will be referred to in Appendix 3 in order to aid the identification of the origin of terms. $E(A'_{i(x)}^2)$ is the correlation between half-sibs with common parent of sex x in generation 2 prior to selection and will be denoted ρ_x and his value $\frac{1}{4}h_0^2(1 - k_x h_0^2)/(1 - \frac{1}{2}k h_0^4)$.

The matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in $P_{i(x),t+1}$ have already been derived and are all of the form D post-multiplied by some diagonal matrix. Therefore the general form for contributions to $E(r_{i(x)}^2)$ is always $\frac{1}{4}(M+F)^{-2} \mathbf{1}^T D T D^T \mathbf{1}$ for some matrix T . If T has elements t_{ij} , then

$$\frac{1}{4}(M+F)^{-2} \mathbf{1}^T D T D^T \mathbf{1} = (F^2 t_{11} + M F t_{12} + M F t_{21} + M^2 t_{22}) / (16 M^2 F^2).$$

(i) Squared mean contributions

For ancestors of sex x the contribution to ΔF is $\frac{1}{4} X(M+F)^{-2} \mathbf{1}^T P_{i(x),3} \boldsymbol{\mu}_{i(x),2} \boldsymbol{\mu}_{i(x),2}^T P_{i(x),3}^T \mathbf{1}$. Since

$$\mu_{i(x)w,2} \mu_{i(x)y,2} = \frac{1}{4} W Y X^{-2} (1 + A'_{i(x)}(i_w + i_y) + A'_{i(x)} i_w i_y), \\ X \boldsymbol{\mu}_{i(x),2} \boldsymbol{\mu}_{i(x),2}^T = \frac{1}{4} X^{-1} \\ \left[\begin{pmatrix} M^2 & MF \\ MF & F^2 \end{pmatrix} + A'_{i(x)} \begin{pmatrix} 2M^2 i_m & MF(i_m + i_f) \\ MF(i_m + i_f) & 2F^2 i_f \end{pmatrix} + A'_{i(x)}^2 \begin{pmatrix} M^2 i_m^2 & M F i_m i_f \\ M F i_m i_f & F i_f^2 \end{pmatrix} \right].$$

Contributions are given in Appendix 3.

(ii) Binomial sampling

In generation 2, if success is taken as a selected offspring of sex y the binomial parameter for ancestor i of sex x is given by equation (22) of Wray & Thompson (1990) as

$$P_{i(x)y,2} = p_y + n_x^{-1} b_{xy,2} A_{i(x)} = p_y (1 + i_y A'_{i(x)}).$$

The variance assuming independent trials is therefore $n_x P_{i(x)y,2} (1 - P_{i(x)y,2})$. The contribution to $E(r_{i(x),2}^2)$ is

$$V_{i(x),2} = \frac{1}{4} X^{-1} T [V_0 + A'_{i(x)} V_1 + A'_{i(x)}^2 V_2]$$

where V_0 , V_1 and V_2 are diagonal matrices:

$$V_0 = \text{diag}(p_m(1 - p_m), p_f(1 - p_f)),$$

$$V_1 = \text{diag}((1 - 2p_m) z_m, (1 - 2p_f) z_f)$$

and

$$V_2 = \text{diag}(-z_m^2, -z_f^2).$$

In subsequent generations, new contributions arise from binomial sampling of gene pathways leading to offspring of sex y in generation t from parents of sex w . The probability of success is given by $p_{i(x)wy,t} = p_y(1 + i_y c_w c^{t-3} A'_{i(x)})$ and each parent has a family size of n_w . The expected contribution of these variances is weighted by the expected number of ancestors $\mu_{i(x)w,t-1}$. The contributions are then summed over w . Thus

$$V_{i(x),t} = 2^{-t} X^{-1} T [V_0 + A'_{i(x)} [i S_{t-3} V_0 + c^{t-2} V_1] + A'^2_{i(x)} [(i c^{t-3} S_{t-4} + g c^{2t-6}) V_1 + \frac{1}{2} (c_m^2 + c_f^2) c^{2t-6} V_2]]$$

$V_{i(x),t}$ can be thus separated into 5 components denoting terms in V_0 , $A'_{i(x)} V_0$, $A'_{i(x)} V_1$, $A'^2_{i(x)} V_1$ and $A'^2_{i(x)} V_2$. The resulting terms are listed in Appendix 3.

(iii) Co-selection of sibs

The original description of co-selection by Wray & Thompson (1990) was incomplete, leading to the omission of important contributions from the co-selection of sibs, including half-sibs. The omissions can contribute up to 8% of the total for the examples of mass selection considered here and can be many times larger than the terms considered by Wray & Thompson (1990). The original description was also incorrect in using $h^2/4$ as a correlation between half-sibs which can lead to important errors as h^2 increases.

The objective is to estimate the variance of family size conditional on the breeding value of an ancestor in generation 1. Co-selection occurs through the covariances of selection probabilities of full- or half-sibs that are not accounted for by regression on the breeding value of the ancestor in generation 1. It is useful to split the consideration of this into two parts; generation 2 and generation 3 onwards, since in generation 2 the ancestor is in fact the parent.

In generation 2, the covariance between half-sibs or full-sibs, arising from the parent of sex x are already fully accounted for in the term describing expected contributions, leaving only those arising in full-sib families through the parent of sex x' (i.e. the sex other than x). The probability of selecting an individual of sex y with parents of breeding value $A_{i(x)}$ and $A_{j(x')}$ is given by $p_y(1 + i_y A'_{i(x)} + i_y A'_{j(x')})$. Thus when mating is at random, and conditional on the breeding value $A_{i(x)}$, the covariance of the selection probabilities of two full-sibs of sex y and z is given by

$$E[p_y p_z (1 + i_y A'_{i(x)} + i_y A'_{j(x')}) (1 + i_z A'_{i(x)} + i_z A'_{j(x')})] - p_y p_z (1 - i_y A'_{i(x)}) (1 + i_z A'_{i(x)}) = p_y p_z i_y i_z \rho_x$$

This covariance will arise between all full-sib pairs for ancestor $i(x)$, i.e. $X^{-1} T(n_f - 1)$ times if $y = z$ and $X^{-1} T n_f$ if $y \neq z$. For simplicity of exposition the $(n_f - 1)$ will be treated as n_f at present. Thus $C_{i(x),2} = \frac{1}{3} X^{-1} T n_f \rho_x C_0$ where $C_0 = (p_m i_m, p_f i_f)^T (p_m i_m, p_f i_f)$. The terms arising from $C_{i(x),2}$ are given in Appendix 3 for a single sex x .

In generation t where $t \geq 3$, for an offspring of a parent of sex w of breeding value A_w with ancestor $i(x)$ in generation 1 and mated at random to an individual of sex w' with breeding value $A_{w'}$, the probability of selection expressed as a regression on $A_{i(x)}$, A_w and $A_{w'}$ is

$$p_y (1 + c^{t-3} c_w i_y A'_{i(x)} + i_y (A'_w - c^{t-3} c_w A'_{i(x)}) + i_y A'_{w'})$$

Thus conditional on the ancestor's breeding value, with random mating the covariance not only arises from w' but from w as well. For two full-sibs of sexes y and z the covariance is

$$p_y p_z i_y i_z [(\rho_w + \rho_{w'}) - c_w^2 c^{2t-6} \rho_x];$$

but for two half-sibs of sexes y and z with common male parent there is also a covariance of the form $p_y p_z i_y i_z [\rho_m - c_m^2 c^{2t-6} \rho_x]$. Strictly speaking in generation $t \geq 3$ the correlation of half-sibs with common parent of sex w is approximately $\frac{1}{2} h_w^2 c_w$, but it will be assumed that for the cases considered this differs little from ρ_w , the correlation of half-sibs in generation 2. With the hierarchical mating structure considered here only male half-sibs are formed.

There are $M^{-1} T n_f (d - 1)$ half-sib pairs in a male parent's family where $d = F M^{-1}$. Thus for a male descendant in generation $t - 1$ where $t \geq 3$

$$C_{i(x)m,t} = 2^{-t} M^{-1} T (n_f (\rho_m + \rho_f) + n_f (d - 1) \rho_m - n_m c_m^2 c^{2t-6} \rho_x) C_0$$

whereas for a female descendant

$$C_{i(x)f,t} = 2^{-t} F^{-1} T (n_f (\rho_m + \rho_f) - n_f c_f^2 c^{2t-6} \rho_x) C_0$$

Following the procedure set out in the section on binomial sampling the terms require weighting by $\mu_{i(x)w,t-1}$ and summed for $w = m$ and f . The terms arising are listed in Appendix 3. Terms in $(c^3)^t$ have been ignored and also simplification has been made for (A 3.24).

5. Application of the prediction equations

In this section, results from using the full prediction equations listed in Appendix 3 will be compared to simulation. In the simulations $M = 20$, and F varied from 20 to 200 with two family sizes $n_f = 3$ and 6. For each combination a range of heritabilities from 0 to 0.99 was considered. To assess the predictive power of the method it was necessary to consider three modifications to the methods presented by Wray & Thompson (1990): (i) replacement of binomial sampling by hypergeometric sampling; (ii) correction for the contribution of the base population; and (iii) accounting for the sampling of the breeding value of the ancestors. These modifications are described in Appendices 4-6. The first has greatest influence when $M(F)$ is small, the second when ΔF increases in size, and the third when heritabilities are high. Results are presented in Table 1.

Table 1. Prediction of ΔF using all terms in Appendix 3 (ΔF_{pred}) compared to rates obtained from simulation (ΔF_{sim}). Values are $100 \times \Delta F$, and % errors are calculated as $(\Delta F_{sim} - \Delta F_{pred})/\Delta F_{pred}$

	F = 20				F = 40			F = 100			F = 200		
	h_0^2	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}}$	% error	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}}$	% error	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}}$	% error	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}}$	% error
$n_f = 3$	0	104	+3	+2.9	83	0	0	71	0	0	67	-1	-1.5
	0.1	119	+4	+3.4	99	-1	-1.0	88	-6	-6.8	85	-3	-3.5
	0.2	130	+3	+2.3	109	+1	+0.9	98	-4	-4.1	96	-10	-10.4
	0.4	142	0	0	120	-2	-1.6	107	-2	-1.9	105	-5	-4.8
	0.6	145	+5	+3.3	122	+1	+0.8	107	-5	-4.6	104	-5	-4.8
	0.99	137	0	0	109	-1	-0.9	90	-1	-1.1	83	-2	-2.4
$n_f = 6$	0	115	-2	-1.8	89	-1	-1.1	73	0	0	68	0	0
	0.1	146	-2	-1.4	116	+1	+0.9	99	-4	-4.0	95	-8	-8.4
	0.2	166	-5	-3.0	134	-4	-3.0	115	-5	-4.3	111	-8	-7.2
	0.4	187	+5	+2.7	150	0	0	128	-4	-3.1	123	-6	-4.9
	0.6	191	+7	+3.7	152	-2	-1.3	127	-4	-3.1	120	-1	-0.8
	0.99	165	-3	-1.8	124	-5	-4.0	98	-3	-3.1	88	-1	-1.1

* Standard errors from simulations vary from 1 to 2.

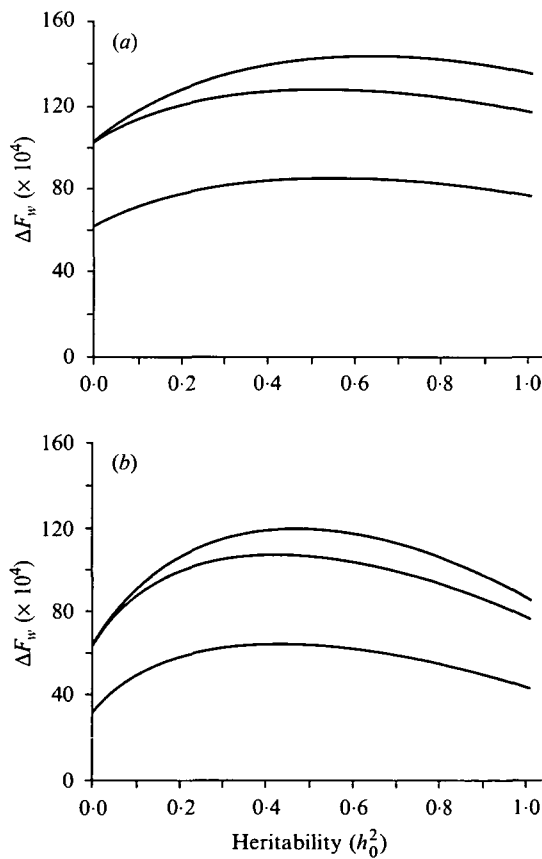


Fig. 1. The relationship of ΔF_w and its three components with heritability for (a) $M = F = 20$, $n_f = 3$ and (b) $M = 20$, $F = 200$, $n_f = 6$. The lower line is the squared mean contribution, the middle line is the sum of the squared mean and sampling contributions and the upper line is ΔF_w . (ΔF_w is the rate of inbreeding uncorrected for base contributions as described in Appendix 5.)

The results show a clear trend; very accurate prediction for $F \leq 40$, but an increasing tendency to overpredict as F increases further. This is also confounded with the increase in FM^{-1} . The possible

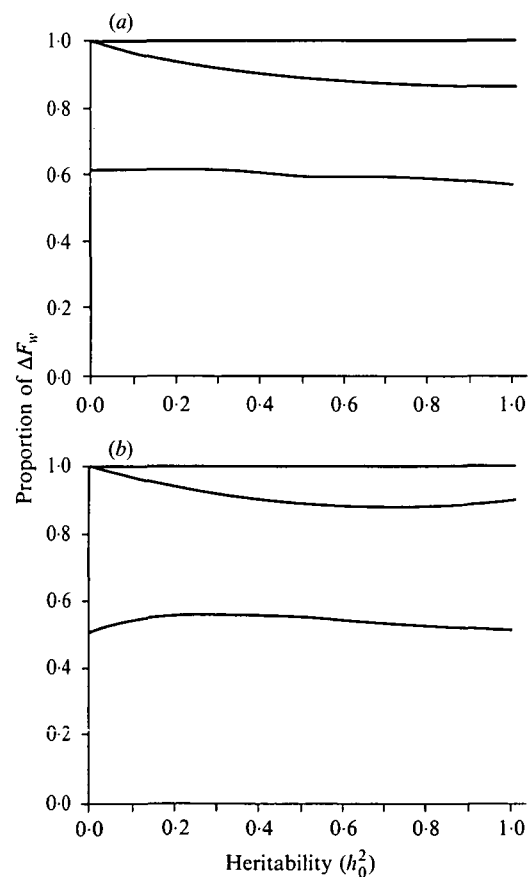


Fig. 2. The proportional contribution of the three components of inbreeding for (a) $M = F = 20$, $n_f = 3$ and (b) $M = 20$, $F = 200$, $n_f = 6$. The lower line is the contribution of the squared mean and the middle line is the sum of the squared mean and sampling contribution.

reasons will be discussed in more detail in a later section.

It has been shown that for mass selection 3 types of contribution to long-term contributions (and hence to ΔF) can be identified and modelled: squared mean,

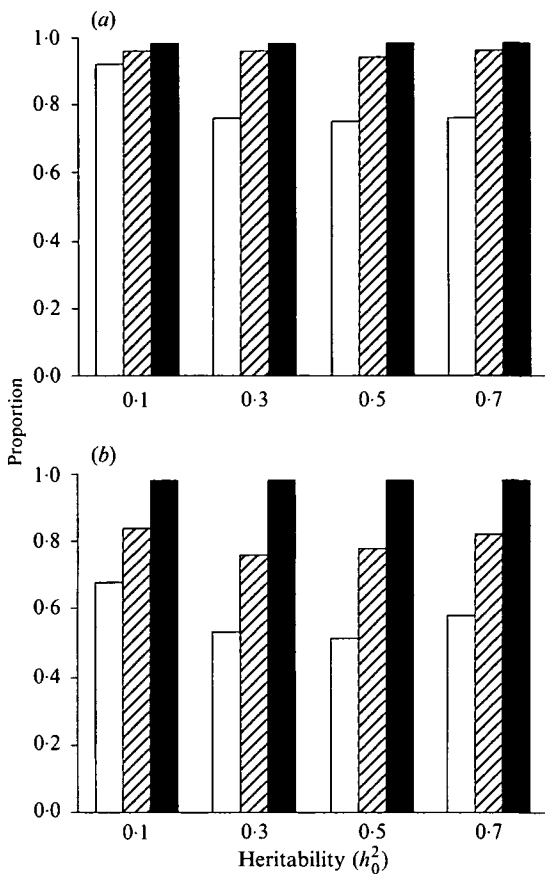


Fig. 3. The proportional contribution of (i) equation (A 3.1) to the squared mean contribution; (ii) equation (A 3.6) to the sampling contribution; and (iii) equations (A 3.13), (A 3.16) and (A 3.17) to the co-selection contribution. Results are shown as (a) $M = F = 20$, $n_f = 3$ and (b) $M = 20$, $F = 200$, $n_f = 6$. (See Appendix 3 for definition of the equations.)

sampling and co-selection of sibs. These sources change in importance as the population parameters change and further, within each type the importance of particular contributions involving inheritance also varies. Among the terms for the squared mean only (A 3.1) occurs independently of selection, and likewise (A 3.6) is the only sampling term to occur in the absence of selection. For the purpose of this paper, the co-selection of sibs occurs through genetic covariance only. However, terms analogous to (A 3.13), (A 3.16) and (A 3.17) will occur when common family variance is encountered through non-genetic means e.g. maternal or environmental factors, but none of the remaining co-selection terms would occur without some mode of inheritance for the trait under selection. Thus for co-selection there is a case for separating the contribution of these three from the remaining terms among (A 3.13) to (A 3.24).

The magnitude of these contributions have been examined using the prediction for two cases $M = F = 20$, $n_f = 3$ and $M = 20$, $F = 200$, $n_f = 6$, and results are presented in Figs 1–3.

Figure 1 shows the change in the expected long-term squared contribution with heritability separated

into three types of contribution, whilst Fig. 2 shows these same data when expressed as a proportion of the total. In the cases considered the predicted $E(r^2)$ reaches its peak for h^2 between 0.4 and 0.6. When $F = 20$, the squared mean contribution remained between 57 and 62% whilst for $F = 200$ the squared mean contribution remained less than 58% of the total. In both examples the proportional contribution from co-selection increased from $h^2 = 0$ to 0.9 and contributed up to 13% of the total.

Figure 3 shows the proportional contribution made by: (i) (A 3.1) to the total squared mean contribution; (ii) (A 3.6) to the sampling contribution; and (iii) (A 3.13), (A 3.16) and (A 3.17) to the co-selection contribution. For h^2 as low as 0.2, approximately 30% of the squared mean contribution was due to inherited advantage, whereas this accounted for up to 20% and only 2% of the sampling and co-selection components respectively. It was also noted that (A 3.10) was large only when $F = 100$ and 200, where ΔF was over-predicted.

6. Approximation and relationship with variance of family size

The following simplifying approximation was strongly suggested by the results.

(i) The terms of the squared mean were all included and when summed give a total contribution of

$$(16M)^{-1}[1 + (K + i^2)\rho_m] + (16F)^{-1}[1 + (K + i^2)\rho_f], \quad (1)$$

where $K = i^2(S_\infty^2 - 1) + 2iqQ_\infty$ (where S_∞ , Q_∞ and q are as defined in Appendix 2, namely $(1 - c^{-1})$, $(1 - c^2)^{-1}$ and $\frac{1}{2}(i_m c_m + i_f c_f) + ic^2 S_\infty$ respectively).

(ii) Only (A 3.6) was included from the sampling terms and this has the form

$$(16M)^{-1} + (16F)^{-1} - (8T)^{-1}. \quad (2)$$

Correction for hypergeometric sampling is achieved by multiplication with $[\frac{1}{2}(1 - M^{-1}) + \frac{1}{2}(1 - F^{-1})]$.

(iii) Only terms (A 3.13), (A 3.16) and (A 3.17) were included from the covariance terms and these combine to give

$$i^2(16F)^{-1}[\rho_m(FM^{-1} - 1) + 3(\rho_m + \rho_f)]. \quad (3)$$

More precisely i^2 should be replaced by

$$\frac{1}{4}((1 - M^{-1})(1 - n_f^{-1})i_m^2 + 2i_m i_f + (1 - F^{-1})(1 - n_f^{-1})i_f^2)$$

although no correction for n_f is necessary for A3.17.

The results from this formula are given in Table 2. Results from Table 2 show that the approximation has a tendency to underpredict for $h^2 = 0.4, 0.6$, but exhibits little loss of accuracy for $F \leq 40$ but a gain in accuracy for $F = 100$ and 200. It appears more robust.

Latter (1959) and Hill (1972) derived an expression for the rate of inbreeding using the variances of family

Table 2. Prediction of ΔF using the approximation compared to rates obtained from simulation. Values are $100 \times \Delta F$, and % errors are calculated as $(\Delta F_{sim} - \Delta F_{pred}) / \Delta F_{pred}$

		F = 20			F = 40			F = 100			F = 200		
n_f	h_0^2	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}} *$	% error	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}} *$	% error	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}} *$	% error	ΔF_{pred}	$\frac{\Delta F_{sim} - \Delta F_{pred}}{\Delta F_{pred}} *$	% error
$n_f = 3$	0	104	+3	+2.9	83	0	0	71	0	0	67	-1	-1.5
	0.1	118	+5	+4.2	97	+1	1.0	84	-2	-2.3	81	+1	+1.2
	0.2	128	+5	+3.9	106	+4	+3.8	93	+1	+1.1	90	-4	-4.4
	0.4	140	+2	+1.4	116	+2	+1.8	102	+3	+2.9	98	+2	+2.0
	0.6	144	+6	+4.2	119	+4	+3.4	103	-1	-1	98	+1	+1.0
	0.99	138	-1	+0.7	109	-1	-0.9	89	0	0	82	-1	-1.2
$n_f = 6$	0	115	-2	-1.7	89	-1	-1.1	73	0	0	68	0	0
	0.1	141	+3	+2.1	111	+6	+5.4	94	+1	+1.1	88	-1	-1.1
	0.2	159	+2	+1.3	126	+4	+3.1	107	+3	+2.8	101	+2	+2.0
	0.4	180	+12	+6.6	142	+8	+5.6	119	+5	+4.2	112	+5	+4.5
	0.6	186	+12	+6.1	146	+4	+2.7	120	+3	+2.5	112	+7	+6.2
	0.99	164	-2	+1.2	124	-5	-4.0	97	-2	-2.1	87	0	0

* Standard errors of mean simulated values range from 1 to 2.

size that was general enough to include environmental covariances between sibs but did not include any framework for selection and the inheritance of selective advantage. Wray (1989) derived a form of this equation (ΔF_{LH}) for populations of the same structure considered here that predicted the increased family size arising from sib covariances due to genetic variation and selection in one generation only. This form, ignoring correction factors of $(1 - n_f^{-1})$, $(1 - M^{-1})$ and $(1 - F^{-1})$ is:

$$\Delta F_{LH} = (8M)^{-1} + (8F)^{-1} - (8T)^{-1} + i^2 \rho_m [(8M)^{-1} + (8F)^{-1}] + i^2 \rho_f (4F)^{-1}$$

This was found to underestimate rates of inbreeding (Wray, 1989; Wray *et al.* 1990). However, if the sum of equations (1) to (3) are denoted ΔF_w [ignoring corrections to (2) and (3)] it can be seen that

$$\Delta F_w = \Delta F_{LH} + K[\rho_m(16M)^{-1} + \rho_f(16F)^{-1}]. \tag{4}$$

In fact, this relationship still holds when the corrections using $(1 - n_f^{-1})$, $(1 - M^{-1})$ and $(1 - F^{-1})$ are made to both ΔF_{LH} and ΔF_w . The term in ΔF_w that is not included in ΔF_{LH} , is part of the squared mean contribution and describes the extra inbreeding arising from selection; this is caused by the interdependence of generations through the inheritance of selected advantage and the consequent expected proliferation of lines arising from superior ancestors in generation 1 at the expense of their inferior contemporaries. Precise prediction will require correction using Appendix 5.

In conclusion, the prediction of inbreeding in mass selection can be shown to approximate closely a prediction involving the variance of family size assuming independent selection processes in each generation plus a single term that describes the cumulative effect of the expected proliferation of lines

from superior ancestors in generation 1 at the expense of lines from the inferior ancestors.

(i) Example

From $M = 20$, $F = 40$, $n_f = 6$ and $h_0^2 = 0.4$: the proportions selected are $p_m = 0.0833$ and $p_f = 0.1667$; the intensities of selection and variance reduction coefficients are (from standard tables) $i_m = 1.839$, $i_f = 1.499$ and $i = \frac{1}{2}(i_m + i_f) = 1.669$, $k_m = 0.838$, $k_f = 0.797$ and $k = \frac{1}{2}(k_m + k_f) = 0.818$. As is the case for estimating progress, accurate estimation of ΔF requires the calculation of h_2^2 and half-sib correlations (ρ_m and ρ_f). For the example chosen these are $h_2^2 = 0.358$, $\rho_m = 0.071$ and $\rho_f = 0.073$, to give $c_m = 0.350$, $c_f = 0.357$ and $c = \frac{1}{2}(c_m + c_f) = 0.354$, and $K = 7.361$.

For simplicity the hypergeometric corrections will be ignored: the squared mean contribution is 0.0081 [equation (1)], the sampling contribution is 0.0042 [equation (2)], and the co-selection contribution is 0.0022 [equation (3)]. This gives a value of $\Delta F_w = 0.0145$, which when corrected using Appendix 5 gives $\Delta F = 0.0149$ (compared to the value of 0.0150 from stimulation). The prediction is slightly higher than that shown in Table 2 for two reasons: firstly, hypergeometric corrections have not applied; and secondly, the values of h_2^2 , ρ_m and ρ_f that were calculated above were not corrected for sampling as described in Appendix 6, but were simply adjusted for the Bulmer effect as defined earlier in the text.

7. Discussion

The work presented has shown that good predictions of rates of inbreeding in mass selection can be made using a straightforward closed expression involving predictable genetic parameters. These genetic parameters involve the intensity of selection and variance

parameters from the second generation of offspring, i.e. heritability, competitiveness ($c = \frac{1}{2}(1 - kh^2)$) and half-sib correlations.

The terms involved predict that to order X^{-1} , the rate of inbreeding in the absence of genetic variation and environmental covariation takes the form $(8M)^{-1} + (8F)^{-1} - (8T)^{-1}$ a term that can be derived for random selection with fixed family sizes by other methods (Burrows, 1984a; Wray, Woolliams and Thompson, 1990). This paper has shown that when viewed from the concept of long-term contributions this has two components, firstly a squared mean contribution of $(16M)^{-1} + (16F)^{-1}$ and secondly a smaller sampling contribution of $(16M)^{-1} + (16F)^{-1} - (8T)^{-1}$. When heritable and non-heritable correlations are present between sibs in a hierarchical scheme then additional terms analogous to the situations considered by Latter (1959) and Hill (1972) are required plus the addition to the squared mean contribution of $K[\rho_m(16M)^{-1} + \rho_f(16F)^{-1}]$ where ρ_x is the half-sib correlation with common parent of sex x , and $K = i^2(S_\infty^2 - 1) + 2iqQ_\infty$. If selection is assumed to be of equal intensity in the sexes $q = icS_\infty$ and $K = i^2(S_\infty - 1)(1 + S_\infty + 2Q_\infty)$. This extra term describes the inbreeding arising from the expected proliferation of lines arising from superior ancestors in generation 1 at the expense of lines from their inferior contemporaries.

The full derivation of the theory presented by Wray & Thompson (1990) was found to be very accurate in the situations studied in this paper for $F \leq 40$, but clearly overestimated for $F \geq 100$ where male half-sib families are large. Overestimation can be firmly ascribed to the inability to predict $E(r^2)$ rather than a failure in the relationship of $E(r^2)$ with ΔF : simulations show that for mass selection, index selection and niche selection with environmental covariances, $E(r^2)$ accurately predicts ΔF when mating is at random (N. R. Wray, unpublished results). Two possible reasons for the inability to predict $E(r^2)$ in these circumstances can be advanced. Firstly when modelling the sib covariance only partial adjustment has been made for hypergeometric sampling; when covariances are added, negative covariances must also be added elsewhere i.e. one families success in another's failure. When the half-sib family size is large, selection becomes more intense ($p_m \leq n_m^{-1}$) and Appendix 1 shows the regression of the selection score of an ancestor on its breeding value becomes steeper: in the next generation, if an ancestor has many descendants the success of one branch of his family is to the detriment of another. Thus the expected increase in pathways is not as great as predicted. A second contributing cause is that in later generations the accumulation of variances has been assumed to depend on the ancestor [equation (29) of Wray & Thompson, 1990]. In fact after generation 2 families of descendants are a mix of ancestors from the same sex and whilst the overall expected gain in pathways is unaffected by

this, the variance of additional contributions between like-sex ancestors will be reduced. One implication of all these considerations is that the robustness of estimation may be less determined by the number of parents than by the number of parents in relation to family size. Nevertheless, the approximation derived is robust over a wide range of parameters, and although possibly benefitting from compensating errors, achieves a great gain in simplicity.

The importance of identifying the components to long-term contributions and quantifying their magnitude in terms of predictable genetic parameters, is that it is only by these means that the value of quanta of information can be assessed for both promoting genetic gain and promoting inbreeding. With this understanding it would then be possible to reconsider selection indices and scheme design to maximise genetic gain while simultaneously constraining inbreeding. The impact of this would clearly depend on the circumstances and chosen constraints. An example of this was given by Woolliams (1989) in which the change from hierarchical to factorial mating designs increased progress when inbreeding was constrained: the foregoing analysis shows that this change left the squared mean and sampling contributions unchanged but substantially reduced that from co-selection. The full potential will, however, require the extension of the foregoing analysis to index selection.

A further finding of the paper is that the expected long-term contribution of an ancestor in mass selection relative to that for random selection will increase by a factor that is linearly related to its breeding value. The factor is of the form $(1 + i(1 - c)^{-1}A'_{i(x)})$, where $A'_{i(x)}$ is $\frac{1}{2}A_{i(x)}\sigma_p^{-1}$, and $A_{i(x)}$ is the ancestors breeding value about the mean of those selected in generation 1. Thus selection will be expected to leave the contribution unchanged only for an average ancestor, and in general the expected contribution will be linearly related to the ancestors breeding value even though in their offspring in generation 2, contributions are equal. Toro & Neita (1984) and Lindgren (1991) have shown that over a single generation when a restriction is placed on diversity, genetic gain is increased by the differential use of individuals, linearly related to their breeding value. An interesting question arises as to whether over many generations this reduces or exacerbates inbreeding since the superior individuals not only have more progeny but their progeny have a selective advantage. This will determine if linear deployment is a strategy for long-term selection or for clonal propagation.

In a wider context it has been shown that a tractable approach to the problem of predicting inbreeding in the presence of some form of selection may be to estimate the expected long-term squared contribution of ancestors. However, these have only been shown to be equivalent for random mating. It would seem a logical and sensible alternative in some circumstances to assess breeding schemes by $\frac{1}{4}(M + F)^{-2}\Sigma E(r_i^2)$ rather

than by ΔF and that this assessment may also prove tractable in the absence of random mating.

J. A. Woolliams and R. Thompson gratefully acknowledge funding by the Ministry of Agriculture, Fisheries and Food, UK.

Appendix 1: Derivation of $b_{x,t}$

$$\begin{aligned}
 D\mathbf{b}_{x,t} &= D\mathbf{b}_{x,t-1} + \frac{1}{4}MX^{-1}n_m c_m c^{t-3}\sigma_P^{-1} \mathbf{z} \\
 &\quad + \frac{1}{4}FX^{-1}n_f c_f c^{t-3}\sigma_P^{-1} \mathbf{z} \\
 &= D\mathbf{b}_{x,t-1} + \frac{1}{2}n_x c^{t-2}\sigma_P^{-1} \mathbf{z}.
 \end{aligned}$$

Since

$$D^t = D, \mathbf{b}_{x,t} = D \left[\mathbf{b}_{x,2} + \frac{1}{2}n_x \sigma_P^{-1} \left(\sum_{i=1}^{t-3} c^i \right) \mathbf{z} \right] + \frac{1}{2}n_x c^{t-2}\sigma_P^{-1} \mathbf{z}.$$

Substituting $\mathbf{b}_{x,2} = \frac{1}{2}n_x \sigma_P^{-1} \mathbf{z}$, $n_x = TX^{-1}$ and $\mathbf{z}_x = X\mathbf{i}_x T^{-1}$ and defining $S_t = \sum_{i=0}^t c^i$, gives

$$\mathbf{b}_{x,t} = \frac{1}{2}X^{-1}\sigma_P^{-1}[S_{t-3}D + c^{t-2}I](Mi_m, Fi_f)^T.$$

Finally $D(Mi_m, Fi_f)^T = i(M, F)^T$, thus

$$\begin{aligned}
 \mathbf{b}_{x,t} &= \frac{1}{2}\sigma_P^{-1}[S_{t-3}i(MX^{-1}, FX^{-1})^T \\
 &\quad + c^{t-2}(i_m MX^{-1}, i_f FX^{-1})].
 \end{aligned}$$

Appendix 2

$$\begin{aligned}
 P_{i(x),t} &= \prod_{j=t}^{\infty} M_{i(x),j} = P_{i(x),t+1} \cdot M_{i(x),t} \\
 &= P_{i(x),t+1}[D + c^{t-3}A'_{i(x)}LDR],
 \end{aligned}$$

where L and R are diagonal matrices with non-zero elements i_m and i_f for L and c_m and c_f for R . Thus

$$\begin{aligned}
 P_{i(x),t} &= P_{i(x),t+2}[D + c^{t-2}A'_{i(x)}LDR][D + c^{t-3}A'_{i(x)}LDR] \\
 &= P_{i(x),t+2}[D + c^{t-2}A'_{i(x)}LDRD + c^{t-3}A'_{i(x)}DLDR \\
 &\quad + c^{t-2}c^{t-3}A'^2_{i(x)}LDRDLDR].
 \end{aligned}$$

These terms can be simplified by noting $DLD = iD$ and $DRD = cD$, and likewise for other diagonal matrices; and, if terms of order $A'^3_{i(x)}$ or higher are ignored, by similar multiplication and collection of terms

$$\begin{aligned}
 P_{i(x),t} &= D + A'_{i(x)} \left[\left(\sum_{j=t-2}^{\infty} c^j \right) icD + c^{t-3}iDR \right] \\
 &\quad + A'^2_{i(x)} \left[\left(\sum_{j=t-2}^{\infty} c^j c^{j+1} \right) icgD \right. \\
 &\quad + \left(\sum_{j \geq t-2, k > j+1}^{\infty} c^j c^k \right) i^2 c^2 D \\
 &\quad \left. + c^{t-3}c^{t-2}igDR + c^{t-3} \sum_{i=1}^{\infty} c^i i^2 cDR \right].
 \end{aligned}$$

where $g_x = i_x c_x$ and $g = \frac{1}{2}(g_m + g_f)$. If

$$\begin{aligned}
 S_{\infty} &= \sum_{j=0}^{\infty} c^j = (1-c)^{-1}, \quad \text{and} \quad Q_{\infty} = \sum_{j=0}^{\infty} (c^j)^2 \\
 &= (1-c^2)^{-1}
 \end{aligned}$$

$$\begin{aligned}
 P_{i(x),t} &= D + A'_{i(x)} ic^{t-3}D[c^2 S_{\infty} I + R] \\
 &\quad + A'^2_{i(x)} ic^{2t-5}[g + ic^2 S_{\infty}]D[c^3 Q_{\infty} I + R].
 \end{aligned}$$

This can be further simplified by defining terms $q = g + c^2 i S_{\infty}$, and diagonal matrices Θ and Λ with elements θ_m, θ_f and λ_m, λ_f where $\theta_w = c_w + c^2 S_{\infty}$ and $\lambda_w = c_w + c^3 Q_{\infty}$, then

$$P_{i(x),t} = D + A'_{i(x)} ic^{t-3}D\Theta + A'^2_{i(x)} ic^{2t-5}qD\Lambda.$$

Note

$$\frac{1}{2}(\theta_m + \theta_f) = cS_{\infty} \quad \text{and} \quad \frac{1}{2}(\lambda_m + \lambda_f) = cQ_{\infty}.$$

Appendix 3

Listing of terms for ΔF derived for sex x . The constants used in terms are

$$B_{\infty} = \sum_{j=0}^{\infty} c^{2j} 2^{-j} \quad \text{and} \quad H_{\infty} = \sum_{j=0}^{\infty} c^j 2^{-j}.$$

Note

$$\sum_{j=0}^{\infty} c^j S_j 2^{-j} = B_{\infty} H_{\infty}.$$

(i) Squared mean contribution

$$v_0 : (16X)^{-1}. \tag{A 3.1}$$

$$v_1 : \rho_x i^2 (16X)^{-1}. \tag{A 3.2}$$

$$v_2 : \rho_x i^2 c^2 S_{\infty}^2 (16X)^{-1}. \tag{A 3.3}$$

$$v_3 : 2\rho_x (i^2 c S_{\infty} + iq) (16X)^{-1}. \tag{A 3.4}$$

$$v_4 : 2\rho_x ic^2 q Q_{\infty} (16X)^{-1}. \tag{A 3.5}$$

(ii) Binomial sampling

Terms arising from V_0

$$v_0 : (M^{-1} + F^{-1} - 2T^{-1})/32. \tag{A 3.6}$$

$$v_2 : \rho_x B_{\infty} i^2 (\theta_m^2 M^{-1} + \theta_f^2 F^{-1} - (\theta_m^2 + \theta_f^2) T^{-1})/64. \tag{A 3.7}$$

$$v_4 : \rho_x B_{\infty} icq (\lambda_m M^{-1} + \lambda_f F^{-1} - 2\lambda T^{-1})/32. \tag{A 3.8}$$

Terms arising from $A'_{i(x)} V_0$

$$v_3 : \rho_x B_{\infty} H_{\infty} i^2 c (\theta_m M^{-1} + \theta_f F^{-1} - 2\theta T^{-1})/64. \tag{A 3.9}$$

Terms arising from $A'_{i(x)} V_1$

$$v_3 : \rho_x B_{\infty} i (\theta_m i_m M^{-1} + \theta_f i_f F^{-1} - 4qT^{-1})/32. \tag{A 3.10}$$

Terms arising from $A'^2_{i(x)} V_1$

$$v_1 = \rho_x B_{\infty} (g + \frac{1}{2}ic^2 H_{\infty}) (i_m M^{-1} + i_f F^{-1} - 4iT^{-1})/128. \tag{A 3.11}$$

Terms arising from $A'^2_{i(x)} V_2$

$$v_1 = -\rho_x (1 + \frac{1}{4}B_{\infty} (c_m^2 + c_f^2)) (i_m^2 + i_f^2) T^{-1}/64. \tag{A 3.12}$$

(iii) Co-selection of full-sibs in generation 2

Terms arising from C_0

$$v_0 : \rho_x i^2 (16F)^{-1}. \tag{A 3.13}$$

$$v_2 = \rho_x \rho_x i^2 q^2 (16F)^{-1}. \tag{A 3.14}$$

$$v_4 = \rho_x \rho_x i^2 cq (\lambda_m i_m + \lambda_f i_f) (16F)^{-1}. \tag{A 3.15}$$

(iv) Co-selection of sibs in generation $t \geq 3$

Terms arising from C_0

$$v_0: (\rho_m + \rho_f) i^2 (16F)^{-1}. \tag{A 3.16}$$

$$v_0: \rho_m (d-1) i^2 (32F)^{-1}. \tag{A 3.17}$$

$$v_0: -B_\infty i^2 (\rho_m c_m^2 M^{-1} + \rho_f c_f^2 F^{-1}) / 64. \tag{A 3.18}$$

$$v_2: \rho_x (\rho_m + \rho_f) B_\infty i^2 c^2 q^2 (32F)^{-1}. \tag{A 3.19}$$

$$v_2: \rho_x \rho_m (d-1) B_\infty i^2 c^2 q^2 (64F)^{-1}. \tag{A 3.20}$$

$$v_4: \rho_x (\rho_m + \rho_f) B_\infty i^2 c^3 q (\lambda_m i_m + \lambda_f i_f) (32F)^{-1}. \tag{A 3.21}$$

$$v_4: \rho_x \rho_m (d-1) B_\infty i^2 c^3 q (\lambda_m i_m + \lambda_f i_f) (64F)^{-1}. \tag{A 3.22}$$

Terms arising from $A'_{i(x)} C_0$

$$v_3: \rho_x (\rho_m + \rho_f) B_\infty H_\infty i^3 c q (16F)^{-1}. \tag{A 3.23}$$

$$v_3: \rho_x \rho_m (d-1) B_\infty H_\infty i^3 c q (32F)^{-1}. \tag{A 3.24}$$

Appendix 4. Hypergeometric sampling

For sampling without replacement the variance of the hypergeometric distribution is more appropriate than that of the binomial. When selecting W individuals at random from a total T without replacement the variance of a total family contribution for a family of $n_w = TW^{-1}$ members is $[1 - (n_w - 1)(T - 1)^{-1}]$ times the binomial variance. When terms in T^{-1} are neglected this factor is approximately $(1 - W^{-1})$ and the variance is then $(1 - W^{-1}) n_w p(1 - p)$ where $p = WT^{-1}$. In derivation p is a function of family and sex but these deviations are assumed to have no effect on the factor $(1 - W^{-1})$. Terms used to derive (A 3.6) to (A 3.12) inclusive require these corrections. In the derivation of terms arising from the co-selection of sibs the simplifying assumption was made that the joint probability of selecting two sibs of like sex is p^2 where again p is a function of the sex and family. This probability is better approximated by $p^2(1 - W^{-1})$ and the correction factor $(1 - W^{-1})$ is used multiplicatively with the correction $(1 - n_w^{-1})$ which is described in the text. Since equations (A 3.13) to (A 3.24) include both like-sex and unlike-sex contributions these correction terms are not simple factors.

Appendix 5. Contributions from the base population

In their derivation Wray & Thompson (1990) show $\Delta F \approx C_1(2 - C_0)^{-1}$ where C_1 and C_0 are the average of the diagonal elements in the genetic contribution matrices for generations 1 and 0 (i.e. the unselected base) respectively. They argued that since C_0 is small, $\Delta F \approx \frac{1}{2}C_1$. However, C_0 can contribute to significant proportional errors even when ΔF is in the range covered by the mass selection examples in their paper. However, by using the relations $\Delta F \approx \frac{1}{4}C_0$ (J. A. Woolliams, N. R. Wray, unpublished results) this problem may be largely overcome and the need for predicting C_0 avoided.

$\Delta F = C_1(2 - C_0)^{-1} = C_1(2 - 4\Delta F)^{-1}$ since $\Delta F \approx \frac{1}{4}C_0$ and expanding to a quadratic gives $4\Delta F^2 - 2\Delta F +$

$C_1 = 0$. Solving for the lowest root, $\Delta F = \frac{1}{4}(1 - (1 - 4C_1)^{\frac{1}{2}})$.

Substituting ΔF_w for $\frac{1}{2}C_1$, the estimate used by Wray & Thompson (1990), gives

$$\Delta F = \frac{1}{4}(1 - (1 - 8\Delta F_w)^{\frac{1}{2}}) \approx \Delta F_w(1 + 2\Delta F_w),$$

the latter approximation arising from expanding $(1 - 8\Delta F_w)^{\frac{1}{2}}$ as Taylor series. Thus when $\Delta F \approx 0.05$, ΔF_w underestimates ΔF by 10%.

Appendix 6. Sampling of ancestral breeding values

The average contribution from sex x to sex y , $r_{i(x)y\infty}$, is constrained to be $\frac{1}{2}YX^{-1}$. The sampling distributions incorporated ensure the error terms obey this constraint. However, no allowance has so far been made for $\mu_{i(x)y,\infty}$; as modelled the average expected contribution is

$$X^{-1} \sum_{i=1}^X \mu_{i(x)y,\infty} = \frac{1}{2}YX^{-1}(1 + iS_\infty \bar{A}'_{i(x)}).$$

The deviation from the constrained value decreases as X increases since $\text{Var}(\bar{A}'_{i(x)})$ becomes smaller and $\bar{A}'_{i(x)}$ lies more surely close to 0.

Two approaches can be adopted to overcome this problem which are equivalent in first order terms. Firstly, the coefficients $\mu_{i(x)y,t}$ and $\mu_{i(x)wy,t}$ can be recalculated with the discrete distribution of genotypes obtained from selection in generation 1 and, making the same approximations as were made for the truncated normal, it is seen that $A'_{i(x)}$ should be replaced by $A'_{i(x)} - \bar{A}'_{i(x)}$. Alternatively, the $\mu_{i(x)y,t}$ and $\mu_{i(x)wy,t}$ may be regarded as fitness coefficients, and the process can be recalculated using relative fitness to replace absolute fitness, by dividing through by mean fitness coefficients at each generation and transition. The use of $A'_{i(x)} - \bar{A}'_{i(x)}$ requires the estimation of its variance.

$$\text{Var}(A_{i(x)} - \bar{A}_{i(x)}) = \text{Var}(A_{i(x)}) - \text{Var}(\bar{A}_{i(x)}),$$

$$\begin{aligned} \text{Var}(\bar{A}_{i(x)}) &= X^{-1} \text{Var}(A_{i(x)}) + X^{-2} \sum_{i \neq j}^X \text{cov}(A_{i(x)}, A_{j(x)}) \\ &= X^{-1} \text{Var}(A_{i(x)}) + (1 - X^{-1}) \\ &\quad [\text{Prob}(i, j \text{ half-sibs} | i, j \text{ selected}) \\ &\quad \times \text{cov}(\text{half-sibs}) \\ &\quad + \text{Prob}(i, j \text{ full-sibs} | i, j \text{ selected}) \\ &\quad \times \text{cov}(\text{full-sibs})]. \end{aligned}$$

After selection in generation 1, $\text{Var}(A_{i(x)}) = h_0^2(1 - k_x h_0^2)$, $\text{cov}(\text{half-sibs}) \approx \frac{1}{4}h_0^2(1 - k_x h_0^2)$ and $\text{cov}(\text{full-sibs}) \approx \frac{1}{2}h_0^2(1 - k_x h_0^2)$. More precise estimates of covariance, useful for high $h^2 > 0.8$ can be obtained by applying the results of Tallis (1961).

The results of Wray *et al.* (1990) are used to approximate $\text{Prob}(i, j \text{ full-sibs} | i, j \text{ selected})$ when sampling without replacement by standard normal probabilities. For a normally distributed trait undergoing

truncation selection with fixed proportion p_x and a correlation ρ_{FS} between the indices of full-sibs, approximation to the required probability is given by $(n_f - 1)(T - 1)^{-1} \text{Prob}(i, j \geq v | i, j \text{ full-sibs}) p_x^{-2}$, where v is the truncation deviate for infinite populations. $\text{Prob}(i, j \geq v | i, j \text{ full-sibs})$ can be re-expressed as $p_x Q_{FS}$ where $Q_{FS} = \text{Prob}(i, j \geq v | j \geq v \text{ and } i, j \text{ full-sibs})$. Thus the required probability is approximately that for $\rho_{FS} = 0$ scaled by the ratio of the conditional and unconditional probabilities of $i \geq v$. Mendell & Elston (1974) show $Q_{FS} \approx \Phi[(i_x \rho_{FS} - v)(1 - k_x \rho_{FS})^{-\frac{1}{2}}]$ to be an accurate approximation. Thus

$$\begin{aligned} \text{Prob}(i, j \text{ full-sibs} | i, j \text{ selected}) \\ \approx (n_f - 1)(T - 1)^{-1} Q_{FS} p_x^{-1} \approx (n_f - 1) Q_{FS} X^{-1}. \end{aligned}$$

Similarly

$$\begin{aligned} \text{Prob}(i, j \text{ half-sibs} | i, j \text{ selected}) \\ \approx (FM^{-1} - 1) n_f Q_{HS} X^{-1} \end{aligned}$$

Since individuals are in generation 1, $\rho_{FS} = \frac{1}{2}h_0^2$ and $\rho_{HS} = \frac{1}{4}h_0^2$.

References

Bulmer, M. G. (1971). The effect of selection on genetic variability. *American Naturalist* **105**, 201–221.
 Burrows, P. M. (1984a). Inbreeding under selection for unrelated families. *Biometrics* **40**, 357–366.
 Burrows, P. M. (1984b). Inbreeding under selection from related families. *Biometrics* **40**, 895–906.

Hill, W. G. (1972). Effective population size of populations with overlapping generations. *Theoretical Population Biology* **3**, 278–289.
 Latter, B. D. H. (1959). Genetic sampling in a random mating population of constant size and sex ratio. *Australian Journal of Biological Science* **12**, 500–505.
 Lindgren, D. (1991). Optimal utilization of genetic resources. *Forest Tree Improvement* **23**, 49–67.
 Mendell, N. C. & Elston, R. C. (1974). Multifactorial traits: genetic analysis and prediction of recurrence risks. *Biometrics* **30**, 41–57.
 Robertson, A. (1961). Inbreeding in artificial selection programmes. *Genetical Research* **2**, 189–194.
 Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society B* **23**, 223–229.
 Toro, M. A. & Neita, B. (1984). A simple method for increasing the response to artificial selection. *Genetical Research* **44**, 347–349.
 Woolliams, J. A. (1989). Modification to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. *Animal Production* **49**, 1–14.
 Wray, N. R. (1989). Consequences of selection in finite populations with particular reference to closed nucleus herds of swine. Ph.D. Thesis, Edinburgh.
 Wray, N. R. & Thompson, R. (1990). Prediction of rates of inbreeding in selected populations. *Genetical Research* **55**, 41–54.
 Wray, N. R., Woolliams, J. A. & Thompson, R. (1990). Methods for predicting rates of inbreeding in selected populations. *Theoretical Applied Genetics* **80**, 503–512.
 Verrier, E. (1989). Prédiction de l'évolution de la variance génétique dans les populations animales d'effectif limité soumises à sélection. Ph.D. Thesis. INA Paris, Grignon, France.