Theor Appl Genet (1991) 82:329–336

Prediction of response with overlapping generations accounting for multistage selection*

B. L. Harris and A. E. Freeman

Department of Animal Science, Iowa State University, Ames, IA 50011, USA

Received August 13, 1990; Accepted December 18, 1990
Communicated by E.J. Eisen

Summary. A generalization of Hill's equations predicting response to selection is developed that accounts for multiple stage selection in either or both sexes. The method accounts for the flow of genes for animals selected at later stages. This allows for the use of genetic gains from later stages, which explains the reduction in variance due to previous selection. Genetic gains from different selection differentials in each reproductive pathway are incorporated into the equations. The asymptotic response to a single cycle of selection is shown to agree with classical selection theory.

The method is applied to a dairy progeny testing scheme representative of an artificial insemination organization in the USA. Two models were compared: (1) the first model accounted for two-stage selection of males, the first stage being based on pedigree information and the second stage on both pedigree and progeny test information; and (2) the second model assumed single-stage male selection. Selection was based on milk volume, milk fat, and milk protein yields. The predicted asymptotic rates for a single cycle of selection were overestimated by 6% and the cumulative response to continuous selection over 20 years was overestimated by 8% by assuming single-stage male selection.

Key words: Multistage selection – Response – Overlapping generations – Dairy cattle

Introduction

With discrete generations, selected individuals are mated at approximately the same time. The generation interval is the interval between matings in successive generations, or when replacements for the next generation are born. When generations overlap, the replacement of parents by selected offspring approaches a continuous process. The generation interval in a population with overlapping generations can be calculated as the average age of parents at the birth of their selected offspring. The original formulae for predicting the rates of response to selection in overlapping generations were obtained by Dickerson and Hazel (1944) and generalized by Rendel and Robertson (1950). In addition to identifying the intensities of selection and changes in genetic variation due to two-stage selection, Dickerson and Hazel (1944) distinguished between selection in the male and female populations. The asymptotic rate of selection in a continuing breeding program essentially equals the ratio of the mean selection intensity of the parents to the mean age of parents when progeny are born.

A number of authors has computed predictions of response to selection before the asymptotic rates have been reached. Searle (1961) examined the improvement in genetic gain from the use of herd testing and artificial insemination (AI). Alternative methods were described by Brascamp (1973), McClintock and Cunningham (1974), and Hill (1974). The methods are of two kinds. The first, attributed to McClintock and Cunningham (1974), considers the number, genetic contribution (one-half to progeny, one-fourth to grand progeny, etc.), and year of birth of descendants derived from an insemination by a single male. The total number of expressions of an individual genotype up to a specified time is computed. These expressions can be discounted. This method is useful when considering the dissemination of genes from a single animal. The second formulation considers the change in the mean breeding value of each age group in the population. Hill (1974) and Elsen and Mocquot (1974) independently

---

* Journal Paper No. J14146 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa; Project No. 1053

described a matrix method for modeling the response to selection with overlapping generations. Several simplifying assumptions were made, most of which were reasonable, for computing the response to selection from artificial selection programs in large domestic species. The heritabilities and genetic correlations were assumed to be unchanged, and thus the phenotypic and genotypic selection differentials remained constant over time. Single-stage selection was assumed in both sexes. Dentine and McDainel (1987) used a matrix approach to compute short-term genetic gain for milk yield in the USA. The method was deterministic and based on the methods of Hill (1974). Ducrocq and Quaas (1988) described a method for optimizing truncation selection across distributions as well as an application of this procedure, utilizing the matrix methods of Hill (1974) in dairy AI breeding plans. These approaches have assumed that a single stage of selection occurs within the population. If multiple stage selection is not taken into account, then these methods may overestimate the selection response. The purpose of this paper is to outline a procedure that can be used when multistage selection occurs.

Prediction of response with overlapping generations

This discussion will use Hill's (1974) matrix notation. An iterative process can be used to compute the proportion of genes at time t that were derived from animals at time 0. Assume those h male age groups and k female age groups are present in the population. The equation is then (Hill 1974; Eq. 2)

$$\mathbf{m}_t = \mathbf{P}\,\mathbf{m}_{t-1}, \tag{1}$$

where $\mathbf{P}$ is an $(h+k)$ by $(h+k)$ matrix corresponding to the paths of gene transmission via reproduction and aging. $\mathbf{m}_t$ is an $(h+k)\times 1$ column vector, with element $\mathbf{m}_t(i)$ being the proportion of genes in males of age i $(i \leq h)$ in year t, or in females of age $i-h$ $(i>h)$ in year t, originating from males in year 0. A similar equation to Eq. 1 can be derived for the genes originating from females in year 0 in animals in born year t by substituting $\mathbf{f}_t$ for $\mathbf{m}_t$ in Eq. 1. Because an animal has all of its own genes at time $t=0$, the vectors $\mathbf{f}_0$ and $\mathbf{m}_0$ are defined broadly as

$$\mathbf{f}_0 = [\mathbf{0}_{1,k}\ \ 1\ \ \mathbf{0}_{1,h-1}] \quad \text{and} \quad \mathbf{m}_0 = [1\ \ \mathbf{0}_{1,k+h-1}].$$

The blocks of $\mathbf{P}$ correspond to the pathways of genes:

$$\left[\begin{array}{c|c} \text{males to males} & \text{females to males} \\ \hline \text{males to females} & \text{females to females} \end{array}\right].$$

The resulting structure of $\mathbf{P}$ is

$$\left[\begin{array}{cccc|cccc}
p_{1,1} & \cdots & \cdots & p_{1,h} & p_{1,h+1} & \cdots & \cdots & p_{1,h+k} \\
1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
0 & 1 & & \vdots & \vdots & & & \vdots \\
\vdots & & \ddots & \vdots & \vdots & & & \vdots \\
0 & 0 & \cdots\ 1 & 0 & 0 & \cdots & \cdots & 0 \\
\hline
p_{h+1,1} & \cdots & \cdots & p_{h+1,h} & p_{h+1,h+1} & \cdots & \cdots & p_{h+1,h+k} \\
0 & \cdots & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\vdots & & & \vdots & 0 & 1 & & \vdots \\
\vdots & & & \vdots & \vdots & & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & 0 & 0 & \cdots\ 1 & 0
\end{array}\right].$$

The matrix $\mathbf{P}$ combines information on reproduction and aging. The i^{th} element of the first row refers to the proportion of genes transferred to males $(i \leq h)$ or to females $(i>h)$ in year t from males of age i in year $t-1$. Similarly, the i^{th} element of the $h+1$ row refers to the proportion of genes transferred to males $(i \leq h)$ or to females $(i>h)$ in year t from females of age i in year $t-1$. The response to selection is the contribution of genes by reproduction alone. To remove the contribution of genes by aging, a matrix $\mathbf{Q}$ is defined $\mathbf{P}$ by setting the elements of rows 1 and $h+1$ of $\mathbf{P}$ to zero.

The response in year t to one cycle of selection occurring at year 1 is (Hill 1974; Eq. 5):

$$r_t = (\mathbf{P}^t - \mathbf{Q}^t)(\mathbf{m}_0\, g_m + \mathbf{f}_0\, g_f), \tag{2}$$

where g_m and g_f are the genetic gains from male and female selection, respectively. The genetic gain vector $\mathbf{s}$ can be defined as

$$\mathbf{s} = (\mathbf{m}_0\, g_m + \mathbf{f}_0\, g_f). \tag{3}$$

Equation 2 assumes that the females and males are used to breed replacements for both sexes. This assumptions is not normally the case in dairy cattle AI programs. Two additional matrices are required to specify the passage of genes by reproduction to males only ($\mathbf{E\,m}$) and the passage of genes by reproduction to females only ($\mathbf{E\,f}$). These two matrices of dimensions $(h+k)$ by $(h+k)$ are formed from

$$e\,m_{i,j} = \begin{cases} p_{1,j} & \text{for} \quad i=1 \\ 0 & \text{for} \quad i \neq 1 \end{cases} \tag{4}$$

and

$$e\,f_{i,j} = \begin{cases} p_{k+1,j} & \text{for} \quad i=k+1 \\ 0 & \text{for} \quad i \neq k+1 \end{cases}, \tag{5}$$

respectively. To compute the response to selection, a genetic gain vector for each sex, $\mathbf{s}_m$ for male and $\mathbf{s}_f$ for females, needs to be defined as:

$$\mathbf{s}_m = (\mathbf{m}_0\, g_{mm} + \mathbf{f}_0\, g_{fm})$$

$$\mathbf{s}_f = (\mathbf{m}_0\, g_{mf} + \mathbf{f}_0\, g_{ff}),$$

where g_i is the genetic gain for the i^{th} pathway. The response to selection in the first time period is

$$r_1 = \mathbf{E\,m\,s}_m + \mathbf{E\,f\,s}_f; \tag{6}$$

and, in general for $t > 1$ (Hill 1974, Eq. 22),

$$r_t = \mathbf{P}\,r_{t-1} + \mathbf{E\,m\,Q}^{t-1}\,\mathbf{s}_m + \mathbf{E\,f\,Q}^{t-1}\,\mathbf{s}_f. \tag{7}$$

The response to continued selection, whether Eqs. 3 or 6 and 7 are used, ist best computed using

$$R_n = \sum_{t=1}^{n} r_t. \tag{8}$$

Equation 8 is an approximation for response to continuous selection in the short term (less than 20 years). Because the effects of inbreeding are ignored, phenotypic and genetic (co)variances are assumed to be constant over time, and the realized genetic gain is assumed to equal the expected genetic gain. For longer periods, the effects of inbreeding depression must be considered. A simple method that accounts for inbreeding depression is

$$r'_t = r_t - \mathbf{B}\,\mathbf{i}_t, \tag{9}$$

where $\mathbf{i}_t$ is the vector of percentage inbreeding for male and female age groups at time t and $\mathbf{B}$ is an $(h+k)$ by $(h+k)$ matrix which is the product of the regression coefficient of percentage inbreeding on response depression and an identity matrix. The percentage of inbreeding can be computed using methods described by Johnson (1977), James (1978), or Hill (1979); however, these methods may underestimate the rate of inbreeding in some situations (see Wooliams and Wilmut 1989).

Response to selection with overlapping generations accounting for multistage selection

To account for multistage selection in the model, equations distinguishing between genes from animals selected at different stages are required. To simplify the notation, the ideas will first be presented for two-stage selection and then extended to n stage selection. Consider an example in which the same selected males and females are used to breed replacements for both sexes. Assume that second-stage selection occurs at year y for males and at year z for females and that the genetic gains from second-stage male and female selections are g_{2m} and g_{2f}, respectively. Two additional vectors, $\mathbf{m}_{t,2}$ and $\mathbf{f}_{t,2}$, need to be defined, which specify the proportion of genes in male and female progeny originating from animals selected at the second stage and born in year 0. In general, $\mathbf{m}_{0,2}$ is specified with zeros in all elements except the y^{th} element, which is set to 1. Similarly, $\mathbf{f}_{0,2}$ is specified with zeros in all elements, except the element corresponding to the z^{th} age group of females. In general,

$$\mathbf{f}_{0,2} = [\mathbf{0}_{1,k+z-1}\;\;1\;\;\mathbf{0}_{1,h-z}]$$

and

$$\mathbf{m}_{0,2} = [\mathbf{0}_{1,y-1}\;\;1\;\;\mathbf{0}_{1,h+k-y-1}]. \tag{10}$$

Three genetic gain vectors can be specified, one for first-stage selection, one for second-stage male selection, and one for second-stage female selection:

$$s_1 = (\mathbf{m}_0\,g_m + \mathbf{f}_0\,g_f)$$

$$s_{2m} = (\mathbf{m}_{0,2}\,g_{2m})$$

$$s_{2f} = (\mathbf{f}_{0,2}\,g_{2f}).$$

The reduction in variance due to first-stage selection should be taken into account when computing the values for genetic gain for second-stage selection. With single-stage selection, the genes from selected individuals are available for dissemination through the population before the selected individuals become parents. The general passage of genes from second-stage selection is not different from that of genes from first-stage selection, with the exception that the passage has a time lag of y or z years for male or female second-stage selections, respectively. Thus, it is possible to use the general form of Eq. 2 to compute the response to selection from a second-stage selection. The general form of the equation for second-stage male selection is:

$$r_t^{2m} = (\mathbf{P}^{t-y+1} - \mathbf{Q}^{t-y+1})\,s_{2m} \quad t \geq y \tag{11}$$

and for female second-stage selection:

$$r_t^{2f} = (\mathbf{P}^{t-z+1} - \mathbf{Q}^{t-z+1})\,s_{2f} \quad t \geq z. \tag{12}$$

The total response to a single cycle of selection, assuming $y > z$, is an extension of Eq. 4 that incorporates Eqs. 11 and 12:

$$r_t = \begin{cases} (\mathbf{P}^t - \mathbf{Q}^t)\,s_1 & t < z \\ (\mathbf{P}^t - \mathbf{Q}^t)\,s_1 + (\mathbf{P}^{t-z} - \mathbf{Q}^{t-z+1})\,s_{2f} & t \geq z \\ (\mathbf{P}^t - \mathbf{Q}^t)\,s_1 + (\mathbf{P}^{t-z+1} - \mathbf{Q}^{t-z+1})\,s_{2f} & \\ \qquad + (\mathbf{P}^{t-y+1} - \mathbf{Q}^{t-y+1})\,s_{2m} & t \geq y. \end{cases} \tag{13}$$

When there is n-stage selection in males and m-stage selection in females, the general form of Eq. 13 holds. In Eq. (13), the quantity r_t^{2m} (the response for males from second-stage selection) is added to the first-stage selection responses for both males and females and to the second-stage selection response for females during the year the male second-stage selection occurred. If a third-stage selection of males had taken place, the response from this selection would be added to the cumulative responses from preceding years for both males and females in the year it occurred and in subsequent years. Thus, the general form of Eq. 13 for n- and m-stage selection occurring at year y_n and z_n ($z_n < y_n$) in males and females, respectively, is

$$r_t=\begin{cases}(\mathbf{P}^t-\mathbf{Q}^t)\,s_1 & t<z\\ (\mathbf{P}^t-\mathbf{Q}^t)\,s_1+(\mathbf{P}^{t-z+1}-\mathbf{Q}^{t-z+1})\,s_{2f} & t\geq z\\ (\mathbf{P}^t-\mathbf{Q}^t)\,s_1+(\mathbf{P}^{t-z+1}-\mathbf{Q}^{t-z+1})\,s_{2f} & \\ \qquad +(\mathbf{P}^{t-y+1}-\mathbf{Q}^{t-y+1})\,s_{2m} & t\geq y\\ \vdots & \vdots\\ (\mathbf{P}^t-\mathbf{Q}^t)\,s_1+,\ldots,+(\mathbf{P}^{t-z_n+1}-\mathbf{Q}^{t-z_n+1})\,s_{nf} & \\ \qquad +(\mathbf{P}^{t-y_n+1}-\mathbf{Q}^{t-y_n+1})\,s_{nm} & t\geq y_n.\end{cases}\tag{14}$$

The same principles are used to define the response to selection for multistage selection when different selection intensities are used to breed replacements for each sex. The only difference is the need to extend the four different genetic gain vectors, namely, $\mathbf{s}_{2mm}$, $\mathbf{s}_{2mf}$, $\mathbf{s}_{2fm}$, and $\mathbf{s}_{2ff}$, in the case of two-stage selection, where

$$s_{2mm}=(\mathbf{m}_{0,2}\,g_{2mm})$$

$$s_{2mf}=(\mathbf{m}_{0,2}\,g_{2fm})$$

$$s_{2fm}=(\mathbf{f}_{0,2}\,g_{2mf})$$

$$s_{2ff}=(\mathbf{f}_{0,2}\,g_{2ff}),$$

where s_{2mm}, s_{2mf}, s_{2fm}, and s_{2ff} correspond to the genetic gain from second-stage selection for each of the four reproductive pathways. The response to second-stage selection can be defined directly from the single-stage Eqs. 4 and 5, allowing for the time lag between stage 1 and stage 2 selections and for the appropriate genetic gain vectors. The response to the second-stage male selection is

$$\begin{aligned} r_t^{2m}&=\mathbf{E}\,\mathbf{m}\,\mathbf{s}_{mm}+\mathbf{E}\,\mathbf{f}\,\mathbf{s}_{mf} && t=y\\ r_t^{2m}&=\mathbf{P}\,r_{t-1}^{2m}+\mathbf{E}\,\mathbf{m}\,\mathbf{Q}^{t-y}\,\mathbf{s}_{mm}+\mathbf{E}\,\mathbf{f}\,\mathbf{Q}^{t-y}\,\mathbf{s}_{mf} && t>y,\end{aligned}\tag{15}$$

with the selection occurring at year y. Similar equations can be derived for second-stage female selection by replacing y (the year second-stage female selection occurs) with z, $\mathbf{s}_{2ff}$ with $\mathbf{s}_{2mf}$, and $\mathbf{s}_{2fm}$ with $\mathbf{s}_{2mm}$. The total response to a single cycle of selection, assuming two-stage selection only, is

$$r_t=\begin{cases} r_t & 1\leq t<z\\ r_t+r_t^{2f} & z\geq t<y\\ r_t+r_t^{2f}+r_t^{2m} & y\geq t.\end{cases}\tag{16}$$

Equation 16 can easily be extended to n-stage selection. Assume that male selection occurs after female selection for each stage, with stage t occurring at years z_t and y_t and with n stage occurring in years q and p for females and males, respectively. Then the response to selection is

$$r_t=\begin{cases} r_t & 1\ \leq t\leq z_2\\ r_t+r_t^{2f} & z_2\leq t<y_2\\ r_t+r_t^{2f}+r_t^{2m} & y_2\leq t<z_3\\ \vdots & \\ r_t+r_t^{2f}+r_t^{2m}+r_t^{3f}+r_t^{3m}+,\ldots,+r_t^{pm} & y_p\leq t\leq z_q\end{cases}\tag{17}$$

To compute the cumulative response from n multiple stage selection Eq. 15 or 17 are used in Eq. 8. It has been assumed that the elements of **P** are constant over time. Ducrocq and Quaas (1988) suggested that the genetic merit of different age groups be taken into account when deriving the elements of **P**. This results in a **P** matrix that varies over time. The equations derived here can easily account for a **P** matrix that varies with time. Extensions to Eqs. 14 and 15 are given in the Appendix.

Asymptotic response to selection

The total response achievable from a single cycle of selection, using the equations developed in the previous sections, can be shown to agree with the methods of Dickerson and Hazel (1944) and Rendel and Robertson (1950), which are commonly used. To compute the asymptotic response, the limit, $\lim_{t\to\infty} r_t$, must be evaluated. Two general results are required:

$$\lim_{t\to\infty}\mathbf{Q}^t=\mathbf{0},\tag{18}$$

when t exceeds h or k, and (Hill 1974; Eqs. 9 and 11)

$$\lim_{t\to\infty}\mathbf{P}^t=\frac{\mathbf{1}\,\mathbf{v}'}{2L},\tag{19}$$

where L is the average generation interval (i.e., $(L_{mm}+L_{mf}+L_{fm}+L_{ff})/4$), where **1** is a vector of 1's with dimensions 1 and $h+k$ and **v** is a vector of the expected gene contribution of the i^{th} age group. The vector **v** of dimension $h+k$ is defined from the elements of the matrix **P** (Hill 1974) as

$$v_i=\begin{cases}\sum\limits_{j=1}^{h}(p_{1,j}+p_{h+1,j}) & i=1,\ldots,h\\ \sum\limits_{j=1}^{h+k}(p_{1,j}+p_{h+1,j}) & i=h+1,\ldots,h+k\end{cases}.$$

The asymptotic response to single-stage selection with the same selected males and females breeding replacements for both sexes is (Hill 1974; Eq. 12)

$$\lim_{t\to\infty} r_t=\lim_{t\to\infty}(\mathbf{P}^t-\mathbf{Q}^t)\,\mathbf{s}=\frac{\mathbf{1}\,\mathbf{v}'\,\mathbf{s}}{2L},\tag{20}$$

which is identical to Rendel and Robertson's (1950) formulation with $G_{ff}=G_{fm}$ and $G_{mm}=G_{mf}$. With multistage selection, the asymptotic response is

$$\begin{aligned}\lim_{t\to\infty} r_t&=\lim_{t\to\infty}(\mathbf{P}^t-\mathbf{Q}^t)\,\mathbf{s}_1+,\ldots,+(\mathbf{P}^{t-z_n+1}-\mathbf{Q}^{t-z_n+1})\,s_{nf}\\ &\qquad+(\mathbf{P}^{t-y_n+1}-\mathbf{Q}^{t-y_n+1})\,s_{nm}\\ &=\frac{\mathbf{1}\,\mathbf{v}'(\mathbf{s}_1+,\ldots+\mathbf{s}_{nf}+\mathbf{s}_{nm})}{2L},\end{aligned}\tag{21}$$

which, for two-stage selection, is identical to the formulation given in Dickerson and Hazel (1944), using Rendel and Robertson's (1950) notation, with $G_{ff}=G_{fm}$ and $G_{mm}=G_{mf}$. Equation 21 can be written in terms of Rendel and Robertson's (1950) formula for two-stage selection as

$$\lim_{t\to\infty} r_t=\frac{G_m^1+G_f^1+p_m G_m^2+p_f G_f^2}{L_{mm}+L_{mf}+L_{fm}+L_{ff}} \tag{22}$$

where G_j^i is the expected genetic gain from the j^{th} sex and i^{th} stage of selection and p_i is the proportion of the i^{th} sex bred from animals selected at the second stage.

When different selection differentials are used for breeders of males and females, the asymptotic response is the same for both sexes. If $\mathbf{v}_m$ and $\mathbf{v}_f$ are vectors defined as the reproductive values of male and female breeders, respectively, these are given (Hill 1974; Eq. 25) as

$$v_{m,i}=\begin{cases}\sum_{j=1}^{h} p_{1,j}) & i=1,\ldots,h\\ \sum_{j=1}^{h+k} p_{1,j} & i=h+1,\ldots,h+k\end{cases}$$

and

$$v_{f,i}=\begin{cases}\sum_{j=1}^{h} p_{h+1,j} & i=1,\ldots,h\\ \sum_{j=1}^{h+k} (p_{h+1,j} & i=h+1,\ldots,h+k.\end{cases}$$

The asymptotic response for single stage selection is (Hill 1974; Eq. 24)

$$\lim_{t\to\infty} r_t=\left[\frac{\mathbf{v}_m'\mathbf{s}_m+\mathbf{v}_f'\mathbf{s}_f}{2L}\right]\mathbf{1}. \tag{23}$$

Equation 23 is identical to Rendel and Robertson's (1950) formula for response to selection. Extension of this result to n-stage selection is straightforward and can be derived directly from Eq. 23 as

$$\lim_{t\to\infty} r_t=\left[\frac{\mathbf{v}_m'(\mathbf{s}_m+,\ldots,+\mathbf{s}_{n,mf}+\mathbf{s}_{n,mm})+\mathbf{v}_f'(\mathbf{s}_f+,\ldots,+\mathbf{s}_{n,fm}+\mathbf{s}_{n,ff})}{2L}\right]\mathbf{1}. \tag{24}$$

This can be written in the form of Dickerson and Hazel's (1944) response to two-stage selection [using Rendel and Robertson's (1950) notation] as

$$\lim_{t\to\infty} r_t=\frac{G_{mm}^1+G_{ff}^1+G_{mf}^1+G_{mf}^1+p_{mm}G_{mm}^2+p_{ff}G_{ff}^2+p_{mf}G_{mf}^2+p_{fm}G_{fm}^2}{L_{mm}+L_{mf}+L_{fm}+L_{ff}}. \tag{25}$$

Application: modelling genetic gain from dairy progeny testing scheme

In the prediction of genetic gain from a dairy progeny testing scheme, three reproductive pathways under direct control of the breeding scheme are of interest

(1) male to male;
(2) male to female;
(3) female to male.

The female-to-female pathways is under the control of individual producers, and hence the selection decisions made may reflect a variety of criteria. In this study a population of the size and structure described in Table 1 was used. These parameters were chosen to be representative of a USA sire proving scheme serving the Holstein breed. Table 2 gives the proportion of males and females that are assumed to breed future males and females by age group. These values were used to form the matrix **P**. In this example **P** is constant over time. Two models are compared the first (model 1) accounting for two-stage selection in males and the second (model 2) assuming one-stage selection for males and failing to account for two-stage selection. Both models have identical population structures, dam selection, and numbers of progeny tested males selected.

Table 1. Population parameters

Number of young sires selected	70
Proportion of young sires bred to females in the population	6%
Total cow population size	300,000
Selection intensities	
1. Young sires (first stage, top 3.00%), model 1	2.268
2. Bull dams (top 5.00%)	2.063
3. Sires of males (10 from 70), model 1	1.553
4. Sires of females (25 from 70), model 1	1.032
5. Sires of males (top 0.43%), model 2	2.947
6. Sires of females (top 1.00%) model 2	2.665
Average number daughters for sires, grand, and maternal grand sires	500
Average number of records for bull dam's dam	5.00
Average number of records for young sire's dam	1.90
Average number of daughters for progeny test of young sires	50

The selection of males in a progeny testing scheme generally occurs in two stages. First, the selection of young males for progeny testing from the population of eligible young sires occurs. The population of eligible young sires can be considered as all young sires produced from all possible matings of parents that meet the proge-

Table 2. Proportion of animals used as breeders in the four reproductive pathways for the numerical example of a progeny testing scheme

Age	Male to male	Female to male	Male to female	Female to female
1	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000	0.0300	0.0000
4	0.0000	0.1000	0.0000	0.1500
5	0.0000	0.2000	0.0000	0.1250
6	0.0000	0.2000	0.0000	0.1000
7	0.0750	0.0000	0.0750	0.0750
8	0.3000	0.0000	0.1500	0.0250
9	0.1250	0.0000	0.1000	0.0250
10	0.0000	0.0000	0.1000	0.0000
11	0.0000	0.0000	0.0450	0.0000

Table 3. Means ($\bar{x}$), genetic standard deviations (SD), and heritabilities (h^2), genetic correlations (r_g), phenotypic correlations (r_p), repeatabilities (R), and economic weights (E.W.) for milk, fat, and protein yield (kg)

	$\bar{x}$	SD	h^2	r_g	r_p	R	E.W.
Milk volume	8,948	905.77	0.26			0.46	0.0274
Milk fat	324	32.30	0.24			0.38	1.4800
Milk protein	290	25.16	0.21			0.42	1.4300
Milk–Fat				0.58	0.82		
Milk–Protein				0.86	0.93		
Fat–Protein				0.70	0.85		

Table 4. Genetic gain (kg) for the three traits and the asymptotic response to selection calculated using Eq. 24

	Model 1		
	Milk volume	Milk fat	Milk protein
Male-to-male first stage	1,216.85	29.52	30.82
Male-to-female first stage	1,216.85	29.52	30.82
Male-to-male second stage	972.09	22.40	23.64
Male-to-female second stage	735.48	16.94	17.89
Female-to-male first stage	1,358.13	32.17	34.37
Asymptotic rate	200.05	4.75	5.00
	Model 2		
	Milk volume	Milk fat	Milk protein
Male-to-male first stage	2,336.91	55.67	58.51
Male-to-female first stage	2,140.41	50.43	52.91
Female-to-male first stage	1,358.13	32.17	34.37
Asymptotic rate	213.91	5.06	5.43

ny schemes criteria (e.g., progeny-tested males, registered females with adequate pedigree information and adequate dairy conformation scores). The second stage is selection of males from the progeny-tested young sires that are to be used widely in the population. The selection of bull dams is single-stage selection. It will be assumed that for model 1 the first stage of selection for sires is based on pedigree information (grand sire, maternal grand sire, sire, and dam), and that second-stage selection is based on the same pedigree information, as well as on progeny test information. The genetic gains from sire selection (model 1) are computed using a two-stage selection index (Cunningham 1975), which accounts for the reduction in variance from first-stage selection when computing genetic gain for the second stage. Sire selection in model 2 is single stage, based on the same pedigree information as well as on progeny test information. Bull dam selection is assumed to be based on the pedigree information, as are selections of young sires and their production. Genetic gains from bull dams and sires in model 2 are computed using a single-stage selection index (Hazel 1943). In the selection index three traits – milk volume, milk fat, and milk protein production – were considered. Heritabilities, genetic and phenotypic correlations, and repeatabilities for the Holstein Freisian breed have been reported recently by Manfredi et al. (1984), de Jager and Kennedy (1987), Schutz et al. (1990), and Welper et al. (1989). The estimates used in this study were derived from these reports and are given in Table 3. Using the appropriate selection intensities from Table 1, the expected gains in the three traits for the three reproductive pathways are given in Table 4. The economic weights used are from the USDA predicted transmitting abilities dollars protein (PTA$P) index; these are given in Table 3.

To predict the response to selection over time with two stages (model 1), Eq. 17 is used. More specifically, if two-stage male selection occurs at the start of year 7 (when young sires are 6 years old), then the equations are

$$
\begin{aligned}
&r_1 = \mathbf{E\,m\,s}_m + \mathbf{E\,f\,s}_f && \\
&r_t = \mathbf{P}\,r_{t-1} + \mathbf{E\,m\,Q}^{t-1}\mathbf{s}_m + \mathbf{E\,f\,Q}^{t-1}\mathbf{s}_f && \qquad (27) \\
&r_t^{2m} = \mathbf{E\,m\,s}_{2mm} + \mathbf{E\,f\,s}_{2mf} && t=7 \\
&r_t^{2m} = \mathbf{P}\,r_{t-1}^{2m} + \mathbf{E\,m\,Q}^{t-7}\mathbf{s}_{2mm} + \mathbf{E\,f\,Q}^{t-7}\mathbf{s}_{2mf} && t>7 \\
&r_t = \begin{cases} r_t & 1 \le t < 7 \\ r_t + r_t^{2m} & t \ge 7 \end{cases} &&
\end{aligned}
$$

P, **E m**, and **E f** can be formed from Table 2. The genetic gain vectors $\mathbf{s}_m$, $\mathbf{s}_f$, $\mathbf{s}_{2mf}$, and $\mathbf{s}_{2mm}$ have dimension 22, with the following form (values for milk are from Table 4)

(1) $s'_m = [1216.48 \quad \mathbf{0}_{1\times 10} \quad 1216.48 \quad \mathbf{0}_{1\times 10}]$,

(2) $s'_f = [735.48 \quad \mathbf{0}_{1\times 21}]$,

(3) $s'_{2mm} = [\mathbf{0}_{1\times 6} \quad 972.09 \quad \mathbf{0}_{1\times 15}]$, and

(4) $s'_{2mf} = [\mathbf{0}_{1\times 6} \quad 735.48 \quad \mathbf{0}_{1\times 15}]$.

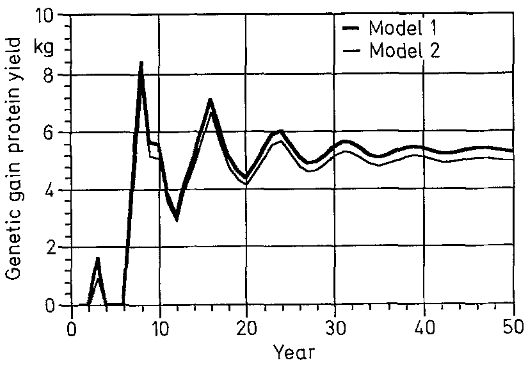

Fig. 1. Response to a single cycle of selection for protein yield

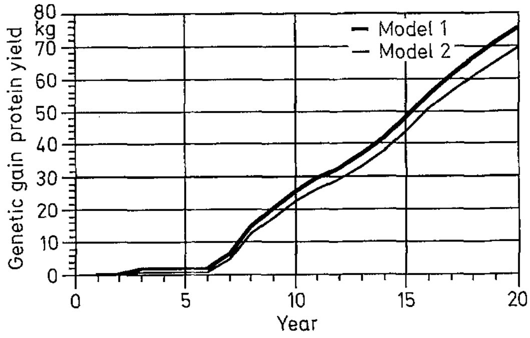

Fig. 2. Response to cumulative selection for protein yield

The response to selection for model 2 is computed using Eqs. 6 and 7, where the appropriate genetic gains vectors are (values for milk are from Table 4):

(1) $s'_m = [2366.91 \quad \mathbf{0}_{1\times10} \quad 2140.41 \quad \mathbf{0}_{1\times10}]$,

(2) $s'_f = [735.48 \quad \mathbf{0}_{1\times21}]$.

The cumulative response to continuous cycles of selection for both models is computed using Eq. 8. The models were run for 20 years and the results for both responses to a single cycle of selection and cumulative responses to continuous selection for the three traits and two models are given in Table 5.

The responses to selection for the three traits and both models show considerable oscillation in the first few years before the asymptotic rates are reached. This is clearly seen in Fig. 1 for protein yield and in Table 5 for all traits. If model 1 is allowed to run until the response from a single cycle of selection stabilizes, the following values are obtained: 200.5, 4.75, and 5.00 for milk, fat, and protein, respectively. These values are identical to those calculated (Table 4) in the terms of classical selection theory. Assuming single-stage selection for males (model 2), the asymptotic rates obtained are 213.74, 5.04, and 5.31 for milk, fat, and protein, respectively, which overestimates the predicted rates accounting for two-stage selection by at least 6%.

Table 5. Response to a single cycle of selection for the three production traits for models 1 and 2 (all amounts in kg)

Year	Model 1			Model 2		
	Milk	Fat	Protein	Milk	Fat	Protein
1	0.0	0.00	0.00	0.00	0.00	0.00
2	0.0	0.00	0.00	0.00	0.00	0.00
3	36.5	0.89	0.92	64.21	1.51	1.59
4	0.0	0.00	0.00	0.00	0.00	0.00
5	0.0	0.00	0.00	0.00	0.00	0.00
6	0.0	0.00	0.00	0.00	0.00	0.00
7	156.0	3.71	3.89	174.13	4.10	4.31
8	305.6	7.27	7.63	337.01	7.93	8.34
9	207.0	4.93	5.17	228.39	5.37	5.65
10	203.0	4.83	5.07	224.38	5.28	5.55
11	142.3	3.38	3.55	155.64	3.66	3.85
12	115.4	2.74	2.89	122.81	2.90	3.06
13	159.5	3.79	4.01	166.14	3.93	4.16
14	193.4	4.59	4.86	202.83	4.78	5.06
15	233.3	5.54	5.84	248.28	5.85	6.17
16	266.3	6.32	6.65	285.73	6.73	7.09
17	225.7	5.36	5.64	243.57	5.73	6.04
18	192.3	4.57	4.80	207.45	4.88	5.14
19	173.1	4.11	4.33	185.59	4.37	4.61
20	164.9	3.92	4.13	175.15	4.13	4.36

The cumulative response to protein yield for both models is illustrated in Fig. 2. Calculating the cumulative response to selection from model 2 rather than model 1 results in cumulative responses for 20 years of 3,022, 71, and 75 kg compared to 2,774, 66, and 69 kg for milk, fat, and protein, respectively, which would overestimate the response by at least 8% for all traits. The extent of overestimation depends upon the nature of selection and the generation intervals in the four reproductive pathways. Because multistage selection increases the generation interval, which increases the degree of overestimation from assuming stage, the breeding equations presented should be used to compute the response to selection. The overestimation of response results from the failure to account for the flow of genes from animals selected in different stages and reduction of variance caused by prior selection in the estimation of genetic gains. The methods of Cunningham (1975) used here to compute genetic gains for two-stage sire selection are only valid for two stages of selection. After two stages of selection, the distribution of selected individuals is no longer normal. For situations where selection involves more than two stages, the effect of the departure from normality on the computation of genetic gains from later stages will require consideration when applying the breeding equations.

Appendix

Equations 14 and 15 are presented assuming **P** is constant over time. In situations where **P** is known a priori to vary with time, or where truncation selection across distributions of individuals is applied (see Ducrocq and Quaas 1988), a simple extension to these equations can be used. Define $\mathbf{P}^{*t}$ as:

$$\mathbf{P}^{*t}=\mathbf{P}_t\times\mathbf{P}_{t-1}\times,..,\times\mathbf{P}_1, \tag{A 1}$$

where $\mathbf{P}_t$ is the $\mathbf{P}$ matrix at time t. The extension to Eq. 14 is:

$$r_t=\begin{cases}(\mathbf{P}^{*t}-\mathbf{Q}^t)\,s_1 & t<z\\ (\mathbf{P}^{*t}-\mathbf{Q}^t)\,s_1+(\mathbf{P}^{*t-z+1}-\mathbf{Q}^{t-z+1})\,s_{2f} & t\geq z\\ (\mathbf{P}^{*t}-\mathbf{Q}^t)\,s_1+(\mathbf{P}^{*t-z+1}-\mathbf{Q}^{t-z+1})\,s_{2f} & \\ \qquad\qquad +(\mathbf{P}^{*t-y+1}-\mathbf{Q}^{t-y+1})\,s_{2m} & t\geq y\\ \vdots & \vdots\\ (\mathbf{P}^{*t}-\mathbf{Q}^t)\,s_1+,\ldots,+(\mathbf{P}^{*t-z_n+1}-\mathbf{Q}^{t-z_n+1})\,s_{nf} & \\ \qquad\qquad +(\mathbf{P}^{*t-y_n+1}-\mathbf{Q}^{t-y_n+1})\,s_{nm} & t\geq y_n.\end{cases} \tag{A 2}$$

The extension to Eq. 15 requires $\mathbf{P}$, $\mathbf{E\,m}$, and $\mathbf{E\,f}$ to be indexed in time:

$$\begin{aligned} r_t^{2m}&=\mathbf{E\,m}_t\,\mathbf{s}_{mm}+\mathbf{E\,f}_t\,\mathbf{s}_{mf} && t=y\\ r_t^{2m}&=\mathbf{P}_t\,r_{t-1}^{2m}+\mathbf{E\,m}_t\,\mathbf{Q}^{t-y}\,\mathbf{s}_{mm}+\mathbf{E\,f}_t\,\mathbf{Q}^{t-y}\,\mathbf{s}_{mf} && t>y. \end{aligned} \tag{A 3}$$

Equations A 2 and A 3 can be substituted for Eqs. 14 and 15 and used in Eqs. 17 and 18.

References

Brascamp EW (1973) Model calculations concerning economic optimization of A. I. breeding cattle. Z Tierz Zuechtungsbiol 90:1–15

Cunningham EP (1975) Multi-stage index selection. Theor Appl Genet 46:55–61

Dentine MR, McDainel BT (1987) Expected genetic gain for milk yield in diary cattle. Theor Appl Genet 74:753–757

Dickerson GE, Hazel LN (1944) Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. J Agric Res 69:459–476

Ducrocq V, Quaas RL (1988) Prediction of genetic response to truncation selection across generations. J Dairy Sci 71:2543–2553

Elsen JM, Mocquot JC (1974) Recherches pour une rationalisation technique des schémas de séléction des bovins et ovins. Bull Tech Dep Génét Anim Inst Natl Rech Agron 17, Jouy-en-Josas, France

Hazel LN (1943) The genetic basis for constructing selection indexes. Genetics 28:476–490

Hill WG (1974) Prediction and evaluation of response to selection with overlapping generations. Anim Prod 18:174–139

Hill WG (1979) A note on effective population size with overlapping generations. Genetics 92:317–322

Jager D de, Kennedy BW (1987) Genetic parameters of milk yield and composition and their relationships with alternative breeding goals. J Dairy Sci 70:1258

James JW (1978) Effective population size in open nucleus breeding schemes. Acta Agric Scand 28:388–392

Johnson DL (1977) Inbreeding in populations with overlapping generations. Genetics 87:581–591

Manfredi EJ, Everett EW, Searle SR (1984) Phenotypic and genetic statistics of milk components and two measures of somatic cell concentration. J Dairy Sci 67:2028–2033

McClintock AE, Cunningham EP (1974) Selection in dual purpose cattle populations: effect of beef crossing and cow replacement rates. Ann Genet Sel Anim 6:227–235

Rendel JM, Robertson A (1950) Estimation of genetic gain in milk yield by selection in a close herd of dairy cattle. J Genet 50:1–8

Schutz MM, Hansen, LB, Steuernagel GR, Reneau JK, Kuck AL (1990) Genetic parameters for somatic cells, protein, and fat in milk of Holsteins. J Dairy Sci 73:494–502

Searle SR (1961) Estimating herd improvement from selection programmes. J Dairy Sci 44:1103–1112

Welper RD, Freeman AE, Thelen AJ (1989) Genetic parameters for production traits of holsteins, including lactose. Presented at ADSA and ASA joint meeting, University of Kentucky, Lexington

Wooliams JA, Wilmut I (1989) Embryo manipulation in cattle breeding and production. Anim Prod 48:3