Lecture notes for the SelAction workshop

February 19 and 20, 2002
Wageningen
The Netherlands

Piter Bijma
Marc Rutten
Animal Breeding and Genetics Group
Wageningen University

WAGENINGEN UNIVERSITY
ANIMAL SCIENCES

OPTIMIZATION OF BREEDING PROGRAMS

SelAction is a tool for strategic optimization of animal breeding programs. The term "strategic" indicates that SelAction can be used to optimize the design of the breeding program, it is not a decision support tool for day-to-day selection decisions. Optimization of the design requires that alternative breeding schemes can be compared and evaluated. Choosing the best breeding scheme among a number of alternatives requires yardsticks to measure the quality of breeding schemes. Such yardsticks can be developed only when there is a well-defined breeding goal. Given that the breeding goal is clearly defined, there are three criteria that summarize the quality of a breeding program. These are:

1. Selection response for the breeding goal.
2. Maintenance of genetic diversity as measured by the rate of inbreeding.
3. Costs of the breeding program.

Selection response for the breeding goal traits is the revenue of a breeding program, whereas loss of genetic diversity and financial costs are the expenses of a breeding program. Selection response, loss of genetic diversity and financial costs are expressed in different units. The problem therefore is to combine them into a single criterion for the quality of a breeding program.

A comparison of breeding schemes based on selection response and the rate of inbreeding can be done as follows. To avoid long-term loss of genetic diversity and to avoid inbreeding depression on health and fitness traits, an upper limit can be set to the rate of inbreeding. Next, alternative breeding schemes can be judged by comparing their selection response at the same rate of inbreeding. The scheme with the highest selection response at the same rate of inbreeding (e.g. 1%/generation) is the best scheme.

It is more difficult to combine selection response and cost into a single criterion. The question is whether the revenues of an increase of selection response, for example in the form of increased market share, makes up for the cost of increased selection response. Hence, this is not a genetic issue but primarily a commercial and marketing issue.

To compare selection response and inbreeding among alternative breeding schemes, expected response and inbreeding of those alternatives needs to be known, i.e. prediction of response and inbreeding is required. The following summarizes the theory used in SelAction to predict selection response and inbreeding.

PREDICTION OF SELECTION RESPONSE

The basic equation: The basic equation for prediction selection response is:

$\Delta G = i\ \rho\ \sigma_A$

where ΔG is selection response, i is the intensity of selection, ρ is the accuracy of selection and σ_A is the additive genetic standard deviation of the trait that is to be improved.

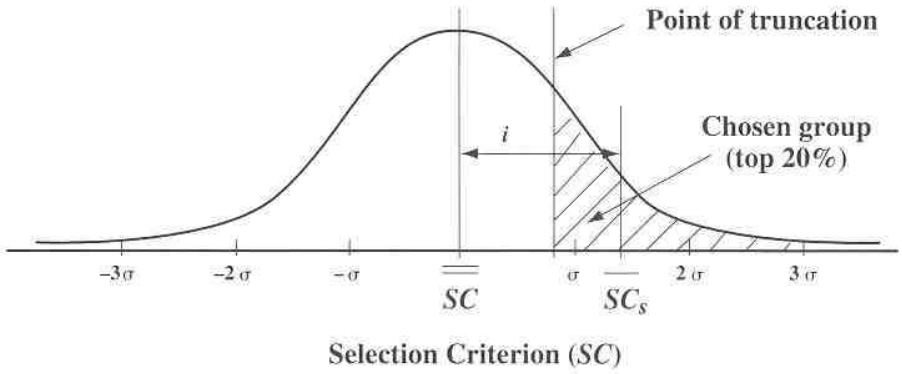

Figure 8.2 – Selection intensity with truncation selection. Saving the top 20% (p = 0.2) of selection candidates results in a selection intensity (i) of 1.4 standard deviation units. (Taken from Bourdon, 2000)

The intensity of selection is the average value of the selection criterion (e.g. index) of the selected parents expressed as a deviation from the population mean and expressed on a standardized scale. In most cases we assume that the selection criterion is normally distributed in which case i can be obtained directly from the proportion of individuals that is selected as parents (selected proportion). Tables of the selection intensity as a function of the selected proportion can e.g. be found in the book of Falconer.

The accuracy of selection is the *correlation* between the selection criterion and the true breeding value for the trait that is to be improved. The squared accuracy is the proportion of the additive genetic variance that is explained by the selection criterion. The accuracy is bounded between 0 and 1.

The additive genetic standard deviation measures the genetic variability of the trait that is to be improved. Thus in the above equation, σ_A is a kind of scaling factor, whereas i and ρ are on a standardized scale.

More complex situations: Though $\Delta G = i\ \rho\ \sigma_A$ is the basic equation, it is obviously too simple for practical breeding schemes. It only applied directly to a population of a single sex with discrete generations. With separate sexes, the equation becomes

$$\Delta G = \tfrac{1}{2} i_m \rho_m \sigma_A + \tfrac{1}{2} i_f \rho_f \sigma_A$$

which is the average of the response due to selection of males (m) and females (f).

Overlapping generations: With overlapping generations, there are multiple age classes and selection response depends on the combined effect of selection in all age classes and on the generation interval. Selection response for overlapping generations is given by

$$\Delta G = \frac{\overline{R}}{\overline{L}}$$

where $\overline{R}$ is the genetic selection differential averaged over age-classes and sexes and $\overline{L}$ is the generation interval averaged over age-classes and sexes. $\overline{R}$ is given by

$$\overline{R} = \tfrac{1}{2} \sum_{x=m,\,f} \sum_{k=1}^{n} i_{x,k} \rho_{x,k} \sigma_A \frac{n_{x,k}}{N_x}$$

where x is the sex and k is the age class. The ½ averages the males and females, $n_{x,k}$ is the number of parents of sex x selected from age class k, N_x is the total number of selected male parents (x=m) or female parents (x=f), thus $n_{x,k}/N_k$ averages the selection differential over age classes within sexes.

The average generation interval is given by

$$\bar{L} = \frac{1}{2} \sum_{x=m,f} \sum_{k=1}^{n} age(k) \frac{n_{x,k}}{N_x}$$

Where age(k) is the age of age-class k at the time when the offspring of selected parents are born. For example, when dams are selected at three years of age, but gestation time is a year, then age(k) is 4 years.

Multitrait selection: With multitrait selection knowledge of the breeding goal is required to predict response to selection. The breeding goal (H) defines the emphasis given to each trait, H is a weighted sum of true breeding values and (economic) weighting factors:

$$H = A_1v_1 + A_2v_2 + \ldots..$$

where A are true breeding values and v are weighting factors for each trait. Because true breeding values cannot be observed H is unknown. The selection criterion is therefore an index of known phenotypes or EBV, constructed so that the expected value of the index is equal to H. Thus an index is used to predict H. If the index is a weighted sum of phenotypes of selection candidates and relatives of selection candidates for multiple traits,

$$I = b_1P_1 + b_2P_2 + ..$$

then the optimal index weights are given by

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gv}$$

where **b** is a vector of index weights, **P** is the variance-covariance matrix of index information sources, **G** is a matrix of covariances between information sources and true breeding values of selection candidates and **v** is a vector of (economic) weights of the breeding goal H. However, in practice the index will probably be a weighted sum of BLUP-EBV of selection candidates:

$$I = b_1EBV_1 + b_2EBV_2 + ..$$

where EBV are the estimated breeding values for the breeding goal traits. When the index is a weighted sum of EBV's and EBV's are estimated using multitrait BLUP, then the optimum index weights are equal to the (economic) weighting factors in the breeding goal:

$$I = v_1EBV_1 + v_2EBV_2 + ..$$

So that the complexity of $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gv}$ can be avoided. Thus when using the BLUP-option in SelAction, the corresponding situation in practice is simply an index that is a weighted sum of EBV and (economic) breeding goal weights.

Selection response with multitrait selection for an index is given by:

$$\Delta\mathbf{G} = \mathbf{b'G}i/\sigma_I$$

where $\Delta\mathbf{G}$ is a vector containing selection response for each trait, **b** and **G** are as above, i is selection intensity and σ_I is the standard deviation of the index. The total gain in the breeding goal (e.g. in monetary units) is given by

$$\Delta H = i\ \sigma_I$$

When selection is on an index of estimated breeding values, $I = v_1EBV_1 + v_2EBV_2 + $, then the selection response in each trait is simply equal to the EBV-selection differential for that trait,

$$\mathbf{\Delta G} = \mathbf{\Delta EBV}$$

SelAction uses $\mathbf{\Delta G} = \mathbf{b'G}i/\sigma_I$. For BLUP-selection an index is used that includes EBV and information sources (see below) which gives the same result as $\mathbf{\Delta G} = \mathbf{\Delta EBV}$.

The above section presents the equations for ΔG. The next step is to derive the elements of those equations, in particular, **P**, **G**, i and σ_I. In principle, those elements can be derived using standard selection index theory. However, there are a number of complicating factors. Those are:

i). Reduction of the genetic variance due to selection, the so-called Bulmer effect. This introduces the problem of predicting the equilibrium value of the variances and covariances.

ii). The effect of correlated EBV of selection candidates on the selection intensity. When EBV are (partly) based on family information, EBV of sibs are similar. Selection therefore is not only between individuals but also partly between families. Since the number of families is generally smaller than the number of individuals, the selection intensity will be reduced when EBV of sibs are similar

iii). The inclusion of full pedigree information on a BLUP context. BLUP breeding value estimation takes account of the full pedigree. In a selection index approach, this situation could be mimicked by including all phenotypic information of ancestors, all the way back to the base generation. However, this approach is impractical because the number of information sources and the dimension of the **P**-matrix would become huge. Thus the pedigree information should ideally be condensed into a single value.

Points *i* to *iii* are addressed below.

i) Reduction of the genetic variance due to selection: The additive genetic variance is the variance of the true breeding values. In a selected population, the parents are a selected group with similar breeding values. Thus the variance of the breeding values of parents is smaller than that of unselected individuals. Because offspring inherit half of the breeding value of their parents, a reduction in the variance of breeding values of parents will also lead to a reduction in the variance of breeding values of offspring.

The variance reduction can be derived as follows. The breeding value of an individual is

$$A = ½A_s + ½A_d + A_{MS}$$

where A_s is the breeding value of the sire, A_d is the breeding value of the dam and A_{MS} is contribution to the breeding value that originates from Mendelian sampling. (A_{MS} causes that full sibs are not genetically identical). To obtain the variances, the ½'s have to be squared, giving

$$\sigma_A^2 = \tfrac{1}{4}\sigma_{A,S}^2 + \tfrac{1}{4}\sigma_{A,S}^2 + \sigma_{MS}^2$$

where $\sigma^2_{A,s}$ is the additive genetic variance of the selected sires, $\sigma^2_{A,d}$ is the additive genetic variance of the selected dams, and σ^2_{MS} is the additive genetic variance due to Mendelian sampling. Due to selection $\sigma^2_{A,s}$ and $\sigma^2_{A,d}$ are smaller than σ^2_A. The Mendelian sampling variance is equal to half of the base generation additive genetic variance, $\sigma^2_{MS} = \frac{1}{2}\sigma^2_{A,0}$. Thus there are to opposite effects, selection of sires and dams has the effect of reducing the genetic variance, but Mendelian sampling has the effect of restoring it. There will be a value where both effects balance each other; this value is the equilibrium genetic variance. The equilibrium genetic variance is generally reached within very few generations, approx. 3.

When selection is for a single trait, the equilibrium variance is approx. 25% lower than the base generation variance, but this depends on the intensity and accuracy of selection. For selection on a single trait, the equilibrium variance can be calculated in an iterative manner, using

$$\sigma^2_{A,s} = \sigma^2_A(1 - k_s\rho_s^2)$$

$$\sigma^2_{A,d} = \sigma^2_A(1 - k_d\rho_d^2)$$

where k_s and k_d are the variance reduction coefficients which take values ranging from 0 to 1, but are close to 0.8 in most cases, and ρ_s and ρ_d are the accuracies of selection for sires and dams. The variance reduction coefficients are a direct function of the selected proportion, $k = i(i - x)$ where i is selection intensity and x is the standardized truncation point (see e.g. book of Falconer). For example, when the selected proportion $p = 0.1$, then $i = 1.755$ and $x = 1.282$ so that $k = 0.83$. The iterative scheme is as follows: *i*) from the base generation genetic variance, $\sigma^2_{A,s}$ and $\sigma^2_{A,d}$ are calculated using the above equations, *ii*) σ^2_A is calculated as $\sigma^2_A = \frac{1}{4}\sigma^2_{A,s} + \frac{1}{4}\sigma^2_{A,s} + \sigma^2_{MS}$ with $\sigma^2_{MS} = \frac{1}{2}\sigma^2_{A,0}$, *iii*) the accuracy is recalculated, $\sigma^2_{A,s}$ and $\sigma^2_{A,d}$ are recalculated using the above equations and step *ii* is executed again. It takes approx. 3 rounds to reach the equilibrium value.

For multitrait selection the equations are a multitrait analogy of the above equations. The basic equation is

$$Cov(A_i, A_j)_{sires} = Cov(A_i, A_j) - \frac{Cov(A_i, I)Cov(A_j, I)k_s}{\sigma^2_I}$$

and for dams k_d should be used instead of k_s. It is beyond the scope of these notes to include all equations for multitrait selection.

Once equilibrium variances are obtained by iteration, the equilibrium values can be used to predict response to selection. Thus for predicting the equilibrium response, the equilibrium values need to be used in the above equations for ΔG.

Summarizing the Bulmer effect:

i) Selection causes a reduction of the genetic variance. An equilibrium value is reached in approx. 3 generations.

ii) Variance reduction is often close to 25%, i.e. ~75% of the initial variance is remaining. Variance reduction is largest with high selection intensity and high accuracy.

iii) When own performance information is important, then reduction of selection response is similar to the reduction of the variance. With selection based on progeny testing response reduction is closer to the square root of the variance reduction.

Accounting for the Bulmer effect is important for two reasons. First, it reduces the absolute response by approx 25%. This is important when comparing costs of programs with selection response. Second, accounting for the Bulmer effect may change the relative ranking of alternative breeding programs. The Bulmer effect reduces the between family variance. As a consequence, full and half-sib information is becoming less important, i.e. sib information explains a smaller proportion of the total additive genetic variance. The Bulmer effect does not reduce the Mendelian sampling variance. Thus information that includes the Mendelian sampling term, such as own performance and progeny information, becomes relatively more important compared to family information. Those effects are important when comparing breeding schemes that rely on sib information with breeding schemes that rely on own performance or progeny information, e.g. MOET dairy cattle breeding programs vs. classical progeny testing programs.

ii) Correlated index values of selection candidates: When EBV are (partly) based on family information, EBV of sibs are similar. Selection therefore is not only between individuals but also partly between families. Since the number of families is generally smaller than the number of individuals, the selection intensity will be reduced when EBV of sibs are similar. In other words, the selection intensity will be reduced when the EBV of selection candidates are correlated.

There is a second effect that becomes important in small breeding schemes with few selection candidates. If the number of selection candidates is small, selection intensity is smaller than indicated by the standard tables. This can be understood as follows: compare selection of the 2 best individuals out of 10 candidates with selection of the 4 best individuals out of 20 candidates. In both cases p = 0.2, but the top 4 out of 20 is expected to be better than the top 2 out of 10. If two sets of 10 would be joined, then the best 2 of each sub set are not necessarily the best 4 out of the total. Maybe the best 3 come from only one of the subsets. This effect is only important if the number of candidates is really small, but SelAction accounts for it.

SelAction calculates the selection intensity as follows (Meuwissen, 1991):

$$i = i_{r,0}(1 - \rho_{avg,sibs})^{\upsilon_{\rho sibs}}$$

where $i_{r,0}$ is the selection intensity accounting for finite population size but ignoring correlated EBV, $\rho_{avg,sibs}$ is the average correlation between EBV of sibs and $\upsilon_{\rho sibs}$ is a power-term depending on the correlation between sibs. Detailed equations are in Meuwissen (1991). This correction is particularly important in breeding schemes that rely heavily on information coming from full and half sibs and where the number of selected parents is small.

iii) The inclusion of full pedigree in a BLUP-context: In modern livestock improvement, breeding value estimation is performed using BLUP. BLUP accounts for the full pedigree of selection candidates. To predict response to selection on BLUP-EBV using selection index, the selection index equations have to include the pedigree information. In principle this could be done by including phenotypes of all ancestors in the selection index, but this would lead to a very large number of information sources in the selection index equations. In addition, the contribution of distant ancestors would be very small.

An elegant solution was proposed by Wray and Hill (1989). The trick is to include estimated breeding values of parents in the selection index of their offspring. With BLUP, the EBV of an individual includes information of all generations up to the present generation. Thus the EBV compresses all pedigree information into a single figure. Therefore a selection index may be composed of the EBV of the parent generation and in addition the phenotypic information of the current generation. Such an index is called a pseudo-BLUP index. Thus the information sources are:

1. phenotypic own performance (P_i)
2. phenotypic information of full sibs (P_{FS})
3. phenotypic information of half sibs (P_{hs})
4. phenotypic information of progeny testing (P_{prog})
5. estimated breeding value of the sire (EBV_s)
6. estimated breeding value of the dam (EBV_d)
7. average estimated breeding values of the dams of the half sibs ($EBV_{hs\text{-}dams}$)

Obviously, information sources 1-4 are only included if those phenotypes are available. Thus e.g. in the absence of progeny testing, information source 4 is not included. Information sources 5 and 6 represent the compressed pedigree information of the parents. Information source 7 may need a bit extra explanation because the dams of the half-sibs are not related to the selection candidate at all. Thus why include their EBV? The reason is that including the EBV of the dams of the half sibs corrects the half-sib performance that is not due to the sire. Because the half sibs are assumed to be paternal they provide information on the breeding value of the sire of the selection candidate. However, not the entire half-sib performance is due to the sire, it is also partly due to the dam and the environment. Nothing is known about the environment, but true-BLUP accounts for the dams of the half sibs. For example, if half sibs have an excellent performance because they had a good dam, true BLUP corrects for this via the relationship matrix. In a pseudo-BLUP selection index the same can be achieved by including the dams of the half sibs as information sources.

Wray and Hill (1989) show by stochastic simulation that the pseudo-BLUP index gives practically identical response to selection as true BLUP. Thus a prediction of selection response using pseudo-BLUP is accurate.

The pseudo-BLUP index can be extended directly to multi-trait selection. In that case, the index contains estimated breeding values of parents for all traits. Thus, for each trait there is

a 7×7 submatrix and the total P-matrix has dimensions of n_{traits} × 7. Mathematical details of a multitrait pseudo-BLUP index can be found in Villanueva *et al*. (1993).

Truncation selection with overlapping generations: Pseudo-BLUP can directly be extended to populations with overlapping generations. With overlapping generations there is a separate index for each age-class because the amount of information differs between age classes. In general, young age classes have less information than older age classes. Because older age classes have more information, they have higher accuracy and the variance of the EBV of older age classes is generally larger than that of young age classes, i.e. $\sigma^2_{ebv} = \rho^2\sigma^2_A$. However, when there is continuous genetic improvement the mean level of older age classes will be lower than that of younger age classes, simply because old animals are born in an earlier generation that has a lower genetic level. Thus the average level of selection candidates in age class *k* is $\overline{EBV}_k = \mu - age(k)\Delta G$. When the aim is to select the individuals with the highest EBV, truncation selection across age classes can be performed. There are two parameters that determine the number of animal selected from each age class; $\overline{EBV}_k$ which is lower for older age classes and σ^2_{ebv} which is higher for older age classes. It is difficult to determine the truncation point analytically. Numerical algorithms are therefore used to determine the truncation point so that the number of selection candidates with EBV above the truncation point is equal to the desired number of selected parents. In SelAction you will notice that when the difference in accuracy is small most parents will be selected from the younger age classes, whereas when differences in accuracy are large (e.g. progeny tested bulls in dairy cattle programs) a larger number may be selected from older age classes. Mathematical details of truncation selection across generations can be found in Ducrocq and Quaas (1988) and Bijma *et al*. (2001) appendix A.

Multistage selection

Multistage selection is a selection strategy where candidates are selected in multiple stages and individuals that are not selected in a particular stage are not candidates for the next stage. For example, in a pig-breeding program the aim is to select 10 parents out of 100 candidates but there are only 50 test stations for feed intake. The individuals that are tested for feed intake may be preselected based on their juvenile growth. Thus there are two stages of selection. The first stage with p_1 = 0.5 based on full pedigree, sib information and own performance for juvenile growth, and a second stage with p_2 = 0.2 based additionally on own performance and sib information for feed intake. Thus the overall selected proportion is p = $p_1 \times p_2$.

From a methodological point of view, prediction of response to multistage selection has two problems, *i*) selection in the first step reduces the variance of traits, which has to be accounted for in the second stage; *ii*) selection in the first stage causes that traits are no longer normally distributed, so that the selection intensity in the second stage cannot simply be taken from the common tables. The first problem can be accounted for by calculating the

reduced variance in a way similar to the calculation of the Bulmer effect. The second problem is more serious and a more statistical approach is needed to solve it.

From a statistical point of view, the index in the first and second stage and the true breeding value follow a multivariate normal distribution (before selection). The genetic selection differential after the second stage is therefore equal to the expectation of the true breeding value given that index1 and index2 are greater than the respective truncation points, R = $E(BV|I_1 > t_1 , I_2 > t_2)$. For each trait the genetic selection differential can be calculated by regression on the index selection differentials,

$$\Delta G = b_1 E(I_1 \mid I_1 > t_1, I_2 > t_2) + b_2 E(I_2 \mid I_1 > t_1, I_2 > t_2) + b_3 .....$$

The regression coefficients b_1, b_2, etc, can be calculated using standard regression theory, but the index selection differential requires to calculate the multivariate normal integrals:

$$E(I_1 \mid I_1 > t_1, I_2 > t_2) = \frac{1}{p} \int_{t_1}^{\infty} \int_{t_2}^{\infty} I_1 f(I_1, I_2) dI_1 dI_2 ,$$

$$E(I_2 \mid I_1 > t_1, I_2 > t_2) = \frac{1}{p} \int_{t_1}^{\infty} \int_{t_2}^{\infty} I_2 f(I_1, I_2) dI_1 dI_2 ,$$

where $f(I_1,I_2)$ is the bivariate normal density of index 1 and 2. Those integrals cannot be calculated analytically; SelAction uses numerical integration algorithms to calculate the index selection differentials. Unfortunately, numerical integration requires a bit of computing time. Further mathematical details can be found in Ducrocq and Colleau (1989).

The general picture with multistage selection is as follows. If the indices in the different stages show a high correlation, then substantial preselection can be applied with minor loss of selection response. The reason is that the majority of candidates that are culled based on index 1 would also have been culled if selection were in a single stage based on index 2. Therefore, if the indices show a high correlation, preselection may be used to reduce the cost of the breeding program. In terms of the above example on feed intake, if the indices in stage 1 and 2 have a high correlation then it will be possible to reduce the number of feeding stations with minor reductions of response to selection. In contrast, if the indices have a low correlation, preselection may result in culling of individuals that would have been selected in the second stage. Thus if the indices have low correlation one needs to be careful with preselection (Saxton, 1983).

In SelAction it is assumed that information available in a certain stage is also available in the next stage. Thus stage 2 cannot have less information than stage 1; information can only be added. It does not make sense to ignore information because including all information gives the highest accuracy. Because information can only be added, the correlation between indices in different stages can be calculated from the accuracies of those stages:

$$\rho_{I1,I2} = \rho_1 / \rho_2$$
$$\rho_{I2,I3} = \rho_2 / \rho_3$$
$$\rho_{I1,I3} = \rho_1 / \rho_3$$

where ρ_i is the accuracy in stage i. Thus the accuracies can be used to calculate the correlations between indices in the different stages.

DRIFT, INBREEDING AND EFFECTIVE POPULATION SIZE

Drift is a process causing a *random* change of allele frequency over time and is caused by random sampling of the alleles transmitted to the next generation. There are two sampling processes that cause drift, sampling of alleles between individuals and Mendelian sampling of alleles within individuals. Sampling between individuals is due to selection of parents; alleles of individuals that are selected as parents are transmitted to the next generation, whereas alleles of non-parents are not. Mendelian sampling refers to the random transmission of one of the two parental alleles to the offspring.

Inbreeding at the individual level is the result of deliberate mating of related individuals, for example mating of two full sibs. Inbreeding at the individual level therefore is often called "deliberate inbreeding", meaning that the inbred matings are created on purpose. In contrast to deliberate inbreeding, inbreeding at the population level is caused by the number of parents being finite and is unavoidable.

To understand inbreeding at the population level, consider a population consisting of N parents (ignoring the distinction between males and females). Suppose that all alleles in the original population are unique so that there are $2N$ distinct alleles, 2 alleles for each of the N parents. The inbreeding level in the next generation is the probability that both alleles at an arbitrary locus in an individual are identical by descent. This probability is equal to $1/(2N)$, because when drawing the second allele of an offspring the probability that it is identical to the first allele is $1/(2N)$ since there are $2N$ distinct alleles. Thus in the next generation the inbreeding level is $1/(2N)$. The general relationship between inbreeding and the number of parents is more complicated but the principle remains the same.

Inbreeding at the population level cannot be avoided. Every population has a steady increase of the inbreeding level because the number of parents is finite. Thus in every population there will come a moment when we have to mate related individuals, simply because there are no unrelated individuals available anymore. To understand that inbreeding cannot be avoided, consider the following argument (Falconer, 1989). "*In a population with two sexes, every individual has two parents, four grandparents, eight great-grandparents, etc., and t generations back it has 2^t ancestors. Not very many generations back the number of individuals required to provide separate ancestors for all the present individuals becomes larger than any real population could contain. Any pair of individuals must therefore be related to each other through one or more common ancestors in the more or less remote past, and the smaller the size of the population in the previous generations the less remote are the common ancestors.*"

The magnitude of inbreeding on the population level is measured by the **rate of inbreeding**, ΔF. The rate of inbreeding is defined as:

$$\Delta F = \frac{\overline{F}_t - \overline{F}_{t-1}}{1 - \overline{F}_{t-1}} \qquad (10.2)$$

where *t* is the generation number and $\overline{F}_t$ is the average inbreeding level of generation *t*. Drift and inbreeding are closely related phenomena. In fact, drift is the cause of inbreeding; the random change of allele frequency due to drift results in inbreeding in later generations

Only in the absence of selection the rate of inbreeding is related directly to the number of sires and dams,

$$\Delta F = \frac{1}{8N_m} + \frac{1}{8N_f} \tag{10.3}$$

where N_m and N_f are the number of male and female parents per generation. The "effective population size" is a parameter that expresses the rate of inbreeding on a scale that is related to the number of parents. Effective population size is defined as:

$$N_E = \frac{1}{2\Delta F} \tag{10.6}$$

It is important to understand that the effective population size is not related directly to the actual number of parents, but that it is merely a way to express the rate of inbreeding on a different scale. The effective population size can be very different from the actual number of parents. The rate of inbreeding in the global dairy cattle population is approximately 1% per generation. The effective size of the global Holstein-Friesian dairy cattle population is therefore equal to $N_E = 1/(2\Delta F) = 1/(2\times0.01) = 50$ individuals. Thus though there are millions of diary cows, effectively the population consists of only ~50 individuals. The global Holstein-Friesian population clearly shows that a large actual number of animals by no means guarantees a large effective number of animals. In other words, populations consisting of many animals may still have high rates of inbreeding. The reason is that, by using artificial insemination, a very small number of sires may contribute the majority of offspring to the next and later generation, whereas other sires contribute hardly any offspring to the next generation.

Predicting the rate of inbreeding in selected populations: *In selected populations the equation $\Delta F = 1/(8N_m) + 1/(8N_f)$ is invalid.* The reason is that with selection, parents contribute unequally to the next generation and also to further generations. For example, a superior sire will have more offspring selected and those offspring will be better than average. Therefore, the sire will also have more selected grand-offspring etc. This process continues until the breeding value of the sire is sufficiently diluted, which takes ~2-5 generations, depending on the selection strategy. In a selected population, therefore, there will be a systematic difference between the contribution of good *vs*. poor individuals. After a number of generations a superior individual will show up frequently in the pedigree of all individuals, whereas a poor individual will hardly show up in the pedigree. This difference in the contribution of ancestors has a booster effect on the rate of inbreeding. Therefore, in a selected population the rate of inbreeding can be much larger than $1/(8N_m) + 1/(8N_f)$.

The rate of inbreeding is directly related to the sum of squared contributions of parents (Wray and Thompson, 1990),

$$\Delta F = \tfrac{1}{4}\sum r^2$$

where ΔF is the rate of inbreeding per generation, r is the (long-term) contribution of a particular parent and the Σ indicates summation across all parents in this generation. The long-term contribution is the proportional contribution of an individual to the genes in the population and can be calculated from the pedigree, but we will not go into that here. Consider an example with 20 selected parents per generation (ignoring that there are two sexes for simplicity). By analyzing the pedigree we can quantify the contribution of each of those 20 parents of a particular generation. Their contribution will sum to 1; genetic contributions always sum to 1 per generation. Consider two extreme cases, one where the contribution of each individual is the same, r = 0.05 for all individuals, and one where contributions differ extremely between individuals, r = 0.25 for the 4 best parents and r=0 for the rest. In the first case, the rate of inbreeding is $\frac{1}{4}(0.05^2 + 0.05^2 + .. 0.05^2) = 0.0125$ = 1.25% per generation. This is the lowest possible rate of inbreeding with 20 parents. In the second case the rate of inbreeding is $\frac{1}{4}(0.25^2 + 0.25^2 + .. 0^2) = 0.0625$ = 6.25% per generation. This example illustrates that variance in the contributions of ancestors causes higher inbreeding.

Sel*A*ction can predict the rate of inbreeding for populations with discrete undergoing selection. The prediction is based on the relationship between the rate of inbreeding and long-term contributions. The rate of inbreeding is predicted by predicting the long-term contribution, taking account of the effect of selection on the contribution.

In short, the procedure is as follows. First, a regression model is used to predict the long-term contribution,

$$E(r) = \alpha + \beta(BV - \overline{BV})$$

In this model E(r) is the expected contribution given that we know the true breeding value of an individual, α is the contribution of an individual with an average breeding value and β is a term that accounts for the increase of the contribution of parents with a higher breeding value. Thus the second term takes account of the fact that parents with high breeding values will have more selected offspring. Both α and β can be derived mathematically, but that is beyond the scope of this text. The next step is to calculate the square of the expected contributions, by squaring the above equation

$$E(r)^2 = \alpha^2 + \beta^2\sigma_A^2(1 - k\rho^2)$$

the right-most term is the genetic variance of the selected parents (see also the section about the Bulmer effect). So far we have calculated E(r)^2, but in fact we need to calculate E(r^2). The difference is that E(r)^2 is the square of the expected contributions whereas E(r^2) is the expectation of the squared (actual) contribution. It can be derived that under certain conditions E(r^2) = 2E(r)^2 , leading to the result that the rate of inbreeding is

$$\Delta F = \tfrac{1}{2} NE(r)^2$$

where N is the number of parents and $E(r)^2$ is the square of the expected contributions. Note that the ¼ is replaced by ½ because we have replaced the square of the actual contributions Σr^2 by the square of the expected contributions, $NE(r)^2$.

The above theory shows that selection strategies that increase the variance of contributions among parents lead to higher rates of inbreeding. Selection strategies that rely heavily on family information are an example. If estimated breeding values are largely based on family information, truncation selection on EBV will lead to between family selection. In that case parents of successful families will have a large contribution whereas parents of non-successful families will have no contribution at all, which increases the rate of inbreeding substantially.

BLUP automatically places large emphasis on family information in cases where individual or progeny information is lacking. Examples are MOET breeding schemes in dairy cattle where EBV are largely based on performance of female full-sibs; selection of boars for litter size in dam lines of pig-breeding programs, selection of roosters for egg number based on female full sibs etc. However, the fact that BLUP puts large emphasis on family information does not mean that BLUP breeding value estimation should be avoided. BLUP yields the most accurate method for breeding value estimation and properly corrects for fixed effects. Therefore, BLUP is the method of choice for breeding value estimation. If EBVs are largely based on family information one needs to include a restriction on the rate of inbreeding in the selection step, EBVs can still be based on BLUP. An excellent method to include a restriction on the rate of inbreeding in the selection step is developed by Meuwissen (1997). In order to maximize the genetic selection differential but restrict the rate of inbreeding, Meuwissen's method calculates the optimum number of offspring of each parent. At present this method is the method of choice to maximize response to selection while restricting the rate of inbreeding.

References

Bijma, P., J. A. M. VanArendonk and J. A. Woolliams (2001). Predicting rates of inbreeding in livestock breeding populations. J. Anim. Sci. **79**: 840-853.

Ducrocq, V. and R. L. Quaas (1988). “Prediction of genetic response to truncation selection across generations.” J. Dairy Sci. **71**(9): 2543-2553.

Ducrocq, V. and J. J. Colleau (1989). “Optimum truncation points for independent culling level selection on a multivariate normal distribution, with application to dairy cattle selection.” Genet. Sel. Evol. **21**: 185-198.

Falconer, D. S., 1989. Introduction to quantitative genetics. Longman, UK.

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.” J. Anim. Sci. **75**: 934-940.

Saxton, A. M. (1983). “A comparison of exact and sequential methods in multistage selection.” Theor. Appl. Genet. **66**: 23-28.

Villanueva, B., N. R. Wray, et al. (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.

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.