

2 Genetic change of multiple traits

Julius van der Werf

In this lecture we consider the genetic change of multiple traits in more detail.

- How to derive selection criteria for multiple traits
 - How to predict multiple trait selection response
 - How to manipulate multiple trait selection response
 - Considerations in order to make optimal genetic change for multiple traits
-

2.1 Introduction

Improvement of efficiency and quality of animal production can be achieved by improving several characteristics simultaneously. For example, we may want to breed the ideal dairy cow that produces not only a lot of milk, but also with a high percentage of components like fat and especially protein. In addition, such high producing animals should preferably have no fertility problems, not suffer from diseases, and have a high roughage intake capacity. We would like pigs that grow fast but do not become too fat, with no leg problems and producing many piglets per litter. In beef cattle we want fast growing animals, but we don't want a too large mature size. We would like merino sheep with very fine wool, yet yielding a lot of it.

Questions to be addressed:

- Is it possible to change animals in any direction we want, and, can different characters be improved simultaneously?
- How does change for each trait depend on economic weights, on correlations and heritabilities
- how can we predict response to multiple trait selection?
- how can we manipulate change in the construction of selection criteria (indexes)?
- What is the most desirable change and how can we achieve this change?

Genetic change takes place through selection. Selection criteria can be derived from information on animals' phenotypes and their relatives. Phenotypic information is generally based on objective measurement of animals' characteristics, but it can also include subjective scoring. Even genotype information about DNA-markers can be put in this category as it contributes to predicting an animals breeding value. Figure 2.1 puts this all in perspective. Information on phenotypes combined with pedigree information allows estimation of BLUP EBV on all animals for all traits. Breeding values can be combined in selection criteria (often a '*selection index*') by using the economic weights. Then selection decisions are based on selection criteria, but not necessarily by truncation selection on index values, as we want to keep some genetic diversity. That's why pedigree information comes in again.

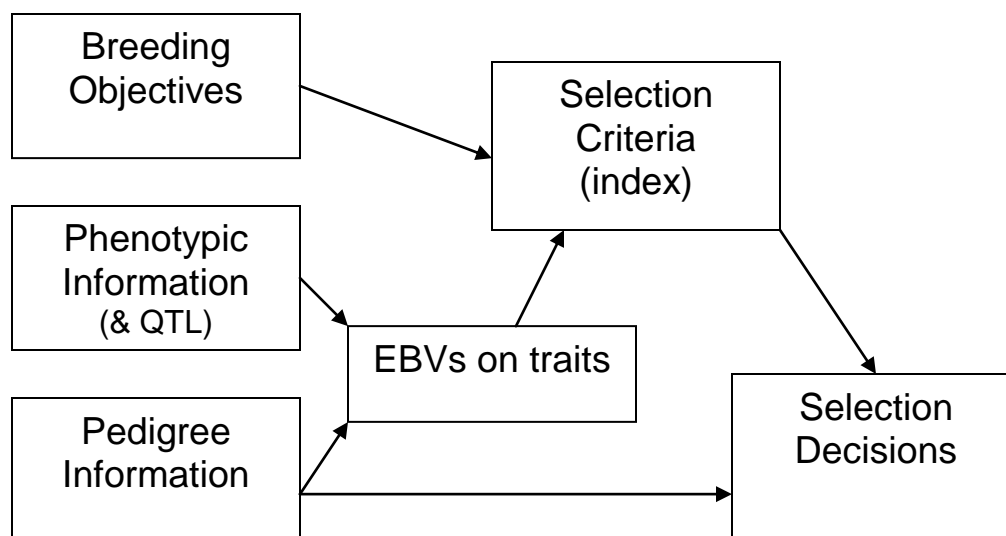
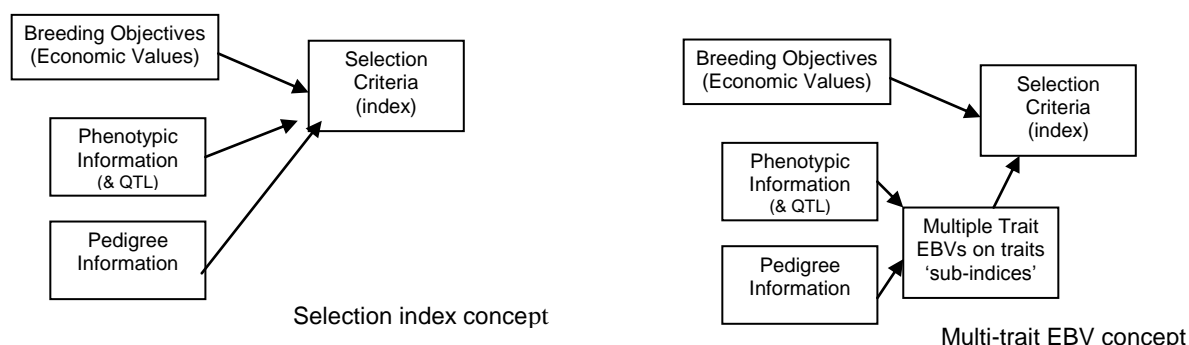


Figure 2.1 Relationship between information, pedigree, index and

Multiple trait selection index theory deals with combining information on phenotypes and pedigree as well as economic weights into one index value, which is the best estimate of the aggregate genotype. However, it is also convenient to see this as a two step process, where first we derive and EBV for each trait separately, but using information on all traits. Subsequently these multi trait EBV are combined into an index. It will be shown that the weights in that case are simply the economic values as defined in the breeding objective.



When explaining multiple trait selection to the industry, the Multi-trait EBV concept is much easier to grasp and relates to the general practice of genetic evaluation systems presenting EBVs for each trait. For example, an index might to select on body weight and feed intake might look like

$$\text{Index} = v_1 \text{EBV}_{\text{BW}} + v_2 \text{EBV}_{\text{FI}}$$

However, with this approach it is more difficult to predict response, or even to see why certain traits would be more difficult to change. A trait such as feed intake is difficult to measure. It might have a high economic value, but with a lack of phenotypic information, it would be changed only slowly. The fact is that such traits will have lowly accurate EBV, and therefore, the variation in these EBV is small. Such traits will therefore contribute less to the overall index ranking of

individuals. A section on 'properties of EBV' is presented to make you more familiar with some of those basic concepts about EBV.

The selection index approach is useful for more theoretical analyses. It allows predicting overall selection response as well as responses for each of the individual traits. In this Chapter we will therefore in some detail go through selection index theory.

When genetic change is (to be) achieved for a certain trait, it is important to consider possible genetic changes for other traits since traits can be phenotypically or genetically correlated. Some of these changes may be desired. For example, dairy cows selected for milk production capacity are able to eat more roughage. Pigs that grow faster have also a better feed conversion. However, other changes may be undesirable, e.g. fast growing pigs have smaller litter sizes, cows with high milk yield have lower fat content and Merino's with more wool will have higher fibre diameter. Hence, traits correlations can be favourable or unfavourable. We will see that situations with unfavourable correlations are difficult to deal with because 1) it harder to make genetic change to each of the traits and 2) the optimal genetic change is much more sensitive to economic weights that are used. These issues in multiple trait selection will be discussed in this chapter.

From an economic viewpoint it is important to predict how correlated characters change when animals are selected for a certain characteristic. Also from a more biological viewpoint it is important to know how animals change after selection. Did they 'overall' become more efficient? How did the higher production change the animals' physiology? Is the higher production also more efficient from a biological point of view. Are there any side effects from long selection for a certain selection index? Optimal selection strategies need to deal with a reality that often involves non-linear profit functions. Discussion about optimal change often results in a 'desired gains approach' where not (only) the economic weight of traits, but also the resulting response determines the most desirable selection outcome. Such approaches and debates will conclude this chapter and be dealt with further in the next.

2.2 Properties of Estimating Breeding Values

Accuracy of EBV

The accuracy is defined as the correlation between true and estimated breeding value. The symbol for accuracy is r_{IA}

Since the EBV is often indicated as an Index (I), - the true breeding value has symbol A and r is a common symbol for correlation.

The accuracy is between 0 and 1 (or 0% and 100%). In the extreme case of no information, the accuracy of a breeding value is 0, and with a very large amount of information, the accuracy will approach 1. The following Table shows examples of accuracy. It illustrates that

- Accuracy is higher when more information is used, e.g. on relatives and progeny
- The accuracy is higher for traits with a higher heritability, but the effect of heritability becomes smaller with more information used
- The accuracy of parent average depends on the parent EBV accuracy and not on heritability (but note that with low heritability it will be harder for a parent to achieve a certain accuracy)
- The accuracy of information from collateral relatives (i.e. siblings) is limited to 0.5 for HS and 0.71 for FS. A progeny test is required to obtain higher accuracies

Table 2.1. Accuracies of EBV depending on source of information used

Information used	$h^2 = 0.10$	$h^2 = 0.30$
Sire EBV ($r_{IA}=0.5$)	0.25	0.25
Sire EBV ($r_{IA}=0.9$)	0.45	0.45
Sire EBV ($r_{IA}=0.5$) + Dam EBV ($r_{IA}=0.5$)	0.35	0.35
Sire EBV ($r_{IA}=0.9$) + Dam EBV ($r_{IA}=0.5$)	0.51	0.51
Own Performance only	0.32	0.55
OP+ Sire EBV ($r_{IA}=0.9$)+ Dam EBV ($r_{IA}=0.5$)	0.57	0.66
Mean of 5 full sibs	0.32	0.48
Mean of 10 half sibs	0.23	0.33
OP + 5 FS + 10 HS	0.43	0.65
Mean of 1000 half sibs	0.49	0.50
Mean of 1000 full sibs	0.70	0.71
Mean of 20 progeny	0.58	0.79
Mean of 100 progeny	0.85	0.94
Mean of 1000 progeny	0.98	0.99

Accuracies can be derived using selection index theory. Here we only give a simple example for the derivation of accuracy of an EBV based own performance

$$EBV = I = h^2P \quad \text{giving} \quad r_{IA} = r_{h2P,A}$$

$$r_{h2P,A} = \frac{Cov(h^2P, A)}{\sqrt{V(h^2P)VA}} = \frac{h^2V_A}{\sqrt{h^4V_PV_A}} = \frac{h^2V_A}{\sqrt{h^2V_AV_A}} = \frac{h^2}{\sqrt{h^2}} = h$$

If the heritability is higher, EBV's based on own performance records become more accurate.

Variance among EBV

The variance among EBVs is of practical value because

- it can give us an indication of the difference in EBV between the highest and lowest animals
- It is used to predict selection differential, e.g. the average EBV of the best 10% of animals
- A trait will be more impacted by selection on an multiple trait index if that index trait has more variation.

Example: Assume EBV's of three rams

case A		carcass weight (kg)	IMF (%)
	Ram A	+10	+2
	Ram B	+5	-1
	Ram C	-2	-4
Case B		carcass weight (kg)	IMF (%)
	Ram A	+10	+0.5
	Ram B	+5	-0.25
	Ram C	-2	-1

For a given set of economic weights, IMF will be less impacted by index selection in the second case

In general, the variation among EBV can be predicted from accuracy and genetic variance

$$\text{Var}(\text{EBV}) = r_{IA}^2 V_A \quad \text{and} \quad \sigma_{\text{EBV}} = r_{IA} \sigma_A$$

where r_{IA} is the accuracy of the EBV. Hence, the variance of the EBV's is equal to the accuracy-squared multiplied by the variance of the true breeding values (additive genetic variance).

It is useful to consider the following

- If $r_{IA} = 0$ then $\text{var}(\text{EBV}) = 0$: all EBVs have the same value (=0)
- If $r_{IA} = 1$ then $\text{var}(\text{EBV}) = 1$: the variance of EBV is equal to the variance of breeding values. All EBV should be equal to the true BV with this accuracy, and there is no prediction error.

$\text{Var}(\text{EBV})$ is generally smaller than V_A

$\text{Var}(\text{EBV})$ becomes larger when accuracy is higher. i.e. the EBV of older animals will be more apart than those of you animals. The same holds for EBV of intensely measured nucleus animals compared to the EBV of base animals that have less information and therefore EBVs closer to each other.

Example: Single trait/own performance case:

$$\text{Var}(\text{EBV}) = \text{Var}(h^2 P) = h^4 V_P = h^2 V_A$$

$$[\dots \text{ as } h^2 = V_A / V_P \quad V_P = V_A / h^2]$$

2.3 Selection Index (single trait EBV)

We can write the EBV as an index, weighing different types of information.

$$\text{EBV} = \text{Index} = b_1X + b_2X_2 + \dots + b_nX_n$$

where b_1, b_2, \dots, b_n are index weights and X_1, X_2, \dots, X_n are phenotypic information sources. For example, X_i could refer to own performance, performance of the sire, or the mean performance of a number of siblings or progeny. P can also refer to an EBV, e.g. the EBV of the sire.

There can be many different information sources, e.g. own performance, (mean) performance of full sib(s), (mean) performance of half sib(s), performance of sire or dam, (mean) performance of progeny. Even more distant relatives provide information (e.g. performance of an aunt), and in principle, different animals may often have different sets of information, and therefore different sets of weighting factors (remember that the weight for a particular information source depends on all other sources). In practice, this would be quite cumbersome to compute, and distant relatives are often ignored in selection index. In fact, selection index weights generally don't have to be calculated since they are automatically derived in the BLUP procedure for estimating breeding values (which we will see later on). Deriving selection index weights has therefore become more a theoretical tool to determine relative importance of different information sources, and accuracy of EBV, rather than a practical tool to obtain EBV's (indexes) of actual animals.

The optimal weights in the index to combine information from different sources are obtained by multiple regression: each weight indicates what proportion of variation in information sources predicts breeding value.

Working out the weights in a selection index requires quantitative genetic theory and some algebra, which can in fact become quite tedious.

Generally, regression of y on x is worked out as

$$b_{xy} = \text{cov}(x,y)/\text{var}(x)$$

In our case we regress breeding value (A) on phenotypic information

$$b_{x,A} = \text{cov}(X_i,A)/\text{var}(X_i)$$

Therefore we need

- the covariance of each information source with the breeding value: $\text{cov}(X_i, A)$
-
- the variance of each information source: $\text{var}(X_i)$

but also

- the covariance between these information sources; $\text{cov}(X_i, X_j)$

because the information sources may not be independent. We have to account for the relationship between phenotypic sources, otherwise we tend to 'double count' the same information. Therefore, when an EBV is predicted from more than one information source, the regression coefficients are predicted from multiple regression and we need matrices to work this out.

Matrices denoting variances and covariances are used to calculate selection index weights. The index weights are written in a vector. The index is a scalar value, being a product of two vectors:

the vector b containing index weights: $b' = [b_1, b_2, \dots, b_n]$

and vector X containing information sources: $X' = [X_1, X_2, \dots, X_n]$

$$\text{EBV} = I = b_1X + b_2X_2 + \dots + b_nX_n$$

$$= b'X$$

The multiple regression coefficients are calculated by regression

$$b = \text{Cov}(X, A) / \text{var}(X)$$

$$b = P^{-1}G$$

This is the basic formula for (single trait) selection index.

- matrix **P** = var(**X**) is a matrix denoting the variance (and covariances) of information sources (not to be confused with the symbol P that we used for an individual observation)
- matrix **G** = cov(**X**, A) is a vector denoting the covariance between each information source and the breeding value (not to be confused with the symbol G that we used for genetic value)

The accuracy of the Index (i.e. of the EBV) is calculated as the correlation between EBV and A:

$$r_{IA} = \text{cov}(I, A) / \sqrt{\text{var}(I), \text{var}(A)}$$

$$\text{cov}(I, A) = \text{cov}(b'X, A) = b' \text{cov}(X, A) = b'G$$

$$\text{var}(I) = \text{var}(b'X) = b' \text{var}(X)b = b'Pb$$

Note that since $b = P^{-1}G$; $Pb = G$ and therefore $b'G = b'Pb$, in other words, the covariance between I and A is equal to the variance of I: $\text{cov}(I, A) = \text{var}(I)$

$$\text{hence: } r_{IA} = \text{cov}(I, A) / \sqrt{\text{var}(I), \text{var}(A)}$$

$$= \text{var}(I) / \sqrt{\text{var}(I), \text{var}(A)}$$

$$= \sqrt{\text{var}(I)} / \sqrt{\text{var}(A)}$$

$$= \sigma_I / \sigma_A$$

i.e. the accuracy is calculated as the SD of the index divided by the genetic SD

This is a proof of what we had before: $\text{var}(\text{EBV}) = \sigma_{\text{EBV}} = r_{IA} \cdot \sigma_A$

Summarizing again the main steps of single trait selection index:

Set up variance-covariance matrices P and G

Calculate Weights: $b = P^{-1}G$

Variance of index $\text{Var}(I) = \sigma_I^2 = b'Pb$

And accuracy $r_{IA} = \sigma_I / \sigma_A$

The main challenge of selection index theory is to work out P and G matrices. Quantitative Genetics theory is needed here and some basic rules will follow. Note that we will now use the term σ^2 for variance, i.e. σ^2_P rather than V_P and σ^2_A rather than V_A

Summarizing the rules for setting up selection index matrices (single trait):

For the P-matrix:

$$\text{Var}(X_i) = \sigma^2_P$$

for the variance of any single measurement

$$\text{Var}(\bar{X}_i) = r \cdot \sigma^2_P + (1-r)/n \cdot \sigma^2_P$$

for the variance of a mean of n records

where r is the intra-class correlation between the measurements in the group, e.g. r = repeatability for the mean of repeated measurements on the same animal, and $r = \frac{1}{4} h^2$ for a half sib or progeny mean

$$\text{Cov}(X_i, X_j) = a_{ij} \cdot \sigma^2_A$$

for the covariance between two single measurements on 2 different individuals, a_{ij} is the additive genetic relationship between these individuals. In some cases the individuals could share a common environment, in which case we add a term $c^2 \cdot \sigma^2_P$, where c^2 is the proportion of variation due to common environment.

$$\text{Cov}(X_i, \bar{X}_j) = a_{ij} \cdot \sigma^2_A$$

for the covariance between a single measurements and the mean of measurements, a_{ij} is the additive genetic relationship between the one individual and each of the members that make up the mean.

In working out selection index equations it is convenient to take a mean of a group of individuals, which is allowed if all members of that group have the same relationship with each other, as well as the same relationship with the animal we calculate an EBV for. Hence, the group needs to be 'homogeneous'.

And for the G-vector

$$\text{Cov}(X_i, A) = a_{ij} \cdot \sigma^2_A$$

for the covariance between a single measurements and the breeding value of the EBV-animal.

$$\text{Cov}(\bar{X}_i, A) = a_{ij} \cdot \sigma^2_A$$

for the covariance between a a mean of measurements and the breeding value of the EBV-animal, a_{ij} is the additive genetic relationship between the EBV-animal and each of the members that make up the mean.

Selection Index Examples (single trait)

Example 1:

Information sources: X_1 = own performance
 X_2 = performance of sire

- variance and covariance of information sources:

$$\text{var} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = P = \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{pmatrix}$$

The individual elements of the matrices can be worked out using quantitative genetic theory:

$\text{var}(X_1) = \sigma_P^2$ is the phenotypic variance

$\text{var}(X_2) = \sigma_P^2$ is the phenotypic variance

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \text{Cov}(A+E, A_s + E_s) \\ &= \text{Cov}(A, A_s) + \text{cov}(A, E_s) + \text{cov}(E, A_s) + \text{cov}(E, E_s) \\ &= \frac{1}{2}\sigma_A^2 + 0 + 0 + 0. \end{aligned}$$

- covariance between information sources and the animal's breeding value

$$\text{cov} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, A = G = \begin{pmatrix} \text{cov}(X_1, A) \\ \text{cov}(X_2, A) \end{pmatrix}$$

$$\text{Cov}(X_1, A) = \text{Cov}(A+E, A) = \text{Cov}(A, A) + \text{cov}(E, A) = \sigma_A^2 + 0$$

$$\text{Cov}(X_2, A) = \text{Cov}(A_s + E_s, A) = \text{Cov}(A_s, A) + \text{cov}(E_s, A) = \frac{1}{2}\sigma_A^2 + 0.$$

such that index weights obtained by regression = covariance/variance:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = P^{-1}G = \begin{pmatrix} \sigma_P^2 & \frac{1}{2}\sigma_A^2 \\ \frac{1}{2}\sigma_A^2 & \sigma_P^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_A^2 \\ \frac{1}{2}\sigma_A^2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2}h^2 \\ \frac{1}{2}h^2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} h^2 \\ \frac{1}{2}h^2 \end{pmatrix}$$

We can now plug in any value of h^2 . Table 2.5 gives the index weights and the accuracy of the EBV for different values of heritability.

It shows that

- The weights generally increase for higher heritability,
- The weight for the sire's phenotype is relatively higher for lower heritability.
- The increase of accuracy from using sire's information is relatively larger for lower heritability.

Table 2.5 Index weights for information on phenotype of an individual and its sire, accuracy of

index (EBV) and increase in accuracy from using sire's information in addition to own phenotype for different values of heritability

Heritability	b_1	b_2	Accuracy	% Increase relative to using X_1 only
0.10	0.098	0.045	0.347	9.7
0.30	0.284	0.107	0.581	6.1
0.50	0.467	0.133	0.730	3.3

Example 2:

Information sources: X_1 = own performance
 X_2 = mean performance of n full sibs

- variance and covariance of information sources:

$$\text{var} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = P = \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{pmatrix}$$

$\text{var}(X_1) = \sigma_P^2$ is the phenotypic variance

$\text{var}(X_2) = r\sigma_P^2 + ((1-r)/n)\sigma_P^2$ where $r = \frac{1}{2}h^2 + c^2$ (c^2 is proportion of variance due to their common environment)

$\text{Cov}(X_1, X_2) = r$

- covariance between information sources and the animal's breeding value

$$\text{cov} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, A = G = \begin{pmatrix} \text{cov}(X_1, A) \\ \text{cov}(X_2, A) \end{pmatrix}$$

$\text{Cov}(X_1, A) = \sigma_A^2$

$\text{Cov}(X_2, A) = \frac{1}{2}\sigma_A^2$

such that index weights obtained by regression = covariance/variance:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = P^{-1}G = \begin{pmatrix} \sigma_P^2 & r\sigma_P^2 \\ r\sigma_P^2 & [r + (1-r)/n]\sigma_P^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_A^2 \\ \frac{1}{2}\sigma_A^2 \end{pmatrix} = \begin{pmatrix} 1 & r \\ r & [r + (1-r)/n] \end{pmatrix}^{-1} \begin{pmatrix} h^2 \\ \frac{1}{2}h^2 \end{pmatrix}$$

We can plug in values of h^2 and c^2 . Table 2.6 gives the index weights and the accuracy of the EBV for different parameter values.

Table 2.6. Index weights for information on own phenotype of and the mean of n full sibs, accuracy of index (EBV) for different parameter values.

h^2	$n=3$		$n=10$	
$c^2 = 0$	b_1	b_2	b_1	b_2
0.10	0.09	0.12	0.08	0.32
0.30	0.26	0.26	0.22	0.49
0.50	0.43	0.29	0.38	0.48
0.70	0.62	0.24	0.57	0.36
h^2	$n=3$		$n=10$	
$c^2 = .15$	b_1	b_2	b_1	b_2
0.10	0.09	0.07	0.08	0.13
0.30	0.26	0.14	0.24	0.21
0.50	0.46	0.11	0.43	0.17
0.70	0.70	0.00	0.70	0.00

Results from the table show that own performance is more important with high heritability and small family size whereas otherwise, family information is more important. Environmental covariances among full sibs decrease the value of their information, as full sib records are more alike, but not due to genetics.

For a heritability above 0.7, and $c^2 = 0.15$, the weight on the FS information become even negative, as the FS mean will serve as a correction for common environment

Selection of animals based on a selection index (using relatives' information, as we see later this is the same as BLUP) tends therefore to look more like mass selection for high heritable traits, and more like family selection for low heritable traits. The important consequence is that selection based on BLUP EBV leads to more inbreeding if heritabilities are low, since we tend to select more related animals as parents for the next generation.

Example 3: Progeny Testing

Using information from a group of progeny is a special case because it is potentially the most accurate way to determine an animals' EBV. We already know that the maximum accuracy attained by very many full sibs is 0.71 (which is $\sqrt{1/2}$, since full sibs have at most 50% of the differences in common with an individual). Similarly, the maximum accuracy of very many full sibs is 0.5 (which is $\sqrt{1/4}$). The only way to obtain an accuracy near 100% is to test progeny.

The index weight for a progeny test is again found by regression:

$$EBV_{\text{sire}} = \text{Index} = b_1 P_1$$

P_1 is the mean of n progeny

b_1 is the index weight

More detail on the derivation of b_1 :

b_1 is the regression of A_{sire} on the progeny mean.

Each progeny's performance can be written as

$$P_{\text{ind. progeny}} = \frac{1}{2}A_{\text{sire}} + [\text{other effects}]$$

The other effects include the effect of dam, other genetic effects (due to Mendelian segregation and dominance) and environmental effects. The 'other effects' are different for each progeny of a sire, but their sire effects is common to all.

The mean of all progeny is

$$P_{\text{progeny mean}} = \frac{1}{2}A_{\text{sire}} + [\text{other effects}]/n$$

and the variance of the progeny mean is

$$V_{PM} = \frac{1}{4}V_A + (V_P - \frac{1}{4}V_A) / n$$

and the covariance between the sire's breeding value and the progeny mean is

$$\text{cov}(A_{\text{sire}}, \frac{1}{2}A_{\text{sire}} + [\text{othereffects}] / n) = \text{cov}(A_{\text{sire}}, \frac{1}{2}A_{\text{sire}}) = \frac{1}{2}V_A$$

Thus the regression coefficient to determine a sire's EBV based on its progeny test is

$$b_1 = \frac{\text{Cov}(A, \frac{1}{2}A)}{V_{PM}} = \frac{\frac{1}{2}V_A}{\frac{1}{4}V_A + (V_P - \frac{1}{4}V_A) / n}$$

$$\dots \text{ times } \frac{n}{\frac{1}{4}V_A} \text{ top and bottom gives } b_1 = \frac{2n}{n + a} \quad \text{where} \quad a = \frac{4 - h^2}{h^2}$$

b_1 depends on the number of progeny and the heritability:

$$b_1 = \frac{2n}{n + a} \quad \text{where} \quad a = \frac{4 - h^2}{h^2}$$

note that $0 < b_1 < 2$.

If a sire has a very large number of progeny who, on average perform 10 units better than the mean, we estimate the sire's EBV as 2 times that deviation (EBV = +20). This makes sense, because we know that if a sire has an EBV of +20, we expect his progeny to receive half of that (with average dams).

Since

$$EBV_{\text{sires}} = b_1 P_{PM}$$

and the expected performance of future progeny: $E(P_{\text{future progeny}}) = \frac{1}{2}EBV_{\text{sire}}$

we can predict the performance of future progeny directly from the current progeny mean:

$$E(P_{\text{future progeny}}) = \frac{1}{2} b_1 P_{PM} = \frac{n}{n + a} P_{PM}$$

We could call the term $\frac{1}{2} b_1 = \frac{n}{n + a} = h^2_{PT}$ is the 'heritability' of the progeny test, it determines which part of differences in observed progeny means can be expected back in future progeny.

EXAMPLE - Dairy cattle, mean annual lactation = 5000 Kg.

$h^2=0.25$ giving $a=15$. A bull has 20 daughters with a mean first lactation yield of 5500 Kg.

1. What is the expected yield of subsequent daughters (True Progeny Mean) ?

$$h^2_{pt} = n/(n+a) = 20/(20+15) = 0.57$$

$$\text{Observed Progeny Mean (OPM)} = +500$$

$$\text{Estimate of sire effect on progeny mean} = 0.57 \times 500 = +286 \text{ Kg}$$

$$\text{Expected yield of future progeny} = 5000 + 268 = 5268 \text{ Kg.}$$

2. What is the estimate of the sire's breeding value ?

$$b_{BV, OPM} \times OPM = 2n/(n+a) \times +500 = +572 \text{ Kg (which is twice 286 Kg)}$$

The accuracy of the progeny test

Like in mass selection, the accuracy r_{IA} is also the square root of a “heritability” - the heritability of the progeny test [$h^2_{pt} = n/(n+a)$].

$$\text{i.e. } r_{IA} = \sqrt{\frac{n}{n+a}}$$

This simple formula allows you to determine the accuracy, for a given progeny test based on n progeny, for a trait with heritability h^2 (where $a = (4-h^2)/h^2$).

Example: Accuracy of progeny test

	Nr of progeny			
h^2	5	25	50	100
0.1	.34	.63	.75	.85
0.3	.54	.82	.90	.94
0.5	.65	.88	.99	.995

2.4 Multiple trait selection

Correlated response to selection

Multiple Trait Selection refers to the situation where information from more than one trait is used for selection. When selecting, even if it is on only one characteristic, there is likely to be a response in other traits not considered.

Direct response for trait 1 when selection is practised on trait 1:

$$R = i \cdot h_1 \sigma_{A1}$$

Correlated response for trait 2 when selection is practised on trait 1 is calculated by regression.

$$CR = i \cdot h_1 r_g \sigma_{A2}$$

More detail (for reference)

This can be derived two ways. We can determine the change of the genotypic mean of trait 2 of a selected group of animals that are phenotypically $S (=i \cdot \sigma_P)$ units of trait 1 better than their population mean. We can also determine how many units the offspring will be better for trait 2 if they were genetically improved for trait 1 by R units. Approaches are equivalent, i.e. they lead to the same answer and in both cases we need a regression coefficient. We only work out the first approach.

Animals are selected with phenotypic selection superiority $S_1 = i \cdot \sigma_{P1}$

We are interested in the genetic response of offspring for trait 2.

We need the regression of phenotypic values on a parent for 1 (P_1) on genetic values for offspring on 2 (O_2). This regression can be calculated as $b_{P_1 O_2}$:

$$\begin{aligned} b_{P_1 O_2} &= \text{cov}(P_{P1}, G_{O2}) / \text{var}(P_{P1}) \\ &= \text{cov}(G_{P1} + E_{P1}, G_{O2}) / \text{var}(P_{P1}) \\ &= \text{cov}(G_{P1}, G_{O2}) / \text{var}(P_{P1}) \\ &= \frac{1}{2} r_g \sigma_{A1} \sigma_{A2} / \sigma_{P1}^2 \end{aligned}$$

and the correlated response in the progeny for trait 2. with phenotypic selection of both parents is therefore CR ,

$$\begin{aligned} CR &= 2b_{P_1 O_2} \cdot S_1 = (r_g \cdot \sigma_{A1} \sigma_{A2} / \sigma_{P1}^2) \cdot i \cdot \sigma_{P1} \\ &= i \cdot r_g \cdot \sigma_{A1} \sigma_{A2} / \sigma_{P1} \\ &= i \cdot h_1 r_g \sigma_{A2} \end{aligned}$$

Example:

Data was measured on cattle for weight at weaning (Kg) and for feed intake (Kg/day).

Assume	Genetic correlation:	0.50
	Phenotypic correlation	0.20
	Heritability of Weight	0.40
	Heritability Feed intake	0.25
	Phenotypic SD	17
	Phenotypic SD	2.0

Selection of parents is on individual phenotype and the selected fraction is 38% for both males and females ($i = 1.0$).

Selection on weight:

Direct Response in Weight: $R = i \cdot h^2 \cdot \sigma_p = 1.0 \cdot 0.4 \cdot 17 = 6.80$ Kg.

Correlated Response in Feed Intake

$$CR = i \cdot r_g \cdot h_A \sigma_{gB} = 1.0 \cdot 0.5 \cdot \sqrt{(.4)} \cdot \sqrt{(0.25 \cdot 2)} = 0.32 \text{ Kg.}$$

Selection on Feed Intake:

Direct Response in Feed Intake $R = -1.0 \cdot .25 \cdot 2.0 = -0.50$ Kg

Correlated Response in Weight

$$CR = -1.0 \cdot 0.5 \cdot \sqrt{(0.25)} \cdot \sqrt{(.4)} \cdot 17 = -2.68 \text{ Kg.}$$

The results are summarized in the following table:

Response to selection per trait for different selection criteria

Selection on	Response	
	Weight	Feed intake
Weight	6.80	0.32
Feed Intake	-2.68	-0.50

Multiple trait selection index

Most selection programs in livestock production aim for the simultaneous improvement of several traits. The theory to optimize selection on multiple traits is based on the selection index principle. The purpose of the selection index method is to combine information from different sources such that an optimal selection criterion is achieved. Information sources can be different measurements on different traits on an animal or measurements on related animals.

Combining information on related animals on a single trait is exactly what is done by BLUP, and this will not be further discussed here. BLUP can also use information on 'multiple traits'. However, it is useful to be able to do this also in a selection index context, to understand and predict the outcome of multiple trait selection. The selection index is not so much used for actual breeding value estimation, - we have BLUP for that -, but mainly to derive weights, to understand the relative value of different sources of information, and to predict response and selection accuracy for different alternatives. For example, selection index theory is used to calculate the additional merit of measuring a particular trait.

Optimal is defined as 'most accurate', or 'giving the highest selection response when selecting on it. If we rank the animals on the index, we have the best chance of ranking the animals according to their true genetic merit. If we consider only selection for one trait, a selection index is nothing else than the best prediction of a breeding value.

In the case of joint selection for more traits it is important to weigh the relative importance of the different traits. We use economic values for this purpose (see previous lecture). The index is then the best ranking for genetic merit for an aggregate genotype for the different traits in the breeding objective.

The index I is written as

$$I = b_1X_1 + b_2X_2 + \dots + b_nX_n$$

where X_i refers to a measurement of the phenotypic performance of an animal (or its relative) and b refers to the appropriate weight. The information sources X are random deviations from an expected mean, caused by random additive genetic and non-genetic effects.

Multiple trait breeding objective

If there are more traits to improve, the breeding objective is a linear combination of genotypic values for each of these traits, and it is defined as an aggregate genotype

$$\text{Aggregate genotype} = H = v_1g_1 + v_2g_2 + \dots + v_mg_m.$$

In the standard notation of animal breeding literature, the breeding objective (aggregate genotype) is usually indicated with the letter H .

With one trait in H , G is a vector. However, with more traits, the G -matrix becomes a matrix with a column for each trait in the breeding objective, containing its covariance with each information source. The number of rows in G is therefore equal to the number of information sources in X .

The b -values for a multiple trait objective are derived with the selection index equations

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

The economic weights of objective traits are included in vector v , so that the b values account for the relative importance of the breeding goal traits.

Note that in this section I will use g for breeding value and σ_g^2 for (additive) genetic variance.

Example 1

Information: individual phenotypic observations for two traits:

weight (X_1)
feed intake (X_2).

The index for an animal is $b_1X_1 + b_2X_2$.

Let the breeding objective be to improve the weight of the animal:

$$H = g_1 \quad \text{i.e. a single trait objective}$$

We can construct the matrix P as

$$P = \text{var} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_{p_1}^2 & r_p \sigma_{p_1} \sigma_{p_2} \\ r_p \sigma_{p_1} \sigma_{p_2} & \sigma_{p_2}^2 \end{pmatrix}$$

With 'the variance of X_1 ' we mean the variance of many possible observations like X_1 . In this case: the variance of phenotypic observations for weight, which is equal to the phenotypic variance for weight (indicated as $\sigma_{p_1}^2$).

Likewise, with the covariance between X_1 and X_2 we need the covariance between animals' phenotypic performances on two traits, which is equal to the phenotypic covariance.

The G-matrix:

$$G = \text{cov} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, g_1 = \begin{pmatrix} \text{cov}(X_1, g_1) \\ \text{cov}(X_2, g_1) \end{pmatrix} = \begin{pmatrix} \sigma_{g_1}^2 \\ r_g \sigma_{g_1} \sigma_{g_2} \end{pmatrix}$$

The covariance between a phenotypic observation and a breeding value contains only genetic components: either the genetic variance (if same trait) or the genetic covariance (if a different trait).

and the solution for optimum weights

$$b = P^{-1}G = \begin{pmatrix} 289 & 6.8 \\ 6.8 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 115.6 \\ 5.38 \end{pmatrix} = \begin{pmatrix} 0.384 \\ 0.692 \end{pmatrix}$$

and the optimal index to select for weight is now : Index = $I = 0.38 X_1 + 0.69 X_2$.

Note that the b-value for X_1 is not equal to heritability, because there is another sources of information which is correlated to X_1 . If this correlation was zero, or if X_1 were the only information source, the b-value would be equal to heritability indeed!

Example 2

Information: individual phenotypic observations for two traits:

weight (X_1)
feed intake (X_2).

The index for an animal is $b_1X_1 + b_2X_2$.

Breeding goal, with economic weights of 1 for weight and -4.0 for feed intake:

$$H = g_1 - 4g_2.$$

The selection index weights are found as $b = P^{-1}Ga$.

The P-matrix is the same as before, the G-matrix is

$$G = \text{cov} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, (g_1 \quad g_2) = \begin{pmatrix} \text{cov}(X_1, g_1) & \text{cov}(X_1, g_2) \\ \text{cov}(X_2, g_1) & \text{cov}(X_2, g_2) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{g_1}^2 & r_g \sigma_{g_1} \sigma_{g_2} \\ r_g \sigma_{g_1} \sigma_{g_2} & \sigma_{g_2}^2 \end{pmatrix} = \begin{pmatrix} 115.6 & 5.68 \\ 5.68 & 1 \end{pmatrix}$$

And the solutions for b become

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = P^{-1}Ga = \begin{pmatrix} 289 & 6.8 \\ 6.8 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 115.6 & 5.68 \\ 5.68 & 1 \end{pmatrix} \begin{pmatrix} 1.0 \\ -4.0 \end{pmatrix} = \begin{pmatrix} 0.331 \\ -0.218 \end{pmatrix}$$

and the optimal index to select for weight is now : Index = $I = 0.331 X_1 - 0.218 X_2$.

Note that the b-value for X_2 is now negative, as we don't want to select animals with high feed intake. In Example 1, we were not interested in Feed Intake (no economic value) and we gave it a positive weight, as heavy eaters will have a higher weight. Note also that in this example, the economic value for feed intake is very highly negative. For more realistic values (like -0.5) the weight for feed intake would not even be negative. In that case, the value of weight is so dominant; we would even select heavy eaters to increase weight.

Accuracy of index selection

The accuracy of the selection index is calculated as the correlation between predicted and true breeding value. It can be calculated as

$$r_{IH} = \sigma_I / \sigma_H = \text{SD_Index} / \text{SD_Breed_Objective}$$

more detail

To obtain this, we can first calculate the variance of the index as

$$\text{var}(I) = \sigma_I^2 = \text{var}(b'X) = b'\text{var}(X)b = b'Pb$$

The variance of the breeding objective ('true breeding values' is

$$\text{var}(H) = \sigma_H^2 = \text{var}(a'g) = a'\text{var}(g)a = a'Ca$$

$$\text{and in case of a single trait objective: } \sigma_H^2 = \text{var}(g) = \sigma_g^2.$$

The covariance between the index and the breeding goal (true breeding value) is

$$\text{cov}(I,H) = \text{cov}(b'X, a'g) = b'\text{cov}(X,g)a = b'Gv = b'Pb.$$

Hence, we see that the covariance between the index and the true genotype is equal to the variance of the index. The correlation is then

$$r_{IH} = \text{cov}(I,H) / (\sigma_I \sigma_H) = \sigma_I^2 / (\sigma_I \sigma_H) = \sigma_I / \sigma_H$$

With high accuracies, the standard deviation of the index is almost equal to the standard deviation of the breeding goal, and with low accuracy the ratio is relatively lower - compare V_A vs $\text{var}(\text{EBV})$.

Generally, the more information (index sources) we use, the higher the accuracy. In practice, measuring animals is usually related to costs. As the increase of the accuracy is directly related to response to selection, the cost of measuring can be contemplated versus the gain from expected extra response to selection.

Response to selection

Suppose selection has taken place on trait A with genotype g_A .

Genetic change for a single trait (direct response):

$$\begin{aligned} R. &= i.h^2.\sigma_P && \text{for selection on individual phenotype} \\ &= i.h_A \sigma_A \end{aligned}$$

$$R. = i.r_{IH}\sigma_A \quad \text{in case of index or BLUP selection}$$

(in both cases males and females are assumed to be selected with equal accuracy and selection intensity).

If selection is on a multiple trait selection index, the average index value of selected parents is

$$\begin{aligned} S &= i.\sigma_I \\ &= i.r_{IH}.\sigma_H \end{aligned}$$

The index is an EBV, and as the 'heritability of an EBV is equal to 1, the index superiority in parents will be fully passed on to progeny. Hence, the average value of progeny is expected to equal the average index value of parents:

$$\begin{aligned} R &= i.\sigma_I \\ &= i.r_{IH}.\sigma_H \end{aligned}$$

Response for trait i becomes $i.b'_i G_i / \sigma_i$ which is a vector (see next page for detail)

These responses are 'per round of selection' and 'per unit of selection intensity'. To obtain the actual genetic change per year, we have to multiply by the selection intensity and the inverse of the generation interval. Also, the index and index accuracy can be different for males and females, as there may be more information measured on one of the sexes.

$$\text{Response per year (in dollars): } \frac{i_m \sigma_{Im} + i_f \sigma_{If}}{L_m + L_f} = \frac{i_m r_{IH-m} + i_f r_{IH-f}}{L_m + L_f} \sigma_H$$

$$\text{Response per year (for each trait – a vector): } \frac{i_m b'_m G_{m-i} / \sigma_{Im}}{L_m + L_f} + \frac{i_f b'_f G_{f-i} / \sigma_{If}}{L_m + L_f}$$

where m and f refer to male and female index, and G_{m-i} is the i^{th} column of the G matrix for the male selection index.

Response in example 1

Selection for weight ($H = g_1$ is “aggregate” genotype), using phenotypes of weight and feed intake:

The index was $I = 0.38 X_1 + 0.69 X_2$.

The variance of the index is $b'Pb = b'G = \begin{pmatrix} 115.6 & 5.38 \\ 0.384 & 0.692 \end{pmatrix} = 48.08 \rightarrow SD_{\text{index}} = 6.93$

and the response to selection (per selection round, with $i = 1.0$) is $R = i \cdot \sigma_I = 6.93 \text{ Kg (weight)}$.

There is a correlated response for feed intake.

More Detail

The correlated response for trait 1 depends on the regression of index values (I) on genetic values for trait 2 (g_2). This regression can be calculated as $b_{I g_2}$;

$$b_{I g_2} = \text{cov}(I, g_2) / \text{var}(I) = \text{cov}(b'X, g_2) / \text{var}(I) = b' \text{cov}(X, g_2) / \text{var}(I)$$

the term $\text{cov}(X, g_2)$ is a vector with covariances of each index information source with the genotype for trait 2, like G , but than for trait 2. The denominator of the regression coefficient is equal to the variance of the index. Hence,

$$b_{I g_2} = [0.384 \quad 0.692] \begin{bmatrix} 5.68 \\ 1 \end{bmatrix} / 6.93 = 0.41 \text{ Kg (for Feed Intake)}$$

Response in Example 2 (Breeding Objective: $H = g_1 - 4 \cdot g_2$)

Variance of index: $\sigma_I^2 = b'Pb = 30.82$ $SD_{\text{index}} = 5.55$

Variance of the breeding obj. $\sigma_H^2 = v'Cv = 88.59$ $SD_{\text{breeding objective}} = 9.41$

Notice that in this example the C-matrix is identical to the G matrix.

$$\text{The accuracy is } r_{IH} = \frac{\text{cov}(I, H)}{\sigma_I \sigma_H} = \frac{\sigma_I^2}{\sigma_I \sigma_H} = \frac{\sigma_I}{\sigma_H} = 0.59$$

Response to Index selection is equal to $R = i \cdot \sigma_I = i \cdot r_{IH} \cdot \sigma_H = 1.0 \cdot 0.59 \cdot 9.41 = \5.55

The unit of this response is in the units of the economic weights, e.g. in dollars.

Response for trait i becomes $i \cdot b'G / \sigma_I$.

Response for weight in the last example:

$$\delta g_1 = i \cdot (b_1 \quad b_2) \begin{pmatrix} \text{cov}(X_1, g_1) \\ \text{cov}(X_2, g_2) \end{pmatrix} / \sigma_I = 1.0 \cdot (0.331 \quad -0.218) \begin{pmatrix} 115.6 \\ 5.68 \end{pmatrix} / 5.55 = 6.68 \text{ Kg}$$

and similarly the correlated response for feed intake is equal to 0.28 Kg.

(note: in spite of a negative weight still a slight increase of feed intake)

Confirm the total response using economic weights: $R = 1 \cdot (6.68) + (-4) \cdot (0.28) = 5.55 \$$

More detail (for reference)

Response for each trait is determined with the regression of each trait on the index (i.e. how much does a trait increase for one unit of increase of the index).

This response for trait i is $\delta g_i = b_{g_i, I} R = [\text{cov}(I, g_i) / \sigma_I^2] \cdot i \cdot \sigma_I$ [1]

Notice that $\text{cov}(I, g_i) = \text{cov}(bX, g_i) = b \cdot \text{cov}(X, g_i) = b \cdot G_i$ [2]

Where G_i is the i -th column of the G matrix.

and substituting [2] in [1] gives the response for trait i becomes $i \cdot b'G_i / \sigma_I$.

Manipulating multiple trait response

The response for multiple traits depends on the economic weight given to the traits, and on the biology of the traits (heritability, correlation structure). These could be referred to as economic and genetic (technical) parameters. For a given production system and genetic resource, where the biology is defined by these genetic parameters, we can manipulate the response in two ways

- 1) by manipulating economic weights
- 2) by varying the information sources used

The second option is less powerful than the first. But generally, measuring more information about a trait gives more power to change that trait genetically. Below is a Table with response to index selection for weight and feed intake for varying economic values. Response is per round of selection ($i = 1$)

TABLE 2.7 Response (R_i) per trait per selection round for different selection index strategies. The economic weights in the breeding goal are v_1 for weaning weight and v_2 for feed intake.

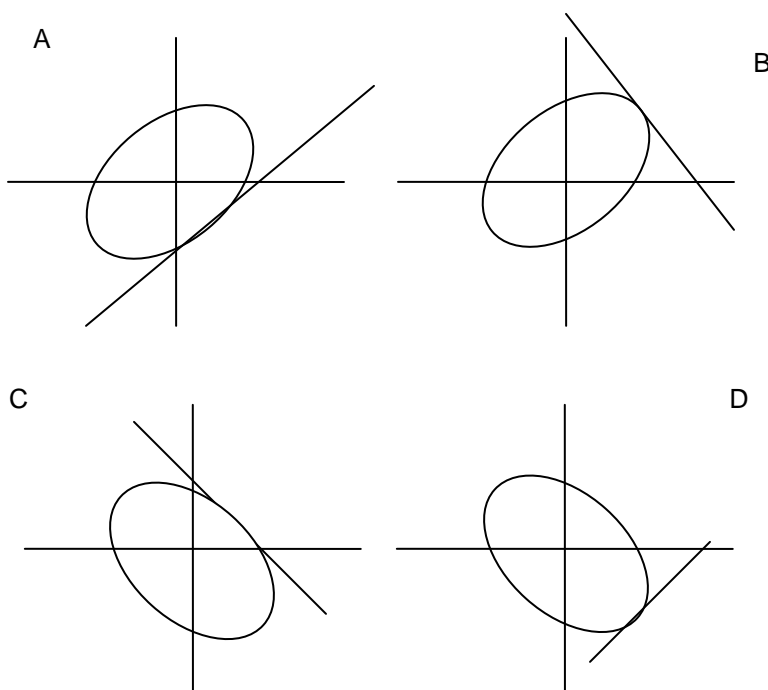
Information on	breeding goal		R. weight	R. feed intake
	v_1	v_2		
Weight	1	0	6.80	0.32
Weight + feed	1	0	6.93	0.40
Weight + feed	1	-1	6.92	0.38
Weight + feed	1	-4	6.68	0.28
Weight + feed	1	-6	6.21	0.19
Weight + feed	1	-8	5.42	0.08
Weight + feed	1	-10	4.29	-0.05
Weight + feed	1	-12	3.00	-0.17
Weight + feed	1	-16	0.66	-0.34
Weight + feed	1	-20	-0.93	-0.43
Weight + feed	0	-1	-5.04	-0.55
Feed	0	-1	-2.69	-0.50

The table shows that index selection (selecting on 2 traits) gives more response than selecting on one trait alone.

If the derived economic weights for weaning weight and feed intake were 1 and -4, respectively, and we wish a decrease of feed intake, then such a change goes to the expense of weight gain. We need a weight for feed intake of -10 or more (negative) to achieve a decrease in Feed Intake. Suppose we applied the economic weight 1 for Weight and -20 for Feed Intake. Indeed, a larger decrease for feed intake is achieved, but the weight is also decreased. This illustrates that the optimal genetic changes that can be made need to be based on economic values, if possible. However, sometimes, very drastic economic values are needed in order to achieve a result that seems desirable. This is a reason why just going for desired gains approaches can imply quite extreme economic weights, and the optimality should always be related to the implicit economic weight that makes such a change optimal. In our example: any economic value for feed of more than -1\$/Kg seems too value feed cost too much and in tis example optimal genetic change should increase weight and allow a correlated increase in feed intake.

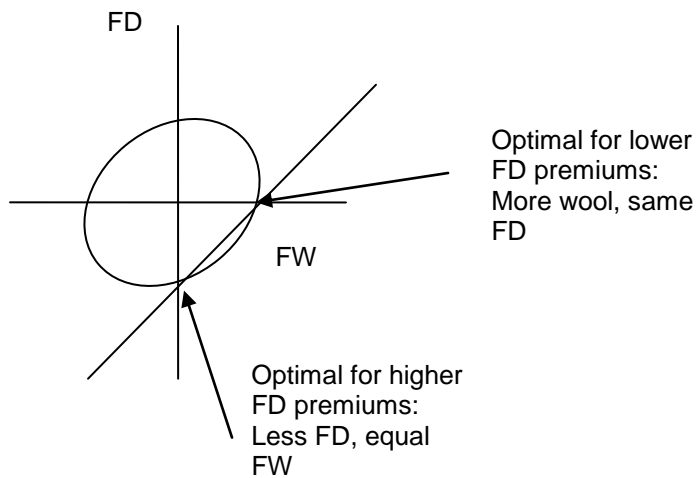
Favourable versus unfavourable correlations

The results in the previous example show that in spite of a positive correlation between Weight and Feed Intake, it is possible to select them in opposite directions. The correlation between these traits is positive and unfavourable, if our breeding goal is to increase Weight while decreasing Feed Intake. When plotting all possible responses, one would obtain an ellipse as in the figure below. The ellipse would be tilting to the right for positive correlations, and to the left if the correlation between traits was negative. The higher the correlation, the flatter the ellipse (and for zero correlations, the 'ellipse' would be a circle). The sloped line in the figures represents an iso-economic line, i.e. a line of equal profit for the different trait combinations. The optimal selection response is determined by the point where the lines touches the ellipse (something that is beautifully and dynamically illustrated in the 'ELLIPSE' module in GENUP. Now there are 4 possible situations for a 2-trait combination, determined by the sign of the correlation (the slope of the ellipse) and whether the economic values are of equal or of opposite sign (the slope of the iso-economic line)



Correlation	Sign of economic weights	
	Equal	Opposite
Positive	Favourable (B)	Unfavourable (A)
Negative	Unfavourable (C)	Favourable (D)

Generally, when correlations are unfavourable, it is much harder to improve both traits jointly in the same direction, and the direction of selection will be quite sensitive to economic values. For example, for fleece weight and fibre diameter (case A), the optimal change can vary from emphasising FD versus emphasising FW depending on changes in price ratio that are quite realistic. Hence, in such cases, it is relatively hard to define an appropriate selection strategy.



Selection Index versus Multiple Trait BLUP

Selection index theory deals with combining information and pedigree into one index value. However, it is also convenient to see this as a two step process, where first we derive EBV for each trait separately, but using information on all traits. Subsequently these EBVs are combined into an index. It will be shown that the weights in that case are simply the economic values as defined in the breeding objective.

In multivariate selection, it is important to distinguish two components, which need to be optimized: breeding value estimation and weighing of traits. We consider selection index methodology, with several traits (say n) included in the index to predict an optimum value for ranking animals with respect to a breeding goal, consisting of a weighted aggregate of several (say m) traits. Traits in the breeding goal may be different from traits in the index. The selection index method combines the information on index traits using genetic and phenotypic parameters to predict a value for an aggregate of traits, using both genetic parameters and economic weights. The second component assumes knowledge about the (economic) weights of the breeding goal traits and some kind of linearity of the profit function. In fact, the last assumption may not always be justified, which could be a reason to use some type of independent culling or desired gains approach rather than selection index as tool for selection.

The BLUP breeding value estimation method only relates to the first component, i.e. it combines optimally the information on animals and their relatives to estimate breeding values for each desired trait. Information can be from several traits if a multitrait estimation method is used. The result of using the BLUP methodology is not a single index value but a set of m estimated breeding values per animal, one for each trait. Multitrait BLUP breeding value estimation is therefore an extension of selection index only in the sense of obtaining more accurate breeding values, combining information from several relatives on several traits and adjusting appropriately for fixed environmental effects. The method, however, is neutral to any procedure of weighing breeding goal traits. It is, therefore, useful to assess the relevance of multivariate BLUP in terms of accuracy of individual breeding values. In selection index terminology these individual breeding values per trait are known as sub-indices.

$$\text{\$-index} = \sum_{i=1}^{\text{nr.of_traits}} \text{BLUP_EBV}_i * v_i \quad [2-1]$$

An optimal index could possibly be produced by weighing the BLUP breeding values by their economic weights (v_i) as used in selection index, or possibly using other weights. Suppose the index is $\mathbf{b}'\hat{\mathbf{g}}$ and the aggregate genotype is $\mathbf{v}'\mathbf{g}$, and $\hat{\mathbf{g}}$ are the estimated breeding values of \mathbf{g} . If $\hat{\mathbf{g}}$ is estimated using the true variance-covariance structure, i.e. using a multivariate model, then

$$\text{var}(\hat{\mathbf{g}}) = \text{cov}(\hat{\mathbf{g}}, \mathbf{g})$$

$$\mathbf{b} = \text{var}(\hat{\mathbf{g}})^{-1} \text{cov}(\hat{\mathbf{g}}, \mathbf{g})\mathbf{v} = \mathbf{v}.$$

Hence the \mathbf{b} -values are equal to the economic weights in \mathbf{v} when multivariate BLUP values are used in an index. Notice that the weight does not depend on the accuracy of each of the EBV's in $\hat{\mathbf{g}}$.

Schneeberger *et al.* (1992) discussed a related problem where breeding values were estimated for a set of index traits \mathbf{u} whereas the selection criteria existed of a set of other traits in \mathbf{g} . They showed that optimal weights for the selection criterion $\mathbf{b}'\mathbf{u}$ were $\mathbf{b} = \mathbf{G}_{11}^{-1}\mathbf{G}_{12}\mathbf{v}$ with $\mathbf{G}_{11} = \text{var}(\mathbf{u})$ and $\mathbf{G}_{12} = \text{cov}(\mathbf{u}, \mathbf{g})$. Hence, when the estimated breeding values are on traits which are different from those in

the breeding goal, the weights depend on the genetic regression from breeding goal trait on index traits and the economic weights, but not on the accuracy of the estimated breeding values. The condition is that estimates have to be available for all index traits.

Notice that in [2.1] no information is needed about family structure or availability of information for each animal. All animals have the same index weights although their estimated breeding values might have very different accuracy.

This result seems inconsistent from what most people experience when using multiple trait selection index (where index weights are different for each animal with different sources of information). However, Multiple Trait BLUP already has accounted for differences in information when predicting breeding values, i.e. animals with little information will have smaller accuracies and therefore will be more regressed towards the mean.

An important condition for [2-1] is often not realised, that is that the estimated breeding values are assumed to be estimated by multiple trait techniques. In practice it might seem tempting to simply calculate single trait evaluations and use those in the aggregate index. Estimated single trait values have to be somehow combined in an aggregate index and in fact this is more difficult than combining multi-trait breeding values because the correlation structure of estimated breeding values has to be accounted for.

If \hat{g}_u were estimated breeding values estimated with univariate analysis, the optimal weighting would be $\mathbf{b} = \text{cov}(\hat{g}_u, \mathbf{a}) (\text{var}(\hat{g}_u))^{-1}$. Covariances between estimated breeding values from single trait analysis are hard to obtain from BLUP analyses. They are not only affected by residual- and genetic correlations, but also by different family structures used for each estimated breeding value. In selection index one could derive the (co)variance for $g_i = b_i x_i$ and $g_j = b_j x_j$ as $b_i' \text{cov}(x_i, x_j) b_j$, where $\text{cov}(x_i, x_j)$ relates to the same submatrix as necessary for multiple trait indices. The efficiency of combining univariate BLUP breeding values as well as the question whether such an index would be more sensitive to the use of incorrect variances and covariances will be discussed in the subsequent paragraphs.

Increased accuracy from using information on correlated traits

Accuracy in mixed model terms is usually expressed in terms of Prediction Error Variances (PEV's), which is $\text{PEV} = \text{var}(\hat{g} - g) = (1 - r_{IH}^2) \sigma_A$ and r_{IH} is the correlation between true and predicted breeding value. The gain to be expected from increased accuracy due to including information on correlated traits can in principle be assessed by r_{IH} values using selection index theory (e.g. Gjedrem, 1967). Thompson and Meyer (1986) give an illustrative table on the efficiency of a selection index for two sets of correlation values and a range of heritabilities. The example is for phenotypic selection (Table 2.8).

The gain in efficiency is mostly dependent on the difference between genetic and environmental correlation and the accuracy for a trait with a low heritability could be significantly improved if the information from a correlated trait with high heritability is included (Table 2.8). Schaeffer (1984) gives examples for more combinations of parameter values with generally the same results.

Table 2.8 Effect of genetic and environmental correlations and heritabilities on ratio of accuracies of multivariate versus univariate evaluation of trait 1 assuming information on individual only (from Thompson and Meyer, 1986)¹.

	h^2_2	h^2_1		
		0.1	0.3	0.5
$r_g=r_e=0.5$	0.1	1.00	1.02	1.03
	0.3	1.09	1.00	1.01
	0.5	1.25	1.02	1.00
$r_g=-r_e=0.5$	0.1	1.40	1.18	1.10
	0.3	1.59	1.23	1.11
	0.5	1.70	1.25	1.12

¹ h^2_1 is heritability of trait 1 and h^2_2 is heritability of correlated trait 2.

It is shown in the table that genetic and residual correlations are equal, and when heritability of the two traits is the same, there is no benefit of using a multivariate analysis. In this case the genetic regression of one trait on another is equal to the phenotypic regression. The correlated trait does not give us any clue in determining which part of an individuals' phenotypic deviation is due to genetic or due to environmental effects. Suppose that G_o and R_o are matrices with residual and genetic covariances, respectively. When heritabilities are equal and $r_e = r_g (= r_p)$, and when all animals have an observation for each trait, G_o can be written as a multiple of R_o . The MT mixed model equations can be written as

$$R_o^{-1} \otimes Z' MZ + G^{-1} \otimes A^{-1} = I_2 \otimes Z' My$$

and when $R_o = \lambda G_o$, pre-multiplying the MT system by R_o gives two independent single trait systems of equations:

$$I_2 \otimes Z' MZ + \lambda I_2 \otimes A^{-1} = I_2 \otimes Z' My$$

Table 1 shows that the marginal benefit from including information on a correlated trait decreases with higher heritabilities, i.e. when single trait evaluations are relatively accurate. Therefore, a table comparable to Table 1 could be presented when each animal has for each trait not only an own record, but in addition a record on each parent and a mean of the full- and half-sib group (Table 2.9).

Table 2.9 Effect of genetic and environmental correlations and heritabilities on ratio of accuracies of multivariate versus univariate evaluation of trait 1 assuming information on relatives known¹

	h^2_2	h^2_1		
		0.1	0.3	0.5
$r_g=r_e=0.5$	0.1	1.00	1.01	1.02
	0.3	1.03	1.00	1.00
	0.5	1.08	1.01	1.00
$r_g=-r_e=0.5$	0.1	1.18	1.08	1.05
	0.3	1.22	1.10	1.06
	0.5	1.25	1.11	1.07

¹ Breeding value is estimated from phenotype of animal itself, its sire and its dam, on 4 full sibs and on 36 paternal half sibs for each trait.

When records from relatives are also included for each trait, the improvement from including information on a correlated trait is reduced. The effect of the heritabilities on the ratio of accuracy from multivariate and univariate models is smaller and the ratio is almost unity if the phenotypic and environmental correlations are rather equal.

For the same example used in Table 2.9 we could also determine single trait indices for each trait and combine those in an index (taking into account the variance and covariances between these univariate estimates).

A combined index of univariate breeding values is as efficient as a multivariate index when $r_e = r_g$ and it is about 3% less efficient when $r_e = -r_g = 0.5$. In the latter case, information is lost from averaging sources of information that are not equivalent. The weights for the different information sources of the correlated trait are dependent on the phenotypic and genetic regression (Thompson and Meyer, 1986) and those regressions are not equal for say an own record and a full sib mean. Noteworthy is that the ratio of accuracy from a multivariate index and a combined univariate index is almost fully dependent on the differences between r_g and r_e , whereas the heritabilities have an insignificant effect.

Effect of using incorrect parameters in genetic evaluation

Because true genetic parameters are not known, genetic evaluations can only be done using parameter estimates. The realised response to selection will be always less or equal to the optimum response and the expected response (these are all equal when parameters used are correct). When P is the VCV matrix among phenotypic observations, and G is the covariance matrix between these observations and the breeding value to be estimated, and the 'hats' refer to the used (estimated) values, then

Optimum response is (in units of i)	$R_{\text{opt}} = \sqrt{G' P^{-1} G}$
Expected response is	$R_{\text{exp}} = \sqrt{\hat{G}' \hat{P}^{-1} \hat{G}}$
Realised response is	$R_{\text{act}} = \frac{\hat{G}' \hat{P}^{-1} G}{\sqrt{\hat{G}' \hat{P}^{-1} P \hat{P}^{-1} \hat{G}}}$

A simple method to study the effect of using incorrect parameters is to try arbitrary values for incorrect parameters in a selection index and compare the response with that obtained with the correct values, and what was predicted using the incorrect values. Examples are given in Figure 2.2 and Figure 2.3.

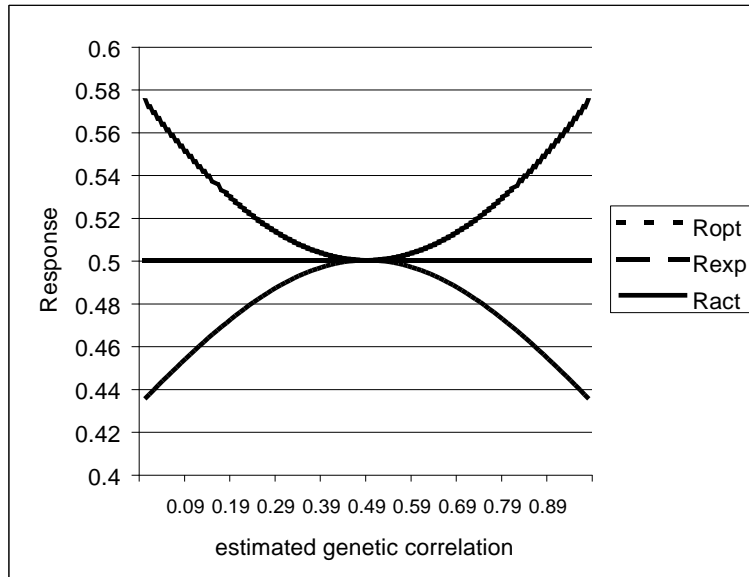


Figure 2.2
Optimal (R_{opt}), Actual (R_{act}) and Expected (R_{exp}) Response to selection for one trait, with varying estimated genetic correlation. Selection is on phenotype for two traits, heritabilities are 0.5, phenotypic and true genetic correlations are 0.5.

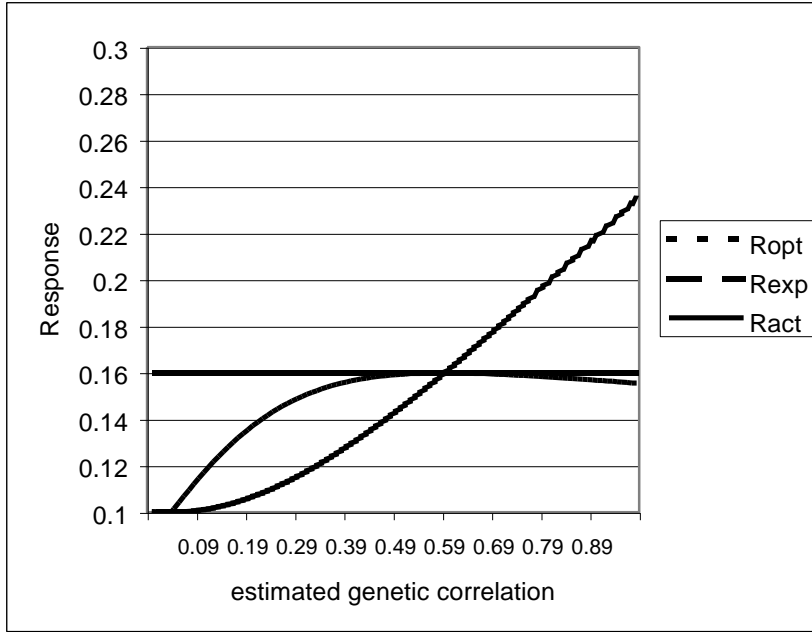


Figure 2.3. Optimal (Ropt), Actual (Ract) and Expected (Rexp) Response to selection for oneh trait, with varying estimated genetic correlation. Selection is on phenotype for two traits, $h_1^2 = 0.1$; $h_2^2 = 0.5$; $r_p = 0.1$ and true $r_g = 0.6$

Schaeffer (1984) used this method of trial values for incorrect genetic parameters and calculated the loss in efficiency of multiple trait BLUP models as the increase in prediction error variance. He noticed that the increase was very much dependent on the Absolute Difference, which he defined as

$$AD = |(\hat{r}_e - r_e) - (\hat{r}_g - r_g)|$$

Where \hat{r}_g and \hat{r}_e are estimates of genetic and error correlation, respectively. The formula shows that the loss of accuracy would be less if both genetic and error correlation would be over- or underestimated. Decrease of accuracy would be large if one correlation was overestimated whereas the other correlation would be underestimated. Increase in PEV varied from 1% for AD=0.1 to 35% for AD=1.8.

The AD-value, however, predicts loss of efficiency particularly quite well if both traits have equal heritability. When heritabilities are different, the absolute difference between estimated and true r_g becomes relatively more important. This is illustrated in Table 5. Traits with lower heritability generally suffer more from an increase in PEV when incorrect parameters are used. Animal models with a reasonable amount of information from relatives available (higher accuracy), and sire models are less sensitive to incorrect correlations.

Table 2.10 Effect of incorrect genetic correlations on ratio of accuracies of multivariate versus univariate evaluation of trait 1 for two sets of correlation values, assuming information on relatives known

r_g	r_e			$h^2_1=0.3$	$h^2_1=.1$
				$h^2_2=0.3$	$h^2_2=.5$
				AD	MT/ST
.5	.5	true values	0	1.00	1.08
.2	.2		0	1.00	1.04
-.2	-.2		0	.97	.84
.7	.3		.4	.98	1.05
.3	.7		.4	.97	1.03
-.2	.2		.4	.94	.74
-.5	.5		1	.93	1.04
.5	-.5	true values	0	1.10	1.25
.7	-.3		0	1.10	1.23
.2	-.2		.6	1.06	1.17
-.2	.2		1.4	.84	.64

¹ Selection is on phenotype of animal itself, its sire and its dam, on 4 full sibs and on 36 paternal half sibs for each trait.

The effect of avoiding multiple trait analysis would fit in the perspective of using incorrect parameter estimates, i.e. a single trait analysis is equivalent to using prior correlation of 0. In other words, a multiple trait with incorrect parameters could be a better alternative than single trait analysis as is illustrated in some examples in Table 2.10.

The previous exercises might give some idea about using incorrect correlations, but in practice, we usually do not know true correlations, and therefore can not make such comparisons. A more relevant questions might be how accurate an estimate of a certain parameter needs to be, to use it with good confidence in multi trait genetic evaluation. Sales and Hill (1976a) proposed a method to determine the proportional loss in response when estimated rather than true values were used for the parameters in a single trait selection index. The method was based on a Taylor's series approximation of the function for actual response dependent on the accuracy of the parameter estimates. They found that for single trait selection, the loss of efficiency from using incorrect parameters (i.e. heritability) is very small, even for large differences between true and estimated heritability.

Sales and Hill (1976b) applied the same method also to a selection index using information from correlated traits and showed the proportional loss from the optimal response when selection was on phenotypical values for each trait. They considered an economically important trait, and the value of using information from a second correlated trait (with no economic value). They made the loss of response dependent on the sample size that was used to estimate the parameters (i.e. on the accuracy of the parameter estimate rather than on an actual (but arbitrary) difference between true and estimated parameter). Very inaccurate parameter estimates indeed may make single trait analysis more beneficial than using (unreliable) information from a correlated trait.

Generally they found for a wide range of parameters the proportional loss to depend critically on h_1^2 (the heritability of the economic trait) but much less on the correlations and the heritability of the correlated trait. Table 2.11 (from Table 2 from Sales and Hill) illustrates what should be the minimum size of an experiment to estimate parameters to obtain equal expected response using 2 traits (compared to only using trait 1). They considered selection on individual phenotypes only.

Table 2.11

Number of observations required to get an increase in expected response using two traits, from an initial experiment with s half sib families of size n (Sales and Hill, 1997b)

h_1^2	h_2^2	R_g	R_e	R_p	R/R_{ST}	T (n=4)	T (n=16)
0.2	0.5	0.0	0.16	0.10	1.005	17800	11632
		0.0	0.50	0.32	1.054	1696	1056
		0.5	-0.5	-0.16	1.386	208	80
		0.5	0.0	0.16	1.187	456	208
		0.5	0.5	0.47	1.062	1432	832
0.5	0.5	0.0	0.2	0.10	1.005	3080	2672
		0.0	0.5	0.25	1.035	476	400
		0.5	-0.5	0.0	1.118	124	80
		0.5	0.0	0.25	1.035	456	336
		0.5	0.5	0.5	1.000	-	-

R = optimum response (if values were the true ones), R_{ST} = single trait response, T = sample size of experiment to estimate genetic parameters so that expected $R > R_{ST}$.

Generally, when the correlated trait adds more information, there's a smaller amount of observations required. If little extra information is added by the correlated trait, very large sample sizes would be required to obtain estimates that are accurate enough to include in a multi trait index. Sample sizes need to be a bit larger if family sizes are very small ($n=4$). A special case occurs when the additional trait adds no value to the economic trait, since incorrect parameters are likely to have most effect. It should be reminded that a second trait's contribution is zero if genetic and phenotypic correlations are 0, but more generally if $r_g h_2 = r_p h_1$ (phenotypic regression = genotypic regression). So even if phenotypic correlations are positive (easy to measure), one needs to know genetic correlations as well about decisions to incorporate a second trait.

The use of incorrect parameters has no effect on bias in genetic evaluation (Henderson, 1975) with no selection. In case of selection, however, not all selection bias would be removed by analysis with a multiple trait model that does not have the correct parameters and therefore such analyses would be biased.

Schaeffer (1984) also discussed the problem of invalid variance-covariance matrices because of inconsistencies in parameter values. This problem can be avoided by checking the eigenvalues of the variance-covariance matrices; they should be positive, i.e. If the estimated variance-covariance is not positive definite, the values should be adjusted, e.g. using the bending algorithm described by Hayes and Hill (1976).

Other issues

Which traits to include in the selection index?

Criteria Traits: All traits that add information and that can be measured cost effectively
Breeding Objective Traits: All traits that have economic value

Although in principle there could be many traits that could have an economic value different from zero, the breeding goal contains usually only a limited number of traits. For traits to be included in selection index, phenotypic, genetic and economic parameters have to be known. Such parameters are usually estimated only for 'obviously measurable' traits like production and reproduction traits. From other traits there are usually no accurate parameters known and it is difficult and unpractical to include such traits in the breeding goal. Moreover, the additional gain from including more traits may be small. The question whether or not certain traits should be included in the breeding goal can be assessed by selection index theory. Another important

question, more related to costs, that can be solved by selection index is whether or not certain traits should be included in the information index, i.e. whether or not they should be measured..

The selection index approach is therefore a very general method to optimize selection on multiple traits, and to weigh the information on several relatives and for several traits optimally in an index.

If there is no phenotypic (or marker) information available for a certain trait, we could still include it in the breeding objective. However, generally we would not directly produce an EBV for such a trait. Then how would we construct a selection criterion, based on economic weights times EBV for all traits in the breeding objective? Firstly, we could calculate an EBV of such a trait, based on a genetic regression on the other traits that we have an EBV on. For example, consider trait j with g_j . EBV for all other traits are represented by the vector \hat{g} . \hat{g}_j can be estimated as

$\text{cov}(g_j, g)/\text{var}(g)$, which is a genetic regression resulting in a vector, say w . Hence, for each

individual we estimate the EBV as $\hat{g}_j = \sum_{i=1}^m w_i \hat{g}_i$ where \hat{g}_i are BLUP EBV on other traits.

The selection index becomes then

$$I = \sum_{i=1}^m v_i \hat{g}_i + v_j \hat{g}_j$$

However, we would get an equivalent index value if the economic value for each trait was adjusted for the unmeasured trait. The index can also be written as

$$I = \sum_{i=1}^m v_i \hat{g}_i + v_j \sum_{i=1}^m w_i \hat{g}_i = \sum_{i=1}^m (v_i + v_j w_i) \hat{g}_i$$

A practical example of this: suppose selection is for more Weight and less Feed Intake. And suppose only Weight is measured. We can now either calculate an EBV for Feed Intake by a genetic regression on the EBV of Weight, and subsequently use these 2 EBV in an index with the appropriate economic weights for each trait. However, we could also calculate the index based on just one EBV for weight, and correct the economic weight for Weight by adjusting it for the amount of extra feed eaten if an animals grows bigger.

Selection index assumed linearity in the breeding objective

The linearity of the breeding goal implies that a negative value for one trait can be compensated by a positive value for another trait in the breeding goal. From the perspective of an individual breeding company, who has to consider the competitive value of his livestock, this compensation does not always hold. If the animals are low for a certain trait, it may be under a certain acceptable level for the consumers. In that case, the improvement should be solely focused on this trait that deteriorates the competitive value, until it is above an acceptable level. Non linearity of economic value is for many breeders a reason to make recourse to other selection methods. A simple alternative is independent culling, a method that selects animals only if they are above a minimum value for a certain set of traits. However, for small deviation of linearity, selection index is still an optimal method, and independent culling is in most cases not as optimal as selection index.

Selection Index versus desired gains

Sometimes the changes predicted per trait may be surprising or not satisfying. We may wish a trait to increase, but the optimal index shows that an optimum change should be a decrease. The reason could be that the trait is negatively correlated with another trait that needs to be improved. The change in the individual traits depends not only on economic weights but also on heritability and genetic and phenotypic correlations, and is sometimes hard to predict.

Sometimes, the knowledge about the economic weights is not accurate but we might have an idea in which direction traits need to be changed. The previous table shows that if we put more weight on feed intake in the breeding goal, a much larger change could be achieved. In principle, we could choose a certain ratio of changes for both traits and choose the “economic” weights accordingly. This approach is called “selection index with desired gains”. One particular example of a desired change index is to choose the weight so that one breeding goal trait does not change at all. Such an index is called a “restricted index”. A restricted index could be applied if the trait has an optimum value, i.e. when the economic weight is not linear with the trait mean. In general, it is suboptimal to choose for a desired gains approach, since one puts potentially very much emphasis on traits that are difficult to improve (e.g. due to a low heritability). The small ‘desired’ improvement that is achieved goes to the expense of a large loss of potential genetic gain for another trait that was easier to improve. Therefore, if economic weights could be derived without too much uncertainty, straightforward selection index is the most efficient approach.

Indirect selection

If the breeding goal is the improvement of a certain character, say trait A, an obvious approach is to improve the trait by direct selection on trait A. An alternative strategy is to select on a correlated trait B. to achieve progress in trait A. Such a strategy is called “indirect selection”. The motivation for indirect selection can be

- it is cheaper to measure B rather than A.
- it is impossible to measure A
- indirect selection can be more efficient.

Examples of cheaper indirect selection are the selection on backfat to determine carcass quality. A very expensive but more direct measurement (on a live animal) would be a computer tomography scan. Slaughter quality measured on the carcass is an example of a trait impossible to measure on breeding animals since usually we prefer not to slaughter breeding animals. For this example, another alternative to direct selection is also used, being selection not on a trait correlated to carcass value, but to measure carcass value on related animals.

It can be easily determined for which cases indirect selection is more accurate as direct selection. Let us consider only selection on individual phenotype. The direct response by selecting on trait A is

$R_A = i \cdot h_A \cdot \sigma_{gA}$ and the correlated response for trait A if selection is on trait B is

$CR_A = i \cdot h_B \cdot r_{g \cdot} \cdot \sigma_{gA}$

Indirect selection is therefore more efficient as direct selection if $CRA > RA$, i.e. if $h_B \cdot r_{g \cdot} > h_A$. In words; indirect selection is more efficient as direct selection if the heritability of the indirect character is much higher and the correlation with the trait under selection is also reasonably high.

Genotype-environment interaction

We base selection decisions on the performances of animals measured under certain circumstances. We select therefore in a certain environment. It is possible that the best genotypes of the animals we choose in one environment are not the best in another environment. An example is the selection of dairy bulls based on daughter performances under circumstances with optimal feeding- and health conditions and using such bulls to improve cattle under tropical conditions. Besides the fact that in another environment other traits might be important (e.g.

specific disease resistance), also the trait under direct selection might be expressed differently in different environments. If differences between genotypes are not the same in different environments, there is genotype-environment interaction. The question is then under which environment should we select the animals?

It was generally assumed that selection should take place under the most optimal conditions, since animals were then able to optimally show their genetic potential. However, Falconer (1960) described experiments with selection in mice that selection under suboptimal conditions resulted in a better overall performance. Furthermore he concluded that if the environments are very different, it is generally better to select the animals under the environment in which they (or their offspring) are going to produce.

Falconer (1960) introduced an interesting concept of genotype-environment interaction. He considered performance for a trait in different environments as performances of two different traits with a genetic correlation between them. For example, the growth performance can be measured under ad lib feeding, and under a restricted feeding regime. It is possible that animals would rank differently when they were ranked under the ad lib regime performance than as if they had been measured under restricted feeding. Under ad lib conditions, the feed intake capacity will be important for growth, and under restricted feeding it is more important to improve the efficiency utilizing feed. Both underlying traits are possibly not affected by the same genes, but measurements of growth under the two conditions will be correlated. If the genetic correlation of measurements of the same trait in different environments is equal to 1, there is no genotype-environment interaction. The response to selection for a trait in the 'other' environment can be predicted based on the indirect selection principles (see previous paragraph). It is important that the breeding goal trait is defined for the environment that animals will have to produce under in the near future.

REFERENCES

- Gjedrem, T., 1967. Selection indexes compared with single trait selection. I. The efficiency of including correlated traits. *Acta. Agric. Scand.* 17:263-268.
- Hayes, J.F. and W.G. Hill. 1981. Modification of estimates of parameters in the construction of genetic selection indices ("bending"). *Biometrics* 37:483.
- Jensen, P. and P. Barton-Gade, 1985. Performance and carcass characteristics of pigs with known genotypes for halothane susceptibility. In: *Stress susceptibility and meat quality in pigs*. EAAP Publ. no. 33: 80.
- Sales, J. and W.G. Hill. 1976a. Effect of sampling errors on efficiency of selection indices. 1. Use of information from relatives for single trait improvement. *Anim. Prod.* 22:1-7.
- Sales, J. and W.G. Hill. 1976b. Effect of sampling errors on efficiency of selection indices. 2. Use of information on associated traits for improvement of a single important trait. *Anim. Prod.* 23:1-14.
- Schaeffer, L.R. 1984. Sire and cow evaluation under multiple trait models. *J. Dairy Sci.* 67:1567-1580
- Schneeberger, M., S.A. Barwick, G.H. Crow and K. Hammond. 1992. Economic indices using breeding values predicted by BLUP. *J. Anim. Breed. Genet.* 109:180-187.