Long-term Genetic Contributions

Prediction of Rates of Inbreeding and Genetic Gain in Selected Populations

Promotoren: dr. ir. E. W. Brascamp
Hoogleraar Fokkerij en Toegepaste Genetica
Wageningen Universiteit

dr. ir. J. A. M. Van Arendonk
Persoonlijk hoogleraar bij de leerstoelgroep Fokkerij en Genetica
Wageningen Universiteit

Co-promotor: dr. J. A. Woolliams, MA, Dip. Math. Stats., DSc.
Principal Investigator
Roslin Institute (Edinburgh), Roslin, United Kingdom

Piter Bijma

Long-term Genetic Contributions

Prediction of Rates of Inbreeding and Genetic Gain in Selected Populations

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
dr. C. M. Karssen,
in het openbaar te verdedigen
op dinsdag 27 juni 2000
des namiddags te half twee in de Aula.

Long-term genetic contributions: prediction of rates of inbreeding and genetic gain in selected populations

Bijma, Piter

Het in dit proefschrift beschreven onderzoek is financieel gesteund door de Technologiestichting STW en is begeleid door het Gebied Aard- en Levenswetenschappen (ALW).

ISBN 90-5808-241-5

Printed by Universal Press, Veenendaal, The Netherlands.

Bijma, P. Long-term genetic contributions: prediction of rates of inbreeding and genetic gain in selected populations. This dissertation focuses on the prediction of long-term genetic contributions, rates of inbreeding and rates of gain in artificially selected populations. The long-term genetic contribution (r_i) of ancestor i born at time t_1, is defined as the proportion of genes from i that are present in individuals in generation t_2 deriving by descent from i, where $(t_2 - t_1) \rightarrow \infty$. The long-term genetic contribution of an individual was predicted by linear regression on the selective advantage of the individual. With overlapping generations, long-term genetic contributions were predicted using a modified gene flow approach. A novel definition of generation interval was introduced, which states that the generation interval is the length of time in which long-term genetic contributions sum to unity. It was shown that the rate of inbreeding is proportional to the sum of squared of expected long-term genetic contributions and that the rate of genetic gain is proportional to the sum of cross products of long-term genetic contributions and Mendelian sampling terms. Accurate predictions of rates of inbreeding were obtained for populations with discrete or overlapping generations undergoing either mass selection or selection on Best Linear Unbiased Prediction of breeding values. The method was applied to crossbreeding systems, which showed that the use of crossbred information may increase the rate of genetic gain, but measures to restrict the rate of inbreeding are required.

Ph.D. thesis, Animal Breeding and Genetics Group, Department of Animal Sciences, Wageningen University, P. O. Box 338, 6700 AH Wageningen, The Netherlands.

Voorwoord

Dit proefschrift is tot stand gekomen bij de Leerstoelgroep Fokkerij en Genetica van Wageningen Universiteit. Van degenen die aan dit proefschrift hebben bijgedragen wil ik een aantal personen met name noemen.

Ten eerste, Johan, bedankt voor je vertrouwen, je altijd positieve instelling en de ruimte die je me hebt gegeven om het onderzoek naar mijn eigen inzicht uit te voeren. John, you have contributed very substantially to the scientific content of this thesis. I learned a lot during my visits to the Roslin Institute, both technically as well as important aspects of good scientific practice. I learned to respect you as an excellent scientist and a good friend. Pim, bedankt voor de kritische kijk op mijn artikelen en je prikkelende opmerkingen tijdens de koffiepauzes. Verder wil ik Ab en Theo bedanken voor de kritische opmerkingen op mijn artikelen.

De gebruikerscommissie, bestaande uit Gerard Albers, Sijne van der Beek, Huub Huizenga en Egbert Knol, bedank ik voor de kritische opmerkingen, de link met de praktijk en de mij gegeven onderzoeksvrijheid. Ik ben er van overtuigd dat dit proefschrift, in combinatie met het door Marc Rutten ontwikkelde computerprogramma, recht doet aan zowel de theoretische als de praktische doelstellingen van het oorspronkelijke onderzoeksproject.

Verder wil ik mijn leerstoelgroepsgenoten en mede aio's bedanken voor de plezierige werkomgeving. Jascha, jij bent de perfecte kamergenoot. Marco, betanke foar de krityske oanmerkings en foar it trochlêzen fan in tal fan myn artikels; ik sil tenei goed om de komma's tinke. Anna, bedankt voor de inspirerende discussies over inteelt. Ricardo Pong-Wong, many thanks for the pleasant discussions during my visits to the Roslin Institute. Verder wil ik mijn huisgenoten bedanken voor de plezierige leefomgeving, die zonder meer heeft bijgedragen aan de kwaliteit van dit proefschrift.

Heit en mem en fierdere famylje, betanke foar jimme stipe en foar de frijheid dy't ik altyd krigen haw om it ûnderwiis te folgjen dat ik sels woe. En tenslotte, Liesbeth, bedankt.

Piter

Contents

Introduction

In animal breeding, tools to evaluate breeding schemes in the short-term are well established. Best Linear Unbiased Prediction (HENDERSON 1963, 1973, 1975, 1976) is widely used to estimate breeding values and selection index theory (HAZEL, 1943) is the common tool to evaluate breeding programs. Response to selection on Best Linear Unbiased Prediction of breeding values can be predicted accurately, by including estimated breeding values of parents in the selection index (WRAY and HILL, 1989) and accounting for reduction of the genetic variance due to the Bulmer effect (BULMER, 1971).

Apart from the Bulmer effect, little attention has been paid to the long-term aspects of selection in animal breeding theory. When the selected trait is heritable, selection in the current generation will favor offspring of superior parents of previous generations, thus inducing a certain degree of selection between families of previous generations. Selection, therefore, reduces the effective number of grandparents and earlier ancestors, which results in decreased genetic variation in the long term. This process has not explicitly been modeled in animal breeding theory. For example, selection theory has not generally assessed how the number of descendants of an individual grows or reduces over time in relation to the properties of the population and the selection strategy. Though HILL (1974) modeled the flow of genes through a population, his method ignores the effect of selection and does not consider the individual animal. No theoretical framework has been developed to model the inheritance of selective advantage from parents to offspring, with the exception of ROBERTSON (1961), who introduced the concept of accumulation of selective advantage. There is no general theory that provides a model to describe the effect of different selection strategies on pedigree development and relates it to rates of inbreeding.

In classical selection theory (see, *e.g.*, FALCONER and MACKAY, 1996), genetic gain is expressed as a selection differential, which is a conditional expectation of a subset of the population, *i.e.*, it is a statistical measure of genetic progress. Classical theory does not explicitly show how selection response is related to the selective success of individuals in relation to their genetic superiority. It seems obvious that sustained genetic gain can only be achieved when the individuals contributing to the population on the long-term have an above average Mendelian sampling term. Nevertheless, apart from WOOLLIAMS and THOMPSON (1994), no theory has been developed that explicitly shows this relation.

This thesis focuses on the effects of selection on the development of pedigree, with particular emphasis on the rate of inbreeding (ΔF). The central concept in this thesis is the "long-term genetic contribution", which was introduced by JAMES and MCBRIDE (1958). The long-term genetic contribution (r_i) of ancestor i born at

time t_1, is defined as the proportion of genes from i that are present in individuals in generation t_2 deriving by descent from i, where $(t_2 - t_1) \rightarrow \infty$ (WOOLLIAMS *et al.*, 1993). In other words, the long-term contribution of an individual is its proportional contribution to the genetic make-up of the population in the long term. Because long-term genetic contributions are proportions, they sum to unity per generation. In the remainder of this introduction, long-term genetic contributions will be referred to as long-term contributions.

Besides chance effects, the long-term contribution of an individual is affected in a systematic manner by the superiority of the individual. For example, when selection is for estimated breeding values (EBV), individuals with a high EBV are expected to have more selected offspring, which will increase their long-term contribution. The EBV, therefore, is a measure of the selective advantage of an individual. Throughout this thesis, the term "selective advantage" may refer to any variable that affects the long-term contribution of an individual, by affecting the selective success of its offspring and more distant descendants.

There are two mechanisms that affect the long-term contribution of an individual (WRAY and THOMPSON, 1990). First the relation between the number of selected offspring and the selective advantage of their parents, which determines the expected number of selected offspring. Second, the inheritance of selective advantage from parents to *selected* offspring, which affects the selective advantage of the next generation of parents. In this thesis, a general theory will be developed to predict long-term contributions, by modeling those two mechanisms. This theory enables prediction of the expected development of pedigree, which has not been possible so far.

WOOLLIAMS and THOMPSON (1994) stated that the rate of gain is proportional to the sum of cross-products of long-term contributions and Mendelian sampling terms, without giving a formal derivation. In this thesis a formal theory will be developed, explicitly showing that genetic gain arises from creating a covariance between long-term contributions and Mendelian sampling terms. Predictive equations will be developed to implement the theory and to demonstrate the relation to classical selection theory.

The rate of inbreeding, or equivalently, effective population size [$\Delta F = 1/(2N_e)$], is the key parameter that measures the genetic size of a population. It determines the variance of gene frequency due to drift, the increase in homozygosity by descent, the fixation probability of favorable mutants and the equilibrium state of the mutation-selection-drift balance. (FALCONER and MACKAY, 1996; LYNCH and WALSH, 1998). WRAY and THOMPSON (1990) showed that rates of inbreeding are proportional to the sum of squared long-term contributions. Subsequently, WOOLLIAMS *et al.* (1993) and WRAY *et al.* (1994) further developed this approach, but in particular the prediction of the variance of long-term contributions proved to be difficult. SANTIAGO and CABALLERO (1995) predicted rates of inbreeding by modeling the variance of gene frequency, without using long-term contributions. This thesis will show that, under

certain conditions, the rate of inbreeding can directly be predicted from the expectation of the long-term contribution, making a separate prediction of the variance redundant. Predictive equations will be developed for animal breeding populations, which, for the first time, enable a computationally feasible optimization of breeding schemes with respect to rates of genetic gain and inbreeding.

Outline of the thesis

This thesis can be divided into three main parts. First, CHAPTERS 2 to 4 deal with the prediction of long-term contributions and their relation to genetic gain and generation interval. Second, CHAPTERS 5 to 8 deal with the prediction of rates of inbreeding based on long-term contributions. Finally, CHAPTERS 9 and 10 deal with the application of the theory to Combined Crossbred Purebred Selection.

CHAPTER 2 to 4: CHAPTER 2 develops a general theory to predict long-term contributions and formally derives the relation between long-term contributions and genetic gain. Long-term contributions will be predicted by linear regression of contributions on selective advantages. The regression coefficients will be derived by modeling the relation between selective advantage and the number of selected offspring and by modeling the inheritance of selective advantage. With overlapping generations the long-term contribution will be predicted by modifying conventional gene flow theory (HILL, 1974) in order to account for selection.

CHAPTER 3 shows how the general theory developed in CHAPTER 2 can be implemented for populations with overlapping generations undergoing mass selection. Particular emphasis will be given to the generation interval. The theory of CHAPTER 2 will be compared to conventional gene flow theory, which ignores the effect of selective advantage on long-term contributions.

CHAPTER 4 is a short note, which discusses the relation between gene flow theory and genetic gain in an intuitive manner. Particular emphasis will be given to the different concepts underlying gain predicted from conventional gene flow theory and gain predicted from long-term contributions.

CHAPTER 5 to 8: CHAPTER 5 deals with the relationship between long-term contributions and rates of inbreeding. First, it will be shown that rates of inbreeding are proportional to squared long-term contributions. WRAY and THOMPSON (1990) already derived this relation, using properties of the relationship matrix. In this thesis the relation will be derived directly from identity by descent, which enhances intuitive understanding.

Second, it will be shown that, with Poisson family size, rates of inbreeding are directly related to squared *expected* long-term contribution, making a separate prediction of the variance of long-term contributions redundant. Finally, the theory will be applied to sib-indices in discrete generations. Together, CHAPTER 2 and 5 represent a unified theory of rates of gain and inbreeding.

In CHAPTER 6, equations will be developed to predict rates of inbreeding for populations with either discrete or overlapping generations undergoing mass selection, which shows how

the theory described in CHAPTERS 2 and 5 can be implemented. Furthermore, CHAPTER 6 shows how the prediction of ΔF based on long-term contributions relates to previous predictions for mass selection based on the variance of gene frequency, as described by SANTIAGO and CABALLERO (1995) for discrete generations and by NOMURA (1996) for a special case of overlapping generations.

CHAPTER 7 shows how rates of inbreeding may be predicted for populations that are selected on Best Linear Unbiased Prediction (BLUP) of breeding values. Specific attention will be given to the relation between ΔF and population parameters, such as the number of parents and selection intensity.

Finally, CHAPTER 8 shows how rates of inbreeding may be predicted for typical livestock breeding populations with overlapping generations, BLUP selection and progeny testing.

CHAPTERS 9 and 10: CHAPTERS 9 and 10 deal with Combined Crossbred Purebred Selection (CCPS) in crossbreeding schemes. CHAPTER 9 shows how short term rates of genetic gain may be predicted with CCPS and BLUP selection, following the approach of WEI and VAN DER WERF (1994) and WRAY and HILL (1989). Furthermore, CHAPTER 9 describes the optimization of CCPS breeding schemes, ignoring rates of inbreeding.

CHAPTER 10 describes the optimization of CCPS breeding schemes when the rate of inbreeding is restricted, and shows how the theory developed in CHAPTERS 2 to 8 can be used to balance rates of gain and inbreeding for animal breeding schemes in a computationally feasible manner.

GENERAL DISCUSSION: The General Discussion addresses the relevance of this thesis for quantitative genetic theory and for applied animal breeding. Finally, the relevance of inbreeding in future breeding programs will be discussed.

Literature Cited

BULMER, M. G., 1971. The effect of selection on genetic variability. Am. Nat. 105: 201-211.

FALCONER, D. S. and T. F. C. MACKAY, 1996. Introduction to quantitative genetics. 4th Ed. Longmand Sci. and Tech., Harlow, UK.

HAZEL, L. N., 1943. The genetic basis for constructing selection indices. Genetics 28: 476-490.

HENDERSON, C. R., 1963. Selection index and the expected genetic advance. *In* W.D. HANSON and H. F. ROBINSON (eds.), *Statistical genetics in plant and animal breeding*, pp. 141-163. Natl. Acad. Sci., Natl. Res. Council Publ. No. 982, Washington, D. C.

HENDERSON, C. R., 1973. Sire evaluation and genetic trends. *In* Proceedings of the animal breeding and genetics symposium in honor of Dr. J. L. Lush, pp. 10-41. Am. Soc. Anim. Sci., Champaign, IL.

HENDERSON, C. R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423-447.

HENDERSON, C. R., 1976. A simple method for the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32: 69-83.

HILL, W. G., 1974. Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

JAMES, J. W. and G. MCBRIDE, 1958. The spread of genes by natural and artificial selection in a closed poultry flock. J. Genet. 56: 55-62.

LYNCH, M. and B. WALSH, 1998. Genetics and analyses of quantitative traits. Sinauer Associates, Inc., Sunderland, M.A., U.S.A.

NOMURA, T., 1996. Effective size of selected populations with overlapping generations. J. Anim. Breed. Genet. 113: 1-16.

ROBERTSON, A., 1961. Inbreeding in artificial selection programmes. Genet. Res. 2: 189-194.

SANTIAGO, E. and A. CABALLERO, 1995. Effective size of populations under selection. Genetics 139: 1013-1030.

WEI, M. and J. H. J. VAN DER WERF, 1994. Maximizing genetic response in crossbreds using both purebred and crossbred information. Anim. Prod. 59: 401-413.

WOOLLIAMS, J. A. and R. THOMPSON, 1994. A theory of genetic contributions. Proc. 5th World Congr. Genet. Appl. Livestock Prod. 19: 127-134.

WOOLLIAMS, J. A., N. R. WRAY and R. THOMPSON, 1993. Prediction of long-term genetic contributions and inbreeding in populations undergoing mass selection. Genet. Res. 62: 231-242.

WRAY, N. R. and W. G. HILL, 1989. Asymptotic rates of response from index selection. Anim. Prod. 49: 217-227.

WRAY, N. R. and R. THOMPSON, 1990. Prediction of rates of inbreeding in selected populations. Genet. Res. 55: 41-54.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1994. Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. 87: 878-892.

Expected Genetic Contributions and their Impact on Gene Flow and Genetic Gain

John A. Woolliams[1], Piter Bijma[2] and Beatriz Villanueva[3]

[1]Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K. [2]Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University,6700 AH Wageningen, The Netherlands and [3]Scottish Agricultural College, West Mains Road, Edinburgh, EH9 3JG, U. K.

Abstract – Long-term genetic contributions (r_i) measure lasting gene flow from an individual i. By accounting for linkage disequilibrium generated by selection both within and between breeding groups (categories), assuming the infinitesimal model, a general formula was derived for the expected contribution of ancestor i in category q ($\mu_{i(q)}$), given its selective advantages ($s_{i(q)}$). Results applied to overlapping generations with multiple modes of inheritance and selection indices. Genetic gain was related to the covariance between r_i and the Mendelian sampling deviation (a_i), thereby linking gain to pedigree development. When $s_{i(q)}$ includes a_i, gain was related to $E[\mu_{i(q)}a_i]$, decomposing it into independent components attributable to within- and between-families, within each category, for each element of $s_{i(q)}$. The formula for $\mu_{i(q)}$ was consistent with previous index theory for predicting gain in discrete generations. For overlapping generations, accurate predictions of gene flow were obtained among and within categories, in contrast to previous theory that gave qualitative errors among categories, and no predictions within. The generation interval was defined as the period for which $\mu_{i(q)}$, summed over all ancestors born in that period, equalled 1. Predictive accuracy was supported by simulation results for gain and contributions with sib-indices, BLUP selection and selection with imprinted variation.

Selection theory has not generally addressed how the number of descendants from an individual grows or reduces over time, in relation to properties of the population. This is perhaps surprising, since the development of the pedigree over generations provides the framework for the passage of genes through the population, forming the link between our understanding of individual genotypes and the way such genotypes influence the population. Such an understanding provides answers to, for example: the relative importance of individuals within a generation; where genetic change has arisen; how quickly the change generated has spread through the population; with what precision are we able to predict this change; how is genetic change related to the loss of variation; and how does genetic change in one generation relate to that in a subsequent generation. These questions have no general framework within which they can be answered although some special cases have been investigated (e.g. VILLANUEVA *et al.* 1996; BIJMA and WOOLLIAMS 1999).

The objective of this study is to describe the expectations for the proliferation of genetic lines using the concept of genetic contributions. The generation of linkage disequilibrium during selection changes the impact of selective advantages, and this must be accounted for in order to predict the flow of an individual's genes through a population over time. These changes affect the comparative gene flow of different breeding groups or categories, and of different individuals within categories. The general development will build upon the pioneering work of WRAY and THOMPSON (1990), and more latterly on the studies of WOOLLIAMS *et al.* (1993; mass selection), WRAY *et al.* (1994; sib indices) and WOOLLIAMS and THOMPSON (1994). Firstly, the concept of genetic contributions will be considered in relation to genetic gain, and a general formula for gain will be proved. The expected genetic contribution of an individual to subsequent generations will be derived, and the relationship of the long-term genetic contribution with gain will be used to show the consistency between the theory developed and classical theory (*e.g.*, BULMER 1980). The concept of the generation interval will be re-evaluated as a natural extension of the contribution theory. Many of the detailed results will be derived assuming an equilibrium. The use of the formulae developed will be shown in examples of selection applied to discrete generations using sib indices, using best linear unbiased predictors (BLUP), with imprinted variation, and with overlapping generations.

Methods

Definitions and basic notation: Table 1 shows the notation for the principle parameters. The concept of genetic contributions was introduced by JAMES and MCBRIDE (1958) and was developed by WRAY and THOMPSON (1990) for the prediction of rates of inbreeding (ΔF). Given the fundamental nature of the concept to this paper, the definition will be re-stated. The genetic contribution of an ancestor i born at time u to an individual j born at time t ($>u$), is the proportion of the genes of j that are expected to derive by descent from ancestor i. This is different from the definition used by WRAY and THOMPSON (1990), who multiplied this proportion by X_m+X_f (where X_m and X_f are the number of male and female parents in a generation). However, as shown by WOOLLIAMS *et al.* (1993), a contribution is more usefully defined without this re-scaling. It is also distinct from the numerator genetic relationship which considers shared genes, not only those restricted to descent; so full-sibs make no genetic contribution to each other although they have a genetic relationship >0.

The notation will be defined to allow extension to overlapping generations. Therefore contributions will be defined within and between categories, where the categories are defined by both age and sex and, potentially, breeding use (*e.g.*, nucleus females and other females). Over its lifetime, an individual will move through various categories. An initial objective is to show the relationship between contributions and rate of gain, and for this there is no need to identify details of the category of an individual and what is happening to the different categories over time. For this objective, it is only necessary to consider the observed contribution by whatever means it is achieved. However, to develop the concept of gene flow, which is important for understanding the dynamics of overlapping generations, the

TABLE 1.- The notational conventions for the principal parameters.

t, u	time variables
p, q	one of a total of n_c categories defined by sex and age
i, j, $i(q)$	individuals in the population, $i(q)$ denotes individual in category q
G_t, ΔG	genetic merit of population at time t and rate of genetic gain
T_m, T_f	number of male and female candidates available for selection
$r_{i(q)}$	long-term contribution of i in category q
a_i, A_i	Mendelian sampling term and breeding value of i
S_i	selection score for i, 0 or 1 according to i selected or not
$s_{i(q)}$	vector of selective advantages for $i(q)$, of length n_s; mean over all selected in category q is $\bar{s}_q$
$\mu_{i(q)}$	expected long-term contribution, assumed to be linear regression on $s_{i(q)}$ of form $\alpha_{i(q)} + \boldsymbol{\beta}_q^T (s_{i(q)} - \bar{s}_q)$
α, $\boldsymbol{\beta}$	vectors of the coefficients for $\mu_{i(q)}$, of lengths n_c and $n_c n_s$ respectively
X_p	number of parents in category p; the n_c x n_c diagonal matrix N has elements X_p
g_{pq}	proportion of genes of selected individuals in category p that derive from parents of category q; the n_c x n_c matrix G has elements g_{pq}; the n_c x n_c matrix G_p has elements other than the pth row equal to zero.
$g_{0,pq}$	proportion of genes among the newborn from which category p are selected that derive from parents of category q
λ_{pq}	regression coefficients of proportion selected in category p on $s_{i(q)}$ for parent in category q; has dimension 1 x n_s; the n_c x $n_c n_s$ matrix $\boldsymbol{\Lambda}$ has elements λ_{pq}
π_{pq}	regression coefficients of $s_{j(p)}$ on $s_{i(q)}$ for parent in category q; has dimension n_s x n_s; the $n_c n_s$ x $n_c n_s$ matrix $\boldsymbol{\Pi}$ has elements π_{pq}
h^2, h_0^2	heritability of trait in candidates, and heritability in unselected base generation
σ_A^2, σ_P^2, σ_I^2	additive genetic, phenotypic and index variance
ι_q, k_q	standardised selection intensity, and variance reduction coefficient for category q
ρ, L	index accuracy and generation interval
d_{pq}	for refining α: $d_{pq} = E[s_{j(p)} \mid j(p)$ *has category* q *parent*$] - \bar{s}_p$; has dimension n_s x 1; the $n_c n_s$ x n_c matrix D has elements d_{pq}
$b_q(p,t)$, $c_q(p,t)$	for ancestor i in category q at time 0, the genetic contribution to selected individuals in category p time t is $c_q(p,t) + b_q(p,t)\,(A_{i(q)} - \bar{A}_q)$, with vector of coefficients for all categories p denoted by $\boldsymbol{c}_q(t)$ and $\boldsymbol{b}_q(t)$
$f_q(p,t)$	regression of $A_{j(p)}$ for selected $j(p)$ at time t on $A_{i(q)}$ for ancestor $i(q)$ at time 0, with vector for all categories p denoted by $\boldsymbol{f}_q(t)$
τ_m, τ_f, τ_w	for sib indices: regression of the index on the sire's and dam's transmitting ability and on the candidate's Mendelian sampling term;
τ, ι, k	for sib indices: $\tau = \frac{1}{2}(\tau_m + \tau_f)$; $\iota = \frac{1}{2}(\iota_m + \iota_f)$; $k = \frac{1}{2}(k_m + k_f)$
z, κ	for sib indices: $z = \rho\sigma_A$; $\kappa = [k\tau + \frac{1}{8}(\tau_m - \tau_f)(k_m - k_f)]$
$\hat{A}_i$, $\delta\hat{A}_i$, $\hat{e}_i$	for BLUP: the EBV of i when selected, its EBV at time of offspring selection, and the remaining prediction error
A^-, A^+	for imprinted variation: the hidden and expressed breeding value

tracking of categories is required. Therefore, in order to keep notation minimal at any given stage, the notation for contributions will be developed through the paper, and a balance between consistency and simplicity has been attempted.

The following notation will be used initially: $r_{i,u}(j,t)$ is the contribution of ancestor i that was born at time u to individual j born at time t; $r_{i,u}(t)$ is the mean contribution over all the newborn cohort at time t (*i.e.*, ½ of the mean for newborn males plus ½ of the mean for newborn females). The long-term contributions of i, $r_{i,u} = r_{i,u}(t)$ as $t \to \infty$. For long-term contributions there will be less need to specify u and r_i is used. T_m males and T_f females are scored in each cohort at random, and only scored individuals are candidates for breeding opportunities.

The populations will be assumed to mix over time. With mixing, the contribution a particular ancestor makes to later-born individuals will tend to a value that is the same for all individuals in later cohorts, *i.e.*, for each i, the variance of $r_{i,u}(j,t)$ among j tends to 0 as $t \to \infty$ (WRAY and THOMPSON 1990). This value is the long-term contribution r_i, and will differ between individual ancestors, depending upon the lifetime breeding use of i, its breeding value and other selective advantages both genetic and non-genetic, and chance factors. WRAY and THOMPSON (1990) and GRUNDY *et al.* (1998) describe in more detail the relationship between the long-term contribution and the numerator relationship matrix.

The full development presented in this paper will assume the infinitesimal model with negligible rates of inbreeding, since this will satisfy the principle requirement for a period of equilibrium in the population structure. This study will use Mendelian sampling terms to mean the deviation of the breeding value of an individual from the mean of its parents' breeding values, and Mendelian sampling variance to mean the variance of these deviations.

Rates of gain

The breeding value of an individual may be decomposed into a sum of independent terms, involving the breeding values of the base generation and Mendelian sampling terms of all other ancestors. This may be done by observing that: (*i*) the breeding value of an individual j born at time t can be expressed as the average of its parental breeding values plus a deviation (the Mendelian sampling term) which is independent of its parental breeding values, *i.e.*, $A_{j,t} = ½A_{sire} + ½A_{dam} + a_{j,t}$; and, (*ii*) by going backwards through the pedigrees, this substitution can be repeated for each generation of ancestors until the base generation is reached. The coefficients for these terms are the genetic contributions of the ancestors to individual j born at time t. Therefore,

$$A_{j,t} = \sum_{u=1}^{t} \sum_{i} r_{i,u}(j,t) a_{i,u} + \sum_{i} r_{i,0}(j,t) A_{i,0}$$

The second term is to allow for the base population, not necessarily unselected, where it is assumed that parents are unknown and so all the genetic information prior to $t = 0$ is contained in this base information. Let G_t, the genetic merit of the population at time t, be the average of the breeding values of the newborn males and females, *i.e.*,

$$G_t = ½ \sum_{j\ males} T_m^{-1} A_{j,t} + ½ \sum_{j\ females} T_f^{-1} A_{j,t};$$

then

$$G_t = \sum_{u=1}^{t} \sum_i r_{i,u}(t)\, a_{i,u} + \sum_i r_{i,0}(t)\, A_{i,0}.$$

Since $E[a_{i,u}] = 0$, the cross-product $r_i a_i$ is related to the covariance between r_i and a_i. Thus sustained genetic gain is related to the creation of a covariance between contributions and Mendelian sampling terms.

Let the gain made by selection in cohort t be $\Delta G_t = G_{t+1} - G_t$, and $\Delta r_{i,u}(t) = r_{i,u}(t+1) - r_{i,u}(t)$ then:

$$\Delta G_t = \sum_{u=1}^{t} \sum_i \Delta r_{i,u}(t)\, a_{i,u} + \sum_i \Delta r_{i,0}(t)\, A_{i,0} \quad (1)$$

Since the population is assumed to mix, the terms $\Delta r_{i,u}(t) \to 0$ as $t \to \infty$ and so $\Delta r_{i,u}(t)\, a_{i,u} \to 0$ as $t \to \infty$ for a fixed u, and, in particular, the terms for the base population terms in Equation 1 tend to 0. Therefore, for large t, summing over males ($i(m)$) and females ($i(f)$) separately and taking expectations,

$$E[\Delta G_t] = \sum_{u=1}^{t} \sum_{q=m,f} T_q E[\Delta r_{i(q),u}(t)\, a_{i(q),u}] \quad (2)$$

If an equilibrium is approached (as will be the case with the infinitesimal model when inbreeding is ignored), the expected *change* in covariance between r_i and a_i will depend only on t-u and not on u *per se*, *i.e.*, only on the elapsed time since the ancestor's birth, and not on the actual time of birth. So $E[\Delta r_{i(q),u}(t)\, a_{i(q),u}] = E[\Delta r_{i(q),u+\delta t}(t+\delta t)\, a_{i(q),u+\delta t}]$.

After making these substitutions, ΔG_t may be expressed as a sum of changes in contributions of individual ancestors, *i.e.*, $\sum_{u=1}^{t} E[\Delta r_{i(q),u}(t)\, a_{i(q),u}] = \sum_{t=0}^{u-1} E[\Delta r_{i(q),u}(u+t)\, a_{i(q),u}]$.

For u large enough, the right hand side will approach its equilibrium value $E[r_{i(q),u} a_{i(q),u}]$. Therefore, for a sufficiently large t, $E[\Delta G_t] = E[\Delta G_{eq}]$ and substitution of these results into Equation 2 gives:

$$E[\Delta G_{eq}] = T_m E[r_{i(m)} a_{i(m)}] + T_f E[r_{i(f)} a_{i(f)}] \quad (3)$$

or equivalently, $E[\Delta G_{eq}] = T_m \mathrm{cov}(r_{i(m)}, a_{i(m)}) + T_f \mathrm{cov}(r_{i(f)}, a_{i(f)})$. An equivalent expression to Equation 3 can be given as a continuous function of time (available from the authors).

Comparison of Equation 3 with other expressions of gain: The traditional formula for quantitative genetic gain expresses gain as the product of selection intensity (i), accuracy (ρ) and genetic standard deviation (σ_A), defined in a single generation. Equation 3 makes explicit and clear that: (*i*) genetic gain must arise from 'good' ancestors contributing more genes; (*ii*) this process of contributing genes concerns more than a single generation; (*iii*) sustained gain depends on utilizing the new variation, *i.e.*, the Mendelian sampling variation, entering the population each generation; and (*iv*) quantitatively, the covariance of r_i with a_i gives a complete description of the process involved in items (*i*) to (*iii*).

The traditional expression for gain may be the most tractable form for calculation in most schemes, but it is unclear that this will always be the case, *e.g.*, with quadratic indices as described by MEUWISSEN (1997) and GRUNDY *et al.* (1998). However, it is shown later that formulae

developed in the next sections and used in Equation 3 lead to estimates for rates of gain that are precisely equivalent to the traditional expression for important cases. Therefore, the main outcome of Equation 3 is that the rate of gain has been connected to the pedigree, which is not apparent with $i\rho\sigma_A$. Equation 3 is useful for decomposing achieved gain, but its usefulness for prediction is limited because r_i is observed. It is therefore necessary to develop expectations for r_i.

Expected long-term contributions

Framework for general solution: As described above, one reason for deriving expected long-term contributions is to exploit the relationships between the long-term contributions and rates of gain, by replacing the observed r_i with its expectation. There are other reasons that are perhaps more important. First, the expected contributions are involved in predicting rates of inbreeding (ΔF) in selected populations, using the relationship between ΔF and the sum of squared contributions (WRAY and THOMPSON, 1990; WOOLLIAMS *et al.* 1993). Second, the expected long-term contributions represent the expected gene flow in the population, and in complex population structures (with overlapping generations and breeding pyramids) this information is essential for scheme design. To develop expected contributions, it will be necessary to modify slightly the notation used. In particular, it will be necessary for breeding categories (*i.e.*, ages and sexes) to be explicit, so $i(q)$ will denote an ancestor in category q.

The expected long-term contribution of individual $i(q)$ will be defined conditional on a vector of n_s selective advantages, $s_{i(q)}$. The $s_{i(q)}$ are expressed as deviations from the average of the selected contemporaries $\bar{s}_q$. The selective advantages influence the success of the offspring, and (or) may influence the selection of subsequent descendants, *i.e.*, $\mu_{i(q)} = E[r_{i(q)} \mid s_{i(q)}]$. For example, an estimated breeding value (EBV) of an ancestor at the time of selection of its own offspring will influence the number of offspring that are selected, and will play a role in the number of grand-offspring selected. In contrast, the corresponding prediction error of the EBV will not influence selection of offspring but will influence selection of grand-offspring. The conditional expectation expresses the expected contribution as a function of the selective advantages. If a linear model for the conditional expectation is assumed, then $\mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^T(s_{i(q)} - \bar{s}_q)$. If an equilibrium is assumed, then the coefficients α_q and $\boldsymbol{\beta}_q$ will not change over generations and the same coefficients can be used for both the ancestor and the selected offspring. The expected lifetime long-term contribution of an individual i will be the sum of the expected long-term contributions for all categories that i belonged to over its lifetime.

The objective of the following section is to define a set of achievable steps which can be followed to derive formulae for α_q and $\boldsymbol{\beta}_q$, in order to obtain expected contributions even in complex breeding schemes. The starting point will be to note that the long-term contribution of individual i is given by:

$$r_i = \tfrac{1}{2} \sum_{\text{offspring } j} r_j \qquad (4)$$

where the sums are taken over its male and female offspring. Since unselected offspring have no long-term contribution, these sums may be restricted to the selected offspring. Taking expectations conditional on $s_{i(q)}$ and summing over categories p,

$$\mu_{i(q)} = \tfrac{1}{2} \sum_{\text{categories } p} E[\text{no. offspr. selected in cat. } p \mid s_{i(q)}] \times E[r_{j(p)} \mid s_{i(q)}] \quad (5)$$

Let the population have n_c categories which describe sex, age and breeding purpose. Discrete generations are a special case with only two categories, males and females. Initially $s_{i(q)}$ will be assumed to be a single variable (n_s = 1), namely the breeding value $A_{i(q)}$. This was assumed for mass and sib-index selection by WOOLLIAMS *et al.* (1993) and WRAY *et al.* (1994). In this situation, β_q is a single number. The expected long-term contributions for individual i in category q can then be represented by $\mu_{i(q)} = \alpha_q + \beta_q (A_{i(q)} - \bar{A}_q)$.

The solutions will be obtained from four steps: (*i*) for overlapping generations only, to determine the gene flow from the parents (*sic*) in previous periods to selected individuals in the current period; (*ii*) to regress the expected number of offspring selected for a parent upon its selective advantage(s), with the regression coefficients λ_{pq} forming an n_c x n_c matrix $\mathbf{\Lambda}$; (*iii*) to regress the selective advantage(s) of a selected offspring upon those of its parent, with the coefficients π_{pq} forming an n_c x n_c matrix $\mathbf{\Pi}$; (*iv*) from these steps calculate the vectors of α_q and β_q for all categories, *i.e.*, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots \alpha_{n_c})^T$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_{n_c})^T$, both of dimension n_c x 1.

Step 1, defining the gene flow matrix $\boldsymbol{G}$: The concept of gene flow (HILL 1974) is used, but the development of HILL does not account for the inheritance of selective advantage, which is critical for selection. A consequence of this selective advantage is that the probability that the parent of a selected individual in category p comes from category q will depend on the selection intensity in category p and on the selective advantage of category q over other categories contributing candidates for category p. If category q has a selective advantage over other categories, then its offspring will have increasing success as selection becomes more intense. Consider an example where dams from age 1 have a higher genetic merit than those of age 2, and the two ages contribute equally to a group of newborn individuals. If selection among this newborn group is at random, then those chosen are expected to come equally from 1 and 2 year old females. However, if there is selection in this group, offspring of females of age 1 would be expected to be favoured.

In the standard gene flow matrix (HILL 1974), the key elements are $g_{0,pq}$, representing the proportions of genes in the newborn cohort from which category p will be selected (at some time in their life) that arise from category q parents. To obtain the expected long-term contributions, a modified matrix is required ($\boldsymbol{G}$, of dimension n_c x n_c) in which each row represents a category of *selected* individuals (rather than newborn), and with the elements g_{pq} of each row representing the proportions of genes in the *selected* individuals transferred through breeding from the parents in the different categories q. With discrete generations and the standard two pathways, $\boldsymbol{G}$ = (½,½ | ½,½) always. Deterministic procedures to obtain $\boldsymbol{G}$ are described in detail by BIJMA and WOOLLIAMS (1999), and briefly in the application concerning overlapping generations in this paper.

Step 2, defining and deriving $\mathbf{\Lambda}$: A model is required for the regression of the number of offspring (the expected selection score) of a parent in category q that are selected to breed in

category p, on the breeding value of their parent. With random selection, the proportion of the X_p selected in category p that are expected to have category q parents is $2g_{0,pq}$, and these are divided equally among the X_q parents in category q. In this case, the expected selection score for a parent in category q is simply a constant, $2X_p g_{0,pq} X_q^{-1}$, and does not depend upon $A_{i(q)}$. With selection, APPENDIX A shows that this expectation is of the form $2X_p g_{pq} X_q^{-1}$ $[1 + \lambda_{pq}(A_{i(q)} - \overline{A}_q)]$. The elements λ_{pq} form an n_c x n_c matrix $\boldsymbol{\Lambda}$. For mass selection, $\lambda_{pq} = \frac{1}{2}\imath_p \sigma_P^{-1}$, where $\imath_p$ is the intensity of selection in category p, and σ_P is the phenotypic standard deviation.

Step 3, defining and deriving $\boldsymbol{\Pi}$: A second regression model is required for the regression of the breeding value of the *selected* offspring j on the breeding value of the parent i. In principle these, depend on both the category of offspring and parent, giving an n_c x n_c matrix $\boldsymbol{\Pi}$, with π_{pq} representing the coefficient for offspring category p and parent category q. Thus $E[A_{j(p)} - \overline{A}_p] = \pi_{pq}(A_{i(q)} - \overline{A}_q)$. Appendix B gives a general derivation for $\boldsymbol{\Pi}$, which is used in all the applications. For the case of mass selection with only the breeding value conferring selective advantage, $\pi_{pq} = \frac{1}{2}(1 - k_p h^2)$, where k_p is the variance reduction coefficient for selection in category p and h^2 is the heritability in the candidates.

Step 4, solutions: Using Equation 5 with: (*i*) the breeding value replacing $s_{i(q)}$ as the selective advantage; (*ii*) the *E[number of selected offspring* $|A_{i(q)}]$ replaced by $2X_p g_{pq} X_q^{-1}[1 + \lambda_{pq}(A_{i(q)} - \overline{A}_q)]$; (*iii*) the assumption of equilibrium justifying the use of the same $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for both parent and offspring; (*iv*) $(A_{j(p)} - \overline{A}_p)$ in $E[r_{j(p)} \mid s_{i(q)}]$ replaced by $\pi_{pq}(A_{i(q)} - \overline{A}_q)$; and collecting terms independent of $A_{i(q)}$ and those linearly dependent upon $A_{i(q)}$ separately, gives:

$$\alpha_q = \sum_p X_p g_{pq} X_q^{-1} \alpha_p \qquad (6a)$$

$$\beta_q (A_{i(q)} - \overline{A}_q) =$$

$$\sum_p (X_p g_{pq} X_q^{-1} \lambda_{pq} \alpha_p + X_p g_{pq} X_q^{-1} \beta_p \pi_{pq}) (A_{i(q)} - \overline{A}_q) \qquad (6b)$$

The quadratic terms have been neglected and this will be addressed in DISCUSSION. If N is the diagonal matrix with elements X_p then the matrix form of Equations 6a and 6b are:

$$(N\boldsymbol{\alpha}) = G^T(N\boldsymbol{\alpha}) \qquad (7a)$$

$$(N\boldsymbol{\beta}) = (I - G^T \circledast \boldsymbol{\Pi}^T)^{-1}(G^T \circledast \boldsymbol{\Lambda}^T)(N\boldsymbol{\alpha}) \qquad (7b)$$

where $\circledast$ denotes element-by-element multiplication of the matrices.

Therefore, $N\boldsymbol{\alpha}$ is a right eigenvector of G^T with eigenvalue 1 (this eigenvector exists since all rows of G sum to 1). This only defines $\boldsymbol{\alpha}$ up to a scalar. Let L be the generation interval, defined as the period of time for the population to renew itself. Then, (*i*) over its lifetime, a single cohort has a total long-term contribution of $\sum_p X_p \alpha_p$ and so $L\sum_p X_p \alpha_p = 1$; (*ii*) the average age at which the long-term contributions are made is given by $L = (\sum X_p \alpha_p)^{-1} \sum X_p \alpha_p age(p)$, where $age(p)$ is the age of individuals in category p. Combining these two formulae, gives the constraint $\sum_p X_p \alpha_p age(p) = 1$, and this is sufficient to define $\boldsymbol{\alpha}$ uniquely. Note

$L = (\sum X_p \alpha_p)^{-1}$. For discrete generations with the standard two pathways, $\boldsymbol{\alpha} = (\frac{1}{2}X_m^{-1}, \frac{1}{2}X_f^{-1})^T$ and $L = 1$ always.

The vector $N\boldsymbol{\beta}$ is completely determined once $\boldsymbol{G}$, $\boldsymbol{\Pi}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\alpha}$ are defined. If we consider a simple case with a single category, which may occur with a monoecious population with X parents, then all the terms become scalars and $\beta = (1-\pi)^{-1}\lambda\alpha$ and $\alpha = X^{-1}$. For more than one category the g_{pq} act as weighting factors across the categories for the different values of π_{pq} and λ_{pq}.

Extension to multiple variables (s): With multiple variables (n_s) conferring selective advantage, $\mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^T(s_{i(q)} - \bar{s}_q)$. $\boldsymbol{\alpha}$ remains a vector of length n_c, but $\boldsymbol{\beta}$ is a vector of length $n_c n_s$ of the form $(\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots \boldsymbol{\beta}_{n_c}^T)^T$. Each element λ_{pq} becomes a 1 x n_s sub-matrix $\boldsymbol{\lambda}_{pq}$, and each element π_{pq} becomes an n_s x n_s sub-matrix $\boldsymbol{\pi}_{pq}$. The matrix $\boldsymbol{\Lambda}$ is of order n_c x $n_c n_s$, and $\boldsymbol{\Pi}$ is of $n_c n_s$ x $n_c n_s$. The solution for $\boldsymbol{\alpha}$ remains unchanged (Equations 6a, 7a). To obtain the equation analogous to Equation 6b, let $s_{j(p(v))}$ and $s_{i(q(w))}$ represent variables v and w in $s_{j(p)}$ and $s_{i(q)}$ respectively, so $1 \le v, w \le n_s$:

$$\beta_{q(w)}(s_{i(q(w))} - \bar{s}_{q(w)}) = (\sum_{p=1}^{n_c} X_p g_{pq} X_q^{-1} \lambda_{pq(w)} \alpha_p + \sum_{p=1}^{n_c} X_p g_{pq} X_q^{-1} \sum_{v=1}^{n_s} \beta_{p(v)} \pi_{p(v)q(w)}) \times (s_{i(q(w))} - \bar{s}_{q(w)}) \quad (8)$$

The matrix form of the equations for multiple variables in $s_{i(q)}$ (not shown) are the same as in Equation 7, but with: (*i*) the definition of $\circledast$ being extended to mean the multiplication of the sub-matrices $\boldsymbol{\pi}_{pq}$ and $\boldsymbol{\lambda}_{pq}$ by the element g_{pq}; and (*ii*) in Equation 7b, $N\boldsymbol{\beta}$ is replaced by $N\circledast\boldsymbol{\beta}$, *i.e.*, each sub-vector $\boldsymbol{\beta}_q$ is multiplied by X_q.

A further refinement of $\boldsymbol{\alpha}$: This section is not essential to the overall development, but it can prove important for good approximation in complex structures and it is used in RESULTS. The section describes an improvement in the estimation of $\boldsymbol{\alpha}$, which corresponds to a second order approximation.

The g_{pq} account for the different selective advantages among the categories of the parents at the time of selection, but the advantages or disadvantages are inherited in part by the selected offspring. From Equation 6a $\alpha_q = \sum_{p=1}^{n_c} X_p g_{pq} X_q^{-1}(\alpha_p + \boldsymbol{\beta}_p^T \boldsymbol{d}_{pq})$, where $\boldsymbol{d}_{pq} = E[s_p \mid \text{category } q \text{ parent}] - \bar{s}_p$. After rearranging terms in Equations 6a and 6b:

$$(N\boldsymbol{\alpha}) = [\boldsymbol{G}^T + (\boldsymbol{G}^T \circledast \boldsymbol{D}^T)(\boldsymbol{I} - \boldsymbol{G}^T \circledast \boldsymbol{\Pi}^T)^{-1} (\boldsymbol{G}^T \circledast \boldsymbol{\Lambda}^T)](N\boldsymbol{\alpha}) \quad (9)$$

where $\boldsymbol{D}$ is dimension ($n_c n_s$ x n_c), with sub-matrix pq equal to $\boldsymbol{d}_{pq}$. Although $\boldsymbol{\alpha}$ is still defined as a right eigenvector of a matrix with eigenvalue one, the matrix is now more complex. The constraint to define $\boldsymbol{\alpha}$ uniquely is unchanged. When generations are discrete and with the standard two pathway model $\boldsymbol{D} = 0$.

Expected long-term contributions and rates of gain: For any one individual i, the total long-term contribution is the sum of its long-term contributions as it moves through the different categories over its lifetime, *i.e.*, $r_i = \sum_{q=1}^{n_c} r_{i(q)}$.

Define $S_{i(q)} = 1$ if i is selected in category q, and 0 otherwise, then:

$$E[r_i \mid S_{i(q)},\ q=1,\ldots n_c] =$$

$$\sum_{q=1}^{n_c} S_{i(q)} E[r_{i(q)} \mid S_{i(q)} = 1] = \sum_{q=1}^{n_c} S_{i(q)} \mu_{i(q)}$$

When the expected long-term contribution is expressed in terms of the components of the breeding value, in particular the Mendelian sampling term, the expected long-term contribution is sufficient for the prediction of genetic gain, since the remaining part ($r_{i(q)} - \mu_{i(q)}$) has no covariance with the Mendelian sampling term. Within a category q, the sum of $S_{i(q)}$ over all candidates is X_q, and so application of equation (3) gives:

$$E[\Delta G_{eq}] = \sum_{q=1}^{n_c} X_q E[\mu_{i(q)} a_{i(q)}] \qquad (10)$$

where now the expectations are conditional on being selected as a parent, rather than unconditional as was the case in Equation 3. Equation 10 is expressed solely in terms of the selected individuals and in terms that are predictable rather than simply observed.

If $\mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^T (s_{i(q)} - \bar{s}_q)$, then Equation 10 immediately decomposes the gain into two components. The first, $\sum_{q}^{n_c} X_q \alpha_q E[a_{i(q)}]$, is the expected gain from selection within families, which occurs at the time of selection of the ancestor; whilst the second, $\sum_{q=1}^{n_c} X_q \boldsymbol{\beta}_q^T E[(s_{i(q)} - \bar{s}_q) a_{i(q)}]$, represents the expected between-family gain, and describes the changes in contribution of selected ancestors from the time of their selection until convergence in the long-term. Since the between-family gain is explicitly defined in terms of the selective advantages, the gain can be decomposed into components arising from each category and each selective advantage within categories.

The covariance between the Mendelian sampling term $a_{i(q)}$ and $(s_{i(q)} - \bar{s}_q)$ following the selection of the ancestor can be calculated using standard index theory. Note that, since this is a covariance with the deviation from a sample mean, adjustments using $(1 - X_q^{-1})$ should result in increased precision. For simplicity, this has *not* been applied in the results presented. The predicted increase in precision can be confirmed from the results shown.

Development of contributions over time: This section is not essential to the overall development, but describes the solution to an important application of gene flow. In complex population structures, it is often useful to predict how quickly improvement in one part of the population diffuses through to other parts of the population, or what proportion of the gene flow arises from particular pathways (*e.g.*, by male descent alone). This requires methods to predict the rate of convergence of genetic contributions over time.

To simplify the notation, the development of contributions over time is given for the single selective advantage, the breeding value, A. It is assumed that when $t = 0$, the population is already in equilibrium. For category q, a *selected* individual at time 0 has a vector (dimension n_c x 1) of contributions to *selected* individuals in category p at time t given by $c_q(p,t) + b_q(p,t)\ (A_{i(q)} - \bar{A}_q)$. This is a similar form to the long-term contribution, but before convergence it will differ between categories p and so needs to

be defined for each p. Let $\boldsymbol{c}_q(t) = [c_q(1,t), c_q(2,t), \ldots c_q(n_c,t)]^T$, and $\boldsymbol{b}_q(t) = [b_q(1,t), b_q(2,t), \ldots b_q(n_c,t)]^T$. Then $\boldsymbol{c}_q(0) = 0$, except for X_q^{-1} in the qth position, and $\boldsymbol{b}_q(0) = 0$. A further vector of regressions is required, $\boldsymbol{f}_q(t)$, for which the pth element is the regression of the breeding value of the selected individual in category p at time t on the breeding value of an ancestor in category q. By definition, $\boldsymbol{f}_q(0) = 0$ except for the qth position where it is 1.

It is critical to note that the contributions at time t to the selected individuals in category p of age *age(p)*, will depend on the consequences of the selection upon the parental gene pool at time *t-age(p)*. The more intense the selection, the more those parent categories with greater selection advantages will dominate. In a selection scheme, a group of newborn individuals will typically be subject to different selection intensities as they become older. Therefore, the complete spectrum of contributions among the selected individuals in the different categories at time t will depend on states back to t - *maxage*, where *maxage* is the maximum age of the parents in the breeding scheme. Define $\boldsymbol{G}_p$ to be the n_c x n_c matrix consisting of zeros, except for the single row corresponding to category p which is identical to the pth row of $\boldsymbol{G}$. Then

$$\boldsymbol{c}_q(t) = \sum_{p=1}^{n_c} \boldsymbol{G}_p \, \boldsymbol{c}_q(t-age(p)) \qquad (11a)$$

$$\boldsymbol{b}_q(t) = \sum_{p=1}^{n_c} \boldsymbol{G}_p \, \boldsymbol{b}_q(t-age(p)) + \sum_{p=1}^{n_c} X_q^{-1}(\boldsymbol{G}_p \otimes \boldsymbol{\Lambda}) \, \boldsymbol{f}_q(t-age(p)) \qquad (11b)$$

$$\boldsymbol{f}_q(t) = \sum_{p=1}^{n_c} (\boldsymbol{G}_p \otimes \boldsymbol{\Pi}) \, \boldsymbol{f}_q(t-age(p)) \qquad (11c)$$

Equation 11a describes the contribution of category q to each category at each time t, with element p of the sum describing the contributions of category q ancestors (at time $t = 0$) to category p parents at time t, accounting for the selection in category p through the matrix $\boldsymbol{G}_p$. Equation 11b describes the relationship of contributions from ancestors *within* category q (at time $t = 0$) to each category at each time t to the selective advantage. This arises from two processes, the first, analogous to Equation 11a, from the transfer of differential contributions among ancestors of category q that were accumulated up to and including time $t-1$, and the second from further differential contributions from selective advantages among the candidates at time t due to ancestors in category q at time $t = 0$. Equation 11c describes the changes in the selective advantages among the candidates at time t due to ancestors of category q at time $t = 0$.

When t becomes large, the mixing assumption for the population ensures that both $\boldsymbol{c}_q(t)$ and $\boldsymbol{b}_q(t)$ converge to a vector with all elements equal, namely $\alpha_q\mathbf{1}$ and $\beta_q\mathbf{1}$ respectively, where $\mathbf{1}=(1, \ldots, 1)^T$. Furthermore, $\boldsymbol{f}_q(t) \to 0$, since the eigenvalues of $\boldsymbol{G}\otimes\boldsymbol{\Pi}$ are <1 and >-1, and this reflects the diminishing effect of ancestors over time on the selective advantage of their descendants. By re-defining the state vector at time t to include not only $\boldsymbol{c}_q(t)$ but also $\boldsymbol{c}_q(t\text{-}1), \ldots \boldsymbol{c}_q(t\text{-}maxage+1)$, Equation 11a can be re-formulated (results not shown), so that the state vector at time t is the product of a square stochastic matrix of order n_c × *maxage* and the state vector at time $t-1$. Using this re-formulation and the properties of stochastic matrices (described in Appendix 1 of

HILL 1974), it can be demonstrated that Equations 11 are consistent with Equations 7 and the constraint $\sum_p X_p \alpha_p age(p) = 1$ (results not shown).

The discrete time contributions with the refinement in estimating α is given in APPENDIX C. An example of application is given in the results.

Applications of Models and Results

Expected long-term contributions and genetic gain for general sib indices in discrete generations: A general sib-index of the form $I = b_1(P-\overline{P}_F) + b_2(\overline{P}_F-\overline{P}_H) + b_3\overline{P}_H$ was studied by WRAY *et al.* (1994; WWT), where I is the index, P is the phenotype of the candidate, $\overline{P}_F$ and is the mean of the full-sib family (size n_F) including the candidate, and $\overline{P}_H$ is the mean of the half-sib family (size n_H) including the candidate and full-sibs. Mass selection is a special case with $b_1 = b_2 = b_3 = 1$. For simplicity, the only selective advantage considered in this paper $s_{i(q)}$ will be the breeding value $A_{i(q)}$, with other forms of environmental influences that are often considered (*e.g.*, litter effects) omitted, and random mating will be assumed. With discrete generations there are just two categories, males and females. In an unselected base generation the phenotypic variance (σ_P^2) is 1 and the additive genetic variance is h_0^2. The categories are, $q = m$ for male and f for female. The notation is included in Table 1.

The regression models required are derived from APPENDICES A and B: $\lambda_{pq} = \iota_p\tau_q(2\sigma_I)^{-1}$ and $\pi_{pq} = \frac{1}{2}(1-k_p\tau_q\rho\sigma_A\sigma_I^{-1})$, where $\tau_m = b_3$ and $\tau_f = b_2(1-X_mX_f^{-1}) + b_3X_mX_f^{-1}$ and $\tau = \frac{1}{2}(\tau_m + \tau_f)$. The τ_q values were used by WWT, and are twice the regression of the index of the candidate on the breeding value of the parent of sex q, σ_I^2 is the variance of the index, and ρ is the accuracy of the index.

After simplification of Equation 7, (see APPENDIX D for further details):

$$\alpha_q = \tfrac{1}{2}X_q^{-1},$$

$$\beta_q = \tfrac{1}{4}\iota(\tau+\tau_q)(\sigma_I+\kappa z)^{-1}X_q^{-1} \qquad (12)$$

where $\kappa = [k\tau + \tfrac{1}{8}(\tau_m-\tau_f)(k_m-k_f)]$ and $z = \rho\sigma_A$. This form is nearly equivalent to that given by WWT, but their derivation proceeded on different (and more complex) lines. Three points of difference should be noted. First, WWT do not include the small $\frac{1}{8}(\tau_m-\tau_f)(k_m-k_f)$ term in κ that arises when *both* the selection intensity and the regression on the parental breeding value differs between the sexes. Second, the indices of WWT are explicitly scaled so that the regression of the breeding value of the candidate on its index is 1 (*i.e.*, $\rho\sigma_A\sigma_I^{-1} = 1$), but scaling does not change $\tau_q\sigma_I^{-1}$ and so α and β do not change with scaling). Finally, in this paper, predictions in equilibrium are obtained using equilibrium parameters.

Rate of gain from sib indices: The decomposition of the rate of gain is achieved using Equation 10 and standard index theory. Within-family gain is given by:

$$\sum_q^{n_c} X_q\alpha_q E[a_{i(q)} \mid i \text{ selected}]$$

$$= \sum_{q=m,f} \tfrac{1}{4}h_0^2\iota_q\tau_w\sigma_I^{-1} = \tfrac{1}{2}h_0^2\iota\tau_w\sigma_I^{-1}$$

since $\alpha_q = \frac{1}{2}X_q^{-1}$, and $E[a_{i(q)} | i \text{ selected}] = \frac{1}{2}h_0^2 \iota_q \tau_w \sigma_I^{-1}$, where τ_w is the regression of the index I on $a_{i(q)}$ [$\tau_w = b_1(1-n_F^{-1}) + b_2(n_F^{-1}-n_H^{-1}) + b_3 n_H^{-1}$]. The total between-family gain is given by:

$$\sum_{q=1}^{n_c} X_q \beta_q^T E[(s_{i(q)} - \bar{s}_q) a_{i(q)} | i \text{ selected}] =$$

$$\sum_{q=m,f} \frac{1}{8}h_0^2 \iota(\tau+\tau_q)(1-k_q\tau_w z\sigma_I^{-1})(\sigma_I + \kappa z)^{-1}$$

since $cov(a_{i(q)}, A_{i(q)}) = \frac{1}{2}h_0^2(1-k_q\tau_w z\sigma_I^{-1})$ for the selected individuals in category q.

The total gain, summed over both sexes and including both between- and within- family gain is, after simplification,

$$\Delta G_{eq} = \frac{1}{2}h_0^2 \iota(\tau_w + \tau)(\sigma_I + \kappa z)^{-1} \qquad (13)$$

This uses the result $k_m\tau_m + k_f\tau_f =$

$$\frac{1}{2}(k_m+k_f)(\tau_m+\tau_f) + \frac{1}{2}(k_m-k_f)(\tau_m-\tau_f) =$$

$$2k\tau + 1/2(k_m-k_f)(\tau_m-\tau_f).$$

Consistency with other approaches: Equation 13 for equilibrium ΔG_{eq} can be compared to the standard formula $\Delta G = \iota\rho\sigma_A = \iota z$. Equation 13 comes from considering the gain achieved from a single cohort over all subsequent generations, whereas the standard formula comes from considering the gain achieved by all previous generations over a single cohort. For an equilibrium, the two forms must be equal, and equating them results in a quadratic equation for z:

$$\kappa z^2 + \sigma_I z - \frac{1}{2}h_0^2[\tau_w + \tau] = 0 \qquad (14)$$

Equation 14 can be obtained as an equilibrium condition when using standard index theory with:

$$\sigma_A^2 = \frac{1}{2}h_0^2 + \frac{1}{4}\sigma_A(1-k_m\rho^2) + \frac{1}{4}\sigma_A^2(1-k_m\rho^2)$$

and $\text{cov}(A,I) = \rho\sigma_A\sigma_I$.

This demonstrates a consistency of the methods presented in this paper (in particular those detailed in APPENDICES A and B) with results from classical index theory for discrete generations. Thus the decision to neglect the second order correction for the Bulmer effect when deriving π_{pq} in APPENDIX B (*i.e.*, omitting the correcting of the genetic variance of the selected parents for selection among their offspring) is also implicit in standard index theory.

Equation 14 can be used to give reasonable estimates of equilibrium gain for indices, even when using unselected base parameters, since many of the terms are constant over time. To use Equation 14 only the base generation value of σ_I is required to solve the quadratic equation for z, and then gain is estimated by ιz. Using Equation 14 to obtain z results in underestimates, rather than the overestimates obtained using base parameters and ignoring linkage disequilibrium. However, the magnitude of the errors from Equation 14 are smaller (Table 2). Estimates from Equation 14 are not expected to be precise, since they assume σ_I constant. Further improvements to Equation 14 would require an iterative scheme in combination with $\sigma_I^2 = \sigma_I^2 - \frac{1}{4}z^2([b_2^2(1-X_mX_f^{-1}) + b_3^2X_mX_f^{-1}](k+k_f) + b_3^2(k+k_m))$. The consistency with standard index theory shows that this will lead to the same result as the usual procedures for deriving equilibrium

TABLE 2.–Approximations to equilibrium rates of gain for general half-sib indices using formulae developed in the text, either with initial or equilibrium parameters. The schemes assumed 20 male and 40 female parents with eight offspring per litter. Initial heritability was assumed equal to 0.4. (z is the product of the index accuracy and the genetic standard deviation).

Index			Rates of gain using initial parameters			Equilibrium parameters
b_1	b_2	b_3	iz	Equation 13[b]	Equation 14[c]	
1	1	0	0.353	0.344	0.344	0.344
1	1	1	0.584	0.444	0.466	0.480
1	2	2	0.650	0.437	0.481	0.511
1	3	6	0.658	0.403	0.457	0.502

[a]Uses z calculated in the base generation. [b]Uses both z and σ_I calculated in the base generation. [c]Uses σ_I calculated in the base generation.

gain (*i.e.*, by iterating on the index accuracy and the genetic variance among the parents).

Expected long-term contributions for Best Linear Unbiased Predictors (BLUP): The analysis of individual long-term contributions can be extended to BLUP evaluation and indices based directly upon it. This analysis will be in discrete generations. An initial consideration for the model proposed, is what form $s_{i(q)}$ will take. In sib-indices, $s_{i(q)}$ was simply the breeding value $A_{i(q)}$ since it is the only means by which a parent may influence its offspring over multiple generations (in the absence of common environmental effects *etc.*). With BLUP this is no longer the case, since the parental information contributes to the evaluation of the offspring, and so a parent's phenotypic record, which also includes its environmental components, will continue to affect selection of the offspring. The information available is summarized in the BLUP estimate $\hat{A}$ (the EBV), which is updated over time. This does not entirely summarize $s_{i(q)}$, since the remaining prediction error at a given time will have a residual effect influencing the subsequent selection success of descendants.

However, some assessment of what should be included in the model may be made. Here three terms are considered for individual i in category q: $\hat{A}_{i(q)}$, the 'initial EBV' at the point of selection of i; $\delta\hat{A}_{i(q)}$, the 'increment' in the EBV at the point of selection of its offspring; and $\hat{e}_{i(q)}$, the remaining 'prediction error' of the parent at the selection of offspring. Selection of i itself is determined by $\hat{A}_{i(q)}$, the selection of the offspring is influenced by $\hat{A}_{i(q)}$ and $\delta\hat{A}_{i(q)}$, whilst selection of grand-offspring and subsequent generations is influenced by all three. Therefore, $\mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^T(s_{i(q)} - \bar{s}_q)$, where $s_{i(q)}$ is a vector comprising $\hat{A}_{i(q)}$, $\delta\hat{A}_{i(q)}$ and $\hat{e}_{i(q)}$. APPENDIX E shows the derivation of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Pi}$ using APPENDICES A and B. $\mu_{i(q)}$ is then determined from Equation 7. The decomposition of the rates of gain was calculated using results on the covariance of the Mendelian sampling terms with $s_{i(q)}$, which are also given in APPENDIX E.

An example of the application is given in Table 3, where predictions are compared to simulations with selection based upon pseudo-BLUP as described by WRAY and HILL (1989). Excellent agreement was found between simulations and predictions, both for the regressions and the total gain (0.508 units for simulation, 0.518 units for predicted). Gain can also be predicted by using

TABLE 3.–A comparison of simulated and predicted responses, and β, for selection using BLUP in discrete generations. The scheme has 20 male and 40 female parents with eight offspring per litter and $h^2 = 0.4$. Simulation results are from 400 replicates.

	Simulation		Equilibrium	
	Male	Female	Male	Female
$\beta(\hat{A})$	0.056	0.032	0.057	0.029
$\beta(\delta\hat{A})$	0.103	0.068	0.103	0.067
$\beta(\hat{e})$	0.011	0.006	0.012	0.006
ΔG(within)	-	-	0.135	0.104
$\Delta G(\hat{A})$	-	-	0.012	0.016
$\Delta G(\delta\hat{A})$	-	-	0.110	0.094
$\Delta G(\hat{e})$	-	-	0.020	0.027
Total	0.507		0.518	

the formulae $\iota\rho\sigma_A$, which was very close to 0.518 when using equilibrium parameters.

The results show that, with BLUP, the primary source of between-family selection among ancestors is the increment in the EBV between their own selection and that of their offspring. The initial EBV plays the least important role, with slightly more between-family gain (in this example) coming from the prediction error. The magnitude of gain arising from the prediction error was a major source of discrepancy between using equilibrium and base parameters, since the latter parameters predicted very little gain from this source.

Extensions to other inheritance modes in the absence of allelic interactions: Extensions of the model to other inheritance modes, such as additive maternal effects or X-linked variation, are made by defining the variables in $s_{i(q)}$ and their impact on λ_{pq} and π_{pq}. As an example, results with maternal imprinted variation are given, where the passage of genes from parent to offspring follows normal Mendelian inheritance, but only the alleles passed to the offspring by the dam are expressed and affect the phenotype. For maternal imprinting, the breeding value can be split into the 'expressed' breeding value (A^+) inherited from the dam, and the 'latent' breeding value (A^-) inherited from the sire and not expressed.

Define $s_{i(q)} = (A_{i(q)}^-, A_{i(q)}^+)$, with discrete generations giving two categories, m for males and f for females. In this case, λ_{pm} will be zero, since the genes passed by the sire do not affect selection of its offspring. However, λ_{pf} will depend on both breeding values, since although A^- is not expressed in the dam, it is expressed in its offspring. For π_{pq}, there will be a dependence on both breeding values: genes passed by the sire only affect A^-, and genes passed by the dam only affect A^+. Since genes passed by the sire are not expressed, the regression of offspring on parent is unaffected by selection. Therefore, applying APPENDICES A and B,

$$G = (\tfrac{1}{2}, \tfrac{1}{2} \mid \tfrac{1}{2}, \tfrac{1}{2})$$

$$\mathbf{A} = (0.0, 0.0, \tfrac{1}{2}\iota_m\sigma_P^{-1}, \tfrac{1}{2}\iota_m\sigma_P^{-1} \mid 0.0, 0.0, \tfrac{1}{2}\iota_f\sigma_P^{-1}, \tfrac{1}{2}\iota_f\sigma_P^{-1})$$

$$\mathbf{\Pi} = [\tfrac{1}{2}, \tfrac{1}{2}, 0.0, 0.0 \mid 0.0, 0.0, \tfrac{1}{2}(1-k_m h^2), \tfrac{1}{2}(1-k_m h^2) \mid \tfrac{1}{2}, \tfrac{1}{2}, 0.0, 0.0 \mid 0.0, 0.0, \tfrac{1}{2}(1-k_f h^2), \tfrac{1}{2}(1-k_f h^2)]$$

where $h^2 = \mathrm{Var}(A^+)/\sigma_P^2$, and the phenotypic variance, σ_P^2, is the sum of the variance of A^+ and the environmental variance. Equation 7 was used to obtain β.

Predictions were made using variance parameters obtained after iteration to equilibrium. To calculate ΔG, the expected values of the Mendelian sampling terms for selected individuals and the covariance with $s_{i(q)}$ for

TABLE 4.–A comparison of simulated and predicted responses, and β, for selection with maternally imprinted variation in discrete generations. The scheme has 20 male and 40 female parents with six offspring per litter and $h^2 = 0.4$. Standard errors of simulation estimates are given in parenthesis.

	Sex	Simulation	parameters used	
			Equilibrium	Initial
β (×100)[a]	male	0.69 (0.041)	0.73	0.71
	female	1.07 (0.016)	1.09	1.06
ΔG (between families)	male	0.023 (0.013)	0.025	0.024
	female	0.073 (0.014)	0.076	0.072
ΔG (within families)	male	0.076 (0.0012)	0.077	0.075
	female	0.055 (0.0008)	0.056	0.055
Total ΔG		0.229 (0.0014)	0.234	0.223

[a]The predictions of β for A^- and A^+ were identical and simulations were not significantly different. Therefore results have been pooled.

selected individuals were calculated using standard index theory.

$$E[(a_m^-, a_m^+, a_f^-, a_f^+)] =$$

$$(0.0,\ \tfrac{1}{2}h_0^2\iota_m\sigma_P^{-1},\ 0.0,\ \tfrac{1}{2}h_0^2\iota_f\sigma_P^{-1})$$

$$cov[s_{i(q)}, (a_{i(q)}^- + a_{i(q)}^+)] =$$

$$\tfrac{1}{2}h_0^2\ [1.0, (1.0-k_m h^2), 1.0, (1.0-k_f h^2)]^T$$

Since this is imprinted variation, half the genes from an ancestor will be expressed, and half will be latent in the long term. Gains predicted from Equation 3, therefore, should be halved.

Table 4 shows very close agreement between simulation results and predictions. The gains within families shown in Table 4 entirely arise through the expressed breeding value of the candidates (achieved at the time of selection among the candidates). For each sex, the predictions for the regression coefficients of long-term contributions upon A^- and A^+ were identical, and this was supported by the simulations. This is an expression of the fact that the gene expression depended upon the parent but not upon the grandparent. Approximately 0.6 of the between-family gains shown in Table 4 arise from selection on the latent breeding value (A^-) of the candidates, and since the regressions are identical, this effect may be ascribed to the larger genetic variance associated with this term (it is not reduced through the initial selection of the ancestors). The regression of long-term genetic contributions on breeding values was greater in females than in males, despite their greater number, which is not surprising given the mode of inheritance.

Prediction using the base generation heritability and phenotypic variance was also accurate, indeed it appeared more accurate than with equilibrium parameters. However, in the results presented in Table 4, the covariances which yield the between-family selection predictions have not been reduced for finite numbers of parents, *i.e.*, $(1-X_q^{-1})$, which would result in increased precision for predictions using the equilibrium parameters and increased bias for predictions using the base generation parameters.

Overlapping generations: An example of application with overlapping generations is

presented for mass selection, with a fixed number of parents selected at each age, in a two-path scheme (*i.e.*, there was no subdivision of breeding individuals into males to breed males, males to breed females *etc.*). The general approach is explained in more detail by BIJMA and WOOLLIAMS (1999). The steps will be illustrated using a scheme with three categories: 20 males breeding at one year of age, 20 females breeding at one year of age and 20 females breeding at three years of age respectively. The number of offspring per litter was eight and the trait was assumed to have a heritability of 0.4. The age groups not used for parents will be omitted, so the categories are: males aged one (category 1), females aged one (category 2), and females aged three (category 3).

1. The genetic make-up of the newborn are described by $g_{0,p1}$, $g_{0,p2}$, and $g_{0,p3}$. These are 0.5, 0.25, and 0.25 respectively for all categories p. From the $g_{0,pq}$, and the number of parents and family sizes, the selection intensities (ι_p) and variance reduction coefficients (k_p) were calculated for each category: ι_p=1.647, k_p=0.817, *i.e.*, the same for all three categories.

2. An initial ΔG was assumed as a starting point for iteration. In the following, the starting point was ΔG calculated from standard gene flow (HILL 1974). After iterating to equilibrium, this was $\Delta G = 0.412$.

3. The genetic value of selected parents in category p was $\iota_p h^2 \sigma_P - [age(p)-1]\Delta G$. Deviations from the overall means of the selected males and females was $\boldsymbol{\delta} = (0, +0.412, -0.412)$, *i.e.*, the one-year-old female parents had breeding values 0.412 units above average, and the three-year-old female parents had breeding values 0.412 units below average.

4. Before selection, genetic variance in category p was calculated using the pooled variance within categories plus between categories plus the Mendelian sampling variance:

$$\tfrac{1}{2}h_0^2 + \sum_q \left[\tfrac{1}{4}\sigma_A^2(2g_{0,pq})(1-k_q h^2) + \tfrac{1}{4}(2g_{0,pq})\delta_q^2 \right]$$

This was 0.370 for all p, and the phenotypic variance was, $\sigma_P^2 = 0.970$ for all p.

5. $\boldsymbol{G}$ was calculated using an truncation algorithm to find a truncation point for a given upper-tail probability for a mixture of Normal distributions. The algorithm was used twice for the selection of candidates in each category, firstly to obtain the genetic make-up from sire categories and then to obtain the genetic make-up from dam categories. For category p candidates, the mixing proportions for the Normal distributions were $2g_{0,pq}$ (q = 1, 2, 3), *i.e.*, the frequency of the candidates with parent category q. The means of the Normal distributions were the deviations of the candidates with parent category q from the mean of all like-sexed candidates, *i.e.*, $\tfrac{1}{2}\delta_q$. The variance was assumed independent of parent category q, and was the phenotypic variance adjusted for the component of genetic variance between categories of the same sex as parent category q, *i.e.*, $\sigma_P^2 - \sum_{q^* \text{ same sex as } q} \tfrac{1}{4}(2g_{0,pq^*})\delta_{q^*}^2$. In the first iteration, each row of $\boldsymbol{G}$ was (0.5, 0.336, 0.164), thus indicating that, although the dams of ages one and three provided equal numbers of candidates, the candidates with dams of age one were expected to be twice as successful in having selected offspring.

6. The $\boldsymbol{\Lambda}$ and $\boldsymbol{\Pi}$ matrices were constructed according to APPENDICES A and B respectively.

For mass selection, $\pi_{pq} = 0.5(1-k_p h^2)$, and $\lambda_{pq=}0.5\iota_p\sigma_P^{-1}$. In the first iteration, $\mathbf{\Pi} = 0.344\ \mathbf{1}\mathbf{1}^T$ where $\mathbf{1}^T = (1, 1, 1)$, $\mathbf{\Lambda} = 0.836\ \mathbf{1}\mathbf{1}^T$, and $\boldsymbol{D} = \mathbf{1}$ (0,0.092,-0.188). The result for $\boldsymbol{D}$ indicates that the breeding value of a selected individual (of any category p) with a dam of age one is expected to be 0.28 greater than a selected individual of the same category with a dam of age three.

7. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were calculated according to Equations 7b and 9. In the first iteration, $(\boldsymbol{N\alpha})^T$ = (0.395, 0.289, 0.106) and $(\boldsymbol{N\beta})^T$ = (0.503, 0.338, 0.165).

8. The covariance of the Mendelian sampling term with the breeding values were calculated and ΔG was updated using Equation 11. This uses the result that $E[a_{i(q)}] = \frac{1}{2}h_o^2\iota_q\sigma_P^{-1}$, and after selection, $cov(a_{i(q)},A_{i(q)}) = \frac{1}{2}h_o^2(1- k_q h^2)$.

9. Steps 3 through 8 were repeated until convergence.

Results after convergence were, $\boldsymbol{\alpha}$ = (0.0200, 0.0149, 0.0050)^T and $\boldsymbol{\beta}$ = (0.0255, 0.0171, 0.0084)^T. Predicted gain within families was (0.134, 0.100, 0.034), and predicted gain between families was (0.067, 0.045, 0.022), giving a total gain of 0.402. At equilibrium $\boldsymbol{G}$ was 1(0.500, 0.335, 0.165). This was compared to simulation results for 1000 replicates, giving $\boldsymbol{\alpha}$ = (0.0197, 0.0145, 0.0052)^T with a maximum *s.e.* of 0.0009, $\boldsymbol{\beta}$ =(0.0249, 0.0175, 0.0071)^T with a maximum *s.e.* of 0.0004, and a total gain of 0.398 with a *s.e.* of 0.001. Thus very close agreement between simulations and predictions was obtained. As in discrete generations, the gain from mass selection was evenly divided between males and females. The gene flow predicted using HILL (1974) is, $\boldsymbol{\alpha}$ = (0.0167, 0.0083, 0.0083). HILL (1974) makes no prediction of $\boldsymbol{\beta}$.

The generation interval, defined by the time taken to turn over the genes once, was predicted from $(\Sigma X_q\alpha_q)^{-1}$ to be 1.25 (*cf.* 1.26 with *s.e.* 0.01 in the simulations), which was notably shorter than the average age of the parents. This was because of the cumulative effect of the selective advantage of the younger age group of females. Although they produced equal numbers of offspring, they produced more than twice as many parents. However, the generation interval was not predictable from the equilibrium $\boldsymbol{G}$ alone (*i.e.*, accounting for a single generation of selective advantage), since this would have predicted an interval of 1.33 (*i.e.*, 0.5×1 + 0.335×1 + 3×0.165).

To obtain the time course of the contributions, APPENDIX C was used. APPENDIX C needs the following matrices based on $\boldsymbol{G}$,

$\boldsymbol{G}_1$ (0.500, 0.335, 0.165 | 0.0, 0.0, 0.0 | 0.0, 0.0, 0.0)

$\boldsymbol{G}_2$ (0.0, 0.0, 0.0 | 0.500, 0.335, 0.165 | 0.0, 0.0, 0.0)

$\boldsymbol{G}_3$ (0.0, 0.0, 0.0 | 0.0, 0.0, 0.0 | 0.500, 0.335, 0.165)

The results are shown in Table 5, for the time course of contributions from category 2. The contributions converged in cohort 10.

Discussion

This study has developed a framework for predicting the expected genetic contributions of individuals and categories of individuals, under a wide range of selection and inheritance models. This framework allows selection to be more properly accounted for compared to existing gene-flow methods for overlapping generations and multiple breeding groups (such as that presented by HILL 1974). Furthermore, it advances understanding by considering the differential gene flow among individuals within

TABLE 5.–The time course of expected contributions from an individual female parent of age one at $t = 0$. The breeding scheme has mass selection with 20 male parents of age 1, 40 female parents at ages one and three (20 at each age), eight offspring per litter and heritability 0.4. The expected contribution is $c(t) + b(t)(A_i - \overline{A})$.

	To males age one		To females age one		To females age three	
Time	$c(t)$	$b(t)$	$c(t)$	$b(t)$	$c(t)$	$b(t)$
$t = 1$	0.0167	0.0140	0.0167	0.0140	0	0
$t = 2$	0.0151	0.0157	0.0151	0.0157	0	0
$t = 3$	0.0132	0.0146	0.0132	0.0146	0.0167	0.0140
$t = 6$	0.0148	0.0168	0.0148	0.0168	0.0145	0.0160
$t = 10$	0.0149	0.0170	0.0149	0.0170	0.0149	0.0170

categories, an extension not hitherto achieved except in some special cases. The framework has been constructed by first modelling the selection process and the transfer of selective advantages within a single generation of selection, and second, extending this to multiple generations. Two regression models are required, both of which are derived using standard index theory. First, a model describing the expected number of selected offspring that a parent may have (A), and second, a model describing the relationship of the selective advantages of a selected offspring with those of its parent (Π). Predictions of genetic gain directly follow from the expected long-term contributions. Unlike $i\rho\sigma_A$, the relationship between gain and contributions (Equations 3 and 10) shows that gain comes from generating a covariance between the long-term contributions and the new variance arising in the population (*i.e.*, the Mendelian sampling variation) in each cohort, thus changing the description of gain from a statistical one to a genetical one.

The framework has been developed to describe the expected genetic contribution over all time horizons, from the short-term to the long-term. The novel, closed formulae (Equations 7 and 9), developed to predict the expected long-term contribution of an ancestor, rely on the assumption of equilibrium in the selection process. If there is no equilibrium, the error will depend on the relative degree of departure in relation to the timescale of convergence of the contributions (approximately five generations). However, this assumption is not necessary for the use of Equations 11, where contributions are predicted over finite time periods, but more effort may be required to define the changes in the necessary parameters if there is no equilibrium.

In the development of the framework, the effects of inbreeding on parameters and progress have been neglected, but this is not a serious problem. First, the timescale for the convergence of contributions is small in comparison to the timescale for the effects of inbreeding on parameters in breeding schemes, especially where inbreeding is controlled to be at reasonable levels. The impact of individuals within a cohort is very largely decided within five generations, and even within this period, the scope for controlling an individual's contribution declines exponentially (the scope can be measured by the variance of an individual's contribution within the population). A second reason is that schemes will most usefully be compared at the same rates of inbreeding, and so the neglect of inbreeding is less likely to bias the comparisons made.

The expected long-term contribution has been described in a general linear form, $\alpha_q + \boldsymbol{\beta}_q^T(s_{i(q)} - \bar{s}_q)$, where $s_{i(q)}$ is a vector of selective advantages for an ancestor i. Judged by the accuracy of the results in this study, the omission of quadratic terms from the model has not led to serious errors in predicting the rates of gain, or in the linear component of relationship between the long-term contribution and the selective advantages. Quadratic terms in s do not affect the prediction of rates of gain, unless terms of the order $E[s^2a]$ are significant (which will involve the skewness of a after selection), and will not influence the predicted rate of inbreeding unless higher moments than the variance of s are considered (WOOLLIAMS and THOMPSON 1994). The linear approximations used in the applications, and presented in the APPENDICES, were robust.

The $\boldsymbol{\alpha}$ represents the proportion of genes that derive from the various categories as a whole, and these differ qualitatively from predictions using HILL (1974), since the earlier study does not account for the inheritance of selective advantages. The impact of this may be particularly great where breeding structures that are subject to selection, are subdivided, with migration taking place between the subdivisions. In these circumstances, ignoring the selective advantage between groups will overestimate the impact of groups of lesser merit and underestimate the impact of groups of greater merit. The consequences of these errors may be the maintenance and use of subdivisions that have little potential to contribute in the long-term, and a greater rate of inbreeding in the population than had been anticipated (BIJMA *et al.* 2000) . The framework presented here and that of HILL (1974) give the same prediction of $\boldsymbol{\alpha}$ when selection is at random, since (*i*) elements of $\boldsymbol{G}$ are identical to $g_{0,pq}$, (*ii*) $\boldsymbol{\Pi} = 0$, and (*iii*) $\boldsymbol{\Lambda} = 0$.

The genetic contribution of an individual represents the expected impact that its Mendelian sampling term has on the population. Within a cohort, the magnitude of the contribution made by an individual will depend upon the breeding categories in which it is included over its lifetime. In any newborn cohort, even when generations overlap, the males are expected to have a total long-term contribution equal to those of the females, *i.e.*, $\sum_{male\ categories} X_q\alpha_q = \sum_{female\ categories} X_q\alpha_q$. When generations are discrete, these sums are equal to ½, but when generations overlap the sums will be less than ½.

The sum of the total contributions from any one cohort, including both sexes, is a natural measure of the rate at which genes in the population are renewed. In particular the rate measured by the $\Sigma X_q\alpha_q$ places an emphasis upon those contributions that are destined to remain in the population in the long-term. Thus $(\Sigma\, X_q\alpha_q)^{-1}$ is the period of time for the population to complete a cycle of renewal, and is a measure of the generation interval L. The generation interval defined by the long-term contributions is shorter than the traditional 'average age of the parents at the birth of their offspring' for the examples considered, because the younger breeding groups had a selective advantage so that the progeny of older parents were less likely to be selected. The need for a modified generation interval arising from the inheritance of the selective advantage has been considered previously (BICHARD *et al.* 1973; JAMES 1977). BICHARD *et al.* (1973) argued that the traditional generation interval

might be usefully modified to account for non-randomness among parental age-groups in the survival of their offspring to produce grand-offspring. This is what occurs with the inheritance of selective advantage between categories of different ages. For example, such a modification would exclude from the calculation of generation interval those parents whose sole purpose is to produce a commercial cohort outside the breeding population. JAMES (1977) moved the argument forward by considering the generation interval calculated from only those parents with selected offspring, and showed that for the purposes of calculating rates of genetic gain either definition of L would suffice, providing the calculation of the selection differential is matched to the definition of the generation interval.

The average age of the parents might generally be considered to refer to the age at the birth of unselected offspring. The definition of JAMES (1977) considers the average age of the parents at the birth of the selected offspring, who will then produce the unselected *grandoffspring*. These definitions may be viewed as a one-generation estimate of the generation interval and an iteration beyond this respectively, whereas the calculation from long-term contributions represents the converged estimate. The definition of the generation interval from long-term contributions avoids any debate on what parents should or should not be included. The average age of the parents at the birth of their unselected offspring will remain of operational significance to breeding schemes, but the generation interval defined by the long-term contributions is an unambiguous *genetic* property of a population.

The consistency of the framework with other approaches for estimating gain in discrete generations is important, but this consistency does not extend to overlapping generations. The main approach for prediction of gain in overlapping generations is that of RENDEL and ROBERTSON (1950). The formula obtained by RENDEL and ROBERTSON was also obtained by HILL (1974) as a consequence of deriving the traditional gene flow, and this apparent consistency added credence to both the approach and the wider results of traditional gene flow. However, this study shows that this consistency is not justified. The estimates of equilibrium gain using contributions and RENDEL and ROBERTSON differ slightly from each other. The estimate of gain from contributions arises from the prospective analysis of the impact of a single cohort to the future population over the long-term. In contrast, the estimate of gain from RENDEL and ROBERTSON (1950) arises from a retrospective analysis of the impact of selection in the whole population to a single cohort. (See CHAPTER 4 for further discussion.) One reason why differences between these approaches might be expected with overlapping generations lies in the calculation of selection differentials, since each cohort is a mixture of many truncated Normal distributions.

The second component of the expected long-term contribution is the linear regression on the selective advantages of an individual (β). These terms describe the expected differential contributions within a category, that will occur during the selection process as a result of the differences in selective advantages. These differential contributions represent the success of one ancestor's descendants over those from another ancestor, and therefore measure the expected extent of between-family selection. The between-family selection is responsible for the

greater rates of inbreeding that can occur when selection is practised, and the control of the magnitude of the regression coefficients (and the components of s) is an important aspect of methods to optimize genetic gain with constrained inbreeding rates (*e.g.*, VERRIER *et al.* 1993; VILLANUEVA and WOOLLIAMS 1997).

The between-family selection may develop very quickly, so that its extent is largely established in the selection of the progeny, or more slowly. This time-course is controlled by $\boldsymbol{G}\circledast\boldsymbol{\Pi}$ and powers of $\boldsymbol{G}\circledast\boldsymbol{\Pi}$, which describe the decay of the ancestor's selective advantage through progeny [see Equation 11c for $f_q(t)$]. This rate of decay is controlled by the eigenvalues of $\boldsymbol{G}\circledast\boldsymbol{\Pi}$. In the example given for BLUP, the maximum eigenvalue of $\boldsymbol{G}\circledast\boldsymbol{\Pi}$ was 0.18, which may be compared to 0.36 for mass selection with the same numbers of parents and the same initial heritability. Therefore, it is clear that a higher proportion of the ultimate between-family selection, generated by selection with BLUP, is achieved in the first and second generations after the ancestor, than is the case with mass selection. This difference has a consequence for the accuracy of the prediction of rates of inbreeding using techniques accounting for co-selection in one- and two- generations (WRAY *et al.* 1990), and explains why these methods are notably more accurate with BLUP selection than with mass selection (T.H.E. MEUWISSEN personal communication).

The importance of predicting the development of genetic contribution is that risks in breeding schemes, measured by parameters such as ΔF (MEUWISSEN and WOOLLIAMS 1994), cannot be described without a knowledge of the dynamics of individual contributions. The importance of the expected genetic contribution is made greater by the result of WOOLLIAMS (1998), who indicated that ΔF may be predicted from the expectation alone. The framework presented here provides a step-by-step recipe for predicting this expected genetic contribution over multiple generations. In providing the results, particular approaches have been described to derive the necessary regression models (APPENDICES A and B). In other situations, such as the use of quadratic indices (MEUWISSEN 1997; GRUNDY *et al.* 1998) the formulae given in the Appendices, based upon truncation selection, may not be appropriate, whereas the results given in Equations 7, 9, 10 and 11 may remain valid. Therefore, it is important to recognize that the details of these Appendices are not an integral part of the recipe, and other approaches could replace them in the recipe to suit the needs of a particular study.

JAW gratefully acknowledges the Ministry of Agriculture, Fisheries and Food (UK) for funding this work, the encouragement of Dr. P.M. Visscher, Dr. B. J. McGuirk, and Prof. A. Maki-Tanila and Prof. B. Kinghorn for providing opportunities to develop and complete it. PB gratefully acknowledges financial support from the Netherlands Technology Foundation (STW), and BV gratefully acknowledges financial support from the Biotechnology and Biological Sciences Research Council.

Literature Cited

BICHARD, M., A. H. R. PEASE, P. H. SWALES and K. ÖZKÜTÜK, 1973 Selection in a population with overlapping generations. Animal Production 17: 215-227.

BIJMA, P. and J. A. WOOLLIAMS, 1999 Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151: 1197-1210.

BIJMA, P., J. A. M. VAN ARENDONK and J.A. WOOLLIAMS, 2000 Prediction of rates of inbreeding in populations with

overlapping generations under mass selection. Genetics 154.

BULMER, M. G. 1971. The effect of selection on genetic variability. American Naturalist 105: 201-211.

BULMER, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

DEKKERS, J. C. M. 1992. Asymptotic response to selection on best linear unbiased predictors of breeding values. Animal Production 54: 351-360.

GRUNDY, B., B. VILLANUEVA and J. A. WOOLLIAMS, 1998 Dynamic selection procedures for constrained inbreeding and their consequence for pedigree development. Genetical Research (Cambridge) 72: 159-168.

HILL, W. G., 1974 Prediction and evaluation of response to selection with overlapping generations. Animal Production 18: 117-139.

JAMES, J. W., 1977 A note on the selection differential and generation length when generations overlap. Animal Production 24: 109-112.

JAMES, J. W. and G. MCBRIDE, 1958 The spread of genes by natural and artificial selection in a closed poultry flock. Journal of Genetics 56: 55-62.

MEUWISSEN, T. H. E. and J. A. WOOLLIAMS, 1994 Response versus risk in breeding schemes. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production 18: 236-243.

MEUWISSEN, T. H. E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. Journal of Animal Science 75: 934-940.

RENDEL, J. M. and A. ROBERTSON, 1950 Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. Journal of Genetics 50: 1-8.

VERRIER, E., J.J. COLLEAU and J.L. FOULLEY, 1993 Long term effects of selection based on the animal model BLUP in a finite population. Theoretical and Applied Genetics 87: 446-454.

VILLANUEVA, B. and J. A. WOOLLIAMS, 1997 Optimization of breeding programmes under index selection and constrained inbreeding. Genetical Research (Cambridge) 69: 145-158.

VILLANUEVA, B., J.A. WOOLLIAMS and B. GJERDE, 1996. Optimum designs for breeding programmes under mass selection with an application in fish breeding. Animal Science 63: 563-576.

WOOLLIAMS, J. A., 1998 A recipe for the design of breeding schemes. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production 25: 427-430.

WOOLLIAMS, J. A. and R. THOMPSON, 1994 A theory of genetic contributions. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, 19: 127-134.

WOOLLIAMS, J. A., N. R. WRAY and R. THOMPSON, 1993 Prediction of long-term contributions and inbreeding in populations undergoing mass selection. Genetical Research (Cambridge) 62: 231-242.

WRAY, N. R. and W. G. HILL, 1989 Asymptotic rates of response from index selection. Animal Production 49: 217-227.

WRAY, N. R. and W. G. HILL, 1989. Asymptotic rates of response from index selection. Animal Production 49: 217-227.

WRAY, N. R. and R. THOMPSON, 1990 Prediction of rates of inbreeding in selected populations. Genetical Research (Cambridge) 55: 41-54.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1990 Methods for predicting rates of inbreeding in selection populations. Theoretical and Applied Genetics 80: 503-512.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1994 Prediction rates of inbreeding in populations undergoing index selection. Theoretical and Applied Genetics 87: 878-892.

Appendix A

A general approximation to λ_{pq}: The regression of selection score of the unselected candidates of category p on the index I is given by $\omega_p \iota_p / \sigma_I$ (WRAY and THOMPSON 1990), where ω_p is the selection proportion for category p. For a parent i of category q, the regression of the candidate index on $s_{i(q)}$ for all the parents of category p that are of the same sex as category q was derived by standard index theory appropriate to the inheritance model under consideration (denote the coefficients for the regression on $(s_{i(q)} - \bar{s})$ by w).

For each offspring of the parent from group q the probability of selection can then be approximated by $\omega_p(1 + \iota\sigma_I^{-1} w^T(s_{i(q)} - \bar{s}))$. The expected number of offspring for a parent of category p is then $n_p\omega_p(1 + \iota\sigma_I^{-1} w^T(s_{i(q)} - \bar{s}))$ where n_p is the number of candidates in category p per parent. $n_p\omega_p$ is equal to

or $2g_{0,pq}X_pX_q^{-1}$, where g_0 the proportion of genes among the newborn category p that derive from category q.

Considering only category q parents, they have an average selective advantage given by $\bar{s}_q$ so the expectation is $2g_{0,pq}X_pX_q^{-1}$ $[1+\iota\sigma_I^{-1}w^T(s_{i(q)}-\bar{s}_q)$ + $+\iota\sigma_I^{-1}w^T(\bar{s}_q-\bar{s})]$. For sufficiently small deviations this is approximately $2g_{0,pq}X_pX_q^{-1}$ $[1+\iota\sigma_I^{-1}w^T(s_{i(q)}-\bar{s}_q)]$ $[1+\iota\sigma_I^{-1}w^T(\bar{s}_q-\bar{s})]$ where the last term in the product may be viewed as the additional selective advantage of category q, and so $g_{0,pq}(1+\iota\sigma_I^{-1}w^T(\bar{s}_q-\bar{s})) \approx g_{pq}$ and $\lambda_{pq} \approx \iota\sigma_I^{-1}w$.

Appendix B

Derivation of π_{pq}: Let $s_{i(q)}$ be the vector of deviations of explanatory variables from their mean for a parent in category q and $s_{j(p)}$ for an unselected progeny in category p and likewise $I_{j(p)}$ be the index upon which will decide the selection, or otherwise, of $j(p)$. Let $s = (s_{i(q)}^T \mid s_{j(p)}^T \mid I_{j(p)})$ have the partitioned (co)variance matrix

$$\begin{pmatrix} V_{qq} & V_{pq}^T & \nu_q \\ V_{pq} & V_{pp} & \nu_p \\ \nu_q^T & \nu_p^T & \sigma_I^2 \end{pmatrix}$$

Before selection among candidates in category p, $s_{i(q)}$ and $s_{j(p)}$ can be expressed as regressions on $I_{j(p)}$:

$$s_{j(p)} = \sigma_I^{-2}\nu_p\, I_{j(p)} + \varepsilon_{j(p)}$$

$$s_{i(q)} = \sigma_I^{-2}\nu_q\, I_{j(p)} + \varepsilon_{i(q)}$$

Equating $E[s_{i(q)}s_{i(q)}^T]$ to V_{qq} gives $E[\varepsilon_{i(q)}\varepsilon_{i(q)}^T] = V_{qq}-\sigma_I^{-2}\nu_q\nu_q^T$ and, similarly $E[\varepsilon_{j(p)}\varepsilon_{i(q)}^T] = V_{pq}-\sigma_I^{-2}\nu_p\nu_q^T$. After selection, Normal distribution theory infers that the regression coefficients on $I_{j(p)}$ are unchanged, but other regression coefficients are changed. Therefore, after selection

$$V_{pq}^* = (V_{pq}-k_p\sigma_I^{-2}\nu_p\nu_q^T),\ V_{qq}^* = (V_{qq}-k_p\sigma_I^{-2}\nu_q\nu_q^T).$$

Let π_{pq} be the matrix of coefficients of $s_{j(p)}$ on $s_{i(q)}$ after selection, then $\pi_{pq} = V_{pq}^* V_{qq}^{*-1}$.

In the applications described this will be approximated by $\pi_{pq} = V_{pq}^* V_{qq}^{-1}$. This is for three reasons: (i) simpler forms; (ii) it coincides with preceding published theory on genetic contributions; and (iii) such an assumption is implicit in standard index theory.

As an example with more than a single variable consider mass selection in discrete generations with random mating, where the vector of selective advantages explicitly includes the breeding value of the mate as well as the individual. There are two categories, males and females. In this case $s_{i(q)}$ has 2 variables for each parent in category q, $(A_{i(q)}-\bar{A}_q, A_{i(q')}-\bar{A}_{q'})$, where $A_{i(q)}$ is the breeding value of i in category q, and $A_{i(q')}$ is the breeding value of its mate, and define $s_{j(p)}$ similarly for the selected progeny $j(p)$. V_{pq}= $(\frac{1}{2}\sigma_A^2(1-k_qh^2), \frac{1}{2}\sigma_A^2(1-k_qh^2) \mid 0, 0)$, ν_p= $(\sigma_A^2 \mid 0)$, ν_q= $(\frac{1}{2}\sigma_A^2(1-k_qh^2) \mid \frac{1}{2}\sigma_A^2(1-k_qh^2))$, V_{qq}= diag$(\sigma_A^2(1-k_qh^2), \sigma_A^2(1-k_qh^2))$, resulting in π_{pq}= $(\frac{1}{2}(1-k_ph^2), \frac{1}{2}(1-k_ph^2) \mid 0, 0)$. These are results of WRAY and THOMPSON (1990). In this example it was chosen to obtain a fuller description of the expected long-term contribution by explicitly including the mate; ignoring the mate is valid for considering genetic gain, providing the matrices are appropriately constructed, *e.g.*, if mating had been assortative rather than random the covariance between parent and offspring breeding value would need to account for the mate implicitly.

Appendix C

Contributions over finite time when α is estimated as a right eigenvector of $(G^T+(G^T\otimes D^T)(I-G^T\otimes\Pi^T)^{-1}(G^T\otimes\Lambda^T))$: Adjustment of equations (7) will be done assuming, for simplicity, the only selective advantage is the breeding value. For category q, a *selected* individual at time 0, the vector of contributions to *selected* individuals in categories at time t is given by $c_q(t) + b_q(t)\,(A_{i(q)}-\bar{A}_q)$.

The approach taken is to use a modified form of Equation 4:

$$r_{i(q)}(t) = \tfrac{1}{2} \sum_{offspring\, j \in category\, p} r_{j(p)}(t-age(p))$$

Therefore the expected contribution after t cohorts is calculated by considering the expected contributions of selected offspring in category p, for t-$age(p)$ cohorts.

Firstly, $\boldsymbol{c}_q(t)$ and $\boldsymbol{b}_q(t)$ are calculated according to equations (11). Then the following iterative scheme is applied where $\boldsymbol{c}^*$ and $\boldsymbol{b}^*$ are the solutions from the previous iteration.

$$\boldsymbol{c}_q(t) = \sum_p X_q^{-1} X_p g_{pq} (\boldsymbol{c}_p^*(t-age(p)) + \boldsymbol{b}_p^*(t-age(p)) d_{pq})$$

$$\boldsymbol{b}_q(t) = \sum_p X_q^{-1} X_p (g_{pq}\lambda_{pq}) \boldsymbol{c}_p^*(t-age(p)) + \sum_p X_q^{-1} X_p (g_{pq}\pi_{pq}) \boldsymbol{b}_p^*(t-age(p))$$

This is repeated until convergence.

Appendix D

The expected long-term contributions for sib indices in discrete generations: For the sib-indices described by WRAY *et al.* (1993): $G = (\tfrac{1}{2}, \tfrac{1}{2} \mid \tfrac{1}{2}, \tfrac{1}{2})$ and $\pi_{pq} = \tfrac{1}{2}(1 - k_p\, \tau_q\, z\, \sigma_I^{-1})$, $\lambda_{pq} = (2\sigma_I)^{-1} \iota_p \tau_q$, where $z = \rho\sigma_A$. From equation (7a) $\alpha = (\tfrac{1}{2}X_m^{-1}, \tfrac{1}{2}X_f^{-1})^T$.

To apply equation (7b) the determinant of $(\boldsymbol{I} - \boldsymbol{G} \otimes \boldsymbol{\Pi})$ is required: $Det(\boldsymbol{I} - \boldsymbol{G} \otimes \boldsymbol{\Pi}) = 1 - \tfrac{1}{2}(\pi_{mm} + \pi_{ff}) + \tfrac{1}{4}(\pi_{mm}\pi_{ff} + \pi_{mf}\pi_{fm}) = \tfrac{1}{2} + \text{terms in } z\sigma_I^{-1}$, where the terms in $z\sigma_I^{-1}$ are simplified by noting $k = \tfrac{1}{2}(k_m + k_f)$, $\tau = \tfrac{1}{2}(\tau_m + \tau_f)$, with the result that: $(k_m\tau_m + k_f\tau_f)/4 - (k_m\tau_m + k_f\tau_f - k_m\tau_f - k_f\tau_m)/16 = \tfrac{1}{2}[k\tau + \tfrac{1}{8}(\tau_m - \tau_f)(k_m - k_f)]$ and $Det(\boldsymbol{I} - \boldsymbol{G}^T \otimes \boldsymbol{\Pi}^T) = \tfrac{1}{2}(1 + \kappa\, z\, \sigma_I^{-1})$, where $\kappa = [k\tau + \tfrac{1}{8}(\tau_m - \tau_f)(k_m - k_f)]$.

$$(\boldsymbol{I} - \boldsymbol{G}^T \otimes \boldsymbol{\Pi}^T)^{-1} = Det(\boldsymbol{I} - \boldsymbol{G}^T \otimes \boldsymbol{\Pi}^T)^{-1} \begin{matrix} (\tfrac{3}{4} + \tfrac{1}{4}k_f\tau_f z, & \tfrac{1}{4} - \tfrac{1}{4}k_f\tau_m z \mid \\ \tfrac{1}{4} - \tfrac{1}{4}k_m\tau_f z, & \tfrac{3}{4} + \tfrac{1}{4}k_m\tau_m z). \end{matrix}$$

$$(\boldsymbol{G}^T \otimes \boldsymbol{\Lambda}^T) = (2\sigma_I)^{-1} \begin{matrix} (\tfrac{1}{2}\iota_m\tau_m, & \tfrac{1}{2}\iota_f\tau_m \mid \\ \tfrac{1}{2}\iota_m\tau_f, & \tfrac{1}{2}\iota_f\tau_f) \end{matrix}$$

Multiplying these matrices according to equation (7b) gives:

$$\boldsymbol{\beta} = \tfrac{1}{4}\iota(\sigma_I + \kappa z)^{-1}(X_m^{-1}(\tau + \tau_m),\ X_f^{-1}(\tau + \tau_f))^T$$

Appendix E

An approximation for Λ and Π when selection is based upon BLUP: The approach is based upon the approximation to BLUP proposed by WRAY and HILL (1989). For a discrete scheme, an EBV for a candidate was calculated using a selection index constructed from six pieces of information: (*i*) the sire's EBV at the time of his selection, (*ii*) the dam's EBV at the time of her selection; (*iii*) the mean EBV of the d mates of the sire; (*iv*) the phenotypic mean of the paternal half-sib family of size dn, including the candidate; (*v*) the phenotypic mean of the full-sib family of size n, including the candidate; and (*iv*) the candidate's phenotype. Denote the (co)variance matrix for these sources by $\boldsymbol{V}$, which is derived by standard index theory (WRAY and HILL, 1989). The parameters used to define $\boldsymbol{V}$, and $\boldsymbol{V}$ itself, are iterated, accounting for the effects of linkage disequilibrium on genetic variance (BULMER, 1971) and the build up of pedigree information (DEKKERS, 1992) until an equilibrium is reached.

The index coefficients for the sire and dam at the time of selection of the candidate offspring, and the candidate itself are given by $\boldsymbol{b}_x = \boldsymbol{V}^{-1}\boldsymbol{g}_x$, where $\boldsymbol{g}_x$ is the covariance between the appropriate breeding value and the information sources, thus, for the sire:

$$\boldsymbol{g}_1 = (v_1,\ 0,\ 0,\ \tfrac{1}{2}\sigma^2_{a(1)},\ \tfrac{1}{2}\sigma^2_{a(1)},\ \tfrac{1}{2}\sigma^2_{a(1)}),$$

for the dam:

$$\boldsymbol{g}_2 = (0,\ v_2,\ d^{-1}v_2,\ d^{-1}\tfrac{1}{2}\sigma^2_{a(2)},\ \tfrac{1}{2}\sigma^2_{a(2)},\ \tfrac{1}{2}\sigma^2_{a(2)}),$$

and for the candidate:

$$\boldsymbol{g}_3 = (\tfrac{1}{2}v_1,\ \tfrac{1}{2}v_2,\ \tfrac{1}{2}d^{-1}v_2,\ \sigma^2_{HS},\ \sigma^2_{FS},\ \sigma^2_P),$$

where σ_A^2 and σ_I^2 are the equilibrium genetic variance and index variance respectively; $\sigma_{a(q)}^2$ is the additive genetic variances among sires (q=1) and dams (q=2) respectively; σ_{HS}^2 and σ_{FS}^2 are the phenotypic variance of the half-sib and full-sib family means respectively; $v_q = (1-k_q)\sigma_I^2$, is the variance of the EBVs following selection among sires (q=1) and dams (q=2) where k_q is the variance reduction coefficient. The indices with coefficients $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ for the sire, dam and candidate at the time of selection of the candidate have

a 3 x 3 (co)variance matrix denoted **W** where $w_{ij} = \boldsymbol{g}_i^T \mathbf{V}^{-1} \boldsymbol{g}_j$, note $w_{33} = \sigma_I^2$.

λ_{pq} *assuming random mating*: Following Appendix A of WOOLLIAMS *et al.* (1999), the regressions are: for $\hat{A}_{i(q)}$, $\frac{1}{2}\iota_p\sigma_I^{-1}$; for $\delta\hat{A}_{i(q)}$, $\frac{1}{2}\iota_p\sigma_I^{-1}(w_{q3}-\frac{1}{2}v_q)/(w_{qq}-v_q)$; and 0 for $\hat{e}_{i(q)}$.

π_{pq} *assuming random mating*: The terms conferring selective advantage to a parent is its estimated breeding value at the time of its selection ($\hat{A}_{i(q)}$; the 'initial' EBV), its increment at the selection of its offspring ($\delta\hat{A}_{i(q)}$; the 'increment'), and remaining prediction error ($\hat{e}_{i(q)}$; the 'prediction error'). These terms are mutually independent. They are defined as deviations from their selected contemporaries within category q. For discrete generations there are two categories, sires and dams. The regressions of these terms for a selected $j(p)$ of category p on its parent $i(q)$ of category q are required. Note that: (i) the prediction error of the parent is independent of the initial EBV of the progeny (otherwise the prediction of the EBV of the parent could be improved); and (ii) the increment and the prediction error of the progeny must be independent of the information on the parent at the time of progeny selection (i.e. the parent's initial EBV plus its increment), or otherwise the EBV of the progeny could be improved. Therefore it is easiest to consider these regressions in two parts: the regression of $\hat{A}_{j(p)}$ on $\hat{A}_{i(q)}$ and $\delta\hat{A}_{i(q)}$, and the regressions of $\delta\hat{A}_{j(p)}$ and $\hat{e}_{j(p)}$ on $\hat{e}_{i(q)}$. Remaining coefficients will be zero.

The three terms $\hat{A}_{i(q)}$, $\delta\hat{A}_{i(q)}$, and $\hat{A}_{j(p)}$, have a (co)variance matrix before selection given by (v_q, 0, $\frac{1}{2}v_q$ | 0, $w_{qq} - v_q$, $w_{q3} - \frac{1}{2}v_q$ | $\frac{1}{2}v_q$, $w_{q3} - \frac{1}{2}v_q$, w_{33}), where q=1 if i is a sire and 2 if i is a dam. Appendix B of WOOLLIAMS *et al.* (1999) can be followed to obtain:

$$\hat{A}_{j(p)} = \tfrac{1}{2}(1-k_p)\hat{A}_{i(q)} + (w_{q3}-\tfrac{1}{2}v_q)(w_{qq}-v_q)^{-1}(1-k_p)\delta\hat{A}_{i(q)} + \varepsilon$$

Note both regression coefficients have $(1-k_p)$ in them as factors, which makes these coefficients small whatever the value of h^2.

The regression coefficients for $\hat{e}_{i(q)}$ are less immediate. However they can be obtained from the following argument. Let $\sigma_{A(q)*}^2 = \sigma_A^2 - k_q\sigma_I^2$ be the genetic variation among parents of category q. By noting the breeding value is equal to $\hat{A} + \delta\hat{A} + \hat{e}$, before selection of the candidates:

$$\tfrac{1}{2}\sigma_{A(q)*}^2 =$$

$$cov[(\hat{A}_{i(q)}+\delta\hat{A}_{i(q)})+\hat{e}_{i(q)}, (\hat{A}_{j(p)}+\delta\hat{A}_{j(p)})+\hat{e}_{j(p)} \mid i(q)\ selected]$$

$$= cov[(\hat{A}_{i(q)}+\delta\hat{A}_{i(q)}), \hat{A}_{j(p)} \mid i(q) selected] + cov[\hat{e}_{i(q)}\ ,\ (\delta\hat{A}_{j(p)}+\hat{e}_{j(p)})]$$

The first of these terms is w_{q3}, so $cov(\hat{e}_{i(q)}\ ,\ (\delta\hat{A}_{j(p)}+\hat{e}_{j(p)})) = \frac{1}{2}\sigma_{A(q)*}^2 - w_{q3}$. Denote this covariance by $v_{e(q)}$. Selection of the offspring $j(p)$ will not effect $v_{e(q)}$ since all components are independent of the initial EBV for $j(p)$, which is the index used for the selection. Next consider the index for the offspring of $j(p)$, i.e. the grandoffspring of $i(q)$: the covariance of the prediction error of $j(p)$ at the time of its selection (i.e. $\delta\hat{A}_{j(p)} + \hat{e}_{j(p)}$) with its candidate offspring included in its index is ½ of the prediction error variance, since it is a component of the genetic variation passed onto the offspring. The covariance of $\hat{e}_{i(q)}$ with these candidate offspring will be $\frac{1}{2}v_{e(q)}$. Therefore the covariances of the prediction error of $i(q)$ with the information from the offspring of $j(p)$ are given by $\boldsymbol{g}_{e(p)}$ = (0, 0, 0, $\frac{1}{2}v_{e(q)}$, $\frac{1}{2}v_{e(q)}$, $\frac{1}{2}v_{e(q)}$) if j is male, and (0, 0, 0, $\frac{1}{2}v_{e(q)}d^{-1}$, $\frac{1}{2}v_{e(q)}$, $\frac{1}{2}v_{e(q)}$) if j is female. Then, since $\boldsymbol{b}_p$ is the updated index of $j(p)$, $\mathrm{cov}(\hat{e}_{i(q)},\delta\hat{A}_{j(p)})$ will be $\boldsymbol{g}_{e(p)}^T\boldsymbol{b}_p$, and by difference $\mathrm{cov}(\hat{e}_{i(q)},\hat{e}_{j(p)}) = v_{e(q)} - \boldsymbol{g}_{e(p)}^T\boldsymbol{b}_p$. Therefore, since $\mathrm{Var}(\hat{e}_{i(q)}) = \sigma_{A(q)*}^2 - w_{qq}$:

$$\delta\hat{A}_{j(p)} = \boldsymbol{g}_{e(p)}^T\boldsymbol{b}_p(\sigma_{A(q)*}^2 - w_{qq})^{-1}\ \hat{e}_{i(q)}$$

$$\hat{e}_{j(p)} = (\tfrac{1}{2}\sigma_{A(q)*}^2 - w_{q3} - \boldsymbol{g}_{e(p)}^T\boldsymbol{b}_p)(\sigma_{A(q)*}^2 - w_{qq})^{-1}\ \hat{e}_{i(q)}$$

Covariances between selective advantages and Mendelian terms: Define $\boldsymbol{g}_a$ = (0, 0, 0, $\frac{1}{2}h_0^2X_m/T$, $\frac{1}{2}h_0^2X_f/T$, $\frac{1}{2}h_0^2$) which is the vector of covariances of the candidate's Mendelian sampling term with the information sources for its index. Before selection, the covariance of the index of $i(q)$ with $a_{i(q)}$ is $\boldsymbol{g}_a^T\boldsymbol{b}_3$. The mean of the Mendelian sampling terms for selected ancestors is $E[a_{i(q)}] = \iota_q\boldsymbol{g}_a^T\boldsymbol{b}_3/\sigma_I^2$, and the covariance for selected ancestors $\mathrm{cov}(a_{i(q)}, \hat{A}_{i(q)})$ is $(1-k_q)\boldsymbol{g}_a^T\boldsymbol{b}_3$. Therefore $cov(a_{i(q)}, \delta\hat{A}_{i(q)}+\hat{e}_{i(q)}) = v_{a(q)} = \frac{1}{2}h_0^2 -$

$\boldsymbol{g}_a^T\boldsymbol{b}_3$. To determine the separate covariances for $\delta\hat{A}_{i(q)}$ and $\hat{e}_{i(q)}$, consider the covariances of the prediction error of the Mendelian sampling term for $i(q)$ with the sources of index information for an offspring, $\boldsymbol{g}_a^*$: if $i(q)$ is male, $\boldsymbol{g}_a^* = (\ 0,\ 0,\ 0,\ \tfrac{1}{2}v_{a(q)},\ \tfrac{1}{2}v_{a(q)},\ \tfrac{1}{2}v_{a(q)})$; and if $i(q)$ is female, $\boldsymbol{g}_a^* = (0,\ 0,\ 0,\ \tfrac{1}{2}v_{a(q)}d^{-1},\ \tfrac{1}{2}v_{a(q)},\ \tfrac{1}{2}v_{a(q)})$. Then, since $\boldsymbol{b}_q$ is the index for the updated index of $i(q)$, $\text{cov}(a_{i(q)},\ \delta\hat{A}_{i(q)}) = \mathbf{g}_a^{*T}\boldsymbol{b}_q$, and, by difference, $\text{cov}(a_{i(q)},\ \hat{e}_{i(q)}) = v_{a(q)} - \mathbf{g}_a^{*T}\boldsymbol{b}_q$.

Example: (X_m=20, X_f=40, h^2=0.4, 8 offspring per litter). The matrices are presented in the following row and column order: for dimension 6 the order is ($\hat{A}_{i(m)}$, $\delta\hat{A}_{i(m)}$, $\hat{e}_{i(m)}$, $\hat{A}_{i(f)}$, $\delta\hat{A}_{i(f)}$, $\hat{e}_{i(f)}$), whilst for dimension 2 the order is males then females. $\mathbf{G}\circledast\mathbf{A}$ = (1.16, 2.07, 0, 1.16, 2.71, 0 | 0.90, 1.60, 0, 0.90, 2.09, 0), $\mathbf{G}\circledast\mathbf{\Pi}$ = (0.046, 0.082, 0, 0.046, 0.107, 0 |0, 0, 0.049, 0, 0, 0.053 | 0, 0, 0.081, 0, 0, 0.087 |0.060, 0.108, 0, 0.060, 0.141, 0 | 0, 0, 0.032, 0, 0, 0.035 |0, 0, 0.098, 0, 0, 0.106). E[$a_{i(q)}$]=(0.270,0.208), $cov(a_{i(q)},\ \hat{A}_{i(q)})$ = (0.011, 0.014), $cov(a_{i(q)},\ \delta\hat{A}_{i(q)})$ = (0.054, 0.035), $cov(a_{i(q)},\ \hat{e}_{i(q)})$ = (0.088,0.107).

Prediction of Genetic Contributions and Generation Intervals in Populations with Overlapping Generations under selection

Piter Bijma* **and John A. Woolliams†**

* Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands and †Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K.

Abstract – Long-term genetic contributions of ancestors to future generations were predicted, for a population with overlapping generations undergoing mass or sib index selection, following a recently developed theory. This theory provides insight into the mechanisms determining the flow of genes through selected populations, and takes account of selection by modelling the long-term genetic contribution as a linear regression on the breeding value. Total genetic contributions of age classes are modelled using a modified gene flow approach and long-term predictions are obtained assuming equilibrium genetic parameters. Generation interval was defined as the time in which genetic contributions sum to unity, which is equal to the turn over time of genes. Accurate predictions of long-term genetic contributions of individual animals, as well as total contributions of age classes were obtained. Due to selection, offspring of young parents had an above average breeding value. Long-term genetic contributions of youngest age classes were therefore higher than expected from the age class distribution of parents, and generation interval was shorter than the average age of parents at birth of their offspring. Due to an increased selective advantage of offspring of young parents, generation interval decreased with increasing heritability and selection intensity. The method was compared to conventional gene flow theory and showed more accurate predictions of long-term genetic contributions.

Most natural and artificial populations have overlapping generations. When generations overlap, the generation interval differs from the cohort interval. In quantitative genetics, generation intervals are generally defined as the average age of parents at birth of their offspring. In this definition, generation interval is based on the contributions of parental age classes to newborn offspring; *i.e.*, the average age of parents is calculated as the sum of ages at birth of offspring weighted by the contribution of each age class to newborn offspring. This approach is adopted in the well-known gene flow procedure (HILL, 1974). However, if selective advantage (*e.g.*, breeding value) is partly inherited, selection in subsequent generations may affect the genetic contribution of parental age classes to future generations. Thus there may be a difference between generation interval based on contributions to newborn offspring, and generation interval based on contributions to future generations. It has been suggested, therefore, to calculate generation intervals

based on selected offspring only (BICHARD *et al.* 1973). However, contributions of ancestors to future generations may still deviate from contributions to selected offspring.

Recently, WOOLLIAMS *et al.* (1999) found significant differences between generation interval calculated as the average age of parents at the time of birth of a cohort of offspring and generation interval based on the concept of long-term genetic contributions. The latter concept was first introduced by JAMES and MCBRIDE (1958) and developed further for the prediction of inbreeding by WRAY and THOMPSON (1990) and WOOLLIAMS *et al.* (1993). Predictions for more advanced selection systems, however, resulted in complicated expressions (WRAY *et al.* 1994) due to the recursive nature of the prediction procedure. Working on the infinitesimal model (FISHER, 1918), WOOLLIAMS *et al.* (1999) obtained a simple closed-form approximation for the prediction of long-term genetic contributions by considering BULMER's (1971) equilibrium genetic parameters, which makes a recursive algorithm redundant. The method of WOOLLIAMS *et al.* (1999) covers both discrete and overlapping generations and is applicable to mass selection, index selection and BLUP selection.

The aim of the current article is twofold. First, long-term genetic contributions will be predicted for populations with overlapping generations and mass selection, following the approach of WOOLLIAMS *et al.* (1999). This approach illustrates the mechanisms that determine the development of pedigree, the contribution of different categories to the genetic makeup of the population in the long-term and the turnover time of genes. The dependency of long-term genetic contributions and generation intervals on selective advantage will be illustrated in populations with overlapping generations under mass or sib index selection, assuming the infinitesimal model (Fisher, 1918).

Second, predictions following the approach of WOOLLIAMS *et al.* (1999) will be compared to predictions of long-term genetic contributions and generation intervals based on contributions to unselected newborn offspring, as obtained from conventional gene flow theory (HILL, 1974). Both methods will be compared to results obtained from simulated data.

Accurate predictions of long-term genetic contributions are an important step towards the prediction of rates of inbreeding in selected populations (WOOLLIAMS, 1998). The current article focuses on the prediction of genetic contributions and generation intervals, the prediction of rates of inbreeding is in a subsequent article. To show the power of theory of WOOLLIAMS *et al.* (1999), predictions of genetic gain based on long-term genetic contributions will also be presented, but this is not the main item, because accurate predictions of genetic gain are already well established (*e.g.*, VILLANUEVA *et al.* 1993).

Methods

Here we will first describe the population structure that was used. Subsequently we will describe the concept of long-term genetic contributions and the method of WOOLLIAMS *et al.* (1999) for the prediction of long-term

genetic contributions in populations with overlapping generations, followed by a description of the relationship between generation interval and genetic contributions. Finally, we will derive iterative deterministic and stochastic equations to implement the theory.

Population model: This section describes the genetic model, population structure and selection strategy for which predictions of genetic contributions were made. The trait considered was assumed to be determined by the infinitesimal model (FISHER, 1918). Phenotypic values (P) were the sum of additive genetic values (A, breeding values) and environmental values (E), *i.e.*, $P = A + E$. The population consisted of overlapping generations and selection was based upon a sib index for a single trait. With parents up to maximum of c_{max} of age there are $2c_{max}$ categories, one for each sex and age of parent. Categories are indexed by k or by l, so $k = 1$ .. c_{max} are males, and $k = c_{max}+1$ .. $2c_{max}$ are females. Let $age(k)$ denote the age of category k (so $age(1) = 1 = age(c_{max}+1)$) and let n_k be the number of parents selected from category k. The total number of male and female parents equaled: $N_m = \sum_{k=1}^{c_{max}} n_k$ and $N_f = \sum_{k=c_{max}+1}^{2c_{max}} n_k$ respectively. Using random mating, each sire was mated to d dams ($d = N_f/N_m$), and each dam produced a total of n_o offspring ($\frac{1}{2}n_o$ of each sex) so that the total number of offspring in a cohort equaled $T = n_o N_f$. Before reproductive age, the phenotype of individuals was recorded and a selection index was calculated. Because index weights were constant over time and no additional phenotypic information was included in the index at later ages, the index of individuals remained constant over time and the ranking of animals within categories remained unchanged over time. Within categories, individuals were ranked on the index, and the highest ranking n_k individuals were selected. The number of parents selected from each category was determined in advance and remained constant over time, as in conventional gene flow (HILL, 1974). Selection on estimated breeding value across categories, which gives the highest genetic level of the offspring in the next generation (JAMES, 1987) was not applied. The selection index was:

$$I = b_1(P - \overline{P}_{FS}) + b_2(\overline{P}_{FS} - \overline{P}_{HS}) + b_3\overline{P}_{HS},$$

where P is the phenotype of the individual, $\overline{P}_{FS}$ is the mean of n_o full-sib records (including the individual) and $\overline{P}_{HS}$ is the mean of $n_o d$ half-sib records (including individual and its full-sibs). This form was used by WRAY *et al.* (1994) and is convenient because the three sources of information are independent, which simplifies expressions such as the accuracy of selection. Note that mass selection is a special case of this index, where $b_1 = b_2 = b_3$. Different sets of index weights were chosen to allow for different selection strategies, *i.e.*, for a varying emphasis on family information.

Basic approach for prediction of long-term genetic contributions: This section introduces the concept of long-term genetic contributions. The long-term genetic contribution (r_i) of ancestor i in cohort t_1 is defined as the

proportion of genes present in all individuals in cohort t_2 deriving by descent from i, where $(t_2-t_1) \rightarrow \infty$ (WOOLLIAMS *et al.* 1993). In other words, the long-term genetic contribution of an ancestor is the ultimate proportional contribution of the ancestor to generations in the distant future. After several generations, genetic contributions of ancestors stabilise (long-term contributions are reached) and become equal for all individuals in that and subsequent generations of descendants, but values will differ between ancestors (WRAY and THOMPSON, 1990).

In the remainder of the current article, long-term genetic contributions of ancestors will be referred to as "genetic contributions", unless explicitly stated otherwise. Following the approach of WOOLLIAMS *et al.* (1999), contributions of ancestors are predicted by conditioning on the selective advantage of those ancestors. Since sib indices are used here, the selective advantage is equal to the true breeding value of the ancestor [the only parental effect affecting selection of the offspring is the breeding value of the parent (WRAY *et al.* 1994)]. For an individual in category l, $E(r_{i(l)}|A_{i(l)}) \approx u_{i(l)} = \alpha_l + \beta_l (A_{i(l)} - \overline{A}_l)$, where α_l is the expected contribution of an average parent in category l, β_l is the regression of the contribution of i on its breeding value ($A_{i(l)}$) and $\overline{A}_l$ is the mean breeding value of selected contemporaries of i in category l. For discrete generations, the complication of categories can be ignored and α is obtained directly from the number of parents: $\alpha = (2N_x)^{-1}$, ($x = m$, f; WRAY and THOMPSON 1990). For both discrete and overlapping generations, solutions for β can be obtained from two regression models (WOOLLIAMS, 1998; WOOLLIAMS *et al.* 1999). First, the regression of the number of selected offspring on the breeding value of the parent (λ), and second, the regression of the breeding value of *selected* offspring on the breeding value of the parent (π). Both λ and π can be computed using known parameters; a derivation is in APPENDIX A. With equilibrium genetic parameters (BULMER, 1971), regression coefficients (α, β, λ, π) are equal for the parental and offspring generation, allowing for the following closed form expression to compute β, instead of a recursive algorithm (WOOLLIAMS, 1998):

$$\beta = (1-\pi)^{-1}\lambda\alpha$$

Prediction of expected long-term genetic contributions in populations with overlapping generations: This section describes how the theory of WOOLLIAMS *et al.* (1999) can be implemented to predict long-term genetic contributions for populations with overlapping generations. For ancestor i in category l, the expected long-term genetic contribution was predicted from $u_{i(l)} = \alpha_l + \beta_l(A_{i(l)} - \overline{A}_l)$. Predictions of genetic contributions are obtained using a modified gene flow matrix (**G**) of dimension $2c_{max} \times 2c_{max}$, which identifies the origin of genes of *selected* instead of newborn offspring. If the conventional gene flow matrix (HILL, 1974) is denoted by $\mathbf{G}_0$, elements g^0_{kl} represent the proportion of genes currently in category k that were in category l one time unit ago. In the

modified gene flow matrix, elements g_{kl} of **G** represent the proportion of genes in the n_k *selected* individuals in category k, that were contributed by parents in category l. (Contributed by a parent in category l refers to contribution via offspring that were *born* when the parent was in category l.) Because **G** represents the parental origin of the genes of *selected* individuals, it is affected by the degree of selection that is taking place, and this may vary with age. Because selected individuals may be born up to c_{max} years ago (and the age of parents *at birth* of offspring is relevant), **G** has a memory of c_{max} years, whereas $\mathbf{G}_0$ has only one year memory.

Solutions for α and β were obtained from (WOOLLIAMS, *et al.* 1999),

$$\alpha_l = \sum_{k=1}^{2c_{max}} n_l^{-1} g_{kl} n_k \alpha_k \qquad (1)$$

$$\beta_l (A_{i(l)} - \overline{A}_l) = \sum_{k=1}^{2c_{max}} n_l^{-1} g_{kl} n_k \alpha_k \lambda_{kl} (A_{i(l)} - \overline{A}_l) + \sum_{k=1}^{2c_{max}} n_l^{-1} g_{kl} n_k \beta_k \pi_{kl} (A_{i(l)} - \overline{A}_l), \qquad (2)$$

where λ_{kl} is the regression coefficient of the selected number of offspring in category k on the breeding value of the parent in category l and π_{kl} is the regression of the breeding value of *selected* offspring in category k on the breeding value of the parent in category l. An intuitive understanding of Equation (1) and (2) can be gained by noticing that $2n_l^{-1}g_{kl}n_k$ represents the average number of selected offspring in category k of an ancestor in category l. Therefore, in (1), α_l is equal to ½ times the sum of the average contributions of all selected offspring. (The other ½ originates from the other parent.) In (1) it is implicitly assumed that the contribution of an average selected offspring (α_k) is independent of the category of the parent (l). In (2), the first summation represents the change of contributions due to deviations of the selected number of offspring from the average. The second summation represents changes of genetic contributions of ancestors due to deviations of the breeding value of selected offspring of this ancestor from the average breeding value of their selected contemporaries. In matrix form, combining Equations (1) and (2) for all categories l (WOOLLIAMS *et al.* 1999),

$$\mathbf{N}\alpha = \mathbf{G}^{\mathrm{T}}\mathbf{N}\alpha \qquad (3)$$

$$\mathbf{N}\beta = (\mathbf{I} - \mathbf{G}^{\mathrm{T}} * \mathbf{\Pi}^{\mathrm{T}})^{-1}(\mathbf{G}^{\mathrm{T}} * \mathbf{\Lambda}^{\mathrm{T}})(\mathbf{N}\alpha), \qquad (4)$$

where * denotes the element by element multiplication, T denotes the transpose of matrices, **I** is a $2c_{max}\times 2c_{max}$ identity matrix, **N** is a $2c_{max}\times 2c_{max}$ diagonal matrix of elements n_k, $\mathbf{\Pi}$ is a $2c_{max}\times 2c_{max}$ matrix of elements π_{kl}, $\mathbf{\Lambda}$ is a $2c_{max}\times 2c_{max}$ matrix of elements λ_{kl}, α is a $2c_{max}$ vector of elements α_l, and β is a $2c_{max}$ vector of elements β_l. Throughout the article, matrices follow the gene flow notation, *i.e.*, rows represent offspring categories and columns represent parent categories. Prediction of genetic contributions using Equations 3 and 4 is referred to as *Method M* in RESULTS.

Improved modified gene flow: A first-order correction to Equation 1 was derived by taking account of differences among average breeding values of parental subgroups present in the

selected offspring (WOOLLIAMS *et al.* 1999). When newborn offspring are grouped according to the category of their parents, mean breeding values may differ between those groups. Selection then favours offspring descending from parental categories with a higher breeding value, increasing the genetic contribution of these categories. This phenomenon is fully accounted for by the modified gene flow matrix **G**, which identifies the origin of selected offspring. However, after selection, mean breeding values of selected offspring may still differ between parental category subgroups. This affects the contribution of categories, which was ignored in Equation 1. Improved prediction equations were obtained by conditioning on the parental category in Equation 1 (WOOLLIAMS *et al.* 1999),

$$\alpha_l = \sum_k n_l^{-1} g_{kl} n_k \{\alpha_k + \beta_k \mathrm{E}[(A_{(i)k} - \bar{A}_k) \text{ given } i \text{ has category } l \text{ parent}]\}, \quad (5)$$

where $\mathrm{E}[(A_{(i)k} - \bar{A}_k)$ given i has category l parent] is the expected breeding value of a selected offspring in category k descending from a category l parent, as deviation from the mean of selected contemporaries in category k. Substituting Equation 2 for β, the resulting expression is (WOOLLIAMS *et al.* 1999),

$$\mathbf{N}\alpha = [\mathbf{G}^T + (\mathbf{G}^T * \mathbf{D}^T)(\mathbf{I} - \mathbf{G}^T * \mathbf{\Pi}^T)^{-1}(\mathbf{G}^T * \mathbf{\Lambda}^T)]\mathbf{N}\alpha, \quad (6)$$

where **D** is a $2c_{max} \times 2c_{max}$ matrix of elements $d_{kl} = \mathrm{E}[(A_{(i)k} - \bar{A}_k)$ given i has category l parent$]$. Therefore, $\mathbf{N}\alpha$ is obtained as a right eigenvector of the $2c_{max} \times 2c_{max}$ matrix $[\mathbf{G}^T + (\mathbf{G}^T * \mathbf{D}^T)(\mathbf{I} - \mathbf{G}^T * \mathbf{\Pi}^T)^{-1} (\mathbf{G}^T * \mathbf{\Lambda}^T)]$ with an eigenvalue of one (WOOLLIAMS *et al.* 1999). Solutions for β are still obtained from (4). Predictions of genetic contributions using Equation 4 and 6 will be referred to as *Method P* in RESULTS.

Generation interval: Generation interval (L) is defined as the turnover time of genes, *i.e.*, the average time interval between two meioses that an average gene in the population is involved in. This interval is equal to the time in which long-term genetic contributions sum to unity, *i.e.*, the genetic contribution summed over all ancestors entering the population over a time period of L years equals unity: $\sum_L u_i = 1$. The generation interval (in years) is therefore equal to the reciprocal of the total long-term genetic contribution per year, *i.e.*, summed over all ancestors per year. In $u_{i(l)} = \alpha_l + \beta_l (A_{i(l)} - \bar{A}_l)$, and the term $\beta(A_{i(l)} - \bar{A}_l)$ is zero on average, so that the sum of genetic contributions equals $\sum_{k=1}^{2c_{max}} n_k \alpha_k$, and generation interval is (WOOLLIAMS *et al.* 1999),

$$L = 1 \Big/ \sum_{k=1}^{2c_{max}} n_k \alpha_k \quad (7)$$

Generation intervals from this definition were compared to generation intervals defined as the average age of parents at birth of their offspring.

Deterministic prediction procedure: Elements of Equation 3 through 7 were obtained using an iterative procedure, which is described in this section. The iterative procedure is needed because elements (*e.g.*, variances, genetic gain and genetic

contributions) are mutually dependent and BULMER's (1971) equilibrium parameters can only be reached by iteration. [Predictions can also be obtained using base generation parameters, but more accurate predictions are obtained using equilibrium parameters (WOOLLIAMS *et al.* 1999)]. Predictions of genetic contributions shown in the RESULTS are based on BULMER's (1971) equilibrium parameters. A numerical example is in APPENDIX C.

Phenotypic variance in year t was the sum of additive genetic variance and environmental variance, $\sigma^2_{P,t} = \sigma^2_{A,t} + \sigma^2_E$. Environmental variance was constant over time. Additive genetic variance in an unselected cohort born at year t was calculated as:

$$\sigma^2_{A,t} = \sigma^2_{A(m),t} + \sigma^2_{A(f),t} + \tfrac{1}{2}\sigma^2_{A0},$$

where $\sigma^2_{A(m),t}$ and $\sigma^2_{A(f),t}$ are the between sire and between dam-family additive genetic variance in unselected newborn offspring, and σ^2_{A0} is the base generation additive genetic variance. Because genetic contributions are mainly determined in the first few generations, they are hardly affected by the rate of inbreeding. Therefore, no effect of inbreeding on the within family variance was modeled.

Between sire-family additive genetic variance was calculated from

$$\sigma^2_{A(m),t} = \tfrac{1}{4}\sum_{l=1}^{c_{max}} 2g^0_{lt}[\sigma^2_{A,t-1}(1-\kappa_l\rho^2_{t-1}) + (\mu_{l,t-1} - \mu_{(m),t-1})^2]$$

where $2g^0_{lt}$ is the proportion of offspring descending from sires in category l ($2g^0_{lt} = n_l/N_m$), κ_l is PEARSON's (1903) variance reduction coefficient, ρ_t is the accuracy of selection in year t (WRAY *et al.* 1994), $\mu_{l,t}$ is the average breeding value of selected sires in category l and $\mu_{(m),t}$ is the average breeding value of all selected sires, *i.e.*, $\mu_{(m),t} = \sum_{l=1}^{c_{max}} 2g^0_{lt}\mu_{l,t}$. Between dam-family additive genetic variance was calculated in the same way. For the calculation of $(\mu_{l,t-1} - \mu_{(m),t-1})^2$, only differences between breeding values of selected individuals are important and breeding values can be expressed relative to an arbitrary base. The genetic level of unselected animals at birth was taken as base here, and therefore, $\mu_{l,t} = (i_l\rho_t\sigma_{A,t} - age(l)\Delta G_t)$, where i_l is the selection intensity in category l (not distinguishing between subgroups within categories and ignoring deviations from normality) and ΔG_t is the rate of genetic gain in year t. (It is assumed here that the difference between consecutive age classes is equal to ΔG from the last iteration, because this assumption decreases the number of iterations needed to reach equilibrium values which are not affected by the assumption).

To calculate elements of the modified gene flow matrix, we need to find how the predefined selected proportion of individuals in category k (p_k) is distributed across the parental age subgroups. The k^{th} row of **G** therefore, was obtained by finding a common index truncation point for all parental subgroups represented among the selection candidates in category k (separate for male and female parents). The solution for the common truncation point has to

satisfy the equations (omitting subscript t for simplicity),

$$p_k = \sum_l 2 g_{ll}^0 p_{kl}$$

$$p_{kl} = 1 - \Phi\left(\frac{I_k - \frac{1}{2}\tau_{(x)}\mu_l}{\sigma_{I,l}}\right),$$

where p_{kl} is the selected proportion in the subclass descending from parents in category l, I_k is the index truncation point common for all offspring in category k, $\sigma_{I,l}$ is the standard deviation of the selection index of individuals descending from parents in category l, Φ denotes the cumulative normal density, and $\tau_{(x)}$ is twice the regression of the index of the offspring on the breeding value of the parent of sex x ($x = m, f$) (WRAY *et al.* 1994), *i.e.*, the term $\frac{1}{2}\tau_{(x)}\mu_{(x)l}$ represents the average index value of offspring descending from parents in category l. A solution for the common truncation point was obtained numerically using the algorithm RIDDR_ROOT from NUMERICAL RECIPES (PRESS *et al.* 1992). Next, elements of **G** were calculated from

$$g_{kl} = p_{kl} g_{ll}^0 p_k^{-1}.$$

Elements of **D** are: d_{kl} = E[$(A_{(i)k} - \overline{A}_k)$ given i has category l parent] and were calculated as (omitting subscript t for simplicity)

$$d_{kl} = \tfrac{1}{2}\mu_l + i_{kl}\rho\sigma_{A_l} - \overline{d}_k,$$

with $\overline{d}_k = \sum_l 2 g_{kl}[\frac{1}{2}\mu_l + i_{kl}\rho\sigma_{A_l}]$ calculated separately for each sex; where i_{kl} is the selection intensity in subclass kl, and σ_{A_l} is the additive genetic variance among unselected offspring descending from parents in category l.

Elements of Π were calculated as $\pi_{kl} = \frac{1}{2}(1 - \kappa_k\tau_{(x)}\rho\sigma_A\sigma_I^{-1})$. Elements of Λ were calculated as $\lambda_{kl} = \frac{1}{2} i_k \tau_{(x)} \sigma_I^{-1}$ (see APPENDIX A). A general procedure to derive Π and Λ is in WOOLLIAMS *et al.* (1999).

As described in the section on prediction of long-term genetic contributions, α can be obtained as a right eigenvector from Equation 3 for the 'modified gene flow" and from Equation 6 for the "improved modified gene flow". In general, eigenvectors can be scaled, *i.e.*, if **x** is an eigenvector of matrix **A** with an eigenvalue γ, then n**x** will also be an eigenvector of **A** with the same eigenvalue γ. With the same eigenvalue, therefore, different eigenvectors can be obtained from Equation 3 or 6, and an additional constraint has to be imposed. Because contributions have to sum to unity per generation, the eigenvector was scaled accordingly. Therefore, first generation interval was calculated as the average age at birth of offspring weighted by the long-term genetic contribution of the categories ($n_k\alpha_k$): $L = \sum_{k=1}^{2c_{max}} age(k)\, n_k\alpha_k \Big/ \sum_{k=1}^{2c_{max}} n_k\alpha_k$. Subsequently, α was scaled so that $\sum_{k=1}^{2c_{max}} n_k\alpha_k = L^{-1}$, *i.e.*, α is defined per year and by definition, the generation interval is the time in which contributions sum to unity.

Using $E(\Delta G) = \sum_{k=1}^{2c_{max}} n_k E[r_{i(k)} a_{i(k)}]$, where $a_{i(k)}$ is the Mendelian sampling value of i (WOOLLIAMS *et al.* 1998), genetic gain was predicted from

$$E(\Delta G) = \tfrac{1}{2}\sigma^2_{A_0}[\tau_w \sigma_I^{-1} \sum_{k=1}^{2c_{max}} n_k \alpha_k i_k + \sum_{k=1}^{2c_{max}} n_k \beta_k (1 - \kappa_k \tau_w \rho \sigma_A \sigma_I^{-1})],$$

where τ_w is the regression of the index on the Mendelian sampling effect of the individual. A derivation is in APPENDIX B.

Stochastic Simulation: To draw inferences on the accuracy of predicted genetic contributions, the breeding scheme described in the *population model* section was simulated stochastically and genetic contributions were estimated from simulated data. A noninbred and unselected base population of the appropriate family structure was generated. Breeding values of base population animals were taken from $N(0,\sigma^2_{A0})$, and environmental values from $N(0,\sigma^2_E)$. Within categories, individuals were ranked on the index, and the highest ranking n_k individuals were selected from the k^{th} category. Breeding values of offspring were obtained as $\frac{1}{2}A_m + \frac{1}{2}A_f + a$, where A_m, A_f and a are the sire and dam breeding values and the Mendelian sampling value. No effect of inbreeding on the Mendelian sampling variance was simulated, *i.e.*, $a \sim N(0,\frac{1}{2}\sigma^2_{A0})$.

For the calculation of genetic contributions, an ancestor cohort t_1 was chosen when BULMER's (1971) equilibrium genetic parameters were reached. Repeated cycles of selection and random mating were performed until genetic contributions were converged and then a descendent cohort t_2 was chosen. Convergence time of genetic contributions ($t_2 - t_1$) was approximately equal to $7c_{max}$. The long-term genetic contribution of ancestor i in category l in cohort t_1, to individuals in cohort t_2 was obtained by summing contributions via all pedigree paths leading from i to individuals in t_2, $r_{i(l)} = T^{-1} \sum_{j=1,T} r_{i(l),j}$, where $r_{i(l),j}$ is the contribution to individual j in cohort t_2, which was calculated as $r_{i(l),j} = \sum_{paths} \frac{1}{2}^{\breve{n}-1}$, where $\breve{n}$ is the total number of animals (including i and j) in a pedigree path from i to j.

Genetic contributions were analyzed using the model: $r_{i(l)} = \alpha_l + \beta_l(A_{i(l)} - \overline{A}_l) + e_{i(l)}$, α was estimated as: $\hat{\alpha}_l = n_l^{-1} \sum_{i=1,n_l} r_{i(l)}$, and β as: $\hat{\beta}_l = \sum_{i=1,n_l} r_{i(l)}(A_{i(l)} - \overline{A}_l) \Big/ \sum_{i=1,n_l} (A_{i(l)} - \overline{A}_l)^2$.

Asymptotic rate of genetic gain was calculated as $\Delta G = (G_{t2} - G_{t1})/(t_2 - t_1)$, where G_t is the average breeding value of all animals born in cohort t. Generation interval was calculated as $L = 1 \Big/ \sum_{k=1}^{2c_{max}} n_k \hat{\alpha}_k$. Results were averaged over 500 replicates and standard errors were calculated from the variance among replicates.

Results

In this section a comparison will be made between results from conventional gene flow theory (*Method C*; HILL, 1974), simple modified gene flow (*Method M*, Equations 3

TABLE 1.–Genetic contributions of categories ($n_k\alpha_k$) with mass selection

	Method C[a]			Method M[a]			Method P[a]			simulation[b]		
$h_0^{2\,c}$	$n_1\alpha_1$	$n_4\alpha_4$	$n_6\alpha_6$	$n_1\alpha_1$	$n_4\alpha_4$	$n_6\alpha_6$	$n_1\alpha_1$	$n_4\alpha_4$	$n_6\alpha_6$	$n_1\alpha_1$	$n_4\alpha_4$	$n_6\alpha_6$
0.01	0.2857	0.0714	0.2143	0.2863	0.0726	0.2137	0.2869	0.0738	0.2131	0.2871	0.0758	0.2135
	-0.0014	-0.0044	+0.0008	-0.0008	-0.0032	+0.0002	-0.0002	-0.0020	-0.0004	0.0008	0.0012	0.0012
0.20	0.2857	0.0714	0.2143	0.2979	0.0957	0.2021	0.3097	0.1194	0.1903	0.3107	0.1172	0.1924
	-0.0250	-0.0458	+0.0219	-0.0128	-0.0215	+0.0097	-0.0010	+0.0022	-0.0021	0.0013	0.0021	0.0021
0.40	0.2857	0.0714	0.2143	0.3122	0.1244	0.1878	0.3363	0.1726	0.1637	0.3340	0.1663	0.1640
	-0.0483	-0.0949	+0.0503	-0.0218	-0.0419	+0.0238	+0.0023	+0.0063	-0.0003	0.0016	0.0025	0.0026
0.60	0.2857	0.0714	0.2143	0.3314	0.1629	0.1686	0.3666	0.2332	0.1334	0.3579	0.2200	0.1449
	-0.0722	-0.1485	+0.0694	-0.0265	-0.0571	+0.0237	+0.0087	+0.0133	-0.0115	0.0018	0.0030	0.0027
0.80	0.2857	0.0714	0.2143	0.3615	0.2230	0.1385	0.3996	0.2992	0.1004	0.3893	0.2766	0.1104
	-0.1036	-0.2052	+0.1039	-0.0278	-0.0536	+0.0281	+0.0103	+0.0226	-0.0100	0.0018	0.0031	0.0026
0.99	0.2857	0.0714	0.2143	0.4115	0.3229	0.0885	0.4324	0.3648	0.0676	0.4188	0.3381	0.0818
	-0.1331	-0.2667	+0.1325	-0.0073	-0.0152	+0.0067	+0.0136	+0.0267	-0.0142	0.0017	0.0027	0.0024

For $\mathbf{N} = diag\{20,0,0,10,0,30\}$, eight tested offspring per dam (n_o=8), ancestor cohort (t_1) = 10 and descendent cohort (t_2) = 35. [a]Deviations from simulation results are on every second line. [b]Standard errors are on every second line. [c] h_0^2 denotes heritability.

and 4) and improved modified gene flow (*Method P*, Equations 4 and 6), for mass and sib-index selection.

Mass selection

Accuracy of α: Table 1 shows long-term genetic contributions of categories ($n_k\alpha_k$), obtained from *Method C*, *Method M*, *Method P* and from simulation, for a population with three age classes, 20 sires in age class 1, 10 dams in age class 1 and 30 dams in age class 3, *i.e.*, $\mathbf{N} = diag\{20,0,0,10,0,30\}$. This scheme, with a high proportion of dams selected from the oldest age class, was chosen because it clearly illustrates the effect of selective advantage on contributions of categories.

Results from *Method C* are independent of heritability (h_0^2), but results from *Method M*, *Method P* and from simulation are not. For h_0^2 = 0.01, results from all methods are practically identical because heritable effects play a minor role in that case. For higher heritabilities, *Method C* shows considerable overestimates of contributions from three-year-old dams ($n_6\alpha_6$), whereas *Method M* and *P* are significantly closer, and from these, *Method P* is most accurate. For high heritabilities (>0.6), absolute differences between *Method P* and simulation are roughly only 10% of the errors from *Method C*, and for this particular scheme in the opposite direction. The large differences between *Method C* and simulation are partly caused by the distribution of parents across age classes in Table 1. Because most dams are selected from the oldest category, offspring from these dams will have a low breeding value which will reduce their genetic contribution. When parents are selected by truncation across age classes, differences between *Method C* and simulation will be much smaller (see DISCUSSION).

Comparing *Method M* and *P* to simulation results, shows that the first-order correction improves the accuracy of the predicted long-term genetic contributions. In Equation 3, differences between selective advantage of

TABLE 2.–Regression coefficients of long term genetic contributions on breeding values (β_i) under mass selection

	Method M[a]			Method P[a]			simulation[b]		
h_0^2 [c]	β_1	β_4	β_6	β_1	β_4	β_6	β_1	β_4	β_6
0.01	0.0228	0.0116	0.0114	0.0229	0.0116	0.0114	0.0233	0.0102	0.0128
	-0.0005	+0.0014	-0.0014	-0.0004	+0.0014	-0.0014	0.0013	0.0017	0.0008
0.20	0.0211	0.0136	0.0095	0.0221	0.0144	0.0100	0.0205	0.0159	0.0093
	+0.0006	-0.0023	+0.0002	+0.0016	-0.0015	+0.0007	0.0005	0.0008	0.0003
0.40	0.0201	0.0161	0.0081	0.0220	0.0182	0.0086	0.0217	0.0205	0.0078
	-0.0016	-0.0044	+0.0003	+0.0003	-0.0023	+0.0008	0.0005	0.0009	0.0002
0.60	0.0199	0.0197	0.0067	0.0222	0.0233	0.0070	0.0199	0.0246	0.0065
	+0.0000	-0.0049	+0.0002	+0.0023	-0.0013	+0.0005	0.0006	0.0010	0.0002
0.80	0.0204	0.0254	0.0052	0.0226	0.0298	0.0051	0.0210	0.0291	0.0053
	-0.0006	-0.0037	-0.0001	+0.0016	+0.0007	-0.0002	0.0007	0.0012	0.0002
0.99	0.0220	0.0347	0.0031	0.0230	0.0372	0.0029	0.0246	0.0362	0.0047
	-0.0026	-0.0015	-0.0016	-0.0016	+0.0010	-0.0018	0.0012	0.0021	0.0003

For $\mathbf{N} = diag\{20,0,0,10,0,30\}$, eight tested offspring per dam ($n_o = 8$), ancestor cohort (t_1) = 10 and descendent cohort (t_2) = 35. [a] Deviations from simulation results are on every second line. [b]Standard errors are on every second line. [c] h_0^2 denotes heritability.

selected offspring from different parental categories (d_{kl}) are ignored, resulting in underprediction of contributions of young categories and in overprediction of contributions of older categories (except for $h_0^2 = 0.99$, probably due to deviations from normality for this extreme case, which is of little practical importance).

Accuracy of β: Table 2 shows the regression coefficients of contributions on breeding values (β), from *Method M*, *Method P* and from simulation, for $\mathbf{N} = diag\{20,0,0,10,0,30\}$. Most predictions from *Method P* are within 3 times the standard error of simulation results, and the trends in predictions agree well with simulation results. *Method P* was slightly more accurate than *Method M*, particularly when modeling the differences between one and 3-year-old females, *i.e.*, β_4 and β_6. In *Method C*, the effect of selective advantage is not modeled, *i.e.*, β is implicitly zero.

Accuracy of genetic gain and generation interval: Table 3 shows genetic gain per year and generation interval from *Method C*, *Method M*, *Method P* and from simulation, for $\mathbf{N} = diag\{20,0,0,10,0,30\}$. Generation interval was calculated from Equation 7. For *Method C*, generation interval from Equation 7 is identical to the average age of parents when their progeny are born and is obtained from $\mathbf{G}_0$. Generation intervals based on the average age of parents of selected offspring, as suggested by BICHARD *et al.* (1973), are obtained from $\mathbf{G}$ (see example in APPENDIX C) and are also in Table 3. *Method C* does not account for the effect of selection on genetic contributions and therefore results in higher generation intervals than simulation. For the scheme in Table 3,

TABLE 3.–Rate of genetic gain (ΔG) and generation interval (L) with mass selection

	Method C[a]		Method M[a]		Method P[a]		selected offspring	simulation[b]	
$h_0^{2\,c}$	ΔG	L	ΔG	L	ΔG	L	L	ΔG	L
0.01	0.0090	1.750	0.0091	1.746	0.0092	1.743	1.746	0.0089	1.735
	+0.0001	+0.015	+0.0002	+0.011	+0.0002	+0.008	+0.011	0.0005	0.005
0.20	0.1751	1.750	0.1687	1.679	0.1770	1.615	1.678	0.1702	1.612
	+0.0049	+0.138	-0.0015	+0.067	+0.0068	+0.003	+0.066	0.0005	0.007
0.40	0.3386	1.750	0.3222	1.602	0.3513	1.487	1.595	0.3363	1.505
	+0.0023	+0.245	-0.0141	+0.097	+0.0150	-0.018	+0.090	0.0008	0.008
0.60	0.4960	1.750	0.4774	1.509	0.5325	1.364	1.500	0.5044	1.384
	-0.0084	+0.366	-0.0270	+0.125	+0.0281	-0.020	+0.116	0.0009	0.008
0.80	0.6518	1.750	0.6533	1.383	0.7224	1.251	1.381	0.6749	1.288
	-0.0231	+0.462	-0.0216	+0.095	+0.0475	-0.037	+0.093	0.0009	0.006
0.99	0.8024	1.750	0.8691	1.215	0.9098	1.156	1.244	0.8420	1.192
	-0.0396	+0.558	+0.0271	+0.023	+0.0678	-0.036	+0.052	0.0009	0.005

For $\mathbf{N} = diag\{20,0,0,10,0,30\}$, eight tested offspring per dam ($n_o = 8$), ancestor cohort (t_1) = 10 and descendent cohort (t_2) = 35. [a]Deviations from simulation results are on every second line. [b] Standard errors are on every second line. [c] h_0^2 denotes heritability.

most dams are selected from the oldest category, which increases differences between *Method C* and *Method P*. Even when the numbers of females selected were exchanged, however, *i.e.*, $\mathbf{N} = diag\{20,0,0,30,0,10\}$, there were differences between generation intervals from *Method C* and *Method P*. (See Figure 2). *Method M* showed systematic overprediction of generation intervals, which agrees with the overprediction of contributions of older categories (see Table 1). Predicted generation intervals based on the average age of parents of selected offspring, *i.e.*, from $\mathbf{G}$ rather than $\mathbf{G}_0$ were very close to generation intervals from *Method M*. Generation intervals from *Method P* were close to simulation results, only showing minor underprediction for high heritabilities.

For this particular scheme, genetic gain from *Method C* was more accurate than gain from *Method P*. However, this was not a general result, *e.g.*, for $\mathbf{N} = diag\{20,0,0,30,0,10\}$ (results not shown) it was the other way around. In general, both methods showed similar accuracies for predicting genetic gain.

Effect of heritability and selection intensity on α: The effect of heritability and selection intensity on average genetic contributions of categories ($n_k\alpha_k$) was studied using *Method P*. Figure 1 shows the predicted long-term genetic contribution of one-year-old females as a proportion of the total contribution of females $[n_4\alpha_4/(n_4\alpha_4 + n_6\alpha_6)]$, for two different breeding schemes and for two selection intensities. The breeding schemes where S_1: $\mathbf{N} = diag\{20,0,0,30,0,10\}$ and S_2: $\mathbf{N} = diag\{20,0,0,10,0,30\}$. Selection intensity was varied by varying the number of tested offspring per dam, *i.e.*, n_o was 4 or 20. To illustrate the relation between genetic contributions and generation interval, Figure 2 shows the corresponding generation interval. In S_1 and S_2, males are selected from a single age,

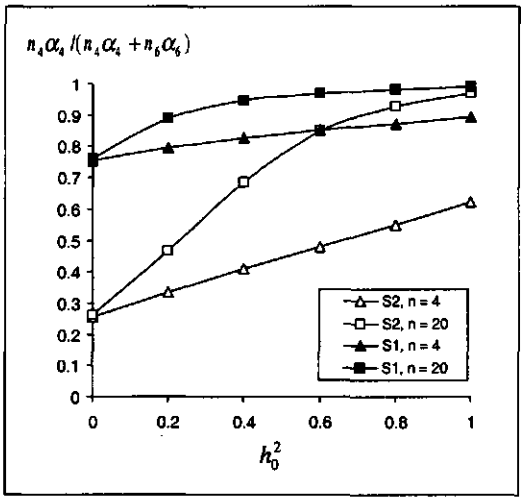

FIGURE 1.–Predicted long term genetic contributions from *Method P* of one year old females as a proportion of the total contribution of females $n_4\alpha_4/(n_4\alpha_4+n_6\alpha_6)$, as a function of heritability, for S_1: **N** = *diag*{20,0,0,30,0,10) and S_2 : **N** = *diag*{20,0,0,10,0,30}, for n_o = 4 or n_o = 20 tested offspring per dam, and mass selection ($b_1 = b_2 = b_3$).

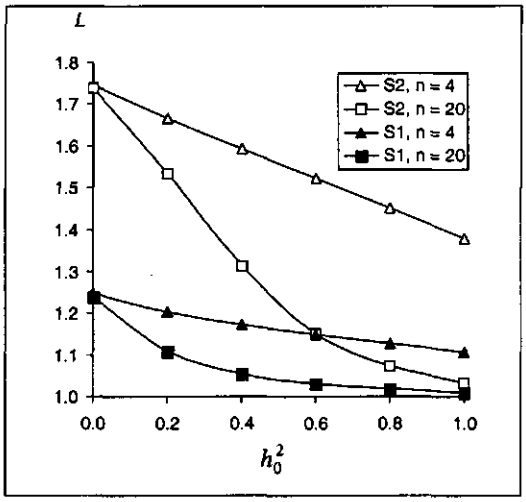

FIGURE 2.–Predicted generation interval (L) from *Method P*, as a function of heritability, for S_1: **N** = *diag*{20,0,0,30,0,10) and S_2 : **N** = *diag*{20,0,0,10,0,30}, for n_o = 4 or n_o = 20 tested offspring per dam, and mass selection ($b_1 = b_2 = b_3$).

and L is directly related to $n_4\alpha_4 / (n_4\alpha_4 + n_6\alpha_6)$. Results from *Method C* are identical to results for $h_0^2 = 0$.

Figure 1 clearly shows an increased contribution of one-year-old females when heritability increases, which is due to an increased selective advantage of offspring descending from one-year-old dams when h_0^2 increases. When heritability increased from 0.2 to 0.8, genetic gain per year increased from 0.232 to 0.977 units σ_p for **N** = *diag*{20,0,0,10,0,30} and n_o = 20. Consequently, the difference between average breeding values of offspring from one and from three-year-old dams increased from 0.277 to 1.153. This selective advantage resulted in an increased proportion of offspring selected from one-year-old dams when h_0^2 increased. When h_0^2 increased from 0.2 to 0.8, it showed that among the one-year-old selected females, the proportion descending from one-year old dams increased from 0.386 to 0.894. [These proportions were determined from the **G** matrix (not shown)].

The relative long-term genetic contribution of one-year-old females also increased with n_o (see Figure 1), *i.e.*, with selection intensity. This is partly due to increased genetic gain resulting in an increased selective advantage of newborn offspring of one-year-old dams, in the same way as when h_0^2 increases, but also due to a decreased overall selected proportion moving the common truncation point of subclasses to the right. When a common truncation point for two normal distributions with different means is

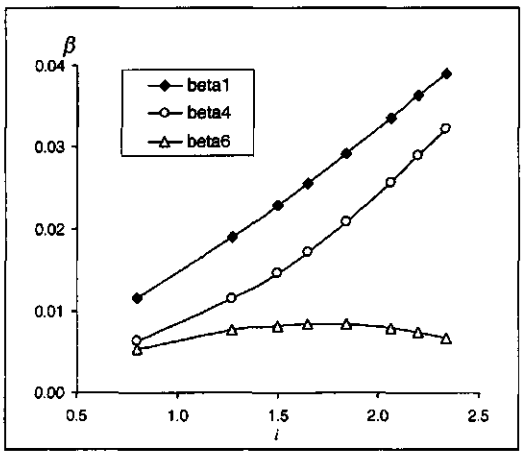

FIGURE 3.–Predicted regression coefficients of long term genetic contributions on breeding values (β) from *Method P* as a function of selection intensity (i), for age class one males (β_1), age class one females (β_4) and age class three females (β_6), for $h_0^2 = 0.4$, N = *diag*{20,0,0,20,0,20} and mass selection ($b_1 = b_2 = b_3$).

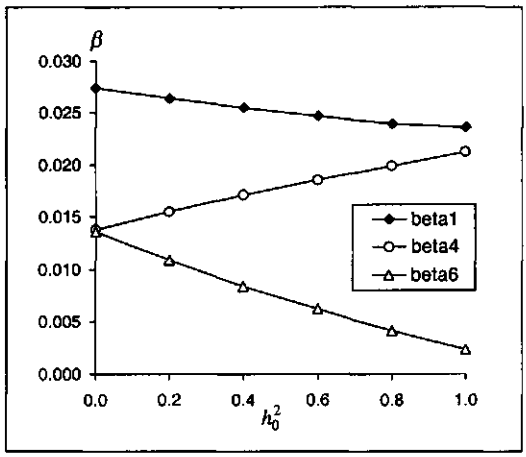

FIGURE 4.–Predicted regression coefficients of long term genetic contributions on breeding values (β) from *Method P* as a function of heritability (h_0^2), for age class one males (β_1), age class one females (β_4) and age class three females (β_6), for N = *diag*{20,0,0,20,0,20}, $n_o = 8$ tested offspring per dam and mass selection ($b_1 = b_2 = b_3$).

shifted to the right, the smaller upper-tail proportion of the two will decrease more rapidly than the larger upper-tail proportion, due to the non-linear relation between truncation point and selected proportion, therefore decreasing the relative contribution of three-year-olds. This effect can be illustrated by comparing the relative contribution of one-year-old females between schemes with different selection intensities at the same ΔG, because at the same ΔG the difference between mean breeding values of one and three-year-old dams will be the same. For N = *diag*{20,0,0,10,0,30}, $n_o = 20$ and $h_0^2 = 0.4$, ΔG was 0.4854, and the same ΔG can be obtained with identical N, but with $n_o = 4$ and $h_0^2 = 0.77$. However, the relative contribution of one-year-old females differed considerably; 0.685 for $n_o = 20$ compared to 0.540 for $n_o = 4$ (see Figure 1), mainly due to different selection intensities.

Effect of selection intensity on β: Figure 3 shows the relation between selection intensity and β for a scheme with N = *diag*{20,0,0,20,0,20} using *Method P*. Selection intensity is equal for all categories in this scheme, and was varied by varying the number of tested offspring per dam from $n_o = 2$ ($i = 0.798$) to $n_o = 40$ ($i = 2.336$).

Figure 3 shows an increase in β_1 and β_4 with increasing selection intensity. On average, β is expected to increase with selection intensity because the regression of selected number of offspring on breeding value (λ) increases with selection intensity (see APPENDIX A) and β is positively related to λ (see Equation 2), explaining the trend for β_1 and β_4. For β_6 the increase with selection intensity is counteracted

TABLE 4.–Regression coefficients of long-term genetic contributions on breeding values (β) for a sib index with positive weight on family information

	Method P[a]			simulation[b]		
h_0^2 [c]	β_1	β_4	β_6	β_1	β_4	β_6
0.01	0.0391	0.0188	0.0181	0.0359	0.0149	0.0175
	+0.0032	+0.0039	+0.0006	0.0019	0.0021	0.0010
0.20	0.0346	0.0251	0.0132	0.0310	0.0255	0.0113
	+0.0036	-0.0004	+0.0019	0.0008	0.0012	0.0004
0.40	0.0325	0.0321	0.0096	0.0307	0.0295	0.0081
	+0.0018	+0.0026	+0.0015	0.0008	0.0013	0.0003
0.60	0.0311	0.0389	0.0064	0.0280	0.0339	0.0077
	+0.0031	+0.0050	-0.0013	0.0008	0.0014	0.0003
0.80	0.0299	0.0448	0.0037	0.0295	0.0397	0.0053
	+0.0004	+0.0051	-0.0016	0.0010	0.0017	0.0003
0.99	0.0293	0.0498	0.0044	0.0293	0.0404	0.0040
	-0.0000	+0.0094	+0.0004	0.0011	0.0021	0.0004

For $N = diag\{20,0,0,10,0,30\}$, $n_o = 8$, $b_1 = 1.0$, $b_2 = 1.5$, $b_3 = 2$, ancestor cohort (t_1) = 10, and descendent cohort (t_2) = 35. [a]Deviations from simulation results are on every second line. [b] Standard errors are on every second line. [c] h_0^2 denotes heritability.

by the reduced total contribution of 3-year-old dams (see Figure 1). For other heritabilities (results not shown) the relation between β and selection intensity was similar.

Effect of heritability on β: Figure 4 shows the relation between β and heritability using *Method P*. For $h_0^2 = 0$, $\beta_4 = \beta_6 = \frac{1}{2}\beta_1$, which is to be expected from (2) when selection intensity is equal for all categories and $g_{kl} = g_{kl}^0$ because $h_0^2 = 0$. When h_0^2 increases, genetic gain increases, resulting in a higher proportion of *selected* offspring descending from one-year-old parents, *i.e.*, for all k, $g_{k4} > g_{k6}$ for $h_0^2 > 0$. When $g_{k4} > g_{k6}$ and selection intensity is equal for all categories, it can be inferred from (2) that $\beta_4 > \beta_6$ as in Figure 4.

It is a general conclusion for mass selection, therefore, that β of younger categories will increase with h_0^2, whereas β of older categories will decrease with h_0^2. The interpretation of this relation is, that with mass selection the contributions of young animals will increasingly be determined by their breeding values when h_0^2 increases, whereas for older animals the effect of breeding value on contributions will decrease with increasing h_0^2. An intuitive way of looking at this is, that for influential animals (which are young animals when h_0^2 is high) a change of breeding value gives a larger (absolute) change of genetic contributions than it does for unimportant animals. The same reasoning holds for the relation between β and

TABLE 5.–Regression coefficients of long-term genetic contributions on breeding values (β) for a sib index with negative weight on family information

	Method P[a]			simulation[b]		
h_0^2 [c]	β_1	β_4	β_6	β_1	β_4	β_6
0.01	0.0015	0.0023	0.0023	0.0013	0.0007	0.0018
	+0.0002	+0.0016	+0.0005	0.0011	0.0012	0.0006
0.20	0.0016	0.0025	0.0024	0.0016	0.0023	0.0023
	+0.0000	+0.0002	+0.0001	0.0002	0.0003	0.0001
0.40	0.0017	0.0028	0.0024	0.0016	0.0030	0.0024
	+0.0001	-0.0002	+0.0000	0.0002	0.0002	0.0001
0.60	0.0018	0.0032	0.0025	0.0016	0.0034	0.0025
	+0.0002	-0.0002	+0.0000	0.0002	0.0002	0.0001
0.80	0.0019	0.0037	0.0026	0.0021	0.0045	0.0026
	-0.0002	-0.0008	+0.0000	0.0002	0.0003	0.0001
0.99	0.0021	0.0045	0.0027	0.0024	0.0044	0.0026
	-0.0003	+0.0001	+0.0001	0.0002	0.0003	0.0001

For $\mathbf{N} = diag\{20,0,0,10,0,30\}$, eight tested offspring per dam ($n_o = 8$), $b_1 = 1.0$, $b_2 = 0.5$, $b_3 = 0$, ancestor cohort (t_1) = 10 and descendent cohort (t_2) = 35. [a]Deviations from simulation results are on every second line. [b] Standard errors are on every second line. [c] h_0^2 denotes heritability.

selection intensity, explaining the different trend of β_4 and β_6 in Figure 3.

The regression coefficient for one-year-old males (β_1) shows only minor variation with h_0^2 because males are selected from a single category in Figure 3. Therefore, category one always contributes 50% of the genes of selected offspring ($g_{11} = g_{41} = g_{61} = 0.5$) regardless of heritability, and variation of β_1 with h_0^2 is only due to variation in λ and π.

Selection on a sib index

Long-term genetic contributions of categories ($n_k\alpha_k$) are mainly dependent on the modified gene flow matrix. For a sib index, **G** is determined by genetic gain and selected proportions, in the same way as for mass selection. The main differences between sib index and mass selection are, therefore, in the regressions λ and π, resulting in different predictions for β. Results for a sib index, therefore, focus on β, though α will also differ from results for mass selection.

Accuracy of β: Predictions for a sib index are compared to simulation results for two opposite schemes: a scheme with positive weight on family information and a scheme with negative weight on family information. The weights used are different from the classical selection index weights (HAZEL, 1943), but as shown by VILLANUEVA and WOOLLIAMS (1997), optimum index weights for intermediate and long-term response are generally different from classical index weights.

For positive weight on family information, Table 4 shows β from *Method P* and from simulation, for $\mathbf{N} = diag\{20,0,0,10,0,30\}$, b_1 =

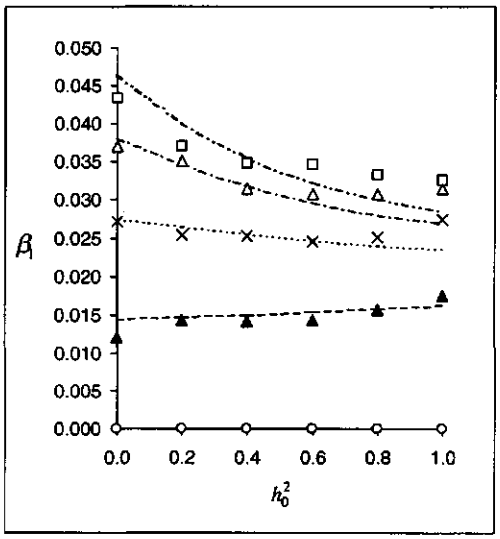

FIGURE 5.–Predicted and simulated regression coefficients of long term genetic contributions on breeding values for one year old males (β_1) from *Method P*, for different index weighting factors (b) and a range of heritabilities (h_0^2). For N = $diag\{20,0,0,20,0,20\}$ and n_o = 8. Lines indicate predictions, markers indicate results from simulation. —○—, b = 1, 0, 0; - - ▲ - - , b = 1, ½, ½; ····×····, b = 1, 1, 1; -·-Δ -·-, b = 1, 1½, 1½; -···□-···-, b = 1, 2, 2.

1, b_2 = 1.5 and b_3 = 2 (*i.e.*, $I = P + \frac{1}{2}\overline{P}_{FS} + \frac{1}{2}\overline{P}_{HS}$). In Table 4, *Method P* shows the same trend as simulation results, but tends to slightly overestimate regression coefficients for one-year-old parents (β_1 and β_4). Predictions of α (results not shown) were close to simulation results and showed similar trends as for mass selection.

For negative weight on family information, Table 5 shows β from *Method P* and from simulation, for N = $diag\{20,0,0,10,0,30\}$, b_1 = 1, b_2 = 0.5 and b_3 = 0 (*i.e.*, $I = P - \frac{1}{2}\overline{P}_{FS} - \frac{1}{2}\overline{P}_{HS}$). In Table 5, *Method P* shows the same trend as simulation results, and is accurate. Predictions for α (results not shown) were very accurate.

Effect of index weights on β: Figure 5 shows the effect of a varying emphasis on family information in the selection index on the regression coefficients of long-term genetic contributions on breeding values, for one-year-old male parents (β_1), from *Method P* (lines) and from simulation (markers) for N = $diag\{20,0,0,20,0,20\}$. For this scheme, β_1 gives a good impression of the average level of β, because males are selected from a single category, so that there is no competition between categories. In Figure 5, the index weights vary from b_1 = 1, b_2 = b_3 = 0, representing complete within family selection, to b_1 = 1, b_2 = 2, b_3 = 2 , which is identical to $I = P + \overline{P}_{FS}$.

For within family selection, β equals zero because offspring are selected on their Mendelian sampling term, which by definition is independent of the parental breeding value. Therefore, selective advantage is not inherited and results (both α and β) are identical to results from *Method C*.

When index weights on family information increased, β_1 increased because selection of offspring was increasingly affected by the parental breeding value. Similar relations between the average level of β and weight given to family information were found for other distributions of parents across categories (including schemes with competition between categories).

When weight on family information increases, selection tends to selection of families instead of individuals, whereas λ is derived assuming a continuous linear change. Accuracy of predictions decreased therefore, when weight given to family information became high, which is shown by the increased difference between lines and markers in Figure 5.

Discussion

This article has studied in detail two methods proposed by WOOLLIAMS *et al.* (1999) for the prediction of long-term genetic contributions of individuals in selected populations with overlapping generations. The methods enable accurate the prediction of long-term genetic contributions of individual animals and of categories, using simple linear models. Predictions of genetic contributions within categories were first shown by WOOLLIAMS *et al.* (1999) but never studied in detail. Genetic contributions were predicted conditional on the breeding value and category of the ancestor, using a modified gene flow approach. The method accounts for the inheritance of selective advantage both between and within categories, resulting in more accurate predictions of genetic contributions and generation intervals than methods based on contributions to newborn offspring in the next cohort.

Some trends in the prediction errors remained (*e.g.*, Table 1, Figure 5), but this is merely a matter of improving the relevant regression equations, they do not undermine the basic ideas underlying the theory. Conventional methods ignore the effect of selection on genetic contributions and, therefore, underestimate contributions of younger categories and overestimate generation interval. Thus, improved methods were generally necessary.

Accurate predictions of long-term genetic contributions for overlapping generation schemes facilitate deterministic prediction of rates of inbreeding for these schemes (WOOLLIAMS, 1998), and consequently a computationally feasible optimization of breeding schemes with restricted inbreeding. The modified gene flow approach enables prediction of *individual* long-term genetic contributions [by including $\beta_k(A_{i(k)} - \overline{A}_k)$ in the model for expected contributions], whereas conventional gene flow theory only enables prediction of average genetic contributions (*i.e.*, assuming $\beta = 0$). For the prediction of rates of inbreeding it is crucial to account for the effect of selection between individuals (WRAY *et al.* 1989), and conventional gene flow is therefore not suitable for prediction of rates of inbreeding.

In the present study, generation interval was defined as $L = 1/\Sigma n_k \alpha_k$, *i.e.*, the generation interval is the time in which long-term genetic contributions sum to unity. Intuitively, this is a sensible definition, one generation is the time in which the genes are turned over once. The definition of generation interval as the time in which contributions sum to unity is general and is also applicable to generation intervals based on newborn offspring or on selected offspring. For example, generation interval based on newborn offspring, *i.e.*, the average age of

parents when their offspring are born, can also be calculated as: $L_0 = 1/\Sigma\alpha_0$, where α_0 are contributions obtained from conventional gene flow theory. Generation interval based on contributions to selected offspring only (L_1), *i.e.*, the average age of parents of selected offspring, can be obtained from the modified gene flow matrix **G** (see APPENDIX C) and was close to results from simple modified gene flow. When genetic gain is made and selective advantage is inherited, generation interval based on long-term genetic contributions is shorter than both L_0 and L_1, because selective advantage is partly passed on to more distant offspring.

Whereas L_0 and L_1 are based on contributions at a specific time point, *i.e.*, before and immediately after selection of the offspring, L is based on converged, *i.e.*, asymptotic long-term genetic contributions of parental categories, which are an invariable property of a population once contributions have converged. Therefore, the definition of generation interval based on long-term genetic contributions is equal to the turnover time of genes that are destined to stay in the population, *i.e.*, it is the average time interval between two meioses, and it is of a more genetical and less operational nature than L_0 and L_1.

In the present study, results are only presented for situations where the selection index of an animal was constant across ages. In practice, animals in different categories will often have different amounts of information, affecting the variance of the selection index. This will mainly affect the **G** matrix, but is easily accounted for by using index variances specific to categories in the equations presented in the METHODS section. The problem is more complex for the prediction of rates of inbreeding, since in that case the lifetime genetic contribution of an ancestor, *i.e.*, its contribution summed over all categories it belonged to over its entire life, is relevant, which requires the probability that the same animal was selected in multiple categories.

Large differences were found between predicted genetic contributions from conventional and from modified gene flow in the present article. These differences were partly caused by the distribution of parents across categories, *i.e.*, in Table 1 and 3 the majority of the dams were selected from the oldest category. When animals are selected by truncation across categories, differences in generation interval between the two methods will be much smaller. For example, for h_0^2 = 0.5, n_o = 4, N_m = 20, N_f = 40, truncation selection across categories resulted in **N** = *diag*{18,2,33,7}, predicted generation interval from conventional gene flow was 1.138 and from modified gene flow 1.129 (simulation: L = 1.130). An advantage of modified gene flow is that it gives accurate predictions of generation interval for an arbitrary distribution of parents across categories, it is not limited to truncation selection across categories.

In the present article, the within family variance was assumed to be constant over time, which is not strictly true when inbreeding is accumulating. However, genetic contributions are mainly determined in the first few generations, where the inbreeding effects on descendants are still small. Long-term genetic

contributions are, therefore, hardly affected by a reduction of variance due to inbreeding. Furthermore, ignoring the effect of inbreeding on the variance allows for the assumption of BULMER's (1971) equilibrium variances (assuming the infinitesimal model), which greatly simplifies prediction equations for long-term genetic contributions (WOOLLIAMS *et al.* 1999). For extremely small populations, *e.g.*, with less than five parents per sex, it may become important to account for the effect of inbreeding when predicting long term genetic contributions.

The number of parents is no guarantee for the genetic constitution of populations on the long term, because selective advantage of parents is inherited to offspring. This is a point of concern for conservation genetics where genetic improvement is also being sought. Simply increasing the number of parents may not safeguard the genetic diversity of a population when offspring of the additional parents have a low chance of being selected.

The inheritance of selective advantage is crucial in the prediction of long term genetic contributions, and thus for the prediction of rates of inbreeding (WRAY and THOMPSON, 1990). Recently, NOMURA (1997) studied rates of inbreeding in open nucleus breeding systems with discrete generations, assuming that genetic contributions of parental groups (nucleus and commercial animals) to progeny remain unchanged after selection. As recognized by NOMURA (1997), this is a critical assumption, and especially in populations with overlapping generations it is likely to be strongly violated.

Asymptotically, response from conventional gene flow is equal to response obtained using the well-known result of RENDEL and ROBERTSON (1950) (HILL, 1974). When comparing gain obtained from conventional and modified gene flow to simulation results, predictions from both methods showed similar accuracy. For the prediction of genetic gain, the ratio of selection differential over generation interval is crucial, rather than the definitions of selection differential and generation intervals separately. When generation interval is defined as the average age of parents of all offspring, and selection differential is defined as the deviation of selected parents from the overall mean, valid predictions for genetic gain are obtained (JAMES, 1977). Conventional gene flow therefore is a valid method for predicting genetic gain. The relevance of the current theory lies in predicting the development of pedigree, *i.e.*, of the origin and turnover rate of genes, and in predicting rates of inbreeding; the primary aim is not to revise methods for predicting rates of gain.

Johan A. M. Van Arendonk is gratefully acknowledged for encouraging and giving PB the opportunity to visit JAW, and for giving useful comments on this manuscript. One author (JAW) gratefully acknowledges: the Ministry of Agriculture, Fisheries and Food (UK) for financial support. Jack C. M. Dekkers is acknowledged for giving very useful comments on this manuscript. This research was financially supported by the Netherlands Technology Foundation (STW), and was coordinated by the Life Sciences Foundation (SLW).

Literature Cited

BICHARD, M., A. H. R. PEASE, P. H. SWALES and K. ÖZKÜTÜK, 1973. Selection in a population with overlapping generations. Anim. Prod. 17: 215-227.

BULMER, M. G., 1971. The effect of selection on genetic variability. Am. Nat. 105: 201-211.

COCHRAN, W. G., 1951. Improvement by means of selection, pp. 449-470 in *Proceedings of the 2nd Berkeley Symposium on Mathematics, Statistics and Probability,* edited by J. NEYMAN. University of California Press, Berkeley.

DEMPSTER, E. R. and I. M. LERNER, 1950. Heritability of threshold characters. Genetics 35: 212-236.

FISHER, R. A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Roy. Soc. (Edinburgh), Trans. 52: 321-341.

HAZEL, L. N., 1943. The genetic basis for constructing selection indexes. Genetics 28: 476-490.

HILL, W. G., 1974. Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

JAMES, J. W., 1977. A note on selection differential and generation length when generations overlap. Anim. Prod. 24: 109-112.

JAMES, J.W., 1987. Determination of optimal selection policies. J. Anim. Breed. Genet. 104: 23-27.

JAMES, J. W. and G. MCBRIDE, 1958. The spread of genes by natural and artificial selection in a closed poultry flock. J. Genet. 56: 55-62.

NOMURA, T., 1997. Prediction of effective population size in open nucleus breeding systems. J. Anim. Breed. Genet. 114: 333-347.

PEARSON, K., 1903. On the influence of natural selection on the variability and correlation of organs. Roy. Soc. London Phil. Trans. A, Vol. 200: 1-66.

PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING and B. P. FLANNERY, 1992. *Numerical Recipes in FORTRAN: The Art of Scientific Computing.* Cambridge University Press, Cambridge.

RENDEL, J. M. and A. ROBERTSON, 1950 Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J. Genet. 50: 1-8.

VILLANUEVA, B. and J. A. W. WOOLLIAMS, 1997 Optimisation of breeding programmes under index selection and constrained inbreeding. Genet. Res. Camb. 69: 145-158.

VILLANUEVA, B., N. R. WRAY, and R. THOMPSON, 1993 Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. Anim. Prod. 57: 1-13.

WOOLLIAMS, J. A., 1998 A recipe for the design of breeding schemes. In *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production,* vol. 25, pp. 427.

WOOLLIAMS, J. A. and R. THOMPSON, 1994 A theory of genetic contributions. In *Proceedings of the5th World Congress on Genetics Applied to Livestock Production,* vol. 19, pp. 127-134.

WOOLLIAMS, J. A., N. R. WRAY and R. THOMPSON, 1993 Prediction of long term contributions and inbreeding in populations undergoing mass selection. Genet. Res. Camb. 62: 231-242.

WOOLLIAMS, J. A., P. BIJMA and B. VILLANUEVA, 1999. Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153: 1009-1020.

WRAY, N. R. and R. THOMPSON, 1990. Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55: 41-54.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1990. Methods for predicting rates of inbreeding in selected populations. Theor. Appl. Genet. 80: 503-512.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1994. Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. 87: 878-892.

Appendix A

Derivation of Λ: Elements λ_{kl} are obtained as: $\lambda_{kl} = p_k^{-1} \frac{1}{2}\tau_x \, b_{S_k I_k}$, where $\frac{1}{2}\tau_x$ is the regression of the index of the offspring on the breeding value of the parent of sex x (WRAY *et al.* 1994), and $b_{S_k I_k}$ is the regression of the selection score (selected or not selected, *i.e.*, $S = 1$ or 0) of the offspring on its index. In Equation 2, λ_{kl} is expressed per selected offspring, whereas S is an expression per selection candidate, the difference being on average a factor p_k^{-1}. Tau is

obtained as: $\tau_x = 2\mathrm{Cov}(A_x, I_{offsp})/\mathrm{Var}(A_x)$ resulting in: $\tau_m = b_3$, $\tau_f = b_2(1-\frac{1}{d}) + b_3/d$. From a result of Robertson (appendix in DEMPSTER and LERNER, 1950), $b_{S_k I_k} = p_k i_k \sigma_I^{-1}$. Resulting expressions are $\lambda_{kl} = \frac{1}{2} b_3 i_k \sigma_I^{-1}$ for male parents, and $\lambda_{kl} = \frac{1}{2}[b_2(1-\frac{1}{d}) + b_3/d] i_k \sigma_I^{-1}$ for female parents.

Derivation of Π: Elements π_{kl} are obtained as: π_{kl} = $\mathrm{Cov}(A_{i(l)}, A_{j(k)})^*/\mathrm{Var}(A_{i(l)})^*$, where * denotes (co)variances after selection of the offspring. Using (COCHRAN, 1951): $\sigma^*_{AB} = \sigma_{AB} - \frac{\sigma_{AI}\sigma_{BI}}{\sigma_I^2}\kappa$ for the calculation of $\mathrm{Cov}(A,B)$ after selection on I gives:

$$\mathrm{Cov}(A_{i(l)}, A_{j(k)})^* = \mathrm{Cov}(A_{i(l)}, A_{j(k)}) - \frac{\mathrm{Cov}(A_{i(l)}, I_{j(k)})\mathrm{Cov}(A_{j(k)}, I_{j(k)})}{\mathrm{Var}(I_{j(k)})}\kappa_k,$$

where $\mathrm{Cov}(A_{i(l)}, A_{j(k)}) = \frac{1}{2}\sigma^{2*}_{A(x)l}$, $\mathrm{Cov}(A_{i(l)}, I_{j(k)}) = \frac{1}{2}\tau_x\sigma^{2*}_{A(x)l}$, $\mathrm{Cov}(A_{j(k)}, I_{j(k)}) = \rho\sigma_A\sigma_I$, $\mathrm{Var}(I_{j(k)}) = \sigma_I^2$, $\sigma^{2*}_{A(x)l}$ is the additive genetic variance among selected parents, and x denotes the sex of parent i. Assuming that $\mathrm{Var}(A_{i(l)})$ is little affected by selection among the offspring, *i.e.*, $\mathrm{Var}(A_{i(l)})^* = \sigma^{2*}_{A(x)l}$, the resulting expression becomes: $\pi_{kl} = \frac{1}{2}(1 - \kappa_k\tau_x\rho\sigma_A\sigma_I^{-1})$.

Appendix B

Derivation of ΔG: Genetic gain is obtained from $\mathrm{E}(\Delta G) = \sum_{l=1}^{2c_{max}} n_l E[r_{i(l)}a_{i(l)}]$ (WOOLLIAMS *et al.* 1999), where $\mathrm{E}[r_{i(l)}a_{i(l)}]$ is the expectation of $r_{i(l)}a_{i(l)}$ among selected individuals in category l. With $\mathrm{E}[r_{i(l)}a_{i(l)}] = \mathrm{E}[(u_{i(l)}+e_{i(l)})a_{i(l)}] = \mathrm{E}[u_{i(l)}a_{i(l)}]$, it follows that $\mathrm{E}[r_{i(l)}a_{i(l)}] = \alpha_l \mathrm{E}[a_{i(l)}] + \beta_l \mathrm{E}[a_{i(l)}(A_{i(l)} - \overline{A}_l)]$, where expectations are conditional on selection in category l. Furthermore, $\mathrm{E}[a_{i(l)}] = \frac{\mathrm{Cov}(a_{i(l)}, I_{i(l)})}{\sigma_I^2} i_l \sigma_I = \frac{1}{2}\tau_w\sigma^2_{A0} i_l \sigma_I^{-1}$, where τ_w is the regression of the index on the Mendelian sampling effect, $\tau_w = b_1(1-1/n_o) + b_2[1/n_o - 1/(n_o d)] + b_3/(n_o d)$. With $\mathrm{E}[A_{i(l)} - \overline{A}_l] = 0$, $\mathrm{E}[a_{i(l)}(A_{i(l)} - \overline{A}_l)] = \mathrm{Cov}[a_{i(l)}, (A_{i(l)} - \overline{A}_l)]^* = \mathrm{Cov}[a_{i(l)}, (A_{i(l)} - \overline{A}_l)] - \frac{\mathrm{Cov}(a_{i(l)}, I_{i(l)})\mathrm{Cov}[(A_{i(l)} - \overline{A}_l), I_{i(l)}]}{\sigma_I^2}\kappa_l =$ $\frac{1}{2}\sigma^2_{A0}(1 - \kappa_l\tau_w\rho\sigma_A\sigma_I^{-1})$. Summing elements over categories, the resulting expression for genetic gain becomes

$$\mathrm{E}(\Delta G) = \tfrac{1}{2}\sigma^2_{A_0}[\tau_w\sigma_I^{-1}\sum_{l=1}^{2c_{max}} n_l\alpha_l i_l + \sum_{l=1}^{2c_{max}} n_l\beta_l(1 - \kappa_l\tau_w\rho\sigma_A\sigma_I^{-1})].$$

Appendix C

Example for mass selection: Consider a mass selection scheme ($b_1 = b_2 = b_3 = 1$) with three age classes, $\mathbf{N} = diag\{20,0,0,10,0,30\}$, $h_0^2 = 0.4$, and $n_o = 8$. Selected proportions, selection intensities and variance reduction coefficients are: $p_1 = 0.1250$, $p_4 = 0.0625$, $p_6 = 0.1875$, $i_1 = 1.6467$, $i_4 = 1.9668$, $i_6 = 1.4357$, $\kappa_1 = 0.8171$, $\kappa_4 = 0.8504$, $\kappa_6 = 0.7877$. Tau equals (see equations in APPENDIX A and B): $\tau_w = \tau_m = \tau_f = 1$. The conventional gene flow matrix equals

$$\mathbf{G}_0 = \begin{bmatrix} 0.5 & 0 & 0 & 0.125 & 0 & 0.375 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.125 & 0 & 0.375 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Contributions and generation interval from conventional gene flow are $n_1\alpha_1 = 0.2857$, $n_4\alpha_4 = 0.0714$, $n_6\alpha_6 = 0.2143$, $L = 1/\Sigma\alpha = 1.75$. Equations in the *deterministic prediction procedure* section

were iterated until equilibrium variances were reached, resulting in $\sigma^2_{A(m)} = 0.0630$, $\sigma^2_{A(f)} = 0.1013$, $\sigma^2_A = 0.3643$, $\sigma^2_I = 0.9643$, $\rho = 0.3778$. Based on equilibrium variances, **G**, **Λ**, **Π** and **D** are:

$$\mathbf{G} = \begin{bmatrix} 0.5 & 0 & 0 & 0.205 & 0 & 0.295 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.223 & 0 & 0.277 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.193 & 0 & 0.307 \end{bmatrix},$$

$$\mathbf{\Lambda} = \begin{bmatrix} 0.838 & 0 & 0 & 0.838 & 0 & 0.838 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1.001 & 0 & 0 & 1.001 & 0 & 1.001 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.731 & 0 & 0 & 0.731 & 0 & 0.731 \end{bmatrix},$$

$$\mathbf{\Pi} = \begin{bmatrix} 0.346 & 0 & 0 & 0.346 & 0 & 0.346 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.339 & 0 & 0 & 0.339 & 0 & 0.339 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.351 & 0 & 0 & 0.351 & 0 & 0.351 \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0.187 & 0 & -0.130 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.171 & 0 & -0.138 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.198 & 0 & -0.124 \end{bmatrix}.$$

Categories without parents are not relevant, and have zeroes. **G** identifies the origin of selected offspring, *e.g.*, $g_{14} = 0.205$ means that a proportion of $2 \times 0.205 = 0.410$ of the selected one-year-old males (category 1) descends from one-year-old dams (category 4), *i.e.*, were born when their dam was one-year-old. From **G**, the generation interval based on selected offspring equals: $L_1 = ½ \{0.5 + 0.205 + 3 \times 0.295\} + ½\{10/40 \times (0.5 + 0.223 + 3 \times 0.277) + 30/40 \times (0.5 + 0.193 + 3 \times 0.307)\} = 1.595$. Matrix **D** represents the breeding value of *selected* subgroups as deviation from the total selected group, *e.g.*, $d_{46} = -0.138$ means that one-year-old selected females descending from three-year-old dams, have an average breeding value of 0.138 units below the average of all selected one-year-old females.

Solutions from *Method M* were: $(\mathbf{N}\alpha)^T = (0.312, 0, 0, 0.124, 0, 0.188)$, $\beta^T = (0.0201, 0, 0, 0.0161, 0, 0.0081)$, $L = 1.602$, $\Delta G = 0.3222$. Solutions from *Method P* were: $(\mathbf{N}\alpha)^T = (0.336, 0, 0, 0.173, 0, 0.164)$, $\beta^T = (0.0220, 0, 0, 0.0182, 0, 0.0086)$, $L = 1.487$, $\Delta G = 0.3513$.

A Note on the Relationship between Gene Flow and Genetic Gain

Piter Bijma* **and John A. Woolliams†**

* Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands and †Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K.

Abstract – In conventional gene flow theory, the rate of genetic gain is calculated as the summed products of genetic selection differential and asymptotic proportion of genes deriving from sex-age groups. Recent studies show that asymptotic proportions of genes predicted from conventional gene flow theory may deviate considerably from true proportions. However, the rate of genetic gain predicted from conventional gene flow theory was accurate. The current note shows that the connection between asymptotic proportions of genes and rate of genetic gain that is embodied in conventional gene flow theory is invalid, even though genetic gain may be predicted correctly from it.

In conventional gene flow theory (HILL, 1974), the rate of genetic gain is calculated as the summed products of genetic selection differential and asymptotic proportion of genes deriving from sex-age groups. Recent studies (BIJMA and WOOLLIAMS, 1998; WOOLLIAMS *et al.* 1999) show that asymptotic proportions of genes predicted from conventional gene flow theory may deviate considerably from true proportions. However, the rate of genetic gain predicted from conventional gene flow theory was accurate. The aim of the current note is to clarify this apparent contradiction.

Conventional gene flow theory

Conventional gene flow theory (HILL, 1974) is a method to predict responses and discounted returns from selection in populations with overlapping generations. In conventional gene flow theory, asymptotic response from a single cycle of selection is calculated as the sum of the products of the asymptotic proportion of genes deriving from the different age-sex classes and their genetic selection differential [equation 12 of HILL (1974)]. Since the result agrees with the ratio of mean genetic selection differential ($\bar{S}$) to mean generation interval (L) (RENDEL and ROBERTSON, 1950), HILL (1974) concluded that the product of asymptotic proportion of genes and selection differential is equal to the rate of genetic gain.

HILL (1974) presented two alternative formulations, which are equivalent. First, asymptotic proportions of genes of sex-age groups were defined as the sum of proportions due to current and subsequent matings, and selection differentials were expressed relative to the previous mating. Second, asymptotic proportions of genes were defined as the proportion due to the current mating only, and

selection differentials were expressed as a deviation from the mean of the whole contemporary sex-age group (Hill refers to this as "cumulative selection differential"). Here we will use the second formulation, so that genetic gain predicted from conventional gene flow theory equals ["alternative formulation of (12)", p. 125 of HILL (1974)]:

$$\Delta G = \sum_k r_{k,\infty} S_k \qquad (1)$$

where $r_{k,\infty}$ is the asymptotic proportion of genes deriving from the k^{th} sex-age group, S_k is the genetic selection differential for the k^{th} sex age-group expressed as a deviation from mean of the whole contemporary sex-age group and the sum is taken over all sex-age groups. Furthermore, in conventional gene flow theory, asymptotic proportions of genes are predicted from the proportional contributions of sex-age groups to the newborn offspring. With equal reproductive rates for all age groups, the asymptotic proportions follow directly from the number of parents selected from the respective sex-age groups (equation 11 and 12 in HILL (1974)]:

$$r_{k,\infty} = \tfrac{1}{2} n_k / (N_{\text{sex}(k)} L) \qquad (2)$$

where n_k is the number of parents selected from the k^{th} age-sex group, $N_{\text{sex}(k)}$ is the total number of parents of sex(k), L is the generation interval calculated as the average age of parents when their offspring are born and the "½" makes asymptotic proportions of genes sum to 0.5 per sex per generation.

Asymptotic proportions of genes

Recently, BIJMA and WOOLLIAMS (1998) showed that, in an ongoing breeding program, asymptotic proportions of genes predicted from Equation 2 deviate systematically from true asymptotic proportions. This will be illustrated here by simulated data. Table 1 shows $r_{k,\infty}$ predicted from conventional gene flow theory (Equation 2) and $r_{k,\infty}$ observed in simulated data. The population consisted of 10 one-year-old sires, 30 two-year-old sires, 20 one-year-old dams and 20 two-year old dams, with 3 offspring of each sex per dam. Mass selection was performed for a trait with an initial heritability of 0.4. Additional results are in BIJMA and WOOLLIAMS (1998).

In Table 1, $r_{k,\infty}$ predicted from conventional gene flow theory differs substantially from simulation results. In particular, the true asymptotic proportion of genes from one-year-old parents was higher than the value predicted from Equation 2. The deviations of asymptotic proportions of genes from those predicted by conventional gene flow theory arise from the inheritance of selective advantage (BIJMA and WOOLLIAMS, 1998; WOOLLIAMS *et al.* 1999), an effect ignored in conventional gene flow theory. For example, when one-year-old selected sires have a higher mean breeding value than their selected male contemporaries, offspring of those one-year-old sires will have an increased probability of being selected which increases the asymptotic proportion of genes deriving from one-year-old sires. In such a case, $r_{1,\infty}$ will be higher than the expected proportion based on the contribution of one-year-old sires to the newborn offspring. Therefore, in an

TABLE 1.—Asymptotic proportions of genes deriving from sex-age groups ($r_{k,\infty}$), and rate of genetic gain from Equation 1 ($\Delta G_{\text{Eqn.1}}$) using $r_{k,\infty}$ from conventional gene flow theory (*i.e.* Equation 2) or using $r_{k,\infty}$ from simulation, and genetic gain observed in simulated data (ΔG_{sim}).

	Conventional gene flow theory	Simulation†
$r_{1,\infty}$	0.0769	0.124 (0.002)
$r_{2,\infty}$	0.2308	0.206 (0.002)
$r_{3,\infty}$	0.1538	0.193 (0.002)
$r_{4,\infty}$	0.1538	0.135 (0.002)
$\Delta G_{\text{Eqn.1}}$	0.313	0.345
ΔG_{sim}	-	0.309 (0.001)

For a scheme with 10 one-year-old sires, 30 two-year-old sires, 20 one-year-old dams, 20 two-year-old dams, 3 offspring of each sex per dam, base generation heritability = 0.4, unity phenotypic variance and mass selection. †s.e. are between brackets. Values are based on Bulmer's equilibrium genetic parameters $\sigma^2_{A,eq.} = 0.34$, $h^2_{eq.} = 0.36$, so that genetic selection differentials were: 0.646, 0.446, 0.526, 0.526.

ongoing selection program, Equation 2 is invalid.

For the scheme in Table 1, ΔG predicted from conventional gene flow theory (*i.e.*, Equation 1 and 2) was 0.313, which is close to the 0.309 observed in simulated data. Thus, for the scheme in Table 1, conventional gene flow theory yields an accurate prediction of genetic gain, even though asymptotic proportions of genes predicted from Equation 2 deviate considerably from the true values. However, ΔG predicted from Equation 1 using $r_{k,\infty}$ observed in simulated data differed from simulated ΔG (0.345 vs. 0.309), which indicates that using true asymptotic proportions in Equation 1 does not yield a valid prediction of genetic gain.

Why conventional gene flow theory yields a valid prediction of ΔG

Although Equation 2 is not generally valid, conventional gene flow theory yields a valid prediction of ΔG. This follows from substituting Equation 2 into Equation 1, which gives: $\Delta G = \sum_k r_{k,\infty} S_k = \frac{1}{2L}\sum_k n_k S_k / N_{\text{sex}(k)} = \bar{S}/L$. This equation is identical to the well-known result of RENDEL and ROBERTSON (1950) and yields a valid prediction of the rate of genetic gain as shown unmistakably by JAMES (1977). Therefore, Equation 1 is valid only when $r_{k,\infty}$ is calculated from Equation 2, even though this means that $r_{k,\infty}$ differs from the true asymptotic proportion of genes.

It can be understood intuitively why $r_{\infty,k}$ should refer to the contribution of sex-age groups to newborn offspring the next cohort in Equation 1. As indicated above, the difference between asymptotic proportions predicted from conventional gene flow theory and true asymptotic proportions is due to inheritance of selective advantage, which changes the proportions in subsequent cycles of selection. Though we are concerned with the asymptotic proportion of genes from specific sex-age groups, part of this proportion arises due to subsequent cycles of selection and should therefore not be attributed to genetic gain originating from a single cycle of selection. Because selection in subsequent generations favors descendents of parents with an above average breeding value, the use of true asymptotic proportions in Equation 1 results in an overprediction of the rate of genetic gain.

HOPKINS and JAMES (1979) studied rates of genetic gain based on contributions of parental age groups to *selected* offspring in the next cohort. However, true asymptotic proportions of genes are not only affected by selection among the offspring, but also by subsequent cycles of selection (BIJMA and WOOLLIAMS, 1998; WOOLLIAMS *et al.* 1999). Therefore, asymptotic proportions of genes that can be calculated using methods in HILL (1974) will deviate systematically from true asymptotic proportions. The predicted ΔG of HILL (1974), however, is valid, as shown by JAMES (1977).

Another approach

By decomposing breeding values into Mendelian sampling terms, WOOLLIAMS *et al.* (1999) have shown that the annual rate of genetic gain is equal to the product of the asymptotic proportion of genes deriving from an individual and its Mendelian sampling term, summed over all parents per year,

$$\Delta G = \sum r_{i,\infty} a_i \qquad (3)$$

where $r_{i,\infty}$ is the asymptotic proportion of genes deriving from *individual* i (*i.e.*, its long term genetic contribution), a_i is the Mendelian sampling contribution to the genotype of individual i and the sum is taken over all the parents in a year. Note that Equation 3 is expressed on an individual level, whereas Equation 1 is expressed on a sex-age class level. In Equation 3, genetic gain is attributed to the cohort in which the newly arising variation (*i.e.*, the Mendelian sampling term) is generated. The product of the long-term genetic contribution and the Mendelian sampling term quantifies the impact of an individual on the population mean in the long term. Contrary to Equation 1, genetic gain originating from a specific individual or group accumulates over generations in Equation 3. The convergence of genetic contributions to their equilibrium values takes several cycles of selection. During the first cycles the summed product of genetic contributions and Mendelian sampling terms is lower than ΔG. Selection, however, favors contributions that go together with above average Mendelian sampling terms and, during subsequent cycles of selection, the product increases until genetic contributions stabilize and $\sum r_{i,\infty} a_i = \Delta G$. Equation 3, therefore, considers the gain arising from a single cohort over all subsequent cycles of selection, whereas Rendel and Robertson's equation considers the gain from selection in a single cycle arising from all previous cohorts. Both are valid, whereas Equation 1, in considering gain from all previous cohorts over all subsequent cycles of selection, results in double counting. Using a modified gene flow procedure, BIJMA and WOOLLIAMS (1998) and WOOLLIAMS *et al.* (1999) show how asymptotic proportions of genes can be predicted accurately, either on an individual or on a group level.

Conclusion

This note has shown that rate of genetic gain differs from the summed product of asymptotic proportions of genes and selection differentials. The connection between asymptotic proportions of genes and rate of genetic gain that is embodied in conventional gene flow theory is invalid, even though genetic gain may be correctly predicted from it. Thus the rate of

genetic gain may be expressed in two manners. First, from conventional gene flow theory, $\Delta G = \Sigma r_{k,0} S_k$, in which case $r_{k,0}$ denotes the proportional contribution of the k^{th} sex-age group to the newborn offspring in the next cohort, as given by Equation 2. Second, $\Delta G = \sum r_{i,\infty} a_i$ where $r_{i,\infty}$ is the true individual asymptotic contribution. Both expressions are valid and give similar results (BIJMA and WOOLLIAMS, 1998). The first expression is based on contributions to the next generation and is valid since it is identical to Rendel and Robertson's result, whereas the second is truly based on asymptotic proportions of genes. Furthermore, conventional gene flow theory can still be used to calculate discounted returns from a single cycle of selection, since differences between $r_{k,0}$ and true asymptotic proportions originate from subsequent cycles of selection and should, therefore, not be attributed to a single cycle of selection.

This research was financially supported by The Netherlands Technology Foundation (STW) and was coordinated by the Earth and Life Sciences Foundation (ALW). JAW gratefully acknowledges the Ministry of Agriculture, Fisheries and Food for financial support.

Literature Cited

BIJMA, P. and J. A. WOOLLIAMS, 1998. Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151: 1197-1210.

HILL, W. G., 1974. Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

HOPKINS, I. R. and J. W. James, 1979. Genetic responses in the early years of selection programmes using genetic differences between generations. Anim. Prod. 28: 65-77.

JAMES J. W., 1977. A note on selection differentials and generation length when generations overlap. Anim. Prod. 24: 109-112.

RENDEL J. M. and A. ROBERTSON, 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J. Genet. 50: 1-8.

WOOLLIAMS J. A., P. BIJMA and B. VILLANUEVA, 1999. Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153: 1009-1020.

Predicting Rates of Inbreeding in Populations undergoing Selection

John A. Woolliams[†] and Piter Bijma

[†]Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K. and [*] Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands

Abstract – Tractable forms of predicting rates of inbreeding (ΔF) in selected populations with general indices, non-random mating and overlapping generations were developed, with the principal results assuming a period of equilibrium in the selection process. An existing theorem concerning the relationship between squared long-term genetic contributions and rates of inbreeding was extended to non-random mating and to overlapping generations. ΔF was shown to be approximately $\frac{1}{4}(1-\omega)$ times the expected sum of squared lifetime contributions, where ω is the deviation from Hardy-Weinberg proportions. This relationship cannot be used for prediction since it is based upon observed quantities. Therefore the relationship was further developed to express ΔF in terms of expected long-term contributions, which are conditional on a set of selective advantages that relate the selection processes in two consecutive generations, and are predictable quantities. With random mating, if selected family sizes are assumed to be independent Poisson variables then the expected long-term contribution could be substituted for the observed providing $\frac{1}{4}$ (since $\omega = 0$) was increased to $\frac{1}{2}$. Established theory was used to provide a correction term to account for deviations from the Poisson assumptions. The equations were successfully applied, using simple linear models, to the problem of predicting ΔF with sib indices in discrete generations since previously published solutions had proved complex.

WRAY and THOMPSON (1990) proved a fundamental relationship between the sum of squared long-term genetic contributions of ancestors and rates of inbreeding for random mating populations in discrete generations. One consequence of this relationship was that rates of inbreeding were tied to the numerator relationship matrix for the first time. This narrowed the conceptual gap between the central parameter for genetic evaluation of individuals using Best Linear Unbiased Prediction and one of the key properties of a breeding scheme. Another important consequence was to set out, in a formal way, a model for the mechanics of inheritance of selected advantage, a concept that ROBERTSON (1961) had introduced but had left unclarified. An achievement of the methods of WRAY and THOMPSON (1990) was to obtain, for the first time, accurate predictions of ΔF in mass selection through modelling pathway extensions. However, this was done by using a recursive algorithm, so that although the

mechanics were clear, the overall structure of the prediction remained obscure.

WOOLLIAMS *et al.* (1993) advanced the understanding of the structure of the prediction by obtaining a closed form for the prediction of ΔF. It was shown to have terms involving variances of family size in one generation, with additional terms for the proliferation or reduction of ancestral lines over many generations that could be predicted as a result of the selective advantage of the ancestor. Furthermore it was clear that under equilibrium conditions, the model would lend itself to geometric summation of terms across generations. This led to simple forms for the expected long-term contribution of an ancestor. WRAY *et al.* (1994) extended the methods to index selection, although the form of the model is a hybrid of the approach of WOOLLIAMS *et al.* (1993) and HILL (1972), since the conditional arguments of pathway extension that had been carried out for mass selection were found to be too complex for index selection. Nevertheless, worthwhile predictions were made available in a tractable form.

SANTIAGO and CABALLERO (1995) used an approach to predict ΔF in mass selection, that made no direct reference to the theory of contributions. They obtained a neater closed form for ΔF than that derived by WOOLLIAMS *et al.* (1993), through an argument based on total drift relating the change through selection to loss of genetic variance. Unlike the previous work of WRAY and THOMPSON (1990) and WOOLLIAMS *et al.* (1993), who had considered the population in relation to an unselected base generation, SANTIAGO and CABALLERO (1995) developed predictions based upon equilibrium genetic variance. NOMURA (1996) extended the approach of SANTIAGO and CABALLERO (1995) to mass selection with overlapping generations, but with the important restriction that the parents selected from a cohort remain the same in both number and identity throughout the breeding life of the cohort.

This paper examines the issues raised by the work described above. First, the relationship between ΔF and the realised long-term genetic contributions is extended to include non-random mating and overlapping generations. Second, an important result for the prediction of ΔF is developed by demonstrating a relationship between ΔF and the expected squared long-term contribution conditional on the selective advantages for random mating. Finally, as an example of application, predictions of ΔF for sib indices, previously considered by WRAY *et al.* (1994), are re-examined using the equilibrium methods for expected long-term contributions developed by WOOLLIAMS *et al.* (1999) and compared to results from simulation.

Relationship between ΔF and Long-term Genetic Contributions

This section discusses the relationship between ΔF and realised long-term genetic contributions. In doing so, it will derive the expected increase in homozygosity at the level of a neutral locus, in contrast to the matrix method of WRAY and THOMPSON (1990). The notation that will be used is shown in Table 1. The model for the population is assumed, for

TABLE1.–Notation used to derive Equations 1 to 27

t, u	Time variables
n_c, q	Number of breeding categories, indexed by q
m, f	Denotes the two sexes in discrete generations, *i.e.*, $q = m$ (male) or f (female)
r_i, $r_{i(q)}$	Observed long-term genetic contribution of individual i (in category q)
$r_{i,u}(q,t)$	The genetic contribution of individual i born at time u to selected parents of sex q born at time t
F_t, ΔF	Inbreeding coefficient at time t, and rate of inbreeding
ω	Deviation from Hardy-Weinberg equilibrium
X_q, X	Number of parents in category q, and a simple monoecious population respectively
$C_u(t)$	Sum of squared contributions for individuals born at time u to selected parents at time t
C	Converged sum of squared contributions, independent of time in an equilibrium
L	Generation interval
$s_{i(q)}$	Set of selective advantages for individual i in category q
$\mu_{i(q)}$	Expected contribution of i in category q conditional upon $s_{i(q)}$
$\sigma_{i(q)}^2$	Variance of contribution of i in category q conditional upon $s_{i(q)}$
n_i	Number of selected offspring of i
$\theta_{n,i}$	Expected number of selected offspring of i conditional upon $s_{i(q)}$
$V_{n,i}$	Variance of the number of selected offspring of i, conditional upon $s_{i(q)}$
$V_{n,dev,i}$	Deviation of $V_{n,i}$ from Poisson, *i.e.*, $V_{n,dev,i} = V_{n,i} - \theta_{n,i}$.
α_q, β_q	Linear model for $\mu_{i(q)} = \alpha_q + \beta_q^T(s_{i(q)} - \bar{s}_q)$

the present, to have discrete generations with X_m male parents and X_f female parents. For the calculation of inbreeding coefficients, every allele is considered as unique in the base population ($t = 0$). It does not matter if the base generation has the structure of an unselected and unrelated population.

Discrete generations: Consider one of these alleles in the base population at a neutral locus (say allele B). Let the gene frequency at time t, in the parents of sex q that have been selected to produce generation $t + 1$, be denoted by $P_B(q,t)$. The gene frequency can be described in terms of genetic contributions similar to Equation (1) of WOOLLIAMS *et al.* (1999). Let A_i be the gene frequency of an allele B in individual i, where $A_i = 1$, ½ or 0 if i is BB, $B\bullet$, or $\bullet\bullet$ respectively (where $\bullet$ represents any other allele). Then the individual gene frequencies can be treated as breeding values for frequency. The average of the gene frequency in the parents of sex q in generation t is given by:

$$P_B(q,t) = \sum_i r_{i,0}(q,t)\, A_{i,0} + \sum_{u=1}^{t} \sum_i r_{i,u}(q,t)\, a_{i,u} \quad (1)$$

where $r_{i,u}(q,t)$ is the genetic contribution of individual i born at time u to the parents of sex q at time t, with breeding value for frequency of allele B given by $A_{i,u}$, and Mendelian sampling terms $a_{i,u} = A_{i,u} - ½(A_{\text{sire}} + A_{\text{dam}})$. Equation (1) separates out the base generation which provides the foundation alleles, and subsequent generations which influence the frequency of the allele through the Mendelian sampling of their parent alleles. The variance of the

Mendelian sampling terms will depend on A_{sire} and A_{dam}; $\mathrm{Var}(a_{i,u}) = 0$ if both A_{sire} and A_{dam} are homozygotes, ⅛ if they are both heterozygotes, or 1/16 otherwise. Since B is unique, $A_{i,0}$ is 0 for all individuals except for one individual for which $A_{i,0} = ½$. The genetic contribution of an individual to the generation of its birth is $r_{i,t}(m,t) = X_m^{-1}$ if i is male, or 0 if i is female, and $r_{i,t}(f,t) = X_f^{-1}$ if i is female, or 0 if i is male.

Initially assume that there is random mating. For any generation the probability of homozygotes for B is obtained from the product of the gene frequencies in the male and in female parents, and is $P_B(m,t)P_B(f,t)$. The inbreeding coefficient F_t for the neutral locus is then the sum over all distinct alleles at the locus.

$$F_t = \sum_{alleles} \sum_{i} r_{i,0}(m,t-1)\, r_{i,0}(f,t-1)\, A_{i,0}^2 + \sum_{alleles} \sum_{u=1}^{t-1} \sum_{i} r_{i,u}(m,t-1)\, r_{i,u}(f,t-1)\, a_{i,u}^2 \quad (2)$$

where $r_{i,u}(q,t\text{-}1)$ is the average contribution to parents of sex q at time t-1. (Note that breeding values and Mendelian sampling terms will depend on the allele, but this dependence has not been made explicit to spare notation). For each allele, the cross-product terms in $A_{i,0}A_{j,0}$ are zero since $A_{i,0} = 0$ except for a single individual. Since the Mendelian sampling terms from different individuals are independent of all other terms for a neutral locus, all cross-products of the Mendelian sampling terms are zero.

More precisely, for each allele and each ancestor, the term $\sum_i r_{i,u}(m,t-1)\, r_{i,u}(f,t-1)\, a_{i,u}^2$ should be the sum of products of contributions of the ancestor to each male and female mating pair:

$$\sum_i \sum_{\text{mates}[j(m),j^*(f)]} r_{i,u}[j(m),t-1]\, r_{i,u}[j^*(f),t-1]\, a_{i,u}^2 \quad (3)$$

This will account for any non-random mating of parents. For a neutral locus, the covariance between r_i and a_i will be 0 (WOOLLIAMS and THOMPSON, 1994; WOOLLIAMS *et al.* 1999), and the expectation of Equation 3 is $\mathrm{E}[\sum_i \sum_{\text{mates}[j(m),j^*(f)]} r_{i,u}(j(m),t-1)\, r_{i,u}(j^*(f),t-1)] \times \mathrm{E}[a_{i,u}^2]$. Let the first of these, the expectation of the cross-products of contributions to mates, be $C_u(t\text{-}1)$. Note: (*i*) $C_{t-1}(t-1) = 0$ since an individual without offspring cannot contribute to both sexes; and (*ii*) the first term in Equation (2) is $½C_0(t\text{-}1)$ since $A_{i,0}^2$ has a value ¼ for each of its 2 alleles, and 0 otherwise.

Assume equilibrium values for (*i*) the deviation from Hardy-Weinberg frequencies arising from non-random mating (ω, equivalent to α_1 of CABALLERO and HILL, 1992), and (*ii*) ΔF, are attained by generation 2 (this assumption will be removed later). Then Equation 2 can be further simplified using results given in APPENDIX A, namely: $\sum_{\text{alleles}} \mathrm{E}[a_{i,u}^2] = ¼$ for $u = 1$ and $¼(1-\omega)(1-\Delta F)^{u-1}$ for $u \geq 2$. Therefore:

$$\mathrm{E}[F_{t+1}] = ½C_0(t) + ¼C_1(t) + \sum_{u=2}^{t-1} ¼(1-\omega)(1-\Delta F)^{u-1} C_u(t) \quad (4)$$

$$E[F_t] = \tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t-1) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} C_u(t-1) \quad (5)$$

Subtracting Equation 5 from 4 and re-arranging terms:

$$E[F_{t+1} - F_t] = \tfrac{1}{2}C_0(t) - \tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t) - \tfrac{1}{4}C_1(t-1) + \tfrac{1}{4}(1-\omega)(1-\Delta F)C_2(t) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} \times [(1-\Delta F)C_{u+1}(t) - C_u(t-1)] \quad (6)$$

Assuming equilibrium, then a steady state of pedigree development will occur and the expectation of the cross-products will be determined by the number of generations over which they have developed, *i.e.*, $C_u(t) = C_{u-1}(t-1)$, since both terms represent contributions $t-u$ generations after the birth of the ancestor. This is not a strong assumption in the context of the problem, since in the absence of an equilibrium there would be no single ΔF to predict.

Therefore, the terms in $C_u(t)$ can be modified to terms in $C_{u-1}(t-1)$, and each term of the sum within the square brackets of Equation 6 can be reduced to $-\Delta F\, C_u(t-1)$. After repeating this process for the $C_2(t)$ term [and temporarily neglecting the term in $\omega \Delta F C_1(t-1)$]:

$$E[F_{t+1} - F_t] = \tfrac{1}{2}C_0(t) - \tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t) - \tfrac{1}{4}\omega C_1(t-1) - \Delta F[\tfrac{1}{4}C_1(t-1) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} C_u(t-1)] \quad (7)$$

For large enough t, the terms in $C_u(t)$ will converge for a given u. Therefore $\tfrac{1}{2}C_0(t) \approx \tfrac{1}{2}C_0(t-1)$, and $\tfrac{1}{4}C_1(t) - \tfrac{1}{4}\omega C_1(t-1) \approx \tfrac{1}{4}(1-\omega)C_1(t-1)$. Then, by adding and subtracting the term $\tfrac{1}{2}\Delta F\, C_0(t)$,

$$E[F_{t+1} - F_t] = \tfrac{1}{2}\Delta F\, C_0(t-1) + \tfrac{1}{4}(1-\omega)C_1(t-1) - \Delta F[\tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t-1) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} C_u(t-1)] \quad (8)$$

Finally, note $E[F_{t+1} - F_t] = \Delta F\, E[1 - F_t]$, and that the term in square brackets in Equation 8 is $E[F_t]$, giving:

$$\Delta F = \tfrac{1}{4}(1-\omega)C_1(t-1)[1 - \tfrac{1}{2}C_0(t)]^{-1} \quad (9)$$

This result holds for t large enough for contributions from early generations to have converged. If it is assumed that the base generation used for defining the inbreeding coefficients was chosen to be part of a period of equilibrium, then $C_1(t\text{-}1) = C_0(t) = C$, and

$$\Delta F = \tfrac{1}{4}(1-\omega)C[1 - \tfrac{1}{2}C]^{-1} \approx \tfrac{1}{4}(1-\omega)C \quad (10)$$

where C is the sum of squared converged contributions for a generation, chosen arbitrarily within the period of equilibrium. Including the term neglected between Equations 6 and 7 would replace $[1 - \tfrac{1}{2}C]^{-1}$ by $[1 - (\tfrac{1}{2} + \tfrac{1}{4}\omega)C]^{-1}$. For random mating, omitting the term $[1 - \tfrac{1}{2}C]^{-1}$ leads to an underestimate with a fractional error of approximately $\tfrac{1}{2}C$, which in turn is approximately $2\Delta F$.

Since

$$C = \mathrm{E}[\sum_i \sum_{\text{mates}[j(m),j^*(f)]} r_{i,u}(j(m),t)\ r_{i,u}(j^*(f),t)]$$

for large $u<<t$, for any i the terms $r_{i,u}[j(m),t]$ and $r_{i,u}[j(f),t]$ converge to the same value for all j in generation t, providing the population mixes. This value will be the long-term contribution of ancestor i to the population, denoted by r_i. This will occur with or without random mating. Thus $C = \mathrm{E}[\sum_i r_i^2]$ for a generation of ancestors, which leads to:

$$\Delta F \approx \tfrac{1}{4}(1-\omega)\,\mathrm{E}[\sum_i r_i^2] \qquad (11)$$

$$\mathrm{E}[\Delta F] \approx \tfrac{1}{4}(1-\omega)(X_m \mathrm{E}[r_{i(m)}^2] + X_f \mathrm{E}[r_{i(f)}^2]) \qquad (12)$$

In Equation 12, the expectations are conditional on the individual i being a selected ancestor. However, since r_i=0 for an unselected ancestor, Equation 12 can also be given as:

$$\mathrm{E}[\Delta F] \approx \tfrac{1}{4}(1-\omega)(T_m \mathrm{E}[r_{i(m)}^2] + T_f \mathrm{E}[r_{i(f)}^2]) \qquad (13)$$

where T_m and T_f are the number of candidates for selection in each sex and the expectation is for a candidate (*i.e.*, it is not conditional on i being selected). ($\mathrm{E}[\Delta F]$ is used in Equations 12 and 13, rather than simply ΔF, to emphasise that the result is an expectation over replicate populations.)

This result was obtained for $\omega = 0$ by WRAY and THOMPSON (1990), but the derivation differs in several aspects. First, in the derivation of WRAY and THOMPSON (1990) the base was unselected and therefore not in equilibrium at the start of the selection process, which led to the impression that the contributions used for estimating rates of inbreeding must be the generation after an unselected base. It is now evident that the choice of generation on which the estimate is obtained is arbitrary, except that it is at the start of some period of local equilibrium during which some 'equilibrium ΔF' may exist. Second, the derivation using the probability of homozygosity for an assumed allele is of value, since the proof of WRAY and THOMPSON (1990) is heavily based upon the properties of the numerator relationship matrix. Third, it extends the result to incorporate non-random mating, although the result was given without proof by WOOLLIAMS and THOMPSON (1994). CABALLERO and HILL (1992) noted that the result of WRAY and THOMPSON (1990) was a poor predictor of ΔF with non-random mating and it is now clear why this is so.

Even though the development of the pedigree may be in equilibrium (which will imply the genetic variance being selected upon is in equilibrium), this does not imply that equilibrium values of ω and ΔF for the alleles defined in the arbitrary base are immediately attained. Equation (4), using APPENDIX A, assumes that these parameters were in equilibrium for the Mendelian sampling in generation 2. However, the following argument shows that this does not affect the result. Assume the equilibrium conditions have not been attained by generation 2, then for this generation plus a small number of generations following (*i.e.*, up to attainment of equilibrium) there will be terms of the form $\delta C_u(t)$ in Equation 4 and $\delta C_u(t\text{-}1)$ in Equation 5. Providing t is sufficiently large compared to the period of attainment, these terms will cancel in

Equation 6 since $C_u(t)$ is a convergent series. Thus Equations 10 to 13 will hold for the equilibrium values of ω and ΔF.

Overlapping generations: If ΔF is taken per unit time then the structure of the preceding proof holds. The reduction in the variance of the Mendelian sampling term over initial cohorts, before an equilibrium ΔF per unit of time is established, is not straightforward since it will depend upon the age structure of the population but the previous argument used to overcome deviations from equilibrium can be applied. However, one distinction in overlapping generations is that the base generation will contain the equivalent of L cohorts, where L is the period of time over which the long-term contributions sum to one, since this is the period required for the population to turn over a generation for those genes destined to remain in the population in the long-term. WOOLLIAMS *et al.* (1999) show this genetic generation interval is different from the average age of the parents when there are selective advantages between groups (see also BIJMA and WOOLLIAMS, 1999). In order to balance Equation 8, there is a need to add and subtract terms of magnitude $\frac{1}{2}C_0(t)(\Delta F$/ generation) or equivalently $\frac{1}{2}C_0(t)L(\Delta F$/ unit time), where L is the generation interval. Thus the error term in Equation 10 is $[1-\frac{1}{2}CL]^{-1}$, and consequently ignoring this term results in an underestimate with a fractional error of $2 \times (\Delta F$ per generation). Equation 11 is obtained by summing over all individuals born in a single cohort. With overlapping generations, individual ancestors within cohorts will have different life-histories, since they will be used at different breeding ages or for different purposes. If X_q is the number of individuals with a lifetime breeding profile categorised by q, then the approximation will be:

$$\mathrm{E}[\Delta F/\text{unit time}] \approx \tfrac{1}{4}(1-\omega) \sum_{\text{categories } q} X_q \mathrm{E}[r^2_{i(q)}] \qquad (14)$$

where the expectations are over the squared contributions from a single cohort, and are conditional on selection in category q. Although the approach is different, Equation 14 is equivalent to the result of HILL (1972, 1979) when random selection and random mating is assumed. However, Equation 14 clearly shows that the rate of inbreeding is related to the sum of squared lifetime contributions irrespective of selection and non-random mating.

Relationship between ΔF and Expected Contributions

Since ΔF is proportional to $\mathrm{E}[r_i^2]$, the task of predicting ΔF in selected populations would be made easier if tractable and general methods for calculating expected squared contributions were available. However, $\mathrm{E}[r_i^2] = \mu_i^2 + \sigma_i^2$, and consequently there is a need to predict both the mean and variance of the contributions. Commonly, prediction of means is a simpler task than prediction of variances. General methods for predicting expected long-term contributions in selected populations have been developed by WOOLLIAMS *et al.* (1999). The objective of the following section is to obtain a relationship between the variance of the long-term contributions and their expectations, which

will then permit development of general methods for the prediction of $E[r_i^2]$, and consequently for ΔF. The relationship will need to assume random mating and is developed by conditioning on the selective advantage(s), s_i, for an ancestor. The selective advantage(s) of the ancestor, if inherited, will partly determine the breeding success of its descendants, with diminishing impact over generations. The proof uses the result $E[r_i^2] = E_s[r_i^2 | s_i] = E_s[\mu_i^2] + E_s[\sigma_i^2]$, where $\mu_i = E[r_i | s_i]$ and $\sigma_i^2 = \mathrm{Var}[r_i | s_i]$, and the subscript s on the E indicates that the expectation is being taken over the selective advantages.

Monoecious population: The proof is simplest in the case of a monoecious diploid population of X parents in discrete generations without selfing. Random mating is assumed ($\omega = 0$). Extension to overlapping generations and to two sexes follows by analogy, but is complicated by the need for matrices, and so this extension is made in APPENDIX B. The long-term contribution of individual i is given by:

$$r_i = \tfrac{1}{2} \sum_{\text{offspring } j} r_j \qquad (15)$$

These sums may be restricted to the selected offspring, since unselected offspring have no long-term contribution. It will be assumed that conditional on the selective advantage s_i of the parent i, the genetic contribution of the offspring is independent of the number of offspring selected from parent i (denote this number by n_i). Then from Equation (15),

$$E[r_i | s_i, n_i] = \tfrac{1}{2} n_i E[r_j | s_i, j \text{ offspring of } i] \qquad (16)$$

$$\mathrm{Var}[r_i | s_i, n_i] = \tfrac{1}{4} n_i \mathrm{Var}[r_j | s_i, j \text{ offspring of } i] \qquad (17)$$

Equation (17) requires random mating. Let $\theta_{n,i}$ and $V_{n,i}$ be the mean and variance of $n_i | s_i$, then:

$$\mu_i = \tfrac{1}{2} \theta_{n,i} E[r_j | s_i, j \text{ offspring of } i] \qquad (18)$$

The derivation of μ_i in a general genetic framework was described by WOOLLIAMS *et al.* (1999).

The variance σ_i^2 is derived using the statistical result that the unconditional variance is the expectation of the conditional variance plus the variance of the conditional expectation. Applying this result to Equations 16 and 17 gives:

$$\sigma_i^2 = \tfrac{1}{4} \theta_{n,i} \mathrm{Var}[r_j | s_i, j \text{ offspring of } i] + \tfrac{1}{4} V_{n,i} E[r_j | s_i, j \text{ offspring of } i]^2 \qquad (19)$$

Assume now that the number selected from parent i has a Poisson distribution. For example, this would be the case if litter size before selection had a Poisson distribution. Then $\theta_{n,i}$ can replace $V_{n,i}$ in the second term of Equation 19, to obtain:

$$\sigma_i^2 = \tfrac{1}{4} \theta_{n,i} (\mathrm{Var}[r_j | s_i, j \text{ offspring of } i] + E[r_j | s_i, j \text{ offspring of } i]^2) \qquad (20)$$

which can be recognised as

$$\sigma_i^2 = \tfrac{1}{4} \theta_{n,i} E[r_j^2 | s_i, j \text{ offspring of } i] \qquad (21)$$

If expectations are now taken over s_i, APPENDIX B shows that, by assuming an equilibrium, there is no covariance between $\theta_{n,i}$ and $E[r_j^2 | s_i, j \text{ offspring of } i]$. A heuristic explanation is that, if there were a covariance, then this would result in selection for increased

squared contributions, breaking the assumption of equilibrium. Taking expectations over $\mathbf{s}_i$, the right hand side of Equation 21 is equal to $\frac{1}{2}\mathrm{E}_s[r_i^2 \mid \mathbf{s}_i]$, since $\mathrm{E}_s[\theta_{n,i}] = 2$. Therefore,

$$\begin{aligned}\mathrm{E}_s[\sigma_i^2] &= \tfrac{1}{2}\mathrm{E}_s[r_i^2 \mid \mathbf{s}_i] \\ &= \tfrac{1}{2}\mathrm{E}_s[\mu_i^2] + \tfrac{1}{2}\mathrm{E}_s[\sigma_i^2] \qquad (22)\end{aligned}$$

which leads to the result that:

$$\mathrm{E}_s[\sigma_i^2] = \mathrm{E}_s[\mu_i^2] \qquad (23)$$

Finally, if X is the number of parents in each generation, then:

$$\begin{aligned}\mathrm{E}[\Delta F] &\approx \tfrac{1}{4} X\, \mathrm{E}_s[r_i^2] \\ &= \tfrac{1}{4} X\, (\mathrm{E}_s[\mu_i^2] + \mathrm{E}_s[\sigma_i^2]) \\ &= \tfrac{1}{2} X \mathrm{E}_s[\mu_i^2] \qquad (24)\end{aligned}$$

The power of this result is that it requires only the mean conditional on the selective advantages to be modelled, which can be done for a wide class of genetic structures using the methods of WOOLLIAMS *et al.* (1999). Note that the set of selective advantages used for conditioning must completely describe the inter-relationship between one generation of selection and the next. This is embodied in the assumption that conditioning on the selective advantage $\mathbf{s}_i$ removes associations between the number of offspring selected and the subsequent success of the offspring. For example, the mates of the individual provide a selective advantage that must be accounted for (WOOLLIAMS and THOMPSON, 1994; SANTIAGO and CABALLERO, 1995).

Corrections for deviation from Poisson: One of the critical assumptions of the proof leading to Equation 24 is that the selected family sizes are distributed as a Poisson variable. However departures from this will occur, for example: (*i*) when the litter sizes are not Poisson; (*ii*) when negative covariances between full-sibs and between half-sibs are induced by using sib indices for selection; (*iii*) when selection intensity becomes large; (*iv*) when there are common environmental variances associated with litters. (The occurrence of last two causes will depend on the model chosen for $\mathbf{s}_i$, which is addressed in DISCUSSION.)

To account for this deviation, let $V_{n,i} = \theta_{n,i} + V_{n,dev,i}$ in Equation 19, where $V_{n,dev,i}$ may be positive or negative according to the circumstances. Then the component in $\theta_{n,i}$ can be treated as previously, and Equation 21 becomes:

$$\begin{aligned}\sigma_i^2 &= \tfrac{1}{4}\theta_{n,i}\,\mathrm{E}[r_j^2 \mid \mathbf{s}_i, j \text{ offspring of } i] \\ &\quad + \tfrac{1}{4} V_{n,dev,i}\,\mathrm{E}[r_j \mid \mathbf{s}_i, j \text{ offspring of } i]^2 \qquad (25)\end{aligned}$$

and Equation 23 becomes:

$$\begin{aligned}\mathrm{E}_s[\sigma_i^2] &= \mathrm{E}_s[\mu_i^2] \\ &\quad + \tfrac{1}{2}\mathrm{E}_s\{V_{n,dev,i}\,\mathrm{E}[r_j \mid \mathbf{s}_i, j \text{ offspring of } i]^2\} \qquad (26)\end{aligned}$$

with the result:

$$\begin{aligned}\Delta F &= \tfrac{1}{2} X\, \mathrm{E}_s[\mu_i^2] + \\ &\quad \tfrac{1}{8} X\, \mathrm{E}_s\{V_{n,dev,i}\,\mathrm{E}[r_j \mid \mathbf{s}_i, j \text{ offspring of } i]^2\} \qquad (27)\end{aligned}$$

Anticipating an observed result, the terms involving $\mathbf{s}_i$ in $\mathrm{E}[r_j \mid \mathbf{s}_i, j \text{ offspring of } i]$ contribute little to the second term of Equation 27 and only the constant term, independent of $\mathbf{s}_i$, will be considered. In the current context, $\mathrm{E}[r_j \mid \mathbf{s}_i, j \text{ offspring of } i] \approx X^{-1}$ and the second term in Equation 27 becomes $\frac{1}{8}\mathrm{E}_s[V_{n,dev,i}]/X$.

For example, in mass selection with fixed litter sizes SANTIAGO and CABALLERO (1995) used the approximation that $E_s[V_{n,dev,i}] \approx -n_o^{-1}$, where n_o is the number of offspring per parent, with the result that the correction for the deviation from Poisson is $(-8T)^{-1}$ where T is the total number of individuals born.

One of the benefits of Equation 24 is that the rate of inbreeding can be obtained from predicting means, often using regression techniques. Accounting for deviations from the Poisson distribution introduces the need for estimating variances of family size to obtain Equation 27. Nevertheless, the multi-generational problem of estimating the variance of a long-term genetic contribution has been reduced to estimating the variance of family size after selection in a single generation.

Extension to overlapping generations: With overlapping generations, individuals within a cohort that are selected to breed at any point in their lifetime can be divided into breeding categories. These categories are defined by the age of breeding, and by how often and for what purpose the individual breeds. Categories are particularly important in selection. As an example, consider mass selection where all selected individuals can have progeny born at ages one, two or three. If the population is making genetic progress, the average merit of individuals born 3 years ago is less than the average merit of an individual born 1 year ago. Thus an offspring of a 3-year-old parent will have a selective disadvantage compared to an offspring of a 1-year-old parent, and so is expected to make a smaller genetic contribution in the long-term (see BIJMA and WOOLLIAMS, 1999). If an individual is a parent at all ages, then its genetic contribution is expected to be greater than an individual chosen for breeding only at a single age. Breeding purpose is also important; if one group of parents are given more mating opportunities, then these would be expected to have more offspring and, other factors being equal, ultimately a greater long-term genetic contribution.

For these reasons, partitioning of the selected individuals into categories is necessary to obtain the general result. It is assumed that the categories are defined so that an individual belongs to a single category that describes its lifetime genetic contribution. To continue the example of mass selection, where the only distinction among parents is the breeding age, there would be potentially 7 categories. If $\{x\}$ denote age x at breeding, then these categories are $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$). The number of categories will inevitably depend on the complexity of the breeding scheme, but the essential point is that they can be defined and enumerated. Let n_c be the number of categories indexed from $q = 1 \ldots n_c$, and $\mu_{i(q)}$ be the expected long-term contribution of individual i in category q conditional on its selective advantage $s_{i(q)}$, with variance $\sigma_{i(q)}^2$. The steps given above in Equations 16 to 27 for a single category remain the same, but changes are needed since terms must be re-defined as vectors and matrices. The notation to develop the argument becomes more complex, therefore, but the result remains simple. For this reason the proof is given in APPENDIX B. The conclusion is that if family sizes after selection

are assumed to be distributed as independent Poisson variables, then:

$$E[\Delta F] = \tfrac{1}{2} \sum_q X_q E_s[\mu^2_{i(q)}] \qquad (28)$$

This simple result shows that the rate of inbreeding, when approximated by the sum of squared contributions, is equal to one half of the sum of the squares of expected lifetime contributions. Instead of using the observed contribution, as in Equation 12, the expected contribution can be substituted provided that the coefficient is changed from ¼ to ½.

Correction for deviations from Poisson: As previously, for a parent from category q, define the matrix $\mathbf{V}_{n(q),\text{dev}}$ of size n_c x n_c to be the (co)variance matrix for the number of selected offspring in each of the n_c categories, expressed as deviations from independent Poisson variances. For each q, neglecting terms in s_i (for empirical reasons given earlier), there will be a term δ_q defined by $\boldsymbol{\alpha}^T \mathbf{V}_{n(q),\text{dev}} \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the vector with qth element equal to the expected long-term contribution for an individual from category q, *i.e.*, $E_s[\mu_{i(q)}] = \alpha_q$. Note that δ_q may be negative, since it is a variance deviation and is not a variance. This term is introduced in Equation B6 of APPENDIX B. From APPENDIX B, we arrive at:

$$E[\Delta F] = \tfrac{1}{2} \sum_q X_q E_s[\mu^2_{i(q)}] + \tfrac{1}{8} \sum_q X_q \delta_q \qquad (29)$$

Although the proof has been based upon a monoecious diploid organism with no selfing, the extension to a dioecious organism is clear from the proof for overlapping generations. Discrete generations with two sexes is identical to having two categories, *i.e.*, males and females. Finally note that, other than assuming an equilibrium and random mating, there have been no assumptions on the type of selection index used, the nature of the genetic variation or the population structure.

Application and Results

Random selection: With random selection and mating in discrete generations and Poisson litter sizes, $\mu_{i(m)}=1/(2X_m)$ and $\mu_{i(f)}=1/(2X_f)$. Application of Equation 28 gives $\Delta F = 1/(8X_m) + 1/(8X_f)$, the formula of WRIGHT (1931). For fixed numbers of offspring per litter (n per sex, giving a total T offspring of each sex), the correction for deviation from Poisson must be included. The true distribution of number selected is hypergeometric, but the variance can be approximated by a binomial for n small compared to T (See appendix C of CHAPTER 7 for the hypergeometric solution). The deviation matrix will be $\mathbf{V}_{n(q),\text{dev}} = (-np_m^2, 0 | 0, -np_f^2)$ for both males ($q = m$) and females ($q = f$), where $p_m = X_m T^{-1}$ and $p_f = X_f T^{-1}$. Application of Equation 29 gives the correction $-1/(8T)$, which is a well-established result.

Sib indices in discrete generations: The theory will be illustrated by selection on a general sib index of the form $I = b_1(P - \overline{P}_{fs}) + b_2(\overline{P}_{fs} - \overline{P}_{hs}) + b_3\overline{P}_{hs}$, where P is the phenotype of the candidate, $\overline{P}_{fs}$ is the phenotypic mean of its full-sibs (including candidate) and $\overline{P}_{hs}$ is the phenotypic mean of its half- sibs (including candidate and full-sibs). Mass selection is a special case, with $b_1 = b_2 = b_3 = 1$ (or any constant >0). This formulation

TABLE 2.–Genetic parameters for a population selected with a sib index.

X_m, X_f, d	Number of male and female parents, and mating ratio $d=X_f/X_m$
n_m, n_f, n_o	Number of male and female offspring in a full-sib family, $n_o = n_m + n_f$
P, $\overline{P}_{fs}$, $\overline{P}_{hs}$	Phenotype of candidate, and its full- and half-sib family means
I, b_1, b_2, b_3	Index and weights for selection $I = b_1(P - \overline{P}_{fs}) + b_2(\overline{P}_{fs} - \overline{P}_{hs}) + b_3\ \overline{P}_{hs}$
p_m, p_f	Selected proportions for males and females: $p_m = (n_m d)^{-1}$, $p_f = n_f^{-1}$
v_m, v_f, i_m, i_f, k_m, k_f	Truncation points, intensities of selection and variance reduction coefficients.
σ_I^2, σ_A^2, ρ	Variance of the index, total genetic variance and accuracy of selection.
σ_{Am}^2, σ_{Af}^2, σ_e^2	Genetic variance among selected sires and dams, and residual variance.
$A_{i(hs)}$, $A_{j(fs)}$	Mean breeding value of the half-sib family of sire i, and the mean breeding value of the full-sib family of dam j expressed as a deviation from the half-sib family.
$v(A_{i(hs)})$	Variance of $A_{i(hs)}$, and similarly defined for $A_{i(fs)}$: $v(A_{i(hs)}) = \sigma_{Am}^2 + \sigma_{Af}^2/d$, $v(A_{i(fs)}) = \sigma_{Af}^2(1 - d^{-1})$
ρ_{HS}, ρ_{FS}	Correlation of indices among full-sibs and half-sibs respectively: $\rho_{HS} = [b_3^2 v(A_{i(hs)}) - b_2^2 v(A_{i(fs)})(d-1)^{-1}]/\sigma_I^2$ $\rho_{FS} = [b_3 v(A_{i(hs)}) + b_2 v(A_{i(fs)}) - b_1^2(\tfrac{1}{2}h_0^2 + \sigma_e^2)n_f^{-1}]/\sigma_I^2$

was used also by WRAY *et al.* (1994) in their study of rates of inbreeding. Every generation, the highest ranking X_m sires and X_f dams are selected as parents for the next generation. Each sire is mated at random to $d = X_f/X_m$ dams and each dam produces a total of n_o offspring, n_m male and n_f female, which are available for selection in the next generation. The unselected base population is assumed to have a phenotypic variance of 1 with a heritability of h_0^2 for the selected trait. Additional notation used for the sib index is shown in Table 2. An example is given at each step, for a selection scheme with $X_m = 20$, $X_f = 60$, $n_m = n_f = 4$, and index weights $b_1 = 1$, $b_2 = 1.5$, and $b_3 = 2$. The principal parameters for this scheme are presented in Table 3 for easy reference.

In WRAY *et al.* (1994) the selective advantages were based on the breeding values $A_{i(x)}$, and this approach will be adopted here, but slightly modified. A sire i has one selective advantage, namely its own breeding value plus the average breeding value of its d mates (*i.e.*, its mate group), and this aggregate value will be denoted by $A_{i(hs)}$. A dam i has two selective advantages: first the selective advantage of its mate ($A_{i(hs)}$) and second, its own breeding value expressed as a deviation from the average breeding value of the mate group to which it belongs (denoted $A_{i(fs)}$). The average breeding value of the full-sib family from dam i is $\frac{1}{2}(A_{i(hs)}+A_{i(fs)})$. Thus, in this scheme, $\mathbf{s}_{i(m)} = (A_{i(hs)})$, and $\mathbf{s}_{i(f)} = (A_{i(hs)},\ A_{i(fs)})^T$. The two selective advantages for a dam are independent.

TABLE 3.–Principal parameters, as described in Table 2, for the example selection scheme used throughout

$X_m = 20$	$X_f = 60$	$d = 3$	$n_m = n_f = 4$	$h_0^2 = 0.4$	$(b_1,b_2,b_3) = (1.0,1.5,2.0)$
$p_m = 0.083$	$p_f = 0.25$	$v_m = 1.383$	$v_f = 0.674$	$i_m = 1.839$	$i_f = 1.271$
$k_m = 0.838$	$k_f = 0.759$	$\sigma_I^2 = 1.331$	$\sigma_A^2 = 0.302$	$\sigma_{Am}^2 = 0.050$	$\sigma_{Af}^2 = 0.052$
$\rho_I = 0.636$	$\rho_{FS} = 0.390$	$\rho_{HS} = 0.205$	$v(A_{i(hs)}) = 0.269$	$v(A_{i(fs)}) = 0.140$	

Expected long-term genetic contributions were modelled following WOOLLIAMS *et al.* (1999) as: $E[r_{i(q)} | \mathbf{s}_{i(q)}] = \mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^T(\mathbf{s}_{i(q)} - \bar{\mathbf{s}}_q)$, where $\mathbf{s}_{i(q)}$ denotes the vector of selective advantages for a selected individual of sex q expressed as a deviation from the mean of its contemporaries $\bar{\mathbf{s}}_q$, $\boldsymbol{\beta}_q$ is the vector of regression coefficients of $r_{i(q)}$ on $\mathbf{s}_{i(q)} - \bar{\mathbf{s}}_q$, α_q is the mean contribution of selected parents of sex q, and ^{T} denotes the transpose. In the parameterisation used, the mean of $A_{i(fs)}$ is always zero. To simplify the notation, it will be assumed that $A_{i(hs)}$ is already expressed as a deviation from the mean of the contemporary group, and so $\bar{\mathbf{s}}_q$ will be omitted from this point onwards.

Step 1. Prediction of expected contributions: The prediction of expected genetic contributions is covered in detail by WOOLLIAMS *et al.* (1999). The current paper will only summarise the procedure for a sib index, without derivation. Prediction of $\mu_{i(q)}$ requires the prediction of $\boldsymbol{\alpha} = (\alpha_m, \alpha_f)^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_m^T, \boldsymbol{\beta}_f^T)$. In discrete generations, $(\alpha_m, \alpha_f) = [1/(2X_m), 1/(2X_f)]$ always. Solutions for $\boldsymbol{\beta}$ are obtained applying the method of WOOLLIAMS *et al.* (1999), using BULMER's (1980) equilibrium genetic variances. A summary of equations used is given in APPENDIX C. For the example, $(\alpha_m, \alpha_f) = (0.0250, 0.0083)$, $\boldsymbol{\beta} = (0.0447, 0.0149, 0.0130)$.

Step 2. Rates of inbreeding assuming Poisson variances: From step 1, $\mu_{i(m)} = [0.0250 + 0.0447]A_{i(hs)}$. The expected squared mean is a simple sum of squared terms: $X_m E[\mu_{i(m)}^2] = X_m [0.0250^2 + 0.0447^2 v(A_{hs})(1 - X_m^{-1})]$. The $(1 - X_m^{-1})$ term accounts for variances about the sample mean of the selected group rather than the true mean.

The terms arising from $X_f E[\mu_{i(f)}^2]$ are calculated analogously. Since the two selected advantages of the females are mutually independent, the expected mean squared is simply the sum of squared terms. The expected long-term contribution of a female parent is:

$$\mu_{i(f)} = [0.0083 + 0.0149 A_{i(hs)} + 0.0130 A_{i(fs)}]$$

and the sum of squared means is: $X_f E[\mu_{i(f)}^2] = X_f[0.0083^2 + 0.0149^2 v(A_{i(hs)})(1 - X_m^{-1}) + 0.0130^2 v(A_{(i)fs})]$. As previously mentioned, $A_{(i)fs}$ is defined as a deviation from the mean over all ancestors, so $v(A_{i(fs)})$ requires no correction.

The rate of inbreeding ignoring deviations from Poisson variances is predicted from: $\Delta F = \frac{1}{2}(X_m E[\mu_{i(m)}^2] + X_f E[\mu_{i(f)}^2]) = (0.0227 + 0.0090)/2 = 0.0158$.

TABLE 4.– Simulated (S) and predicted (P) rates of inbreeding for weights (1.0, 0.75, 0.5) for mating ratio (d) = 3, and (1.0, 0.75, 0.75) for d = 1, with different numbers of sires X_m, total litter size (n_o) and heritability (h^2). Derived from 1000 replicates with standard errors of the means less than 0.0001.

			$X_m = 20$		$X_m = 40$		$X_m = 80$	
d	n_o	h^2	S	P	S	P	S	P
1	4	0.1	0.0092	0.0093	0.0046	0.0047	0.0024	0.0024
		0.4	0.0106	0.0104	0.0052	0.0053	0.0026	0.0027
	8	0.1	0.0119	0.0115	0.0058	0.0059	0.0029	0.0030
		0.4	0.0146	0.0141	0.0071	0.0072	0.0036	0.0036
3	4	0.1	0.0068	0.0069	0.0034	0.0035	0.0017	0.0018
		0.4	0.0076	0.0077	0.0039	0.0039	0.0019	0.0020
	8	0.1	0.0076	0.0078	0.0039	0.0040	0.0019	0.0020
		0.4	0.0093	0.0091	0.0046	0.0045	0.0023	0.0023

Step 3. Correction for deviations of V_n from Poisson variances: Deviations from Poisson variances can be accounted for by correcting the rate of inbreeding using Equation 28, where $\delta_q = \boldsymbol{\alpha}^T \mathbf{V}_{n(q),dev} \boldsymbol{\alpha}$ and $\mathbf{V}_{n(q),dev}$ is the (2×2) matrix with (co)variances of the number of selected offspring of a parent of sex q ($q = m,f$) as a deviation from independent Poisson variances. The calculation of the deviation from Poisson family variance for fixed numbers of selection candidates per full-sib family is described in APPENDIX D. The approach adopted was derived in detail by BURROWS (1984), although extension to two sexes was required and the method was made more flexible by incorporating results from MENDEL and ELSTON (1974). Applying the method to the example gives:

$$\mathbf{V}_{n(m),dev} = (0.186, 0.751 \mid 0.751, -0.079)$$

$$\mathbf{V}_{n(f),dev} = (0.020, 0.159 \mid 0.159, -0.154).$$

The total correction to the predicted ΔF is 0.0016, and the prediction, using Equation 29 is 0.0175. After these adjustments, the mean ΔF derived from 1000 simulations was 0.0183 (*s.e.* 0.0001).

General fit: To test the general fit of the predictions, extensive simulations were carried out with factorial combinations of: X_m = 20, 40, 80; d = 1, 2, 3 (and 5 for X_m = 20, 40); total offspring of 4, 8 and 16 per full-sib family equally divided between sexes; and with h^2 = 0.1, 0.2, 0.4 and 0.6. The weights used were (1.0,0.75,0.5) for d > 1 (changed to (1.0,0.75,0.75) for d = 1) and (1.0,1.5,2.0) for d > 1 (changed to (1.0,1.5,1.5) for d = 1). Classical weights were also examined since these weights were the subject of the study of WRAY *et al.* (1994), although they are sub-optimal for all cases other than the first round of selection from an unselected base population. Results have been summarised in Tables 4, 5 and 6. The accuracy of the subset of data presented was typical of the results as a whole.

With weights (1.0, 0.75, 0.5 or 0.75) the accuracy was excellent for all schemes, with all errors less than 5%, with a general tendency to overestimation (Table 3). With weights (1.0, 1.5, 1.5 or 2.0) accuracy was also very good,

TABLE 5.– Simulated (*S*) and predicted (*P*) rates of inbreeding for weights (1.0, 1.5, 2.0) for mating ratio (d) = 3, and (1.0, 1.5, 1.5) for d = 1, with different numbers of sires X_m, total litter size (n_o) and heritability (h^2). Derived from 1000 replicates with standard errors of the means less than 0.0001.

			$X_m = 20$		$X_m = 40$		$X_m = 80$	
d	n_o	h^2	*S*	*P*	*S*	*P*	*S*	*P*
1	4	0.1	0.0123	0.0123	0.0061	0.0062	0.0030	0.0031
		0.4	0.0135	0.0137	0.0069	0.0070	0.0033	0.0035
	8	0.1	0.0174	0.0177	0.0089	0.0090	0.0046	0.0045
		0.4	0.0215	0.0211	0.0110	0.0107	0.0057	0.0054
3	4	0.1	0.0110	0.0112	0.0055	0.0056	0.0028	0.0029
		0.4	0.0127	0.0126	0.0064	0.0064	0.0033	0.0032
	8	0.1	0.0147	0.0146	0.0076	0.0076	0.0040	0.0038
		0.4	0.0183	0.0174	0.0097	0.0097	0.0049	0.0045

accurately tracking trends with the changes in the parameters (Table 4). The trends in the errors were for overestimation with d = 1, X_m = 20 but with a maximum error of less than 7%, and an underestimate of 8% for d = 3, X_m = 80. The trends in rates of inbreeding were also accurately tracked with classical weights (see Table 5), where several schemes had rates of inbreeding over 0.03. As previously, X_m = 20 tended to lead to overestimates (up to 13% for d = 3, n_o = 4, h^2 = 0.1), although for d = 3, n_o = 8 and h^2 = 0.4, the methods underpredicted the rates of inbreeding by approximately 10% of the observed.

The most serious trend in the errors was a pattern of underprediction characterised by high mating ratio and large family sizes (both of which increase the selection intensity), and increased family weights. More surprisingly, the errors also increased with the numbers of parents at a constant d (*i.e.*, X_m = 20, X_f = 60 compared to X_m = 80, X_f = 240), and also that the errors were not present for h^2 = 0.01 and increased sharply as h^2 increased. To explore these errors further, the long-term contributions for selected males were plotted against $A_{i(hs)}$ for the schemes with d = 3, weights (1.0,1.5,2.0) for: I, X_m = 20, h^2 = 0.4, n_o = 16; II, X_m = 80, h^2 = 0.4, n_o = 16; III, X_m = 80, h^2 = 0.01, n_o = 16; and IV, X_m = 80, h^2 = 0.4 with n_o = 4. The ΔF from simulation (S) and prediction (P) was: I, S 0.0231, P 0.0220; II, S 0.0070, P 0.0058; III, S 0.0028, P 0.0029; IV, S 0.0037, P 0.0037. Note that Scheme II is simply Scheme I with 4 times the number of parents and expected long-term contributions of I are consequently 4 times bigger than II. The prediction of ΔF for Scheme II is close to (but not precisely) ¼ of that for I. However, the ratio of the simulated ΔF for Scheme II compared to I was closer to ⅓, *i.e.*, much greater than would be expected from scaling. Serious prediction errors occurred only for Scheme II.

Figure 1 shows that the accuracy of prediction with low h^2 (Scheme III) is because the linear model used is a good fit (*i.e.*, the contributions are a simple linear regression on the selective advantage), and similarly for low selection intensity (Scheme IV). However, for both the other two schemes, the linear model

TABLE 6.– Simulated (S) and predicted (P) rates of inbreeding for classical weights, with different mating ratios (d), numbers of sires X_m, total litter size (n_o) and heritability (h^2). Derived from 1000 replicates with standard errors of the means less than 0.0001.

			$X_m = 20$		$X_m = 40$		$X_m = 80$	
d	n_o	h^2	S	P	S	P	S	P
1	4	0.1	0.0163	0.0171	0.0081	0.0086	0.0042	0.0044
		0.4	0.0154	0.0164	0.0077	0.0082	0.0038	0.0042
	8	0.1	0.0362	0.0385	0.0192	0.0196	0.0098	0.0099
		0.4	0.0300	0.0305	0.0158	0.0156	0.0079	0.0079
3	4	0.1	0.0157	0.0171	0.0083	0.0088	0.0043	0.0044
		0.4	0.0142	0.0144	0.0073	0.0074	0.0037	0.0037
	8	0.1	0.0315	0.0336	0.0177	0.0171	0.0091	0.0086
		0.4	0.0232	0.0231	0.0125	0.0119	0.0067	0.0060

predicts a substantial proportion of the selected males to have negative contributions, although rates of inbreeding are accurately predicted in one case (Scheme I) but not in the other (Scheme II).

Closer replicate by replicate analysis shows that despite the expectation, the substantially greater variance of contributions (approximately proportional to $\Delta F/X_m$) in Scheme I obscures the non-linearity in the majority of replicates. When both linear and quadratic terms for the selective advantage were included in a regression model for observed contributions, the quadratic term was not statistically significant (defined here as $P < 0.01$) in over 60% of the replicates. In contrast, for Scheme II, this percentage was less than 15%. Thus the accuracy of prediction depends on the goodness-of-fit of the linear model within a replicate, so more parents may promote greater proportional prediction errors, even though these errors will be associated with lower rates of inbreeding.

The pattern of the correction for deviations from Poisson distribution for selected family sizes is worth noting. These corrections are negative for b_2, $b_3 < 1$, reduce in size as the index weights increase and were generally positive for b_2, $b_3 > 1$. For mass selection, $b_1 = b_2 = b_3 = 1$, the correction is of the order of $-1/(8T)$.

Discussion

The theory described in this paper provides a powerful tool for predicting rates of inbreeding in selected populations, and insights into the forces that contribute to the rate of loss of variation. The relationship of WRAY and THOMPSON (1990) has been derived directly from consideration of identity by descent and has been extended to cover overlapping generations and non-random mating. Applicability was then advanced by showing how expected long-term contributions, which are predictable by general methods, can be used in place of observed long-term contributions to predict the rates of inbreeding, if random mating was assumed. Finally, the methods were applied to sib indices in discrete generations,

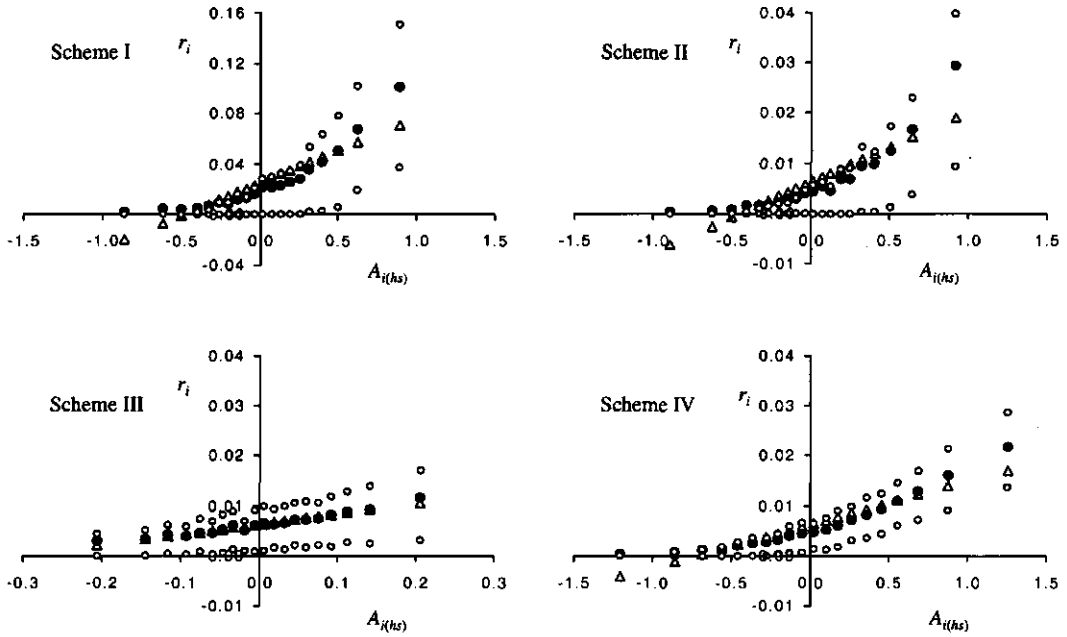

FIGURE 1.–The expected long-term contribution, lower and upper quartiles obtained from simulation as a function of the selective advantage $A_{i(hs)}$, together with the expected long-term contribution predicted assuming a linear model for four example schemes. The curves obtained from simulation are the result of sampling of 8000 individuals. The schemes all have $d = 3$, with weights (1.0, 1.5, 2.0) for: I, $X_m = 20$, $h^2 = 0.4$, $n_o = 16$; II, $X_m = 80$, $h^2 = 0.4$, $n_o = 16$; III, $X_m = 80$, $h^2 = 0.01$, $n_o = 16$; I, $X_m = 80$, $h^2 = 0.4$, $n_o = 4$. (Δ linear prediction; • simulated expectation; ° lower and upper quartiles)

for which the previous solutions were complex (WRAY *et al.* 1994). In doing so, some insight was gained into the origin of the prediction errors, and these appeared to arise from the goodness-of-fit of the models used to implement the theory, rather than those used to derive it.

Theory

The first theorem relating the rate of inbreeding in a population to the squared long-term contributions was previously derived by WRAY and THOMPSON (1990), but the proof here has several useful extensions. In contrast to WRAY and THOMPSON (1990), the proof is direct in using identity by descent rather than properties of the numerator relationship matrix, and it incorporates non-random mating and overlapping generations. The simplest relationship ($\Delta F \approx \frac{1}{4}\sum r_i^2$) is not exact and was shown to underestimate the rate of inbreeding by a fraction of the order of ($2\Delta F$), providing there was no major deviation from random mating and the error is therefore small for any practical scheme. In overlapping generations, with rates of inbreeding per unit time and per generation both of interest, it is shown that this error is $2(\Delta F$/generation), where the generation

interval was defined by the period over which the long-term genetic contributions sum to 1.

The importance of the relationship between rates of inbreeding and squared genetic contributions is that it holds for selected populations, with no assumptions on the form of selection, providing (*i*) the genes are ultimately mixed, and (*ii*) an equilibrium exists over which a stable ΔF may be defined. A further caveat is that the rate obtained applies to a neutral, unlinked gene.

The extension of other relationships to predict ΔF in selected populations does not always hold. For example, using the relationship $\mathrm{Var}(\delta q) = q(1\text{-}q)\Delta F$, where q is the frequency of a neutral gene and δq is the change in frequency per unit time, will not hold if selection is not random, since it assumes mutual independence of δq over consecutive intervals. The increments, δq, are also correlated for overlapping generations due to the many intervals over which the progeny of a single parent may be selected. As a consequence, the justification for the proof by HILL (1979) for ΔF with overlapping generations is invalid, even in the absence of genetic selection, although the result is correct and agrees with the previous proof of HILL (1972). Closer examination of HILL (1979) shows that its justification lies in an intuitive argument for the relationship that was to be proved later by WRAY and THOMPSON (1990). Consequently, the methods derived here may be seen to arise as a natural development of the results of HILL (1972, 1979) for selected populations.

The form of Equation 4 shows that the sum of squared long-term contributions for any given cohort may be usefully interpreted in the absence of an equilibrium. The sum of squared contributions for a cohort is the proportion of the new variation (the Mendelian sampling variance) arising from within that cohort, that is lost to the population in the long-term. This will include all mutational variance arising in prior generations, since the choice of base is arbitrary. Therefore, the sum of squared contributions of cohorts (particularly those still to converge) are important, irrespective of equilibrium, and provide a meaningful measure of risk, and merit attention in both breeding and conservation schemes. The operational tools described by GRUNDY *et al.* (1998) are based upon controlling sums of squared contributions of cohorts, and have meaning and validity beyond the infinitesimal model (*e.g.*, VILLANUEVA *et al.* 1999). However, there are clearly greater problems in providing deterministic predictive tools to analyse population dynamics if the assumption of equilibrium is removed, and those provided by WOOLLIAMS *et al.* (1999) assume this equilibrium.

The second novel theorem derived in this paper is concerned with showing how the formulae with observed long-term contributions may be translated into formulae with expected long-term contributions. The latter are advantageous since they use predictable entities. The major change is that the expected can be substituted for the observed providing the constant of proportionality is increased from ¼ to ½. The critical step in the proof is that the error variance of a long-term contribution given the selective advantage is related to the square

of its mean, *i.e.*, the coefficient of variation is relatively constant. Apart from random mating, the scope of this proof is very broad and is applicable to overlapping generations. The validity of the derivation was checked using general sib-indices as an example in discrete generations, and a companion paper (BIJMA *et al.*, 2000) provides verification in overlapping generations with mass selection and lifetime selection, so removing a serious restriction of NOMURA (1996). The limitation to random mating arises from Equation (17), although in the special case of partial full-sib mating with no selection, the analysis can be completed (using results of GHAI, 1975) and agrees with results of CABALLERO and HILL (1992) (see CHAPTER 11). This provides an indirect verification of Equation (13) for non-random mating.

WOOLLIAMS *et al.* (1999) show how the expected long-term contribution may be calculated in general for different inheritance models (*e.g.*, imprinted variation, maternal additive, sex-linked variation) and with different selection indices (sib indices, Best Linear Unbiased Predictors). Using long-term contributions follows the path of WRAY and THOMPSON (1990) and WOOLLIAMS *et al.* (1993), and differs from SANTIAGO and CABALLERO (1995; mass selection in discrete generations) and NOMURA (1996; a special case of mass selection with overlapping generations) who base their predictions on genetic variation transmitted to descendants. This is because the approach using genetic variation cannot be sustained for general selection schemes. SANTIAGO and CABALLERO (1995) suggest (their Equation 13) that a change in covariance between a general selective advantage and a neutral gene following selection is determined by the reduction in genetic variation. This is true for mass selection, where the index of selection is solely a function of the total breeding value and residual error, but will not be true in general (WOOLLIAMS *et al.* 1999). BIJMA *et al.* (2000) show why there is agreement between the two approaches for mass selection in discrete generations, and also why the current methods are required to cope with overlapping generations.

Prediction

Usable predictions were obtained by WRAY *et al.* (1994), and VILLANUEVA and WOOLLIAMS (1996) used an alternative form based upon Wray et al. (1994). However the method of WRAY *et al.* (1994) was complicated, although it attempted to model the expected proliferation of ancestral lines. The authors believe the proposed method is conceptually simpler than that of WRAY *et al.* (1994), and is open to development.

In any attempt to obtain prediction formulae, a balance has to be achieved between accuracy and simplicity. We have used simple linear models to interpret the theory. Thus in application the prediction consists of two elements: (*i*) the squared expected contribution; (*ii*) the deviation from independent Poisson families. The first of these elements was applied precisely as described by WOOLLIAMS *et al.* (1999), with corrections for finite numbers only being used to obtain the sample variance of selective advantages. No other modifications

were needed, because the other terms in the squared expected contribution were estimates of regression coefficients, which were assumed to be relatively robust to finite sampling. This assumption may be justified in part by the excellent agreement obtained by WOOLLIAMS *et al.* (1999) and BIJMA and WOOLLIAMS (1999) between simulations and deterministic predictions of expected long-term contributions. The second element, calculating the deviation from independent Poisson families, only required extension of the method of BURROWS (1984) to two sexes.

The choice of selective advantages has as an objective the minimum number needed to make the selective processes in different time periods independent. Using sib indices as an example, the authors considered both the method presented, where only breeding values were included as selective advantages, and an alternative definition in which the selective advantages were the half-sib mean and deviation of the full-sib mean from the half-sib mean. The potential benefit from the alternative parameterisation is that the environmental covariances in the index arising from the sib means are accounted for within the expected long-term contribution. Conditioning on the sib means is more than is strictly necessary for conditional independence between generations. However, whilst results using the alternative parameterisation were as accurate in most cases (results not shown), the underestimates explored in the results tended to be more severe. One reason for this is that terms included in the expected long-term contribution are modelled by linear functions, whereas modelling the environmental correlations by the method of BURROWS (1984) allows part of the non-linearity to be accounted for.

Non-linear relationships between the selective advantage and long-term contributions occurred when high selection intensities were combined with moderate heritabilities, large numbers of parents and high mating ratios. Results from including quadratic terms in the model for the expected long-term contribution (unpublished) confirm that the serious prediction errors arise from the assumption of linearity, rather than from Equation 29.

There are good reasons to believe that these departures from linearity should not prove a major problem where the objective is to design effective breeding schemes. First, on pragmatic grounds, the curvilinear relationship shown in Figure 1 suggests that 15% of selected males were being used with no expectation of long-term contribution to the population (this percentage is even higher if the contributions were plotted against the observed half-sib mean). The resources used to keep and breed these animals are clearly wasted. In an ideal selection scheme, an ancestor's long-term contributions will be zero or, once its Mendelian sampling term is above a critical threshold, linearly related to the Mendelian sampling term (WOOLLIAMS and THOMPSON, 1994; GRUNDY *et al.* 1998). Consequently, it would be expected that, in an ideal scheme, the long-term contribution of a selected ancestor will show an approximately linear relation with its breeding value. This argument suggests that, if the design objective is for a scheme to generate gain efficiently from the resources

available, a linear model for the relationship between the long-term contribution and the selective advantage should prove sufficient. If so, then the need for improved deterministic models to cater for the schemes with large prediction errors would be removed. The viewpoint that the schemes with large prediction errors were inefficient is supported by the results of VILLANUEVA and WOOLLIAMS (1997), who showed that when using sib indices, efficient schemes had $d \leq 2$ for which the methods presented here had a good fit.

In conclusion, this paper has: (*i*) established a broader theorem (compared to WRAY and THOMPSON, 1990) concerning the relationship between squared long-term genetic contributions and rates of inbreeding, in particular extending the theorem to non-random mating and to overlapping generations; (*ii*) shown that, for random mating, the relationship can be generalised from long-term contributions that are simply observed to encompass expected long-term contributions which can be predicted; and (*iii*) shown how these equations might be interpreted with simple linear models in the context of predicting rates of inbreeding with sib indices in discrete generations. Together with the findings of WOOLLIAMS *et al.* (1999), the findings of this study show how rates of inbreeding may be predicted in general populations with complex structures and genetic models.

JAW gratefully acknowledges financial support from the Ministry of Agriculture Fisheries and Food (UK), and the support and encouragement of Prof. A. Maki-Tanila who gave an opportunity for this work to be initiated. The contribution of PB was financially supported by the Netherlands Technology Foundation (STW) and co-ordinated by the Earth and Life Science Foundation (ALW).

Literature Cited

BIJMA, P. and J. A. WOOLLIAMS, 1999 Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics, 151: 1197-1210.

BIJMA, P., J. A. M. VAN ARENDONK and J. A. WOOLLIAMS, 2000 Prediction of rates of inbreeding in populations with discrete or overlapping generations under mass selection. Genetics, 154: 1865-1877.

BULMER, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

BURROWS, P. M., 1984 Inbreeding under selection from related families. Biometrics 40: 895-906.

CABALLERO, A. and W. G. HILL, 1992 Effective size of nonrandom mating populations. Genetics 130: 909-916.

GHAI, G. L., 1965 Structure of populations under mixed random and sib mating. Theoretical and Applied Genetics 39: 179-182.

GRUNDY, B., B. VILLANUEVA and J. A. WOOLLIAMS, 1998 Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res., Camb. 72: 159-168.

HILL, W. G., 1972 Effective size of populations with overlapping generations. Theor. Pop. Biol. 3: 278-289.

HILL, W. G., 1979 A note on effective population size with overlapping generations. Genetics 92: 317-322.

MENDELL, N. C. and R. C. ELSTON, 1974. Multifactorial traits: genetic analysis and prediction of recurrence risks. Biometrics 30: 41-57.

NOMURA, T., 1996 Effective size of selected populations with overlapping generations. J. Anim. Breed. Genet. 113: 1-16.

ROBERTSON, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. 2: 189-194.

SANTIAGO, E. and A. CABALLERO, 1995 Effective size of populations under selection. Genetics 139: 1013-1030.

VILLANUEVA, B. and J. A. WOOLLIAMS, 1997 Optimization of breeding programmes under index selection and constrained inbreeding. Genet. Res. Camb. 69: 145-158.

VILLANUEVA, B., R. PONG-WONG, B. GRUNDY and J. A. WOOLLIAMS, 1999 Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genetics Selection Evolution 31: 115-133.

WOOLLIAMS, J. A., P. BIJMA and B. VILLANUEVA, 1999 Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153: 1009-1020.

WOOLLIAMS, J. A. and R. THOMPSON, 1994 A theory of genetic contributions. In Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, vol. 19: 127-134.

WOOLLIAMS, J. A., N. R. WRAY and R. THOMPSON, 1993 Prediction of long term contributions and inbreeding in populations undergoing mass selection. Genet. Res. Camb. 62: 231-242.

WRAY, N. R. and R. THOMPSON, 1990 Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55: 41-54.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1994 Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. 87: 878-892.

WRIGHT, S., 1931. Evolution in Mendelian Populations. Genetics 16: 97-159.

Appendix A

The expected Mendelian sampling variance: The expected Mendelian sampling variance in generation 1 summed over all alleles in the founders can be calculated using the following argument. For the progeny of the carrier founder i^* of the allele, the gene frequency has mean ¼, *i.e.*, half of the gene frequency in carrier (½) plus half of that in mate (0), with $\sigma_a^2 = {}^1/_{16}$. For progeny of other parents, $\sigma_a^2 = 0$. Therefore, for a single allele, the Mendelian sampling variance is $\sigma_a^2 = n_{i^*}/(16X)$ where n_{i^*} is the number of offspring of i^* selected in generation 1, and where X is the total number selected. Summing over all alleles (2 per base individual) and, since the sum of the number of offspring selected over all parents is $2X$, the expected variance is ¼.

At generation 2 and later, with true random mating, the Mendelian sampling variance will be reduced. For dioecious species, this will be delayed by a generation and in general the expected variance is $¼(1-\omega)(1-\Delta F)^{u-1}$ in generation $u > 1$, where ΔF is the rate of inbreeding among the parents.

Appendix B

The covariance of number of offspring and the squared contributions of offspring: Let j denote those selected in the offspring generation and i denote the parents,

$$2\Sigma r_j^2 = \sum_{\text{parents } i} \sum_{\text{offspring } j} r_j^2$$

The factor of 2 is present since each offspring is counted once for each parent. The unconditional expectation of the left-hand side is $2X\mathrm{E}[r_j^2]$ for the offspring generation, which, from the assumption of equilibrium, is equal to $2X\mathrm{E}[r_i^2] = 2X\mathrm{E}[r_i^2 \mid \mathbf{s}_i]$ for the parent generation. For the right hand side, the unconditional expectation can be taken in steps: first expectations from joint conditioning on the number of offspring (n_i) and $\mathbf{s}_i$; second taking expectations conditional on $\mathbf{s}_i$ alone; and finally taking the expectation over $\mathbf{s}_i$ and summing over parents. This gives consecutively:

Step 1: $\sum_{\text{parents } i} n_i \,\mathrm{E}[r_j^2 \mid n_i, \mathbf{s}_i, j \text{ offspring } i]$

Step 2: $\sum_{\text{parents } i} \theta_{n,i}\,\mathrm{E}[r_j^2 \mid \mathbf{s}_i, j \text{ offspring } i]$

Step 3: $XE_s\{\theta_{n,i}\,\mathrm{E}[r_j^2 \mid \mathbf{s}_i, j \text{ offspring } i]\}$

Note that Step 2 uses the assumption of the independence of n_i and $\mathrm{E}[r_j^2 \mid \mathbf{s}_i, j \text{ offspring } i]$. The required result is obtained from Step 3:

$$2X\mathrm{E}[r_i^2] = XE_s\{\theta_{n,i}\,\mathrm{E}[r_j^2 \mid \mathbf{s}_i, j \text{ offspring } i]\}$$

Since $\mathrm{E}[\theta_{n,i}] = 2$, this is equivalent to:

$$E_s\{\theta_{n,i}\, E[r_j^2 \mid s_i,\ j \text{ offspring } i]\} = E_s[\theta_{n,i}]\, E_s\{E[r_j^2 \mid s_i,\ j \text{ offspring } i]\}$$

A simple interpretation of the result is that, in an equilibrium, there can be no preferential selection of offspring with the highest squared contributions, otherwise, we would be making selection progress in squared contributions over time and the equilibrium assumption is broken.

For a single cohort in overlapping generations:

$$2 \sum_{\text{categories } p} \sum_{j(p)} r_{j(p)}^2 = \sum_{\text{categories } q} \sum_{\text{parents } i(q)} \sum_{\text{categories } p} \sum_{\text{offspring } j(p)} r_{j(p)}^2$$

The left-hand side is $2 \sum_{\text{categories } p} X_p E[r_{j(p)}^2]$ which from the assumption of equilibrium is $2 \sum_{\text{categories } q} X_q E_s[r_{i(q)}^2 \mid s_{i(q)}]$. By using the same steps as previously, the right hand side is equal to

$$\sum_{\text{categories } q} X_q \sum_{\text{categories } p} E_S\{\theta_{n,i(q),p}\, E[r_{j(p)}^2 \mid s_{i(q)},\ j(p) \text{ offspring } i(q)]\}$$

which is the result required for overlapping generations.

Extension of the proof relating expected contributions to rates of inbreeding to include overlapping generations: Let X_q be the number of parents in category q and for convenience define a diagonal matrix $\mathbf{N}$ with elements X_q. The prediction for ΔF in overlapping generations is given by Equation 14. Let $\mu_{i(q)} = E[r_{i(q)} | s_{i(q)}]$ and $\sigma_{i(q)}^2 = \mathrm{Var}[r_{i(q)} | s_{i(q)}]$, so that:

$$\Delta F = \tfrac{1}{4} \sum_q X_q (E_s[\mu_{i(q)}^2] + E_s[\sigma_{i(q)}^2]) \quad \text{(B1)}$$

Let $\boldsymbol{m}$ be the vector with element q equal to $E_s[\mu_{i(q)}^2]$, $\boldsymbol{v}$ be the vector with element q equal to $E_s[\sigma_{i(q)}^2]$, and $\mathbf{1}$ be the vector with all elements equal to 1, so that $\Delta F = \tfrac{1}{4}\mathbf{1}^T\mathbf{N}\mathbf{m} + \tfrac{1}{4}\mathbf{1}^T\mathbf{N}\mathbf{v}$. In Equation 16, the n_i is no longer a single number but is a vector of numbers $\boldsymbol{n}_{i(q)}$ where the pth element $n_{p,i(q)}$ is the number of offspring of $i(q)$ selected that belong to category p.

Thus Equation 16 becomes:

$$E[r_{i(q)} \mid s_{i(q)}, \mathbf{n}_{i(q)}] = \tfrac{1}{2} \sum_p n_{p,i(q)}\, E[r_{j(p)} \mid s_{i(q)},\ j(p) \text{ offspring of } i(q)] \quad \text{(B2)}$$

$$\mathrm{Var}[r_{i(q)} \mid s_{i(q)}, \mathbf{n}_{i(q)}] = \tfrac{1}{4} \sum_p n_{p,i(q)}\, \mathrm{Var}[r_{j(p)} \mid s_{i(q)},\ j(p) \text{ offspring of } i(q)] \quad \text{(B3)}$$

To obtain Equations 18 and 19, we need to define $\boldsymbol{\theta}_{n,i(q)} = E[\boldsymbol{n}_{i(q)}]$ with element p given by $\theta_{n,i(q),p}$, and $\mathbf{V}_{n,i(q)}$ to be the (co)variance matrix for the elements $n_{i(q)}$, and to simplify the expressions, define $\boldsymbol{\gamma}$ to be the vector with elements $\gamma_p = E[r_{j(p)} \mid s_{i(q)},\ j(p)$ offspring $i(q)]$, and $\boldsymbol{\eta}$ to be a vector with elements $\eta_p = \mathrm{Var}[r_{j(p)} \mid s_{i(q)}, j(p)$ offspring $i(q)]$. This results in:

$$\mu_{i(q)} = \tfrac{1}{2}\, \boldsymbol{\theta}_{n,i(q)}^T \boldsymbol{\gamma} \quad \text{(B4)}$$

$$\sigma_{i(q)}^2 = \tfrac{1}{4} \boldsymbol{\theta}_{n,i(q)}^T \boldsymbol{\eta} + \tfrac{1}{4} \boldsymbol{\gamma}^T \mathbf{V}_{n,i(q)} \boldsymbol{\gamma} \quad \text{(B5)}$$

It is now possible to make the same assumption as for discrete generations, *i.e.*, that the number in category p selected from parent $i(q)$ has a Poisson distribution:

$$\sigma_{i(q)}^2 = \tfrac{1}{4} \sum_p \theta_{n,i(q),p} (\eta_p + \gamma_p^2) = \tfrac{1}{4} \sum_p \theta_{n,i(q),p}\, E[r_{j(p)}^2 \mid s_{i(q)},\ j(p) \text{ offspring of } i(q)] \quad \text{(B6)}$$

Then by taking expectations over $s_{i(q)}$ in (B6), and using the equilibrium property that $E_s[r_{i(q)}^2]$ is unchanged from generation to generation,

$$\sum_{\text{categories } q} X_q E_s[\sigma_{i(q)}^2] = \tfrac{1}{2} \sum_{\text{categories } q} X_q E[r_{i(q)}^2 \mid s_{i(q)}] \quad \text{(B7)}$$

$$\sum_{\text{categories } q} X_q E_s[\sigma_{i(q)}^2] = \sum_{\text{categories } q} X_q E_s[\mu_{i(q)}^2] \quad \text{(B8)}$$

Thus $\mathbf{1}^T\mathbf{Nm} = \mathbf{1}^T\mathbf{Nv}$, and substitution into Equation 17 gives the result:

$$\Delta F = \tfrac{1}{2}\mathbf{1}^T\mathbf{Nm} = \tfrac{1}{2}\sum_q X_q \, E_s[\mu^2_{i(q)}] \quad \text{(B9)}$$

Appendix C

Prediction of expected genetic contributions for sib indices: Expected genetic contributions were calculated using equilibrium genetic parameters. The genetic parameters were obtained by iterating rounds of selection, starting from an unselected base generation with additive genetic variation h_0^2 and phenotypic variance 1. The iterative equations were $\sigma^2_{A,t+1} = \sigma^2_{Am,t} + \sigma^2_{Af,t} + \tfrac{1}{2}h_0^2$ and $\sigma^2_{Ax,t} = \tfrac{1}{4}\sigma^2_{A,t}(1-k_x\rho_t^2)$ where $\sigma^2_{A,t}$ is the additive genetic variance in generation t, ρ_t denotes the accuracy of selection in generation t (see VILLANUEVA *et al.* 1997), and where x = m or f as appropriate. Equilibrium variances were obtained in five iterations. The notation used in this appendix is given in Table 2.

Calculation of the expected long-term genetic contributions followed the methods of WOOLLIAMS *et al.* (1999). Briefly, these methods depend upon defining two regression models: the first describes the relative fitness of a parent as a linear function of its selective advantages; the second regression model describes the relationship of the selective advantages of the selected offspring with those of its parent. In discrete generations, these models will depend only upon the sex of the parent and the sex of the selected offspring (in overlapping generations they may also depend on age).

For discrete generations, the values of α_m and α_f are simply $(2X_m)^{-1}$ and $(2X_f)^{-1}$ respectively, and so the only term that needs more detailed description is the calculation of $\boldsymbol{\beta}$. $\boldsymbol{\beta}$ is a vector of 3 regression coefficients, the first (β_1) describing the regression of the long-term contribution of a selected male on its selective advantages $A_{i(hs)}$ and the remaining (β_2, β_3) describing the regression of the long-term contribution of a selected female on its two selective advantages $(A_{i(hs)}, A_{i(fs)})$. In the remainder of this Appendix, the selective advantages will be indexed 1 to 3 as above.

$\boldsymbol{\beta}$ is derived from the formula of WOOLLIAMS *et al.* (1999), which has been simplified for application to discrete generations:

$$\boldsymbol{\beta} = \mathbf{N}^{-1}(\mathbf{I} - \tfrac{1}{2}\boldsymbol{\Pi}^T)^{-1}(\tfrac{1}{2}\boldsymbol{\Lambda}^T)(\tfrac{1}{2}\ \tfrac{1}{2})^T$$

where $\mathbf{N}$ is a diagonal matrix with elements (X_m, X_f, X_f), $\mathbf{I}$ is the identity matrix, and where $\boldsymbol{\Pi}$ and $\boldsymbol{\Lambda}$ are described below.

$\boldsymbol{\Lambda}$ is a (2x3) matrix, where λ_{i1} is the regression coefficient for the relative fitness of a male parent on its selective advantage, and where λ_{i2}, λ_{i3} are the corresponding coefficients for the selective advantages of a female parent. When i = 1 the relative fitness is for having male offspring selected, and i = 2 for having female offspring selected. These coefficients will depend on the index of selection used and on the selection intensity. The coefficients are derived using APPENDIX A of WOOLLIAMS *et al.* (1999). The elements are $\lambda_{11} = \lambda_{12} = i_m b_3 \sigma_I^{-1}$, $\lambda_{21} = \lambda_{22} = i_f b_3 \sigma_I^{-1}$, $\lambda_{13} = i_m b_2 \sigma_I^{-1}$, and $\lambda_{23} = i_f b_2 \sigma_I^{-1}$.

$\boldsymbol{\Pi}$ is a (3x3) matrix, with π_{ij} being the regression coefficient of selective advantage i of a selected offspring on the selective advantage j of the parent. This matrix describes exactly how the selection process in one generation is related to the same process in the next generation. The elements of $\boldsymbol{\Pi}$ are derived by standard selection theory (described in detail in APPENDIX B of WOOLLIAMS *et al.* 1999) and account for the effects of selection. Let $z = \rho\sigma_A/\sigma_I$ then the elements of $\boldsymbol{\Pi}$ are: $\pi_{11}=\tfrac{1}{2}(1-k_m b_3 z)$, $\pi_{12}=\tfrac{1}{2}(1-k_m b_3 z)$, $\pi_{13}=\tfrac{1}{2}(1-k_m b_2 z)$; $\pi_{21}=\tfrac{1}{2}(1-k_f b_3 z)/d$, $\pi_{22}=\tfrac{1}{2}(1-k_f b_3 z)/d$, $\pi_{23}=\tfrac{1}{2}(1-k_f b_2 z)/d$; $\pi_{31}=\tfrac{1}{2}(1-k_f b_3 z)(1-d^{-1})$, $\pi_{32}=\tfrac{1}{2}(1-k_f b_3 z)(1-d^{-1})$, $\pi_{33}=\tfrac{1}{2}(1-k_f b_2 z)(1-d^{-1})$.

Example: For $X_m = 20$, $X_f = 60$, $h_0^2 = 0.4$, $n_m = n_f = 4$, with weights $b_1 = 1.0$, $b_2 = 1.5$, $b_3 = 2.0$. The principal parameters for the scheme are given in Table 3. Using the formulae given above:

$2\Lambda =$ (0.797, 0.797, 0.598 | 0.551, 0.551, 0.413)

$2\Pi =$ (0.123, 0.123, 0.155 | 0.045, 0.045, 0.055 | 0.090, 0.090, 0.109)

The solutions are α = (0.0250, 0.0083) and β = (0.0447, 0.0149, 0.0130).

Appendix D

The variances of family size after selection when litter sizes are constant: The variances of family size when litter sizes are constant will be derived by combining results of BURROWS (1984) and MENDEL and ELSTON (1974), which will extend and formalise results used by WOOLLIAMS *et al.* (1993). For simplicity, litters are assumed to have n males and n females, and there are T candidates for selection in each sex. The basic approach of using factorial moments, *i.e.*, $\mathrm{E}[n_{ij}(q)(n_{ij}(q)-1)]$, where $n_{ij}(q)$ is the number of sex q (*i.e.*, $q = m$ or f) selected from the full-sib family with sire i and dam j, was described in detail by BURROWS (1984). Since BURROWS (1984) was working in the context of forestry, only a single sex was considered and hence some extension to two sexes is necessary. The approach of BURROWS (1984) has been preferred since it results in elegant formulae.

Denote $n_{ij}(q)$ as the number of offspring selected of sex q from the full-sib family of sire i and dam j, and $n_{i*}(q)$ as the number selected from sire i (*i.e.*, summed over all its mates). Note that the variance of family size can be simply expressed in terms of the factorial moments.

$$Var[n_{ij}(q)] = \mathrm{E}[n_{ij}(q)(n_{ij}(q)-1)] - \mathrm{E}[n_{ij}(q)](E[n_{ij}(q)]-1) \quad \text{(D1)}$$

$$Var[n_{i*}(q)] = \mathrm{E}[n_{i*}(q)(n_{i*}(q)-1)] - \mathrm{E}[n_{i*}(q)](E[n_{i*}(q)]-1) \quad \text{(D2)}$$

To obtain deviations of the variance from Poisson family size, the term $\mathrm{E}[n_{ij}(q)](\mathrm{E}[n_{ij}(q)]-1)$ in Equation D1 is replaced by $\mathrm{E}[n_{ij}(q)]^2$ and a similar change is made in Equation D2.

BURROWS (1984) derived the following asymptotic form (equations 4 to 12 of BURROWS):

$$\mathrm{E}[n_{ij}(q)(n_{ij}(q)-1)] \approx [n(n-1)X_q(X_q-1)]/[T(T-1)R(p_q,\rho_{FS})] \quad \text{(D3)}$$

where X_q is the total number of that sex selected and T is the total number of candidates, p_q the proportion selected (*i.e.*, X_q/T) and ρ_{FS} the correlation between full-sibs. $R(p,\rho)$ is the ratio $p^2/\Phi(v,v;\rho)$, where $\Phi(v,v;\rho)$ is the upper-quadrant probability that both variables of a standardised bivariate Normal distribution with correlation coefficient ρ exceed v; v is defined by $\Phi(v) = 1-p$; and $\Phi(v)$ is the distribution function of the standard univariate Normal distribution. The ratio is essentially the probability of two sibs being above the index truncation point when the index correlation among sibs is zero (as in random selection) divided by the probability with correlation ρ. BURROWS (1984) uses tabulated values for the co-selection ratio $R(p,\rho)$, but these values can be approximated closely by results from MENDEL and ELSTON (1974): $R(p,\rho) \approx p/\Phi[(i\rho - v)(1-k\rho^2)^{-\frac{1}{2}}]$. This approximation is used throughout. To allow extension to two sexes, we shall denote $R(p,\rho)$ by $R(p,p,\rho)$. [It seems more natural to the authors to use a term equal to $1/R(p,p,\rho)$ in the formula, which describes the proportional increase in co-selection probability. However, we have used $R(p,p,\rho)$ to maintain continuity of notation with BURROWS (1984)].

BURROWS (1984) derived the following additional result to use for the variance of half-sib family sizes. In this paper only paternal half-sib families are considered, $q = m$:

$$E[n_{ij}(m)\, n_{ik}(m)] = [n^2 X_m (X_m - 1)] / [T(T-1) R(p_m, p_m, \rho_{HS})] \quad \text{(D4)}$$

where j and k are distinct mates to a common parent i, and ρ_{HS} is the correlation between half-sibs. Therefore, for a sire with d mates:

$$E[n_{i*}(m)(n_{i*}(m)-1)] = \sum_j E[n_{ij}(m)(n_{ij}(m)-1)] + \sum_{j \neq k} E[n_{ij}(m)\, n_{ik}(m)]$$
$$= [d(1-n^{-1}) / R(p_m, p_m, \rho_{FS}) + d(d-1) / R(p_m, p_m, \rho_{HS})] [n^2 X_m (X_m - 1)] / [T(T-1)] \quad \text{(D5)}$$

The covariances of male and female family size are:

$$\mathrm{Cov}[n_{ij}(m), n_{ij}(f)] = E[n_{ij}(m)\, n_{ij}(f)] - E[n_{ij}(m)]\, E[n_{ij}(f)] \quad \text{(D6)}$$

$$\mathrm{Cov}[n_{i*}(m), n_{i*}(f)] = E[n_{i*}(m)\, n_{i*}(f)] - E[n_{i*}(m)]\, E[n_{i*}(f)] \quad \text{(D7)}$$

The expected cross-products are derived analogously to the variance, and are given by:

$$E[n_{ij}(m)\, n_{ij}(f)] = [n^2 X_m X_f] / [T^2 R(p_m, p_f, \rho_{FS})] \quad \text{(D8)}$$

$$E[n_{i*}(m)\, n_{i*}(f)] = [d / R(p_m, p_f, \rho_{FS}) + d(d-1) / R(p_m, p_f, \rho_{HS})] [n^2 X_m X_f / T^2] \quad \text{(D9)}$$

The rationale of the term $R(p_m, p_f, \rho)$ as a ratio of probabilities for random selection and with correlation ρ remains unchanged, but has been extended to the situation with two sexes with different selection proportions. This ratio is calculated from MENDEL and ELSTON (1974), using $R(p_m, p_f, \rho) = p_f / \Phi[(i_m \rho - v_f)(1 - k_m \rho^2)^{-1/2}]$, which was found by WRAY *et al.* (1994) to be the more accurate implementation of their results (there are two possible implementations, since the approximation is asymmetric in male and female parameters).

To obtain the variances and covariances conditional upon the selective advantage, the regression model derived for the expected number of offspring selected is used (see APPENDIX C).

Thus for a dam family:

$$E[n_{ij}(m)]^2 = d^{-2}[1 + \lambda_{12}^2 v(A_{i(hs)}) + \lambda_{13}^2 v(A_{i(fs)})] \quad \text{(D10)}$$

$$E[n_{ij}(m)] E[n_{ij}(f)] = d^{-1}[1 + \lambda_{12}\lambda_{22} v(A_{i(hs)}) + \lambda_{13}\lambda_{23} v(A_{i(fs)})] \quad \text{(D11)}$$

$$E[n_{ij}(f)]^2 = 1 + \lambda_{22}^2 v(A_{i(hs)}) + \lambda_{23}^2 v(A_{i(fs)}) \quad \text{(D12)}$$

and for a sire family:

$$E[n_{i*}(m)]^2 = 1 + \lambda_{11}^2 v(A_{i(hs)}) \quad \text{(D13)}$$

$$E[n_{i*}(m)] E[n_{i*}(f)] = d[1 + \lambda_{11}\lambda_{21} v(A_{i(hs)})] \quad \text{(D14)}$$

$$E[n_{i*}(f)]^2 = d^2[1 + \lambda_{21}^2 v(A_{i(hs)})] \quad \text{(D15)}$$

Example: For the scheme in Table 3, where the λ-values are derived in APPENDIX C, co-selection ratios are: $R(p_m, p_m, \rho_{FS}) = 0.350$, $R(p_m, p_m, \rho_{HS}) = 0.546$, $R(p_m, p_f, \rho_{FS}) = 0.482$, $R(p_m, p_f, \rho_{HS}) = 0.656$, $R(p_f, p_f, \rho_{FS}) = 0.589$, $R(p_f, p_f, \rho_{HS}) = 0.742$. Applying the results of this APPENDIX gives: $\mathbf{V}_{n(m),dev}$ = (0.186, 0.751 | 0.751, -0.079) and $\mathbf{V}_{n(f),dev}$ = (0.020, 0.159 | 0.159, -0.154).

A General Procedure to Predict Rates of Inbreeding in Populations undergoing Mass Selection

Piter Bijma*, Johan A. M. Van Arendonk* and John A. Woolliams†

* Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands and †Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K.

Abstract – Predictions of rates of inbreeding (ΔF), using the concept of long-term genetic contributions assuming the infinitesimal model, are developed for populations with discrete or overlapping generations and mass selection. Phenotypes are assumed to be recorded prior to reproductive age and to remain constant over time. The prediction method accounts for inheritance of selective advantage both within and between age classes and for changing selection intensities with age. Terms corresponding to previous methods that assume constant selection intensity with age are identified. Predictions are accurate (relative errors < 8%), except for cases with extreme selection intensities in females in combination with high heritability. With overlapping generations ΔF reaches a maximum when parents are equally distributed over age classes, which is mainly due to selection of the same individuals in consecutive years. ΔF /year decreases much more slowly compared to ΔF /generation as the number of younger individuals increases, whereas the decrease is more similar as the number of older individuals increases. The minimum ΔF (/year or /generation) is obtained when most parents were in the later age classes, which is mainly due to increased numbers of parents per generation. With overlapping generations, the relation between heritability and ΔF depends on the age structure of the population.

In the absence of selection and with a Poisson distribution of family size, expected rates of inbreeding are related directly to the number of parents, $E(\Delta F) = 1/(8N_m) + 1/(8N_f)$ (WRIGHT, 1969 p. 212). In selected populations, however, superior families contribute more offspring to the next generation than average families. This increases the rate of inbreeding of a selected population compared to an unselected population. Prediction of rates of inbreeding in selected populations is difficult, because selection decisions are correlated over generations due to the inheritance of selective advantage. Methods accounting for only one or two generations of selection (*e.g.*, BURROWS, 1984a,b), therefore, generally underestimate the rate of inbreeding (WRAY *et al.* 1990; see CABALLERO 1994 for a review).

Two approaches to prediction of rates of inbreeding for selected populations can be distinguished. First, rates of inbreeding can be predicted based on the variance of allele frequency, using the idea of accumulation of selective advantages over generations (ROBERTSON, 1961). Using this approach and equilibrium genetic variances, SANTIAGO and CABALLERO (1995) obtained accurate predictions for populations with discrete

generations undergoing mass selection. NOMURA (1996) extended that method to populations with overlapping generations and equal numbers of parents per sex selected from each age class. Second, rates of inbreeding can be predicted using the concept of long-term genetic contributions. Rates of inbreeding are proportional to the sum of squared long-term genetic contributions of ancestors (WRAY and THOMPSON, 1990). WRAY and THOMPSON (1990) obtained accurate predictions of rates of inbreeding for populations with discrete generations undergoing mass selection, using iterative regression methods. For discrete generations and mass selection a closed form expression was obtained by WOOLLIAMS *et al.* (1993). For more complicated selection schemes, however, predictions became unmanageable due to the recursive nature of the procedure and the need for predicting the variance of long-term genetic contributions (WRAY *et al.* 1994).

Recently, WOOLLIAMS and BIJMA (1999) showed that the variance of long-term genetic contributions is related to their squared expectation, making a separate prediction of the variance redundant. Furthermore, WOOLLIAMS *et al.* (1999) obtained general predictions of expected genetic contributions using equilibrium genetic variances instead of second generation genetic variances (WOOLLIAMS *et al.* 1993). Using the approach of WOOLLIAMS *et al.* (1999), BIJMA and WOOLLIAMS (1999) obtained accurate predictions of genetic contributions for populations with overlapping generations undergoing mass or sib-index selection. However, they did not develop predictions of rates of inbreeding for those schemes.

The aim of this article is twofold. First, explicit prediction equations for rates of inbreeding in populations with discrete or overlapping generations undergoing mass selection will be developed, on the basis of the theory of WOOLLIAMS *et al.* (1999) and WOOLLIAMS and BIJMA (1999). These predictions are valid for any distribution of parents across age classes, overcoming the restriction of NOMURA (1996) to give a general and practical method for mass selection with overlapping generations. The present method will be compared to methods of SANTIAGO and CABALLERO (1995) for discrete generations, and to methods of NOMURA (1996) for the special case of equal numbers of parents per age class with overlapping generations. The accuracy of predictions will be examined using simulation. Second, relationships between rates of inbreeding and genetic or population parameters will be examined, and differences between populations with discrete and overlapping generations will be presented and discussed.

Derivation of Expressions

Population model: This section describes the population and the selection procedures for which rates of inbreeding will be predicted. This model will also be used in the simulation. The trait considered is assumed to be determined by an infinite number of additive loci, each having an infinitesimal effect (infinitesimal model; FISHER, 1918). Phenotypic

values are the sum of additive genetic values (breeding values) and environmental values, $P = A + E$. A population with either discrete or overlapping generations undergoing mass selection is modeled. With parents up to a maximum age of c_{max} there are $2c_{max}$ categories, one for each sex and age of parent. Categories are indexed by k or by l, so $k,l = 1 \,..\, c_{max}$ are males, and $k,l = c_{max}+1 \,..\, 2c_{max}$ are females. With discrete generations, there are only two categories, males and females, which are indexed by $s = m$ or f. Before reproductive age, phenotypes of individuals are recorded and remain unchanged over time, so that ranking of individuals within categories is constant over time. Within categories, individuals are ranked on their phenotype and each year the highest-ranking n_k individuals are selected from the k^{th} category, to produce the next cohort. The total number of male and female parents of each cohort is, $N_m = \sum_{k=1}^{c_{max}} n_k$ and $N_f = \sum_{k=c_{max}+1}^{2c_{max}} n_k$, respectively. Each sire is mated at random to d dams ($d = N_f/N_m$), and each dam produces a fixed number, n_o, of offspring ($\frac{1}{2}n_o$ of each sex), so that for each sex the total number of offspring born in a cohort is, $T = \frac{1}{2}n_o N_f$. The unit of age, *i.e.*, the interval between consecutive age classes, was one year. Genetic contributions and rates of inbreeding per year, therefore, are equal to genetic contributions and rates of inbreeding per cohort.

General: The prediction of ΔF is based on the concept of long-term genetic contributions (JAMES and MCBRIDE, 1958). The long-term genetic contribution (r_i) of ancestor i in cohort t_1 is defined as the proportion of genes present in individuals in cohort t_2 deriving by descent from i, where $(t_2-t_1) \rightarrow \infty$ (WOOLLIAMS *et al.* 1993). In the remainder of the current article, long-term genetic contributions of ancestors are referred to as "genetic contributions", or simply as "contributions".

Rates of inbreeding are predicted from (WOOLLIAMS and BIJMA, 1999):

$$E(\Delta F) = \tfrac{1}{2}\sum_s n_s E(u_{i,s}^2) + \tfrac{1}{8}\sum_s n_s \delta_s \qquad (1)$$

where $\sum_s$ denotes summation over all exclusive categories, $u_{i,s}$ is the expected *lifetime* long-term contribution of individual i in category s conditional on its selective advantage (which in mass selection is the breeding value), and δ_s a correction factor for deviations of the variance of family size ($\mathbf{V}_n$) from independent Poisson variances. Throughout the paper, family size refers to the number of selected offspring of a parent, not to the number of candidates. With mass selection and fixed n_o, δ_s takes negative values, showing that ΔF for fixed n_o is less than for n_o~Poisson. In Equation 1, categories are exclusive, *i.e.*, individuals are in only one category, and categories are therefore indexed by s instead of k.

To calculate $E(u_{i,s}^2)$, the selective advantage of the mate has to be included, since the mate affects the contribution of an ancestor. With random mating and mass selection, however, the selective advantages of mates are independent, and it is therefore possible to ignore the mate when calculating $u_{i,s}$ and add the mate term when calculating $E(u_{i,s}^2)$. The advantage of this is that the selective advantage contains only one term (the breeding value of

the individual) which simplifies the prediction of $u_{i,k}$.

Rates of inbreeding will be predicted in three steps. First, expected genetic contributions will be predicted using the method of WOOLLIAMS *et al.* (1999). Second, $E(u_{i,s}^2)$ will be derived and third, δ_s will be derived. Discrete and overlapping generations will be treated separately.

The difference between the current prediction and the method of WOOLLIAMS *et al.* (1993) is: 1. The current prediction is based on equilibrium genetic variances which simplifies the prediction of $u_{i,s}$ (WOOLLIAMS *et al.* 1999) 2. The variance of genetic contributions is not predicted separately, since it is related to the mean (WOOLLIAMS and BIJMA, 1999). 3. The current prediction covers populations with overlapping generations.

Discrete generations

Step 1, prediction of expected long-term genetic contributions: Expected genetic contributions of ancestors are obtained from the following linear model (BIJMA and WOOLLIAMS, 1999),

$$E(r_{i,s}|A_{i,s}) = u_{i,s} = \alpha_s + \beta_s(A_{i,s} - \overline{A}_s) \qquad (2)$$

where s denotes males or females, α_s is the expected contribution for an average ancestor of sex s, and β_s is the regression coefficient of the contribution on the breeding value ($A_{i,s}$) of the ancestor as a deviation from the average of the selected group ($\overline{A}_s$) for sex s. In discrete generations, $\alpha_s = \frac{1}{2N_s}$ and $\beta_s = \frac{\alpha_s \lambda}{1-\pi}$, where $\lambda = \frac{1}{2} i \sigma_P^{-1}$, which is the average regression coefficient of the number of selected male and female offspring on the breeding value of the parent, and $\pi = \frac{1}{2}(1 - \kappa h^2)$, which is the average regression coefficient of the breeding value of selected male and female offspring on the breeding value of the parent (BIJMA and WOOLLIAMS, 1999). Here, i = ½(i_m + i_f) is selection intensity and, κ = ½(κ_m + κ_f) is PEARSON's (1903) variance reduction coefficient, and $h^2 = \sigma_A^2 / \sigma_P^2$, where σ_A^2 and σ_P^2 are BULMER's (1971) equilibrium genetic and phenotypic variance.

Step 2: derivation of $E(u_{i,k}^2)$: Substituting Equation (2) and with terms added for the mate:

$$E(u_{i,m}^2) = \alpha_m^2 + \beta_m^2 E[(A_{i,m} - \overline{A}_m)^2] + \sum_{j=1}^{d} \beta_f^2 E[(A_{j,f} - \overline{A}_f)^2] \qquad (3)$$

$$E(u_{i,f}^2) = \alpha_f^2 + \beta_f^2 E[(A_{i,f} - \overline{A}_f)^2] + \frac{1}{d^2} \beta_m^2 E[(A_{j,m} - \overline{A}_m)^2] \qquad (4)$$

where j denotes the mate and

$$E[(A_{i,s} - \overline{A}_s)^2] = (1 - 1/N_s)\sigma_A^2(1 - \kappa_s h^2) \qquad (5)$$

From Equation 1, ignoring the second term, $E(\Delta F)$ = ½[$N_m E(u_{i,m}^2)$ + $N_f E(u_{i,f}^2)$]. Next, substituting Equation 3 and 4 and the equations for β_s, λ and π, predicted ΔF equals (see APPENDIX A),

$$E(\Delta F) = \frac{1}{8N_m} + \frac{1}{8N_f} +$$

$$\frac{i^2h^2}{4(1+\kappa h^2)^2}\left[(1-\kappa_m h^2)(1-\frac{1}{N_m})(\frac{1}{2N_m}+\frac{1}{2N_f}) + (1-\kappa_f h^2)(1-\frac{1}{N_f})\frac{1}{N_f}\right] \quad (6)$$

For $N_m = N_f = \frac{1}{2}N$, the result simplifies to:

$$E(\Delta F) = \frac{1}{2N} + \frac{1}{N}\left[\frac{i^2h^2(1-\kappa h^2)(1-2/N)}{(1+\kappa h^2)^2}\right] \quad (7)$$

The assumption for Equations 6 and 7 is that, conditional on the selective advantage (*i.e.*, conditional on $(A_{i,s}-\overline{A}_s)$ in mass selection) family size follows a Poisson distribution (WOOLLIAMS and BIJMA, 1999), which is approximately the case with mass selection when n_o~Poisson. A numerical example is in APPENDIX A.

Step 3: Correction of $E(\Delta F)$ for deviations of $\mathbf{V}_n$ from Poisson variances: With fixed n_o, family size follows a hypergeometric distribution (BURROWS, 1984b) and a correction is required according to the second term of Equation 1. In this article, we will approximate the hypergeometric variance by a binomial variance, which simplifies the prediction. For more complicated selection strategies, *e.g.*, index selection, a hypergeometric variance may be required (WOOLLIAMS and BIJMA, 1999).

With discrete generations, the second term of Equation (1) reduces to: $\frac{1}{8}[N_m\delta_m + N_f\delta_f]$, where $\delta_s = \boldsymbol{\alpha}^T\mathbf{V}_{n(s),dev}\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^T = (\alpha_m\ \alpha_f)$ and $\mathbf{V}_{n(s),dev}$ is the 2×2 matrix of deviations of the (co)variance of family size from Poisson variances for sex s (WOOLLIAMS and BIJMA, 1999). Diagonal elements of $\mathbf{V}_{n(s),dev}$ are obtained as: $\mathbf{V}_{n(s),dev} = \mathbf{V}_{n(s)} - \mathbf{V}_{n(s),poisson}$, which are of the form $np(1\text{-}p) - np = -np^2$ where n is the number of candidates and p is the selected proportion. Off-diagonal elements of $\mathbf{V}_{n(s),dev}$ are zero. For discrete generations the total correction equals (APPENDIX A):

$$\tfrac{1}{8}\sum_s n_s\delta_s = \frac{-1}{8T} \quad (8)$$

Relation to SANTIAGO and CABALLERO (1995): The prediction equation of SANTIAGO and CABALLERO (1995) (denoted SC95) can be related directly to the current prediction. With random mating and assuming $\alpha_{I,s} = \alpha_O = 0$ (see SC95 for notation), equations 21 and 36 of SC95 reduce to (see APPENDIX A), $\frac{1}{2}N_m\left[\alpha_m^2 + \alpha_m^2Q^2C_m^2\right] + \frac{1}{2}N_f\left[\alpha_f^2 + \alpha_f^2Q^2C_f^2\right]$. This can be equated directly to the first term of Equation 1, which shows that $E(u_{i,s}^2)$ corresponds to $\left[\alpha_s^2 + \alpha_s^2Q^2C_s^2\right]$, and also that $\alpha_s^2 + \alpha_s^2Q^2C_s^2$ corresponds to $\beta_s^2E[(A_{i,s}-\overline{A}_s)^2]$. SANTIAGO and CABALLERO (1995) use $Q = \frac{1}{1-\frac{1}{2}(1-\kappa h^2)}$, which is identical to our $\frac{1}{1-\pi}$. Furthermore, they use $C_s^2 = \frac{1}{2}i^2h^2(1-\kappa h^2)$ which is identical to our $2\lambda^2E[(A_{i,s}-\overline{A}_s)^2]$, where the 2 accounts for the mate.

The correction for deviations of $\mathbf{V}_n$ from Poisson variances can also be related to Equation 36 of SC95. SANTIAGO and

CABALLERO (1995) use $\mathbf{V}_{n(s)}(s',s') = \frac{N_{s'}}{N_s}\left[1-\frac{N_{s'}}{\tilde{n}_s N_s}\right]$, (see Equation 30 of SC95 and ignoring the term C_{sm}^2) where $\tilde{n}_s$ is the number of selection candidates per sex of a parent of sex s ($\tilde{n}_m = \frac{1}{2}n_o d$, $\tilde{n}_f = \frac{1}{2}n_o$) and s' denotes the sex of the offspring. This is a binomial variance. The deviation from a Poisson variance (*i.e.*, $N_{s'}/N_s$) equals: $\mathbf{V}_{n(s),dev}(s',s') = N_{s'}^2/(N_s^2\tilde{n}_s)$. From Equation (36) of SC95, the total correction of ΔF equals $-1/(8T)$, which is identical to Equation 8 (see APPENDIX A). Therefore, with random mating, equations 21, 30 and 36 of SC95 are identical to the current prediction for mass selection. A numerical difference between both methods exists because SC95 omit the correction for a finite number of parents when calculating their $C_{ss'}^2$, which would be equivalent to omitting the $(1-1/N_s)$ in Equation 5 of the current prediction.

Overlapping generations

Step 1, prediction of expected long-term genetic contributions: Genetic contributions are predicted using Equation 2 again, but now categories refer to sex-age class combinations which are indexed by k instead of s; so that $k = 1...2c_{max}$ and $u_{i,k}$ is the expected genetic contribution of individual i originating from its selection in category k. Solutions for α_k and β_k are obtained from (WOOLLIAMS *et al.* 1999)

$$\mathbf{N\alpha} = [\mathbf{G}^T + (\mathbf{G}^T * \mathbf{D}^T)(\mathbf{I} - \mathbf{G}^T * \mathbf{\Pi}^T)^{-1}(\mathbf{G}^T * \mathbf{\Lambda}^T)]\mathbf{N\alpha} \quad (9)$$

$$\mathbf{N\beta} = (\mathbf{I} - \mathbf{G}^T * \mathbf{\Pi}^T)^{-1}(\mathbf{G}^T * \mathbf{\Lambda}^T)\mathbf{N\alpha} \quad (10)$$

where * denotes element by element multiplication, T denotes the transpose of matrices, **I** is the $2c_{max}\times 2c_{max}$ identity matrix, **N** is a $2c_{max}\times 2c_{max}$ diagonal matrix containing the numbers of parents selected from each category (n_k), **Π** is a $2c_{max}\times 2c_{max}$ matrix with each element, π_{kl}, being the regression coefficient of the breeding value of a selected offspring in category k on the breeding value of the parent in category l, **Λ** is a $2c_{max}\times 2c_{max}$ matrix with each element, λ_{kl}, being the regression coefficient of the number of selected offspring in category k on the breeding value of the parent in category l, **G** is a $2c_{max}\times 2c_{max}$ modified gene flow matrix connecting selected offspring to parental categories, **D** is a $2c_{max}\times 2c_{max}$ matrix of deviations of breeding values from the mean of the selected group, **α** is a $2c_{max}$ vector of elements α_l, and **β** is a $2c_{max}$ vector of elements β_l. Generation interval (L) was calculated as the time interval in which genetic contributions sum to one: $L = 1/[\sum_{k=1}^{2c_{max}} n_k\alpha_k]$ (WOOLLIAMS *et al.* 1999). More details and a numerical example are in BIJMA and WOOLLIAMS (1999).

Contributions predicted using Equations 9 and 10 are per year, *i.e.*, they are the long-term contribution originating from a single cohort, not of a total generation. Rates of inbreeding predicted from these contributions are, therefore, also per year.

Step 2, derivation of $E(u_{i,s}^2)$: For the calculation of $E(u_{i,s}^2)$ one needs to find the *lifetime* expected genetic contribution, *i.e.*, one has to account for the fact that individuals may

be selected in multiple categories. With c_{max} age classes per sex and the ranking of individuals within age classes remaining constant, there are $2c_{max}$ exclusive categories which will be indexed by s, *i.e.*, individuals selected once, twice, up to c_{max} times for each sex. Therefore, $s = 1 \ldots c_{max}$ denotes males selected 1 through c_{max} times and $s = c_{max}+1 \ldots 2c_{max}$ denotes females selected 1 through c_{max} times. The expected lifetime contribution for these categories is $u_{i,s} = \sum_k u_{i,k}$, where the sum is taken over the age-sex categories k from which i is selected. Thus individuals are indexed in two different ways, *i.e.*, by whether or not they were selected at a specific age, denoted by k; and by how many times they were selected throughout their lifetime, denoted by s.

The first term of Equation 1 is: $\frac{1}{2}\sum_{s=1}^{c_{max}} n_s E(u_{i,s}^2) + \frac{1}{2}\sum_{s=c_{max}+1}^{2c_{max}} n_s E(u_{i,s}^2)$, with the first term denoting males and the second females. The summation over exclusive categories s can be written in terms of the categories k. For males,

$$\sum_{s=1}^{c_{max}} n_s E(u_{i,s}^2) = \sum_{k=1}^{c_{max}} n_k E(u_{i,k}^2) + 2\sum_{k=1}^{c_{max}-1}\sum_{l=k+1}^{c_{max}} \min(n_l, n_k) E(u_{i,k}u_{i,l}) \quad (11)$$

and for females,

$$\sum_{s=c_{max}+1}^{2c_{max}} n_s E(u_{i,s}^2) = \sum_{k=c_{max}+1}^{2c_{max}} n_k E(u_{i,k}^2) + 2\sum_{k=c_{max}+1}^{2c_{max}-1}\sum_{l=k+1}^{2c_{max}} \min(n_l, n_k) E(u_{i,k}u_{i,l}) \quad (12)$$

where $\min(n_k, n_l)$ denotes the minimum of n_k and n_l (See also example in APPENDIX B).

These summations can be written in matrix form, so that for Poisson family size, the rate of inbreeding per year is:

$$E(\Delta F_Y) = \tfrac{1}{2}\mathbf{1}^T\mathbf{N}_o\mathbf{U}_o\mathbf{1} \quad (13)$$

where $\mathbf{1} = (1\ 1\ \ldots\ 1)^T$, $\mathbf{N}_o$ is similar to $\mathbf{N}$ but has a reordering of age classes within sexes so that they go from large to small according to the number of parents, and $\mathbf{U}_o$ is a $2c_{max}\times 2c_{max}$ matrix containing a lower triangular sub-matrix for each sex (with categories ordered as in $\mathbf{N}_o$), with $E(u_{i,k}^2)$ on the diagonal and $2E(u_{i,k}u_{i,l})$ as off-diagonals in the lower triangular sub-matrices (See example APPENDIX B).

Note that, although Equation 1 uses the exclusive categories s, we have expressed ΔF_Y in terms of the age-sex categories k in Equation 13. Therefore, the expected genetic contributions for the categories k can be used directly in Equation 13. Rates of inbreeding per generation were calculated as $E[\Delta F_L] = LE[\Delta F_Y]$.

As with discrete generations, $E(u_{i,k}^2)$ has to include terms for the mates. With overlapping generations, the mate term consists of two elements. The first element is due to the category of the mate as a deviation of the average category for the sex of the mate, $\alpha_l - \overline{\alpha}_{\mathrm{sex}(l)}$. The second term is due to the selective advantage of the mate within its category, $\beta_l(A_{i,l} - \overline{A}_l)$. Therefore, for males, $u_{i,k} = \alpha_k + \beta_k(A_{i,k} - \overline{A}_k) + \sum_{j=1}^{d}[(\alpha_l - \overline{\alpha}_{\mathrm{sex}(l)}) + \beta_l(A_{j,l} - \overline{A}_l)]$ and for females, $u_{i,k} = \alpha_k + \beta_k(A_{i,k} - \overline{A}_k) + [(\alpha_l - \overline{\alpha}_{\mathrm{sex}(l)}) + \beta_l(A_{j,l} - \overline{A}_l)]/d$, where j

denotes the mate, l the category of the mate and *sex* the sex of the mate. For Equation 11 and 12, expectations of squared contributions are obtained for males as

$$E(u_{i,k}^2) = \alpha_k^2 + (1-1/n_k)\beta_k^2\sigma_A^2(1-\kappa_k h^2) + d\left\{\overline{\alpha_f^2} - \overline{\alpha}_f^2 + \sigma_A^2\overline{[(1-1/n_l)\beta_l^2(1-\kappa_l h^2)]}_f\right\} \quad (14)$$

where $k = 1 \,..\, c_{max}$, and for females as

$$E(u_{i,k}^2) = \alpha_k^2 + (1-1/n_k)\beta_k^2\sigma_A^2(1-\kappa_k h^2) + \frac{1}{d^2}\left\{\overline{\alpha_m^2} - \overline{\alpha}_m^2 + \sigma_A^2\overline{[(1-1/n_l)\beta_l^2(1-\kappa_l h^2)]}_m\right\} \quad (15)$$

where $k = c_{max}+1 \,\ldots\, 2c_{max}$ and bars with subscripts m or f denoting weighted averages over mate categories.

Cross-products in Equations 11 and 12 arise only from the individuals selected in both categories, which are all the individuals selected from the smallest category [*i.e.*, $\min(n_k,n_l)$]. Thus cross-products are,

$$E(u_{i,k}u_{i,l}) = \alpha_k\alpha_l + [1-1/\min(n_l,n_k)]\beta_k\beta_l\sigma_A^2[1-\max(\kappa_k,\kappa_l)h^2] + \alpha_{\min}\beta_{\max}E[\,\overline{A}_{\min} - \overline{A}_{\max}\,] \quad (16)$$

where subscript min denotes the category with the lower number of parents and subscript max denotes the category with the higher number of parents. (With random mating there is no covariance between different mates of i, so that there is no mate term in the cross-product). A numerical example is in APPENDIX B.

Step 3, correction of $E(\Delta F_y)$ for deviations of $\mathbf{V}_n$ from Poisson variances: The second term of Equation 1 is $\frac{1}{8}\sum_k n_k\delta_k$, where $\delta_k = \boldsymbol{\alpha}^T\mathbf{V}_{n(k),dev}\boldsymbol{\alpha}$, and $\mathbf{V}_{n(k),dev}$ is a $2c_{max}\times 2c_{max}$ matrix with deviations from Poisson variances (WOOLLIAMS and BIJMA, 1999). Similar to the discrete generation case, $\mathbf{V}_{n(k)}$ is approximated by a binomial variance. Elements of $\mathbf{V}_{n(k),dev}$ and a numerical example are given in APPENDIX B.

Relation to NOMURA (1996): NOMURA (1996) developed predictions for the special case of equal numbers of parents per sex selected from each age class (denoted n_m and n_f), *i.e.*, for constant selection intensity with age. With those schemes, every parent is selected in every category (except for categories with zero parents) and there are only two exclusive categories, *i.e.*, males selected always and females selected always. In this respect, schemes with equal numbers of parents selected from each age class are like discrete generations, *i.e.*, there are only two categories that do not compete for being selected.

APPENDIX B shows that equations 30 and 31 of NOMURA (1996) reduce to $\Delta F_Y = \frac{1}{2}n_m\left[\alpha_m^2 + \alpha_m^2Q^2C_m^2\right] + \frac{1}{2}n_f\left[\alpha_f^2 + \alpha_f^2Q^2C_f^2\right]$, which is equivalent to the first term of Equation 1. This result is a re-scaling of discrete generations, *i.e.*, with discrete generation, $\alpha_s = 1/(2N_s)$, with overlapping generations and two exclusive categories each contribution half, $\alpha_s = 1/(2n_sL)$, where L is the generation interval. Summation of contributions over the number of parents per generation shows that they sum to one, $\sum_{i=1}^{n_mL} 1/(2N_mL) + \sum_{i=1}^{n_fL} 1/(2N_fL) = 1$.

Furthermore, NOMURA (1996) calculated Q using $(\mathbf{I}-\mathbf{P})^{-1}$ [$\mathbf{P}$ is a gene flow matrix identifying the contribution of parental age

groups to selected offspring multiplied by the proportion of genetic variance remaining after selection, appendix in NOMURA (1996)], which, for his special case, is equivalent to our $(\mathbf{I}-\mathbf{G}^{T}*\mathbf{\Pi}^{T})^{-1}$ (see Equation 10, WOOLLIAMS *et al.* 1999). Analogous to SANTIAGO and CABALLERO (1995), NOMURA (1996) calculated C_s^2 omitting the $(1\text{-}1/n_s)$. Contrary to SANTIAGO and CABALLERO (1995) and to the present study, NOMURA (1996) included a term C_{ss}^2 in the calculation of $\mathbf{V}_{n(s)}$ [the first term in equation 22 of NOMURA (1996)]. Finally, NOMURA (1996) considered only one generation inheritance of selective advantage when he calculated the total contribution of age classes. [See equation 8 of NOMURA (1996), which is equivalent to solving α from $\mathbf{N}\alpha = \mathbf{G}^{T}\mathbf{N}\alpha$ instead of using Equation 9 (WOOLLIAMS *et al.* 1999)].

Stochastic simulation: To examine the accuracy of the prediction equations, the breeding scheme described in the "population model" section was simulated and rates of inbreeding were calculated from simulated data. The simulation procedure is described in BIJMA and WOOLLIAMS (1999). In the simulated data, an ancestor cohort t_1 and a descendent cohort t_2 were chosen (BIJMA and WOOLLIAMS, 1999). Inbreeding coefficients of individuals in cohorts t_1 and t_2 were calculated from the simulated pedigree, using the algorithm of MEUWISSEN and LUO (1992). Rates of inbreeding per year were calculated as $\Delta F_y = 1-\left(\frac{1-\overline{F}_{t_2}}{1-\overline{F}_{t_1}}\right)^{(t_2-t_1)^{-1}}$, where $\overline{F}_{t_1}$ and $\overline{F}_{t_2}$ are the average inbreeding coefficients in cohorts t_1 and t_2 respectively. Rates of inbreeding per generation were calculated as $\Delta F_L = L\Delta F_Y$. Results were averaged over 500 replicates.

Results

Discrete generations: For examination of the accuracy of predictions and to identify the origin of prediction errors, Table 1 shows simulated and predicted ΔF. Two types of predictions are in Table 1, ΔF_{pred}* is the prediction using α and β estimated from simulation, and ΔF_{pred} is the full deterministic prediction using α and β from Equations 9 and 10. Differences between ΔF_{pred} and ΔF_{pred}^{*} reflect prediction errors originating from the prediction of β (in discrete generations, $\alpha_s = 1/(2N_s)$ is known). Differences between ΔF_{sim} and ΔF_{pred}^{*} reflect errors in Equation 1.

Generally, errors of the full prediction in Table 1 are small, most errors are below 5%, maximum errors are up to 8.1% and trends agree well between simulations and predictions. Though errors are small, some trend can be observed. Most errors are positive and errors tend to be highest for $N_m = 10$, but errors tend to be negative for $n_o = 8$ and $N_m = 100$. Prediction errors are partly due to errors in the prediction of β, *i.e.*, ΔF_{pred}^{*} is generally more accurate than ΔF_{pred}. Because we have approximated the hypergeometric variance of family size by a binomial variance, positive errors for small numbers of parents were expected. The correction for hypergeometric variances becomes larger with fewer parents (BURROWS, 1984b), whereas a binomial

Table 1.–Rates of inbreeding from simulation (ΔF_{sim}) and from prediction (ΔF_{pred}*, ΔF_{pred})[a] for populations with discrete generations[b].

		$h^2 = 0.2$						$h^2 = 0.5$					
		$n_o = 4$			$n_o = 8$			$n_o = 4$			$n_o = 8$		
N_m	d	ΔF_{sim}	ΔF_{pred}*	ΔF_{pred}	ΔF_{sim}	ΔF_{pred}*	ΔF_{pred}	ΔF_{sim}	ΔF_{pred}*	ΔF_{pred}	ΔF_{sim}	ΔF_{pred}*	ΔF_{pred}
10	1	0.0222	0.0225	0.0225	0.0291	0.0300	0.0306	0.0235	0.0243	0.0243	0.0323	0.0323	0.0339
	2	0.0186	0.0195	0.0195	0.0226	0.0230	0.0244	0.0204	0.0208	0.0211	0.0261	0.0265	0.0269
	5	0.0166	0.0173	0.0174	0.0191	0.0196	0.0203	0.0179	0.0185	0.0187	0.0216	0.0210	0.0222
40	1	0.0056	0.0057	0.0057	0.0077	0.0078	0.0078	0.0059	0.0062	0.0062	0.0086	0.0087	0.0087
	2	0.0047	0.0049	0.0049	0.0062	0.0061	0.0062	0.0052	0.0054	0.0054	0.0071	0.0069	0.0069
	5	0.0043	0.0043	0.0044	0.0053	0.0051	0.0052	0.0048	0.0048	0.0048	0.0061	0.0057	0.0057
100	1	0.0022	0.0023	0.0023	0.0031	0.0031	0.0032	0.0024	0.0025	0.0025	0.0036	0.0036	0.0035
	2	0.0019	0.0020	0.0020	0.0025	0.0025	0.0025	0.0021	0.0022	0.0022	0.0029	0.0028	0.0028
	5	0.0017	0.0018	0.0018	0.0022	0.0021	0.0021	0.0019	0.0019	0.0020	0.0025	0.0023	0.0023

[a]ΔF_{pred}* = prediction using α and β from simulation, ΔF_{pred} = full prediction, h_0^2 = base generation heritability, N_m = no. of selected sires, d = mating ratio, n_o = number of offspring per dam. [b]Standard errors of simulation result were approx. 1% of the estimate.

correction is unaffected by the number of parents. Because the correction is a negative value, a binomial correction results in an over-prediction for small numbers of parents. The current prediction was compared to the prediction of SANTIAGO and CABALLERO (1995; results not shown). As expected from the close agreement between equations of both methods, both methods gave very similar results.

Relationship of ΔF with heritability and selection intensity: Figure 1 shows the relationship between ΔF and heritability (h_0^2), for $N_m = N_f = 20$ and for three selection intensities (n_o = 2, 8 or 32 $\rightarrow$ i = 0, 1.271 or 1.967). Though relationships of ΔF with heritability and selection intensity can be inferred from other studies (*e.g.*, WRAY and THOMPSON 1990), they have never been explored in detail.

Figure 1 shows that ΔF has a maximum for intermediate heritabilities, (except for n_o = 2) and changes in ΔF are more pronounced with greater selection intensity. The maximum of ΔF for intermediate h^2 is due to the Bulmer effect. When the Bulmer effect is ignored in Equation 7 (*i.e.*, $\kappa = 0$), the rate of inbreeding increases with h^2 over the whole range. The logic behind this is that with increasing h^2 the reduction of between family variance increases, reducing the importance of the family component in the phenotype. Note also that the intraclass correlation between full sibs [$\rho = \frac{1}{2}h^2(1-\kappa h^2)$] has a maximum for $h^2_{max} = 1/(2\kappa)$, which, for a common value of $\kappa = 0.8$, equals $h^2_{max} = 0.625$. For $h_0^2 = 0$ and with Poisson family size, Equation (1) reduces to $E[\Delta F] = 1/(8N_m) + 1/(8N_f) = 0.0125$ (WRIGHT, 1969 p. 212).

With n_o = 2, one male and one female offspring are selected from every pair of parents, which gives zero variance of family size, $\beta = 0$ and minimal inbreeding. Expected long-term genetic contributions are equal for all parents and the variance of the contributions is zero, *i.e.*, expected and realized contributions are equal. The absence of variance of family

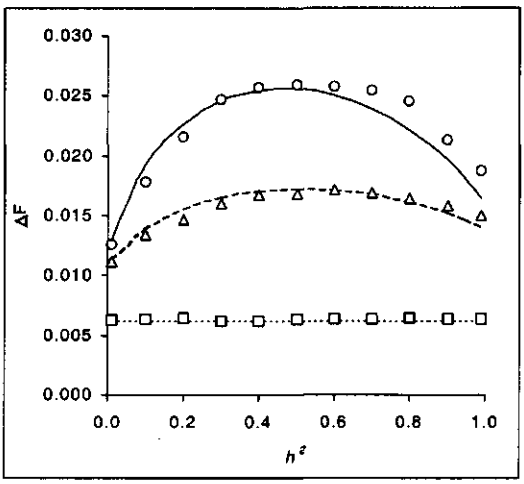

FIGURE 1.–Relation of predicted (lines) and simulated (symbols) rates of inbreeding (ΔF) with heritability h_0^2 for populations with discrete generations, with 20 sires and 20 dams and varying number of offspring per dam (n_o). n_o = 2: □ ΔF_{sim}; ····· ΔF_{pred}. n_o = 8: Δ ΔF_{sim}; – – – ΔF_{pred}. n_o = 32: o ΔF_{sim}; —— ΔF_{pred}.

size with n_o = 2 is taken into account by the correction of ΔF for deviations of V_n from Poisson variances. Without this correction, ΔF_{pred} is equal to a situation with h_0^2 = 0 and Poisson family size, resulting in ΔF_{pred} = 0.0125. The correction halves the prediction to 0.00625. This is an established result (FALCONER and MACKAY, 1996, p. 69). In the absence of variance of family size, which can only be achieved for d = 1, effective population size equals twice the actual population size, N_e = $2(N_m + N_f)$ and $E(\Delta F)$ = $1/2N_e$ = 1/(4*40) = 0.00625.

With higher selection intensities, (n_o = 8 or 32), ΔF increases considerably with heritability. For example, for h^2 = 0.6, ΔF increases by 54% compared to random selection (*i.e.*, h^2 = 0) for n_o = 8, and by 105% for n_o = 32. The large increase of ΔF with selection intensity originates from the regression of the number of selected offspring on the breeding value of the parent, which is linear in i ($\lambda = \frac{1}{2} i \sigma_P^{-1}$), giving a quadratic term in ΔF (Equation 7). Large values of λ indicate that the population descends for a large proportion from only a few ancestors.

For practical selection intensities (n_o = 2, 8), there is close agreement between ΔF_{pred} and ΔF_{sim}. For large selection intensities, errors are larger (*e.g.*, for n_o = 200, $N_m = N_f$ = 40 and h^2 = 0.4, an error of -18% was found). Large errors for extreme selection intensities do not undermine the general theory, *i.e.*, Equation 1 is still valid, but the linear model (Equation 2) may be insufficient to predict expected genetic contributions (WOOLLIAMS and BIJMA, 1999).

Overlapping generations: Table 2 shows simulated and predicted rates of inbreeding per generation and generation intervals. Predictions of ΔF using α and β from simulation (such as ΔF_{pred}^* in Table 1) are not included, because standard errors on β were too large to draw conclusive inferences. Because the potential number of alternative schemes is very large with overlapping generations, a wide range of schemes was evaluated. Only schemes 1,3,5,6,7 are within the scope of NOMURA (1996). Schemes 1 to 5 represent a situation with two age classes with gradually increasing ages of females. Scheme 6 is similar to scheme 5 but with a mating ratio of two. Scheme 7 has equal numbers of parents in all categories. With schemes 8 and 9, parents were ranked across age classes on estimated breeding values

Table 2.–Rates of inbreeding per generation from simulation (ΔF_{sim}) and from prediction (ΔF_{pred}) and generation intervals from simulation (L_{sim}) and prediction (L_{pred}) for populations with overlapping generations[a,b]

	$h_0^2 = 0.2$				$h_0^2 = 0.5$			
diag **N**	ΔF_{Lsim}	ΔF_{Lpred}	L_{sim}	L_{pred}	ΔF_{Lsim}	ΔF_{Lpred}	L_{sim}	L_{pred}
1. {20,0,20,0}	0.0150	0.0156	1.00	1.00	0.0169	0.0173	1.00	1.00
2. {20,0,15,5}	0.0178	0.0185	1.11	1.11	0.0200	0.0207	1.10	1.10
3. {20,0,10,10}	0.0194	0.0206	1.23	1.22	0.0229	0.0235	1.20	1.18
4. {20,0,5,15}	0.0158	0.0165	1.35	1.34	0.0205	0.0209	1.32	1.30
5. {20,0,0,20}	0.0097	0.0104	1.50	1.50	0.0112	0.0115	1.50	1.50
6. {20,0,0,40}	0.0080	0.0083	1.50	1.50	0.0091	0.0091	1.50	1.50
7. {10,10,10,10}	0.0237	0.0247	1.48	1.43	0.0285	0.0290	1.41	1.36
8. {19,1,38,2}	0.0132	0.0133	1.05	1.05	0.0152	0.0148	1.04	1.04
9. {18,2,33,7}[a]	0.0117	0.0123	1.14	1.13	0.0133	0.0137	1.13	1.13
10. {20,0,0,10,0,30}	0.0103	0.0106	1.63	1.62	0.0163	0.0159	1.48	1.43
11. {20,0,0,30,0,10}	0.0125	0.0130	1.17	1.16	0.0153	0.0150	1.10	1.10
12. {10,5,5,0,20,10,5,5}	0.0217	0.0217	1.66	1.58	0.0267	0.0254	1.46	1.41

[a]For $n_o = 8$, **N** = distribution of parents over age classes, h_0^2 = base generation heritability. [a] $n_o = 4$ for this scheme. [b]Standard errors of ΔF_{sim} were approx. 1% of the estimate.

$[EBV_{i,k} = h^2 (P_{i,k} - \overline{P}_k)]$ and the highest ranking N_m males and N_f females were selected across age classes, which gives the highest genetic level of the offspring in the next cohort (JAMES, 1987). This strategy resulted in **N** = diag{19,1,38,2} for $n_o = 8$ and **N** = diag{18,2,33,7} for $n_o = 4$ (for both $h_0^2 = 0.2$ and 0.5). Furthermore some arbitrary schemes with three and four age classes were evaluated to show that predictions are also accurate for more than two age classes. Prediction errors of ΔF_L were small, with most less than 5%. The maximum error was 6.6% and most errors were positive. Similar to the case with discrete generations, positive errors for small numbers of parents were expected due to the binomial approximation for the variance of family size.

Generation intervals are systematically underpredicted in Table 2 (except for schemes with only one reproductive category per sex, in which case L is fixed; schemes 1, 5 and 6). The underprediction is entirely explained by the way L_{sim} is calculated, *i.e.*, $L_{sim} = 1/n_{repl.} \sum_{k=1}^{n_{repl}} L_k$, where $L_k = 1 \Big/ \sum_{l=1}^{2c_{max}} n_l \alpha_l$, *i.e.*, generation interval is calculated per replicate as the time in which genetic contributions sum to unity, and subsequently averaged over replicates (BIJMA and WOOLLIAMS, 1999). However, if α was averaged over replicates and L_{sim} was calculated from the average, *i.e.*, $L_{sim} = 1 \Big/ \sum_{k=1}^{2c_{max}} n_k \overline{\alpha}_k$, then L_{pred} and L_{sim} were in very close agreement (results not shown). This result was expected from the non-linear relationship between L and α, so that $E[L]$ differs from $1 / \sum_{k=1}^{2c_{max}} n_k E[\alpha_k]$.

Results from the current prediction were compared to results from the prediction of

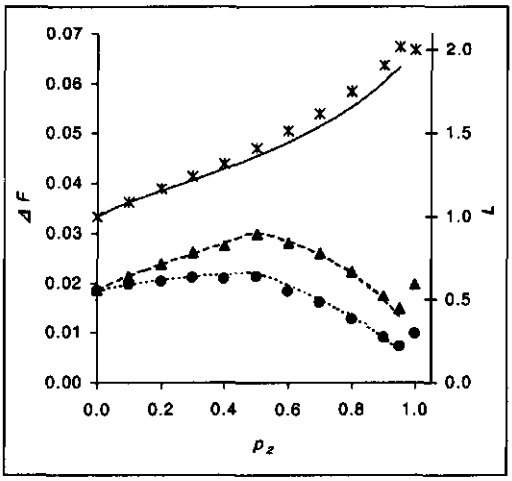

FIGURE 2.–Relation of the proportion of parents from the second age class (p_2) with predicted (lines) and simulated (symbols) generation intervals (L) and with rates of inbreeding per year (ΔF_Y), and per generation (ΔF_L), for a population with two age classes, $N_m = 20$, $N_f = 20$, $h_0^2 = 0.4$ and $n_o = 10$. ● $\Delta F_{Y,sim}$; ····· $\Delta F_{Y,pred}$; ▲ $\Delta F_{L,sim}$; – – – $\Delta F_{L,pred}$; ✳ L_{sim}; —— L_{pred}.

NOMURA (1996) for the special case of equal numbers of parents selected from every age class (results not shown). As expected from theory, results from both methods were similar.

Relationship between ΔF and distribution of parents over age classes: Figure 2 shows the relationship between the rate of inbreeding (per year and per generation) and the proportion of parents selected from the second age class (p_2), for a population with two age classes, $N_m = N_f = 20$, $h_0^2 = 0.4$ and $n_o = 10$. With the exception of $p_2 = 0$, 0.5 and 1.0, these schemes are beyond the scope of NOMURA (1996). Generation interval was also included in Figure 2. On the horizontal axis, parents are shifted from all parents in the first age class ($p_2 = 0$, N = diag{20,0,20,0}) to all parents in the second age class ($p_2 = 1$, N = diag{0,20,0,20}). For $p_2 = 1$, no predictions are presented (*i.e.*, no lines, only symbols), because in this scheme there are two distinct sub-populations that do not mix, *i.e.*, individuals born in odd numbered cohorts are one population and individuals born in even numbered cohorts are the other population. This scheme violates the assumption of one randomly mating population in the derivation of ΔF_{pred}. The populations should, therefore, be treated separately, which resulted in accurate predictions. Despite the complex relationship between ΔF and p_2 in Figure 2, where for example ΔF_Y is nearly constant before declining sharply, accurate predictions were obtained throughout. The rate of inbreeding per year has a flat curve with a maximum for $p_2 = 0.5$, because the increase of ΔF_L with p_2 is counteracted by an increase in the generation interval, and as a result, $\Delta F_Y = \Delta F_L/L$ shows only slight increase before $p_2 = 0.5$ and steep decrease after $p_2 = 0.5$.

For random selection, HILL (1979) showed that the rate of inbreeding with overlapping generations is related to the lifetime variance of family size and the number of parents entering the population per generation. The same pattern can be observed in Figure 2, which shows that ΔF_L has a maximum when parents are equally distributed over age classes, *i.e.* for N = diag{10,10,10,10}, where the 10 parents selected in age class 1 the first year are the same as the 10 parents selected in age class 2 the next year. Thus only 10 distinct parents are selected from every cohort for this scheme, and with $L = 1.41$ the number of parents entering the population per generations equals only 14.1. For N = diag{20,0,20,0}, twenty distinct

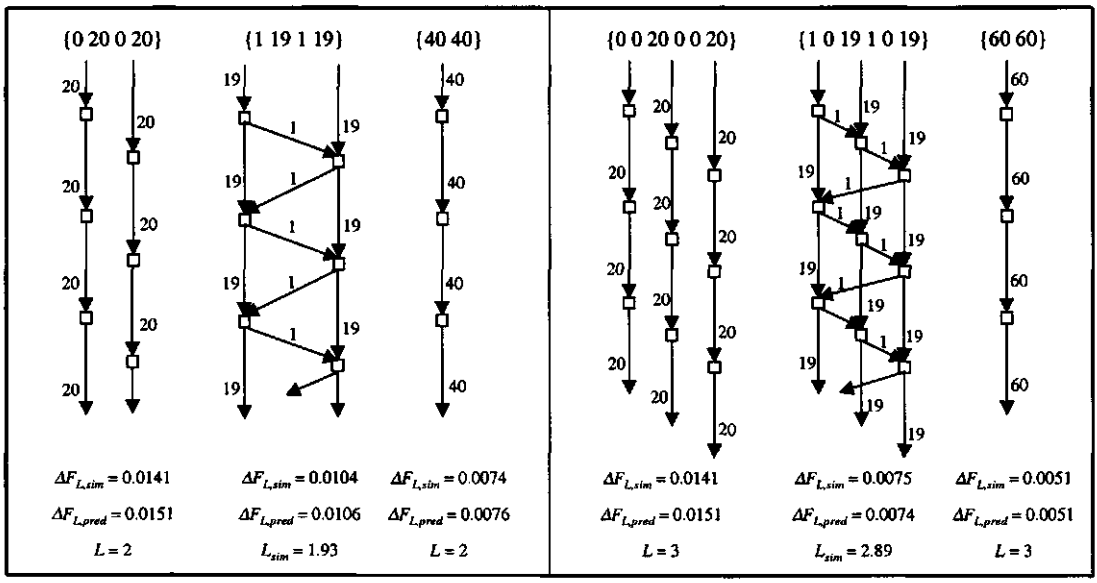

FIGURE 3.–Comparison between full line subdivision, partial migration and one line, with an equal total number of parents per year. Numbers at lines represent the number of parents per sex, $h_0^2 = 0.4$, $n_o = 6$.

parents are selected from every cohort and, with $L = 1$, twenty parents enter the population per generation. The rate of inbreeding per generation reaches a minimum for $p_2 = 0.95$ (**N** = diag{1,19,1,19}). At first glance, this result may seem counter intuitive, *i.e.*, one might expect approximately equal rates of inbreeding per generation for **N** = diag{19,1,19,1} and for **N** = diag{1,19,1,19}. However, for **N** = diag{1,19,1,19}, nineteen distinct individuals are selected from every cohort and, with L = 1.90, the number of parents per generation equals 36.1.

Line subdivision and migration: As mentioned earlier, the scheme with **N** = diag{0,20,0,20} has two non-mixing lines. Changing this scheme to **N** = diag{1,19,1,19} is equivalent to allowing some migration between both lines. Figure 3 shows a comparison between full line subdivision, line subdivision with migration and one single line, for schemes with 2 or 3 age classes. Note that the total number of parents per year is equal per comparison. The comparison shows that allowing some migration between lines substantially reduces ΔF_L (*i.e.*, 0.0104 *vs.* 0.0141 and 0.0075 vs. 0.0141). The smallest ΔF is obtained when lines are joined together, giving **N** = *diag*{40, 40} with a cohort interval of 2 years and **N** = *diag*{60, 60} with a cohort interval of 3 years. When comparing these rates of inbreeding, it must be realized, however, that the schemes with full line subdivision accumulate a between line genetic variance equal to $2(1-1/n_{\text{lines}})F\sigma^2_{A0}$, where the $(1-1/n_{\text{lines}})$ accounts for the fact that the mean is estimated from the sample, *i.e.*, the variance is

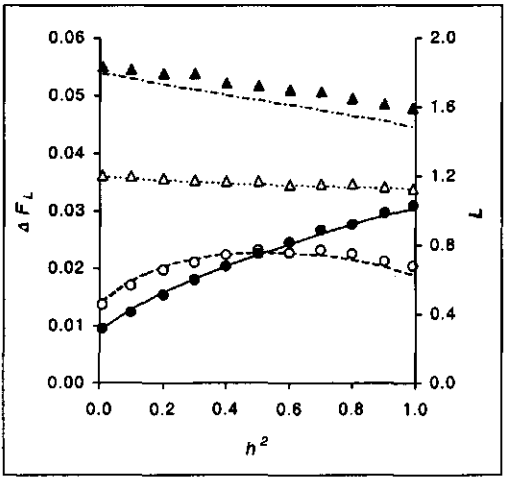

FIGURE 4.–Relation of heritability with simulated (symbols) and predicted (lines) generation interval (L) and with the rate of inbreeding per generation (ΔF_L), with n_o = 8; for two different breeding schemes, S_1: **N** = diag{16,4,16,4} and S_2: **N** = diag{4,16,4,16}. S_1: o $\Delta F_{L,sim}$; – – – $\Delta F_{L,pred}$; Δ L_{sim} ; ····· L_{pred}. S_2: ● $\Delta F_{L,sim}$; —— $\Delta F_{L,pred}$; ▲ L_{sim} ; – · – L_{pred}.

the observed variance in the sample (FALCONER and MACKAY, 1996, p. 265). The total genetic variance at time t, *i.e.*, $\sigma^2_{A,t} = \sigma^2_{A,\text{between}} + \sigma^2_{A,\text{within}}$ equals σ^2_{A0} for **N** = diag{0,20,0,20} and $\sigma^2_{A0}(1+\frac{2}{3}F_t)$ for **N** = diag{0,0,20,0,0,20}. Thus the total genetic variance is larger with full line subdivision.

Relationship between ΔF and heritability: Figure 4 shows the relationship between h^2_0 and ΔF_L, for two breeding schemes. The first scheme (S_1) has most parents in the first age class, **N** = diag{16,4,16,4}, whereas the second scheme (S_2) has most parents in the second age class, **N** = diag{4,16,4,16}. With S_1, ΔF_L has a maximum for h^2_0 = 0.5-0.6, similar to the discrete generation case (See Figure 1). With S_2 however, ΔF_L increases with heritability over the whole range. The increase of ΔF_L with h^2_0 for S_2, is mainly due to an increased contribution of parents in age class one at high heritabilities. With high heritability, genetic gain is large, which gives offspring of one-year-old parents an increased selective advantage. This increases the contribution of parents in age class one relative to the contribution of parents in age class two. For example, with S_2 and h^2_0 = 0.5, expected genetic contributions of average parents are: α^T = [0.027 0.012 0.027 0.012], whereas for h^2_0 = 0.9, expected genetic contributions of average parents are α^T = [0.040 0.011 0.040 0.011]. This result shows that, with increasing h^2_0, the genetic contributions become distributed more unequally over parents, resulting in a higher sum of squared contributions and therefore in an increased ΔF. Furthermore, with S_2, β increases with heritability, resulting in increased differences between genetic contributions of different parents selected from the same category, which further increases ΔF.

Rates of inbreeding per year can be obtained from Figure 4 as $\Delta F_Y = \Delta F_L/L$, which shows the same trends with h^2_0 as ΔF_L. In conclusion, results from Figure 4 show that, in contrast to the case of discrete generations, no general pattern can be observed in the relationship between ΔF and h^2_0 with overlapping generations.

Discussion

Explicit prediction equations for rates of inbreeding in populations with either discrete or overlapping generations undergoing mass selection were developed, following the approach of WOOLLIAMS and BIJMA (1999) and WOOLLIAMS *et al.* (1999). Except for extreme selection intensities in females, predictions were accurate for discrete as well as for overlapping generations. Though based on a different approach, the current method extends the method of NOMURA (1996) to populations with overlapping generations and an arbitrary distribution of parents across age classes, removing the stringent restriction of NOMURA (1996). Relationships between rates of inbreeding and genetic and population parameters were also presented. General relationships apparent in discrete generations could not be extended to overlapping generations. For the prediction of rates of inbreeding in overlapping generations, it is crucial to account for the inheritance of selective advantage both between and within categories. For discrete generations with only two categories (males and females) which do not compete for selection, only competition between selection candidates within categories is relevant.

The current method was compared to methods based upon the proportion of genetic variance transmitted to the offspring, which showed that with random mating, the equations of both SANTIAGO and CABALLERO (1995) and NOMURA (1996) can be reduced to simple expressions in terms of expected genetic contributions. SANTIAGO and CABALLERO (1995) suggested that the differences between their results using the reduced genetic variance and those of WOOLLIAMS *et al.* (1993) using long-term contributions were due to the difference in approach. The present results show that the differences obtained previously were most likely due to errors in the derivations involving complex pathways over multiple generations, that were needed by WOOLLIAMS *et al.* (1993). These complexities were avoided by SANTIAGO and CABALLERO (1995). However, WOOLLIAMS and BIJMA (1999) were able to derive the present results using long-term contributions by modelling the transfer of selective advantages in a single generation by assuming an equilibrium. The idea of basing the prediction on Bulmer's equilibrium variances was introduced by SANTIAGO and CABALLERO (1995). However, their approach to modeling the inheritance of selected advantage by the proportion of genetic variance retained is correct only for mass selection, whereas the approach of WOOLLIAMS *et al.* (1999) is general.

Prediction errors became large when the number of selection candidates per dam became extremely large (Figure 2), but these situations are out of the scope of most artificial selection programs. Certain species (*e.g.*, fish or chicken) are able to produce many offspring per dam, but the number of selection candidates per dam is generally lower. High selection intensities in males can easily be obtained with a limited number of selection candidates per dam when the mating ratio is large. For these situations predictions were accurate (see Table 1, schemes with $d = 5$, $n_o = 8 \rightarrow i = 2.063$). The errors with

large n_o were not present for low h^2 (results not shown), which indicates that the current method is also applicable to species with a large number of offspring when natural directional selection acts on a trait with low heritability.

In this article, equations for predicting rates of inbreeding were developed assuming a model of truncation selection on a normally distributed trait controlled by an infinitesimal model of gene effects. The predicted rate of inbreeding relates to homozygosity (by descent) at a neutral locus, unlinked to genes affecting the trait under selection (WOOLLIAMS and BIJMA 1999). When the infinitesimal model does not hold and the number of genes affecting the trait is large or when the number of chromosomes is small, it is questionable whether neutral and unlinked loci exist at all. When loci are non-neutral, or linked to non-neutral loci, predicted rates of inbreeding can not be related directly to the homozygosity at the locus, because a covariance between the genetic contribution and the gene frequency will arise due to selection (WOOLLIAMS and BIJMA 1999). However, the rate of inbreeding can still be related to rates of inbreeding obtained by analyzing pedigrees using WRIGHT's (1922) path coefficient method, or MALECOT's (1948) coefficient of kinship, and also to estimates of inbreeding depression based on inbreeding levels calculated from the pedigree. Recently, SANTIAGO and CABALLERO (1998) extended prediction methods for effective population size to populations with linked loci undergoing mass selection, but for discrete generations only.

In general, to obtain accurate predictions of ΔF one needs to account for more than one generation of inheritance of selective advantage between categories. It was sufficient for NOMURA (1996) to account for only a single generation because of the special case of equal numbers of parents per age class. In that case, shifting contributions between age classes has only a minor effect on ΔF because the contributions will remain with the same individuals with the same relative fitness, because every individual is selected in every category. The lifetime contribution will not be affected therefore. For schemes where the number of parents differs between age classes, shifting of contributions between categories means shifting to other individuals (at least partly), which will affect the lifetime contribution. Consider, for example, scheme 10 in Table 2 with $h_0^2 = 0.5$. Accounting for only one generation of inheritance (*i.e.*, calculating α from $\mathbf{N}\alpha = \mathbf{G}^{\mathrm{T}}\mathbf{N}\alpha$, WOOLLIAMS *et al.* 1999) gives $\Delta F_{\mathrm{pred}} = 0.0128$, an error of –21%, whereas using Equation 9 gives $\Delta F_{\mathrm{pred}} = 0.0159$, an error of only –2%.

The use of long-term genetic contributions to predict rates of inbreeding has several appealing properties. First, the derivation of the relationship between rates of inbreeding and genetic contributions is based directly on the probability of alleles being identical by descent, which enhances the intuitive understanding (WOOLLIAMS and BIJMA, 1999). Furthermore, rates of genetic gain can easily be obtained from the covariance between the genetic contribution and the Mendelian sampling component of the breeding value (WOOLLIAMS and THOMPSON, 1994; WOOLLIAMS *et al.* 1999), which integrates methods for predicting

genetic gain and rates of inbreeding. Finally, the prediction procedure for genetic contributions describes mechanisms determining the impact of current individuals on future populations and the turn over rate of genes, and gives therefore an understanding of the mechanisms determining the development of the pedigree (WOOLLIAMS *et al.* 1999; BIJMA and WOOLLIAMS, 1999). Because the approach is general, it is clear how prediction equations can be extended to other situations.

With a fixed total number of parents selected per year, populations showed maximum rates of inbreeding (per year and per generation) when the number of parents entering the populations per generation was least, which occurred with an equal number of parents in every age class. Rates of inbreeding were smallest when most parents were in the older age classes, because those schemes had the largest number of parents entering the population per generation. This result broadly resembles the results of HILL (1974) for random selection in overlapping generations, although selected populations have an additional component of inbreeding arising from the expected differential contributions within age classes which will modify this relationship. Schemes with most parents in the later age classes resembled population subdivision with some migration between lines. Because the selective advantage of categories depends on heritability, genetic contributions of categories are strongly affected when heritability changes (BIJMA and WOOLLIAMS, 1999), *i.e.*, contributions generally shifted to the younger age classes when heritability increased. Therefore, no general relationship between heritability and rate of inbreeding could be observed with overlapping generations.

In this article, equations were developed to predict rates of inbreeding for diploid populations with two sexes under controlled selection. The results are therefore primarily relevant for populations under artificial selection, for example in animal breeding or in selection experiments. Though this article focuses on mass selection within age classes, results for mass selection across age classes can easily be accomplished by choosing the appropriate N, as in scheme 8 and 9 in Table 2. An extension to situation where individuals in older age classes have more information, *e.g.*, progeny information, only requires the calculation of probabilities of selecting the same individual at different ages, which can be done using standard index theory. The method can also be extended to other selection strategies and modes of inheritance (*e.g.*, index selection and imprinting), using the key results of WOOLLIAMS and BIJMA (1999) and WOOLLIAMS *et al.* (1999).

In animal breeding, optimization of breeding programs has focussed for a long time on the maximization of genetic gain for the short term, partly because methods to predict long-term response were not available. When rates of inbreeding in selected populations can be predicted, predictions of long-term response under the infinitesimal model become available. This article enables methods for the optimization of breeding schemes on the long-term (*e.g.*, VILLANUEVA *et al.* 1996, VILLANUEVA and WOOLLIAMS, 1997) to be

extended to populations with overlapping generations and mass selection.

This research was financially supported by the Netherlands Technology Foundation (STW), and was co-ordinated by the Earth and Life Science Foundation (ALW). One author (JAW) gratefully acknowledges: the Ministry of Agriculture, Fisheries and Food (UK) for financial support. Ab Groen is acknowledged for giving useful comments on the manuscript, and Tetsuro Nomura for generously sending us a copy of his programs.

Literature Cited

BIJMA, P. and J. A. WOOLLIAMS, 1999 Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics, 151: 1197-1210.

BULMER, M. G., 1971 The effect of selection on genetic variability. Am. Nat. 105: 201-211.

BURROWS, P. M., 1984a Inbreeding under selection from unrelated families. Biometrics 40: 357-366.

BURROWS, P. M., 1984b Inbreeding under selection from related families. Biometrics 40: 895-906.

CABALLERO, A. 1994 Developments in the prediction of effective population size. Heredity 73: 657-679.

FALCONER, D. S. and T. F. C. MACKAY 1996 *Introduction to quantitative genetics*. Longman, Essex, UK.

FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Roy. Soc. (Edinburgh), Trans. 52: 321-341.

HILL, W. G., 1974 Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

HILL, W. G., 1979 A note on effective population size with overlapping generations. Genetics 92: 317-322.

JAMES, J. W., 1987 Determination of optimal selection policies. J. Anim. Breed. Genet. 104: 23-27.

JAMES, J.W. and G. MCBRIDE, 1958 The spread of genes by natural and artificial selection in a closed poultry flock. J. Genet. 56: 55-62.

MALECOT, G., 1948 Les Mathématiques de l'Hérédité. Masson et Cie, Paris.

MEUWISSEN, T. H. E. and Z. LUO, 1992 Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24: 305-313.

NOMURA, T., 1996 Effective size of selected populations with overlapping generations. J. Anim. Breed. Genet. 113: 1-16.

PEARSON, K., 1903 On the influence of natural selection on the variability and correlation of organs. Roy. Soc. London Phil. Trans. A, Vol. 200: 1-66.

ROBERTSON, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. 2: 189-194.

SANTIAGO, E. and A. CABALLERO, 1995 Effective size of populations under selection. Genetics 139: 1013-1030.

SANTIAGO, E. and A. CABALLERO, 1998 Effective size and polymorphism of linked neutral loci in populations under directional selection. Genetics 149: 2105-2117.

VILLANUEVA, B. and J. A. WOOLLIAMS, 1997 Optimisation of breeding programmes under index selection and constrained inbreeding. Genet. Res. Camb. 69: 145-158.

VILLANUEVA, B., J. A. WOOLLIAMS and B. GJERDE, 1996 Optimum designs for breeding programmes under mass selection with an application in fish breeding. Anim. Sci. 63: 563-576.

WOOLLIAMS, J. A., 1998 A recipe for the design of breeding schemes. In *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, vol. 25, pp. 427. Armidale, Australia

WOOLLIAMS, J. A. and R. THOMPSON, 1994 A theory of genetic contributions. In *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, vol. 19, pp. 127-134 Guelph, Canada.

WOOLLIAMS, J. A. and P. BIJMA, 1999 Predicting rates of inbreeding in populations undergoing selection. Genetics 154: 1851-1864.

WOOLLIAMS, J. A., N. R. WRAY and R. THOMPSON, 1993 Prediction of long-term contributions and inbreeding in populations undergoing mass selection. Genet. Res. Camb. 62: 231-242.

WOOLLIAMS, J. A., P. BIJMA and B. VILLANUEVA, 1999 Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153: 1009-1020.

WRAY, N. R. and R. THOMPSON, 1990 Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55: 41-54.

WRAY, N. R., J. A. WOOLLIAMS AND R. THOMPSON 1990 Methods for predicting rates of inbreeding in selected populations. Theor. Appl. Genet. 80: 503-512.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1994 Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. 87: 878-892.

WRIGHT, S., 1922 Coefficients of inbreeding and relationship. Am. Nat. 56: 330-338.

WRIGHT, S. 1969 *Evolution and the genetics of populations*. Vol. 2 The Theory of Gene Frequencies. Univ. of Chicago, Chicago.

Appendix A

Discrete generations

Derivation of Equation 6: Starting from Equations 3 and 4, and substituting: $\alpha_s^2 = \frac{1}{4N_s^2}$, $\beta_s^2 = \frac{\alpha_s^2\lambda^2}{(1-\pi)^2}$, $\lambda = \frac{1}{2}i\sigma_P^{-1}$, $\pi = \frac{1}{2}(1-\kappa h^2)$ and $E[(A_{i,s} - \bar{A}_s)^2] = (1-1/N_s)\sigma_A^2(1-\kappa_s h^2)$, it follows that: $E[u_{i,m}^2] = \frac{1}{4N_m^2} + \frac{1}{4N_m^2}\frac{i^2}{\sigma_P^2}\frac{(1-1/N_m)\sigma_A^2(1-\kappa_m h^2)}{(1+\kappa h^2)^2} + d\frac{1}{4N_f^2}\frac{i^2}{\sigma_P^2}\frac{(1-1/N_f)\sigma_A^2(1-\kappa_f h^2)}{(1+\kappa h^2)^2}$, and $E[u_{i,f}^2] = \frac{1}{4N_f^2} + \frac{1}{4N_f^2}\frac{i^2}{\sigma_P^2}\frac{(1-1/N_f)\sigma_A^2(1-\kappa_f h^2)}{(1+\kappa h^2)^2} + \frac{1}{d^2}\frac{1}{4N_m^2}\frac{i^2}{\sigma_P^2}\frac{(1-1/N_m)\sigma_A^2(1-\kappa_m h^2)}{(1+\kappa h^2)^2}$.

Substituting those expressions into $E[\Delta F] = \frac{1}{2}N_m E[u_{i,m}^2] + \frac{1}{2}N_f E[u_{i,f}^2]$ and using $\sigma_A^2/\sigma_P^2 = h^2$ and $d = N_f/N_m$ gives Equation 6.

Derivation of Equation 8: With a binomial distribution of family size, the deviation from a Poisson variance equals: $np(1-p) - np = -np^2$, where n is the number of candidates ($\frac{1}{2}n_o d$ for sires and $\frac{1}{2}n_o$ for dams) and p is the selected proportion [$1/(\frac{1}{2}n_o d)$ for male offspring and $1/(\frac{1}{2}n_o)$ for female offspring]. Elements of $\mathbf{V}_{n(s),dev}$ are therefore: $\mathbf{V}_{n(m),dev} = [-1/(\frac{1}{2}n_o d),\ 0;\ 0,\ -d/(\frac{1}{2}n_o)]$ and $\mathbf{V}_{n(f),dev} = [-1/(\frac{1}{2}n_o d^2),\ 0;\ 0,\ -1/(\frac{1}{2}n_o)]$. From $\delta_s = \alpha^T\mathbf{V}_{n(s),dev}\alpha$ it follows that: $\delta_m = -1/[(\frac{1}{2}n_o d)(4N_m^2)] - d/[(\frac{1}{2}n_o)(4N_f^2)]$ and $\delta_f = -1/[(\frac{1}{2}n_o d^2)(4N_m^2)] - 1/[(\frac{1}{2}n_o)(4N_f^2)]$. The total correction equals $\frac{1}{8}N_m\delta_m + \frac{1}{8}N_f\delta_f = -1/(8n_oN_f) - 1/(8n_oN_f) = -1/(8T)$, where $T = \frac{1}{2}n_oN_f$.

Example: For $N_m = 20$, $N_f = 60$, $n_o = 8$ and $h^2 = 0.4$, selected proportions, selections intensities and variance reduction coefficients are $p_m = 0.083$, $p_f = 0.250$, $i_m = 1.839$, $i_f = 1.271$, $i = 1.555$, $\kappa_m = 0.839$, $\kappa_f = 0.759$, $\kappa = 0.799$. BULMER's (1971) equilibrium genetic variance and heritability are $\sigma_A^2 = 0.314$, $h^2 = 0.343$. From Equation 6 ΔF for a Poisson variance of family size equals: $\Delta F_{poisson} = 0.00625 + 0.00208 + 0.1277[0.02255 + 0.01212] = 0.01276$. From Equation 8, the correction equals $-1/1920 = -0.00052$ and the final prediction is $E[\Delta F] = 0.0122$.

Relation to Santiago and Caballero (1995): Equation 36 of SC95 equals:

$$N_{es} = 4N_s \Big/ \left\{\left[\frac{1}{\mu_{sm}} + \frac{1}{\mu_{sf}}\right](1-\alpha_{1,s}) + \left[\left(\frac{S_{sm}^2}{\mu_{sm}^2} + \frac{2S_{sm,sf}}{\mu_{sm}\mu_{sf}} + \frac{S_{sf}^2}{\mu_{sf}^2}\right) + 4Q^2C_s^2\right](1+\alpha_{1,s}+2\alpha_0)\right\}$$

where $S_{ss'}^2$ is the variance of family size due to random sampling. Assuming $\alpha_{1,s} = \alpha_0 = 0$ with random mating, and with Poisson family size, $\mu_{ss'} = N_{s'}/N_s$, $S_{ss'}^2 = N_{s'}/N_s$, and $S_{sm,sf} = 0$ this reduces to: $1/N_{es} = \frac{1}{4N_s}[2 + 2N_s/N_{s'} + 4Q^2C_s^2]$, where s'

denotes the sex other than s. With $E[\Delta F] = 1/(8N_{em}) + 1/(8N_{ef})$ it follows that $E[\Delta F] = [1/(8N_m) + 1/(8N_m)Q^2C_m^2] + [1/(8N_f) + 1/(8N_f)Q^2C_f^2] = \frac{1}{2}N_m\left[\frac{1}{4N_m^2}+\frac{Q^2C_m^2}{4N_m^2}\right] + \frac{1}{2}N_f\left[\frac{1}{4N_f^2}+\frac{Q^2C_f^2}{4N_f^2}\right] = \frac{1}{2}N_m\left[\alpha_m^2+\alpha_m^2Q^2C_m^2\right] + \frac{1}{2}N_f\left[\alpha_f^2+\alpha_f^2Q^2C_f^2\right]$, where $\alpha_s^2 = 1/(4N_s^2)$.

Deviations of $V_{n(s)}$ from Poisson variances: With a constant number of selection candidates available per litter, $S_{ss'}^2$ is hypergeometric and SC95 use a binomial approximation, $S_{ss'}^2 = (N_{s'}/N_s)[1-(N_{s'}/(N_s\tilde{n})]$. The difference from a Poisson variance is, $\Delta S_{ss'}^2 = -N_{s'}^2/(N_s^2\tilde{n})$, and with $\alpha_{l,s} = \alpha_0 = 0$ the change in effective population size per sex equals $\Delta N_{es} = 4N_s/\{\Delta S_{sm}^2/\mu_{sm}^2 + \Delta S_{sf}^2/\mu_{sf}^2\}$. The correction of ΔF equals $1/(8\Delta N_{em}) + 1/(8\Delta N_{ef})$. Substituting $\mu_{ss'} = N_{s'}/N_s$, the total correction of ΔF equals $-1/(8T)$.

Appendix B

Overlapping Generations

Corrections for deviations of $V_{n(k)}$ from Poisson variances: From Equation 1, the correction equals $\frac{1}{8}\sum_k n_k\delta_k$, where $\delta_k = \alpha^T V_{n(k),dev}\alpha$ and $V_{n(k),dev}$ is a $2c_{max}\times 2c_{max}$ matrix with deviations of $V_{n(k)}$ from Poisson variances (WOOLLIAMS and BIJMA, 1999). Similar to the case of discrete generations, deviations from Poisson variances are $-np^2$, where n is the number of candidates ($\frac{1}{2}n_o d$ for sires and $\frac{1}{2}n_o$ for dams) and p is the selected proportion. The selected proportion in subclass kl, *i.e.*, among offspring in category k descending from parents in category l, equals $p_{kl} = p_k g_{kl}/g_{ll}^0$, where p_k is the selected proportion in category k ($p_k = n_k/T$), and g_{kl} and g_{ll}^0 are elements of the modified gene flow matrix (**G**) and the conventional gene flow matrix ($\mathbf{G}_0$) respectively. The element g_{kl} represents the proportion of *selected* offspring in category k descending from parents in category l, and g_{ll}^0 represents the proportion of candidates for selection in category k descending from parents in category l (WOOLLIAMS *et al.* 1999). Therefore, $V_{n(l),dev}(k,k)$ equals $-\frac{1}{2}n_o dp_k^2 g_{kl}^2/[g_{ll}^0]^2$ when the parent is a male, and $-\frac{1}{2}n_o p_k^2 g_{kl}^2/[g_{ll}^0]^2$ when the parent is a female. Off-diagonal elements of $V_{n(k),dev}$ are zero with binomial family size.

Example: For $\mathbf{N} = diag\{12,8,15,25\}$, $h_0^2 = 0.4$, and $n_o = 4$; selected proportions, selection intensities and variance reduction coefficients are, p = (0.1500 0.1000 0.1875 0.3125), i = (1.5544 1.7546 1.4357 1.1331), κ = (0.8051 0.8297 0.7877 0.7306). Predicted alpha, beta, generation interval and BULMER's (1971) equilibrium genetic variance and heritability are (see BIJMA and WOOLLIAMS, 1999 for an example of the prediction of α and β), α^T = (0.01974 0.01454 0.01171 0.00710), β^T = (0.02228 0.01829 0.01251 0.00904), $L = 1.416$, $\sigma_A^2 = 0.3355$, $h^2 = 0.3586$. The conventional and modified gene flow matrix are (BIJMA and WOOLLIAMS, 1999),

$$\mathbf{G}_0 = \begin{bmatrix} 0.3 & 0.2 & 0.1875 & 0.3125 \\ 1 & 0 & 0 & 0 \\ 0.3 & 0.2 & 0.1875 & 0.3125 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{G} = \begin{bmatrix} 0.3245 & 0.1755 & 0.2291 & 0.2709 \\ 0.3276 & 0.1724 & 0.2347 & 0.2653 \\ 0.3227 & 0.1773 & 0.2258 & 0.2742 \\ 0.3180 & 0.1820 & 0.2175 & 0.2825 \end{bmatrix}.$$

For $\mathbf{N} = diag\{12,8,15,25\}$, there are four *exclusive* categories: 1. males selected both at one and two years of age (*i.e.*, the eight highest-ranking males), for which $E[u_{i,s=1}^2] = E[(u_{i,k=1}+u_{i,k=2})^2]$, 2. males selected only at one year of age (*i.e.*, males ranking 9

through 12), for which $E[u^2_{i,s=2}] = E(u_{i,k=1})^2$, 3. females selected both at one and two year of age (*i.e.*, the 15 highest-ranking females), for which $E[u^2_{i,s=3}] = E[(u_{i,k=3}+u_{i,k=4})^2]$, and 4. females selected only at two years of age (*i.e.*, females ranking 16 through 25), for which $E[u^2_{i,s=4}] = E(u^2_{i,k=4})$. Summation of expectations of squares and cross-products over categories s, gives, for males, $\sum_{s=1}^{2} n_s E(u^2_{i,s}) = 12\,E(u^2_{i,k=1}) + 8\,E(u^2_{i,k=2}) + 2\times8\,E(u_{i,k=1}u_{i,k=2})$, and for females, $\sum_{s=3}^{4} n_s E(u^2_{i,s}) = 15\,E(u^2_{i,k=3}) + 25\,E(u^2_{i,k=4}) + 2\times15\,E(u_{i,k=3}u_{i,k=4})$ (See also Equations 11 and 12).

From Equation 14, $E(u^2_{i,1}) = \alpha_1^2 + (1-1/n_1)\beta_1^2\sigma_A^2(1-\kappa_1 h^2) + d\{\overline{\alpha_f^2} - \overline{\alpha}_f^2 + \sigma_A^2\overline{[(1-1/n_l)\beta_l^2(1-\kappa_l h^2)]}_f\} =$ 0.0003897 + 0.0001086 + 0.0000606 = 0.0005589. (Bars with subscript f denote averages over female categories weighted by the number of dams in the categories, *e.g.*, $\overline{\alpha_f^2} = (15 \times 0.01171^2 + 25 \times 0.00710^2)/40 = 0.829\times10^{-4}$). Similarly, for females $E(u^2_{i,k})$ is calculated from Equation 15. From Equation 16, $E(u_{i,3}u_{i,4}) = \alpha_3\alpha_4 + (1-1/n_3)\beta_3\beta_4\sigma_A^2(1-\kappa_3 h^2) + \alpha_3\beta_4 E[\bar{A}_3 - \bar{A}_4] =$ 0.0000831 + 0.0000254 + 0.0000111 = 0.0001196.

Using Equation 13 (notice the re-ordering), with $\mathbf{N}_o = diag\{12,8,25,15\}$ and

$$\mathbf{U}_o = E\begin{bmatrix} u_1^2 & 0 & 0 & 0 \\ 2u_1u_2 & u_2^2 & 0 & 0 \\ 0 & 0 & u_4^2 & 0 \\ 0 & 0 & 2u_4u_3 & u_3^2 \end{bmatrix} =$$

$$10^{-3}\begin{bmatrix} 0.5589 & 0 & 0 & 0 \\ 0.7868 & 0.3408 & 0 & 0 \\ 0 & 0 & 0.0946 & 0 \\ 0 & 0 & 0.2393 & 0.1972 \end{bmatrix},$$ the rate of inbreeding with Poisson family size is $E(\Delta F_Y) = \frac{1}{2}\mathbf{1}^T\mathbf{N}_o\mathbf{U}_o\mathbf{1} = 0.0123$. Instead of using the matrix form of Equation 13, one can also use $E(\Delta F_Y) = \frac{1}{2}\sum_{s=1}^{c_{max}} n_s E(u^2_{i,s}) + \frac{1}{2}\sum_{s=c_{max}+1}^{2c_{max}} n_s E(u^2_{i,s})$ with the summations calculated as above.

The correction for deviations of $\mathbf{V}_{n(s)}$ from Poisson variances, *e.g.*, for sires in age class one to selected male offspring in age class two, is $\mathbf{V}_{n(1),dev}(2,2) = -\frac{1}{2}n_o dp_2^2 g_{21}^2/[g_{11}^0]^2 = -0.0477$. The full matrix for sires in age class one equals:

$$\mathbf{V}_{n(1),dev} = \begin{bmatrix} -0.1053 & 0 & 0 & 0 \\ 0 & -0.0477 & 0 & 0 \\ 0 & 0 & -0.1627 & 0 \\ 0 & 0 & 0 & -0.4389 \end{bmatrix}$$

The matrices for other age classes are: $\mathbf{V}_{n(2),dev} = diag\{-0.0692, -0.0297, -0.1105, -0.3235\}$, $\mathbf{V}_{n(3),dev} = diag\{-0.0672, -0.0313, -0.1020, -0.2629\}$, $\mathbf{V}_{n(4),dev} = diag\{-0.0338, -0.0144, -0.0541, -0.15957\}$. Delta is: $\delta^T = [-9.554\ -6.474\ -6.004\ -3.168]\times10^{-5}$, *e.g.*, $\delta_1 = \alpha^T\mathbf{V}_{n(1),dev}\alpha = -9.554\times10^{-5}$. The correction factor is $\frac{1}{8}\sum_k n_k\delta_k = 0.0004$, resulting in $\Delta F_Y = 0.0123 - 0.0004 = 0.0119$ and $\Delta F_L = L\Delta F_Y = 0.0168$.

Relation to NOMURA (1996): Equation 30 of NOMURA (1996) equals:

$$N_{ey,k} = 4n_k L^2 \Bigg/ \left[\left(\frac{1}{\mu_{km}} + \frac{1}{\mu_{kf}} \right)(1-\alpha_k) + \left(\frac{V_{km}}{\mu_{km}^2} + \frac{V_{kf}}{\mu_{kf}^2} + 4Q^2 C_k^2 \right)(1+\alpha_k) \right]$$

Note that the denominator is analogous to SC95. Analogous to SC95, this reduces to $E[\Delta F_y]$ =

$$\frac{1}{L^2}\left\{ \tfrac{1}{2} n_m \left[\frac{1}{4n_m^2} + \frac{Q^2 C_m^2}{4n_m^2} \right] + \tfrac{1}{2} n_f \left[\frac{1}{4n_f^2} + \frac{Q^2 C_f^2}{4n_f^2} \right] \right\}$$

where n_m and n_f are the number of male and female parents selected from every age class. With $L = 1 \Big/ \sum_{k=1}^{2c_{max}} n_k \alpha_k$ (BIJMA and WOOLLIAMS, 1999) and, since in total males and females contribute equally, $\sum_{k=1}^{c_{max}} n_m \alpha_k = \sum_{k=c_{max}+1}^{2c_{max}} n_f \alpha_k$, it follows that $1/L^2 = 4n_m^2 \left(\sum_{k=1}^{c_{max}} \alpha_k \right)^2 = 4n_f^2 \left(\sum_{k=c_{max}+1}^{2c_{max}} \alpha_k \right)^2$. Notice that $\sum_{k=1}^{c_{max}} \alpha_k = \alpha_m$ and $\sum_{k=c_{max}+1}^{2c_{max}} \alpha_k = \alpha_f$, which are the expected lifetime genetic contributions for an average male and for an average female respectively. Substitution gives $E[\Delta F_y] = \tfrac{1}{2} n_m \left[\alpha_m^2 + \alpha_m^2 Q^2 C_m^2 \right] + \tfrac{1}{2} n_f \left[\alpha_f^2 + \alpha_f^2 Q^2 C_f^2 \right]$.

Prediction of Rates of Inbreeding in Populations Selected on Best Linear Unbiased Prediction of Breeding Value

Piter Bijma* **and John A. Woolliams†**

* Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands and †Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K.

Abstract – Predictions for the rate of inbreeding (ΔF) in populations with discrete generations undergoing selection on Best Linear Unbiased Prediction (BLUP) of breeding value were developed. Predictions were based on the concept of long-term genetic contributions using a recently established relationship between expected contributions and rates of inbreeding and a known procedure for predicting expected contributions. Expected contributions of individuals were predicted using a linear model $u_{i(x)} = \alpha_x + \beta_x s_{i(x)}$, where $s_{i(x)}$ denotes the selective advantage as a deviation from the contemporaries, which was the sum of the breeding values of an ancestor and the breeding value of its mates. The accuracy of predictions was evaluated for a wide range of population and genetic parameters. Accurate predictions were obtained for populations of 5 up to 20 sires. For 20 up to 80 sires systematic underprediction of on average 11% was found, which was shown to be related to the goodness of fit of the linear model. Using simulation, it was shown that a quadratic model would give accurate predictions for those schemes. Furthermore, it was shown that, contrary to random selection, ΔF less than halved when the number of parents was doubled, and that in specific cases, ΔF may increase with the number of dams.

In genetic evaluation of individuals, Best Linear Unbiased Prediction (BLUP) (HENDERSON 1963, 1975) of additive genetic merit is an increasingly applied procedure in a variety of fields. Though developed in the context of livestock breeding programs, BLUP is now becoming an integral component of tree breeding (KERR, 1998) and has recently been introduced into fish breeding (GJØEN and GJERDE, 1998). The BLUP procedure utilizes information of all relatives in an optimal way, to give the most accurate prediction of additive genetic merit. BLUP, therefore, has become the method of choice for estimating breeding values of individuals from field records of large and complex pedigrees (LYNCH and WALSH, 1998). Selection on breeding values estimated using BLUP allows for increased genetic selection differentials and gives the highest response from a single cycle of selection (GOFFINET, 1983). For this reason, truncation selection on BLUP of additive genetic merit has often been regarded as the optimal selection procedure (*e.g.*, JAMES, 1987).

In most selections schemes however, a balance needs to be found between short term and long-term selection response. Selection schemes that maximize short term response by

utilizing all available information generally lead to increased rates of inbreeding (*e.g.*, ROBERTSON, 1961; VERRIER *et al.* 1993). High rates of inbreeding (*i.e.*, small effective population size) cause a decrease of genetic variation and a decreased accumulation of mutational variance (*e.g.*, LYNCH and HILL, 1986; KEIGHTLEY and HILL, 1987), resulting in a reduction of long-term selection response and fitness. To safeguard the genetic variation of the population in the long-term, the rate of inbreeding needs to be restricted to an acceptable level. Therefore, besides the expected selection response, one needs to know the expected rate of inbreeding before being able to choose among breeding schemes. This requires a method for predicting rates of inbreeding in populations undergoing BLUP selection, which is currently lacking.

The rate of inbreeding (ΔF) is proportional to the sum of squared long-term genetic contributions, and therefore, ΔF can be predicted using long-term genetic contributions (WRAY and THOMPSON, 1990). WRAY and THOMPSON (1990) obtained accurate predictions of ΔF for populations undergoing mass selection. However, their method was complicated due to the iterative nature of the prediction procedure and the need for predicting the variance of long-term genetic contributions.

Recently, using long-term genetic contributions, a general procedure to predict rates of inbreeding in selected populations was developed by WOOLLIAMS *et al.* (1999) and WOOLLIAMS and BIJMA (1999), simplifying and generalizing the approach of WRAY and THOMPSON (1990). Using that procedure, BIJMA *et al.* (1999) developed predictions of ΔF for populations with discrete or overlapping generations and mass selection. WOOLLIAMS and BIJMA (1999) developed predictions for populations with discrete generations and sib-index selection.

The current article extends the procedure for predicting ΔF to populations with discrete generations that are selected on BLUP of additive genetic merit, using the general approach of WOOLLIAMS *et al.* (1999) and WOOLLIAMS and BIJMA (1999). Predictions of ΔF for BLUP selection have not been developed before. The accuracy of predictions will be evaluated by comparing predictions to rates of inbreeding observed in simulated data. Furthermore, it will be shown that, with BLUP selection, the relationship between ΔF and the size of the breeding scheme and between ΔF and the mating ratio differs qualitatively from those relationships with random selection. Finally, in DISCUSSION, the current prediction method will be compared to an extension of the method of BURROWS (1984a,b).

Derivation of Expressions

Population structure: This section describes the trait and the population structure for which rates of inbreeding will be predicted. The model described in this section will also be used in the simulations (See also BIJMA *et al.* (1999) for details on the simulation procedure). Table 1 shows the notation that will be used. The infinitesimal model is assumed. Phenotypic values were the sum of additive genetic values

TABLE 1.–Notation used

N_m, N_f, d	Number of sires, number of dams, mating ratio $d = N_f/N_m$
n_o, T	Number of offspring born per dam, total number of candidates
p_x, t_x	Selected proportion and standardized truncation point for sex x
i_x, k_x	Selection intensity, variance reduction coefficient for sex x
x	Subscript, $x = m$ or f denoting males or females
P, A, $\hat{A}$	Phenotype, breeding value, estimated breeding value (EBV)
$\mathbf{b}$, $\mathbf{x}_i$	6×1 vector of index weights, 6×1 vector of index information sources
σ_A^2, $\sigma_{\hat{A}}^2$	Additive genetic variance, variance of $\hat{A}$
ρ, h^2	Accuracy of selection, heritability
$\bar{\rho}_{FS}$, $\bar{\rho}_{HS}$	Sample correlation between EBVs of full sibs, and between EBVs of half sibs
ΔF, $r_{i(x)}$	Rate of inbreeding, long term genetic contribution of individual i of sex x
$s_{i(x)}$, $\sigma^2_{s(x)}$	Selective advantage of individual i of sex x, variance of $s_{i(x)}$
$u_{i(x)}$	Expectation of $r_{i(x)}$ conditional on $s_{i(x)}$
α_x, β_x	Linear model for $u_{i(x)} = \alpha_x + \beta_x s_{i(x)}$
Π, π_{xy}	2×2-matrix of regression coefficients of $s_{\text{offspring}}$ on s_{parent}, element of Π
Λ	2×2 matrix of regression coefficients of the number of selected offspring on s_{parent}
λ_{xy}, $\mu_x(y)$	Element of Λ, expected number of offspring of sex y selected from parent of sex x
$\mathbf{V}_{n(x)}$, $\Delta\mathbf{V}_{n(x)}$	2×2 matrix of variance of family size, deviation of $\mathbf{V}_{n(x)}$ from Poisson variance
δ_x	correction term needed when $\Delta\mathbf{V}_{n(x)} \neq 0$, see Equations 1 and 9

(breeding values) and environmental values, $P = A + E$. Heritability was $h^2 = \sigma_A^2/\sigma_P^2$, where σ_A^2 is the additive genetic variance and σ_P^2 is the phenotypic variance.

A population with discrete generations was modeled. Each generation, $\frac{1}{2}T$ male selection candidates and $\frac{1}{2}T$ female selection candidates were ranked on the BLUP of their breeding value (*i.e.*, the estimated breeding value, denoted as $\hat{A}$) and the highest ranking N_m males and N_f females were selected to become sires and dams of the next generation. Each sire was mated at random to $d = N_f/N_m$ dams, and each dam produced n_o offspring ($\frac{1}{2}n_o$ of each sex). The total number of offspring born per generation equaled therefore, $T = n_o N_f$, so that selected proportions were, $p_m = 1/(\frac{1}{2}n_o d)$ and $p_f = 1/(\frac{1}{2}n_o)$. Selection and mating were iterated until equilibrium genetic variances (BULMER, 1971) were reached (See APPENDIX A). The current prediction uses those equilibrium genetic variances.

Pseudo BLUP selection index: To enable deterministic prediction of ΔF, BLUP selection will be approximated by the pseudo-BLUP selection index of WRAY and HILL (1989). As shown by WRAY and HILL (1989), this selection index analogy of BLUP very closely approximates true BLUP selection. The pseudo-

BLUP index was simplified by using an orthogonal re-parameterization of the information sources, so that most information sources are independent. (The re-parameterized index is a BLUP analogy of the WRAY *et al.* 1994 sib index). The advantage is that the (co)variance matrix of the information sources contains only a few non-zero elements. The re-parameterized index for the i^{th} candidate is, $\hat{A}_i = \mathbf{b}^T\mathbf{x}_i$, where ^{T} denotes the transpose, $\hat{A}_i$ is the EBV, $\mathbf{b}$ is a (6×1) vector of weights and $\mathbf{x}_i$ is a 6×1 vector of information sources for the i^{th} candidate. Information sources in $\mathbf{x}_i$ were: 1. $\hat{A}_s$, 2. $(\hat{A}_d - \overline{\hat{A}}_d)$, 3. $\overline{\hat{A}}_d$, 4. $\overline{P}_{HS}$, 5. $(\overline{P}_{FS} - \overline{P}_{HS})$ and 6. $(P_i - \overline{P}_{FS})$, where $\hat{A}_s$ is the EBV of the sire of i, $\hat{A}_d$ is the EBV of the dam of i, $\overline{\hat{A}}_d$ is the average EBV of the d dams mated to the sire, $\overline{P}_{HS}$ is the phenotypic average of the $n_o d$ half-sibs of i (including i and its full-sibs), $\overline{P}_{FS}$ is the phenotypic average of the n_o full-sibs of i (including i) and P_i is the phenotype of candidate i. Information sources 1 and 4, 3 and 4, and 2 and 5 are correlated, the others are mutually independent. Iterative equations for calculating index weights, the accuracy of selection (ρ), the correlation between estimated breeding values of full sibs and of half sibs (intraclass correlation, $\overline{\rho}_{FS}$ and $\overline{\rho}_{HS}$, where the bars denote the finite sample mean) and equilibrium variances (BULMER, 1971) are given in APPENDIX A.

Prediction of rates of inbreeding

General: The prediction method is based on the concept of long-term genetic contributions. The long-term genetic contribution (r_i) of ancestor i in generation t_1 is defined as the proportion of genes from i that are present in individuals in generation t_2 deriving by descent from i, where $(t_2-t_1) \rightarrow \infty$ (WOOLLIAMS *et al.* 1993). In the remainder of this article, long-term genetic contributions are referred to as "genetic contributions", or just "contributions".

Rates of inbreeding were predicted from (WOOLLIAMS and BIJMA, 1999),

$$\mathrm{E}(\Delta F) = \tfrac{1}{2}\left[N_m \mathrm{E}_s(u^2_{i(m)}) + N_f \mathrm{E}_s(u^2_{i(f)})\right] + \tfrac{1}{8}\left[N_m\delta_m + N_f\delta_f\right] \quad (1)$$

where E_s denotes the expectation with respect to the selective advantage, $u_{i(x)}$ is the expected genetic contribution of a parent of sex x conditional on its selective advantage $s_{i(x)}$ (*i.e.*, $u_{i(x)} = \mathrm{E}[r_{i(x)}|s_{i(x)}]$) and δ_x is a term to correct the prediction of ΔF for deviations of the variance of family size ($\mathrm{V}_{n(x)}$, where $x = m$ or f) conditional on the selective advantage from independent Poisson variances. (The second term of Equation 1 will be referred to as "term for deviations from Poisson".) Throughout the article, family size refers to the number of selected offspring of a parent, not to the number of selection candidates. The selective advantage may consist of any term that affects the long-term genetic contribution of an ancestor (*i.e.*, by affecting selection of its offspring or of more distant descendants), *e.g.*, it can be the breeding value.

To compute Equation 1, one needs to decide which elements should be included in the

selective advantage. In the current prediction, the selective advantage of an individual will be the sum of its breeding value and the breeding values of its mate(s), though other choices are possible (see DISCUSSION). With mass selection, a selective advantage consisting of linear terms of the breeding value is sufficient to accurately predict ΔF (BIJMA *et al.* 1999). However, when more emphasis is placed on family information, higher order terms may be required as observed by WOOLLIAMS and BIJMA (1999) for selection on a sib index. Therefore, two models will be evaluated. First, the long-term genetic contribution will be a linear function of the breeding value, denoted "linear model". Second, the long-term genetic contribution will be a quadratic function of the breeding value, denoted "quadratic model". For the quadratic model, components of Equation 1 will be estimated from simulated data, *i.e.*, no fully deterministic prediction will be presented for the quadratic model.

For the linear model, rates of inbreeding will be predicted in three steps. First, expected genetic contributions ($u_{i(x)}$) will be predicted using the method of WOOLLIAMS *et al.* (1999). Second, $E_s(u_{i(x)}^2)$ will be derived which enables calculation of the first term of Equation 1. Finally, δ_m and δ_f will be derived, giving the term for deviations from Poisson.

Linear model: In the linear model the selective advantage of sires was

$$s_{i(m)} = (A_{i(m)} + \overline{A}_d) - \overline{(A_{i(m)} + \overline{A}_d)}, \qquad (2)$$

where $A_{i(m)}$ is the breeding value of sire i, $\overline{A}_d$ is the average breeding value of the d dams mated to sire i and the second term represents subtraction of the average. For dams the selective was

$$s_{i(f)} = [(A_{i(f)} + A_s) - \overline{(A_{i(f)} + A_s)}], \qquad (3)$$

where A_s is the breeding value of the sire (*i.e.*, the mate of dam i). Note that $s_{i(m)}$ and $s_{i(f)}$ are zero on average.

Step 1, prediction of expected contributions: Expected contributions ($u_{i(x)}$) were predicted by linear regression on the selective advantage. For both sexes the model was

$$u_{i(x)} = E(r_{i(x)} \mid s_{i(x)}) = \alpha_x + \beta_x s_{i(x)}. \qquad (4)$$

With discrete generations, $\alpha_m = 1/(2N_m)$ and $\alpha_f = 1/(2N_f)$ always. Solutions for β_m and β_f were obtained from a simplified form of equation 7b and 8 of WOOLLIAMS *et al.* (1999) [since with discrete generations the gene-flow matrix can be replaced by ½, see WOOLLIAMS *et al.* (1999)],

$$\begin{bmatrix} N_m\beta_m \\ N_f\beta_f \end{bmatrix} = \left(\mathbf{I}_2 - \tfrac{1}{2}\boldsymbol{\Pi}^{\mathrm{T}}\right)^{-1}\left(\tfrac{1}{2}\boldsymbol{\Lambda}^{\mathrm{T}}\right)\begin{bmatrix} N_m\alpha_m \\ N_f\alpha_f \end{bmatrix}, \qquad (5)$$

where $\mathbf{I}_2$ is the 2×2 identity matrix, $\boldsymbol{\Pi}$ is a 2×2 matrix of regression coefficients π_{xy}, being the regression coefficient of $s_{i(x)}$ of a selected offspring of sex x on $s_{j(y)}$ of its parent of sex y (*e.g.*, π_{12} is the regression coefficient of $s_{i(m)}$ of a selected male offspring on $s_{j(f)}$ of its dam), $\boldsymbol{\Lambda}$ is a 2×2 matrix of regression coefficients λ_{xy}, being the regression coefficient of the number of selected offspring of sex x on $s_{j,y}$ of the parent of sex y (*e.g.*, λ_{21} is the regression of the number of selected female offspring on $s_{i(m)}$ of their sire). Matrices $\boldsymbol{\Pi}$ and $\boldsymbol{\Lambda}$ are calculated

using the method of WOOLLIAMS *et al.* (1999) as outlined in APPENDIX B of the current article.

Step 2, derivation of $E_s(u^2_{i(x)})$*:* Since all terms of the selective advantage are expressed as a deviation from their mean, expectations of squares are equal to variances, so that $E(s^2_{i(x)}) = \sigma^2_{s(x)}$. Therefore, squaring Equation 4 and taking expectations gives

$$E_s(u^2_{i(x)}) = \alpha^2_x + \beta^2_x \sigma^2_{s(x)}, \qquad (6)$$

and from Equation 2 and 3,

$$\sigma^2_{s(m)} = \sigma^2_A[(1-k_m\rho^2)+(1-k_f\rho^2)/d](1-1/N_m) \qquad (7)$$

$$\sigma^2_{s(f)} = \sigma^2_A[(1-k_m\rho^2)(1-1/N_m) + (1-k_f\rho^2)(1-1/N_f)] \qquad (8)$$

where k_x is PEARSON's (1903) variance reduction coefficient (FALCONER and MACKAY, 1996, p. 201).

Step 3, calculation of δ_m *and* δ_f*:* The term for deviations from Poisson (*i.e.*, the second term of Equation 1) requires the calculation of δ_x. As an approximation, WOOLLIAMS and BIJMA (1999) and BIJMA *et al.* (1999) used $\delta_x = \alpha^T \Delta\mathbf{V}_{n(x)}\alpha$, where $\alpha^T = (\alpha_m\ \alpha_f)$ and $\Delta\mathbf{V}_{n(x)}$ is the 2×2 matrix of deviations of the variance of family size from Poisson variances. For example, $\Delta\mathbf{V}_{n(m)}(1,1)$ is the deviation of the variance of the number of selected male offspring of a sire from the Poisson variance, and $\Delta\mathbf{V}_{n(m)}(1,2)$ is the covariance between the number of selected male and female offspring. The above approximation for δ_x accounts only for the average contribution of an offspring (*i.e.*, α). The effect of the selective advantage of the parent on the contribution of an offspring is ignored. This effect can be included by using [see equation 25 through 27 of WOOLLIAMS and BIJMA (1999)]

$$\delta_x = E_s\left[\mathbf{u}^{*T}_{i(x)} \Delta\mathbf{V}_{n(x),i} \mathbf{u}^{*}_{i(x)}\right] \qquad (9)$$

where $\mathbf{u}^{*T}_{i(x)} = (u^*_{j(m)}, u^*_{j'(f)})_x$, with $u^*_{j(y)} = E(r_{j(y)}|\mathbf{s}_{i(x)}$, j is offspring of i), which is the expected contribution of selected offspring j of sex y given the selective advantage of its parent i of sex x. The terms $\mathbf{u}^*_{i(x)}$ and $\Delta\mathbf{V}_{n(x),i}$ will be assumed independent and $\mathbf{u}^*_{i(x)}$ is calculated from

$$\mathbf{u}^*_{i(x)} = \begin{bmatrix} \alpha_m + \beta_m \pi_{mx} s_{i(x)} \\ \alpha_f + \beta_f \pi_{fx} s_{i(x)} \end{bmatrix}, \qquad (10)$$

where π, as defined in Equation 7, represents the transfer of the selective advantage from the parent to the offspring.

Equation 10 requires the calculation of $\Delta\mathbf{V}_{n(x)}$. With fixed n_o, family size follows a correlated hypergeometric distribution and the variance of family size can be approximated using a result of BURROWS (1984a,b) as described in detail in appendix E of WOOLLIAMS and BIJMA (1999). Here we will only outline the concept for a single sex without giving the derivation. Detailed equations are given in APPENDIX B.

In general, variance of family size equals $\mathrm{Var}(n_i) = E[n_i^2] - E[n_i]^2$, where n_i denotes family size after selection, conditional on the selective advantage of the parent. Diagonal elements of $\mathbf{V}_{n(x)}$ represent the variance of the number of selected offspring of a particular sex, and, with n_i~Poisson, $\mathrm{Var}(n_i) = E(n_i)$, so that deviations from Poisson variances are,

$$\Delta \mathbf{V}_{n(x)} = \mathrm{E}_s\{\mathrm{E}[n_i^2] - \mathrm{E}[n_i]^2 - \mathrm{E}[n_i]\}$$
$$= \mathrm{E}_s\{\mathrm{E}[n_i(n_i-1)]\} - \mathrm{E}_s\{\mathrm{E}[n_i]^2\}. \quad (11)$$

Off-diagonal elements of $\mathbf{V}_{n(x)}$ represent the covariance between the number of selected male and female offspring and are obtained following the same approach as for the diagonal elements (APPENDIX B).

The first term of Equation 11 is (BURROWS, 1984a)

$$\mathrm{E}_s\{\mathrm{E}[n_i(n_i-1)]\} \approx$$
$$[n(n-1)N(N-1)]/[T(T-1)R(p, \overline{\rho}_{fam})], \quad (12)$$

where n is the number of candidates per family, N is the total number selected, T is the total number of candidates and $R(p, \overline{\rho}_{fam})$ is the ratio of the probability of selecting two arbitrary candidates over the probability of selecting two family members, where p is the selected proportion and $\overline{\rho}_{fam}$ is the intra-class correlation between family members. The probability of selecting two family members can be approximated using a result of MENDEL and ELSTON (1974) (See APPENDIX B). WRAY *et al.* (1990) observed that Equation 12 gives substantial bias in cases where the number of parents is small compared to the number of offspring per parent, and suggested to adjust the selected proportion,

$$p_{x,\mathrm{adj}} = (1-\overline{\rho}_{fam})p_x + \overline{\rho}_{fam}\max(p_x, 1/N_m). \quad (13)$$

The reasoning behind this correction is that with large $\overline{\rho}_{fam}$ selection moves towards between family selection and $p_{x,\mathrm{adj}}$ is a weighted sum of the original selected proportion and the selected proportion when selecting between families. For schemes with few parents (N_m = 5 or 10), selection intensities and variance reduction coefficients were re-calculated using $p_{x,\mathrm{adj}}$ and used in the calculation of $R(p_x, p_y, \overline{\rho}_{fam})$.

In Equation 11, the second term, $\mathrm{E}_s\{\mathrm{E}[n_i]^2\}$, denotes the expectation of the square of the expected number of selected offspring given the selective advantage, which can be obtained from $\mathrm{E}_s\{\mathrm{E}[n_i]^2\} = \mathrm{E}_s\{[\overline{\overline{n}}(1+\lambda s)]^2\}$ where $\overline{\overline{n}}$ is the overall expected number of offspring selected per parent (*e.g.*, $\overline{\overline{n}}$ = 1 male offspring per sire and $\overline{\overline{n}}$ = d female offspring per sire since population size is constant over time), and λs represents the change of the number of selected offspring due to the selective advantage of the parent.

The extension of Equations 11 and 13 to two sexes and a hierarchical mating structure is described in detail in appendix E of WOOLLIAMS and BIJMA (1999). The resulting equations for calculating $\Delta \mathbf{V}_{n(x)}$ used in the current prediction, an example of computation and more details on the calculation of Equation 9 and 10 are in APPENDIX B.

Quadratic model: With the quadratic model, the selective advantage consists of two terms. For sires, $\mathbf{s}_{i(m)}^{\mathrm{T}} = (s_{i,1}, s_{i,2})$, where

$$s_{i,1} = (A_{i(m)} + \overline{A}_d) - \overline{(A_s + \overline{A}_d)} \quad (14)$$

$$s_{i,2} = s_{i,1}^2 - \overline{s_{i,1}^2} \quad (15)$$

For dams, $\mathbf{s}_{i(f)}^{\mathrm{T}} = (s_{i,3}, s_{i,4})$, where

$$s_{i,3} = (A_s + A_d) - \overline{(A_s + A_d)} \quad (16)$$

$$s_{i,4} = s_{i,3}^2 - \overline{s_{i,3}^2} \quad (17)$$

TABLE 2.–Rates of inbreeding from simulation (ΔF_{sim}) and corresponding prediction errors[a] for a population with 10 sires[b]

		$n_o = 4$		$n_o = 16$	
N_f	h^2	ΔF_{sim}[c]	error%	ΔF_{sim}[c]	error%
20	0.1	0.0347	–8	0.0917	+11
	0.2	0.0323	–4	0.0794	+8
	0.4	0.0289	–5	0.0623	+0
	0.6	0.0253	–5	0.0474	–4
100	0.1	0.0378	–4	0.0700	–1
	0.2	0.0333	–4	0.0609	–9
	0.4	0.0269	–5	0.0452	–10
	0.6	0.0224	–5	0.0341	–10

[a]Predictions were obtained using $p_{s,adj}$ (Equation 13) [b]N_f = number of dams; h^2 = initial heritability; n_o = number of offspring per dam, error% = 100% × ($\Delta F_{pred} - \Delta F_{sim}$)/$\Delta F_{sim}$; ΔF_{pred} = predicted rate of inbreeding from Equation 1 with linear model.
[c]Standard errors of ΔF_{sim} were ≤ 1% of the estimate.

For the quadratic model, components needed to compute Equation 1 were estimated from simulated data. For step one, β was estimated as the multiple regression of the long-term contribution of ancestors on their selective advantage (*e.g.*, for sires, $\beta^T_{(m)} = (\beta_1, \beta_2)$ was the multiple regression of the long-term contribution of sires on $s_{i,1}$ and $s_{i,2}$). For step two, the (co)variance matrix of $s_{i,1}$ through $s_{i,4}$ was estimated from the simulated data and the first term of Equation 1 was calculated analogous to Equations 6. For step three, $\mathbf{V}_{n(x)}$ and Λ were estimated from simulated data and the term for deviations from Poisson was calculated analogous to Equations 9 and 10.

Results

Accuracy of predictions

Linear model: For the linear model, the accuracy of predictions was tested over a wide range of values. All combination of N_m = 5, 10, 20, 40, 60 or 80; d = 1, 2, 3, 5 or 10; n_o = 4, 8 or 16 and h^2 = 0.1, 0.2, 0.4 or 0.6 were evaluated (due to computational restrictions, N_f was restricted to be ≤ 200, *e.g.*, for N_m = 80, only d = 1 and d = 2 were evaluated).

Three different ranges of results could be identified, exemplified in Tables 2 to 4. First, despite very large rates of inbreeding (up to 12.5%), accurate predictions were obtained for schemes with N_m = 5 or 10 (Table 2). For those schemes, the term for deviations from Poisson was calculated using adjusted selected proportions according to Equation 13. The maximum relative error encountered for schemes with N_m = 5 or 10 was 12%, which occurred with N_m = 5, d = 2, h^2 = 0.1 and n_o = 16. For schemes with N_m = 5 or 10 the average relative error was –2% and the standard deviation of the relative error was 5%.

Second, a range with accurate predictions was found for N_m = 20, (see Table 3). For the schemes in Table 3, most errors were negative with a maximum of –9%. For N_m = 20, d = 10, n_o = 16 and h^2 = 0.1 (data not shown) an

TABLE 3.– Rates of inbreeding from simulation (ΔF_{sim}) and corresponding prediction errors[a] for a population with 20 sires

		$n_o = 4$		$n_o = 8$	
N_f	h^2	ΔF_{sim}[c]	error%	ΔF_{sim}[c]	Error%
20	0.1	0.0182	–5	0.0393	–2
	0.2	0.0174	–2	0.0359	–3
	0.4	0.0158	–1	0.0295	–4
	0.6	0.0143	–2	0.0239	–6
40	0.1	0.0184	–9	0.0361	–3
	0.2	0.0171	–5	0.0314	–3
	0.4	0.0151	–5	0.0251	–6
	0.6	0.0130	–5	0.0201	–9
60	0.1	0.0188	–9	0.0350	–2
	0.2	0.0172	–6	0.0305	–4
	0.4	0.0146	–5	0.0230	–5
	0.6	0.0123	–4	0.0181	–7
100	0.1	0.0199	–7	0.0347	3
	0.2	0.0178	–6	0.0294	–2
	0.4	0.0145	–6	0.0220	–5
	0.6	0.0121	–6	0.0168	–7

N_f, number of dams; h^2, initial heritability; n_o number of offspring per dam; [a]error% = $100\% \times (\Delta F_{pred} - \Delta F_{sim})/\Delta F_{sim}$; ΔF_{pred} = predicted rate of inbreeding from Equation 1 with linear model. [c]Standard errors of ΔF_{sim} were ≤ 1% of the estimate.

overprediction of 37% was encountered which was due to bias in Equation (12) and was reduced to –13% when $p_{x,adj}$ (Equation 13) was used. Note that this is an extreme scheme (*i.e.*, i_m = 2.59, $\overline{\rho}_{FS}$ = 0.86, $\overline{\rho}_{HS}$ = 0.59, ΔF_{sim} = 0.0495).

Third, underpredictions were found for schemes with many sires and n_o = 8 or 16. Table 4 shows the prediction errors for N_m = 80, d = 2 and n_o = 16, where errors up to –19% were found. These were the largest errors encountered throughout the whole range evaluated. To identify the origin of the underprediction, components of Equation 1 were estimated from simulated data (for the linear model) and ΔF was predicted from Equation 1 using those estimates (See also Table 4). However, this did not remove the underprediction. This indicates that components of Equation 1 were predicted accurately for the linear model, but the linear model is insufficient for predicting ΔF when the number of parents is large, irrespective of ΔF.

The accuracy of predictions for schemes that are not included in Table 2, 3 or 4 showed values in the range of the schemes presented in the Tables. For example for N_m = 40, d = 2 and n_o = 4, predictions errors were –9%, –7%, –3% and -5% for h^2 = 0.1, 0.2, 0.4 and 0.6 respectively. The average error for schemes with $N_m \geq 40$ was –10%.

Contribution of the term for deviations from Poisson to E[ΔF]: The prediction procedure would be simplified considerably if the term for deviations from Poisson could be ignored or

TABLE 4.– Rates of inbreeding from simulation (ΔF_{sim}) and corresponding prediction errors for a population with 80 sires, 160 dams and 16 offspring per dam

h^2	ΔF_{sim}[c]	error%[a]	Error%[b]
0.1	0.0210	−17	−19
0.2	0.0171	−19	−19
0.4	0.0114	−17	−16
0.6	0.0079	−15	−12

h^2, initial heritability; [a] error from full prediction, [b] error with components of Equation 1 estimated from simulation; error% = 100% × (ΔF_{pred} − ΔF_{sim})/ΔF_{sim}; ΔF_{pred}, predicted rate of inbreeding from Equation 1 with linear model. [c]Standard errors of ΔF_{sim} were ≤ 1% of the estimate.

TABLE 5.–Rates of inbreeding from simulation (ΔF_{sim}) and corresponding prediction errors with (error%[a]) or without (error%[b]) the correction for deviations from Poisson

n_o	h^2	ΔF_{sim}[c]	error%[a]	error%[b]
4	0.1	0.0184	−9	−25
	0.2	0.0171	−6	−14
	0.4	0.0151	−6	−4
	0.6	0.0130	−5	+5
16	0.1	0.0602	+9	−52
	0.2	0.0511	+2	−42
	0.4	0.0374	−4	−32
	0.6	0.0280	−9	−24

For N_m = 20, N_f = 40; h^2, initial heritability; n_o number of offspring per dam; [a,b]error% = 100% × (ΔF_{pred} − ΔF_{sim})/ΔF_{sim}; [c]Standard errors of ΔF_{sim} were ≤ 1% of the estimate.

simplified. Therefore, ΔF was predicted omitting this term. Prediction errors in Table 5 reveal that the term for deviations from Poisson showed positive values in most cases and became very large for schemes with large n_o and low h^2. For the schemes in Table 5, the term for deviations from Poisson contributed up to 55% of the total value. For $N_m > 20$, $n_o = 16$ and $h^2 = 0.1$ (data not shown) even larger contributions were found. These results show that the term for deviations from Poisson needs to be included. The large values of the term for deviations from Poisson are due to remaining correlations between selection probabilities of sibs after conditioning on the linear effect of the selective advantage (See DISCUSSION).

We investigated whether the term for deviations from Poisson can be simplified by ignoring any terms due to β, in which case Equation B32 and B33 can be omitted. However, this increased the underprediction for schemes with $N_m > 20$, $n_o = 16$ and $h^2 = 0.1$ or 0.2, by *approx.* 8% and 4% respectively. For example, for the schemes in Table 4, prediction errors became −25%, −23%, −18% and −16%. For schemes with $n_o = 4$ or schemes with $h^2 > 0.2$ prediction errors were only slightly affected. Therefore, Equation B32 and B33 are required only for schemes with $n_o > 4$, $N_m > 20$ and $h^2 \leq 0.2$.

Quadratic model: For schemes where the linear model showed underprediction, ΔF was predicted using the quadratic model with components of Equation 1 estimated from simulation. Table 6 shows results for the same

TABLE 6.–Rates of inbreeding from simulation (ΔF_{sim}) and corresponding prediction error for the quadratic model with 80 sires, 160 dams and 16 offspring per dam

h^2	ΔF_{sim}[c]	error%
0.1	0.0210	−7
0.2	0.0171	−6
0.4	0.0114	−5
0.6	0.0079	−4

h^2, initial heritability; error% = 100% × (ΔF_{pred} − ΔF_{sim})/ΔF_{sim}; ΔF_{pred}, predicted rate of inbreeding with components of Equation 1 estimated from simulation. [c]Standard errors of ΔF_{sim} were ≤ 1% of the estimate.

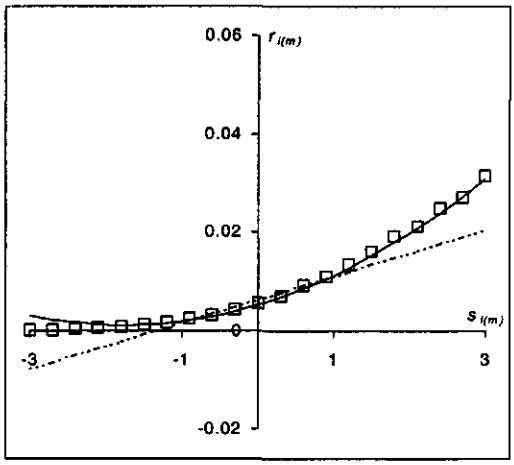

FIGURE 1.–Relation between the genetic contribution ($r_{i(m)}$) and the selective advantage ($s_{i(m)}$) for sires, with N_m = 20, d = 2, h^2 = 0.4 and n_o = 4. ....., linear model; ——, quadratic model; □, observed in simulated data. Note that $s_{i,(m)}$ is in *sd* units.

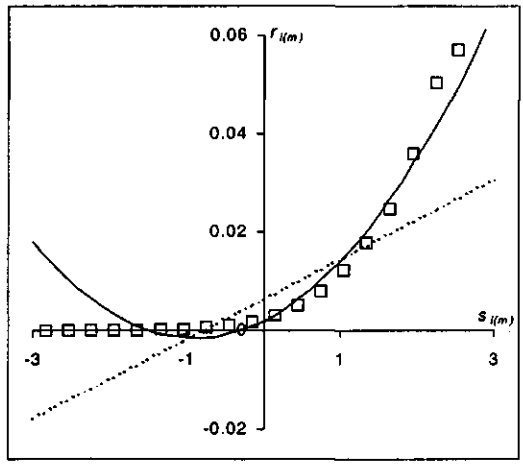

FIGURE 2.–Relation between the genetic contribution ($r_{i(m)}$) and the selective advantage ($s_{i(m)}$) for sires, with N_m = 20, d = 2, h^2 = 0.4 and n_o = 16. ....., linear model; ——, quadratic model; □, observed in simulated data. Note that $s_{i(m)}$ is in *sd* units.

schemes as in Table 4, which reveals a reduction of the prediction error from a maximum of –19% for the linear model to a maximum of –7% for the quadratic model. For schemes with $N_m \geq 40$, the average relative error was only –2% with a standard deviation of 3% for the quadratic model, whereas for the linear model the average relative error was –11% with a standard deviation of 5%.

Goodness of fit: Figure 1 and 2 show the relationship between the selective advantage ($s_{i(m)}$) and the genetic contribution for sires, from the linear model, from the quadratic model and the relationship observed in the simulated data, for N_m = 80, d = 2, h^2 = 0.4 and n_o = 4 or 16. For the linear model, predicted β was almost identical to β estimated from the simulations and, therefore, only the predicted relationship is presented. For n_o = 4 (Figure 1), there is a relatively small difference between the linear and the quadratic fit and the linear model showed only –3% error. (*Approx.* 95% of the individuals was within ± 2sd, so deviations outside this range have limited impact). For n_o = 16 (Figure 2) there is substantial non-linearity, the quadratic fit is better than the linear fit (*e.g.*, the linear model assigned negative contributions to all individuals below –0.8 sd.). For this scheme the linear prediction showed –17% error *vs.* –6% for the quadratic model.

Comparing Figure 1 and 2 shows that with increasing selection intensity, the contributions are increasingly affected by the selective advantage (*i.e.*, the slope of the linear fit increases) and that for positive values of the selective advantage the slope becomes steeper, whereas for negative values the slope becomes flatter. For example, for n_o = 16, all individuals with a negative selective advantage are

TABLE 7.–Relation between the rate of inbreeding (ΔF_{sim}) and the number of parents

	$h^2 = 0.1, n_o = 16$		$h^2 = 0.6, n_o = 4$	
N_m	ΔF_{sim}[b]	Red.[a]	ΔF_{sim}[b]	Red.[a]
5	0.1252	-	0.0483	-
10	0.0917	0.27	0.0253	0.48
20	0.0602	0.34	0.0130	0.49
40	0.0364	0.40	0.0066	0.49
80	0.0210	0.42	0.0033	0.50

For a mating ratio of 2; N_m, number of sires; h^2, initial heritability; n_o, number of offspring per dam; [a]Red = reduction, *e.g.* (0.1252–0.0917)/0.1252 = 0.27; [b]Standard errors of ΔF_{sim} were ≤ 1% of the estimate.

expected to make the same (*i.e.*, almost zero) genetic contribution, whereas for individuals with a positive selective advantage the genetic contribution increases rapidly with the selective advantage. For the schemes in Figure 1 and 2, respectively 31% and 68% of the selected sires made no long-term contribution at all. For low heritabilities the non-linearity was even more extreme, *e.g.*, for $N_m = 80$, $N_f = 160$, $n_o = 16$ and $h^2 = 0.1$, 83% of the selected sires had zero long-term contribution. (The linear model predicted negative contributions for *approx.* 20% of the sires). Not surprisingly, this scheme gives an extremely large rate of inbreeding, $\Delta F_{sim} = 0.0210$ (Table 4), almost the 10-fold of random selection and 6.6 times the rate of inbreeding with mass selection.

Relationship between ΔF and population parameters

Relationship between ΔF and the number of parents: Table 7 shows the relationship between ΔF and the number of sires, for $d = 2$. In the absence of selection, $E(\Delta F) \approx 1/(8N_m) + 1/(8N_f)$ (FALCONER and MACKAY, 1996), showing that, in the absence of selection, the rate of inbreeding halves when the number of parents is doubled. However, for BLUP selection with $h^2 = 0.2$ and $n_o = 16$, the rate of inbreeding less than halves when doubling the number of parents. For example, when N_m increased from 5 to 10, ΔF reduced by only 27%. For $h^2 = 0.6$ and $n_o = 4$ the reduction was closer to 50%. SANTIAGO and CABALLERO (1995) observed a similar pattern for mass selection, but here the effect is much larger.

The difference in the effect of doubling the number of parents with and without selection is due to the effect of a finite number of families on the intraclass correlation between sibs and on the variance of family size. For example, when N_m decreased from 80 to 5, the intra-class correlations between sibs decreased from $\bar{\rho}_{FS} = 0.86$ and $\bar{\rho}_{HS} = 0.55$ to $\bar{\rho}_{FS} = 0.78$ and $\bar{\rho}_{HS} = 0.40$ for schemes with $h^2 = 0.1$ and $n_o = 16$. This reduction of the intra-class correlation was accurately predicted using the current method (equations APPENDIX A). Additionally, for schemes with $N_m = 5$ or 10, the correction of the selected proportions (Equation 16) further reduces $V_{n(x)}$ and this reduction is greater with higher intra-class correlation, which reduces ΔF proportionally more for schemes with large emphasis on family information (*i.e.*, large n_o and low h^2). For such schemes, increasing the number of parents is an inefficient way of reducing ΔF.

Relationship between ΔF and mating ratio: With random selection, ΔF decreases when the number of sires is kept constant and the number

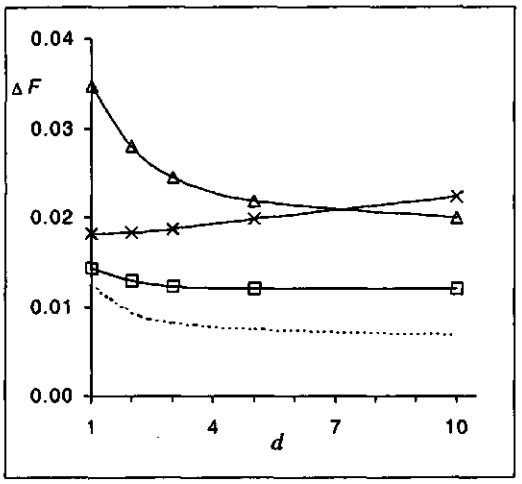

FIGURE 3.–Relation between the rate of inbreeding (ΔF) and the mating ratio (d) for N_m = 20. ....., random selection; —□—, n_o = 4, h^2 = 0.6; —×—, n_o = 4, h^2 = 0.1; —Δ—, n_o = 16, h^2 = 0.6.

of dams is increased. BIJMA and WOOLLIAMS (1999) found a similar pattern for mass selection. Figure 3 shows the relationship between ΔF and the mating ratio for BLUP selection with N_m = 20. (Note that n_o remains constant.) The dotted line represents random selection with $\Delta F = 1/(8N_m) + 1/(8N_f)$ and serves as a reference. Surprisingly, for n_o = 4 and h^2 = 0.1, Figure 3 shows an increase of ΔF when N_f increases. This increase is due to an increased male selection intensity when d increases, *i.e.*, for d = 1, $p_m = 20/(20\times1\times4\times½)$ = 0.5 → i_m = 0.798, whereas for d = 10, p_m = 0.05 → i_m = 2.063. An increased selection intensity results in an increased Λ (see Equations B1 and B3) which increases the term $E(u^2_{i(m)})$ in Equation 1. Additionally, decreased selected proportions result in an increased variance of family size, increasing δ_m. Together, both effects more than compensate the reduction of the terms due to dams (*i.e.*, $N_f E[u^2_{i(f)}]$ and $N_f\delta_f$, which are *approx.* proportional to $1/N_f$) for schemes with low h^2 and low n_o. For high h^2, the effect of selection intensity on the rate of inbreeding is smaller and consequently there was only a small effect of d on ΔF for n_o = 4 and h^2 = 0.6. For n_o = 16 and h^2 = 0.6 the relationship was similar to random selection. For schemes with high n_o, selection intensity among males is already reasonably large, so the increase of i_m with d is limited. Therefore, those schemes showed a decrease of ΔF with increasing d (*e.g.*, this was found for N_m = 20, n_o = 16 and h^2 = 0.1, data not shown). When instead of n_o the total number of offspring was kept constant [*i.e.*, by using $n_o = T/(N_m d)$], the rate of inbreeding always decreased with increasing d.

Discussion

This article has presented a method to predict rates of inbreeding in populations with discrete generations undergoing BLUP selection, which has not been possible up till now. The method is based on the concept of long-term genetic contributions (WRAY and THOMPSON, 1990), using the recently established relationship between rates of inbreeding and expected genetic contributions (WOOLLIAMS and BIJMA, 1999) and the method of WOOLLIAMS *et al.* (1999) for predicting expected genetic contributions.

Quantitative genetics theory: The results have verified the theory developed by WOOLLIAMS and BIJMA (1999), showing that the simple form of the relationship between ΔF

and expected contributions derived in that article for random mating can be applied to challenging selection indices. Examination of the results showed that this relation (*i.e.*, Equation 1) was accurate over a range of ΔF from 0.3 to 12.5%. Even where significant errors were encountered, further examination showed that Equation 1 remained accurate, and that the inaccuracies were due to inadequate parameterization of the models used to implement Equation 1. The issues surrounding the parameterization are discussed later.

Combining the theory of WOOLLIAMS and BIJMA (1999) and WOOLLIAMS *et al.* (1999) enables a general approach to the prediction of rates of inbreeding in selected populations, because it allows for a general model for the inheritance of selective advantage (*i.e.*, Π) and accounts for all generations of inheritance of selective advantage. SANTIAGO and CABALLERO (1995) developed predictions of ΔF for mass selection, using the proportion of additive genetic variance transmitted to offspring as a model for the inheritance of selective advantage. With BLUP or various other selection strategies (*e.g.*, sib-index selection, see WOOLLIAMS and BIJMA, 1999), however, the proportion of genetic variation transmitted to offspring is insufficient to model the inheritance of selective advantage.

Other methods, for example the method of BURROWS (1984a,b), have accounted for only a single generation of inheritance of selective advantage and, therefore, systematically underpredict ΔF (WRAY *et al.* 1990). This can be illustrated by extending the method of BURROWS (1984a,b) to BLUP selection and two sexes (APPENDIX C), which shows a systematic underprediction of on average –16% for $N_m = 20$ (Table 8) to –28% for $N_m = 80$ (not shown). WRAY *et al.* (1990) investigated methods accounting for two generations inheritance of selective advantage. Those methods, however, still rely on simulation to calculate the variance of family size.

TABLE 8.–Prediction errors (%) of rates of inbreeding from an extension of Burrow's method for a population with 20 sires[a]

N_f	h^2	$n_o = 4$	$n_o = 8$
40	0.1	–24	–26
	0.2	–23	–23
	0.4	–21	–22
	0.6	–16	–20
100	0.1	–25	–20
	0.2	–24	–20
	0.4	–20	–18
	0.6	–17	–16

[a]values are $100\% \times (\Delta F_{pred} - \Delta F_{sim})/\Delta F_{sim}$; for corresponding ΔF_{sim} see Table 3; N_f, number of dams; h^2, initial heritability; n_o, number of offspring per dam.

The application of the theory of WOOLLIAMS and BIJMA (1999) to BLUP selection brings our understanding of the forces governing the rate of loss of genetic variation in selected populations into line with our ability to carry out genetic evaluations within populations. However, MEUWISSEN (1997) and GRUNDY *et al.* (1998) have shown that a higher selection response over the medium to long term at the same rate of loss of genetic variation may be obtained by using a quadratic index. In such indices, the rate of inbreeding is constrained directly and does not require prediction. However, for such selection indices, only approximate predictions of rates of gain are available (GRUNDY *et al.* 1998) and it would seem that, in order to keep pace with the

development of breeding technology, quantitative genetic theory may need to return to the problem of predicting rates of gain, rather than rates of inbreeding.

Execution of the methods: In all prediction methods, the models frequently have to strike a balance between simplicity of concept execution and accuracy. In this article, the simplicity of concept lies in the formulae for ΔF and the methods used for their execution being based on regression models that are established within standard selection index theory. The high rates of inbreeding and the high intraclass correlations of sibs tested the predictive methods to the full.

The principle decision affecting the execution is the choice of the selective advantage. This involves two issues, first what components are included in the set of selective advantages and second, whether a linear component is sufficient. Ideally, the selective advantage should account for all covariances between selection probabilities of sibs, so that (*i*) off-diagonal elements of $\Delta \mathbf{V}_{n(x)}$ are zero and the term for deviations from Poisson simplifies to $-1/(8T)$ with constant n_o (BIJMA and WOOLLIAMS, 1999), and (*ii*) generations are independent.

To make selection probabilities of sibs independent, a natural way forward is to condition on selective advantages so as to remove any covariance between EBVs of sibs. With sib-index or BLUP selection, removing covariances between EBVs of sibs involves more than conditioning only on the breeding value of the parent, because the average environmental effect of full and half-sib

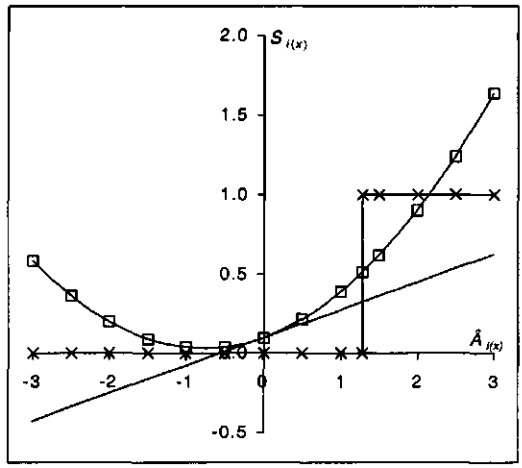

FIGURE 4.–Relation between selection score ($S_{i(x)}$) and estimated breeding value ($\hat{A}$) for a selected proportion of 10%; ——, from linear model ($E[S_{i(x)}] = p_x + i_x p_x \sigma^{-1}_{\hat{A}} \hat{A}_{i(x)}$); —□—, from quadratic model ($E[S_{i(x)}] = p_x + i_x p_x \sigma^{-1}_{\hat{A}} \hat{A}_{i(x)} + \frac{1}{2} i_x p_x t_x \sigma^{-2}_{\hat{A}} \hat{A}^2_{i(x)}$) t = standardized truncation point) ; —×—, true selection score. Note that $\hat{A}$ is in *sd* units.

families also contributes to this covariance. Therefore, an alternative parameterization was tested which has selective advantages consisting of EBV and prediction errors of the parent (see WOOLLIAMS *et al.* 1999), and fully accounts for correlations between the EBV of sibs. However, this alternative was substantially more complicated and resulted in similar accuracy of prediction (results not shown), indicating that residual correlations between EBVs of sibs are not the principle source of the observed underprediction.

Second, and more important, when using a linear model, it is assumed that the conditional mean of the number of selected offspring is a linear function of the selective advantage (*i.e.*, the Λ model). However, the relation between the number of selected offspring and the

selective advantage of the parent is non-linear. The non-linearity originates from the non-linear relation between the EBV of the offspring and its selection score, as shown in Figure 4. Therefore, when assuming that the conditional mean is linear in the selective advantage, sibs will have prediction errors of their selection score (S = 0 or 1, not selected or selected) in common, making their selection probabilities dependent. Thus, although a linear model may fully remove any covariance between EBVs of sibs, the non-linear relation between selection probabilities and EBVs implies that a covariance between the selection probabilities of sibs will remain.

The degree of non-linearity in the relation between selective advantage and the selection probability will depend on the intraclass correlation between EBV of sibs. With mass selection, where the intraclass correlation is low, off-diagonal elements of $\Delta\mathbf{V}_{n(x)}$ can be ignored and the principle purpose of the term for deviations from Poisson is to take account of the fixed value of n_o instead of $n_o \sim$ Poisson (BIJMA and WOOLLIAMS, 1999). With mass selection and fixed n_o, therefore, the term for deviations from Poisson takes negative values, whereas with BLUP selection, off-diagonals of $\Delta\mathbf{V}_{n(x)}$ can be substantial due to residual covariances between sibs not accounted for by the linear model, giving large positive values of the term for deviations from Poisson (Table 5).

The observed non-linearity prompted the consideration of a fully deterministic quadratic model to describe the relationship between the selective advantage and genetic contributions. This proved difficult, since it involves third and fourth moments of truncated multivariate normal distributions. Using a result of TALLIS (1961), accurate predictions of $\sigma^2_{s_1}$ through $\sigma^2_{s_4}$ (see Equations 14 through 17) were obtained (results not shown), but prediction of Λ and Π for the quadratic model remained complex. Prediction of Π, for example, requires the multiple regression of the linear and squared term of the breeding value of a selected offspring on the linear and squared term of the breeding value of the parent, where truncation on the EBV of both parent and offspring has to be taken into account. Hence, this involves 4-variate truncated normal distributions and one needs to derive terms such as $E(A^4_{\text{parent}})$ and $E(A^3_{\text{parent}} A_{\text{offsp.}})$, which are difficult. Although the quadratic model shows a better fit than the linear model (*e.g.*, in Figure 4, the linear model assigns a negative selection score to all individuals below −0.6 *sd.*), it does not fully remove errors in common between sibs. Hence, the term for deviations from Poisson is also needed with the quadratic model.

Given that it is not generally feasible to formulate a selective advantage that removes all covariances between selection probabilities of sibs, the term for deviations from Poisson needs to be included in the general case. When including the Poisson correction, the non-linearity is fully accounted for when calculating $\Delta\mathbf{V}_{n(x)}$ (Equation B9 to B29). However, for the calculation of δ_x, the model for predicting genetic contributions is required again (*i.e.*, Equation 10). The simplest solution is to ignore selective advantage so that $u^*_{j(y)} = \alpha_y$ and δ_x =

$\alpha^T\Delta V_{n(x)}\alpha$, which was used by WOOLLIAMS and BIJMA (1999) for index selection. In that case, the non-linear relationship between EBV and selection score is taken into account when calculating $\Delta V_{n(x)}$, but inheritance of the non-linear part is ignored, resulting in underprediction. For selection on a sib index, WOOLLIAMS and BIJMA (1999) noticed that the underprediction disappeared when selective advantage was not inherited (*i.e.*, $h^2 = 0$). This confirms that the underprediction is not due to the non-linearity *per se*, but due to the necessity of providing a tractable model for the inheritance of selective advantage. The impact of non-linearity was also observed by SANTIAGO and CABALLERO (1995) in their study of rates of inbreeding with mass selection, where the sib correlations encountered were much lower than with BLUP.

Implications: The results indicate that, with BLUP selection, relationships between ΔF and population parameters differ qualitatively from random or mass selection, the main difference being the dominant role of selection intensity compared to the number of parents. For example, with $N_m = 20$, $d = 2$, $n_o = 4$ and $h^2 = 0.1$, simultaneously *increasing* the number of parents and the number of offspring per dam by a factor of four (giving $N_m = 80$, $d = 2$, $n_o = 16$) *increases* the rate of inbreeding from 0.0184 to 0.0210. This shows that the number of candidates per parent may be as or more important than the number of parents. Furthermore, doubling the number of parents fails to halve the rate of inbreeding and, although this was remarked upon by SANTIAGO and CABALLERO (1995) in the context of mass selection, with BLUP the impact of doubling the number of parents is even less, and substantially so. Increasing the number of dams may even increase the rate of inbreeding in particular cases. The dominant role of selection intensity compared to the number of parents with BLUP selection will change perceptions about procedures and designs of breeding schemes to effectively reduce rates of inbreeding.

By understanding the forces governing the rate of inbreeding, perceptions may also be changed upon the desirability of naïve selection on EBVs. The results showed that in some cases 83% of the selected sires failed to contribute in the long-term, which seems to be a waste of resources. This is an indication of the inefficiency of BLUP selection compared to more advanced procedures (MEUWISSEN, 1997; GRUNDY *et al*, 1998). Thus, unless the time horizon is limited to a single generation, it is better to incorporate the good genes from all the parents, rather than spending substantial effort raising offspring from the parents that will not contribute. The cost of raising offspring from parents that are destined not to contribute in the long-term is a hidden cost of the high rates of inbreeding associated with BLUP selection. Nevertheless, for breeding schemes where BLUP selection is being conducted the methodology developed here allows the design of such schemes to maximize genetic gain for a fixed rate of inbreeding [\see for example VILLANUEVA *et al.* (1997) for sib selection]. The systematic nature of the underprediction for schemes with $N_m > 20$ allows for a rule of thumb to correct these predictions. When 11%

is added to the predicted values, all predictions are within ±8% of the simulation results. This is simplistic, but may prove valuable for practical purposes and holds for the wide range of alternatives investigated in this article.

Extensions: The current prediction procedure can be extended directly to populations with multi-trait BLUP selection, using a multi-trait pseudo BLUP index [VILLANUEVA *et al.* (1993) for animal breeding and KERR (1998) for tree breeding], or where the heritability in BLUP evaluations is artificially increased to avoid excessive inbreeding (GRUNDY, 1994). Neither of these extensions requires the development of new theory. With multi-trait selection, the selective advantage may consist of the sum of the true breeding values for the respective traits weighted by their economic value, and the EBV may be replaced by the estimate of the aggregate genotype.

The current procedure can also be extended to populations with overlapping generations, in a similar way as with mass selection (BIJMA *et al.* 1999). BIJMA *et al.* (1999), however, assumed that the phenotype of selection candidates was recorded only once and remained unchanged afterwards. In that case, the number of different categories is twice the number of age classes, *i.e.*, one category for each age-sex combination. With BLUP selection, however, the EBV of selection candidates may change, *e.g.*, when information on their offspring is included in their EBV. This will alter the ranking of the selection candidates throughout their life. Therefore, with BLUP selection, individuals may be selected in any combination of age classes, which increases the number of categories compared to a static index. Additionally, the expected number of individuals in every category needs to be derived, which involves multivariate normality. However, the number of categories is merely a problem of enumeration, and methods to calculate multivariate normal probabilities, needed for the calculation of the expected number of individuals in every category, are available [TALLIS (1961), MENDEL and ELSTON (1974) or DUTT and SOMS (1976), see also DUCROCQ and COLLEAU (1986)].

This research was financially supported by the Netherlands Technology Foundation (STW) and was coordinated by the Life Science Foundation (ALW). JAW gratefully acknowledges the Ministry of Agriculture, Fisheries and Food (United Kingdom) for financial support. Fetsje Bijma is acknowledged for giving helpful suggestions and Johan Van Arendonk for giving PB the opportunity to visit JAW.

Literature Cited

BIJMA, P., and J. A. W. WOOLLIAMS, 1999 Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151: 1197-1210.

BIJMA, P., J. A. M. VAN ARENDONK and J. A. WOOLLIAMS, 1999 A general procedure to predict rates of inbreeding in populations undergoing mass selection. Genetics 154: 1865-1877.

BULMER, M. G., 1971 The effect of selection on genetic variability. Am. Nat. 105: 201-211.

BURROWS, P. M., 1984a Inbreeding under selection from unrelated families. Biometrics 40: 357-366.

BURROWS, P. M., 1984b Inbreeding under selection from related families. Biometrics 40: 895-906.

DEMPSTER, E. R., and I. M. LERNER, 1950 Heritability of threshold characters. Genetics 35: 212-236.

DUCROCQ, V., and J. J. COLLEAU, 1986 Interest in quantitative genetics of DUTT's and DEAK's methods for numerical computation of multivariate normal probability integrals. Genet. Sel. Evol. 18: 447-474.

DUTT, J. E., and A. P. SOMS, 1976 An integral representation technique for calculating general multivariate probabilities with an application to multivariate χ^2. Commun. Stat. 11: 2195-2205.

FALCONER, D. S., and T. F. C. MACKAY 1996 *Introduction to quantitative genetics*. Longman, Essex, UK.

GJØEN, H. M., and B. GJERDE, 1998 Comparing breeding schemes using individual phenotypic values and BLUP breeding values as selection criteria. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Vol. 27, Univ. New England, Armidale, NSW, Australia, pp. 111-114.

GOFFINET, B., 1983 Selection on selected records. Genet. Sel. Evol. 15: 91-98.

GRUNDY, B., A. CABALLERO, E. SANTIAGO and W. G. HILL, 1994 A note on using biased parameter values and non-random mating to reduce rates of inbreeding in selection programmes. Anim. Prod. 59: 465-468.

GRUNDY, B., VILLANUEVA, B. and J. A. WOOLLIAMS, 1998 Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. Camb. 72: 159-168.

HENDERSON, C. R., 1963 Selection index and the expected genetic advance. *In* W. D. HANSON and H. F. ROBINSON (eds.), Statistical genetics and plant breeding, pp. 141-163. Natl. Acad. Sci., Natl. Res. Council Publ. No. 982, Washington, D. C.

HENDERSON, C. R., 1975 Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423–447.

JAMES, J. W., 1987 Determination of optimal selection policies. J. Anim. Breed. Genet. 104: 23-27.

KEIGHTLEY, P. D., and W. G. HILL, 1997 Directional selection and variation in finite populations. Genetics 117: 573-582.

KENDALL, M. G., and A. STUART, 1963 The advanced theory of statistics. Charles Griffin & Company, Ltd., London.

KERR, R. J., 1998 Asymptotic rates of response from forest tree breeding strategies using best linear unbiased prediction. Theor. Appl. Genet. 96: 484-493.

LYNCH, M., and W. G. HILL, 1986 Phenotypic evolution and neutral mutation. Evolution 40: 915-935.

LYNCH, M., and B. WALSH, 1998 Genetics and analyses of quantitative traits. Sinauer Associates Inc. Sunderland, USA.

MENDEL, N.R., and R. C. ELSTON, 1974 Multifactorial qualitative traits: Genetic analyses and prediction of recurrence risks. Biometrics 30: 41-57.

MEUWISSEN, T.H.E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

PEARSON, K., 1903 On the influence of natural selection on the variability and correlation of organs. Roy. Soc. London Phil. Trans. A, Vol. 200: 1-66.

ROBERTSON, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. 2: 189-194.

SANTIAGO, E., and A. CABALLERO, 1995 Effective size of populations under selection. Genetics 139: 1013-1030.

TALLIS, G.M., 1961 The moment generating function of the truncated multi-normal distribution. J. R. Statist. Soc. B. 23: 223-229.

VERRIER, E., J. J. COLLEAU and J. L. FOULLEY, 1993 Long-term effects of selection based on the animal model BLUP in a finite population. Theor. Appl. Genet. 87: 446-454.

VILLANUEVA, B. and J. A. WOOLLIAMS, 1997. Optimization of breeding programmes under index selection and constrained inbreeding. Genet. Res. Camb. 69: 145-158.

VILLANUEVA, B., N. R. WRAY and R. THOMPSON, 1993 Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. Anim. Prod. 57: 1-13.

WOOLLIAMS, J. A., and P. BIJMA, 2000 Predicting rates of inbreeding in populations undergoing selection. Genetics, 154: 1851-1864.

WOOLLIAMS, J. A., N. R. WRAY and R. THOMPSON, 1993 Prediction of long-term contributions and inbreeding in populations undergoing mass selection. Genet. Res. Camb. 62: 231-242.

WOOLLIAMS, J. A., P. BIJMA and B. VILLANUEVA, 1999 Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153: 1009-1020.

WRAY, N. R., and R. THOMPSON, 1990 Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55: 41-54.

WRAY, N. R., J. A. WOOLLIAMS AND R. THOMPSON 1990 Methods for predicting rates of inbreeding in selected populations. Theor. Appl. Genet. 80: 503-512.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1994 Prediction of rates of inbreeding in populations

undergoing index selection. Theor. Appl. Genet. 87: 878-892.

Appendix A

Approximate BLUP selection index: Index weights are calculated as $\mathbf{b} = \mathbf{V}^{-1}\mathbf{g}$, where $\mathbf{V}$ is the 6×6 (co)variance matrix of information sources in $\mathbf{x}_i$ and $\mathbf{g}$ is the 6×1 vector of covariances between information sources in $\mathbf{x}_i$ and the breeding value of the candidate. Non-zero elements of $\mathbf{V}$ are: $V_{11} = \sigma^2_m$, $V_{22} = \sigma^2_f(1-1/d)$, $V_{33} = \sigma^2_f/d$, $V_{44} = [\sigma^2_{A,s} + \sigma^2_{A,d}/d + (½\sigma^2_{A,0}+\sigma^2_E)/(n_o d)]$, $V_{55} = [\sigma^2_{A,d} + (½\sigma^2_{A,0}+\sigma^2_E)/n_o](1-1/d)$, $V_{66} = (½\sigma^2_{A,0}+\sigma^2_E)(1-1/n_o)$, $V_{14} = V_{41} = ½\sigma^2_m$, $V_{34} = V_{43} = ½\sigma^2_f/d$ and $V_{25} = V_{52} = ½\sigma^2_f(1-1/d)$; and $\mathbf{g}^T = \{½\sigma^2_m, ½\sigma^2_f(1-1/d), ½\sigma^2_f/d, \sigma^2_{A,s} + \sigma^2_{A,d}/d + ½/(n_o d), [\sigma^2_{A,d} + ½\sigma^2_{A,0}/n_o](1-1/d), ½\sigma^2_{A,0}(1-1/n_o)\}$, where ^{T} denotes the transpose, $\sigma^2_{A,0}$ is the base generation additive genetic variance, σ^2_E is the environmental variance, σ^2_m and σ^2_f are the variance of the EBV among selected sires and dams respectively, which are $\sigma^2_x = \sigma^2_A \rho^2(1-k_x)$ where $\rho = \sqrt{\mathbf{b}^T\mathbf{g}/\sigma^2_A}$; $\sigma^2_{A,s}$ and $\sigma^2_{A,d}$ are the between sire and between dam family additive genetic variance, $\sigma^2_{A,s} = ¼\sigma^2_A(1-k_m\rho^2)$ and $\sigma^2_{A,d} = ¼\sigma^2_A(1-k_f\rho^2)$. The variance of EBV is $\sigma^2_{\hat{A}} = \rho^2\sigma^2_A$. Each generation, additive genetic variance is calculated from $\sigma^2_A = \sigma^2_{A,s} + \sigma^2_{A,d} + ½\sigma^2_{A,0}$. The above equations are iterated until equilibrium variances are reached (*approx.* 5 iterations).

Intraclass correlations of sibs are corrected for the number of families being finite using an empirical analogy of the sample mean of the bivariate correlation coefficient, which is, $\bar{\rho} \approx \rho - \rho(1-\rho^2)/2n$ (KENDALL and STUART, 1963, p. 390). For the current two-way classification this is extended to $\bar{\rho}_{sibs} = \rho_{sibs} - \rho_{sibs}(1-\rho^2_{sibs})(a/N_m + b/N_f)$, where $\bar{\rho}_{sibs}$ is the sample mean of the intraclass correlation and ρ_{sibs} is the true mean, calculated from $\rho_{FS} = \mathbf{b}^T\mathbf{C}_{FS}\mathbf{b}/\sigma^2_{\hat{A}}$ and $\rho_{HS} = \mathbf{b}^T\mathbf{C}_{HS}\mathbf{b}/\sigma^2_{\hat{A}}$, where $\mathbf{C}_{FS}$ and $\mathbf{C}_{HS}$ are the 6×6 covariance matrices of information sources of full-sibs and of half-sibs respectively. Matrices $\mathbf{C}_{FS}$ and $\mathbf{C}_{HS}$ are identical to $\mathbf{V}$, except for $\mathbf{C}_{FS}(6,6) = -(½\sigma^2_{A,0}+\sigma^2_E)/n_o$, $\mathbf{C}_{HS}(2,2) = -\sigma^2_f/d$, $\mathbf{C}_{HS}(5,5) = -[\sigma^2_{A,d} + (½\sigma^2_{A,0}+\sigma^2_E)/n_o]/d$, $\mathbf{C}_{HS}(2,5) = \mathbf{C}_{HS}(5,2) = -½\sigma^2_f/d$ and $\mathbf{C}_{HS}(6,6) = 0$. Coefficients a and b were determined empirically using simulated data, resulting in $a = 0.8634$, $b = 0.9540$ for full-sibs and $a = 1.4075$, $b = 1.4581$ for half-sibs.

Appendix B

Linear model

Elements of Λ: Elements of Λ, λ_{xy}, are the regression coefficient of the number of selected offspring of sex x on the selective advantage of the parent of sex y. A general procedure to derive Λ is in WOOLLIAMS *et al.* (1999). For the current purpose, single regressions instead of multiple regression can be used because elements of the selective advantage are independent. Elements are, $\lambda = b_{\hat{A}_o,s_p}\, b_{S_o,\hat{A}_o}$, where $b_{\hat{A}_o,s_p}$ is the regression coefficient of the EBV of the offspring on the selective advantage of the parent and $b_{S_o,\hat{A}_o}$ is the regression coefficient of the selection score ($S_o = 0$ or 1, *i.e.*, not selected or selected) of the offspring on its EBV. The first regression coefficient is, $b_{\hat{A}_o,s_p} = \mathbf{b}^T\mathbf{c}_x/\mathrm{Var}(s_{i(x)})$, where $\mathbf{c}_x$ is the 6×1 vector of

covariances between $\mathbf{x}_i$ of the offspring and $s_{i(x)}$ of the parent of sex x. The second regression coefficient is, $b_{S_o,\hat{A}_o} = i_o/\sigma_{\hat{A}}$ (Robertson, appendix in DEMPSTER and LERNER, 1951), where i_o is the selection intensity for the offspring. Resulting equations are:

$$\lambda_{11} = [\mathbf{b}^T\mathbf{c}_m i_m/\sigma_{\hat{A}}]/\sigma^2_{s(m)}, \quad \text{(B1)}$$

$$\lambda_{12} = [\mathbf{b}^T\mathbf{c}_f i_m/\sigma_{\hat{A}}]/\sigma^2_{s(f)}, \quad \text{(B2)}$$

$$\lambda_{21} = [\mathbf{b}^T\mathbf{c}_m i_f/\sigma_{\hat{A}}]/\sigma^2_{s(m)}, \quad \text{(B3)}$$

$$\lambda_{22} = [\mathbf{b}^T\mathbf{c}_f i_f/\sigma_{\hat{A}}]/\sigma^2_{s(f)}, \quad \text{(B4)}$$

where $\mathbf{c}_m^T = [\sigma_m^2(1-1/N_m),\ 0,\ \sigma_f^2(1-1/N_m)/d,\ \tfrac{1}{2}\sigma^2_{s(m)},\ 0,\ 0]$ and $\mathbf{c}_f^T = [\sigma_m^2(1-1/N_m),\ \sigma_f^2(1-1/d),\ \sigma_f^2(1-1/N_f)/d,\ \tfrac{1}{2}\sigma^2_{s(m)},\ \tfrac{1}{2}\sigma_A^2(1-k_f\rho^2)(1-1/d),\ 0]$.

Elements of Π: Elements of Π, π_{xy}, are the regression coefficient of the selective advantage of a selected offspring of sex x on the selective advantage of the parent of sex y. A general procedure to derive Π is in WOOLLIAMS *et al.* (1999). As with A, single regressions can be used so that $\pi_{op} = [\mathrm{Cov}(s_p,s_o) - \mathrm{Cov}(s_p,\hat{A}_o)\mathrm{Cov}(s_o,\hat{A}_o)k_o/\sigma^2_{\hat{A}}]/\mathrm{Var}(s_p)$, where the subscripts p and o denote parent and offspring, and (co)variances are taken before selection of the offspring. With $\mathrm{Cov}(s_p,\hat{A}_o) = \mathbf{b}^T\mathbf{c}_x$ and $s_{i(x)}$ from Equations 2 and 3, resulting equations are:

$$\pi_{11} = \tfrac{1}{2} - \mathbf{b}^T\mathbf{c}_m k_m/\sigma^2_{s(m)}, \quad \text{(B5)}$$

$$\pi_{21} = \tfrac{1}{2} - \mathbf{b}^T\mathbf{c}_m k_f/\sigma^2_{s(m)}, \quad \text{(B6)}$$

$$\pi_{12} = \tfrac{1}{2} - \mathbf{b}^T\mathbf{c}_f k_m/\sigma^2_{s(f)}, \quad \text{(B7)}$$

$$\pi_{22} = \tfrac{1}{2} - \mathbf{b}^T\mathbf{c}_f k_f/\sigma^2_{s(f)}. \quad \text{(B8)}$$

Calculation of δ_x : Calculation of δ_m and δ_f requires the calculation of $\Delta\mathbf{V}_{n(x)}$ and further development of Equation (12) and (13).

Calculation of $\Delta\mathbf{V}_{n(x)}$: Following the approach outlined in the main text and using appendix E of WOOLLIAMS and BIJMA (1999),

$$\Delta\mathbf{V}_{n(m)}(1,1) = \mathrm{E_s}\{n_{i*}(m)[n_{i*}(m)-1]\} - \mathrm{E_s}[\mu_m^2(m)], \quad \text{(B9)}$$

$$\Delta\mathbf{V}_{n(m)}(1,2) = \Delta\mathbf{V}_{n(m)}(2,1) = \mathrm{E_s}[n_{i*}(m)n_{i*}(f)] - \mathrm{E_s}[\mu_m(m)\mu_m(f)], \quad \text{(B10)}$$

$$\Delta\mathbf{V}_{n(m)}(2,2) = \mathrm{E_s}\{n_{i*}(f)[n_{i*}(f)-1]\} - \mathrm{E_s}[\mu_m^2(f)], \quad \text{(B11)}$$

$$\Delta\mathbf{V}_{n(f)}(1,1) = \mathrm{E_s}\{n_{ij}(m)[n_{ij}(m)-1]\} - \mathrm{E_s}[\mu_f^2(m)], \quad \text{(B12)}$$

$$\Delta\mathbf{V}_{n(f)}(1,2) = \Delta\mathbf{V}_{n(f)}(2,1) = \mathrm{E_s}[n_{ij}(m)n_{ij}(f)] - \mathrm{E_s}[\mu_f(m)\mu_f(f)], \quad \text{(B13)}$$

$$\Delta\mathbf{V}_{n(f)}(2,2) = \mathrm{E_s}\{n_{ij}(f)[n_{ij}(f)-1]\} - \mathrm{E_s}[\mu_f^2(f)], \quad \text{(B14)}$$

where $n_{ij}(x)$ is the number of offspring of sex x selected from the i^{th} sire family and the j^{th} dam family, $n_{i*}(x)$ is the total number of offspring of sex x selected from the i^{th} sire family and $\mu_x(y)$ denotes the expected number of selected offspring of sex y of a parent of sex x conditional on its selective advantage, *i.e.*, $\mu_x(y) = \mathrm{E}(n_{i(x)}(y)|s_{i(x)})$. Note that in the main text, $\mu_x(y)$ was abbreviated as $\mathrm{E}(n_i)$. Elements are:

$$\mathrm{E_s}[\mu_m^2(m)] = 1 + \lambda_{11}^2\sigma^2_{s_1} \quad \text{(B15)}$$

$$\mathrm{E_s}[\mu_m(m)\mu_m(f)] = d(1 + \lambda_{11}\lambda_{21}\sigma^2_{s_1}) \quad \text{(B16)}$$

$$\mathrm{E_s}[\mu_m^2(f)] = d^2(1 + \lambda_{21}^2\sigma^2_{s_1}) \quad \text{(B17)}$$

$$\mathrm{E_s}[\mu_f^2(m)] = (1 + \lambda_{12}^2\sigma^2_{s_2})/d^2 \quad \text{(B18)}$$

$$\mathrm{E_s}[\mu_f(m)\mu_f(f)] = (1 + \lambda_{12}\lambda_{22}\sigma^2_{s_2})/d \quad \text{(B19)}$$

$$\mathrm{E_s}[\mu_f^2(f)] = 1 + \lambda_{22}^2\sigma_{s_2} \quad \text{(B20)}$$

$$\mathrm{E_s}\{n_{i*}(m)[n_{i*}(m)-1]\} = d\mathrm{E_s}\{n_{ij}(m)[n_{ij}(m)-1]\} + d(d-1)\mathrm{E_s}[n_{ij}(m)n_{ik}(m)] \quad \text{(B21)}$$

$$\mathrm{E_s}[n_{i*}(m)n_{i*}(f)] = d\mathrm{E_s}[n_{ij}(m)n_{ij}(f)]$$

$$+ d(d-1)E_s[n_{ij}(m)n_{ik}(f)] \quad \text{(B22)}$$

$$E_s\{n_{i*}(f)[n_{i*}(f)-1]\} = dE_s\{n_{ij}(f)[n_{ij}(f)-1]\}$$
$$+ d(d-1)E_s[n_{ij}(f)n_{ik}(f)] \quad \text{(B23)}$$

where

$$E_s\{n_{ij}(m)[n_{ij}(m)-1]\} =$$
$$n_m(n_m-1)N_m(N_m-1)/[T_m(T_m-1)R(p_m,p_m,\overline{\rho}_{FS})] \quad \text{(B24)}$$

$$E_s[n_{ij}(m)n_{ij}(f)] = n_m n_f N_m N_f/[T_m T_f R(p_m,p_f,\overline{\rho}_{FS})] \quad \text{(B25)}$$

$$E_s\{n_{ij}(f)[n_{ij}(f)-1]\} =$$
$$n_f(n_f-1)N_f(N_f-1)/[T_f(T_f-1)R(p_f,p_f,\overline{\rho}_{FS})] \quad \text{(B26)}$$

and

$$E_s[n_{ij}(m)n_{ik}(m)] =$$
$$n_m^2 N_m(N_m-1)/[T_m(T_m-1)R(p_m,p_m,\overline{\rho}_{HS})] \quad \text{(B27)}$$

$$E_s[n_{ij}(m)n_{ik}(f)] = n_m n_f N_m N_f/[T_m T_f R(p_m,p_f,\overline{\rho}_{HS})] \quad \text{(B28)}$$

$$E_s[n_{ij}(f)n_{ik}(f)] =$$
$$n_f^2 N_f(N_f-1)/[T_f(T_f-1)R(p_f,p_f,\overline{\rho}_{HS})], \quad \text{(B29)}$$

where $T_m = T_f = \frac{1}{2}T$ and $n_m = n_f = \frac{1}{2}n_o$ for the current breeding schemes. Furthermore, from MENDEL and ELSTON (1974), $R(p_x,p_y,\overline{\rho}_{sibs}) = p_y/\Phi[(i_x\overline{\rho}_{sibs} - t_y)(1-k_x\overline{\rho}_{sibs}^2)^{-\frac{1}{2}}]$ where Φ is the normal cumulative distribution function and t_y is the standardized truncation point for sex y. When both males and females are involved, the most accurate value is obtained by using $x = m$ and $y = f$ (WRAY *et al.* 1994).

Calculation of Equation (12) and (13): In δ_x, terms due to α are, for sires,

$$\delta_m(\alpha) = \alpha^T\Delta\mathbf{V}_{n(m)}\alpha = \alpha_m^2\,\Delta\mathbf{V}_{n(m)}(1,1) + 2\alpha_m\alpha_f$$
$$\Delta\mathbf{V}_{n(m)}(1,2) + \alpha_f^2\,\Delta\mathbf{V}_{n(m)}(2,2) \quad \text{(B30)}$$

and for dams,

$$\delta_f(\alpha) = \alpha^T\Delta\mathbf{V}_{n(f)}\alpha = \alpha_m^2\,\Delta\mathbf{V}_{n(f)}(1,1) + 2\alpha_m\alpha_f$$
$$\Delta\mathbf{V}_{n(f)}(1,2) + \alpha_f^2\,\Delta\mathbf{V}_{n(f)}(2,2) \quad \text{(B31)}$$

Terms due to β are, for sires,

$$\delta_m(\beta) = [\,\beta_1^2\pi_{11}^2\,\Delta\mathbf{V}_{n(m)}(1,1) + 2\beta_1\beta_2\pi_{11}\pi_{21}\Delta\mathbf{V}_{n(m)}(1,2)$$
$$+ \beta_2^2\pi_{21}^2\,\Delta\mathbf{V}_{n(m)}(2,2)]\,\sigma_{s,1}^2 \quad \text{(B32)}$$

and for dams,

$$\delta_f(\beta) = [\,\beta_1^2\pi_{12}^2\,\Delta\mathbf{V}_{n(f)}(1,1) + 2\beta_1\beta_2\pi_{12}\pi_{22}\Delta\mathbf{V}_{n(f)}(1,2)$$
$$+ \beta_2^2\pi_{22}^2\,\Delta\mathbf{V}_{n(f)}(2,2)]\,\sigma_{s,2}^2 \quad \text{(B33)}$$

So that $\delta_m = \delta_m(\alpha) + \delta_m(\beta)$ and $\delta_f = \delta_f(\alpha) + \delta_f(\beta)$, which gives Equation 12.

Example: With $N_m = 20$, $N_f = 60$, $n_o = 8$, $\sigma_P^2 = 1$, $\sigma_{A,0}^2 = h_0^2 = 0.4$: $p_m = 0.083$, $p_f = 0.25$, $t_m = 1.383$, $t_f = 0.674$, $i_m = 1.839$, $i_f = 1.271$, $k_m = 0.838$ and $k_f = 0.759$. From APPENDIX A, after the first iteration with $\rho^2 = h^2$: $\sigma_{A,s}^2 = 0.066$, $\sigma_{A,d}^2 = 0.070$, $\sigma_m^2 = 0.026$ and $\sigma_f^2 = 0.039$, so that non-zero elements of V are: $V_{11} = 0.026$, $V_{22} = 0.026$, $V_{33} = 0.013$, $V_{44} = 0.123$, $V_{55} = 0.113$, $V_{66} = 0.700$, $V_{14} = V_{41} = 0.013$, $V_{25} = V_{52} = 0.013$ and $V_{34} = V_{43} = 0.006$; $\mathbf{g}^T = (0.013, 0.013, 0.006, 0.098, 0.063, 0.175)$; $\mathbf{b}^T = (0.110, 0.234, 0.110, 0.779, 0.531, 0.250)$ and $\rho^2 = 0.472$. At equilibrium (*approx.* 5 iterations): $\rho^2 = 0.421$, $\sigma_{A,s}^2 = 0.048$, $\sigma_{A,d}^2 = 0.051$, $\sigma_A^2 = 0.299$, $\sigma_m^2 = 0.020$ $\sigma_f^2 = 0.030$ and $\sigma_{\hat{A}}^2 = 0.126$; $V_{11} = 0.020$, $V_{22} = 0.020$, $V_{33} = 0.010$, $V_{44} = 0.099$, $V_{55} = 0.101$, $V_{66} = 0.700$, $V_{14} = V_{41} = 0.010$, $V_{25} = V_{52} = 0.010$ and $V_{34} = V_{43} = 0.005$; $\mathbf{g}^T = (0.010, 0.010, 0.005, 0.074, 0.051, 0.175)$ and $\mathbf{b}^T = (0.137, 0.262, 0.137, 0.726, 0.477, 0.25)$. Covariance matrices of EBV of sibs are equal to V, except for: $C_{FS}(6,6) = -0.100$, $C_{HS}(2,2) = -0.010$, $C_{HS}(2,5) = C_{HS}(5,2) = -0.005$, $C_{HS}(5,5) = -0.050$ and $C_{HS}(6,6) = 0$, so that $\rho_{FS} = 0.603$, $\rho_{HS} = 0.335$, $\overline{\rho}_{FS} = 0.581$ and $\overline{\rho}_{HS} = 0.306$. From Equation 7 and 8, $\sigma_{s(m)}^2 = 0.248$ and $\sigma_{s(f)}^2 = 0.384$.

From APPENDIX B, $\mathbf{c}_m^T = (0.0193, 0, 0.0096, 0.1242, 0, 0)$, $\mathbf{c}_f^T = (0.0193, 0.0202, 0.0099, 0.1242, 0.0679, 0)$, so that $\lambda_{11} = 0.981$, $\lambda_{12} = 0.888$, $\lambda_{21} = 0.678$, $\lambda_{22} = 0.614$; $\pi_{11} = 0.091$, $\pi_{12} = 0.106$, $\pi_{21} =$

0.106 and π_{22} = 0.120. From Equation 5 with α_m = 0.025 and α_f = 0.0083, β_1 = 0.051 and β_2 = 0.016. From Equation 6, $E(u_{i,m}^2)$ = 0.00128 and $E(u_{i,f}^2)$ = 0.000171, so that the first term of Equation 1 equals 0.0179.

From APPENDIX B, $R(p_m, p_m, \overline{\rho}_{FS})$ = 0.23, $R(p_m, p_f, \overline{\rho}_{FS})$ = 0.37, $R(p_f, p_f, \overline{\rho}_{FS})$ = 0.47, $R(p_m, p_m, \overline{\rho}_{HS})$ = 0.42, $R(p_m, p_f, \overline{\rho}_{HS})$ = 0.55 and $R(p_f, p_f, \overline{\rho}_{HS})$ = 0.65. Furthermore, $E\{n_{ij}(m)[n_{ij}(m)-1]\}$ = 0.34, $E[n_{ij}(m)n_{ij}(f)]$ = 0.91, $E\{n_{ij}(f)[n_{ij}(f)-1]\}$ = 1.56; $E[n_{ij}(m)n_{ik}(m)]$ = 0.25, $E[n_{ij}(m)n_{ik}(f)]$ = 0.61, $E[n_{ij}(f)n_{ik}(f)]$ = 1.52; $E\{n_{i*}(m)[n_{i*}(m)-1]\}$ = 2.52, $E[n_{i*}(m)n_{i*}(f)]$ = 6.35, $E\{n_{i*}(f)[n_{i*}(f)-1]\}$ = 13.80 and $E[\mu_m^2(m)]$ = 1.96, $E[\mu_m(m)\mu_m(f)]$ = 4.98, $E[\mu_m^2(f)]$ = 13.11; $E[\mu_f^2(m)]$ = 0.25, $E[\mu_f(m)\mu_f(f)]$ = 0.61, $E[\mu_f^2(f)]$ = 1.58, so that $\Delta V_{(m)}(1,1)$ = 0.56, $\Delta V_{(m)}(1,2) = \Delta V_{(m)}(2,1)$ = 1.37, $\Delta V_{(m)}(2,2)$ = 0.69, $\Delta V_{(f)}(1,1)$ = 0.09, $\Delta V_{(f)}(1,2) = \Delta V_{(f)}(2,1)$ = 0.29 and $\Delta V_{(f)}(2,2)$ = –0.01. Furthermore, $\delta_m(\alpha) = 9.7\times10^{-4}$, $\delta_f(\alpha) = 1.8\times10^{-4}$, $\delta_m(\beta) = (4.912 + 8.861 + 0.830)\times10^{-5}\times0.248 = 3.627\times10^{-5}$, $\delta_f(\beta) = (1.102 + 2.491 - 0.016)\times10^{-5}\times0.384 = 1.374\times10^{-5}$, $\delta_m = 1.004\times10^{-3}$, $\delta_f = 1.929\times10^{-4}$ so that the Poisson correction equals 0.0040 and the final prediction is, $E[\Delta F] = 0.0179 + 0.0040 = 0.0219$.

Appendix C

Extension of Burrows method: The method of BURROWS (1984a,b) is based on calculating the average co-ancestry after a single cycle of selection. Using appendix E of WOOLLIAMS and BIJMA (1999), the extension to two sexes is straightforward. From equation (1) of BURROWS (1984b): $\Delta F_B = \frac{1}{8} Q_{HS} + \frac{1}{4}Q_{FS}$, where Q_{HS} is the probability that two selected offspring are half-sibs and Q_{FS} is the probability that two selected offspring are full-sibs. For two sexes, a distinction has to be made between male and female offspring, so that: $Q_{HS} = \frac{1}{4}Q_{HS}(m,m) + \frac{1}{2}Q_{HS}(m,f) + \frac{1}{4}Q_{HS}(f,f)$ and $Q_{FS} = \frac{1}{4}Q_{FS}(m,m) + \frac{1}{2}Q_{FS}(m,f) + \frac{1}{4}Q_{FS}(f,f)$. Combining BURROWS (1984b) and appendix E of WOOLLIAMS and BIJMA (1999): $Q_{HS}(m,m) = n_m(d-1)/[(T_m-1)R(\alpha_m,\alpha_m,\overline{\rho}_{HS})]$, $Q_{HS}(m,f) = n_m(d-1)/[T_m R(\alpha_m,\alpha_f,\overline{\rho}_{HS})]$, $Q_{HS}(f,f) = n_f(d-1)/[(T_f-1)R(\alpha_f,\alpha_f,\overline{\rho}_{HS})]$, $Q_{FS}(m,m) = (n_m-1)/[(T_m-1)R(\alpha_m,\alpha_m,\overline{\rho}_{FS})]$, $Q_{FS}(m,f) = n_m/[T_m R(\alpha_m,\alpha_f,\overline{\rho}_{FS})]$, $Q_{FS}(f,f) = (n_f-1)/[(T_f-1)R(\alpha_f,\alpha_f,\overline{\rho}_{FS})]$, where $T_m = T_f = \frac{1}{2}T$ and $n_m = n_f = \frac{1}{2}n_o$ for the current breeding schemes and $R(\alpha_x, \alpha_x, \overline{\rho}_{sibs})$ is given in APPENDIX B.

For random selection the result reduces to $\Delta F_B = \Delta F_W + \frac{1}{2}[\Delta F_W - 1/(8T)]/(T-1) \approx \Delta F_W$, where $\Delta F_w = 1/(8N_m) + 1/(8N_f) - 1/(8T)$, which is Wright's equation for fixed n_o; ΔF_B accounts for sampling parents without replacement by using a hypergeometric distribution of family size, whereas ΔF_w uses a binomial distribution.

Predicting Rates of Inbreeding in Livestock Breeding Populations

Piter Bijma*, Johan A. M. Van Arendonk* and John A. Woolliams†

* Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands and †Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K.

Abstract – This paper presents a deterministic method to predict rates of inbreeding (ΔF) for typical livestock breeding populations. The method is based on a recently developed general theory to predict rates of inbreeding, which uses the concept of long-term genetic contributions. A typical livestock breeding population was modelled, with overlapping generations, BLUP selection and an increasing amount of information when animals become older. Two types of selection were practised, animals were either selected by truncation on estimated breeding values (EBV) across age classes, or the number of parents selected from each age class was set to a fixed value and truncation selection was practised within age classes. Bulmer's equilibrium genetic parameters were obtained by iterating on a pseudo-BLUP selection index and ΔF was predicted for the equilibrium situation. Predictions were substantially more accurate than predictions from other available methods, which ignore the effect of selection on ΔF. Predictions were accurate for schemes with up to 20 sires. Predicted ΔF was too low for schemes with more than 20 sires, which was due to the use of simple linear models to predict genetic contributions. In spite of these errors, the present method enables balancing of rates of inbreeding and gain in animal breeding populations on a computationally feasible (*i.e.*, deterministic) manner.

Since the introduction of the Mixed Model Equations (HENDERSON 1963, 1975) and the availability of powerful computers, selection on Estimated Breeding Values (EBV) using Best Linear Unbiased Prediction (BLUP) has become a standard animal breeding practice. However, selection on BLUP-EBV is known to increase rates of inbreeding (*e.g.*, VERRIER *et al.* 1993) and may therefore lead to inbreeding depression and may jeopardize future genetic progress. Because methods that accurately predict rates of inbreeding in livestock breeding populations are lacking, many studies have used Wrights equation, $\Delta F = 1/(8N_m) + 1/(8N_f)$ (WRIGHT, 1969. p. 212), which dramatically underpredicts ΔF for selected populations (BIJMA and WOOLLIAMS, 2000).

WRAY and THOMPSON (1990) introduced methods to predict rates of inbreeding in selected populations, based on the concept of long-term genetic contributions. Recently, WOOLLIAMS *et al.* (1999) and WOOLLIAMS and BIJMA (2000) developed a general theory to predict rates of inbreeding in populations undergoing selection. Subsequently, explicit prediction equations were developed for

Submitted

populations with either discrete or overlapping generations undergoing mass selection (BIJMA *et al.* 2000a), and for populations with discrete generations undergoing BLUP selection (BIJMA and WOOLLIAMS, 2000). Most livestock breeding populations are selected on BLUP-EBV but have overlapping generations.

The aim of this paper is to develop explicit equations to predict rates of inbreeding for typical livestock breeding populations. For this purpose, we will model a population with overlapping generations, selection on BLUP-EBV and progeny testing of male selection candidates.

Methods

In this section we will first describe the population structure for which rates of inbreeding will be predicted. Because deterministic prediction of ΔF requires a deterministic analogy to BLUP, a pseudo-BLUP selection index will be used to provide the necessary equilibrium genetic parameters (BULMER, 1971). Next, the procedure for predicting rates of inbreeding will be outlined in three steps. Finally, we will describe a stochastic simulation procedure, which will be used to evaluate the accuracy of the deterministic prediction of ΔF. Table 1 shows the notation used.

The general theory to predict ΔF is derived in previous papers (*e.g.*, WOOLLIAMS and BIJMA, 2000; BIJMA and WOOLLIAMS, 2000). This paper, therefore, focuses on the implementation of the theory. Throughout the methods section, the meaning of the prediction equations is described in an intuitive manner.

Population structure: Selection was for a trait determined by the infinitesimal model (FISHER, 1918; BULMER, 1971). Phenotypic values (P) were the sum of additive genetic values (A, breeding values) and environmental values (E), $P = A + E$. A closed nucleus population with overlapping generations was modeled and selection was on BLUP-EBV for a single trait. With two sexes and a maximum age of c_{max}, there are $2c_{max}$ categories of animals, one for each sex and age of parent. Categories will be indexed by k or l, so $k,l = 1 \,..\, c_{max}$ are males, and $k,l = c_{max}+1 \,..\, 2c_{max}$ are females.

Phenotypes of selection candidates were recorded prior to reproductive age and BLUP-EBV were calculated. Progeny testing was included for males in the oldest age class ($k = c_{max}$), which had information on n_{prg} progeny included in their EBV. Those progeny were assumed to be born outside the nucleus, so that their dams did not enter the breeding value estimation. Females did not have progeny information.

Within categories, individuals were ranked on their EBV and each year the highest-ranking n_k individuals were selected from the k^{th} category, to produce the next cohort. The number selected from each age class, n_k, was either set to a fixed value in advance, or determined by truncation selection on EBV across age classes. The total number of male and female parents of each newborn cohort was, $N_m = \sum_{k=1}^{c_{max}} n_k$ and $N_f =$

Table 1.–Notation used

N_m, N_f, d , n_{prg}	Number of sires, number of dams, mating ratio $d = N_f/N_m$, number of progeny per sire
n_o, T	Number of selection candidates born per dam, total number of candidates per sex
c_{max}, n_k	Maximum age of parents, number of parents selected from category k
$\mathbf{N}$	$2c_{max}\times 2c_{max}$ diagonal matrix of the number of parents selected from each category k
k, s	Indicator for sex-age class categories, indicator for life time categories
p_k, τ_k	Selected proportion and standardized truncation point for category k
ι_k, κ_k	Selection intensity, variance reduction coefficient for category k
P, A, $\hat{A}$	Phenotype, breeding value, estimated breeding value (EBV)
$\mathbf{b}_1(\mathbf{b}_2)$, $\mathbf{x}_{1,i}(\mathbf{x}_{2,1})$	6×1 (7×1) vector of index weights, 6×1 (7×1) vector of index information sources
σ_A^2, $\sigma_{\hat{A}}^2$	Additive genetic variance, variance of the estimated breeding value
ρ, h^2	Accuracy of selection, heritability
$\overline{\rho}_{FS,kl}$ ($\overline{\rho}_{HS,kl}$)	Sample correlation between EBVs of full sibs (half sibs) in category k and l
ΔF, $r_{i,k}$	Rate of inbreeding, long term genetic contribution of individual i in category k
$s_{i,k}$, $\sigma_{s_k}^2$	Selective advantage of individual i in category k, variance of $s_{i,k}$
$u_{i,k}$, α_k, β_k	Expectation of $r_{i,k}$ conditional on $s_{i,k}$, linear model for $u_{i,k} = \alpha_k + \beta_k s_{i,k}$
$\mathbf{G}$, g_{kl}	$2c_{max}\times 2c_{max}$ gene flow matrix, element of $\mathbf{G}$
$\mathbf{D}$, d_{kl}	$2c_{max}\times 2c_{max}$ matrix of deviations of breeding values from mean, element of $\mathbf{D}$
Π, π_{kl}	matrix of regression coefficients of $s_{\text{offspring}}$ on s_{parent}, element of Π
Λ	matrix of regression coefficients of the number of selected offspring on s_{parent}
λ_{kl}, $n_{i,k}(l)$	Element of Λ, number of selected offspring in category l of parent i in category k
$\mathbf{V}_{n,k}$, $\Delta\mathbf{V}_{n,k}$	matrix of variance of family size, deviation of $\mathbf{V}_{n,k}$ from Poisson variance
δ_k	correction term required when $\Delta\mathbf{V}_{n(x)} \neq 0$

$\sum_{k=c_{max}+1}^{2c_{max}} n_k$, respectively. Each sire was mated at random to d dams ($d = N_f/N_m$), and each dam produced a fixed number, n_o, of offspring ($\frac{1}{2}n_o$ of each sex), so that the total number of selection candidates born in a cohort was, $T = \frac{1}{2}n_o N_f$ for each sex. The unit of age, *i.e.*, the interval between consecutive age classes, was one year. Genetic contributions and rates of inbreeding per year were therefore equal to genetic contributions and rates of inbreeding per cohort.

Pseudo BLUP selection index: A selection index analogy of the BLUP procedure was developed by extending the pseudo-BLUP selection index of WRAY and HILL (1989) to populations with overlapping generations. Because part of the selection candidates may have progeny information, two pseudo-BLUP indices were distinguished. First, *index1* without progeny information, which was used for male selection candidates in categories 1 to $c_{max}-1$ and for all female selection candidates. Second, *index2* with progeny information, which was used only for males in category c_{max}.

For the i^{th} candidate, *index1* was $\hat{A}_{1,i} = \mathbf{b}_1^T\mathbf{x}_{1,i}$, where superscript T denotes the transpose, $\hat{A}_{1,i}$ is the EBV, $\mathbf{b}_1$ is a 6×1 vector of index weights and $\mathbf{x}_{1,i}$ is a 6×1 vector of information sources for the i^{th} candidate. Information sources in $\mathbf{x}_{1,i}$ were, 1. $\hat{A}_m$, the EBV of the sire of i, 2. $\hat{A}_f$, the EBV of the dam of i, 3. $\overline{\hat{A}}_f$, the average EBV of the d dams mated to the sire, 4. $\overline{P}_{HS}$, the phenotypic average of the $n_o d$ half-sibs of i (including i and its full-sibs), 5. $\overline{P}_{FS}$ the phenotypic average of the n_o full-sibs of i (including i) and 6. P_i, the phenotype of candidate i.

In *index2*, progeny information was added, so that for the i^{th} male in category c_{max}, *index2* was, $\hat{A}_{2,i} = \mathbf{b}_2^T\mathbf{x}_{2,i}$, where $\hat{A}_{2,i}$ is the EBV when including progeny information, $\mathbf{b}_2$ is a 7×1 vector of index weights and $\mathbf{x}_{2,i}$ is a 7×1 vector of information sources for the i^{th} candidate, with the first six elements identical to $\mathbf{x}_{1,i}$ and additionally 7. $\overline{P}_{prg}$, which is the phenotypic average of the n_{prg} offspring of a male in category c_{max}.

Using the above pseudo-BLUP indices, selection and mating were iterated until equilibrium genetic parameters were reached (BULMER, 1971). Iterative equations for calculating index weights, the accuracy of selection (ρ), and equilibrium parameters are given in APPENDIX A. Rates of inbreeding were predicted using equilibrium genetic parameters.

Prediction of rates of inbreeding

General: The prediction method is based on the concept of long-term genetic contributions (JAMES and MCBRIDE, 1958; WRAY and THOMPSON, 1990). The long-term genetic contribution (r_i) of ancestor i in generation t_1 is defined as the proportion of genes from i that are present in individuals in generation t_2 deriving by descent from i, where $(t_2-t_1) \rightarrow \infty$ (WOOLLIAMS *et al.* 1993). In other words, the long-term genetic contribution of an individual is its proportional contribution to the genetic make-up of the population in the long-term. In the remainder of this article, long-term genetic contributions are referred to as "genetic contributions", or simply "contributions".

WRAY and THOMPSON (1990) showed that rates of inbreeding per generation are proportional to the sum of squared contributions, $E(\Delta F) = \frac{1}{4}\sum r_i^2$, where r_i is the realized genetic contribution of individual i and the sum is taken over all parents in a generation. Recently, WOOLLIAMS and BIJMA (2000) showed that rates of inbreeding can be expressed in terms of *expected* contributions,

$$E(\Delta F) = \tfrac{1}{2}\sum_s n_s E(u_{i,s}^2) + \tfrac{1}{8}\sum_s n_s \delta_s \,, \qquad (1)$$

where n_s is the number of parents selected from life-time category s, $u_{i,s}$ is the expected lifetime contribution of individual i in lifetime category s conditional on its selective advantage, and δ is a correction factors for deviations of the variance of family size ($\mathbf{V}_n$) from a Poisson variance. When $\mathbf{V}_n$ deviates from Poisson, $\frac{1}{4}\sum r_i^2$ differs from $\frac{1}{2}\sum_s n_s E(u_{i,s}^2)$ and the difference is accounted for by the second term

of Equation 1. In this paper, we will predict ΔF using Equation 1. Throughout the paper, family size refers to the number of selected offspring of a parent, not to the number of candidates. The second term of Equation 1 will be referred to as the 'Poisson correction'.

In Equation 1, it is essential to note that $u_{i,s}$ refers to the *lifetime* contribution of individual i and subscript s denotes the lifetime category to which i belongs. The lifetime contribution of individual i is the sum of all contributions originating from its selection at a specific age, $u_{i,s} = \Sigma u_{i,k}$, where the sum is taken over all age classes k in which individual i is selected. Throughout this paper, lifetime categories are indexed by s, which refers to a specific combination of age classes in which the individual is selected, whereas index k refers to a specific sex-age class. This issue will be addressed below (see "*step* 2").

Components of Equation 1 will be calculated in three steps. In the first step we will predict expected genetic contributions of sex-age classes, $u_{i,k}$, using the approach of WOOLLIAMS *et al.* (1999). In the second step, lifetime contributions, $u_{i,s}$, will be expressed as a function of sex-age class contributions, $u_{i,k}$, and subsequently $E(u_{i,s}^2)$ will be derived. In the third step we will derive δ_s.

In a selected population, a superior individual is expected to have a higher genetic contribution than an average individual, because its offspring and further descendants have a higher probability of being selected. When predicting genetic contributions, we need to explicitly model this superiority, *i.e.*, we need to define the factors that confer selective advantage to an individual. Before proceeding to the prediction of expected contributions, therefore, we will first define the selective advantage.

Selective advantage: In principle, the selective advantage should contain all terms that affect the long-term contribution of an individual (*i.e.*, by affecting the selective success of its descendents). In this paper, we will use the breeding value of the individual plus the breeding value of its mate(s). The mate is included, because its breeding value affects the selective success of the offspring, so that the breeding value of the mate can be regarded as a component of the selective advantage of an individual. Other models for the selective advantage are possible, *e.g.*, instead of using the true breeding value, one may use the EBV together with the prediction error. A discussion on this topic is in BIJMA and WOOLLIAMS (2000). For sire i in category k, the selective advantage was

$$s_{i,k} = (A_i + \overline{A}_f) - \overline{(A_i + \overline{A}_f)_k}, \quad k = 1, c_{max} \quad (2)$$

where A_i is the breeding value of sire i, $\overline{A}_f$ is the average breeding value of the d dams mated to sire i and the second term represents subtraction of the average selective advantage for category k. For dams the selective advantage was

$$s_{i,k} = (A_i + A_m) - \overline{(A_i + A_m)_k}, \; k = c_{max}+1, 2c_{max} \quad (3)$$

where A_i is the breeding value of dam i and A_m is the breeding value of the sire mated to dam i.

Step 1, prediction of expected contributions: This section follows the general

approach of WOOLLIAMS *et al.* (1999) for the prediction of expected genetic contributions. Expected contributions were predicted by linear regression on the selective advantage.

For males, the expected contribution of sire i in category k was,

$$u_{i,k} = \mathrm{E}(r_{i,k} \mid s_{i,k}) = \alpha_k + \beta_k s_{i,k} + \sum_{j=1}^{d} (\alpha_{cat(j)} - \overline{\alpha}_f), \quad k = 1, c_{max} \quad (4)$$

where α_k represents the average contribution of an ancestor in category k, $\beta_k s_{i,k}$ represents the deviation of the contribution from this average due to the selective advantage of ancestor i, j denotes the mates of i, $\alpha_{cat(j)}$ is the category of mate j, and $\overline{\alpha}_f = \sum_{k=c_{max}+1}^{2c_{max}} (n_k \alpha_k)/N_f$, which is the contribution of an average dam. The last term of Equation 4 represents the effect of the categories of the mates on the contribution of individual i, which was not accounted for when defining the selective advantage of individual i, *i.e.*, $\overline{A}_d$ in Equation 2 accounts only for the within category effect of the mate.

For females, the expected contribution of dam i in category k was:

$$u_{i,k} = \mathrm{E}(r_{i,k} \mid s_{i,k}) = \alpha_k + \beta_k s_{i,k} + (\alpha_j - \overline{\alpha}_m)/d, \quad k = c_{max}+1, 2c_{max} \quad (5)$$

Note that, contrary to Bijma and Woolliams (2000), the effect of the mate is included directly in the model for predicting $u_{i,k}$. This difference does not affect the results, but the present approach is more straightforward.

Solutions for $u_{i,k}$ are obtained by predicting α_k and β_k. There are two mechanisms determining α_k and β_k. First, superior parent are expected to have more selected offspring, which is modeled by a regression coefficient λ. Second, offspring partly inherit the selective advantage of their parents, which is modeled by a regression coefficient π.

By modeling these two mechanisms, WOOLLIAMS *et al.* (1999) show that α_k and β_k can be obtained from

$$\mathbf{N\alpha} = [\mathbf{G}^{\mathrm{T}} + (\mathbf{G}^{\mathrm{T}} * \mathbf{D}^{\mathrm{T}})(\mathbf{I} - \mathbf{G}^{\mathrm{T}}\boldsymbol{\Pi}^{\mathrm{T}})^{-1}(\mathbf{G}^{\mathrm{T}} * \boldsymbol{\Lambda}^{\mathrm{T}})]\mathbf{N\alpha} \quad (6)$$

$$\mathbf{N\beta} = (\mathbf{I} - \mathbf{G}^{\mathrm{T}} * \boldsymbol{\Pi}^{\mathrm{T}})^{-1}(\mathbf{G}^{\mathrm{T}} * \boldsymbol{\Lambda}^{\mathrm{T}})\mathbf{N\alpha} \quad (7)$$

where * denotes element by element multiplication, $\mathbf{I}$ is the $2c_{max} \times 2c_{max}$ identity matrix, $\mathbf{N}$ is a $2c_{max} \times 2c_{max}$ diagonal matrix containing the numbers of parents selected from each category (n_k), $\boldsymbol{\Pi}$ is a $2c_{max} \times 2c_{max}$ matrix of elements π_{kl}, being the regression coefficient of the selective advantage ($s_{j,k}$) of selected offspring j in category k on the selective advantage ($s_{i,l}$) of parent i in category l, $\boldsymbol{\Lambda}$ is a $2c_{max} \times 2c_{max}$ matrix of elements λ_{kl}, being the regression coefficient of the number of selected offspring in category k on the selective advantage of parent i in category l, $\mathbf{G}$ is a $2c_{max} \times 2c_{max}$ gene flow matrix of elements g_{kl}, specifying the proportional contribution of parent category l to selected offspring in category k, $\mathbf{D}$ is a $2c_{max} \times 2c_{max}$ matrix of elements d_{kl}, being the average selective advantage of selected offspring in category k descending from parents in category l, expressed as a deviation from the mean

selective advantage in category k, α is a $2c_{max}$ vector of elements α_l, and β is a $2c_{max}$ vector of elements β_l. [See BIJMA and WOOLLIAMS (1999) for a detailed study and an example on the prediction of expected genetic contributions with overlapping generations]. The above matrices follow the gene flow notation (HILL, 1974), so that rows represent offspring categories and columns represent parent categories, and this is the reason for the matrices in (6) and (7) requiring the transpose.

Note that the gene flow matrix, **G**, differs from HILL's (1974) gene flow matrix, which specifies the contribution of parent categories to offspring *before* selection. Here, **G** refers to selected offspring. For example, for c_{max} = 2, g_{41} = 0.35 means that one-year-old sires contribute 35% of the genes of 2-year-old selected females.

Matrices **G**, **Π**, **A** and **D** were derived following the approach of WOOLLIAMS *et al.* (1999) and BIJMA and WOOLLIAMS (1999). Resulting equations for the case studied in this paper are listed in Appendix B.

Note that contributions predicted from Equations (4) through (7) are the contributions of a single cohort (*i.e.*, the group born in a single year, not an entire generation) originating from their selection at the different ages. Rates of inbreeding predicted from these contributions are, therefore, also per year.

Step 2, derivation of $E(u_{i,s}^2)$: The lifetime contribution is the sum of the contributions originating from selection at a specific age, $u_{i,s} = \Sigma u_{i,k}$, where $u_{i,k}$ is obtained from Equation 4 or 5. To calculate $E(u_{i,s}^2)$, therefore, we have to express contributions of lifetime categories, which are indexed by s, in terms of contributions of sex-age class categories, which are indexed by k. Lifetime categories refer to a specific combination of age classes in which the individual is selected, *e.g.*, males selected only at one year of age, which have, $u_{i,s}^2 = u_{i,k=1}^2$; males selected at one and two years of age, which have, $u_{i,s}^2 = (u_{i,k=1} + u_{i,k=2})^2 = u_{i,k=1}^2 + 2u_{i,k=1}u_{i,k=2} + u_{i,k=2}^2$; *etc.* This example shows that the square of the lifetime contribution, $u_{i,s}^2$, can be expressed as a sum of squares, $u_{i,k}^2$, and cross products, $u_{i,k}u_{i,l}$, of sex-age class contributions.

Instead of explicitly specifying all lifetime categories and deriving the corresponding $E(u_{i,s}^2)$, one can directly express $\sum_s n_s E(u_{i,s}^2)$ as a sum of squares and cross products in terms of the categories k, which has the advantage that contributions predicted from Equations 4 and 5 can be used directly (BIJMA *et al.* 2000a). In the following derivation of squared expected contributions, terms contributing to $\sum_s n_s E(u_{i,s}^2)$ will be collected separately for males without progeny testing, for progeny tested males and for females.

For categories without progeny testing, the ranking of animals remains practically unchanged when they move through the age classes, which is the same situation as with mass selection. For those categories, therefore, equations 11 and 12 of BIJMA *et al.* (2000a) can

be used, so that, for male categories without progeny testing,

$$\sum_s n_s E(u_{i,s}^2) = \sum_{k=1}^{c_{max}-1} n_k E(u_{i,k}^2) + 2\sum_{k=1}^{c_{max}-2}\sum_{l=k+1}^{c_{max}-1} \min(n_l, n_k) E(u_{i,k}u_{i,l}), \quad (8)$$

and for all female categories,

$$\sum_s n_s E(u_{i,s}^2) = \sum_{k=c_{max}+1}^{2c_{max}} n_k E(u_{i,k}^2) + 2\sum_{k=c_{max}+1}^{2c_{max}-1}\sum_{l=k+1}^{2c_{max}} \min(n_l, n_k) E(u_{i,k}u_{i,l}), \quad (9)$$

where Σ_s denotes summation over the relevant lifetime categories, and $\min(n_k,n_l)$ denotes the minimum of n_k and n_l (see also the example in BIJMA *et al.* 2000a).

For Equation 8 and 9, $E(u_{i,k}^2)$ is obtained by squaring equation 4 and 5, using $E(s_{i,k}^2) = \sigma_{s,k}^2$ because $E(s_{i,k}) = 0$. For Equation 8 the result is

$$E(u_{i,k}^2) = \alpha_k^2 + \beta_k^2\sigma_{s,k}^2 + d(\overline{\alpha_f^2} - \overline{\alpha}_f^2) \qquad k = 1, c_{max}-1 \quad (10)$$

and for Equation 9,

$$E(u_{i,k}^2) = \alpha_k^2 + \beta_k^2\sigma_{s,k}^2 + (\overline{\alpha_m^2} - \overline{\alpha}_m^2)/d^2 \qquad k = c_{max}+1, 2c_{max} \quad (11)$$

Next, for males, the variance of the selective advantage is (see Equation 2)

$$\sigma_{s,k}^2 = \sigma_A^2(1-\kappa_k\rho_k^2)(1-1/n_k) + \frac{\sigma_A^2}{d}\sum_{l=c_{max}+1}^{2c_{max}} \frac{n_l}{N_f}(1-\kappa_l\rho_l^2)(1-1/n_l) \qquad k = 1, c_{max} \quad (12)$$

and for females (see Equation 3)

$$\sigma_{s,k}^2 = \sigma_A^2(1-\kappa_k\rho_k^2)(1-1/n_k) + \sigma_A^2\sum_{l=1}^{c_{max}} \frac{n_l}{N_m}(1-\kappa_l\rho_l^2)(1-1/n_l) \qquad k = c_{max}+1, 2c_{max} \quad (13)$$

In Equation 12 and 13, the first term is due to the individual itself, the second term due to its mate(s), the term $(1-\kappa\rho^2)$ accounts for reduced variance because the parents are a selected group and the $(1-1/n_k)$ accounts for finite sample size.

For Equation 8 and 9, expectations of cross products are calculated from (BIJMA *et al.* 2000a),

$$E(u_{i,k}u_{i,l}) = \alpha_k\alpha_l + [1-1/n_{min}]\,\beta_k\beta_l\sigma_A^2[1-\kappa_{min}\rho_{min}^2] + \alpha_{min}\beta_{max}E[\overline{A}_{min} - \overline{A}_{max}] \quad (14)$$

where subscript min denotes the category with the lower number of animals, subscript max denotes the category with the higher number of animals and $\overline{A}_k$ is the genetic selection differential in category k, $\overline{A}_k = \iota_k\rho_k\sigma_A$. With random mating, there is no covariance between the selective advantages of two different mates, so that mates do not contribute to the cross product in Equation 14.

For progeny tested males we need to derive the contribution of category c_{max} to the sum of squared lifetime contributions, which is composed of the sum of squared contributions from category c_{max} and the sum of cross products between category c_{max} and the non-progeny tested male categories,

$$\sum_s n_s E(u_{i,s}^2) = n_{c_{max}} E(u_{i,c_{max}}^2)$$

$$+ 2 \sum_{k=1}^{c_{max}-1} n_{k,c_{max}} E(u_{i,k} u_{i,c_{max}}) , \quad (15)$$

where $E(u^2_{i,c_{max}})$ follows from Equation 10 by putting $k = c_{max}$, and $n_{k,c_{max}}$ is the number of parents selected jointly in category k and category c_{max}. The number selected jointly in both categories can be calculated from the number of candidates and the proportion selected jointly in category k and category c_{max}, $n_{k,c_{max}} = \Phi(\tau_k, \tau_{c_{max}}, \rho_{k,c_{max}})T$, where $\Phi(\tau_k, \tau_{c_{max}}, \rho_{k,c_{max}})$ denotes the bivariate normal proportion above truncation points τ_k and $\tau_{c_{max}}$, which are the EBV truncation points for category k and c_{max}, and $\rho_{k,c_{max}}$ is the correlation between $\hat{A}_{i,k}$ and $\hat{A}_{i,c_{max}}$, $\rho_{k,c_{max}} = \sigma_{\hat{A},k} / \sigma_{\hat{A},c_{max}}$. The bivariate normal proportion, $\Phi(\tau_k, \tau_{c_{max}}, \rho_{k,c_{max}})$, was calculated using Dutt's algoritm (DUTT, 1973, 1975; DUTT and SOMS, 1976; DUCROCQ and COLLEAU, 1986). For Equation 15, expectations of cross products follow from Equations 2 and 5, with no contribution due to the mates,

$$E(u_{i,k} u_{i,c_{max}}) =$$
$$\alpha_k \alpha_{c_{max}} + \alpha_k \beta_{c_{max}} [E(A_i) - \overline{A}_{c_{max}}]$$
$$+ \alpha_{c_{max}} \beta_k [E(A_i) - \overline{A}_k]$$
$$+ \beta_k \beta_{c_{max}} E[(A_i - \overline{A}_k)(A_i - \overline{A}_{c_{max}})] \quad (16)$$

where i refers to individuals that are selected both in category k and in category c_{max}, and $E(A_i)$ is the expected breeding value of those individuals. The terms $E(A_i)$ and $E[(A_i - \overline{A}_k)(A_i - \overline{A}_{c_{max}})]$ are calculated using a result of TALLIS (1961) and are given in Appendix C.

Summarizing, the sum of squared expected lifetime contributions is given by equation 8 for males without progeny testing, by equation 9 for females and by equation 15 for progeny tested males. Finally, the first term of Equation 1 is obtained by adding up results from Equations 8, 9 and 15. The remaining task is to obtain the second term of Equation 1, which requires the calculation of δ_k.

Step 3, calculation of δ_k: The calculation of δ_k is a straightforward analogy of the discrete generation case (BIJMA and WOOLLIAMS, 2000). Here we will outline the concept, explicit equations are given in Appendix D.
In Equation 1,

$$\delta_k = \alpha^T \Delta \mathbf{V}_{n,k} \alpha \quad (17)$$

where $\Delta \mathbf{V}_{n,k}$ is the $2c_{max} \times 2c_{max}$ matrix of deviations of the variance of family size from a Poisson variance, for a parent in category k, conditional on its selective advantage $s_{i,k}$ (WOOLLIAMS and BIJMA, 2000). For example, for the full variance of family size conditional on the selective advantage, element $\mathbf{V}_{n,k}(l,l')$ represents the covariance between the number of offspring selected in category l, $n_{i,k}(l)|s_{i,k}$, and the number of offspring selected in category l', $n_{i,k}(l')|s_{i,k}$, of a parent in category k.

For diagonal elements, the deviation from a Poisson variance is obtained by subtracting the mean number of selected offspring from the full variance (with Poisson, $\sigma^2 = \mu$). For diagonal elements, therefore, $\Delta \mathbf{V}_{n,k}(l,l) =$

$E_s\{E[n_{i,k}^2(l)\,|\,s_{i,k}] - E[n_{i,k}(l)\,|\,s_{i,k}]^2 - E[n_{i,k}(l)\,|\,s_{i,k}]\}$, where E_s denotes the expectation with respect to $s_{i,k}$, which gives (see also BIJMA and WOOLLIAMS, 2000),

$$\Delta V_{n,k}(l,l) = E_s\{E[n_{i,k}(l)(n_{i,k}(l)-1)\,|\,s_{i,k}]\} - E_s\{E[n_{i,k}(l)\,|\,s_{i,k}]^2\} \quad (18)$$

For off-diagonal elements, $\Delta V_{n,k}(l,l') = V_{n,k}(l,l')$, because, with an independent Poisson distribution for each category, the covariance between $n_{i,k}(l)|s_{i,k}$ and $n_{i,k}(l')|s_{i,k}$ is zero. For off-diagonal elements, therefore,

$$\Delta V_{n,k}(l,l') = E_s\{E[n_{i,k}(l)n_{i,k}(l')\,|\,s_{i,k}]\} - E_s\{E[n_{i,k}(l)\,|\,s_{i,k}]E[n_{i,k}(l')\,|\,s_{i,k}]\} \quad (19)$$

where $s_{i,k}$ is the selective advantage of parent i in category k and $n_{i,k}(l)$ is the number of offspring of parent i selected in category l.

For Equation 18 and 19, the first term can be approximated using a result of BURROWS (1984a) (see appendix E of WOOLLIAMS and BIJMA, 2000). Analogous to BIJMA and WOOLLIAMS (2000), the second term follows from substituting the Λ-model for predicting the expected number of selected offspring, $E[n_{i,k}(l)\,|\,s_{i,k}) = 2g_{lk}n_l n_k^{-1}(1+\lambda_{lk}s_{i,k})$, where $2g_{lk}n_l n_k^{-1}$ represents the average number of selected offspring in category l descending from a single parent in category k, and $\lambda_{lk}s_{i,k}$ represents the effect of the selective advantage of the parent on the number of selected offspring.

The extension of Equations 18 and 19 to a population with overlapping generations and a hierarchical mating structure is a straightforward analogy of equations for the discrete generations case (BIJMA and WOOLLIAMS, 2000) and resulting equations are given in Appendix D. In summary, the second term of Equation 1 is obtained using Equation 17, where $\Delta V_{n,k}$ is given by equations D1 through D14.

TABLE 2.–Rates of inbreeding per year from simulation (ΔF_{sim}) and corresponding prediction errors for schemes with two age classes, 20 sires and 60 dams selected per year and truncation selection on EBV across age classes[a]

n_o[c]	h^2[d]	diagN[e]	ΔF_{sim}[f]	error%[b]
2	0.1	10 10 43 17	.0089	-3
	0.2	11 9 42 18	.0095	0
	0.4	12 8 41 19	.0100	2
	0.6	14 6 40 20	.0101	2
4	0.1	9 11 54 6	.0131	-10
	0.2	11 9 53 7	.0147	-7
	0.4	14 6 52 8	.0160	-8
	0.6	16 4 51 9	.0151	-9
8	0.1	11 9 58 2	.0221	-3
	0.2	13 7 58 2	.0243	-4
	0.4	16 4 58 2	.0245	-14
	0.6	18 2 57 3	.0195	-10

[a]Sires in age class two have information on 100 progeny included in their EBV. [b]error% = $100\%\times(\Delta F_{pred}-\Delta F_{sim})/\Delta F_{sim}$. [c]$n_o$ = number of offspring per dam. [d]h^2 = heritability. [e]diagN = number of parents selected from each age class, first two elements refer to sires, last two elements refer to dams. [f]Standard errors of simulation results were smaller than 1% of their mean value

Results

Table 2 shows rates of inbreeding per year from simulation (ΔF_{sim}) and corresponding prediction errors for populations with two age classes, where EBV for males in age class two include information on 100 progeny. In Table 2, the number of parents selected from each age class

TABLE 3.–Rates of inbreeding per year from simulation (ΔF_{sim}) and corresponding prediction errors for schemes with three age classes, 20 sires and 60 dams selected per year and truncation selection on EBV across age classes[a]

n_o[c]	h^2[d]	diagN[e]	ΔF_{sim}[f]	error%[b]
2	0.1	12 2 6 40 17 3	0.0093	-6
	0.2	13 2 5 39 17 4	0.0105	-5
	0.4	14 3 3 38 17 5	0.0117	-3
	0.6	15 3 2 37 18 5	0.0112	-2
4	0.1	14 1 5 52 8 0	0.0156	-12
	0.2	16 1 3 51 9 0	0.0177	-11
	0.4	18 1 1 50 9 1	0.0173	-13
	0.6	18 2 0 50 9 1	0.0150	-9
8	0.1	18 0 2 57 3 0	0.0313	-7
	0.2	19 0 1 57 3 0	0.0305	-10
	0.4	20 0 0 57 3 0	0.0242	-9
	0.6	20 0 0 57 3 0	0.0186	-9

[a]Sires in age class three have information on 100 progeny included in their EBV. [b]error% = $100\% \times (\Delta F_{pred} - \Delta F_{sim})/\Delta F_{sim}$. [c]$n_o$ = number of offspring per dam. [d]h^2 = heritability. [e]diagN = number of parents selected from each age class. [f]Standard errors of simulation results were smaller than 1% of their mean value.

is a result of truncation selection on EBV across age classes, showing that selection of males moves towards age class two when h^2 and n_o are low. Predictions are accurate for $n_o = 2$. For $n_o = 4$ or 8, predictions are approximately 8% too low. In spite of the errors, predictions and simulations show the same trend. For example, doubling the number of offspring per parent from $n_o = 4$ to $n_o = 8$ for a scheme with $h^2 = 0.1$, raises ΔF_{sim} by a factor 1.7 whereas the prediction indicates a factor 1.8. Note that the use of $\Delta F = 1/(8N_m) + 1/(8N_f)$ would give the same rate of inbreeding for schemes with $n_o = 4$ *vs.* 8, *i.e.*, a factor of 1.

Table 3 shows rates of inbreeding from simulation and corresponding prediction errors for populations with three age classes, truncation selection across age classes, and where EBV for males in age class three include information on 100 progeny. Accuracy of predictions in Table 3 is comparable to accuracy in Table 2.

With truncation selection across age classes, the majority of the parents is selected from the youngest age class in most cases. To evaluate the accuracy of predictions for any distribution of parents across age classes, the proportion of parents selected from each age class was set to a fixed value and animals were selected by truncation on EBV within age classes. For a population with two age classes, three alternatives for the proportion of parents selected from the different age classes were considered, p = (0.75, 0.25; 0.75, 0.25), p = (0.5, 0.5; 0.75, 0.25) and p = (0.25, 0.75; 0.75, 0.25). For example, p = (0.25, 0.75; 0.75, 0.25) with $N_m = 40$ and $N_f = 80$ gives N = diag{10, 30, 60, 20}. For each of these distributions, all

TABLE 4.–Mean and standard deviation of the prediction error for a range of schemes with two age classes[a]

	p = 0.75 0.25 0.75 0.25[b]		p = 0.5 0.5 0.75 0.25		p = 0.25 0.75 0.75 0.25	
N_m[c]	mean	sd.	mean	sd.	mean	sd.
8	-2	3	2	3	0	4
12	-4	4	0	3	-1	3
20	-5	5	-1	5	-4	3
40	-5	6	-7	6	-7	5
60	-7	9	-7	8	-9	7
80	-7	10	-9	9	-9	9

[a]Sires in age class two have information on 100 progeny included in their EBV. [b]p denotes the proportions of parents selected from each age class. [c]N_m = number of sires. For each N_m alternative, results were averaged over the heritability, mating ratio and number of offspring alternatives (see text).

combinations of schemes were evaluated for N_m = 8, 12, 20, 40, 60 or 80, d = 2, 4 or 8, n_o = 2, 4 or 8 and h^2 = 0.1, 0.2, 0.4 or 0.6, with information on 100 progeny for males in age class two, and the restriction that $N_f \leq 160$ to limit computing time for the stochastic simulations. In total 468 different schemes were evaluated therefore. Within this range, the maximum rate of inbreeding was, ΔF_{sim} = 0.0597 for **N** = diag{6, 2, 12, 4}, n_o = 8 and h^2 = 0.1, with a prediction error of +1%. The minimum rate of inbreeding was, ΔF_{sim} = 0.0021 for **N** = diag{20, 60, 120, 40}, n_o = 2 and h^2 = 0.1, with a prediction error of –5%.

Table 4 shows the average error and the standard deviation of the error for the whole range of schemes mentioned above, where schemes are grouped according to the number of sires. Schemes with up to 20 sires show accurate predictions, *i.e.*, the absolute value of the mean error and the standard deviation of the error are below 5%. Schemes with more than 20 sires show a systematic underprediction of approximately 8% and an increasing standard deviation of the error.

To give some background information on prediction errors in Table 4, Table 5 shows the typical error trend for schemes with many parents. When selection intensity is low (n_o = 2), predictions are accurate. For higher selection intensities (n_o = 4 or 8), predictions are too low. In spite of the errors, predictions give a good indication of the effect of changing population parameters. For example, doubling the number of offspring per parent increases the rate of inbreeding from simulation by a factor 1.96 (h^2 = 0.1, n_o = 4 *vs.* 8). For the same scheme, the prediction indicates an increase by a factor 1.82, which gives a good indication of the tremendous effect of selection intensity on ΔF with BLUP selection. BIJMA and WOOLLIAMS (2000) obtained similar error trends for populations with discrete generations and showed that the underprediction is due to the use of simple linear models to predict expected genetic contributions (Equations 2 through 5).

TABLE 5.–Typical trend of the prediction errors for a scheme with many parents[a,b]

	$n_o = 2^d$		$n_o = 4$		$n_o = 8$	
h^2 [e]	ΔF_{sim}	error%[c]	ΔF_{sim}	error%[c]	ΔF_{sim}	error%[c]
0.1	0.0025	0	0.0046	-13	0.0090	-19
0.2	0.0026	0	0.0047	-13	0.0091	-20
0.3	0.0026	4	0.0046	-11	0.0081	-17
0.4	0.0026	4	0.0042	-7	0.0069	-13

[a]For N = diag{40, 40, 120, 40}. [b]Sires in age class two have information on 100 progeny included in their EBV. [c]error% = $100\%\times(\Delta F_{pred}-\Delta F_{sim})/\Delta F_{sim}$. [d]$n_o$ = number of offspring per dam. [e]h^2 = heritability. [f]ΔF_{sim} = rate of inbreeding per year from simulation

Table 6 shows rates of inbreeding and corresponding prediction errors for schemes with four age classes for each sex. Because the potential number of alternative schemes is very large, results are presented for a limited number of schemes with $N_m = 20$, $d = 3$ and $h^2 = 0.3$, where the distribution of parents across age classes was varied. For most schemes in Table 6, predictions are accurate or show some underprediction due to the same reason as mentioned above. There are, however, three schemes in Table 6 that show a surprising overprediction of 14, 11 and 16%.

Detailed examination of the schemes with 14% and 11% error revealed that the overprediction of ΔF was due to overprediction of the contributions of one-year-old sires (α_1), which in turn was due to overprediction of the selection intensity in category one. In category 1, only two sires are selected and the intraclass correlation between sibs is relatively high ($\bar{\rho}_{FS,11} \approx 0.73$; $\bar{\rho}_{HS,11} \approx 0.33$), indicating that reduction of selection intensity due to finite numbers and correlations between indices of relatives becomes important. Adjusting selection intensities using the method of

TABLE 6.–Rates of inbreeding per year from simulation (ΔF_{sim}) and corresponding prediction errors for a scheme with four age classes[a,b]

diagN[c]	n_o[d]	ΔF_{sim}[f]	error%[e]
10 5 3 2 30 15 10 5	2	0.0132	0
	4	0.0227	–4
	8	0.0379	–4
10 5 3 2 5 10 15 30	2	0.0117	–1
	4	0.0214	–7
	8	0.0377	–7
2 3 5 10 30 15 10 5	2	0.0098	1
	4	0.0171	3
	8	0.0302	14
2 3 5 10 5 10 15 30	2	0.0085	4
	4	0.0156	–4
	8	0.0262	11
5 5 5 5 15 15 15 15	2	0.0125	9
	4	0.0224	16
	8	0.0390	3
2 8 8 2 5 25 25 5	2	0.0102	–6
	4	0.0171	–3
	8	0.0272	–7
8 2 2 8 25 5 5 25	2	0.0111	3
	4	0.0208	2
	8	0.0390	–1
2 8 2 8 5 25 5 25	2	0.0086	–1
	4	0.0156	–9
	8	0.0257	1

[a]Sires in age class four have information on 100 progeny included in their EBV. [b]For $h^2 = 0.3$. [c]diagN = number of parents selected from each age class. [d]n_o = number of offspring per dam. [e]error% = $100\%\times(\Delta F_{pred}-\Delta F_{sim})/\Delta F_{sim}$. [f]Standard errors of simulation results were smaller than 1% of the estimate

MEUWISSEN (1991) reduced intensities from 2.73 to approx. 2.57 for both schemes, and prediction errors reduced from +14% to −4% for the one scheme and from +11% to −3% for the other scheme. Those schemes also required adjustment of the selection intensity to accurately predict genetic gain. This indicates that the occasional need for adjusted selection intensities is not a specific feature of the method to predict ΔF, but is a general requirement for schemes with few parents and high intraclass correlations between EBVs of sibs.

Examination of the scheme with 16% error revealed that the overprediction of ΔF was due to overprediction of the variance of family size. When the number of selected parents is small compared to the number of candidates per family, all parents may be selected from very few families, which can be accounted for by adjusting the selected proportion according to Equation D13 (WRAY *et al.* 1990). In the present paper, adjusted selected proportions were used for all schemes where minimum(n_k; $k = 1,c_{max}$) $\leq 0.75n_o$ (See Appendix D). For the scheme with N = diag{5, 5, 5, 5, 15, 15, 15, 15}, this means that selected proportions were not adjusted for $n_o = 2$ or 4, whereas for $n_o = 8$ selected proportions were adjusted. The scheme with $n_o = 4$ is borderline, *i.e.*, in each male age class the number of parents is small, but just above the threshold that was used for adjusting the selected proportion. Adjusting the selected proportion for the scheme with $n_o = 4$ reduced the prediction error from +16% to +4%.

Discussion

This paper shows how the general procedure of WOOLLIAMS *et al.* (1999) and WOOLLIAMS and BIJMA (2000) for predicting rates of inbreeding in selected populations can be implemented for livestock breeding populations. For this purpose, a livestock breeding population with overlapping generations, BLUP selection, and progeny testing was modeled. Except for methods that ignore selection, there are no other methods available to predict rates of inbreeding in livestock breeding populations at present. Detailed discussions on theoretical issues of the methods have been included in previous papers [WOOLLIAMS *et al.* (1999), WOOLLIAMS and BIJMA (2000), BIJMA and WOOLLIAMS (2000), BIJMA *et al.* (2000)]. In this discussion, therefore, we will primarily address topics related to the implementation.

In spite of the prediction errors, the present method is a substantial improvement over other available methods, which ignore the effect of selection on the rate of inbreeding. In the absence of selection, ΔF of a population with overlapping generations is equal to ΔF of a population with discrete generations having the same number of parents entering the population per generation and the same lifetime variance of family size (HILL, 1972, 1979). Following that approach, ΔF was predicted for the schemes in Table 5 with $n_o = 8$, resulting in $\Delta F_{pred} = 0.0022$. (Note that, when ignoring selection, ΔF_{pred} is independent of heritability). In the worst case, the present method showed an underprediction of 20% (0.0073 *vs.* 0.0091), whereas the prediction ignoring selection gives

an error of –76% (0.0022 *vs.* 0.0091). Thus, for the scheme were the present method performs worst, it still accounts for 80% of the true inbreeding, whereas the method ignoring selection accounts for only 24% of the true inbreeding.

The magnitude and pattern of prediction errors in the present study are in line with prediction errors encountered by BIJMA and WOOLLIAMS (2000) for populations with discrete generations. Schemes with many parents show underprediction, which is due to the use of models where the expected genetic contribution is a linear function of the breeding value. Contrary to BIJMA and WOOLLIAMS (2000), we did not include the effect of β when calculating the Poisson correction, *i.e.,* Equation 17 only includes α. Inclusion of β in the Poisson correction would have reduced the underprediction by approximately 5-7% (BIJMA and WOOLLIAMS, unpublished results), but would also complicate the prediction procedure.

Simulation results in discrete generations indicate that the use of a quadratic model for predicting genetic contributions will give accurate predictions of ΔF (BIJMA and WOOLLIAMS, 2000). We expect that this conclusion extends to overlapping generations, but implementation of such a model requires 3rd and 4th moments of the truncated multivariate normal distribution, which are difficult to derive.

With BLUP selection, prediction of the lifetime contribution (required for "*step 2*") is more complicated than with mass selection. With mass selection, the ranking of selection candidates remains unchanged when animals become older, because no additional information is added at older ages. When the ranking of animals remains unchanged over time, the number of animals selected in any combination of age classes can be derived directly from the number of animals selected in each age class. With mass selection, therefore, the lifetime contribution depends directly on the number of animals selected from each age class (BIJMA *et al.* 2000). With BLUP selection, however, the EBV of selection candidates may change when animals become older, because new information (*e.g.,* progeny) becomes available. With different indices in different age classes, the number of animals selected in a particular combination of age classes depends on the proportion selected jointly in those age classes. In the present paper, such bivariate normal proportions were obtained by numerical integration using Dutt's algorithm (DUTT and SOMS, 1976; DUCROCQ and COLLEAU, 1986). Alternatively, bivariate normal proportions can be approximated using the work of MENDEL and ELSTON (1974). Subsequently, the expectation of the squared selective advantage for those categories was calculated using the moment generating function of the truncated multi normal distribution (TALLIS, 1961; Equations C1 – C3).

When there are more than two different indices, *e.g.,* when progeny information accumulates gradually so that each age class has a different amount of information available, the lifetime contribution can still be predicted using the bivariate normal distribution because cross products between any two age classes, $u_{i,k}u_{i,l}$, involve only two categories at a time. With

different amounts of information for each sex-age class, the present method can still be applied but cross products need to be calculated from Equation 16 for all age classes.

In the present study, selection is for a single trait. Predictions for multi-trait selection can be developed using the same methodology. First, a multi-trait pseudo BLUP selection index for populations with overlapping generations has to be developed. This can be done by extending the work of VILLANUEVA *et al.* (1993) to overlapping generations. With multi-trait selection, the selective advantage would consist of the sum of breeding values for the different traits, weighted by their economic value. Subsequently, steps 1 to 3 remain essentially the same, but equations have to be derived based on a multi-trait index. For example, in Equation 6 and 7, the Π-matrix would represent the regression of the aggregate breeding value of a selected offspring on the aggregate breeding value of the parent. All of these steps can be performed using standard selection index theory.

One needs to take care when applying the present methodology to populations with extremely high correlations between estimated breeding values of sibs. For example, in a dairy MOET scheme, selection of young bulls may solely be based on information from relatives, so that between full sib family selection is practiced (ρ_{FS} = 1). For such cases, the equations for calculating the variance of family size and the linear model for predicting expected genetic contributions may be less accurate.

Up till now, optimizing rates of gain and inbreeding in livestock breeding schemes required computationally demanding stochastic simulation, which restricts the number of alternative schemes considered. With the present method, rates of inbreeding in livestock breeding programs can be predicted within very limited computing time, which shows that the general theory of WOOLLIAMS *et al.* (1999) and WOOLLIAMS and BIJMA (2000) provides a toolbox for considering both rates of genetic gain and inbreeding in selected populations. Recently, BIJMA *et al.* (2000b) optimized rates of gain and inbreeding for crossbreeding schemes, which illustrates how the method to predict ΔF can be connected to an optimization procedure. In their study, approximately 750 alternative breeding schemes were evaluated within 9 CPU seconds. This shows that, even when connected to a numerical search algorithm, the method can be used interactively.

Implications

The common animal breeding practice of selection on Estimated Breeding Values using Best Linear Unbiased Prediction has enabled increased rates of genetic gain but will also lead to increased rates of inbreeding, and thus endangers selection response and genetic diversity in the long term. In this paper, we developed a deterministic method to predict rates of inbreeding for livestock breeding populations. The method enables balancing of rates of genetic gain and inbreeding before a breeding scheme commences, and is therefore an important aid to design sustainable animal

breeding plans. This implies that we no longer have to resort to computationally demanding stochastic simulation to balance short and long-term response in livestock breeding populations.

Vincent Ducrocq is acknowledged for providing routines to calculate multivariate normal probabilities. Pim Brascamp, Ab Groen and Theo Meuwissen are acknowledged for giving useful suggestions on this manuscript. This research was financially supported by The Netherlands Technology Foundation (STW) and was coordinated by the Earth and Life Science Foundation (ALW). One author (JAW) gratefully acknowledges the Ministry of Agriculture, Fisheries and Food for financial support.

Literature Cited

BIJMA, P., and J. A. WOOLLIAMS, 1999. Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151:1197-1210.

BIJMA, P., and J. A. WOOLLIAMS, 2000. Prediction of rates of inbreeding in populations selected on Best Linear Unbiased Prediction of breeding value. Genetics (in press).

BIJMA, P., J. A. M. VAN ARENDONK, and J. A. WOOLLIAMS, 2000a. A general procedure to predict rates of inbreeding in populations undergoing mass selection. Genetics 154: 1865-1877.

BIJMA, P., J. A. M. VAN ARENDONK, and J. A. WOOLLIAMS, 2000b. Genetic gain of pure line selection and combined crossbred purebred selection with constrained inbreeding. submitted.

BULMER, M. G., 1971. The effect of selection on genetic variability. Am. Nat. 105:201-211.

BURROWS, P. M., 1984a. Inbreeding under selection from unrelated families. Biometrics 40:357-366.

BURROWS, P. M., 1984b. Inbreeding under selection from related families. Biometrics 40:895-906.

DUCROCQ, V., and J. J. COLLEAU, 1986. Interest in quantitative genetics of Dutt's and Deak's methods for numerical computation of multivariate normal probability integrals. Genet. Sel. Evol. 18:447-474.

DUTT, J. E., 1973. A representation of multivariate probability integrals by integral transforms. Biometrica 60: 637-645.

DUTT, J. E., 1975. On computing the probability integral of a general multivariate t. Biometrika 62: 201-205.

DUTT, J. E., and A. P. SOMS, 1976. An integral representation technique for calculating general multivariate probabilities with an application to multivariate χ^2. Commun. Stat. 11:2195-2205.

FISHER, R. A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Roy. Soc. (Edinburgh), Trans. 52: 321-341.

HENDERSON, C. R., 1963. Selection index and the expected genetic advance. In: W. D. Hanson, and H. F. Robinson (Ed.) Statistical Genetics and Plant Breeding. pp 141-163. Natl. Acad. Sci., Natl. Res. Council Publ. No. 982, Washington, D. C.

HENDERSON, C. R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447.

HILL, W. G., 1972. Effective size of populations with overlapping generations. Theor. Pop. Biol. 3: 278-289.

HILL, W. G., 1974. Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

HILL, W. G., 1979. A note on effective population size with overlapping generations. Genetics 92: 317-322.

JAMES, J. W., and G. MCBRIDE, 1958. The spread of genes by natural and artificial selection in a closed poultry flock. J. Genet. 56:55-62.

MEUWISSEN, T. H. E., 1991. Reduction of selection differentials in finite populations with a nested full-half-sib family structure. Biometrics 47: 195-203.

MENDEL, N. R., and R. C. ELSTON, 1974. Multifactorial qualitative traits: Genetic analyses and prediction of recurrence risks. Biometrics 30:41-57.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. and B. P. FLANNERY, 1992. Numerical Recipes in Fortran. Cambridge University Press, Cambridge, NY, USA.

RIDDERS, C. J. F., 1979. IEEE Transactions on circuits and systems, vol. CAS-26, pp. 979-980.

TALLIS, G.M., 1961. The moment generating function of the truncated multi-normal distribution. J. R. Statist. Soc. B. 23:223-229.

VERRIER, E., J. J. COLLEAU, and J. L. FOULLEY, 1993. Long-term effects of selection based on the animal model

BLUP in a finite population. Theor. Appl. Genet. 87:446-454.

VILLANUEVA, B., N. R. WRAY, and R. THOMPSON, 1993. Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. Anim. Prod. 57:1-13.

WOOLLIAMS, J. A., and P. BIJMA, 2000. Predicting rates of inbreeding in populations undergoing selection. Genetics 154: 1851-1864.

WOOLLIAMS, J. A., N. R. WRAY, and R. THOMPSON, 1993. Prediction of long-term contributions and inbreeding in populations undergoing mass selection. Genet. Res. Camb. 62:231-242.

WOOLLIAMS, J. A., P. BIJMA, and B. VILLANUEVA, 1999. Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153:1009-1020.

WRAY, N. R. and W. G. HILL, 1989. Asymptotic rates of response from index selection. Anim. Prod. 49: 217-227.

WRAY, N. R., and R. THOMPSON, 1990. Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55:41-54.

WRAY, N. R., J. A. WOOLLIAMS, and R. THOMPSON, 1990. Methods for predicting rates of inbreeding in selected populations. Theor. Appl. Genet. 80:503-512.

WRIGHT, S., 1969. Evolution and the genetics of populations. Vol. 2. The Theory of Gene Frequencies. Univ. Chicago, Chicage, USA.

Appendix A

Pseudo-BLUP selection index

There are two indices, Index1 ($\hat{A}_{1,i}$) without progeny information and Index2 ($\hat{A}_{2,i}$) with progeny information. Weights for Index1 (Index2) are, $\mathbf{b}_1 = \mathbf{V}_1^{-1}\,\mathbf{g}_1$ ($\mathbf{b}_2 = \mathbf{V}_2^{-1}\,\mathbf{g}_2$) where $\mathbf{V}_1$ ($\mathbf{V}_2$) is the 6×6 (7×7) covariance matrix of information sources in $\mathbf{x}_{1,i}$ ($\mathbf{x}_{2,i}$) and $\mathbf{g}_1$ ($\mathbf{g}_2$) is the 6×1 (7×1) vector of covariances between information sources in $\mathbf{x}_{1,i}$ ($\mathbf{x}_{2,i}$) and the breeding value of the candidate. Index1 is identical to the index of Wray and Hill (1989).

Elements of $\mathbf{V}_1$ are, per row: $\mathbf{V}_1(1,\ldots) = (\sigma_m^2,\ 0,\ 0,\ \tfrac{1}{2}\sigma_m^2,\ \tfrac{1}{2}\sigma_m^2,\ \tfrac{1}{2}\sigma_m^2)$; $\mathbf{V}_1(2,\ldots) = (0,\ \sigma_f^2,\ \sigma_f^2/d,\ \tfrac{1}{2}\sigma_f^2/d,\ \tfrac{1}{2}\sigma_f^2,\ \tfrac{1}{2}\sigma_f^2)$; $\mathbf{V}_1(3,\ldots) = (0,\ \sigma_f^2/d,\ \sigma_f^2/d,\ \tfrac{1}{2}\sigma_f^2/d,\ \tfrac{1}{2}\sigma_f^2/d,\ \tfrac{1}{2}\sigma_f^2/d)$; $\mathbf{V}_1(4,\ldots) = (\tfrac{1}{2}\sigma_m^2,\ \tfrac{1}{2}\sigma_f^2/d,\ \tfrac{1}{2}\sigma_f^2/d,\ \sigma_{HS}^2,\ \sigma_{HS}^2,\ \sigma_{HS}^2)$; $\mathbf{V}_1(5,\ldots) = (\tfrac{1}{2}\sigma_m^2,\ \tfrac{1}{2}\sigma_f^2,\ \tfrac{1}{2}\sigma_f^2/d,\ \sigma_{HS}^2,\ \sigma_{FS}^2,\ \sigma_{FS}^2)$; $\mathbf{V}_1(6,\ldots) = (\tfrac{1}{2}\sigma_m^2,\ \tfrac{1}{2}\sigma_f^2,\ \tfrac{1}{2}\sigma_f^2/d,\ \sigma_{HS}^2,\ \sigma_{FS}^2,\ \sigma_{self}^2)$, where σ_m^2 and σ_f^2 are the variance of the EBV among selected sires and dams respectively; σ_{HS}^2 and σ_{FS}^2 are the between full sib family and between half sib family variance, which are, $\sigma_{HS}^2 = \tfrac{1}{4}\sigma_{A,m}^2 + \tfrac{1}{4}\sigma_{A,f}^2/d + (\tfrac{1}{2}\sigma_{A,0}^2 + \sigma_E^2)/n_o d$ and $\sigma_{FS}^2 = \tfrac{1}{4}\sigma_{A,m}^2 + \tfrac{1}{4}\sigma_{A,f}^2 + (\tfrac{1}{2}\sigma_{A,0}^2 + \sigma_E^2)/n_o$ where $\sigma_{A,m}^2$ ($\sigma_{A,f}^2$) is the genetic variance among the selected sires (dams), σ_E^2 is the environmental variance, $\sigma_{A,0}^2$ is the base generation additive genetic variance and $\sigma_{self}^2 = \sigma_A^2 + \sigma_E^2$, where σ_A^2 is the total additive genetic variance. Furthermore, $\mathbf{g}_1^T = (\tfrac{1}{2}\sigma_m^2,\ \tfrac{1}{2}\sigma_f^2,\ \tfrac{1}{2}\sigma_f^2/d,\ \tfrac{1}{4}\sigma_{A,m}^2 + \tfrac{1}{4}\sigma_{A,f}^2/d + \tfrac{1}{2}\sigma_{A,0}^2/n_o d,\ \tfrac{1}{4}\sigma_{A,m}^2 + \tfrac{1}{4}\sigma_{A,f}^2 + \tfrac{1}{2}\sigma_{A,0}^2/n_o,\ \sigma_A^2)$

Matrix $\mathbf{V}_2$ is identical to $\mathbf{V}_1$ but has additional row and column, $\mathbf{V}_2(7,\ldots) = \mathbf{V}_2(\ldots,7) = [\tfrac{1}{4}\sigma_m^2,\ \tfrac{1}{4}\sigma_f^2,\ \tfrac{1}{4}\sigma_f^2/d,\ \mathrm{Cov}(\bar{P}_{HS},\ \bar{P}_{prg}),\ \mathrm{Cov}(\bar{P}_{FS},\ \bar{P}_{prg}),\ \tfrac{1}{2}\sigma_A^2,\ \mathrm{Var}(\bar{P}_{prg})]$, where $\mathrm{Cov}(\bar{P}_{HS},\ \bar{P}_{prg}) = \tfrac{1}{8}\sigma_{A,m}^2 + \tfrac{1}{8}\sigma_{A,f}^2/d + \tfrac{1}{4}\sigma_{A,0}^2/n_o d$, $\mathrm{Cov}(\bar{P}_{FS},\ \bar{P}_{prg}) = \tfrac{1}{8}\sigma_{A,m}^2 + \tfrac{1}{8}\sigma_{A,f}^2 + \tfrac{1}{4}\sigma_{A,0}^2/n_o$, and $\mathrm{Var}(\bar{P}_{prg}) = \tfrac{1}{4}\sigma_A^2(1 + 1/d_{prg}) + (\tfrac{1}{2}\sigma_{A,0}^2 + \sigma_E^2)/n_{prg}$, where d_{prg} is number of dams mated to a single sire when producing

offspring for the progeny test. Assuming the same litter size as in the nucleus, $d_{prg} = n_{prg}/n_o$. Furthermore, $\mathbf{g}_2^T$ is identical to $\mathbf{g}_1^T$, but has additional $\mathbf{g}_2^T(7) = \frac{1}{2}\sigma_A^2$.

Genetic variance among selected parents of sex x was $\sigma_{A,x}^2 = \sum_k \sigma_A^2(1-\kappa_k\rho_k^2)\frac{n_k}{N_x} + \sum_k(\mu_k-\mu)^2\frac{n_k}{N_x}$, with $k = 1, c_{max}$ for sires and $k = c_{max}+1, 2c_{max}$ for dams, where κ_k is Pearsons variance reduction coefficient [$\kappa_k = \iota_k(\iota_k-\tau_k)$, where ι is selection intensity and τ is the standardized truncation point], ρ_k is the accuracy of selection, μ_k is the genetic mean of the parents selected from category k expressed as a deviation from an arbitrary base, $\mu_k = \iota_k\rho_k\sigma_A - k\Delta G_y$, where ΔG_y is the rate of genetic gain per year, and μ is the overall mean of the selected group for sex x, $\mu = \sum_k \mu_k \frac{n_k}{N_x}$. In the equation for $\sigma_{A,x}^2$, the first term represents the weighted sum of the within age-class genetic variance after selection and the second term represents the additional variance due to differences between the mean breeding values of parents selected from different age classes.

The variance of the EBV among selected parents, σ_m^2 and σ_f^2, was $\sigma_x^2 = \sum_k \sigma_{\hat{A},k}^2(1-\kappa_k)\frac{n_k}{N_x} + \sum_k(\mu_k-\mu)^2\frac{n_k}{N_x}$, where $x = m$ or f, $\sigma_{\hat{A},k}^2$ is the variance of the EBV among selection candidates in category k.

Accuracy of selection for category k was, $\rho_k = \sqrt{\mathbf{b}_{index(k)}^T\mathbf{g}_{index(k)}/\sigma_A^2}$, where index($k$) = 1 for categories without progeny testing and index(k) = 2 for categories with progeny testing. The variance of EBV was, $\sigma_{\hat{A},k}^2 = \rho_k^2\sigma_A^2$. Genetic gain per year was, $\Delta G_y = \frac{1}{2}\sigma_A \sum_{k=1}^{2c_{max}} \frac{\iota_k\rho_k n_k}{N_{sex(k)}L_0}$, where L_0 is the generation interval based on the average age of parents of newborn offspring, $L_0 = \frac{1}{2}\sum_{k=1}^{2c_{max}} age(k)n_k / N_{sex(k)}$.

Each generation, additive genetic variance was calculated from $\sigma_A^2 = \frac{1}{4}\sigma_{A,m}^2 + \frac{1}{4}\sigma_{A,f}^2 + \frac{1}{2}\sigma_{A,0}^2$. The above equations were iterated until equilibrium variances were reached (approx. 10 iterations).

Intraclass correlations: Three types of intra-class correlations between EBV of sibs can be distinguished. First, between $\hat{A}_{1,i}$ of two sibs with neither progeny tested, second, between $\hat{A}_{2,i}$ of two progeny tested sibs, and finally between one tested and one not.

Between full sibs in not-progeny tested categories: $\rho_{FS,kl} = \mathbf{b}_1^T\mathbf{C}_{FS,kl}\mathbf{b}_1/\sigma_{\hat{A}_1}^2$, where k and l denote the categories of both individuals, $\mathbf{C}_{FS,kl}$ is the 6×6 covariance matrix between the information sources of an individual in category k and the information sources of its full sib in category l. Matrix $\mathbf{C}_{FS,kl}$ is identical to $\mathbf{V}_1$ except for $\mathbf{C}_{FS,kl}(6,6) = \frac{1}{4}\sigma_{A,m}^2 + \frac{1}{4}\sigma_{A,f}^2$. Between full sibs in progeny tested categories, $\rho_{FS,kl} = \mathbf{b}_2^T\mathbf{C}_{FS,kl}\mathbf{b}_2/\sigma_{\hat{A}_2}^2$, where $\mathbf{C}_{FS,kl}$ is a 7×7 covariance matrix which is identical to $\mathbf{V}_2$ except for $\mathbf{C}_{FS,kl}(6,6) = \frac{1}{4}\sigma_{A,m}^2 + \frac{1}{4}\sigma_{A,f}^2$, $\mathbf{C}_{FS,kl}(6,7) = \mathbf{C}_{FS,kl}(7,6) = \frac{1}{8}\sigma_{A,m}^2 + \frac{1}{8}\sigma_{A,f}^2$ and $\mathbf{C}_{FS,kl}(7,7) = \frac{1}{16}\sigma_{A,m}^2 + \frac{1}{16}\sigma_{A,f}^2$. Between a progeny tested individual and its not-progeny tested full sib, $\rho_{FS,kl} = \mathbf{b}_1^T\mathbf{C}_{FS,kl}\mathbf{b}_2/\sigma_{\hat{A}_1}\sigma_{\hat{A}_2}$, where $\mathbf{C}_{FS,kl}$ is a 6×7 covariance matrix which is identical to the sub-matrix $\mathbf{C}_{FS,kl}(1\text{-}6;1\text{-}7)$ for the case of progeny testing in both categories.

Between half sibs in not-progeny tested categories, $\rho_{HS,kl} = \mathbf{b}_1^T \mathbf{C}_{HS,kl} \mathbf{b}_1 / \sigma^2_{\hat{A}_1}$, where $\mathbf{C}_{HS,kl}$ is identical to $\mathbf{V}_1$ except for $\mathbf{C}_{HS,kl}(5\text{-}6,5\text{-}6) = \frac{1}{4}\sigma^2_{A,m}$, $\mathbf{C}_{HS,kl}(2,5\text{-}6) = \mathbf{C}_{HS,kl}(5\text{-}6,2) = 0$ and $\mathbf{C}_{HS,kl}(2,2) = 0$. Between half sibs in progeny tested categories, $\rho_{HS,kl} = \mathbf{b}_2^T \mathbf{C}_{HS,kl} \mathbf{b}_2 / \sigma^2_{\hat{A}_2}$, where $\mathbf{C}_{HS,kl}$ is identical to $\mathbf{V}_2$ except for $\mathbf{C}_{HS,kl}(5\text{-}6,5\text{-}6) = \frac{1}{4}\sigma^2_{A,m}$, $\mathbf{C}_{HS,kl}(2,5\text{-}6) = \mathbf{C}_{HS,kl}(5\text{-}6,2) = 0$ and $\mathbf{C}_{HS,kl}(2,2) = 0$, $\mathbf{C}_{HS,kl}(2,7) = \mathbf{C}_{HS,kl}(7,2) = 0$, $\mathbf{C}_{HS,kl}(5\text{-}6,7) = \mathbf{C}_{HS,kl}(7,5\text{-}6) = \frac{1}{8}\sigma^2_{A,m}$, $\mathbf{C}_{HS,kl}(7,7) = \frac{1}{16}\sigma^2_{A,m}$.

Between a progeny tested individual and its not-progeny tested half sib, $\rho_{HS,kl} = \mathbf{b}_1^T \mathbf{C}_{HS,kl} \mathbf{b}_2 / \sigma_{\hat{A}_1}\sigma_{\hat{A}_2}$, where $\mathbf{C}_{HS,kl}$ is the (6×7) covariance matrix which is identical to the sub-matrix $\mathbf{C}_{HS,kl}(1\text{-}6;1\text{-}7)$ for the case of progeny testing in both categories.

Intraclass correlations were corrected for the number of families being finite, using the empirical correction of BIJMA and WOOLLIAMS (2000), $\overline{\rho}_{FS,kl} = \rho_{FS,kl} - \rho_{FS,kl}(1-\rho^2_{FS,kl})(0.8643/N_m + 0.9540/N_f)$ and $\overline{\rho}_{HS,kl} = \rho_{HS,kl} - \rho_{HS,kl}(1-\rho^2_{HS,kl})(1.4075/N_m + 1.4581/N_f)$.

Appendix B

Prediction of expected genetic contributions

The Π and Λ matrix: Elements of Π and Λ are a multiple category analogy of the discrete generations case (see BIJMA and WOOLLIAMS, 2000):

$$\pi_{kl} = \tfrac{1}{2} - \mathbf{b}^T_{index(k)}\, \mathbf{c}_{index(k),l} \kappa_k / \sigma^2_{s,l} \qquad k,l = 1, 2c_{max} \qquad \text{(B1)}$$

$$\lambda_{kl} = \mathbf{b}^T_{index(k)}\, \mathbf{c}_{index(k),l} l_k / (\sigma^2_{s,l}\ \sigma_{\hat{A},k}) \qquad k,l = 1, 2c_{max} \qquad \text{(B2)}$$

where $index(k) = 1$ for categories without progeny testing and 2 for categories with progeny testing, and $\mathbf{c}_{index(k),l}$ is a vector of covariances between information sources of an offspring in category k and the selective advantage of its parent in category l. For offspring categories without progeny testing, $\mathbf{c}_{1,l} = [\sigma^2_{\hat{A},l}(1-\kappa_l),\ \overline{\sigma^2_{\hat{A},f}}/d,\ \overline{\sigma^2_{\hat{A},f}}/d,\ \frac{1}{2}\sigma^2_A(1-\kappa_l\rho_l^2)+\frac{1}{2}\overline{\sigma^2_{A,f}}/d,\ \frac{1}{2}\sigma^2_A(1-\kappa_l\rho_l^2) + \frac{1}{2}\overline{\sigma^2_{A,f}}/d,\ \frac{1}{2}\sigma^2_A(1-\kappa_l\rho_l^2) + \frac{1}{2}\overline{\sigma^2_{A,f}}/d]$ for $l = 1, c_{max}$; and $\mathbf{c}_{1,l} = [\overline{\sigma^2_{\hat{A},m}},\ \sigma^2_{\hat{A},l}(1-\kappa_l),\ \sigma^2_{\hat{A},l}(1-\kappa_l)/d,\ \frac{1}{2}\sigma^2_A(1-\kappa_l\rho_l^2)/d + \frac{1}{2}\overline{\sigma^2_{A,m}},\ \frac{1}{2}\sigma^2_A(1-\kappa_l\rho_l^2) + \frac{1}{2}\overline{\sigma^2_{A,m}},\ \frac{1}{2}\sigma^2_A(1-\kappa_l\rho_l^2) + \frac{1}{2}\overline{\sigma^2_{A,m}}]$ for $l = c_{max}+1, 2c_{max}$. For offspring categories with progeny testing, $\mathbf{c}_{2,l} = [\mathbf{c}_{1,l}(1\text{-}6),\ \frac{1}{4}\sigma^2_A(1-\kappa_l\rho_l^2)+\frac{1}{4}\overline{\sigma^2_{A,f}}/d]$ for $l = 1, c_{max}$; and $\mathbf{c}_{2,l} = [\mathbf{c}_{1,l}(1\text{-}6),\ \frac{1}{4}\sigma^2_A(1-\kappa_l\rho_l^2)+\frac{1}{4}\overline{\sigma^2_{A,m}}]$ for $l = c_{max}+1, 2c_{max}$. Furthermore, $\overline{\sigma^2_{A,m}} = \sigma^2_A \sum_{l=1}^{c_{max}} \frac{n_l}{N_m}(1-\kappa_l\rho_l^2)$, $\overline{\sigma^2_{A,f}} = \sigma^2_A \sum_{l=c_{max}+1}^{2c_{max}} \frac{n_l}{N_f}(1-\kappa_l\rho_l^2)$, $\overline{\sigma^2_{\hat{A},m}} = \sum_{l=1}^{c_{max}} \frac{n_l}{N_m}\sigma^2_{\hat{A},l}(1-\kappa_l)$ and $\overline{\sigma^2_{\hat{A},f}} = \sum_{l=c_{max}+1}^{2c_{max}} \frac{n_l}{N_f}\sigma^2_{\hat{A},l}(1-\kappa_l)$.

The G matrix: Elements of G are (BIJMA and WOOLLIAMS, 1999):

$$g_{kl} = \tfrac{1}{2}\frac{p_{kl}}{p_k}\frac{n_l}{N_{sex(l)}} \qquad k,l = 1, 2c_{max} \qquad \text{(B3)}$$

where p_{kl} is the selected proportion among offspring in category k descending from parents in category l. Solutions for p_{kl} were obtained separate for each parent sex, by simultaneously solving the equations $p_k = \sum_l p_{kl} n_l / N_{sex(l)}$ and $p_{kl} = 1 - \Phi[(I_k - \frac{1}{2}\mu_l)/\sigma_{\hat{A},kl}]$, using the algorithm

RIDDR_ROOT (RIDDERS, 1979; Press *et al.* 1992), where $l = 1, c_{max}$ for sires, $l = c_{max}+1, 2c_{max}$ for dams, I_k is the EBV truncation point common to all offspring in category k, $\sigma_{\hat{A},kl}$ is the standard deviation of the EBV for offspring in category k descending from parents in category l, μ_l is given in Appendix A and Φ denotes the cumulative normal distribution function (BIJMA and WOOLLIAMS, 1999). The standard deviation of the EBV, $\sigma_{\hat{A},kl}$, was calculated analogous to $\sigma_{\hat{A},k}$ by using the pseudo-BLUP index (Appendix A), but replacing $\sigma^2_{A,m}$ ($\sigma^2_{A,f}$) by $\sigma^2_A(1-\kappa_l\rho_l^2)$ and σ^2_m (σ^2_f) by $\sigma^2_{\hat{A},l}(1-\kappa_l)$ when the parent is a male (female).

The D matrix: Elements of **D** were obtained, separate for each parent sex, from (BIJMA and WOOLLIAMS, 1999),

$$d_{kl} = \tfrac{1}{2}\mu_l + \iota_{kl}\sigma_{\hat{A},kl} - \bar{d}_k \qquad \text{(B4)}$$

where μ_l is given in Appendix A, ι_{kl} is the selection intensity corresponding to p_{kl} (see above) and the second term represents subtraction of the average: $\bar{d}_k = \sum_l (\tfrac{1}{2}\mu_l + \iota_{kl}\sigma_{\hat{A},kl})$, with $l = 1, c_{max}$ for sires, $l = c_{max}+1, 2c_{max}$ for dams.

Appendix C

Step 2, expectation of squared expected contributions.

Equation 16: $E[(A_i-\bar{A}_k)(A_i-\bar{A}_{c_{max}})] = E(A_i^2) - E(A_i\bar{A}_k) - E(A_i\bar{A}_{c_{max}}) + E(\bar{A}_k\bar{A}_{c_{max}})$, where i refers to one of the $n_{k,c_{max}}$ individuals that are selected both in category k and in category c_{max}. When deriving those expectations, one has to account for the fact that $\bar{A}_k$ and $\bar{A}_{c_{max}}$ are finite sample means. For example, $E(A_i\bar{A}_k) = E[A_i(A_{1,k} + ...+ A_{n_k,k})]/n_k = [(n_k-1)E(A_iA_{j\neq i,k}) + E(A_i^2)]/n_k \approx (n_k-1)E(A_i)\bar{A}_k/n_k + E(A_i^2)/n_k$. Following this approach, it can be shown that:

$$\begin{aligned}E[(A_i-\bar{A}_k)(A_i-\bar{A}_{c_{max}})] = \\ [(n_k n_{c_{max}} - n_k - n_{c_{max}} + n_{k,c_{max}})/(n_k n_{c_{max}})]\,E(A_i^2) \\ - [(n_k-1)/n_k]\,E(A_i)\,\bar{A}_k \\ - [(n_{c_{max}}-1)/n_{c_{max}}]\,E(A_i)\,\bar{A}_{c_{max}} \\ + [1 - n_{k,c_{max}}/(n_k n_{c_{max}})]\,\bar{A}_k\,\bar{A}_{c_{max}} \qquad \text{(C1)}\end{aligned}$$

where $\bar{A}_k = \iota_k\rho_k\sigma_A$. This equation has the desired property that it reduces to zero when $n_k = n_{c_{max}} = n_{k,c_{max}} = 1$ and that it reduces to $E(A_i^2) - E(A_i)\bar{A}_k - E(A_i)\bar{A}_{c_{max}} + \bar{A}_k\bar{A}_{c_{max}}$ when $n_k, n_{c_{max}}, n_{k,c_{max}} \to \infty$.

The terms $E(A_i)$ and $E(A_i^2)$ represent the expectation of the breeding value and the squared breeding value of individuals that are selected both in category k and in category c_{max}, *e.g.*, $E(A_i) = E(A_i|\hat{A}_{i,k} > \tau_k, \hat{A}_{i,c_{max}} > \tau_{c_{max}})$, which can be obtained from the moment generating function of the truncated multivariate normal distribution (TALLIS, 1961). From the first and second equations on page 226 of TALLIS (1961), with, in the notation of TALLIS, $a_1 = A_{.1} = A'_{.1} = -\infty$, it follows that:

$$\begin{aligned}E(A_i) = \sigma_A[\rho_k\phi(\tau_k)\Phi(\psi_{k,c_{max}}) \\ + \rho_{c_{max}}\phi(\tau_{c_{max}})\Phi(\psi_{c_{max},k})]\,p^{-1}_{k,c_{max}} \qquad \text{(C2)}\end{aligned}$$

and

$$\begin{aligned}E(A_i^2) = \sigma_A^2\{p_{k,c_{max}} + \rho_k^2\tau_k\phi(\tau_k)\Phi(\psi_{k,c_{max}}) \\ + \rho^2_{c_{max}}\tau_{c_{max}}\phi(\tau_{c_{max}})\Phi(\psi_{c_{max},k}) \\ + \phi(\tau_k,\tau_{c_{max}},\rho_{k,c_{max}})[\rho_k(\rho_{c_{max}} - \rho_k\rho_{k,c_{max}}) \\ + \rho_{c_{max}}(\rho_k - \rho_{c_{max}}\rho_{k,c_{max}})]\}\,p^{-1}_{k,c_{max}} \qquad \text{(C3)}\end{aligned}$$

where ρ_k is the accuracy of selection in category k, $\rho_{k,c_{\max}}$ is the correlation between the index of an individual in category k and its index in category c_{max}, $\rho_{k,c_{\max}} = \sigma_{\hat{A},k} / \sigma_{\hat{A},c_{\max}}$, $\phi(\tau)$ is the univariate normal density function, $\phi(\tau) = (2\pi)^{-½} e^{-½\tau^2}$, $\phi(\tau_k, \tau_l, \rho_{k,l})$ is the bivariate normal density function, $\phi(\tau_k, \tau_l, \rho_{k,l}) = \left(2\pi\sqrt{1-\rho_{k,l}^2}\right)^{-1} e^{-½q}$ with $q = (\tau_k^2 - 2\rho_{k,l}\tau_k\tau_l + \tau_l^2)/(1-\rho_{k,l}^2)$, $\Phi(\psi)$ is the univariate normal upper tail proportion, $\Phi(\psi) = \int_{\psi}^{\infty} \phi(x)dx$ and $\psi_{k,l} = (\tau_l - \rho_{k,l}\tau_k)(1-\rho_{k,l}^2)^{-½}$. Note that $\psi_{k,l} \neq \psi_{l,k}$.

Appendix D

Calculation of δ

General equations for calculating $\Delta\mathbf{V}_{n,k}$ are given in Appendix E of WOOLLIAMS and BIJMA (2000), and can also be obtained by extending the equations for BLUP selection with discrete generations (BIJMA and WOOLLIAMS, 2000) to populations with overlapping generations. To keep notation as short as possible, the number of selected offspring conditional on the selective advantage, $n_{ij,k}(l)|s_{i,k}$, is abbreviated by $n_{ij,k}(l)$.

For sires, $\Delta\mathbf{V}_{n,k}$ is calculated from:

$$\Delta\mathbf{V}_{n,k}(l,l) = \mathrm{E}_s\{n_{i*,k}(l)[n_{i*,k}(l)-1]\} - \mathrm{E}_s[\mu_k^2(l)] \quad k = 1, c_{max};\ l = 1, 2c_{max} \qquad \text{(D1)}$$

$$\Delta\mathbf{V}_{n,k}(l,l') = \mathrm{E}_s[n_{i*,k}(l)n_{i*,k}(l')] - \mathrm{E}_s[\mu_k(l)\mu_k(l')] \quad k = 1, c_{max};\ l,l' = 1, 2c_{max};\ l \neq l' \qquad \text{(D2)}$$

and for dams from:

$$\Delta\mathbf{V}_{n,k}(l,l) = \mathrm{E}_s\{n_{ij,k}(l)[n_{ij,k}(l)-1]\} - \mathrm{E}_s[\mu_k^2(l)] \quad k = c_{max}+1, 2c_{max};\ l = 1, 2c_{max} \qquad \text{(D3)}$$

$$\Delta\mathbf{V}_{n,k}(l,l') = \mathrm{E}_s[n_{ij,k}(l)n_{ij,k}(l')] - \mathrm{E}_s[\mu_k(l)\mu_k(l')] \quad k = c_{max}+1, 2c_{max};\ l,l' = 1, 2c_{max};\ l \neq l' \qquad \text{(D4)}$$

where $n_{i*,k}(l)$ is the number of selected offspring in category l from the i^{th} sire in category k and $n_{ij,k}(l)$ is the number of selected offspring in category l from the j^{th} dam in category k that is mated to sire i, (*i.e.*, n_{i*} represents the sire family size and n_{ij} represents the dam family size). In Equation D1 and D2, $\mu_k(l) = \mathrm{E}[n_{i*,k}(l)|s_{i,k}]$, which is the expected number of selected offspring in category l of sire i in category k, given its selective advantage. In Equation D3 and D4, $\mu_k(l) = \mathrm{E}[n_{ij,k}(l)|s_{j,k}]$.

Elements of Equations D1 to D4 are:

$$\mathrm{E}_s[\mu_k^2(l)] = 4g_{lk}^2 n_l^2 n_k^{-2}(1+\lambda_{lk}^2\sigma_{s_k}^2) \quad k, l = 1, 2c_{max} \qquad \text{(D5)}$$

$$\mathrm{E}_s[\mu_k(l)\mu_k(l')] = 4g_{lk}g_{l'k}n_l n_{l'} n_k^{-2}(1+\lambda_{lk}\lambda_{l'k}\sigma_{s_k}^2) \quad k, l, l' = 1, 2c_{max};\ l \neq l' \qquad \text{(D6)}$$

$$\mathrm{E}_s\{n_{i*,k}(l)[n_{i*,k}(l)-1]\} = d\mathrm{E}_s\{n_{ij,k}(l)[n_{ij,k}(l)-1]\} + d(d-1)\mathrm{E}_s[n_{ij,k}(l)n_{ij',k}(l)] \quad k = 1, c_{max};\ l = 1, 2c_{max} \qquad \text{(D7)}$$

$$\mathrm{E}_s[n_{i*,k}(l)n_{i*,k}(l')] = d\mathrm{E}_s[n_{ij,k}(l)n_{ij,k}(l')] + d(d-1)\mathrm{E}_s[n_{ij,k}(l)n_{ij',k}(l')] \quad k = 1, c_{max};\ l = 1, 2c_{max};\ l \neq l' \qquad \text{(D8)}$$

where j' is another dam than j,

$$\mathrm{E}_s\{n_{ij,k}(l)[n_{ij,k}(l)-1]\} = \frac{2n_{sex(l)}(n_{sex(l)}-1)g_{lk}n_l(2g_{lk}n_l-1)R(p_{lk}, p_{lk}, \bar{\rho}_{FS,ll})}{n_k N_{sex(k)}^{-1} T_{sex(l)}(n_k N_{sex(k)}^{-1} T_{sex(l)} - 1)} \qquad \text{(D9)}$$

$$\mathrm{E}_s[n_{ij,k}(l)n_{ij,k}(l')] = \frac{4n_{sex(l)}n_{sex(l')}g_{lk}n_l g_{l'k}n_{l'}R(p_{lk}, p_{l'k}, \bar{\rho}_{FS,ll'})}{n_k^2 N_{sex(l)}^{-2} T_{sex(l)} T_{sex(l')}} \qquad \text{(D10)}$$

$$\mathrm{E}_s[n_{ij,k}(l)n_{ij',k}(l)] = \frac{2n_{sex(l)}^2 g_{lk}n_l(2g_{lk}n_l-1)R(p_{lk}, p_{lk}, \bar{\rho}_{HS,ll})}{n_k N_{sex(k)}^{-1} T_{sex(l)}(n_k N_{sex(k)}^{-1} T_{sex(l)} - 1)} \qquad \text{(D11)}$$

$$E_s[n_{ij,k}(l)n_{ij',k}(l')] = \frac{4n_{sex(l)}n_{sex(l')}g_{lk}n_l g_{l'k}n_{l'}R(p_{lk}, p_{l'k}, \bar{\rho}_{HS,ll'})}{n_k^2 N_{sex(k)}^{-2} T_{sex(l)} T_{sex(l')}} \quad \text{(D12)}$$

where $T_{sex(l)}$ is the total number of candidates of the sex of category l, and $n_{sex(l)}$ is the number of offspring of $sex(l)$ born per dam ($T_m = T_f = T$ and $n_m = n_f = \frac{1}{2}n_o$ for the current breeding scheme). Note that Equations D9 and D10 are used both in equations D3 and D4 and in equations D7 and D8; in D3 and D4 for dam categories ($k = c_{max}+1,2c_{max}$) and in D7 and D8 for sire categories ($k = 1, c_{max}$).

Furthermore, from MENDEL and ELSTON (1974), $R(p_{lk}, p_{l'k}, \bar{\rho}_{sibs,ll'}) = p_{lk}/\Phi[(i_{l'k}\bar{\rho}_{sibs,ll'} - \tau_{lk})(1-\kappa_{l'k}\bar{\rho}^2_{sibs,ll'})^{-\frac{1}{2}}]$ where Φ is the cumulative normal distribution function and τ_{lk} is the standardized truncation point for offspring in category l descending from parents in category k. When $l' \neq l$, the most accurate value is obtained by using l for the category with the smallest selection intensity (WRAY *et al.* 1994).

In situations where the number of selected parents is small compared to the number of selection candidates per family, Equations D9 to D12 can give substantial bias. For those cases, accuracy of Equations D9 to D12 can be improved by adjusting the selected proportion according to WRAY et al. (1990). Adjusted selected proportions for sire categories were (see Appendix B for unadjusted selected proportions),

$$p_{kl,adj} = (1 - \bar{\rho}_{sibs,kl})p_{kl} + \bar{\rho}_{sibs,kl}\max(p_{kl}, 1/n_l) \qquad l = 1, c_{max} \quad \text{D(13)}$$

and for dam categories

$$p_{kl,adj} = (1 - \bar{\rho}_{sibs,kl})p_{kl} + \bar{\rho}_{sibs,kl}\max[p_{kl}, 1/(n_l/d)] \qquad l = c_{max}+1, 2c_{max} \quad \text{D(14)}$$

In the present study, this adjustment was applied to all schemes where, for any of the sire categories, $\min(n_k) \leq 1.5n_o$.

Maximising Genetic Gain for the Sire Line of a Crossbreeding Scheme utilising both Purebred and Crossbred Information

Piter Bijma and Johan A. M. Van Arendonk

Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands.

Abstract – A selection index procedure which utilises both purebred and crossbred information was developed for the sire line of a three-way crossbreeding scheme in pigs, to predict response to BLUP selection with an animal model. Purebred and crossbred performance were treated as correlated traits. The breeding goal was crossbred performance, but methods can be applied to other goals. A hierarchical mating structure was used. Sires were mated to purebred dams to produce replacements, and to F1's from the dam line to produce fattening pigs. Generations were discrete and inbreeding was ignored. The selection index included purebred and crossbred phenotypic information of the current generation and estimated breeding values for purebred and crossbred performance of parents and mates of parents from the previous generation. Reduction of genetic variance due to linkage disequilibrium and reduction of selection intensity due to finite population size and due to correlated index values was accounted for. Selection was undertaken until asymptotic responses were reached. The index was used to optimise the number of selected parents per generation and the number of offspring tested per litter, and to make inferences on the value of crossbred information. It was optimal to test a maximum number of offspring per litter, mainly due to increased female selection intensities. Maximum response reductions due to linkage disequilibrium and correlated index values were 32% and 29% respectively. Correcting for correlated index values changed ranking of breeding schemes. Benefit of crossbred information was largest when the genetic correlation between purebred and crossbred performance was low. Due to high correlations between index values in that case, the optimum number of selected sires increased considerably when crossbred information was included.

In animal breeding, optimisation of breeding schemes is an important activity of breeding organisations. This optimisation is complicated by several factors, as for example the use of family information, reduction of genetic variance due to linkage disequilibrium (BULMER, 1971) and correlation between index values of selection candidates. Because of this complexity, stochastic simulation is often used to optimise breeding schemes. However, stochastic simulation is computationally demanding. In this study, deterministic methods are applied in order to allow for a large number of alternative schemes to be investigated within limited computing time.

Breeding value estimation using Best Linear Unbiased Prediction of breeding values (BLUP-EBV) under an animal model is a common

procedure for genetic evaluation nowadays (*e.g.*, KENNEDY *et al.* 1988). A deterministic method for predicting response to selection on BLUP-EBV under an animal model was presented by WRAY and HILL (1989). A multitrait extension of this method was presented by VILLANUEVA *et al.* (1993). The method accounted for reduction of genetic variance due to gametic phase disequilibrium induced by selection (BULMER, 1971). Wray and HILL (1989) and DEKKERS (1992) found that ranking of breeding schemes was little affected by the Bulmer effect when selection was on BLUP-EBV under an animal model and when generations were discreet. However, as shown by MEUWISSEN (1991), ranking of breeding schemes can change due to reduction of selection intensity when index values of selection candidates are correlated.

In pig and poultry breeding, information from crossbred animals is becoming available which has increased the interest of utilising crossbred information in genetic evaluation. To optimise breeding schemes utilising both purebred and crossbred information, methods to predict response under combined crossbred purebred selection (CCPS) are crucial. In these methods, purebred and crossbred performance should be treated as genetically different traits, because empirical evidence for differences between purebred and crossbred genetic parameters has been found (*e.g.*, BROWN and BELL, 1980; VAN DER WERF, 1990; WEI and VAN DER STEEN, 1991). Theoretical explanations of this phenomenon based on one and two locus models were given by WEI *et al.* (1991a,b) and BAUMUNG (1997). Differences between purebred and crossbred traits can also arise from genotype by environment interaction, observed especially for carcass traits (MERKS, 1986). WEI and VAN DER WERF (1994) described a selection index procedure for crossbreeding, which utilises both crossbred and purebred information (CCPS). They showed that this index resulted in higher response than pure line selection (PLS) or reciprocal recurrent selection. However, their index is not expected to predict response to selection on BLUP-EBV under an animal model accurately, because the index did not include pedigree information and the Bulmer effect was ignored.

The set-up of this paper is threefold. First, a selection index procedure that predicts response to selection on BLUP-EBV under an animal model is developed for the sire line of a three-path crossbreeding scheme in pigs. Purebred and crossbred performance were treated as genetically correlated traits. The procedure accounts for reduction of the selection intensity due to finite sample size and correlated index values, and for reduction of the genetic variance due to the Bulmer effect. Second, breeding schemes were optimised for different fixed test capacities in order to make optimum use of the available test capacity. Finally, inferences were made on the effect of using crossbred information on the response and on the layout of optimum breeding schemes.

Methods

Breeding scheme: An index was derived for selection in the sire line of a three-path

crossbreeding scheme in pigs. A hierarchical mating structure was assumed, generations were discrete and inbreeding was ignored. Sires from the sire line were mated to n_{pd} purebred dams of the sire line to produce n_{po} purebred offspring per dam. Sires were also mated to n_{cd} F1-dams from the multiplier level of the dam line, to produce n_{co} crossbred fattening pigs per F1-dam. Dams produced only one litter. Sires were used simultaneously for the production of pure line replacements and for the production of crossbred fattening pigs. Purebred performance was measured on animals within the sire line, crossbred performance was measured on fattening pigs. Purebred and crossbred information was assumed to be available at the same time and was measured on both sexes. Each generation, n_s males and $n_s n_{pd}$ females were selected among the $n_s n_{pd} n_{po}$ purebred selection candidates of the current generation of the sire line, to become sires and dams of the next generation. Selection was performed by truncation on pseudo-BLUP EBV.

Breeding goal: The breeding goal (H) was defined as breeding value for crossbred performance. However, because both purebred and crossbred breeding values were needed as information sources in the selection index, the breeding goal included both purebred and crossbred breeding value and zero economic weight was given to the purebred breeding value. Thus the breeding goal was,

$$H = \mathbf{a}'\mathbf{u},$$

where $\mathbf{a}' = (0\ 1)$, which is a vector of economic weights, and $\mathbf{u}$ is a 2×1 vector containing the breeding value for purebred and crossbred performance. Including both traits in $\mathbf{u}$ also allows for the use of other economic values without having to change the selection index equations.

Selection index: An individual animal model was used with one record per animal. Phenotypic observations consisted of additive genetic, common environmental and individual environmental effects, $X = A + C + E$. Maternal effects were not considered. Phenotypic (co)variances were the sum of additive genetic, common environmental and individual environmental (co)variances. Each generation (t), breeding values were predicted for every purebred individual of the current generation, using the following selection index,

$$I_{(t)} = \mathbf{b}_{(t)}'\mathbf{x}$$

where $\mathbf{b}_{(t)}$ is a vector of index weights and $\mathbf{x}$ is a vector of information sources.

Ten information sources were included in $\mathbf{x}$; 1. individual phenotypic record, 2. purebred full-sib phenotypic mean, 3. purebred half-sib phenotypic mean, 4. crossbred half-sib phenotypic mean, 5. estimated purebred breeding value sire, 6. estimated purebred breeding value dam, 7. average of estimated purebred breeding values of n_{pd} purebred dams mated to the sire, 8. estimated crossbred breeding value sire, 9. estimated crossbred breeding value dam, 10. average of estimated crossbred breeding values of n_{cd} F1-dams mated to the sire.

Purebred and crossbred estimated breeding values refer to estimated breeding values for purebred and crossbred performance. Estimated breeding values of sires and dams were included in $\mathbf{x}$ in order to utilise pedigree information. Estimated breeding values of sires

and dams were equal to their index in the previous generation, *i.e.*, when they were a candidate. Therefore, estimated breeding values of parents did not include progeny information, but this information is included in the index via full-sibs and half-sibs. This index is comparable to index ISD of WRAY and HILL (1989). It is assumed to accurately approximate selection on BLUP-EBV under an animal model, because pedigree information is accounted for by including sire and dam EBV, and phenotypic records of paternal half-sibs are corrected for the genetic level of their dams by including information sources 7 and 10.

Each generation, index weights were calculated according to

$$\mathbf{b}_{(t)} = \mathbf{P}_{(t)}^{-1}\mathbf{G}_{(t)}\mathbf{a},$$

where $\mathbf{P}_{(t)}$ is the 10×10 variance-covariance matrix of information sources in **x**, and $\mathbf{G}_{(t)}$ is a 10×2 matrix of covariances between information sources in **x** and breeding values in **u**.

For each generation, the matrices $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$ are given by,

$\mathbf{G}_{(t)} =$

$$\begin{bmatrix} G_0(1,1)_t & G_0(1,2)_t \\ G_1(1,1)_t & G_1(1,2)_t \\ G_2(1,1)_t & G_2(1,2)_t \\ G_3(1,2)_t & G_3(2,2)_t \\ S(1,1)_t/2 & S(1,2)_t/2 \\ D(1,1)_t/2 & D(1,2)_t/2 \\ D(1,1)_t/2n_{pd} & D(1,2)_t/2n_{pd} \\ S(1,2)_t/2 & S(2,2)_t/2 \\ D(1,2)_t/2 & D(2,2)_t/2 \\ 0 & 0 \end{bmatrix}$$

where "1" refers to the purebred trait and "2" refers to the crossbred trait. Notation follows as closely as possible that of VILLANUEVA *et al.* (1993). Generally speaking, three types of (co)variance elements can be distinguished in $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$; (co)variances between phenotypic information sources denoted $P.(i,j)_t$, (co)variances involving estimated breeding values of sires and dams denoted $S(i,j)_t$ and $D(i,j)_t$, and (co)variances between phenotypic information sources and breeding values of the individual denoted $G.(i,j)_t$, where i and j can take the value 1 or 2, *i.e.*, refer to the purebred or the crossbred trait. Because sires and dams

$\mathbf{P}_{(t)} =$

$$\begin{bmatrix} P_0(1,1)_t & P_1(1,1)_t & P_2(1,1)_t & G_3(1,2)_t & S(1,1)_t/2 & D(1,1)_t/2 & D(1,1)_t/2n_{pd} & S(1,2)_t/2 & D(1,2)_t/2 & 0 \\ & P_1(1,1)_t & P_2(1,1)_t & G_3(1,2)_t & S(1,1)_t/2 & D(1,1)_t/2 & D(1,1)_t/2n_{pd} & S(1,2)_t/2 & D(1,2)_t/2 & 0 \\ & & P_2(1,1)_t & G_3(1,2)_t & S(1,1)_t/2 & D(1,1)_t/2n_{pd} & D(1,1)_t/2n_{pd} & S(1,2)_t/2 & D(1,2)_t/2n_{pd} & 0 \\ & & & P_3(2,2)_t & S(1,2)_t/2 & 0 & 0 & S(2,2)_t/2 & 0 & D_{F1}(2,2)/2n_{cd} \\ & & & & S(1,1)_t & 0 & 0 & S(1,2)_t & 0 & 0 \\ & & & & & D(1,1)_t & D(1,1)_t/n_{pd} & 0 & D(1,2)_t & 0 \\ & & & & & & D(1,1)_t/n_{pd} & 0 & D(1,2)_t/n_{pd} & 0 \\ & & \text{symmetrical} & & & & & S(2,2)_t & 0 & 0 \\ & & & & & & & & D(2,2)_t & 0 \\ & & & & & & & & & D_{F1}(2,2)/n_{cd} \end{bmatrix}$$

were assumed to be unrelated and F1-dams were not related to purebred animals, $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$ contain zero's at the corresponding positions. Elements of $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$ that do not contain crossbred components are identical to the corresponding elements of $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$ derived by VILLANUEVA *et al.* (1993).

Estimated breeding values for trait j were obtained from the index,

$$I_{j(t)} = \mathbf{b}'_{j(t)}\mathbf{x}$$

where

$$\mathbf{b}_{j(t)} = \mathbf{P}^{-1}_{(t)}\mathbf{g}_{j(t)}$$

where $\mathbf{g}_{j(t)}$ is the column of $\mathbf{G}_{(t)}$ corresponding to the j^{th} trait.

Matrix elements: For the calculation of elements of $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$, total phenotypic (co)variances between trait i and j [$\mathrm{Cov}(X_i,X_j)_t$] were partitioned into additive genetic (co)variances [$\mathrm{Cov}(A_i,A_j)_t$], common environmental (co)variances [$\mathrm{Cov}(C_i,C_j)$] and individual environmental (co)variances [$\mathrm{Cov}(E_i,E_j)$] between trait i and j. Additive genetic (co)variances between trait i and j were further partitioned into between sire family (co)variances [$\mathrm{Cov}_s(A_i,A_j)_t$], between dam family (co)variances [$\mathrm{Cov}_d(A_i,A_j)_t$] and within family (co)variances [$\mathrm{Cov}_w(A_i,A_j)$]. Elements of $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$ were calculated as

$$G_0(i,j)_t = \mathrm{Cov}(A_i,A_j)_t$$

$$G_1(i,j)_t = \mathrm{Cov}_s(A_i,A_j)_t + \mathrm{Cov}_d(A_i,A_j)_t + \mathrm{Cov}_w(A_i,A_j)/n_{po}$$

$$G_2(i,j)_t = \mathrm{Cov}_s(A_i,A_j)_t + \mathrm{Cov}_d(A_i,A_j)_t/n_{pd} + \mathrm{Cov}_w(A_i,A_j)/n_{po}n_{pd}$$

$$G_3(i,j)_t = \mathrm{Cov}_s(A_i,A_j)_t$$

$$P_0(i,j)_t = \mathrm{Cov}(X_i,X_j)_t$$

$$P_1(i,j)_t = G_1(i,j)_t + \mathrm{Cov}(C_i,C_j) + \mathrm{Cov}(E_i,E_j)/n_{po}$$

$$P_2(i,j)_t = G_2(i,j)_t + \mathrm{Cov}(C_i,C_j)/n_{pd} + \mathrm{Cov}(E_i,E_j)/n_{po}n_{pd}$$

$$P_3(2,2)_t = \mathrm{Cov}_s(A_2,A_2)_t + [\tfrac{1}{2}\mathrm{Cov}_{F1.B}(A_2,A_2)_{eq} + \mathrm{Cov}(C_2,C_2)]/n_{cd} + [(\tfrac{1}{4}\mathrm{Cov}(A_2,A_2)_{t=0} + \tfrac{1}{4}\mathrm{Cov}_{F1}(A_2,A_2)_{t=0} + \mathrm{Cov}(E_2,E_2)]/(n_{cd}n_{co})$$

$$S(i,j)_t = \mathrm{Cov}(A_i,I_j)_{t-1} - \frac{\mathrm{Cov}(A_i,I)_{t-1}\,\mathrm{Cov}(I_j,I)_{t-1}}{\sigma^2_{I(t-1)}}k_s$$

$$D_{F1}(2,2) = r^2_{IHeq}\,\mathrm{Cov}_{F1}(A_2,A_2)_{eq.},$$

where $\mathrm{Cov}_{F1.B}(A_2,A_2)_{eq}$ is the between family component of additive genetic variance for crossbred performance among F1-dams at equilibrium, $\mathrm{Cov}_{F1}(A_2,A_2)_{t=0}$ is the total additive genetic variance for crossbred performance among F1-dams in the base generation, $\mathrm{Cov}_{F1}(A_2,A_2)_{eq}$ is the total additive genetic variance for crossbred performance among F1-dams at equilibrium and r^2_{IHeq} is the accuracy of breeding value estimation of F1-dams at equilibrium. The element $D_{F1}(2,2)$ represents the variance in indexes of F1-dams, and was calculated as the squared accuracy times the additive genetic variance in equilibrium. It was assumed that F1-dams had equilibrium genetic parameters. Therefore, all covariance components concerning F1-dams were constant. If there is no breeding value estimation on F1-

dams, the last row and column of $\mathbf{P}_{(t)}$ are removed, and $\mathbf{P}_{(t)}$ becomes a 9×9 matrix and $\mathbf{G}_{(t)}$ becomes a 9×2 matrix.

Each generation t, the total phenotypic (co)variance was calculated as the sum of additive genetic, common environmental and individual environmental (co)variance. The total additive genetic (co)variance was calculated as the sum of sire, dam and within family component of additive genetic (co)variance. The common environmental and individual environmental (co)variances were assumed to be constant over time. $\mathrm{Cov}(C_1,C_1)$ and $\mathrm{Cov}(C_2,C_2)$ were calculated as $c_1^2\sigma^2_{p1(t=0)}$ and $c_2^2\sigma^2_{p2(t=0)}$ respectively, where c_1^2 and c_2^2 are the common environmental variance between full-sibs as a proportion of the total phenotypic variance for purebred and crossbred performance in the base generation, and $\sigma^2_{p(t=0)}$ is the total phenotypic variance in the base generation. Because inbreeding was ignored, the within family component of additive genetic variance $[\mathrm{Cov}_w(A_i,A_j)]$ was equal to $\frac{1}{2}\mathrm{Cov}(A_i,A_j)_{t=0}$, and was constant over time. The between sire family component of additive genetic variance was calculated every generation as,

$$\mathrm{Cov}_s(A_i,A_j)_t = \frac{1}{4}\left[\mathrm{Cov}(A_i,A_j)_{t-1} - \frac{\mathrm{Cov}(A_i,I)_{t-1}\,\mathrm{Cov}(A_j,I)_{t-1}}{\sigma^2_{I(t-1)}}k_s\right]$$

where

$$\mathrm{Cov}(A_i,I)_t = \mathbf{b}'_{(t)}\mathbf{g}_{i(t)}$$

$$\sigma^2_{I(t)} = \mathbf{b}'_{(t)}\mathbf{P}_{(t)}\mathbf{b}_{(t)}$$

where $\sigma^2_{I(t)}$ is the variance of the index, $k_s = i_s(i_s-x_s)$, i_s is the selection intensity and x_s is the corresponding standardised truncation point for sires. $\mathrm{Cov}_d(A_i,A_j)_t$ was calculated in the same way as $\mathrm{Cov}_s(A_i,A_j)_t$, but k_d was used instead of k_s. $\mathrm{Cov}(A_i,I_j)_t$ and $\mathrm{Cov}(I_j,I)_t$ were calculated as,

$$\mathrm{Cov}(A_i,I_j)_t = \mathbf{b}'_{j(t)}\mathbf{g}_{i(t)}$$

$$\mathrm{Cov}(I_j,I)_t = \mathbf{b}'_{j(t)}\mathbf{G}_{(t)}\mathbf{a}$$

Elements $D(i,j)_t$ were calculated similar to $S(i,j)_t$, but k_d was used instead of k_s.

The sire line heritability for crossbred performance was defined as,

$$h_c^2 = \mathrm{Cov}(A_2,A_2)_{t=0} / \sigma^2_{p2(t=0)}$$

where $\sigma^2_{p2(t=0)}$ is the phenotypic variance among crossbreds in the base population. The genetic covariance between purebred and crossbred performance, $\mathrm{Cov}(A_1,A_2)_{t=0}$, was defined as four times the covariance between purebred and crossbred half-sibs. The purebred-crossbred genetic correlation was defined as

$$r_{pc} = \frac{\mathrm{Cov}(A_1,A_2)_{t=0}}{\sqrt{\mathrm{Cov}(A_1,A_1)_{t=0}\,\mathrm{Cov}(A_2,A_2)_{t=0}}}$$

Selection intensity: Approximate selection intensities corrected for correlated index values and for finite sample size were calculated using the method of MEUWISSEN (1991),

$$i_r(t_{fs},t_{hs}) = i_r(0,0)\{1-t_{av}(t_{fs},t_{hs})\}^{v(t_{fs},t_{hs})}$$

where $i_r(0,0)$ is the selection intensity for uncorrelated index values corrected for finite sample size, t_{av} is the correlation between index values averaged over all selection candidates, t_{fs} is the correlation between index values of full-

sibs, t_{hs} is the correlation between index values of half-sibs and $v(t_{fs},t_{hs})$ is a weighting factor. This method is a 3-dimensional application of RAWLINGS' (1976) method, and performs well for nested full-half-sib family structures (MEUWISSEN, 1991; PHOCAS and COLLEAU, 1995). The approximation of BURROWS (1972) was used to calculate selection intensities for uncorrelated index values corrected for finite sample size [$i_r(0,0)$]. The correlation between index values averaged over all selection candidates was (MEUWISSEN, 1991),

$$t_{av} = \frac{(n_{po}-1)t_{fs} + n_{po}(n_{pd}-1)t_{hs}}{n_s n_{pd} n_{po} - 1}$$

Correlations between selection index values of relatives, *i.e.*, between full-sibs (t_{fs}) and between half-sibs (t_{hs}), were calculated as (DE BOER and VAN ARENDONK, 1991):

$$t_{rel} = \mathbf{b}'_{(t)}\mathbf{R}_{(t)}\mathbf{b}_{(t)} / \mathbf{b}'_{(t)}\mathbf{P}_{(t)}\mathbf{b}_{(t)}$$

where $\mathbf{R}_{(t)}$ is a 10×10 variance-covariance matrix between information sources of two relatives in generation t.

Selection response: Because the breeding goal only consisted of crossbred performance, response in crossbred performance was identical to response in aggregate breeding goal. The expected response was calculated as:

$$R_{(t)} = i_{(t-1)}\sigma_{I(t-1)}$$

where $i_{(t)}$ is $(i_{s(t)} + i_{d(t)})/2$ and $i_{s(t)}$ and $i_{d(t)}$ are the selection intensities for sires and dams in generation t, corrected for finite sample size and correlated index values. Because the same index was used for males and females, $\sigma_{I(t)}$ was the same for both sexes. Throughout the results and discussion section, response refers to genetic response for crossbred performance.

Base generation animals had no family information and were selected on their own phenotype only. Subsequent generations were selected on the above described index. Selection was undertaken until constant selection responses per generation were obtained (equilibrium). DEKKERS (1992) distinguished two equilibria. The first equilibrium is due to build up of pedigree information. The second equilibrium arises from reduction of additive genetic variance due to selection (BULMER, 1971). To distinguish between reductions of response due to the Bulmer effect and due to effects of correlated index values, responses were calculated for three equilibria. First, pedigree equilibrium response (R_{ped}) ignoring the Bulmer effect and correlated index values was calculated. Second, correlated index values were also taken into account (R_{int}), and finally, equilibrium response taking both effects into account, *i.e.*, the asymptotic response (R_{as}) , was calculated.

Optimisation of number of selected parents: The selection index was used to optimise selection in the sire line, given a fixed test capacity. For a fixed test capacity, the number of selected boars and sows and the number of offspring tested per litter was optimised, keeping the amount of crossbred information per sire constant. Test capacity refers to the total number of purebred animals (boars + gilts) from the sire line that were recorded for their purebred performance per generation. Test capacities were N= 96, 192, 384, 768, 1536 and 3072 animals per

TABLE 1.–Optimum number of selected sires ($n_{s,opt}$) and corresponding asymptotic response in crossbred performance (R_{as}) for a varying number of offspring tested per litter (n_{po}) and different heritabilities (h^2) and test capacities (N)†

	$h^2 = 0.15$			$h^2 = 0.25$			$h^2 = 0.40$		
N	n_{po}	$n_{s,opt}$	R_{as}	n_{po}	$n_{s,opt}$	R_{as}	n_{po}	$n_{s,opt}$	R_{as}
3072	2	6	0.224	2	4	0.324	2	4	0.459
3072	4	8	0.250	4	6	0.361	4	6	0.513
3072	6	12	0.267	6	8	0.385	6	6	0.546
3072	**8**	**12**	**0.278**	**8**	**8**	**0.400**	**8**	**6**	**0.566**
1536	2	6	0.209	2	4	0.302	2	4	0.428
1536	4	8	0.237	4	6	0.342	4	6	0.485
1536	6	8	0.253	6	8	0.365	6	6	0.518
1536	**8**	**12**	**0.264**	**8**	**8**	**0.381**	**8**	**6**	**0.538**
768	2	4	0.194	2	4	0.280	2	4	0.396
768	4	6	0.222	4	6	0.321	4	4	0.457
768	6	8	0.239	6	6	0.345	6	6	0.489
768	**8**	**8**	**0.250**	**8**	**8**	**0.360**	**8**	**6**	**0.509**
384	2	4	0.176	2	4	0.255	2	2	0.364
384	4	6	0.207	4	4	0.299	4	4	0.427
384	6	6	0.224	6	6	0.323	6	4	0.458
384	**8**	**8**	**0.234**	**8**	**8**	**0.337**	**8**	**4**	**0.480**
192	2	4	0.158	2	2	0.229	2	2	0.331
192	4	4	0.190	4	4	0.276	4	4	0.393
192	6	6	0.207	6	6	0.299	6	4	0.425
192	**8**	**8**	**0.217**	**8**	**6**	**0.314**	**8**	**4**	**0.446**
96	2	2	0.138	2	2	0.204	2	2	0.294
96	4	4	0.173	4	4	0.251	4	4	0.355
96	6	4	0.188	6	4	0.273	6	4	0.387
96	**8**	**6**	**0.198**	**8**	**4**	**0.288**	**8**	**4**	**0.409**

† For $r_{pc} = 0.7$ and 64 crossbreds tested per sire. Optimum parameters and maximum responses are printed bold.

generation. The number of selected sires varied from n_s = 2, 4, 6, 8, 12, 24. From all litters, an equal number of animals was tested. The number of tested animals per litter was, n_{po} = 2, 4, 6 or 8. Testing eight animals per litter was assumed to be the biological maximum. In the basic scheme, equal numbers of boars and gilts were tested. The number of selected dams was calculated as N/n_{po}. The amount of crossbred information was kept constant at 64 crossbred offspring tested per sire, *i.e.*, the total number of tested crossbreds varied with the number of sires. Purebred and crossbred heritabilities were assumed to be equal. Heritability alternatives were h^2 = 0.15, 0.25 and 0.40. Phenotypic variance was 1, r_{pc} was 0.7, and ${c_1}^2$ and ${c_2}^2$ were 0.1. Sensitivity of the optimum number of selected parents to the value of r_{pc} and to the amount of crossbred information included was investigated in the section on benefit of including crossbred information.

Benefit of including crossbred information: The selection index was also used to make inferences on the potential benefit of using crossbred information. Genetic responses in crossbred performance under PLS (*i.e.*, no crossbred information) were compared to maximum responses in crossbred performance

that were obtained when a large amount of crossbred information was included. The effect of crossbred information on response depends on assumptions about breeding values of F1-dams, crossbred litter size and on the existence of common environmental effects among crossbred full-sibs. Inferences about the effect of including crossbred information were made assuming unknown breeding values of F1-dams, a crossbred litter size of eight and absence of common environmental effects among crossbreds.

Results

Optimisation of the number of selected parents

Optimum number of offspring tested: Table 1 shows the optimum number of selected sires and the corresponding asymptotic selection response in crossbred performance, for a varying number of offspring tested per litter and for different test capacities. Given the fixed test capacity, fewer litters had to be produced and fewer dams were selected when the number of tested animals per litter increased. For example, for $N = 768$ and $n_{po} = 8$, 768/8 = 96 dams had to be selected from 768/2 selection candidates, and the selected proportion was 0.25. When $n_{po} = 4$, 768/4 = 192 dams had to be selected from 768/2 selection candidates and the selected proportion was 0.5. For this reason selection intensity in dams increased when more animals per litter were tested and highest responses were always obtained when the maximum number of eight animals per litter was tested. Selection response increased with increasing test capacity. However, the increase in response was rather low compared to the increase in test capacity. This was because the selected proportion among females was independent of N. When eight animals per litter were tested, one out of four females was needed for replacement, irrespective of N. Therefore, increase in response with increasing test capacity was mainly due to higher selection intensities in sires.

When the number of offspring tested per litter was below the optimum, *i.e.*, n_{po} was lower than eight, the optimum numbers of selected sires also decreased.

Response in different equilibria: Table 2 shows selection responses for different equilibria, for a varying number of selected sires and eight animals tested per litter. As expected, first generation (not shown) and pedigree equilibrium responses (R_{ped}) were always highest when only two sires were selected due to highest selection intensities. Increase of response due to pedigree information (not shown), *i.e.*, differences between first generation and pedigree equilibrium response, were largest when heritability was low. For this reason, pedigree equilibrium responses increased less than linearly with heritability, whereas first generation responses increased linearly with heritability.

Contrary to R_{ped}, responses corrected for correlated index values (R_{int}) were highest when more than two sires were selected. This means that accounting for reduction of selection intensity due to correlated index values altered the ranking of breeding schemes. The effect of

TABLE 2.–Pedigree equilibrium response (R_{ped}), corrected intensity response (R_{int}) and asymptotic response (R_{as}) in crossbred performance for different numbers of sires selected (n_s) and for different test capacities (N) and heritabilities (h^2) †

		$h^2 = 0.15$			$h^2 = 0.25$			$h^2 = 0.40$		
N	n_s	R_{ped}	R_{int}	R_{as}	R_{ped}	R_{int}	R_{as}	R_{ped}	R_{int}	R_{as}
3072	2	**0.472**	0.339	0.229	**0.658**	0.493	0.351	**0.902**	0.708	0.526
	4	0.451	0.378	0.266	0.629	0.541	0.391	0.863	0.760	0.562
	6	0.437	0.384	0.273	0.611	0.547	0.397	0.838	**0.764**	**0.566**
	8	0.427	**0.387**	0.277	0.598	**0.548**	**0.400**	0.820	0.764	0.565
	12	0.413	0.385	**0.278**	0.579	0.544	0.399	0.794	0.755	0.559
	16	0.402	0.381	0.277	0.565	0.539	0.396	0.775	0.746	0.552
	24	0.387	0.373	0.274	0.544	0.528	0.389	0.747	0.729	0.540
768	2	**0.427**	0.308	0.211	**0.598**	0.448	0.322	**0.820**	0.643	0.480
	4	0.402	0.341	0.243	0.565	0.488	0.356	0.775	**0.685**	**0.509**
	6	0.387	0.344	0.248	0.544	**0.490**	0.359	0.747	0.684	0.509
	8	0.375	**0.344**	**0.250**	0.529	0.489	**0.360**	0.726	0.681	0.506
	12	0.359	0.339	0.248	0.506	0.481	0.355	0.696	0.666	0.496
	16	0.346	0.333	0.246	0.489	0.472	0.350	0.673	0.653	0.486
	24	0.328	0.320	0.238	0.465	0.454	0.338	0.640	0.627	0.468
96	2	**0.346**	0.246	0.176	**0.489**	0.360	0.266	**0.673**	0.517	0.394
	4	0.315	**0.269**	**0.198**	0.446	**0.387**	**0.288**	0.614	**0.541**	**0.409**
	6	0.295	0.265	0.198	0.418	0.379	0.285	0.576	0.527	0.399

† For $n_{po} = 8$, $r_{pc} = 0.7$ and 64 crossbreds tested per sire. Maximum responses are printed bold.

correlated index values on selection response increased with decreasing heritability. When heritability was low, information of relatives contributed more to EBV of selection candidates and correlations between index values increased. Therefore, relative reductions in response due to correlated index values were highest at low heritabilities. Maximum relative reduction of response due to correlated index values was 29% for the scenario with $h^2 = 0.15$, $N = 96$ and $n_s = 2$. Differences between R_{int} and asymptotic responses (R_{as}) represent reductions due to the Bulmer effect. Relative reductions due to the Bulmer effect increased slightly with decreasing heritability. This result seems to be in contradiction with results of DEKKERS (1992), who showed that relative reductions due to the Bulmer effect are only dependent on the selection intensity and not on the heritability. However, in the present study, reduction of between family variance due to the Bulmer effect changed the correlation between index values, and therefore affected the selection intensity. When correlated index values were ignored, the heritability did not affect relative reductions of response due to the Bulmer effect. Compared to R_{int}, the Bulmer effect had little effect on the ranking of breeding schemes in general. Maximum reduction of response due to the Bulmer effect was 32%, for the scenario with $h^2 = 0.15$, $N = 3072$ and $n_s = 2$. Maximum overall reduction in response compared to pedigree equilibrium response was 51% for the scenario of $h^2 = 0.15$, $N = 3072$ and $n_s = 2$.

TABLE 3.–Optimum number of selected sires ($n_{s,opt}$) and corresponding asymptotic response in crossbred performance (R_{as}) for a varying number of males tested per litter (n_{ml}) and for different test capacities (N) and heritabilities (h^2) †

		$h^2 = 0.15$		$h^2 = 0.25$		$h^2 = 0.40$	
N	n_{ml}	$n_{s,opt}$	R_{as}	$n_{s,opt}$	R_{as}	$n_{s,opt}$	R_{as}
3072	2	12	0.272	8	0.392	6	0.554
	3	12	0.267	8	0.385	6	0.546
	4	8	0.255	8	0.370	6	0.525
1536	2	8	0.258	8	0.372	4	0.526
	3	8	0.253	8	0.365	6	0.518
	4	8	0.242	6	0.350	6	0.497
768	2	8	0.243	6	0.351	4	0.497
	3	8	0.239	6	0.345	4	0.489
	4	8	0.228	6	0.330	4	0.469
384	2	6	0.227	6	0.328	4	0.466
	3	6	0.224	6	0.323	4	0.458
	4	6	0.213	6	0.308	4	0.439
192	2	4	0.209	4	0.304	4	0.431
	3	6	0.207	4	0.299	4	0.425
	4	6	0.196	4	0.285	4	0.406
96	2	4	0.191	4	0.277	4	0.391
	3	4	0.188	4	0.273	4	0.387
	4	4	0.179	4	0.261	4	0.370

† For $r_{pc} = 0.7$, $n_{po} = 6$ and 64 crossbreds tested per sire.

Because reductions in response increased with decreasing heritability, the optimum number of selected sires increased when heritability decreased (See also Table 1).

When the test capacity decreased, the optimum number of selected sires decreased only slightly. Compared to maximum pedigree equilibrium response, maximum asymptotic response increased slightly more with increasing test capacity. For example, for a heritability of 0.15, the ratio between maximum pedigree equilibrium response for $N = 3072$ and $N = 96$ equalled, 0.472/0.346 = 1.364, whereas for maximum asymptotic response, the ratio equalled, 0.278/0.198 = 1.404. This means that the benefit of having a large test capacity is larger than judged by maximum pedigree equilibrium response.

Testing different numbers of males and females: To investigate whether it is beneficial to test different numbers of males and females, the number of males and females tested per litter was varied. Testing more offspring from a certain sex per litter results in more selection candidates of this sex. Therefore, selection intensity increases for this sex. Testing on average four animals of the same sex per litter was assumed to be the biological maximum. For this reason, a total of six animals per litter was tested, the number of males tested per litter was varied from 2 to 4, and as a result the number of females tested varied from 4 to 2. The number of selected sires per generation was varied, n_s = 4, 6, 8 or 12, except for $N = 96$, where n_s was 2, 4, 6, or 8.

Table 3 shows the optimum number of selected sires and the corresponding asymptotic responses for a varying number of males tested per litter and for different test capacities and heritabilities. Highest responses were obtained when only two males per litter were tested, and therefore four females per litter were tested. This was because female selection intensity increased relatively more than male selection intensity decreased when going from 2 to 4 females tested and from 4 to 2 males tested. Differences between responses for two and three males tested were small. Differences between responses for three and four males tested were larger because reductions in selection intensity due to correlated index values were more severe for small proportions selected. Responses with six offspring tested were always lower than responses with eight offspring tested per litter with equal proportions tested for males and females (see Table 1). This indicates that it is optimal to test the maximum number of offspring per litter, and especially the maximum number of female offspring per litter.

Benefit of crossbred information

When the amount of crossbred information per sire increased, response increased until there was no additional benefit of extra crossbred information. The response predicted for this situation is referred to as maximum response. Figure 1 shows the ratio of asymptotic maximum response over the asymptotic response obtained without crossbred information, *i.e.*, under pure line selection (PLS), as a function of the purebred-crossbred genetic correlation and for a test capacity of N = 768 purebred offspring. The response in both cases is calculated for the optimum number of selected sires under PLS and under combined crossbred purebred selection (CCPS) respectively, and the ratio expresses the maximum superiority of CCPS over PLS. The optimum number of selected sires differed between PLS and CCPS (See Table 4). As expected, the benefit of crossbred information increased rapidly when the purebred-crossbred genetic correlation decreased, and approached infinity when r_{pc} approached zero. Apart from the fact that crossbred information is measured directly on the trait in the breeding goal, it can also be regarded as an increase of the amount of sib information. For this reason, there was still a small benefit of using crossbred information

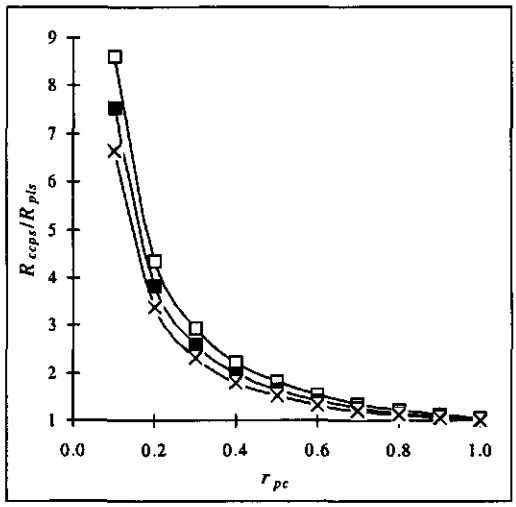

FIGURE 1.–Ratio of asymptotic maximum response (R_{ccps}) over asymptotic response under pure line selection (R_{pls}) for different purebred-crossbred genetic correlations (r_{pc}) and heritabilities (h^2), and for a test capacity (N) of 768 pure-bred offspring. —□— h^2 = 0.15; —■— h^2 = 0.25; —×— h^2 = 0.40.

when r_{pc} = 1, and the benefit of crossbred information was highest for low heritabilities, because low heritable traits benefit more from sib information. For example, for h^2 = 0.15 and r_{pc} = 1, the ratio R_{ccps}/R_{pls} equalled 1.05.

At low values of r_{pc}, the question arises if there is any benefit of including purebred information when the breeding goal is crossbred performance. However, when only crossbred information is used, there is no possibility to distinguish between purebred full-sib, because they all have exactly the same information, *i.e.*, the crossbred breeding value of sire and dam, and the crossbred half-sib performance. For this reason, accuracy decreases and also selection intensity decreases due to higher correlations between index values when only crossbred information is used. For example, for r_{pc} = 0.4, h^2 = 0.25, N = 768, n_s = 12 and information on 25 crossbred litters per sire, response with and without purebred information was 0.334 (Table 4) and 0.308 (not shown) respectively.

Figure 2 shows the effect of the amount of crossbred information on the asymptotic response for a heritability of 0.25 and a crossbred litter size of eight. In this Figure, n_c is the number of crossbred offspring per sire included in the selection index. It was assumed that crossbred litter size was eight, breeding values of F1-dams were unknown and there was no common environment between crossbred half-sibs. When n_c equalled zero, the PLS response was obtained, when n_c was large, response approached the maximum crossbred response. Under PLS, (n_c = 0), the ratio of responses for different values of r_{pc} equalled the ratio of r_{pc}-values, because the purebred

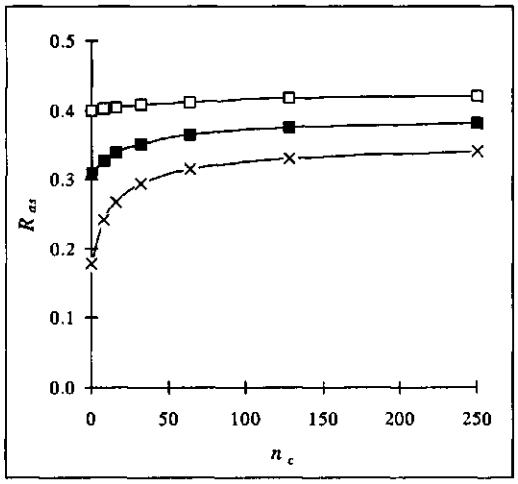

FIGURE 2.–Asymptotic response (R_{as}) achieved with optimum numbers of sires for different numbers of crossbred offspring included in the index (n_c) and different purebred-crossbred genetic correlations (r_{pc}). For N = 768 and h^2 = 0.25. —□— r_{pc} = 0.9; —■— r_{pc} = 0.7; —×— r_{pc} = 0.4.

response was not affected by r_{pc}. For low values of r_{pc}, the response increased rapidly when the number of crossbred offspring increased. For r_{pc} = 0.4, 22 crossbred litters were needed to obtain 90% of the maximum extra response attainable by including crossbred information. For higher values of r_{pc}, larger amounts of crossbred information were needed to obtain a certain percentage of the potential extra response. For example, for r_{pc} = 0.7, more than 64 crossbred litters were needed to obtain 90% of the potential extra response.

Large amounts of crossbred information could not completely compensate for low values of r_{pc}, *i.e.*, the PLS response with r_{pc} = 0.9 could not be reached with r_{pc} = 0.7 and a large amount of crossbred information. This is because crossbred half-sibs do not provide information on within family deviations, *i.e.*, information of

TABLE 4.–Pedigree equilibrium response (R_{ped}), corrected intensity response (R_{int}) and asymptotic response (R_{as}) in crossbred performance, when information on 25 crossbred litters of litter size eight is included, and asymptotic response under PLS (R_{pls}), for a varying number of selected sires (n_s) and a range of purebred-crossbred correlations (r_{pc}) and heritabilities (h^2) †

		$h^2 = 0.15$				$h^2 = 0.25$				$h^2 = 0.40$			
r_{pc}	n_s	R_{ped}	R_{int}	R_{as}	R_{pls}	R_{ped}	R_{int}	R_{as}	R_{pls}	R_{ped}	R_{int}	R_{as}	R_{pls}
0.4	2	0.433	0.285	0.170	0.111	0.585	0.394	0.245	0.170	0.769	0.534	0.348	0.253
	4	0.409	0.331	0.220	**0.121**	0.552	0.453	0.306	**0.178**	0.726	0.605	0.419	**0.256**
	6	0.394	0.338	0.231	0.120	0.532	0.461	0.320	0.176	0.700	0.613	0.433	0.252
	8	0.383	0.342	0.239	0.119	0.518	0.465	0.328	0.174	0.681	0.617	0.441	0.248
	12	0.367	0.341	0.244	0.115	0.496	0.463	0.333	0.168	0.652	0.611	**0.444**	0.239
	16	0.355	0.338	**0.245**	0.112	0.480	0.458	**0.334**	0.163	0.631	0.603	0.443	0.233
	24	0.338	0.328	0.242	0.105	0.456	0.443	0.327	0.155	0.599	0.583	0.432	0.222
0.7	2	0.460	0.326	0.219	0.195	0.631	0.466	0.330	0.297	0.853	0.660	0.488	0.443
	4	0.434	0.365	0.257	**0.211**	0.596	0.512	0.370	**0.310**	0.806	0.708	0.524	**0.448**
	6	0.418	0.369	0.264	0.210	0.575	0.515	0.375	0.307	0.776	0.708	**0.525**	0.440
	8	0.407	0.371	0.267	0.208	0.559	0.515	**0.377**	0.303	0.755	0.705	0.522	0.433
	12	0.389	0.366	**0.267**	0.201	0.536	0.507	0.374	0.293	0.723	0.691	0.513	0.418
	16	0.377	0.361	0.266	0.194	0.518	0.499	0.369	0.285	0.700	0.677	0.504	0.406
	24	0.358	0.349	0.259	0.184	0.492	0.481	0.357	0.271	0.665	0.651	0.485	0.387
0.9	2	0.489	0.362	0.257	0.250	0.678	0.526	0.391	0.381	0.931	0.759	0.582	0.569
	4	0.461	0.396	0.287	**0.271**	0.640	0.564	0.418	**0.399**	0.879	0.794	**0.598**	**0.575**
	6	0.443	0.398	0.291	0.270	0.617	0.563	**0.418**	0.394	0.848	0.788	0.592	0.565
	8	0.431	0.398	**0.291**	0.266	0.599	0.560	0.416	0.389	0.824	0.781	0.584	0.555
	12	0.412	0.391	0.288	0.257	0.574	0.549	0.408	0.376	0.790	0.761	0.569	0.536
	16	0.399	0.384	0.284	0.249	0.556	0.538	0.400	0.365	0.764	0.744	0.555	0.521
	24	0.378	0.370	0.275	0.235	0.528	0.517	0.385	0.347	0.726	0.713	0.533	0.496

† For a test capacity of 768 pure-bred animals ($N = 768$) and a pure-bred litter size of eight ($n_{po} = 8$). Maximum responses are printed bold.

crossbred half-sibs is common to all purebred offspring of a sire. The only within family information source is the individual (purebred) record, and its contribution increases with increasing r_{pc}.

As expected, knowledge of breeding values of F1-dams increased response, whereas presence of common environmental effects decreased response. For a fixed total number of crossbred half-sibs included in the index, it was most efficient to measure only one half-sib per litter, *i.e.*, having a maximum number of F1-dams. However, when breeding values of F1-dams were known without error and common environment was absent, responses were the same for testing one or eight crossbred half-sib per litter given a fixed total number of crossbred half-sibs. In practise, all offspring of a litter will probably be tested because this is likely to be cheaper than testing only one offspring per litter and a large number of litters.

Effect of crossbred information on optimum number of selected parents: Table 4 shows responses at different equilibria and optimum numbers of selected sires when including information on 25 crossbred litters of eight offspring each. For comparison, the asymptotic response obtained under PLS,

denoted as R_{pls}, is also included in Table 4. By comparing R_{as} and R_{pls} for different values of r_{pc}, it can be seen that the benefit of crossbred information increased when r_{pc} decreased, which is in agreement with Figure 1. Inclusion of crossbred information increased the correlation between index values of relatives, because the crossbred information is common to all half-sibs. For example, for $r_{pc} = 0.4$, $h^2 = 0.25$ and 16 sires selected, the average correlation between index values equalled 0.0294 under PLS, whereas it equalled 0.0522 when crossbred information was included. This resulted in a considerable decrease in the selection intensity, which can be seen from the large differences between R_{ped} and R_{int}, especially for low values of n_s. For this reason, the optimum number of selected sires increased considerably when crossbred information was incorporated. Under PLS, the optimum number of selected sires always equalled four, whereas it ranged between 4 and 16 when crossbred information was included. The optimum number of selected sires was largest for $r_{pc} = 0.4$ and $h^2 = 0.15$, because correlations between index values were largest for this situation.

The results in Table 4 are for a situation where information on a fixed number of crossbred litters per sire was included in the index. Therefore, the total amount of crossbred information varied with the number of sires. To make a comparison at the same level of costs, it is interesting to fix the total amount of crossbred information, instead of fixing it per sire. Results of a comparison are listed in Table 5, for a total of 96 or 384 crossbred litters tested. Results are for $h^2 = 0.15$ and $r_{pc} = 0.4$,

TABLE 5.–Asymptotic responses for a fixed total number of 96 (R_{96}) and 384 (R_{384}) crossbred litters tested, for different numbers of selected sires (n_s), and for a heritability of 0.15 and a purebred-crossbred genetic correlation of 0.4†

n_s	R_{96}	R_{384}
2	0.177	0.182
4	0.220	0.233
6	**0.224**	0.243
8	**0.224**	**0.248**
12	0.218	**0.248**
16	0.210	0.245
24	0.193	0.232

†For a text capacity of 768 purebred animals ($N = 768$). Purebred and crossbred litter size equalled eight ($n_{po} = 8$, $n_{co} = 8$). Maximum responses are printed bold.

and can be compared to results for these parameters in Table 4. Compared to Table 4 [$\max(R_{as}) = 0.245$ giving $n_s = 12$], the optimum number of sired with CCPS decreased when the total amount of crossbred information was fixed. This is to be expected, because when the total amount of crossbred information is fixed, the amount per sire will decrease when the number of sires increases, which reduces the accuracy of selection. Especially for low total amounts of crossbred information, this effect is important, because the benefit of additional crossbred information is limited when the amount of information is already high. This is illustrated by the comparison at a low (R_{96}) and a high (R_{384}) total amount of crossbred information. The optimum number of sires was highest for 384 crossbred litters tested (8 to 12 *vs.* 6 to 8), and closer to results for 'fixed per sire' schemes (Table 4, $n_s = 16$). Differences of optimum numbers of sires between 'fixed total' and 'fixed per sire' schemes will decrease when r_{pc} and h_2 increase, because for these situations

the optimum numbers of sires are already closer to PLS optima and because benefit of large sib groups is lower for high heritabilities.

Discussion

In the present study, a selection index was developed which approximates selection on BLUP-EBV for a crossbreeding situation. This index was used to optimise the number of selected parents and the number of purebred offspring tested per litter, and to make inferences on the potential benefit of including crossbred information in the selection index. WRAY and HILL (1989) showed that an index including phenotypic records on an individual and its sibs, and estimated breeding values of parents and of mates of parents can be used to approximate selection on BLUP-EBV under an animal model when generations are discrete. Compared to the index of WRAY and HILL (1989), the index in this paper additionally includes phenotypic information on crossbred half sibs and estimated crossbred breeding values of parents and of mates of parents. WRAY and HILL (1989) found good agreement between response obtained from stochastic simulation and prediction from their selection index. For this reason, it is expected that the index used in the present study will also provide a good approximation of response of BLUP selection, and that there is little need to check the results by means of stochastic simulation. As recommended by WRAY and HILL (1989), ranking of breeding schemes was based on asymptotic response. Because asymptotic responses were reached in five generations and more than 95% of the reductions in response due to the Bulmer effect and due to correlated index values occurred in the first two generations, asymptotic response is a suitable measure for ranking breeding schemes in the present study. An assumption in the prediction method used here, is that fixed effects are known without error. This is not true in practise as fixed effects are estimated from the data simultaneously with the breeding values. For this reason, the test capacity has to be regarded as effective test capacity, *i.e.*, it is corrected for the distribution of observations over fixed effect classes, and it is therefore smaller than the actual number of test places.

In crossbreeding, the final aim is to produce high-quality crossbred fattening pigs. For this reason, the breeding goal was defined at the crossbred level. In a boar line, the number of animals is small compared to the number of crossbred fattening pigs, and purebred performance is therefore of minor importance. However, breeding organisations might want to put some weight on purebred performance. In the presented method, this can easily be accommodated by changing the economic values for purebred and crossbred performance.

In the present study, breeding schemes were optimised assuming a limited test capacity within the pure line, whereas the potential amount of crossbred information was assumed to be unlimited. However, at present, crossbred information is not routinely available to many breeding organisations. In order to use crossbred information, registration at multiplier herds, fattening herds and slaughterhouses has to be integrated and data has to be collected. Given the availability of modern information

technology, this is possible, but it will take considerable effort. For this reason, it is likely that breeding organisations will contract a limited number of test farms and slaughter houses and crossbred information will be collected only there (E. Knol, personal communication).

The present study showed that, given a fixed test capacity, the maximum number of female offspring per litter should be tested. This means that the number of selected sows, *i.e.*, the sow capacity, has to be adjusted to the test capacity. When instead of the test capacity, the sow capacity is limited, *e.g.*, because testing is cheap and keeping sows is expensive, testing the maximum number of offspring per litter will obviously result in the highest response and the test capacity is a result of the sow capacity. Therefore, optimum schemes for fixed test capacity and fixed sow capacity are identical, *i.e.*, the highest response for both situations is obtained when the maximum number of offspring per litter is tested. When only two offspring per litter were tested, the selected proportion in females was 100% and selection intensity equalled zero. However, instead of selecting only tested offspring, the best 50% of the tested offspring could be selected supplemented with untested offspring, which would raise the selection intensity to 0.398 in this case (KIRSCH *et al.* 1962). This scheme is however still inferior to testing eight offspring per litter, and in addition it might be undesirable for breeding organisations to supplement the selected group with untested animals for veterinary or management reasons.

Responses obtained when using both purebred and crossbred information were higher than responses under PLS. Especially for low values of the purebred-crossbred genetic correlation (r_{pc}) the benefit of including crossbred information was very clear (see Figure 1). For $r_{pc} = 0.4$, the maximum response under CCPS was approximately twice the response under PLS. When r_{pc} was high, *i.e.*, > 0.9, the potential benefit was small (Figure 1) and large amounts of crossbred information were needed to obtain extra response (Figure 2). WEI and VAN DER WERF (1994) also compared selection response under PLS to response under CCPS using selection index theory, but ignored the Bulmer effect and pedigree information. They concluded that CCPS was better under all circumstances, even when large errors in estimates of r_{pc} and h^2 were present. This conclusion is not expected to be affected by the correction for correlated index values and for the Bulmer effect and can therefore be extended to the present study.

DE ROO (1987 and 1988) studied breeding schemes for the sire line of a closed pig population using detailed stochastic simulation. The breeding goal consisted of three traits and only own performance was included in the selection index. DE ROO (1988) found considerable reductions in response due to the Bulmer effect and due to correlated index values and finite sample size. Maximum overall reduction in short-term response equalled 34% (DE ROO, 1988, Table 8). In the present study, overall reduction in response ranged from 27% to 51% (Table 2). Due to the use of sib information, reductions in response due to the

Bulmer effect and due to correlated index values were larger than found by DE ROO (1988). DE ROO (1988) found optimum numbers of selected sires around 5 and 15 for time horizons of 5 and 25 years respectively. In the present study, the optimum number of selected sires ranged from 4 to 16 (Table 1 and Table 4) and increased with decreasing heritability and with decreasing purebred-crossbred genetic correlation when crossbred information was included (Table 4). The optimum number of boars for PLS was four (Table 4), which is very close to results of DE ROO (1988) for a time horizon of five year. For longer time horizons, DE ROO (1988) found higher optimum numbers of sires due to accumulated inbreeding. In our study, inbreeding was ignored and optimum numbers of selected sires have to be interpreted as giving the maximum response on short time horizons. For longer time horizons, the number of selected boars may need to be larger because of inbreeding. In the present study, correction of the selection intensity for finite population size and correlated index values was based on the total number of animals per generation, *i.e.*, selection took place only once per generation. In practise, selection takes place weekly as simulated by DE ROO (1988) and selection intensity is overestimated in the present study. Accounting for correlated index values changed the ranking of breeding schemes considerably, which was also found by MEUWISSEN (1991). This means that accounting for correlated index values is crucial when comparing breeding schemes.

Discrete generations were used in the present study, whereas in practise generations overlap. Overlapping generations provide opportunity for sequential culling, *i.e.*, animals can be culled if a better replacement is available. BELONSKY and KENNEDY (1988) found an increased response due to sequential culling of 34% and 57% when selection was on individual phenotype and on BLUP-EBV, respectively. However, sequential culling also increased inbreeding from 0.288 for BLUP selection without sequential culling to 0.383 when sequential culling was practised (BELONSKY and KENNEDY, 1988, Table 3). For this reason, the benefit of sequential culling would be smaller when responses were compared at the same level of inbreeding. In the present study, selected proportion among females was limited to a minimum of 0.25 because all replacements had to be produced from the current generation. When generation overlap, female selection intensity increases due to an increased number of selection candidates. It is not clear from our study if this would influence ranking of breeding schemes.

In practise, the number of selected boars is limited by inbreeding constraints. Inbreeding reduces the variance due to Mendelian sampling and is associated with risk due to drift variance and inbreeding depression when dominance effects are lost. For production traits, inbreeding depression is of minor importance in crossbreeding schemes, because dominance effects are regained in the hybrid slaughter pigs. Due to the presence of inbreeding constraints, it is valuable to compare responses at similar rates of inbreeding. An approximation of the benefit

of including crossbred information when the number of selected sires is limited by inbreeding constraints can be gained by comparing asymptotic responses at equal numbers of selected sires. It follows from Table 4, that especially for low values of r_{pc} the benefit of including crossbred information is larger when compared at equal numbers of sires than when compared at optimum numbers of selected sires. For example, for $r_{pc} = 0.4$ and $h^2 = 0.15$ the ratio of R_{as} over R_{pls} for a fixed number of 16 sires selected equalled 0.245/0.112 = 2.188, whereas the ratio of optimum responses equalled 0.245/0.121 = 2.025. Therefore, the benefit of including crossbred information seems to be larger than that judged by optimum responses when the number of selected sires is limited by inbreeding constraints. However, inclusion of crossbred information increases the correlation between index values, and will therefore also increase the rate of inbreeding, which limits the utility of a comparison between CCPS and PLS at equal numbers of selected sires. A promising approach seems to be the use of crossbred information together with a selection method that imposes restrictions or a cost factor on the rate of inbreeding (*e.g.*, MEUWISSEN, 1997; WRAY and GODDARD, 1994).

This research was financially supported by the Netherlands Technology Foundation (STW), and was co-ordinated by the Earth and Life Sciences Foundation (SLW). The authors want to thank Pim Brascamp, Ab Groen, Egbert Knol, Theo Meuwissen and Richard Spelman for giving helpful suggestions.

Literature Cited

BAUMUNG, R., SÖLKNER, J. and ESSL, A. 1996. Correlation between purebred and crossbred performance under a two-locus model with additive by additive interaction. Journal of Animal Breeding and Genetics 114: 89-98.

BELL, A. E. 1982. Selection for heterosis-results with laboratory and domestic animals. Proceedings of the second world congress on genetic applied to livestock production, Madrid, vol. 6, pp. 206-227.

BELONSKY, G. M. and KENNEDY, B. W. 1988. Selection on individual phenotype and best linear unbiased predictor of breeding value in a closed swine herd. Journal of Animal Science 66: 1124-1131.

BROWN, W. P. and BELL, A. E. 1980. An experimental comparison of selection alternatives to plateaued response. Genetics 94: 477-496.

BULMER, M. G. 1971. The effect of selection on genetic variability. The American Naturalist 105: 201-211.

BURROWS, P. M. 1972. Expected selection differentials for directional selection. Biometrics 28: 1091-1100.

DE BOER, I. J. M. and VAN ARENDONK, J. A. M. 1991. Genetic and clonal responses in closed dairy cattle nucleus schemes. Animal Production 53: 1-9.

DEKKERS, J. C. M. 1992. Asymptotic response to selection on best linear unbiased predictors of breeding values. Animal Production 54: 351-360.

DE ROO, G. 1987. A stochastic model to study breeding schemes in a small pig population. Agricultural Systems 25: 1-25.

DE ROO, G. 1988. Studies on breeding schemes in a closed pig population. 1. Population size and selection intensities. Livestock Production Science 19: 417-441.

DUCROCQ, V. and QUAAS, R. L. 1988. Prediction of genetic response to truncation selection across generations. Journal of Dairy Science 71: 2543-2553.

KENNEDY, B. W., SCHAEFFER, L. R. and SORENSEN, D. A. 1988. Genetic properties of animal models. Journal of Dairy Science 71: 17-26.

KIRSCH, W., FEWSON, D. and FENDER, M. 1962. Prüfungsanstalten als Hilfsmittel für die Selektion in der Schweinezucht. 1. Modellrechnungen über die Effektivität verschiedener Prüfungsmethoden und Prüfungssysteme. Zeitschrift für Tierzucht und Zuchtungsbiologie 77: 388-407.

MERKS, J. W. M. 1986. Genotype × environment interactions in pig breeding programmes. 1. Central test. Livestock Production Science 14: 365-381.

MEUWISSEN, T. H. E. 1989. A deterministic model for the optimisation of dairy cattle breeding based on BLUP breeding value estimates. Animal Production 49: 193-202.

MEUWISSEN, T. H. E. 1991. Reduction of selection differentials in finite populations with a nested full-half-sib family structure. Biometrics 47: 195-203.

MEUWISSEN, T. H. E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. Journal of Animal Science 75: 934-940.

PHOCAS, F. and COLLEAU, J. J. 1995. Approximating selection differentials and variances for correlated selection indices. Genetique, Selection et Evolution 27: 551-565.

RAWLINGS, J. O. 1976. Order statistics for a special class of unequally correlated multinormal variates. Biometrics 32: 875-887.

VAN DER WERF, J. H. J. 1990. Models to estimate genetic parameters in crossbred dairy cattle populations under selection. Doctoral Thesis, Wageningen Agricultural University, Wageningen, The Netherlands.

VILLANUEVA, B., WRAY, N. R. and THOMPSON, R. 1993. Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. Animal Production 57: 1-13.

WEI, M. and VAN DER STEEN, H. A. M. 1991. Comparison of reciprocal recurrent selection with pure-line selection systems in animal breeding (a review). Animal Breeding Abstracts 59: 281-298.

WEI, M., VAN DER STEEN, H. A. M., VAN DER WERF, J. H. J. and BRASCAMP, E. W. 1991a. Relationship between purebred and crossbred parameters. 1. Variances and covariances under the one locus model. Journal of Animal Breeding and Genetics 108: 253-261.

WEI, M., VAN DER WERF, J. H. J. and BRASCAMP, E. W. 1991b. Relationship between purebred and crossbred parameters. 2. Genetic correlation between purebred and crossbred performance under the model with two loci. Journal of Animal Breeding and Genetics 108: 262-269.

WEI, M. and VAN DER WERF, J. H. J. 1994. Maximizing genetic response in crossbreds using both purebred and crossbred information. Animal Production 59: 401-413.

WRAY, N. R. and HILL, W. G. 1989. Asymptotic rates of response from index selection. Animal Production 49: 217-227.

WRAY, N. R. and GODDARD, M. E. 1994. Increasing long-term response to selection. Genetique, Selection et Evolution 26: 431-451.

Genetic Gain of Pure Line Selection and Combined Crossbred Purebred Selection with Constrained Inbreeding

Piter Bijma*, John A. Woolliams† and Johan A. M. Van Arendonk*

* Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen University, 6700 AH Wageningen, The Netherlands and †Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, U.K.

Abstract – Using deterministic methods, rates of genetic gain (ΔG) and inbreeding (ΔF) were compared between pure line selection (PLS) and combined crossbred purebred selection (CCPS), for the sire line of a three-way crossbreeding scheme. Purebred performance and crossbred performance were treated as genetically correlated traits assuming the infinitesimal model. Breeding schemes were compared at a fixed total number of purebred selection candidates, *i.e.*, including crossbred information did not affect the size of the purebred nucleus. Selection was by truncation on estimated breeding values for crossbred performance. Rates of genetic gain were predicted using a pseudo-BLUP selection index. Rates of inbreeding were predicted using recently developed methods based on long-term genetic contributions. Results showed that changing from PLS to CCPS may increase ΔF by a factor 2.14. In particular with high heritabilities and low purebred-crossbred genetic correlations, CCPS requires a larger number of parents than PLS, to avoid excessive ΔF. The superiority of CCPS over PLS was judged by comparing ΔG from both selection strategies at the same ΔF, which revealed that CCPS was superior to PLS and the superiority of CCPS was only moderately reduced compared to the situation without a restriction on ΔF. This paper shows that the long-term genetic contribution theory can be used to optimise ΔF and ΔG in animal breeding schemes within very limited computing time.

In crossbreeding programmes, the primary aim is to increase the performance of the crossbred animals. For this reason, the breeding goal should be defined on the crossbred level, though some weight may be given to purebred performance in particular cases (JIANG and GROEN, 1999). Several studies have shown that selection response in crossbred performance can be increased by including both purebred and crossbred information in the selection criterion (WEI and VAN DER WERF, 1994; BIJMA and VAN ARENDONK, 1998; BAUMUNG *et al.* 1997; UIMARI and GIBSON, 1998). Purebred performance and crossbred performance can be treated as genetically different traits, and selection may be based on an estimated breeding value for crossbred performance (EBV) which utilises both purebred and crossbred information (WEI and VAN DER WERF, 1994; BIJMA and VAN ARENDONK, 1998). Such a selection strategy is referred to as Combined Crossbred Purebred Selection (CCPS), whereas selection based solely on information from the pure line is

Submitted

referred to as Pure Line Selection (PLS). None of the studies mentioned above, however, has considered the effect of CCPS on the rate of inbreeding. As will be argued next, the use of CCPS may substantially increase the rate of inbreeding compared to PLS.

In a CCPS programme, an important information source is the phenotypic information on crossbred half sibs of the selection candidate (WEI and VAN DER WERF, 1994). For example, in a three-way crossbreeding system, sires from the sire line may simultaneously be mated to dams of the sire line and to F1 dams from the multiplier level of the dam line. In that case, sires from the sire line produce two types of offspring; purebred offspring within the sire line, which are the selection candidates for the next generation, and commercial crossbred offspring which provide information on crossbred performance. When estimating breeding values for the purebred selection candidates, the information on their crossbred half sibs can be included in the EBV, which results in a higher accuracy of selection. WEI and VAN DER WERF (1993) and SPILKE *et al.* (1998) show how Mixed Model Equations can be set up to estimate genetic parameters and breeding values with CCPS.

When the genetic correlation between purebred and crossbred performance (r_{pc}) is low, the information coming from crossbred half sibs will dominate the EBVs of selection candidates. As a consequence, there will be a high intraclass correlation between EBVs of full and half sibs present among the purebred selection candidates. Depending on the value of r_{pc}, therefore, selection in a CCPS programme may tend to between family selection, resulting in increased rates of inbreeding. To maintain genetic variation and to avoid inbreeding depression in the pure line, rates of inbreeding have to be restricted in animal breeding programmes. The relevant question, therefore, is whether the superiority of CCPS schemes over PLS schemes can be sustained when rates of inbreeding are restricted. This question has not been addressed so far.

The aim of this paper is to compare CCPS schemes to PLS schemes while restricting the rate of inbreeding. First, we will compare the intraclass correlation between EBVs of sibs for PLS and CCPS schemes. Subsequently, rates of genetic gain and rates of inbreeding will be compared between CCPS and PLS schemes which have the same size of the pure line. Finally, we will maximise genetic gain from PLS and CCPS schemes by optimising the number of selected parents, while constraining the rate of inbreeding. The superiority of CCPS schemes over PLS schemes will be judged by comparing rates of gain from both selection strategies at the same rate of inbreeding.

Methods

Traits: Purebred and crossbred performance are treated as two different traits having a genetic correlation of r_{pc}. It is assumed that both traits are determined by the infinitesimal model, that phenotypic variance equals one for both traits and that heritability is equal for both traits.

Population structure: The population structure was the same as the structure considered by BIJMA and VAN ARENDONK (1998), and will be described briefly here. The sire line of a three-way crossbreeding system in pigs was modelled. Mating structure was hierarchical and generations were discrete. Sires from the sire line were mated to n_{pd} purebred dams from the sire line to produce n_{po} purebred offspring per dam ($\frac{1}{2}n_{po}$ of each sex), so that the total number of selection candidates summed over both sexes equalled $N = n_s n_{pd} n_{po}$. At the same time, the same sires were also mated to F1-dams from the multiplier level of the dam line, to produce n_{cb} crossbred fattening pigs per sire. Each purebred selection candidate, therefore, had n_{cb} crossbred half sibs providing crossbred information. Purebred performance was measured on animals within the sire line, and crossbred performance was measured on fattening pigs. Information on purebred and crossbred individuals was assumed to be available at the same time and was measured on both sexes. Each generation, n_s males and $n_s n_{pd}$ females were selected out of the purebred selection candidates of the current generation, to become parents of the next generation. Selection was by truncation on EBV for crossbred performance, without a restriction on the number of parents contributed by a single family.

Rates of gain: Throughout this paper, genetic gain in crossbred performance with PLS is denoted as ΔG_{pls} and genetic gain in crossbred performance with CCPS is denoted as ΔG_{ccps}. Rates of genetic gain in crossbred performance were predicted using the pseudo-BLUP selection index of BIJMA and VAN ARENDONK (1998). This index is an extension of the single trait pseudo-BLUP index of WRAY and HILL (1989) to CCPS, and combines phenotypic information of the selection candidate and its full and half sibs together with EBVs of parents into a pseudo-BLUP EBV for crossbred performance of the selection candidate. Detailed equations of the index are described by BIJMA and VAN ARENDONK (1998). BIJMA and VAN ARENDONK (1998) explicitly modelled the F1-dams, but here we will assume that EBVs of F1-dams are not available, which is likely to be the situation in practise. Consequently, the F1-dams were omitted from the index, so that the total number of information sources equalled 9, whereas the index of BIJMA and VAN ARENDONK (1998) contained 10 information sources.

Selection intensities accounted for finite population size and intraclass correlations between EBVs of relatives, and were calculated using the method of MEUWISSEN (1991). Selection and mating were iterated until BULMER's (1971) equilibrium parameters were reached, using the equations described by BIJMA and VAN ARENDONK (1998). Rates of genetic gain were predicted for the equilibrium situation.

Rates of inbreeding: Rates of inbreeding (ΔF) were predicted deterministically using the long-term genetic contribution theory (WOOLLIAMS *et al.* 1999; WOOLLIAMS and BIJMA, 2000). The application of this theory to CCPS is a direct analogy of the application to single trait BLUP selection, as described by BIJMA and WOOLLIAMS (2000). Here we will

only outline the main steps involved in predicting ΔF, detailed equations can be found in BIJMA and WOOLLIAMS (2000).

The procedure for predicting ΔF consisted of the following steps. First, BULMER's (1971) equilibrium genetic parameters were obtained by iterating on the pseudo-BLUP index of BIJMA and VAN ARENDONK (1998). Subsequently, the intraclass correlation between EBVs of full and half sibs were calculated from the selection index equations, following the approach described in appendix A of BIJMA and WOOLLIAMS (2000). Finally, rates of inbreeding were predicted using equations 1 through 13 and appendix B of BIJMA and WOOLLIAMS (2000). In equations 2 and 3 of BIJMA and WOOLLIAMS (2000), the single trait breeding value was replaced by the breeding value for crossbred performance. For single trait BLUP selection, an example of computation is given in appendix B of BIJMA and WOOLLIAMS (2000). In the present paper, the accuracy of the predicted rate of inbreeding will be evaluated by means of stochastic simulation for a number of CCPS schemes.

Optimisation of breeding schemes: Breeding schemes were optimised by maximising genetic gain while restricting the rate of inbreeding. Optimisation was performed at different fixed values of the mating ratio (n_{pd}) and of the total number of selection candidates (N). The optimisation variable was the number of sires, *i.e.*, the number of sires was increased until the inbreeding constraint was achieved. The number of dams and the number of offspring per dam was a result of the number of sires and the mating ratio, *i.e.*, the number of dams was $n_s n_{pd}$, and the number of offspring per dam was, $n_{po} = N/(n_s n_{pd})$. The number of offspring per dam was allowed to be a non-integer value, because the primary aim of this paper is to examine trends, and not to present values for specific practical breeding schemes. The optimisation is in one dimension, *i.e.* the only variable is the number of sires, and from $n_s \approx 12$ and higher, ΔF and ΔG are continuously decreasing functions of n_s, which makes it easy to identify the optimum scheme. With increasing n_s, the optimum scheme, *i.e.*, the optimum n_s, is the first scheme where ΔF is smaller than or equal to the constraint, and the maximum possible ΔG given the constraint is the rate of gain from this scheme. The optimality of CCPS *vs.* PLS was judged by comparing ΔG_{pls} and ΔG_{ccps} for the optimum schemes of both selection strategies, at the same rate of inbreeding and for the same total number of selection candidates.

Results

Intraclass correlation between EBVs of sibs: Figure 1 shows the intraclass correlation between EBVs of full and half sibs with CCPS, as a function of the purebred-crossbred genetic correlation (r_{pc}) and for two different heritabilities, $h^2 = 0.2$ or $h^2 = 0.6$. (Note that purebred and crossbred performance were assumed to have the same heritability.) With $r_{pc} = 1$, purebred and crossbred performance are the same trait, so that for this situation, CCPS corresponds to single trait selection. With $h^2 = 0.2$ and $r_{pc} = 1$, intraclass correlations are high, showing that with low heritability single trait

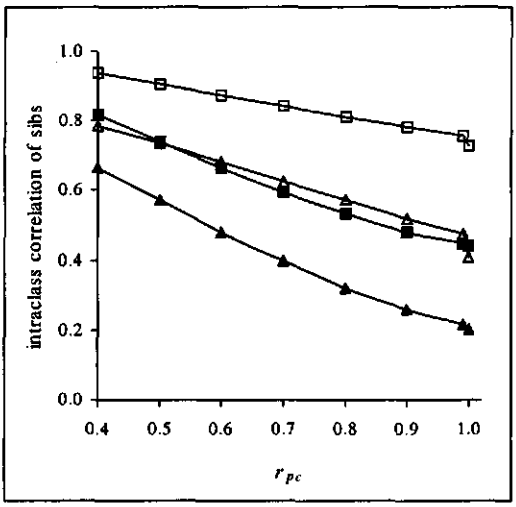

FIGURE 1.–Intraclass correlations between EBVs of full and half sibs with CCPS, as a function of r_{pc}, for n_s = 20, n_{pd} = 3, n_{po} = 4 and n_{co} = 60. —□— full sibs, h^2 = 0.2 ; —Δ— half sibs, h^2 = 0.2 ; —■— full sibs, h^2 = 0.6 ; —▲— half sibs, h^2 = 0.6; The symbols at r_{pc} = 1 represent intraclass correlations for PLS.

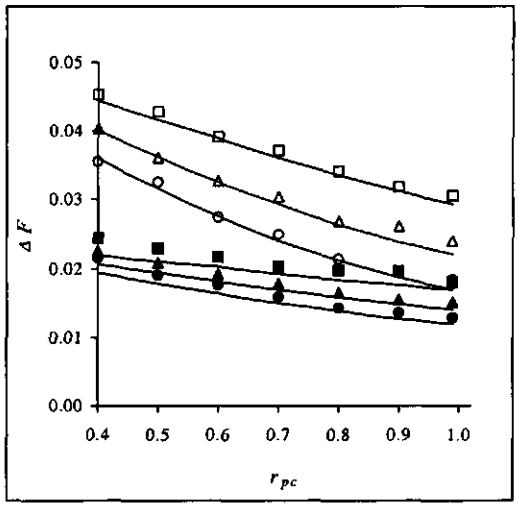

FIGURE 2.–Rates of inbreeding (ΔF) with CCPS as a function of r_{pc}, for n_s = 20, n_{pd} = 3 and n_{co} = 60. —□— n_{po} = 8, h^2 = 0.2; —Δ— n_{po} = 8, h^2 = 0.4; —○— n_{po} = 8, h^2 = 0.6; —■— n_{po} = 4, h^2 = 0.2; —▲— n_{po} = 4, h^2 = 0.4; —●— n_{po} = 4, h^2 = 0.6. Lines represent predictions, symbols represent results from simulation.

BLUP selection puts substantial emphasis on family information. With h^2 = 0.6 and r_{pc} = 1, intraclass correlations are substantially lower, because BLUP-EBV are largely determined by the own performance when heritability is high. With decreasing r_{pc}, the intraclass correlation increases rapidly for h^2 = 0.6 because emphasis shifts from the purebred own performance to family information for crossbred performance. For h^2 = 0.2, the increase is smaller because family information already receives substantial emphasis at r_{pc} = 1. With PLS, the intraclass correlation is slightly smaller than with CCPS and r_{pc} = 1, because with CCPS there are an additional 60 half-sibs included in the EBV. In conclusion, intraclass correlations increase with decreasing r_{pc} and the increase is largest for high heritabilities.

Rates of inbreeding and gain: Figure 2 shows the rate of inbreeding with CCPS as a function of r_{pc}, for different heritabilities and for two different selection intensities, n_{po} = 4 or 8. To evaluate the accuracy of the deterministic predictions (lines in Figure 2), Figure 2 also shows ΔF estimated from stochastic simulation (symbols in Figure 2), which reveals that predictions and simulations are in close agreement.

Figure 2 shows that ΔF increases with decreasing r_{pc} and the relative increase is largest with high heritabilities and high selection intensities. For example, with n_{po} = 8 and h^2 = 0.6, the rate of inbreeding increased by a factor of 2.14 when r_{pc} decreased from 1 to 0.4, whereas with h^2 = 0.2 the rate of inbreeding increased by a factor of 1.51 only. The large

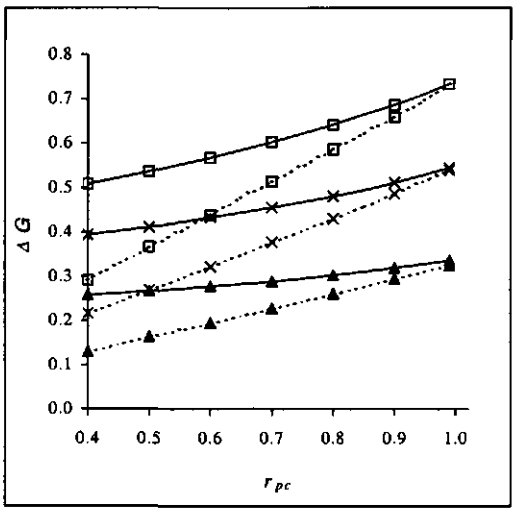

FIGURE 3.–Rates of genetic gain in crossbred performance (ΔG) as a function of r_{pc}, for n_s = 20, n_{pd} = 3, n_{po} = 8 and n_{co} = 60. For CCPS: —□— h^2 = 0.6; —×— h^2 = 0.4; —▲— h^2 = 0.2. For PLS: ---□--- h^2 = 0.6; ---×--- h^2 = 0.4; ---▲--- h^2 = 0.2.

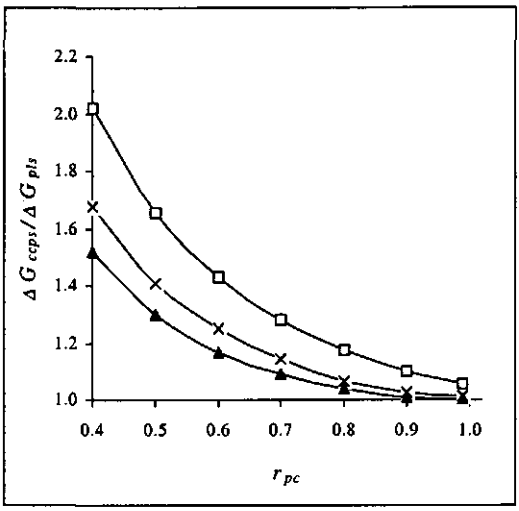

FIGURE 4.–Ratio of genetic gain with CCPS over genetic gain with PLS ($\Delta G_{ccps}/\Delta G_{pls}$) as a function of r_{pc}, for optimum schemes with ΔF < 1% and with N = 768, n_{pd} = 5 and n_{co} = 60. —□— h^2 = 0.15; —×— h^2 = 0.4; —▲— h^2 = 0.6.

increase of ΔF for high heritabilities agrees with the trend in the intraclass correlations observed in Figure 1.

Figure 2 also shows a rapid increase of ΔF when selection intensity increases. For example, with h^2 = 0.2 and r_{pc} = 1, ΔF increased by a factor 1.73 when the number of purebred offspring per dam was doubled from 4 to 8 and, for r_{pc} = 0.4, ΔF increased even by a factor 2.03. This shows that, with BLUP selection, doubling the number of selection candidates while keeping the number of parents constant may double the rate of inbreeding. Obviously, the rate of inbreeding with PLS (results not shown) is independent of r_{pc} and was slightly lower than ΔF for CCPS with r_{pc} = 1. In conclusion, Figure 2 shows that, when keeping the number of parents constant, changing from PLS to CCPS may substantially increase ΔF, in particular for small r_{pc}, high h^2 and high selection intensity.

Figure 3 shows ΔG_{pls} and ΔG_{ccps}, for n_{po} = 8 and for different heritabilities. With PLS, genetic gain in crossbred performance is equal to the product of $\Delta G_{purebred}$ and r_{pc}, and, therefore, ΔG_{pls} increases linearly with r_{pc}. As expected from previous studies (*e.g.*, WEI and VAN DER WERF, 1994), CCPS gives more genetic gain than PLS and the superiority of CCPS over PLS increases with decreasing r_{pc}. The relative superiority of CCPS over PLS is largest for low heritabilities. For example, with r_{pc} = 0.4 and h^2 = 0.2, using CCPS instead of PLS increased genetic gain by a factor 1.98, whereas for h^2 = 0.6, genetic gain increased only by a factor 1.73.

Combining Figure 2 and 3 shows, that for high heritabilities together with high selection

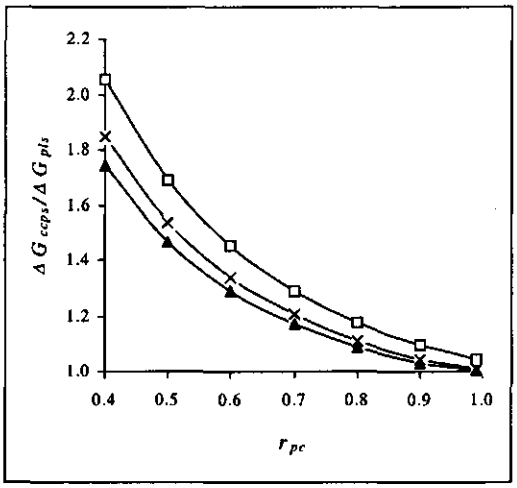

FIGURE 5.–Ratio of genetic gain with CCPS over genetic gain with PLS ($\Delta G_{ccps}/\Delta G_{pls}$) as a function of r_{pc}, without a restriction on ΔF, for n_s = 30, n_{pd} = 5, n_{po} = 6, and n_{co} = 60. —□— h^2 = 0.15; —×— h^2 = 0.4; —▲— h^2 = 0.6.

intensities, ΔF increases more rapidly than ΔG when changing from PLS to CCPS. For example, for h^2 = 0.6, r_{pc} = 0.4 and n_{po} = 8, changing from PLS to CCPS increased ΔG by a factor 1.73 and ΔF by a factor 2.14.

PLS vs. CCPS with constrained inbreeding: To make inferences on the superiority of CCPS over PLS when the rate of inbreeding is restricted, the ratio $\Delta G_{ccps}/\Delta G_{pls}$ was calculated for schemes where genetic gain was maximised while restricting ΔF. Figure 4 shows $\Delta G_{ccps}/\Delta G_{pls}$ for optimum schemes with a fixed total number of 768 pure line selection candiates, where ΔF was restricted to be smaller than 1%. To validate the deterministic optimisation, a limited number of schemes was optimised using stochastic simulation (results not shown), which revealed close agreement between predictions and simulations. Results in Figure 4 show that, with a restriction on the rate of inbreeding, CCPS remains superior over PLS, *i.e.*, the ratio $\Delta G_{ccps}/\Delta G_{pls}$ exceeds one. The superiority of CCPS over PLS increases with decreasing r_{pc} and with decreasing h^2. The ratio $\Delta G_{ccps}/\Delta G_{pls}$ was also evaluated for other values of N and d, but the relation was almost identical, which indicates that with constrained inbreeding, the superiority of CCPS over PLS is determined almost completely by r_{pc} and h^2. Results for other values of N and d, therefore, are not shown.

Figure 5 shows the superiority of CCPS over PLS without a restriction on the rate of inbreeding, for a scheme with n_s = 30, n_{pd} = 5 and n_{po} = 6. For h^2 = 0.15, the ratio $\Delta G_{ccps}/\Delta G_{pls}$ is very similar to the situation where ΔF is restricted, *e.g.*, with r_{pc} = 0.4 and no restriction on ΔF the ratio was 2.06 (Figure 5), whereas with ΔF<1% the ratio was 2.02. For low heritabilities, therefore, restricting ΔF hardly affects the superiority of CCPS over PLS. For higher heritabilities the superiority of CCPS over PLS decreased when ΔF was restricted, *e.g.*, with h^2 = 0.6, r_{pc} = 0.4 and no restriction on ΔF the ratio was 1.74 (Figure 5), whereas with ΔF<1% the ratio decreased to 1.51. Compared to the large increase of ΔF when changing from CCPS to PLS (Figure 2), the reduction of $\Delta G_{ccps}/\Delta G_{pls}$ when ΔF is restricted is strikingly small.

Table 1 shows the design of optimum breeding schemes, *i.e.*, the optimum number of selected sires, for different constraints on the rate of inbreeding. With PLS, the optimum number of sires is always lower than with CCPS and decreases when h^2 increases. With CCPS, the optimum number of sires increases with

TABLE 1.–Optimum number of selected sires for CCPS and PLS schemes with constrained inbreeding (ΔF)†

h^2	r_{pc}	$\Delta F < 2\%$	$\Delta F < 1\%$	$\Delta F < 0.5\%$
0.15	0.4	32	44	61
	0.5	31	44	60
	0.6	30	43	59
	0.7	29	42	58
	0.8	28	41	57
	0.9	27	40	56
	0.99	26	39	55
	PLS‡	26	38	53
0.4	0.4	30	42	59
	0.5	28	41	58
	0.6	27	39	56
	0.7	25	37	54
	0.8	24	36	53
	0.9	22	34	51
	0.99	21	32	49
	PLS‡	21	32	49
0.6	0.4	28	41	58
	0.5	27	39	56
	0.6	24	37	54
	0.7	22	34	51
	0.8	20	32	49
	0.9	19	30	46
	0.99	17	28	44
	PLS‡	17	28	44

† For $N = 768$, $n_{pd} = 5$ and $n_{co} = 60$, h^2 = heritability, r_{pc} = purebred-crossbred genetic correlation. ‡With PLS the optimum design is independent of r_{pc}.

decreasing h^2 and r_{pc}. With decreasing r_{pc}, the required number of sires increased rapidly for $h^2 = 0.6$, whereas for $h^2 = 0.15$ there is only a small increase. This result agrees with the observation that the increase of intraclass correlations is largest at high heritabilities (Figure 1). Results in Table 1 indicate that, with a constraint on ΔF, the difference between the optimum number of sires with PLS and CCPS increases with heritability. For low heritabilities, this difference is relatively small because single trait BLUP selection already puts substantial emphasis on family information.

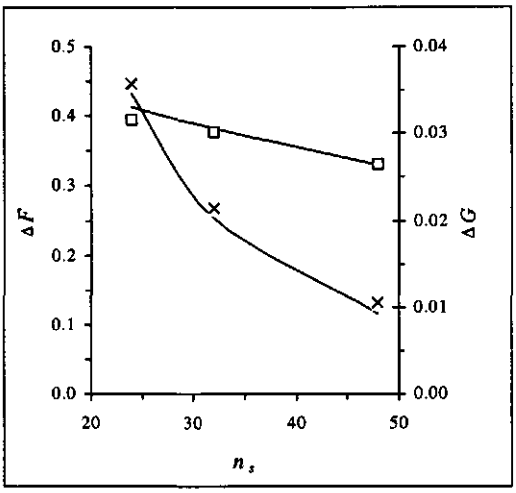

FIGURE 6.–Relation of ΔF and ΔG with the number of sires for a fixed test capacity, for $N = 768$, $n_{pd} = 4$, $h^2 = 0.4$, $r_{pc} = 0.4$ and $n_{co} = 60$; —□— ΔG; —×— ΔF;Lines represent predictions, symbols represent results from simulation.

The comparison of CCPS *vs.* PLS at fixed rates of inbreeding involves a trade off between accuracy and intensity of selection. Table 1 shows that with CCPS the constraint on ΔF is achieved by selecting more parents, which is at the expense of selection intensity. On the other hand, CCPS benefits from a higher accuracy of selection due to the use of crossbred information. The superiority of CCPS *vs.* PLS at fixed rates of inbreeding (Figure 4) indicates that the increase of gain due to increased accuracy exceeds the decrease of gain due to decreased selection intensity.

Given the large differences in ΔF between CCPS and PLS (Figure 2), the difference between the optimum number of sires for both selection strategies (Table 1) is surprisingly

small. This indicates that a small increase of the number of sires is sufficient to substantially reduce ΔF. Figure 6 shows the relation of ΔF and ΔG with the number of selected sires, for $h^2 = 0.4$, $r_{pc} = 0.4$ and $N = 768$, which shows that ΔF drops substantially when the number of sires increases, whereas ΔG decreases only moderately. There are two distinct mechanisms causing the sharp decrease of ΔF. First, in the absence of selection, ΔF approximately halves when the number of parents is doubled (WRIGHT, 1969). Second, with fixed N, selection intensity decreases when the number of parents increases, which further reduces ΔF (See appendix B of BIJMA and WOOLLIAMS 2000). For the scheme in Figure 6, both mechanisms contribute approximately equal to the reduction of ΔF. The different relation of ΔF and ΔG with the number of sires explains why the superiority of CCPS over PLS, *i.e.*, $\Delta G_{ccps}/\Delta G_{pls}$, reduces only moderately when ΔF is restricted (Figure 4 *vs.* Figure 5).

Discussion

In this paper, we have compared CCPS to PLS with particular emphasis on the rate of inbreeding. When keeping the number of parents constant, changing from PLS to CCPS may substantially increase ΔF. In particular with high h^2 and low r_{pc}, CCPS requires a larger number of parents than PLS, to avoid excessive rates of inbreeding. When compared at the same rate of inbreeding, CCPS was superior to PLS and the superiority of CCPS was only moderately reduced compared to a situation without a restriction on ΔF.

Up till recently, balancing rates of gain and inbreeding in animal breeding programmes required computationally demanding stochastic simulation, which seriously restricted the number of alternative schemes involved in the optimisation. This paper shows that the use of the long-term genetic contribution theory for predicting ΔF (WOOLLIAMS *et al.* 1999; WOOLLIAMS and BIJMA, 2000; BIJMA and WOOLLIAMS, 2000) enables the optimisation of breeding schemes within very limited computing time with no restriction on the number of alternative schemes involved. For example, results in Figure 2, which involves the evaluation of approximately 750 alternative breeding schemes, were generated within nine seconds.

Following the approach of WEI and VAN DER WERF (1994), we have treated purebred performance and crossbred performance as correlated traits, assuming the additive infinitesimal model. Other studies have used finite locus models with several degrees of dominance, and generally with equal effects for all loci (UIMARI and GIBSON, 1998; BAUMUNG *et al.* 1997). Neither of these models is fully realistic, *i.e.*, the true number of genes must be finite, but on the other hand we have little knowledge of the number of genes, their effects and the interactions between genes.

Prediction of short-term selection response by means of selection index theory does not require the infinitesimal assumptions, providing gene effects are not very large and gene frequency of favourable alleles is near 0.5 (HILL, 1998). When phenotypes and breeding values follow a multivariate normal distribution,

the regression of breeding values on phenotypes, *i.e.*, the selection index, is unbiased. With multivariate normality, therefore, selection index theory correctly predicts response to a single cycle of selection. Though selection may induce deviations from normality, TURELLI and BARTON (1994) showed that the prediction of genetic gain by regression on phenotypes is robust against deviations from normality, even with strong truncation selection. Therefore, prediction of response by means of selection index theory is expected to be adequate in the short term.

In the long-term, selection response depends on the nature of the genetic variation, *i.e.*, the number of genes, their effects *etc*. However, for livestock populations we have little knowledge of those parameters, and it will be difficult to convince practitioners to sacrifice short-term response (*i.e.*, by not using CCPS) when it remains doubtful whether doing so will indeed yield more long-term response. The most practical approach, therefore, seems to be the use of CCPS while restricting the rate of inbreeding, so as to maximise short-term response and simultaneously avoid rapid erosion of genetic variation. To account for changing genetic parameters due to changing gene frequency, genetic parameters should frequently be re-estimated so that optimality of the CCPS EBV is guaranteed.

In this study, breeding schemes were optimised given a fixed total number of purebred selection candidates, whereas the number of recorded crossbred individuals varied between schemes. In other studies, crossbred individuals were tested at the expense of purebred individuals, thus reducing the number of selection candidates (WEI and VAN DER WERF, 1994). Whether testing crossbred individuals is at the expense of purebred individuals depends on whether testing of crossbred individuals is part of the nucleus-breeding programme, or, alternatively, information from the commercial population is collected, *e.g.*, at the slaughterhouses. Collecting information at the slaughterhouse has the advantage that one can directly measure the traits that determine the carcass price, *e.g.*, carcass grade and lean%, instead of an ultrasonic measurement of backfat. Additionally, measuring crossbred performance on individuals of the commercial population has the advantage that G×E interaction is accounted for. When crossbred information is collected outside the nucleus, PLS and CCPS are not compared at the same total amount of resources, so the relevant question is whether the additional gain arising from using CCPS instead of PLS makes up for the cost of measuring crossbred individuals.

CCPS and PLS can be compared at the same level of resources by specifying the ratio of the costs of testing a purebred *vs.* a crossbred individual, and evaluating breeding schemes at the same total costs. In that case, testing crossbred individuals will reduce the number of purebred selection candidates, which reduces selection intensity and has a decreasing effect on ΔF (Figure 2). When compared at the same level of recourses, therefore, changing from PLS to CCPS will give smaller increases of ΔF than the values presented in Figure 2. Methods used in this study can as well be used to

optimise rates of inbreeding and gain for crossbreeding schemes with a fixed total level of resources.

In this study, constraints on ΔF were achieved by increasing the number of parents. Alternatively, one can use dynamic selection algorithms that directly constrain ΔF (MEUWISSEN, 1996; GRUNDY *et al.* 1998). Though dynamic selection algorithms are expected to give higher ΔG at the same ΔF, optimisation of breeding schemes using those algorithms can only be implemented by means of computationally demanding stochastic simulation, which limits the number of alternative schemes considered in the optimisation. Our methods provide good insight into the impact of different parameters and as such provide a good starting point for the optimisation of breeding schemes.

This research was financially supported by the Netherlands Technology Foundation (STW), and was co-ordinated by the Life Sciences Foundation (SLW). J.A.W. acknowledges the Ministry of Agriculture, Fisheries and Food (United Kingdom) for financial support.

Literature Cited

BIJMA, P. and VAN ARENDONK, J. A. M. 1998. Maximizing genetic gain for the sire line of o crossbreeding scheme utilizing both purebred and crossbred information. Animal Science 66: 529-542.

BIJMA, P. and WOOLLIAMS, J. A. 2000. Prediction of rates of inbreeding in populations selected on Best Linear Unbiased Prediction of breeding value. Genetics *in press.*

BAUMUNG, R., SÖLKNER, J. and ESSL, A. 1997. Selection response according to the use of pure- and/or crossbred information in crossbred animals with varying levels of dominance and additive by additive effects. In: Book of abstracts of the 48th Annual Meeting of the European Association for Animal Production (EAAP), August 25-28, Vienna, Austria.

BELL, A. E. 1982. Selection for heterosis-results with laboratory and domestic animals. Proceedings of the second World Congress on Genetic Applied to Livestock Production, Madrid, vol. 6, pp. 206-227.

BULMER, M. G. 1971. The effect of selection on genetic variability. The American Naturalist 105: 201-211.

GRUNDY, B., VILLANUEVA, B. and WOOLLIAMS, J. A. 1998. Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genetical Reseach (Cambridge) 72: 159-168.

HILL, W. G. 1998. Inferences from evolutionary biology to livestock breeding. In: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, January 11-16, Armidale, Australia.

JIANG, X. and GROEN, A. F. 1999. Combined crossbred and purebred selection for reproduction traits in a broiler dam line. Journal of Animal Breeding and Genetics 116: 111-125.

MEUWISSEN, T. H. E. 1991. Reduction of selection differentials in finite populations with a nested full-half-sib family structure. Biometrics 47: 195-203.

MEUWISSEN, T. H. E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. Journal of Animal Science 75: 934-940.

SPILKE J., GROENEVELD, E., and MIELENZ, N. 1998. Joint purebred and crossbred (co)variance component estimation with a pseudo multiple trait model: loss in efficiency. Journal of Animal Breeding and Genetics 115: 341-350.

UIMARI, P. and GIBSON, J. P. 1998. The value of crossbreeding information in selection of poultry under a dominance model. Animal Science 66: 519-528.

WEI, M. and VAN DER WERF, J. H. J. 1993. Animal model estimation of additive and dominance variances in egg production traits of poultry. Journal of Animal Science 71: 57-65.

WEI, M. and VAN DER WERF, J. H. J. 1994. Maximizing genetic response in crossbreds using both purebred and crossbred information. Animal Production 59: 401-413.

WOOLLIAMS, J. A., BIJMA, P. and VILLANUEVA, B. 1999. Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153: 1009-1020.

WOOLLIAMS, J. A. and BIJMA, P. 2000. Predicting rates of inbreeding in populations undergoing selection. Genetics 154: 1851-1864.

WRAY, N. R. and HILL, W. G. 1989. Asymptotic rates of response from index selection. Animal Production 49: 217-227.

WRIGHT, S. 1969. Evolution and the genetics of populations. Vol. 2 The Theory of Gene Frequencies. Univ. of Chicago, Chicago.

General Discussion

This thesis has primarily focussed on developing methods to predict rates of inbreeding in (artificially) selected populations, using the concept of long-term genetic contributions (WRAY and THOMPSON, 1990). CHAPTER 2 presents a general approach to predict long-term genetic contributions using simple linear models and the assumption of an equilibrium. CHAPTER 5 relates rates of inbreeding to expected genetic contributions making a separate prediction of the variance of genetic contributions redundant. Together, CHAPTER 2 and 5 present a general approach to predict rates of inbreeding in randomly mating populations undergoing selection. Following the general theory, prediction equations for the rate of inbreeding were derived for mass selection and BLUP selection, for populations with either discrete or overlapping generations. For mass selection, it was shown how the current prediction equations relate to previous equations (SANTIAGO and CABALLERO, 1995; NOMURA, 1996)

Besides prediction of rates of inbreeding, substantial attention was paid to prediction of genetic contributions of current individuals to the population in the long-term (long-term genetic contribution) and to the relationship between rates of genetic gain and genetic contributions. It was shown how selective advantage affects the development of pedigree and the turn over rate of the population, *i.e.*, the reciprocal of the generation interval. A novel definition of generation interval was presented, which states that the generation interval is the turnover time of genes that are destined to stay in the population. For overlapping generations, genetic gain was described in a gene flow context, which, contrary to conventional gene flow theory (HILL, 1974), is based on truly asymptotic proportions of genes.

Finally, CHAPTERS 9 and 10 demonstrate how the theory developed in previous chapters may be implemented to optimize breeding schemes with combined crossbred purebred selection.

Many topics have already been addressed in the Discussion sections of the relevant chapters and will not be repeated here. In this General Discussion, I will address the relevance of this thesis for quantitative genetic theory and for practical animal breeding. I will discuss the thesis primarily from an animal breeder's perspective, *i.e.*, no attention will be given to natural populations and constant population size will be assumed. (Note that, in this discussion, the term "inbreeding" is used to indicate effective population size [*i.e.*, $\Delta F = 1/2N_e$], it does not refer to the preferential mating of relatives.)

Quantitative Genetic Theory

Rate versus Increment of Inbreeding

In animal breeding, it is sometimes argued that the level of inbreeding (F_t) or the increment (F_t–F_{t-1}) is the crucial parameter, rather than the rate of inbreeding [$\Delta F = (F_t–F_{t-1})/(1-F_{t-1})$]. The difference between ΔF on the one hand and F_t

or F_t–F_{t-1} on the other hand, is that the level and increment depend on the definition of a base population, whereas ΔF does not.

In most animal breeding populations, however, base populations are hypothetical, primarily constructed to facilitate our understanding of populations or to enable prediction of breeding values (BLUP). In reality there is no such base population since the evolution of populations is an ongoing process. When populations are kept at a constant size, a balance between selection, mutation and drift will arise resulting in equilibrium genetic parameters. This equilibrium state depends on the rate of inbreeding (or effective population size), indicating that the rate of inbreeding is the essential population parameter. Furthermore, the rate of fixation of favorable mutations and the selection against deleterious mutations, depend on the rate of inbreeding and not on the increment (*e.g.*, LYNCH and WALSH, 1998; HILL, 2000).

Though helpful in particular cases, the concept of a base population can be misleading. For example, in the study of WRAY and THOMPSON (1990), the concept of a base population introduced a problem with the reference point for the genetic variance. The assumption of equilibrium genetic parameters (SANTIAGO and CABALLERO, 1995) was a natural way forward to solve this problem. Furthermore, certain definitions of a base population are incompatible with the assumption of an equilibrium. For example, in the absence of selfing, defining a base population by a group of non-inbred and unrelated individuals implies that the rate of inbreeding from the base population to the next generation is zero, which violates the assumption of equilibrium. In an equilibrium, a base population should be defined by a group of non-inbred individuals having an average coefficient of kinship of ΔF, or by considering their parents as a group of unrelated individuals with an average inbreeding coefficient of $-\Delta F$.

Prediction of Rates of Inbreeding

History: Prediction of rates of inbreeding has a long history, starting with the pioneering work of Wright (WRIGHT, 1922, 1931, 1938, 1939; see CABALLERO, 1994 for a review). For populations undergoing selection for a heritable character, predictions become complicated due to the inheritance of selective advantage. BURROWS (1984a,b) developed predictions that were strictly valid for a single cycle of selection. Hence, Burrows' methods under-predict ΔF in ongoing selection programs. Predictions that account for the inheritance of selective advantage were initiated by ROBERTSON (1961), who introduced the concept of accumulation of selective advantages. When selective advantage is halved each generation, the cumulative selective advantage equals, $Q = 1 + ½ + ¼ + 1/8 + ... = 2$. However, due to the reduction of genetic variance in selected populations (BULMER, 1971), selective advantage is more than halved each generation and $Q < 2$. Robertson's method, therefore, gives an overprediction of the rate of inbreeding (WRAY and THOMPSON, 1990; SANTIAGO and CABALLERO, 1995). Based on Robertson's idea, SANTIAGO and CABALLERO (1995) developed predictions

where the inheritance of selective advantage depended on the remaining proportion of genetic variance after selection, showing that, for mass selection and random mating, Q equals $2/(1+kh^2)$. Important advancements of SANTIAGO and CABALLERO (1995) were the use of equilibrium genetic parameters (BULMER, 1971) and the inclusion of non-random mating.

A different approach was initiated by WRAY and THOMPSON (1990), who showed a fundamental relationship between the rate of inbreeding and the sum of squared long-term genetic contributions and, for the first time, obtained accurate predictions of ΔF in populations undergoing mass selection. Their method, however, was complicated due to the recursive nature of the prediction procedure and the need for predicting the variance of long-term genetic contributions. WOOLLIAMS *et al.* (1993) showed that the recursive equations of WRAY and THOMPSON (1990) lend themselves to vertical summation over generations, which enabled a closed form equation. However, prediction equations were still difficult, sufficiently so that WRAY *et al.* (1994) were only able to make limited use of them in developing predictions for selection on sib indices. As a result, predictions were complex, both conceptually and algebraically.

This thesis: This thesis has shown that the variance of long-term genetic contributions is related to the mean, which considerably simplifies predictions (CHAPTER 5). Furthermore, using equilibrium genetic parameters and a gene flow approach, a simple closed form equation for the prediction of expected long-term genetic contributions was derived, both for discrete and overlapping generations (CHAPTER 2). For the first time, prediction of rates of inbreeding has been extended to populations with overlapping generations undergoing selection. CHAPTER 6 shows that, for mass selection with random mating, the prediction equation of SANTIAGO and CABALLERO (1995) is equivalent to the prediction equations based on long-term genetic contributions. The same equivalence was found for the special case of mass selection with overlapping generations and an equal number of parents selected from each age class (CHAPTER 6; NOMURA, 1996). Together, CHAPTER 2 and 5 present a general approach to predict rates of inbreeding in randomly mating populations undergoing artificial selection.

There are a number of important issues in this thesis. First, the importance of the relationship between squared long-term genetic contributions and ΔF (WRAY and THOMPSON, 1990; CHAPTER 5) is that it holds for selected populations, with no assumptions on the form of selection, provided that genes are ultimately mixed and that an equilibrium exists over which a stable ΔF may be defined. In CHAPTER 5, the relation between squared genetic contributions and ΔF was derived directly from identity by descent, which enhances intuitive understanding compared to the derivation of WRAY and THOMPSON (1990).

The current method assumes independent segregation of the locus for which the rate of inbreeding is predicted and the loci affecting the selected trait, *i.e.*, the prediction relates to the expected increase in homozygosity of a neutral gene, unlinked to genes affecting the

selected trait. The current prediction also relates to the rate of inbreeding estimated from the pedigree of a population. For livestock species, which have many chromosomes, linked segregation is probably of academic interest since the average gene may be practically unlinked to genes with large effect. When there are some genes with a large effect, ΔF for those genes and for the chromosomal region surrounding those genes will differ from ΔF for the rest of the genome. This indicates that there is no single measure for the rate of inbreeding of a population and an *a priori* choice has to be made for which loci one would like to predict the rate of inbreeding. Following the approach of SANTIAGO and CABALLERO (1995), SANTIAGO and CABALLERO (1998) developed predictions of effective population size for neutral loci linked to the selected loci. With respect to inbreeding depression, which is mainly observed for traits related to fitness, the rate of inbreeding of a neutral gene is probably the most relevant one, because genetic correlations between fitness and production traits are generally small (RAUW *et al.*, 1998). Furthermore, for practical purposes, the value of a gene dependent rate of inbreeding seems to be rather limited, because, for the majority of the genes, we have little knowledge about their effect on the phenotype.

An important feature of the current approach is that it contains a general model for the inheritance of selective advantage (Π) and for the effect of the selective advantage on the number of selected offspring (Λ), so that it can be applied to a variety of selection strategies and modes of inheritance (*e.g.*, imprinting) without the need for modifying the basic theory. The main task for applying the prediction method to other selection systems and modes of inheritance is to derive the regression coefficients for the Π and Λ matrices. For certain selection systems this task may be difficult (*e.g.*, the quadratic model for BLUP selection, CHAPTER 7), but the concept remains clear. This viewpoint is supported by the observation that prediction errors for the cases studied originated from the goodness of fit of the prediction model for genetic contributions, whereas the basic equations remained valid. From the equivalency of the present method and the drift method (SANTIAGO and CABALLERO, 1995) in special cases (see CHAPTER 6), it follows that the complexities involved in modeling Π and Λ are not a peculiarity of the present method, but will also be encountered if an attempt were made to extend the drift approach, *e.g.*, to BLUP selection.

In CHAPTER 6, for the first time, a simple closed form equation was given, for the rate of inbreeding with mass selection and discrete generations, expressed in commonly used genetic and population parameters. [Though the same equation can be derived directly from the work of SANTIAGO and CABALLERO (1995)]. Such a simple equation is likely to enhance the application of the genetic contribution theory, which has often been hampered by the perception that it is very complicated.

Overlapping generations: With overlapping generations, the generation interval (L) was defined as the length of time in which genetic contributions sum to unity. This definition specifies the genetic time unit of a population as

the average time interval between two meioses for genes that are destined to stay in the population. The reciprocal of L is a measure of the turnover rate of the population, *i.e.*, $1/L$ is the average number of meioses per unit of time. Compared to the average age of parents at birth of their offspring, the current definition of generation interval is of a more genetical and less operational nature.

With discrete generations, effective population size can be partitioned into contributions due to each sex, $1/N_e = 1/(4N_{em}) + 1/(4N_{ef})$ (CROW and DENNISTON, 1998; SANTIAGO and CABALLERO, 1995). The analogous equation based on genetic contributions is $\Delta F = \frac{1}{4}N_m E(r_m^2) + \frac{1}{4}N_f E(r_f^2)$. Because each sex contributes half of the genes, there is no competition between the sexes and both terms can be treated separately.

With overlapping generations, there are more than just two categories, and within sex there is competition between the categories. Thus categories within sex cannot be treated separately, *i.e.*, when predicting genetic contributions, one has to account for competition between categories. In this thesis, competition between categories is accounted for by simultaneously solving the system of regression equations for predicting genetic contributions (CHAPTER 2, Equations 7 and 9).

Non-random mating: In this thesis, rates of inbreeding were predicted for randomly mating populations. Non-random mating has two opposite effects on the rate of inbreeding (CABALLERO and HILL, 1992). On the one hand, avoiding mating of relatives increases the average heterozygosity and, therefore, increases the within family drift (*i.e.* drift due to Mendelian sampling). On the other hand, avoiding mating of relatives accelerates the mixing of ancestral lines within the population, so that the genes of the different ancestors are equally represented among descendants at an earlier stage. Once genes are equally represented among descendants, between family drift (*i.e.*, differential fitness of families) no longer affects the genetic contributions of the ancestor generation. Avoiding mating of relatives, therefore, reduces ΔF due to between family drift. WANG and HILL (2000) derived equations for ΔF in the absence of selection, which explicitly show the two different sources of genetic drift.

In the absence of selection, the extension of the present method to populations with a certain proportion, p, of full-sib mating is straightforward. Following the approach of CHAPTER 5, where conditioning on the selective advantage can be omitted: $E(r_i^2) = \mu^2 + \sigma^2$ with $\sigma^2 = E_n[\mathrm{Var}(r_i|n_i)] + \mathrm{Var}_n[E(r_i|n_i)]$, where n_i denotes family size. Considering the second term of σ^2, it follows from $r_i = \frac{1}{2}\sum_{offspring} r_j$ that $E(r_i|n_i) = \frac{1}{2}n_i E(r_j|n_i)$, which, with Poisson family size (*i.e.*, $\mathrm{Var}(n_i)=2$) and r_j independent of n_i, gives $\mathrm{Var}_n[E(r_i|n_i)] = \frac{1}{2}E(r_j)^2 = \frac{1}{2}\mu^2$ in an equilibrium. Considering the first term of σ^2, when mating two full sibs, both full sibs will have the same long-term contribution, to that $\mathrm{Cov}(r_j,r_{j'}|n_i) = \mathrm{Var}(r_j|n_i)$. With a proportion p of full sib mating, therefore, $\mathrm{Var}(r_j|n_i) = \frac{1}{4}(1+p)n_i\mathrm{Var}(r_j|n_i)$, and with $E(n_i) = 2$ it follows that $E_n[\mathrm{Var}(r_i|n_i)] = \frac{1}{2}(1+p)\sigma^2$. Adding both terms gives: $\sigma^2 = \frac{1}{2}(1+p)\sigma^2 + \frac{1}{2}\mu^2$, so that $\sigma^2 =$

$\mu^2/(1-p)$, which shows that preferential mating of full sibs increases the variance of genetic contributions. Next, with a mating ratio of one, so that $N_m = N_f = \frac{1}{2}N$ and using $\Delta F = \frac{1}{4}(1-\omega)N\mathrm{E}(r^2)$ and $\mu = 1/N$ it follows that $\Delta F = (1-\omega)[(4N)^{-1} + (4N)^{-1}(1-p)^{-1}]$, where ω is the correlation between genes within individuals. Substituting $p = 4\omega(1+3\omega)^{-1}$ (GHAI, 1969; CABALLERO and HILL, 1992) shows that, in the absence of selection, $\Delta F = \frac{1-\omega}{4N} + \frac{1+3\omega}{4N}$. This can be simplified to $\Delta F = \frac{1}{2N}(1+\omega)$, which is identical to the result of CABALLERO and HILL (1992). Note that, in the absence of selection, ΔF is independent of the proportion of full sib mating when $\mathrm{Var}(n_i) = 2/3$, in which case the two opposite effects of mating on the rate of inbreeding cancel each other out.

In the above equation for ΔF, the first part represents the minimum possible rate of inbreeding, $1/(4N)$, which is inbreeding solely due to within family drift, and is multiplied by a factor $(1-\omega)$ when mating is non-random. The second part represents the additional inbreeding due to between family drift, and is multiplied by a factor $(1+3\omega)$ when mating is non-random. Those two terms nicely illustrate the two opposite effects of non-random mating on the rate of inbreeding.

The result of SANTIAGO and CABALLERO (1995) for selected populations suggests that, in the presence of selection, ΔF with non-random mating is a straightforward analogy of ΔF in the absence of selection. Analogous to CABALLERO and HILL (1992) and to the equation presented above, the final equation of SANTIAGO and CABALLERO (1995) consists of two parts. The first part is $(1-\omega)/4N$, and the second part is the additional inbreeding due to between family drift, which is multiplied by $(1+3\omega)$. The agreement between the result of SANTIAGO and CABALLERO (1995) and CABALLERO and HILL (1992) suggests that the effect of mating on the variance of the contributions acts independently of the effect of selection. Unfortunately, in the presence of selection, a general derivation of explicit prediction equations for ΔF with non-random mating, using genetic contributions, appears to be less straightforward than in the absence of selection.

Genetic Gain

In CHAPTER 2, it has been shown that the rate of genetic gain is proportional to the expectation of the product of long-term genetic contribution and Mendelian sampling term of an ancestor, $\mathrm{E}[\Delta G] \propto \mathrm{E}[ra]$. WOOLLIAMS and THOMPSON (1994) already presented this relationship, but no formal derivation was given.

An intuitive understanding of this relationship can be gained by noticing that breeding values of any generation can be decomposed into a sum of Mendelian sampling terms of individuals in previous generations, weighted by their long-term genetic contribution. With n generations of selection, the total genetic advance equals $n\Delta G$. With $n \rightarrow \infty$, this total genetic advance must be equal to the weighted sum of Mendelian sampling terms over all ancestors in those n generations. Assuming an equilibrium, the total genetic advance should be divided equally among the n generations.

Analogously, the weighted sum of Mendelian sampling terms should be divided equally among the n generations, to that the share conferred to a single generation is equal to the weighted Mendelian sampling terms summed over the number of parents in a single generation. Therefore, the rate of genetic gain per generation must be equal to the product of Mendelian sampling term and the genetic contribution, summed over all ancestors in a generation.

Additionally, CHAPTER 4 shows that the expression $\Delta G = \Sigma ra$ represents the genetic gain originating from Mendelian sampling of a single generation summed over all subsequent cycles of selection, whereas $\Delta G = i\rho\sigma_A$ represents the genetic gain of a single cycle of selection originating from Mendelian sampling of all previous generations. Both methods give valid results.

Predictions for ΔG can be obtained by replacing the genetic contribution by its expectation given the selective advantage: $u_{i,k} = E[r_{i,k} | s_{i,k}]$, so that $r_{i,k} = u_{i,k} + e_{i,k}$. This prediction is expected to be accurate when $Cov(a_{i,k},e_{i,k}) = 0$, *i.e.*, conditioning on the selective advantage should remove any covariance between the Mendelian sampling term and the genetic contribution. CHAPTER 2 shows that equations resulting from this approach agree with established results. For practical purposes, however, it is probably easier to use conventional selection index theory, *e.g.*, the work of VILLANUEVA *et al.* (1993).

Consistency with classical theory: CHAPTER 2 shows that, for a general sib index, ΔG predicted from the genetic contribution theory is consistent with classical theory. This consistency is best illustrated with mass selection in discrete generations, using the notation of CHAPTER 7.

With equal selection intensities in both sexes, $N_m = N_f = \frac{1}{2}N$ and $u = \alpha + \beta(A - \bar{A})$, it follows that $\Delta G = N\mathrm{E}(ua) = N\alpha\mathrm{E}(a) + N\beta\mathrm{E}[a(A-\bar{A})]$, where $\mathrm{E}(a) = \frac{1}{2}\sigma^2_{A_0} i/\sigma_P$ and $\mathrm{E}[a(A-\bar{A})] = \frac{1}{2}\sigma^2_{A_0}(1-kh^2)$, where $\sigma^2_{A_0}$ is the base generation additive genetic variance and h^2 is the equilibrium heritability (CHAPTER 3, APPENDIX B). Next, with $\alpha = 1/N$, $\beta = \alpha\lambda/(1-\pi)$, $\lambda = \frac{1}{2}i/\sigma_p$, and $\pi = \frac{1}{2}(1-kh^2)$ (CHAPTER 3, APPENDIX A) it follows that $\Delta G = \frac{1}{2}\sigma^2_{A_0} i/\sigma_P + \frac{1}{2}i\sigma^2_{A_0}(1-kh^2)/[\sigma_P(1+kh^2)] = i\sigma^2_{A_0}/[\sigma_p(1+kh^2)]$. With Bulmer equilibrium, $\sigma^2_A = \frac{1}{2}\sigma^2_A(1-kh^2) + \frac{1}{2}\sigma^2_{A_0}$, so that $\sigma^2_{A_0} = \sigma^2_A(1+kh^2)$ (see also VILLANUEVA and KENNEDY, 1990). Substitution gives $\Delta G = ih^2\sigma_P$, which is the "breeders equation" for equilibrium genetic gain (FALCONER, 1989, p.192; LYNCH and WALSH, 1998, p.50). Therefore, ΔG predicted from the genetic contribution theory is identical to ΔG predicted from classical theory for mass selection with discrete generations.

The Relationship between Gain and Inbreeding

Under the infinitesimal model, inbreeding reduces genetic variance, which in turn reduces genetic gain. Furthermore, when inbreeding depression is present, fitness of the population may reduce to an extent where it affects the

selection differentials, *i.e.*, indirectly inbreeding depression may also reduce genetic gain. In the short-term, inbreeding and genetic gain have an unfavorable relation, in the sense that measures that increase genetic gain, as selecting only a few parents and using all information of relatives (*e.g.*, BLUP), generally increase the rate of inbreeding. Maximizing short-term response, therefore, reduces long-term response and involves substantial risk (*e.g.*, VERRIER *et al.*, 1993).

The genetic contribution theory shows a fundamental relation between the rate of inbreeding and genetic gain, *i.e.*, $\Delta F = \frac{1}{4}\Sigma r^2$ and $\Delta G = \Sigma ra$. When genetic gain is to be maximized with a constraint on the rate of inbreeding, the theoretically optimum solution is for all those with a Mendelian sampling term below a certain level to have no long-term contribution and, then, for the long-term contribution to increase linearly with the Mendelian sampling term (BONDESSON, 1989; WOOLLIAMS and THOMPSON, 1994; GRUNDY *et al.*, 1998). Generally, this optimum can not be achieved, since we can neither set long-term contributions of current individuals to their desired values independently of previous generations, nor observe the Mendelian sampling terms without error. However, genetic gain for the theoretical optimum can be regarded as an upper bound and $\Delta G_{realized}/\Delta G_{optimum}$ can be used as a genetical measure of the efficiency of breeding schemes (GRUNDY *et al.*, 1998).

Figure 1 and 2 in CHAPTER 8 show that, with BLUP selection, the relationship between the breeding value and the genetic contribution is non-linear. Since the optimum relationship between the contribution and the Mendelian sampling term is linear, BLUP selection is not an optimum strategy for maximizing gain while restricting inbreeding. (Note that the breeding value and the Mendelian sampling term follow a bivariate normal distribution and are, therefore, linearly related.)

Up to this point, discussion has ignored the generation of new variation due to mutation. Mutation is generally ignored in animal breeding, since it is regarded of academic interest and relevant only in the very long-term. However, this seems to be a misconception, because populations starting from an inbred base population show continued response (*e.g.*, MACKAY *et al.*, 1994) and estimates of the newly arising variation due to mutation are surprisingly high. For example, for 6-week weight of mice, CABALLERO *et al.* (1995) estimated a mutational heritability, $h_m^2 = \sigma_{m0}^2 / \sigma_E^2$, of 0.0034, where σ_{m0}^2 is the per generation newly arising variation due to mutation. Other estimates for the mouse are generally higher (KEIGHTLEY, 1998; See table 12.1, LYNCH and WALSH, 1998). Assuming that mutations have small effects, so that their fate is determined only by drift (neutral model, LYNCH and HILL, 1986) a mutational heritability of 0.0034 is sufficient for maintaining heritability ($h^2 = \sigma_A^2 / \sigma_P^2$) at relatively high values. In an equilibrium, the loss of genetic variance due to inbreeding and the newly arising variance due to mutation cancel each other out, so that $\sigma_A^2(1-\Delta F) + \sigma_{m0}^2 = \sigma_A^2$ giving $\sigma_A^2 = \sigma_{m0}^2 / \Delta F$ and $h^2 = h_m^2 / (\Delta F + h_m^2)$ (WEI *et*

al., 1996; LYNCH and WALSH, 1998, pp. 330). Note that σ_A^2 and h^2 refer to an unselected population, *i.e.*, they are not reduced by the Bulmer effect.

With a rate of inbreeding of *approx.* 1% (*e.g.*, mass selection with N_m = 20, N_f = 40, 4 offspring per dam and $h^2 \approx 0.25$), a mutational heritability of 0.0034 gives and asymptotic heritability of 0.25, indicating that common values of h^2 can be sustained in relatively small populations. The relationship between ΔF and σ_A^2 in the equilibrium situation, shows that, contrary to the short-term, at equilibrium, lower ΔF gives higher ΔG. Increasing short-term gain, *e.g.*, by increasing selection intensity, will shift the equilibrium to a lower value of σ_A^2.

The above section is based on strong assumptions, *i.e.*, it holds for additive gene action, and only when mutations are selectively neutral or when the infinitesimal model is assumed. Nevertheless, it is useful as a reference point. Furthermore, though the equilibrium heritability may correctly assess the genetic variance, prediction of the equilibrium genetic gain requires knowledge of the distribution of effects and frequency of mutations. For example, when most mutations are deleterious, the newly arising variance due to mutation hardly contributes to genetic progress. Asymmetric responses observed in selection experiments indicate that the majority of mutations may be in the same direction (MACKAY *et al.*, 1994). It would seem logical that the proportion of mutants that are deleterious increases with the mean value of the trait, *i.e.*, when the trait reaches a high level, most mutations are probably deleterious. This idea is supported by the fact that the majority of mutations is deleterious for fitness (*e.g.*, LYNCH and WALSH, 1998, Chapter 12), a trait which has been under selection for a very long time.

Since the majority of mutations is deleterious for fitness, there is a conflict between artificial selection and natural selection when mutations are pleiotropic (ROBERTSON, 1967; KEIGHTLEY and HILL, 1990). Fixation of mutations that are advantageous for the selected trait, but have deleterious effects on fitness will reduce selection response due to negative correlated response for fitness (HILL and MBAGA, 1998). Such negative correlations between fitness and production traits are widely observed in animal breeding populations (RAUW *et al.*, 1998). On the phenotypic level, negative correlations between production and fitness traits may be explained from a resource allocation point of view (BEILHARZ *et al.*, 1993; LUITING *et al.*, 1999). However, a shift of resources from fitness to production traits is likely to be due to an increasing frequency of pleiotropic genes. Both explanations, therefore, do not contradict each other. An overview of factors influencing the maintenance of genetic variance in animal breeding populations is in HILL (2000).

Animal Breeding

Optimization of Breeding Schemes

Conceptual approaches: In the literature there are two main approaches to balance short and long-term response. The first approach is to maximize the cumulative (discounted) economic return, either up to a certain time

horizon, or over all future generations, accounting for the effect of inbreeding on selection response, both due to variance reduction and due to inbreeding depression. In this approach, a breeding scheme with a higher rate of inbreeding is accepted, as long as economic effects of a higher selection response outbalance those of inbreeding. The second approach is to maximize selection response while directly constraining the rate of inbreeding (*e.g.*, MEUWISSEN 1997a). The essential difference between both approaches is that the first approach considers the cost of inbreeding and, therefore, is a purely economical procedure, whereas the second approach considers inbreeding in itself and is more related to risk.

There are a number of reasons why the second approach is to be preferred. First, accurate prediction of cumulative selection response requires detailed knowledge of the nature of the genetic variation and of all effects of inbreeding. Lack of knowledge of certain effects of inbreeding is likely to result in below optimal weight given to the rate of inbreeding (MEUWISSEN 1997b). Given our limited knowledge, the first approach seems to be an optimistic one.

Economic optimization, *e.g.*, maximizing the net present value, may lead to high rates of inbreeding, in particular for species with long generation interval (JAMES, 1972). Though such a scheme may be optimal from an economical perspective, it will not be adapted in practice because the rate of inbreeding is considered as being too high. Instead, the value of the cost factor will be questioned, and subsequently the cost factor will be increased until acceptable rates of inbreeding are achieved.

Additionally, the more sophisticated implementations of time horizon methods suffer from unrealistic solutions in the final generations. For example, in a study on mixed inheritance, the major gene was fixed in one big step in the final generation (DEKKERS and VAN ARENDONK, 1998). The analogy with the problem of balancing gain and inbreeding would be to select only one sire and one dam in year 19, since this is expected to give the highest response in year 20. Obviously, however, in year 19 one would look another 20 years ahead, not just a single year. In realistic implementations of time horizon methods, therefore, the solution proposed for a specific generation should not be a direct function of the time span until the horizon. The solution may of course be an indirect function of time, for example when it depends on a gene frequency that changes over time.

It is interesting to note that there is a contradiction between risk assessed from an economic point of view and risk assessed from a genetic point of view. In an economic context, high discount rates may be associated with low risk, *i.e.*, returns in the distant future are less certain than immediate returns, so that increasing the discount rate decreases risk (SMITH, 1978). In a genetic context, however, high discount rates lead to breeding schemes with high rates of inbreeding, which involves increased risk associated with small population size.

Because, in the vast majority of cases, there are no readily available replacements for the

currently used breeds or lines, selection decisions should aim at maximizing economic return under the condition that the line can be maintained. Selection decisions should, therefore, consider the balance of mutation, inbreeding and (natural) selection. This requires putting a direct constraint on the rate of inbreeding, as is done in the second approach.

Acceptable levels of ΔF: To put a direct constraint on the rate of inbreeding, one needs to make a decision on what level is acceptable. Given our limited knowledge of the nature of genetic variation, it is difficult to formally derive the acceptable rate of inbreeding. However, some insight may be gained from selection experiments and quantitative genetic theory.

When mutation is ignored, acceptable rates of inbreeding are primarily determined by the extent of inbreeding depression for fitness. MEUWISSEN and WOOLLIAMS (1994) studied the balance between natural selection and inbreeding depression. Using Fisher's (1929) fundamental principle of natural selection, they showed that the minimum effective size to prevent a decline in fitness equals $N_e = D/\sigma^2_{WA}$, where D is the inbreeding depression of fitness with complete inbreeding and σ^2_{WA} is the additive genetic variance for fitness. Values of D for major components of fitness can be measured relatively easily, and are generally between 0.5 and 1% inbreeding depression per percent inbreeding (*e.g.*, FALCONER, 1989, p. 249; LYNCH and WALSH, 1998, p. 272; WIENER *et al.* 1992). Values of σ^2_{WA}, however, are more difficult to measure, particularly because artificial selection interferes with natural selection. In the absence of a correlation between fitness and the trait under selection, MEUWISSEN and WOOLLIAMS (1994) obtained effective population sizes ranging from 250 to 31 per generation, (ΔF = 0.2% to 1.6%). Furthermore, they concluded that, with a negative correlation between fitness and the selection criterion, a decline in fitness could not be prevented.

GODDARD (1992) studied the optimal effective size of the global population of black and white dairy cattle, using a separate discount rate to account for risk associated with small population size. The optimum solution maximized the net present value of all future benefits from the breeding program. Optimum effective size primarily depended on the amount of inbreeding depression and on the discounting factor used. The rate of inbreeding for the optimum schemes ranged from 0.2% to 1.3% per generation.

Another important aspect of the rate of inbreeding is the probability of fixation of favorable alleles, which can be approximated by $u(q) \approx 1 - \exp(-2N_e sq)$, where s is the selective value of the allele and q is the allele frequency (KIMURA, 1962; CABALLERO and SANTIAGO, 1998; HILL, 2000). It turns out that favorable alleles are relatively safe, once they have a reasonable frequency. For example, for a livestock breeding program with $\Delta F \approx 1\%$, the fixation probability of an additive allele, with an effect of 0.07 phenotypic standard deviation and a frequency of 0.2, exceeds 90%. Approximately 750 of those alleles would be required to make up a heritability of 0.3, and

fixation of 90% of them would give a huge selection response of approximately 36 phenotypic standard deviations. Fixation of favorable alleles that already have a reasonable frequency, therefore, does not impose stringent restrictions on ΔF. [See HILL (2000) for further discussion.]

When considering mutation and assuming the infinitesimal model, restricting ΔF to 1% seems to be sufficient to maintain heritability on reasonable levels (see above). Furthermore, fixation of favorable mutants and selection against deleterious mutants both profit from larger effective population size. To quantify those effects, however, one needs to have knowledge of the distribution of effects and frequency of mutations, which is not available for livestock breeds at present.

Given the above considerations, the acceptable level of inbreeding seems to be determined primarily by the extent of inbreeding depression on fitness. Detailed knowledge of the relevant parameters to determine the level of the constraint is lacking, but the different approaches all point towards a value between 0.5% and 1% per generation.

When the commercial product is a crossbred, inbreeding depression may seem less important. However, inbreeding depression is most important for traits related to fitness, like reproductive performance. Reproduction traits are clearly important when considering the risk of losing the pure line and particularly for the efficiency of the multiplication required in crossbreeding systems. It does not matter, therefore, whether or not the commercial product is a crossbred.

Many studies indicate that ΔF can be reduced substantially with no or minor loss of ΔG, particularly when using advanced selection algorithms (*e.g.*, MEUWISSEN 1997; GRUNDY *et al.*, 1998). Furthermore, high rates of inbreeding involve substantial risk, especially for dairy cattle breeding which is dominated by a single breed. Thus risk can be reduced substantially at low cost, and a risk adverse attitude should be adopted. Breeding companies should, therefore, target low rates of inbreeding by implementing advanced selection algorithms and fitness should be included in the breeding goal to avoid a conflict between natural and artificial selection (HILL and MBAGA, 1988).

Implementing the present method: The present methods for predicting rates of inbreeding in selected populations facilitate a computationally feasible (*i.e.*, deterministic) optimization of short and long-term response for static breeding schemes. The term "static breeding schemes" refer to schemes like mass selection and selection on BLUP-EBV, where selection decisions do not take account of the relation of specific individuals with the rest of the population.

Regarding the balancing of inbreeding and genetic gain, the present method can answer questions like: what is the optimum number of parents, is it worthwhile to increase the number of selection candidates, is it worthwhile to collect progeny information or should selection be based on sib information. The main application of the present methodology, therefore, is the optimization of the *design* of breeding schemes, not the optimization of the selection criterion.

Other methods: Many studies have addressed the balance between rates inbreeding and gain, mostly using stochastic simulation. There are two main approaches to reduce ΔF. First, the selection criterion can be optimized by reducing weight given to family information (*e.g.*, GRUNDY and HILL, 1993; VERRIER *et al.*, 1993; GRUNDY *et al.*, 1994; VILLANUEVA and WOOLLIAMS, 1997). Second, ΔF can be reduced by using factorial or minimum coancestry mating (*e.g.*, WOOLLIAMS, 1989; CABALLERO *et al.*, 1996). Combining both methods may further reduce ΔF (*e.g.*, VILLANUEVA *et al.*, 1994; LUO *et al.*, 1995). Furthermore, WRAY and GODDARD (1994) and BRISBANE and GIBSON (1995) reduced inbreeding by putting a cost factor on the average relationship of the selected animals. None of these studies, however, developed a theoretical concept of optimality.

MEUWISSEN (1997a) introduced a dynamic selection tool to maximize genetic gain while restricting the rate of inbreeding (see also GRUNDY *et al.*, 1998). Given the available selection candidates, this method maximizes the genetic level of the selected group while constraining the average coefficient of coancestry, and it is therefore equivalent or superior to other selection criteria. In combination with factorial and minimum coancestry mating strategies, this method is state of the art (SONESSON and MEUWISSEN, 2000). Implementation of the method results in a dynamic breeding program, where the number of parents and the number of offspring per parent may vary, depending on the candidates available in a particular generation.

Deterministic optimization of dynamic breeding schemes, where the rate of inbreeding is constrained in advance, requires methods to predict rates of gain for such schemes. This means that the selection differential needs to be predicted for a subset of the population of which the average coefficient of coancestry is constrained. A general solution to this problem is expected to be difficult, particularly for populations with overlapping generations.

Dynamic selection rules (*e.g.*, MEUWISSEN 1997a; GRUNDY *et al.*, 1998) are more efficient in maximizing gain while restricting inbreeding than static selection rules. (An additional advantage is that dynamic selection tools are likely to reduce the variance of the rate of inbreeding). Dynamic selection rules are often regarded as being selection tools during the course of a breeding scheme, whereas static selection tools are primarily suitable for a priori design of breeding schemes (VILLANUEVA and WOOLLIAMS, 1997; GRUNDY *et al.*, 1998). However, this distinction is false. The ideal approach would be to select using the tool that gives the highest genetic gain while constraining the rate of inbreeding. (This will be a dynamic selection rule.) Next, the design of the breeding scheme (*e.g.*, the number of selection candidates, which information should be measured *etc.*) should be optimized given this specific selection tool, *i.e.*, the *a priori* optimization of the design of breeding schemes should take account of the selection tool that is going to be used. *A priori* optimization of the design of breeding schemes assuming a static selection tool and the subsequent use of a dynamic selection tool makes the static

predictions invalid and may result in a sub optimal breeding scheme. The current problem is that we cannot deterministically predict genetic gain for dynamic selection procedures. With *a priori* optimization of the design using dynamic selection tools, therefore, one has to resort to computationally demanding stochastic simulation, which limits the number of alternative schemes that can be considered. However, this is a methodological problem, it does not indicate that one cannot *a priori* optimize breeding schemes while accounting for the use of dynamic selection rules.

Characteristics of optimum schemes: Maximizing genetic gain while constraining inbreeding will change the layout of breeding schemes compared to simply maximizing genetic gain. Following the pioneering work of NICHOLAS and SMITH (1983) and subsequently many others (*e.g.*, MEUWISSEN and WOOLLIAMS, 1993), breeding schemes for dairy cattle have moved towards selection based on sib information which enables higher short-term genetic gain. However, for species where the trait of interest cannot be measured on the selection candidate, maximizing gain while restricting inbreeding is likely to move optimum selection schemes back to progeny testing, in particular when population size is small and the constraint on ΔF is stringent (MEUWISSEN and SONESSON, 1998).

As mentioned above, the degree of optimality of breeding schemes can be judged by the extent to which long-term genetic contributions are set to their desired values, which depends on the Mendelian sampling term. When selection is based solely on sib information, no information on the Mendelian sampling term of the candidate itself is obtained, so that the selection is between Mendelian sampling terms of the parents and more distant ancestors. Since the number of parents is generally much lower than the number of candidates, sib selection can be considered as a reduction of the selection differential among Mendelian sampling terms.

Optimum breeding schemes for maximizing gain while restricting inbreeding are likely to be characterized by: the use of BLUP to estimate breeding values; the use of dynamic selection tools rather than simply truncation selection; decreased importance of sib information; increased importance of progeny testing; increased generation intervals; increased accuracy at the expense of selection intensity; and a mating strategy that accelerates the mixing of ancestral lines within the population, *e.g.*, factorial mating and/or minimum coancestry mating.

Animal Breeding as a Commercial Enterprise

Animal breeding has become a commercial enterprise. Breeding companies often hesitate to sacrifice short-term response in order to safeguard long-term response, since their competitive position may deteriorate. Particularly in dairy cattle breeding, where estimated breeding values of bulls from different companies can be compared directly and genetic material is freely accessible on the market, there is a strong stimulus to focus on short-term response. Obviously, breeding companies are not willing to sacrifice short-term response especially if their competitor is

going to benefit from the genetic variation they saved. Hence, there is a clear contradiction between collective interest and private interest and, in practice, collective interest is likely to get the worst of it.

The consequences of short-term selection are demonstrated by the findings that 50% of the almost 5000 bulls from 18 countries born in 1990, evaluated by the Interbull Center, were bred by only five sires (WICKHAM and BANOS, 1998). Examination of the top-cow and top-heifer data sets of the NRS (Royal Dutch Cattle Syndicate), which are used to select bull dams, revealed that the effective number of sires was below 8 and that the effective number of great-grandsires was around 4 to 5. A single bull (Carlin M. Ivanhoe Bell) accounted for 50% of the genes contributed by males (PRINS, 1999).

The fact that we do not observe dramatically increased rates of inbreeding in the commercial population is because the commercial population lags behind the nucleus population, and also because there is a certain degree of avoidance of mating or relatives. Additionally, in many countries, local black and white breeds have been replaced by the Holstein-Friesian (HF) breed, which has temporarily reduced inbreeding levels and hidden the between family selection that is going on. By the time increased rates of inbreeding are observed in the commercial population, genetic contributions of ancestors will be largely established. Hence, there will be little that can be done. Monitoring inbreeding rates in the commercial population, therefore, is of little use compared to managing genetic variation in the elite population. On the contrary, publication of inbreeding levels in the commercial population may serve as a false reassurance.

To safeguard genetic variation of the Holstein-Friesian breed, commitment of the leading AI companies is essential. In the end this will benefit everybody, but nothing will change unless initiative is taken by one of the companies.

In pig and poultry breeding, breeding companies generally own the lines and the competitor does not have access to the genetic material. Maintaining genetic variation of the line, therefore, benefits the breeding company itself and not its competitors. In pig and poultry breeding, therefore, a restriction on the rate of inbreeding is easier to achieve than in dairy cattle breeding.

The Relevance of Inbreeding in Future Breeding Schemes

Future changes in animal breeding programs will mainly be due to new developments in reproductive technology and molecular biology (genetic markers and knowledge of individual genes; BRASCAMP, 1998). Driving forces for implementing new technology are an increased commercial attitude among breeding companies and direct competition on a global market.

Reproductive Technology

Reproductive technologies may facilitate more efficient breeding schemes for maximizing genetic gain while restricting inbreeding. For example, the use of ovum pick up technologies in combination with in-vitro maturation of embryo's may enable complete factorial

designs, which give higher genetic gain at the same rate of inbreeding (WOOLLIAMS, 1989; WOOLLIAMS and WILMUT, 1989; KINGHORN *et al.* 1991; DE BOER and VAN ARENDONK, 1994; SONESSON and MEUWISSEN, 2000).

Contrary to the expectation of BRASCAMP (1998), progeny testing remains an important selection tool for situations where the trait of interest can not be measured on the selection candidate itself (*e.g.*, bulls in dairy cattle breeding), because selection based on progeny testing may yield higher genetic gain at the same rate of inbreeding (LUO *et al.*, 1995; MEUWISSEN and SONESSON, 1998).

When cloning becomes possible, in dairy cattle a completely different structure may evolve, which is closer to the pig and poultry breeding structures. In such a situation, the commercial product may be a crossbred clone, *e.g.*, HF×Jersey, to exploit non-additive genetic variance, primarily for traits related to fitness. To ensure good quality of the commercial product, a number of clone genotypes may be evaluated in a similar way as the current progeny testing schemes. Since the reliability of clone testing is: $\rho^2 = nh^2/[h^2(n\text{-}1)+1]$, with a heritability of 0.3, only 21 individuals would be required to reach a reliability of 90%. In such a situation, the performance of the commercial population may exceed that of the nucleus population. Selection in the pure line may include progeny testing of males on a limited number of testing herds, and aim at maximizing gain while restricting inbreeding. DE BOER and VAN ARENDONK (1994) showed that, with cloning, higher rates of gain at the same rate of inbreeding can be obtained by selecting fewer dams than sires.

When the commercial product is a crossbred genotype and pure line testing takes place on a limited number of testing herds, pure-line genetic material is no longer available to the competitor and, therefore, it becomes easier to restrict the rate of inbreeding (see above). Additionally, the use of crossbred clones may lead to the reemerging of isolated strains within the HF population, which is desirable since it would increase the between breed diversity and thus reduce risk. In the past, isolated strains were based on geographical separation, whereas in the future they may be based on commercial separation. These considerations indicate that, contrary to common belief, the use of cloning may increase the genetic variation within the HF breed. A broader discussion on the potential benefits of cloning in relation to genetic diversity is in WOOLLIAMS and WILMUT (1999).

Molecular Biology

One of the first applications of molecular biology in animal breeding is the possibility to select for know genes or accurately mapped QTL. GIBSON (1994) showed that there is a conflict between short and long-term gain when selecting for an identified gene. DEKKERS and VAN ARENDONK (1998) and PONG-WONG and WOOLLIAMS (1998) confirmed the result of GIBSON (1994) and showed that it is more efficient to fix a favorable allele in a number of small steps instead of a few big steps. However, they did not restrict the rate of inbreeding. VILLANUEVA *et al.* (1999) showed that, when maximizing genetic gain while

restricting inbreeding, this conflict between short and long-term response was negligible. This result indicates that, when maximizing gain while restricting inbreeding, there is little interaction over time; *i.e.*, with a constraint on the rate of inbreeding, current selection decisions do not jeopardize future possibilities for making genetic progress. This conclusion is expected to be rather robust. (However, one should probably allow for some recombination in the region adjoining the major gene, to avoid loss of potentially favorable genes in this region that have not yet been identified. The major gene, therefore, should not be fixed in very few generations).

When knowledge of the genome is increasing and techniques for genetic modification of individuals become common tools, one may wonder whether inbreeding remains relevant at all, since favorable alleles may be introduced into the population artificially. However, most quantitative traits seem to be highly polygenic, so there will be many genes with small effect. It is very unlikely that we will be able to identify all of those genes and accurately estimate their effect, primarily since this would require huge amounts of phenotypic data. Development of new molecular technology, *e.g.*, for studying the expression of genes, will not solve this problem because one still needs to validate and accurately estimate the effect of the gene within the population in which it is supposed to be used. In particular the effect of recent mutations will be difficult to estimate, since it may take some time before the number of individuals that carry the mutation is large enough to have sufficient phenotypic data. Furthermore, when epistatic effects are important (*e.g.*, MACKAY, 1998), we cannot accurately predict the result of an introgression program in advance. Therefore, our knowledge of the genome is insufficient to identify and preserve all the important genes and a restriction on the rate of inbreeding remains essential. Furthermore, restricting the rate of inbreeding will also preserve genes affecting traits that are currently perceived as being unimportant, but which may become important in the future. Finally, natural selection is becoming increasingly important, but it will only be effective in a population of sufficient size. In conclusion, restricting the rate of inbreeding is a general and risk averse way of safeguarding genetic variation within a population, and is likely to be more robust than relying solely on molecular technology.

On the other hand, developments in molecular biology may be used to more effectively maintain genetic variation in a population. For example, markers may be used in a conservation program that aims at minimizing the loss of alleles (TORO *et al.*, 1998; LAMBERSON, 1998). The essential benefit of using markers to reduce the rate of inbreeding is that one can observe the drift due to Mendelian sampling. WANG and HILL (2000) show that, in principle, N_e can be increased indefinitely by restricting the sampling between maternally and paternally inherited genes within individuals, using genetic markers. For species with many chromosomes, however, the required amount of marker information and number of offspring per family restricts the efficiency of the method. In selected populations, the benefit of using markers to reduce drift due to

Mendelian sampling will be smaller, because the amount of inbreeding due to Mendelian sampling is only a small proportion of the total inbreeding.

Methods for predicting rates of inbreeding may be extended to populations with mixed inheritance. With mixed inheritance, selection will affect the gene frequency of the major gene so that there will not be an equilibrium. Note that in the absence of equilibrium, there is no single ΔF to predict. In principle, genetic contributions can also be predicted in the absence of an equilibrium, but this requires iterative equations (CHAPTER 3, Equation 11, APPENDIX C). It is doubtful whether this is worth the effort. It is probably more practical to approximate the rate of inbreeding assuming an equilibrium. Furthermore, new major genes may come up during the course of the selection program. Assuming an equilibrium, therefore, may be more valid than assuming a single major gene which will be fixed over time, so that in the end only polygenic variance remains.

Gerard Albers, Johan van Arendonk, Sijne van der Beek, Henk Bovenhuis, Pim Brascamp, Henk Geertsema, Ab Groen, Egbert Knol, Jascha Leenhouwers, Theo Meuwissen, Liesbeth van der Waaij, and John Woolliams are acknowledged for giving useful comments on this manuscript. This research was financially supported by The Netherlands Technology Foundation (STW) and was coordinated by the Earth and Life Science Foundation (ALW).

Literature Cited

BEILHARZ, R. G., B. G. LUXFORD and J. L. WILKINSON, 1993. J. Anim. Breed. Genet. 110: 161-170.

BONDESSON, L., 1989. Mathematical principles of optimal utilization of clones under diversity constraints. Swedish University of Agricultural Sciences, Section of Forest Biometry, Report 27, Umeå, Sweden.

BRASCAMP, E. W., 1998. The development of animal breeding enterprises in the next decades: reasons to change the structure of breeding programmes. Acta. Agric. Scand. 28: 61-67.

BRISBANE, J. R. and J. P. GIBSON, 1995. Balancing selection response and inbreeding by including predicted stabilized genetic contributions in selection decisions. Genet. Sel. Evol. 27: 541-549.

BULMER, M. G., 1971. The effect of selection on genetic variability. Am. Nat. 105: 201-211.

BURROWS, P. M., 1984a. Inbreeding under selection from unrelated families. Biometrics 40: 357-366.

BURROWS, P. M., 1984b. Inbreeding under selection from related families. Biometrics 40: 895-906.

CABALLERO, A., 1994. Developments in the prediction of effective population size. Heredity 73: 657-679.

CABALLERO, A. and W. G. HILL, 1992. Effective size of non-random mating populations. Genetics 89: 591-614.

CABALLERO, A. and SANTIAGO, E., 1998. Survival rates of major genes in selection programmes. Proc. 6th World Congress on Genetic Applied to Livestock Production, Armidale, Australia. 26: 5-12.

CABALLERO, A., P. D. KEIGHTLEY and W. G. HILL, 1995. Accumulation of mutations affecting body weight in inbred mouse lines. Genet. Res. 65: 145-149.

CABALLERO, A., E. SANTIAGO, and M. A. TORO, 1996. Systems of mating to reduce inbreeding in selected populations. Anim. Sci. 62: 431-442.

CROW, J. F. and C. DENNISTON, 1998. Inbreeding and variance effective population numbers. Evolution 42: 482-495.

DE BOER, I. J. M. and J. A. M. VAN ARENDONK, 1994. Additive response to selection adjusted for effects of inbreeding in a closed dairy cattle nucleus assuming a large number of gametes per female. Anim. Prod. 58: 173-180.

DEKKERS, J. C. M. and J. A. M. VAN ARENDONK, 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations. Genet. Res. Camb. 70: 257-275.

DEKKERS, J. C. M., 1992. Structure of breeding programmes to capitalize on reproductive technology for genetic improvement. J. Dairy Sci. 75: 2880-2891.

FALCONER, D. S., 1989. *Introduction to quantitative genetics*. Longman, Harlow, United Kingdom.

FISHER, R. A., 1929. The general theory of natural selection. Dover Publications, New York.

GIBSON, J. P., 1994. Short-term gain at the expense of long-term response with selection of identified loci. Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph, Canada. 21: 201-204.

GHAI, G. L., 1969. Structure of populations under mixed random and sib mating. Theor. Appl. Genet. 39: 179-182.

GODDARD, M. E., 1992. Optimal effective population size for the global population of black and white dairy cattle. J. Dairy. Sci. 75: 2902-2911.

GRUNDY, B. and W. G. HILL, 1993. A method for reducing inbreeding with Best Linear Unbiased Prediction. Anim. Prod. 56: 427 (abstr.)

GRUNDY, B., A. CABALLERO, E. SANTIAGO, and W. G. HILL, 1994. A note on using biased parameter values and non-random mating to reduce rates of inbreeding in selection programmes. Anim. Prod. 59: 465-468.

GRUNDY, B., B. VILLANUEVA and J. A. WOOLLIAMS, 1998. Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. Camb. 72: 159-168.

HILL, W. G., 1974. Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18: 117-139.

HILL, W. G., 2000. Maintenance of genetic variation in animal breeding programmes. Livest. Prod. Sci. in press.

HILL, W. G. and S. H. MBAGA, 1998. Mutation and conflicts between artificial and natural selection for quantitative traits. Genetica 102/103: 171-181.

JAMES, J. W., 1972. Optimum selection intensity in breeding programmes. Anim. Prod. 14: 1-9.

KEIGHTLEY, P. D., 1998. Genetic basis of response to 50 generations of selection on body weight in inbred mice. Genetics 148: 1931:1939.

KEIGHTLEY, P. D. and W. G. HILL, 1990. Variation maintained in quantitative traits with mutation-selection balance: Pleiotropic side effects on fitness traits. Proc. Roy. Soc. Lond. B242: 95-100.

KIMURA. M., 1962. On the probability of fixation of mutant genes in a population. Genetics 47: 713-719.

KINGHORN, B., P. and J. C. M. DEKKERS, 1991. Potential genetic gains in dairy cattle with gamete harvesting and in vitro fertilization. J. Dairy Sci. 74: 611-622.

LAMBERSON, W. R., 1998. Marker allele frequency index selection as a method for reducing loss of genetic variation in small populations. In: Proc. of the 6th WCGALP. University of New England, Armidale, Australia.

LUO, Z. W., J. A. WOOLLIAMS and R. THOMPSON, 1995. Controlling inbreeding in dairy MOET nucleus schemes. Anim. Sci. 60: 379-387.

LYNCH M. and W. G. HILL, 1986. Phenotypic evolution and neutral mutation. Evolution 40: 915-935.

LYNCH, M. and B. WALSH, 1998. Genetics and Analyses of *Quantitative Traits*. Sinauer Associates, Inc. Sunderland Massachusetts, USA.

LUITING, P., O. VANGEN, W. M. RAUW, P. W. KNAP and R. BEILHARZ, 1999. Book of abstracts of the 48th EAAP, p.40. Wageningen Pers, Wageningen, The Netherlands.

MACKAY, T. F. C., 1998. The nature of quantitative genetic variation: lessons from Drosophila. Proc. 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia. 26: 69-80.

MACKAY, T. F. C., R. F. Fry, R. F. LYMAN, and S. V. NUZHDIN, 1994. Polygenic mutation in Drosophila Melanogaster. Genetics 111: 351-374.

MEUWISSEN, T.H.E. and J. A. WOOLLIAMS, 1993. Responses of multitrait selection in open nucleus schemes for dairy cattle breeding. Anim. Prod. 56: 293-299.

MEUWISSEN, T. H. E. and J. A. WOOLLIAMS, 1994. Effective sizes of livestock populations to prevent a decline in fitness. Theor. Appl. Genet. 89: 1019-1026.

MEUWISSEN, T. H. E., 1997a. Maximizing response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75: 934-940.

MEUWISSEN, T. H. E., 1997b. Selection schemes for small populations. Paper presented at the 58th Annual Meeting of the EAAP, Session G2-2, Vienna, Austria.

MEUWISSEN, T.H.E. and A. K. SONESSON, 1998. Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations. J. Anim. Sci. 76: 2575-2583.

NICHOLAS, F. W. and C. SMITH, 1983. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. Anim. Prod. 36: 341-353.

NOMURA, T., 1996. Effective size of selected populations with overlapping generations. J. Anim. Breed. Genet. 113: 1-16.

PONG-WONG, R. and J. A. WOOLLIAMS, 1998. Response to mass selection when an identified major gene is segregating. Genet. Sel. Evol. 30: 313-317.

PRINS, A., 1999. Optimalisatie van het nucleus programma gericht op de internationale markt. Msc. thesis, Animal Breeding and Genetics Group, Wageningen University, Wageningen, The Netherlands.

RAUW, W. M., E. KANIS, E. N. NOORDHUIZEN-STASSEN and F.J. GROMMERS, 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livest. Prod. Sci. 56: 15-33.

ROBERTSON, A., 1961. Inbreeding in artificial selection programmes. Genet. Res. 2: 189-194.

ROBERTSON, A., 1967. The nature of quantitative genetic variation. In: BRINK, A. (Ed.), Heritage from Mendel. Univ. Wisc. Press, Madison, pp. 265-280.

SANTIAGO, E. and A. CABALLERO, 1995. Effective size of populations under selection. Genetics 139: 1013-1030.

SANTIAGO, E. and CABALLERO, A. 1998. Effective size and polymorphism of linked neutral loci in populations under directional selection. Genetics 149: 2105-2117.

SMITH, C., 1978. The effect of inflation and form of investment on the estimated value of genetic improvement in farm livestock. Anim. Prod. 26: 101-110.

SONESSON, A. K. and T. H. E. MEUWISSEN, 2000. Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genet. Sel. Evol. in press.

TORO, M. L., L. SILIÓ, J. RODRIGÁÑES and C. RODRIGUEZ, 1998. The use of molecular markers in conservation programmes of live animals. Genet. Sel. Evol. 30: 585-600.

VERRIER, E., J. J. COLLEAU and J. L. FOULLEY, 1993. Long-term effects of selection based on the animal model BLUP in a finite population. Theor. Appl. Genet. 87: 446-454.

VILLANUEVA, B. and B. W. KENNEDY, 1990. Effect of selection on genetic parameters of correlated traits. Theor. Appl. Genet. 80: 746-752.

VILLANUEVA, B. and J. A. WOOLLIAMS, 1997. Optimization of breeding programmes under index selection and constrained inbreeding. Genet. Res. Camb. 69: 145-158.

VILLANUEVA, B., N. R. WRAY and R. THOMPSON, 1993. Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. Anim. Prod. 57: 1-13.

VILLANUEVA, B., J. A. WOOLLIAMS, and G. SIMM, 1994. Strategies for controlling rates of inbreeding in MOET nucleus schemes for beef cattle. Genet. Sel. Evol. 26: 517-535.

VILLANUEVA, B., R. PONG-WONG, B. GRUNDY and J. A. WOOLLIAMS, 1999. Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genet. Sel. Evol. 31: 115-133.

WANG, J. and W. G. HILL, 2000. Marker assisted selection to increase effective population size by reducing Mendelian Segregation Variance. Genetics 154: 475-489.

WEI, M., A. CABALLERO and W. G. HILL, 1996. Selection response in finite populations. Genetics 144: 1959-1973.

WICKHAM, B. W. and G. BANOS, 1998. Impact of international evaluations on dairy cattle breeding programmes. Proc. 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia. 23: 315-322.

WIENER, G., G. J. LEE and J. A. WOOLLIAMS, 1992. Effects of rapid inbreeding and crossing of inbred lines on conception rate, prolificacy and ewe survival in sheep. Anim. Prod. 55: 115-121.

WOOLLIAMS, J. A., 1989. Modifications to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. Anim. Prod. 49: 1-14.

WOOLLIAMS, J. A. and R. THOMPSON, 1994. A theory of genetic contributions. In: Proceedings of the 5th World Congres on Genetics Applied to Livestock Production, vol. 19, Guelph, Canada, pp. 127-134.

WOOLLIAMS, J. A. and I. WILMUT, 1989. Embryo manipulation in dairy cattle breeding and production. Anim. Prod. 48: 3-30.

WOOLLIAMS, J. A. and I. WILMUT, 1999. New advances in cloning and their potential impact on genetic variation in livestock. Anim. Sci. 68: 245-256.

WOOLLIAMS, J. A., N. R. WRAY and R. THOMPSON, 1993. Prediction of long-term contributions and inbreeding in

populations undergoing mass selection. Genet. Res. Camb. 62: 231-242.

WRAY, N. R. and R. THOMPSON, 1990. Prediction of rates of inbreeding in selected populations. Genet. Res. Camb. 55: 41-54.

WRAY, N. R. and M. E. GODDARD, 1994. Increasing long-term response to selection. Genet. Sel. Evol. 26: 431-451.

WRAY, N. R., J. A. WOOLLIAMS and R. THOMPSON, 1994. Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. 87: 878-892.

WRIGHT, S., 1922. Coefficients of inbreeding and relationship. Am. Nat. 56: 330-339.

WRIGHT, S., 1931. Evolution in Mendelian populations. Genetics 16: 97-159.

WRIGTH, S., 1938. Size of population and breeding structure in relation to evolution. Science 87: 430-431.

WRIGHT, S., 1939. Statistical genetics in relation to evolution. Exposès de Biomètrie et de Statistique Biologique. Herman&Cie, Paris, France.

Summary

This thesis focuses on the prediction of long-term genetic contributions, rates of inbreeding and rates of genetic gain, in populations undergoing selection. The long-term genetic contribution (r_i) of ancestor i born at time t_1, is defined as the proportion of genes from i that are present in individuals in generation t_2 deriving by descent from i, where $(t_2 - t_1) \rightarrow \infty$. In other words, the long-term contribution of an individual is its proportional contribution to the genetic make-up of the population in the long term.

This thesis can be divided into three main parts. First, CHAPTERS 2 to 4 deal with the prediction of long-term contributions and their relation to genetic gain and generation interval. Second, CHAPTERS 5 to 8 deal with the prediction of rates of inbreeding using long-term contributions. Finally, CHAPTERS 9 and 10 deal with the application of the theory to Combined Crossbred Purebred Selection.

CHAPTER 2 to 4: CHAPTER 2 formally shows that the rate of genetic gain per generation is related to the cross-product of long-term contributions and Mendelian sampling terms, $\Delta G = \Sigma r_i a_i$, where a_i is the Mendelian sampling contribution to the genotype of individual i and the sum is taken over all individuals in a generation.

Furthermore, CHAPTER 2 develops a general theory to predict long-term contributions. Long-term contributions were predicted by linear regression of contributions on selective advantages, $E[r_{i,k}|\mathbf{s}_{i,k}] = u_{i,k} = \alpha_k + \boldsymbol{\beta}_k^T \mathbf{s}_{i,k}$, where $r_{i,k}$ is the genetic contribution of individual i in category k and $\mathbf{s}_{i,k}$ is a vector of selective advantages of individual i in category k. With discrete generations, $\alpha_k = 1/(2N_k)$, where N_k is the number of parents of sex k. If $\mathbf{s}_{i,k}$ consists of a single element, then $\beta_k = \alpha\lambda(1-\pi)^{-1}$, where λ is the regression coefficient of the number of selected offspring on $s_{i,k}$, and π is the regression coefficient of $s_{j,l}$ of a selected offspring on $s_{i,k}$ of its parent. With overlapping generations, α_k and $\boldsymbol{\beta}_k$ were predicted using a modified gene flow approach.

CHAPTER 3 studies methods to predict long-term contributions, in populations with overlapping generations undergoing mass selection or selection on a sib index. Results were compared to classical gene flow theory. Due to selection, offspring of younger parents had an above average breeding value. Long-term contributions of the youngest age classes, therefore, were higher than expected from the proportional contribution of those age classes to the group of parents. Generation interval was defined as the time in which long-term contributions sum to unity, and was generally shorter than the average age of parents when their offspring are born.

CHAPTER 4 deals with the relation between gene flow and the rate of genetic gain, and shows that the relation between asymptotic proportions of genes and rates of genetic gain that is embodied in conventional gene flow

theory is invalid. It is shown that conventional gene flow theory considers the gain from a single cycle of selection arising from all previous cohorts, whereas the equation $\Delta G = \Sigma r_i a_i$ considers the gain arising from a single cohort over all subsequent cycles of selection. Both methods yield valid and similar predictions of the rate of genetic gain.

CHAPTER 5 to 8: CHAPTER 5 deals with the relationship between long-term contributions and rates of inbreeding. First, the relation $\Delta F = \frac{1}{4}(1-\omega)\sum r_i^2$, where ω is the deviation from Hardy-Weinberg equilibrium, has been derived directly from identity by descent, whereas previous derivations were based on properties of the relationship matrix. Second, CHAPTER 5 shows that rates of inbreeding may be predicted from *expected* long-term contributions, which makes a separate prediction of the variance of long-term contributions redundant. The prediction equation was: $E[\Delta F] = \frac{1}{2}\sum_k n_k E[u_{i,k}^2] + \frac{1}{8}\sum_k n_k \delta_k$, where $u_{i,k}$ is the expected long-term contribution of individual i in category k, n_k is the number of parents selected from category k and δ_k is a correction term for deviations of the variance of family size from a Poisson variance. Using this equation, tractable predictions of ΔF in selected populations were developed. Finally, CHAPTER 5 shows an application of the theory to populations with discrete generations selected on a sib-index.

CHAPTER 6 develops prediction equations for rates of inbreeding in populations with either discrete or overlapping generations undergoing mass selection, following the general theory presented in CHAPTERS 2 and 5. In the simplest case, with discrete generations and an equal number of sires and dams, the rate of inbreeding equals $E[\Delta F] = \frac{1}{2N} + \frac{1}{N}\left[\frac{\iota^2 h^2(1-\kappa h^2)(1-2/N)}{(1+\kappa h^2)^2}\right] - \frac{1}{8T}$, where N is the total number of parents, ι is selection intensity, κ is Pearson's variance reduction coefficient, h^2 is heritability and T is the total number of selection candidates per sex. Furthermore, CHAPTER 6 shows that, with mass selection and discrete generations and for the special case of overlapping generations with an equal number of parents selected from each age class, the present prediction equation is identical to previously derived equations which were based on the variance of gene frequency. Accurate predictions of ΔF were obtained both for populations with discrete generations as well as for populations with overlapping generations. The rate of inbreeding reached a maximum when an equal number of parents was selected from each age class, and a minimum when all parents except one were selected from the oldest age class.

CHAPTER 7 shows how rates of inbreeding may be predicted for populations that are selected on Best Linear Unbiased Prediction (BLUP) of breeding values. Results showed that, with BLUP selection, an increase of selection intensity dramatically increased the rate of inbreeding.

Finally, CHAPTER 8 shows how rates of inbreeding may be predicted for typical livestock breeding populations with overlapping

generations, BLUP selection and progeny testing. Predictions were considerably more accurate than predictions from other available methods that ignore the effect of selection on ΔF. Predictions were accurate for schemes with up to 20 sires. Predicted ΔF was too low for schemes with more than 20 sires, which was due to the use of simple linear models to predict genetic contributions.

CHAPTERS 9 and 10: CHAPTERS 9 and 10 deal with Combined Crossbred Purebred Selection (CCPS) in crossbreeding schemes. In CHAPTER 9, a pseudo-BLUP selection index is developed to predict response to multi-trait BLUP selection for crossbred performance. Correction of selection intensity for correlated index values of relatives proved to be essential to obtain the correct ranking of breeding schemes. Utilizing crossbred information increased selection response, in particular for low values of the purebred-crossbred genetic correlation.

CHAPTER 10 describes the optimization of CCPS breeding schemes when the rate of inbreeding is restricted, and shows how the theory developed in CHAPTERS 2 to 8 can be used to balance rates of genetic gain and inbreeding for animal breeding schemes in a computationally feasible manner. Results indicate that changing from pure-line selection to CCPS, while keeping the number of parents constant, may substantially increase the rate of inbreeding. When compared at the same rate of inbreeding, CCPS was superior to pure-line selection and the superiority was only slightly reduced compared to a situation without a restriction on the rate of inbreeding.

GENERAL DISCUSSION: The GENERAL DISCUSSION addresses the relevance of this thesis for quantitative genetic theory and for applied animal breeding, and discusses the relevance of the rate of inbreeding for future animal breeding programs.

Samenvatting

Dit hoofdstuk bevat een gepopulariseerde samenvatting van het proefschrift, en is met name gericht op personen die niet bekend zijn met het vakgebied Veefokkerij. Een wetenschappelijk verslag kunt u vinden in de Engelstalige samenvatting.

Inleiding

Het vakgebied Veefokkerij richt zich op het verbeteren van de genetische aanleg van landbouwhuisdieren, opdat de toekomstige generaties van landbouwhuisdieren de gewenste producten op een efficiëntere manier voortbrengen. Met de term "fokkerij" wordt dus niet "vermeerderen", maar "veredelen" bedoeld. Het verbeteren van de genetische aanleg vindt plaats door het selecteren van de beste dieren uit de huidige generatie, en deze vervolgens als ouders te gebruiken voor de volgende generatie.

Veefokkerij rust op twee pijlers. De eerste is het schatten van de genetische aanleg van dieren. Dit wordt de fokwaardeschatting genoemd. Omdat we de genetische aanleg van een dier niet direct kunnen meten, schatten we de fokwaarde aan de hand van de uiterlijke kenmerken van een dier en van zijn bloedverwanten. In het eenvoudigste geval wordt alleen gebruik gemaakt van het kenmerk gemeten aan het dier zelf. In dat geval heeft het dier met de hoogste waarde voor het kenmerk ook de hoogste geschatte fokwaarde. Bijvoorbeeld, de koe die de meeste melk geeft heeft in dat geval ook de hoogste geschatte fokwaarde. In de veefokkerij wordt tegenwoordig algemeen gebruik gemaakt van een geavanceerde methode voor fokwaardeschatting, waarbij informatie van alle bloedverwanten in de fokwaardeschatting betrokken wordt. Dit wordt de BLUP-procedure genoemd.

De tweede pijler onder de veefokkerij is het fokprogramma. Het fokprogramma is de structuur of de organisatie waarbinnen de genetische verbetering plaatsvindt. Belangrijke karakteristieken van een fokprogramma zijn: de kenmerken die worden gemeten en de keuze aan welke dieren wordt gemeten, het totaal aantal selectiekandidaten, het aantal ouders dat iedere generatie wordt geselecteerd, de leeftijd waarop de dieren worden geselecteerd, *etc.*

In de afgelopen 50 jaar zijn zowel de methoden voor fokwaardeschatting als de fokprogramma's sterk verbeterd, hetgeen heeft geleid tot een sterke toename van de genetische vooruitgang. Tegelijkertijd heeft dit echter ook geleid tot een afname van de genetische diversiteit in populaties van landbouwhuisdieren. De afname van de genetische diversiteit in fokprogramma's wordt gemeten aan de hand van de inteelt toename. Inteelt is synoniem met het paren van bloedverwanten. De paring van een broer met zijn zus bijvoorbeeld is een vorm van inteelt. We zeggen dan dat de nakomeling uit een dergelijke paring is ingeteeld.

Behalve door het opzettelijk paren van verwanten ontstaat inteelt ook als een populatie klein is, omdat we in dat geval niet kunnen voorkomen dat verwanten met elkaar gepaard

worden. Ter illustratie twee extreme voorbeelden. Als een populatie uit slechts twee dieren bestaat, een mannetje en een vrouwtje, dan zijn alle nakomelingen die hieruit geboren worden volle broers en zussen. In de volgende generatie moeten we dus een broer met zijn zus paren. Als de populatie uit vier dieren bestaat, kunnen we in de eerste generatie inteelt voorkomen door partners te selecteren die uit twee verschillende ouderparen komen. In de tweede generatie moeten we echter een paring maken van een neef met zijn volle nicht. Ook bij grotere populaties treedt dit effect op, maar dan zijn de paringen tussen verwanten van een kleinere verwantschapsgraad.

Inteelt heeft overwegend negatieve gevolgen. Ten eerste veroorzaakt inteelt een afname van de genetische diversiteit, waardoor het in de toekomst steeds moeilijker wordt om genetische vooruitgang te boeken. Ten tweede veroorzaakt inteelt een afname van de fitness van dieren, hetgeen zich met name uit in een groter aantal vruchtbaarheidsproblemen. Tenslotte leidt inteelt tot het vaker voorkomen van erfelijke ziekten in de populatie. Gezien de nadelige gevolgen van inteelt is het dus van belang om inteelt zoveel mogelijk te beperken. Het is belangrijk om te beseffen dat inteelt nooit helemaal voorkomen kan worden, omdat elke populatie een eindige omvang heeft. Uiteindelijk zullen er altijd verwanten gepaard moeten worden, zodat er altijd sprake is van inteelttoename.

Dit proefschrift

In dit proefschrift is gekeken naar de gevolgen van verschillende selectiestrategieën op de inteelttoename in populaties van landbouwhuisdieren. Er zijn methoden ontwikkeld om de verwachte inteelttoename van een populatie vooraf te voorspellen, zodat een fokprogramma gekozen kan worden dat een acceptabele inteelttoename oplevert.

Het centrale concept in dit proefschrift is de "genetische bijdrage". De genetische bijdrage van een individu is het aandeel van de genen in de populatie dat van dit individu afkomstig is. Bijvoorbeeld, in een populatie die uit 50 vaderdieren en 50 moederdieren bestaat, is de gemiddelde genetische bijdrage van een vaderdier ½×1/50 = 0,01. Dat wil zeggen, vaders en moeders dragen elk de helft van de genen bij en binnen de vaders draagt ieder dier gemiddeld 1/50 bij. De werkelijke genetische bijdrage van een dier zal afwijken van 0,01 omdat het ene vaderdier meer nakomelingen krijgt dan het andere. De gemiddelde bijdrage zal dus 0,01 bedragen, maar de werkelijke bijdrage zal variëren tussen de dieren.

In de hoofdstukken 2 tot en met 4 zijn methoden ontwikkeld om de genetische bijdragen van individuen te voorspellen. Bij het voorspellen van genetische bijdragen in populaties van landbouwhuisdieren moet rekening worden gehouden met het fokprogramma. In het algemeen zullen dieren die een hogere fokwaarde hebben ook een hogere genetische bijdrage krijgen. Bijvoorbeeld, als er in een fokprogramma geselecteerd wordt voor een zo hoog mogelijke melkproductie, zullen we van een koe met een hoge melkproductie meer nakomelingen selecteren dan van een gemiddelde koe. Daardoor levert een dergelijke koe een

bovengemiddelde genetische bijdrage aan de populatie. Genetische bijdragen verschillen dus systematisch tussen dieren met een hoge en dieren met een lage fokwaarde.

Er zijn twee mechanismen waardoor de genetische bijdrage van een dier beïnvloed wordt. Ten eerste worden van een dier met een hogere fokwaarde gemiddeld meer nakomelingen geselecteerd. Ten tweede hebben nakomelingen van ouders met een hoge fokwaarde zelf ook weer een bovengemiddelde fokwaarde. In de hoofdstukken 2 tot en met 4 zijn methoden ontwikkeld om genetische bijdragen te voorspellen door middel van het modelleren van deze beide mechanismen. Tevens laat hoofstuk 2 zien dat de genetische vooruitgang per generatie gelijk is aan de afwijking van de fokwaarde van een dier van het gemiddelde van zijn ouders, vermenigvuldigd met de genetische bijdrage van dit dier en vervolgens gesommeerd over alle dieren in de betreffende generatie.

De hoofdstukken 5 tot en met 8 behandelen de relatie tussen de genetische bijdrage en de inteelttoename. Hoofdstuk 5 laat zien dat de inteelttoename per generatie gelijk is aan de helft van de som van de gekwadrateerde voorspelde genetische bijdragen per generatie. Bijvoorbeeld, als in de bovengenoemde populatie met 50 vaderdieren en 50 moederdieren elk dier een voorspelde genetische bijdrage heeft van 0,01, dan bedraagt de inteelttoename $\frac{1}{2}\times100\times0{,}01^2$ = 0,005. Dit betekent dat er elke generatie 0,5% van de genetische diversiteit verloren gaat. In een geselecteerde populatie zullen dieren met een hogere fokwaarde een hogere voorspelde genetische bijdrage hebben. Bijvoorbeeld, als de 10 beste vaderdieren en de 10 beste moederdieren elk een voorspelde genetische bijdrage hebben van 0,02 en de rest van de dieren heeft een voorspelde bijdrage van 0,0075, dan bedraagt de inteelttoename $\frac{1}{2}\times(20\times0{,}02^2 + 80\times0{,}0075^2)$ = 0,00625, zodat er iedere generatie 0,625% van de genetische diversiteit verloren gaat. Dit eenvoudige voorbeeld laat zien op welke manier selectie leidt tot een verhoging van de inteelttoename.

In de hoofdstukken 6 tot en met 8 zijn vervolgens formules afgeleid voor het voorspellen van de inteelttoename in verschillende fokprogramma's. De belangrijkste conclusie van deze hoofdstukken is dat, in een populatie waarbij dieren geselecteerd worden voor een bepaald kenmerk, de inteelttoename veel hoger is dan in een populatie waarbij willekeurige dieren geselecteerd worden als ouders voor de volgende generatie. Vooral het gebruik van geavanceerde methoden voor fokwaardeschatting, zoals de bovengenoemde BLUP-procedure, en het selecteren van dieren op basis van deze geschatte fokwaarden, leidt tot een forse verhoging van de inteelttoename. Voor de veefokkerij betekent dit dat er, bij toepassing van deze geavanceerde methoden, maatregelen genomen moeten worden om de inteelttoename te beperken.

De hoofdstukken 9 en 10 behandelen een speciale toepassing van de methode om inteelt te voorspellen bij populaties waarin gebruik wordt gemaakt van kruising. In fokprogramma's voor varkens en kippen wordt geselecteerd in zogenaamde zuivere lijnen. Dit zijn lijnen die uit één ras bestaan. De uiteindelijke

productiedieren, bijvoorbeeld slachtvarkens, ontstaan uit het kruisen van een aantal van deze zuivere lijnen. Hoofdstuk 9 laat zien dat het meten van kenmerken aan deze gekruiste productiedieren een bijdrage kan leveren aan de genetische vooruitgang van het fokprogramma in de zuivere lijn. Vervolgens laat hoofdstuk 10 zien dat het benutten van de informatie van gekruiste productiedieren niet alleen leidt tot een toename van de genetische vooruitgang, maar tevens kan leiden tot een forse verhoging van de inteelttoename. Door het verhogen van het aantal geselecteerde vaderdieren kan de inteelt echter tot aanvaardbare proporties worden teruggebracht en is de genetische vooruitgang nog steeds hoger dan in een situatie waarin geen gebruik wordt gemaakt van informatie afkomstig van gekruiste dieren.

In de algemene discussie wordt ingegaan op het belang van het huidige proefschrift voor de kwantitatief-genetische theorie en voor de praktische veefokkerij. De belangrijkste bijdrage aan de kwantitatief-genetische theorie is de ontwikkeling van methoden voor het voorspellen van genetische bijdragen en inteelt in geselecteerde populaties. In de praktische veefokkerij kunnen deze methoden gebruikt worden om een balans te vinden tussen genetische vooruitgang en inteelttoename. Tot voor kort was dat alleen mogelijk met behulp van tijdrovende computersimulaties.

Curriculum Vitae

Piter Bijma werd op 13 juni 1970 geboren te Hijum. In 1986 behaalde hij het MAVO diploma aan de Tsjerkwalles MAVO te Leeuwarden. In 1989 behaalde hij het MTS diploma werktuigbouwkunde aan de MTS te Leeuwarden, en in 1993 het HTS diploma werktuigbouwkunde aan de Noordelijke Hogeschool te Leeuwarden. In hetzelfde jaar begon hij met de studie Zootechniek aan de Landbouwuniversiteit te Wageningen. In 1996 studeerde hij met lof af, met als oriëntatie Veefokkerij. In hetzelfde jaar werd hij aangesteld als Assistent in Opleiding (AIO) bij de leerstoelgroep Fokkerij en Genetica van Wageningen Universiteit, alwaar hij het in dit proefschrift beschreven onderzoek verrichtte. Sinds 1 mei 2000 is hij werkzaam als postdoc bij bovengenoemde leerstoelgroep.