

# 2. Principles of Estimation of Breeding Values

Julius van der Werf

## Learning objectives

On completion of this topic you should be able to:

- Understand the principle of estimation of breeding value (EBV)
- Understand accuracy of EBV
- Predict how much EBVs vary among individuals
- Predict how much EBV might change with new information
- Predict response to selection based on EBV

## Key terms and concepts

Estimated Breeding Value (EBV), Accuracy of EBV, Prediction Error Variance, Variance of EBV, Predicting Response to Selection on EBV.

## Introduction to the topic

Within animal breeding systems, we would like to rank and select animals based on their true breeding values (TBV or "A": additive genetic value). Unfortunately we don't have this perfect knowledge - we cannot see genes and breeding values, and so we must use observed phenotypes to obtain estimated breeding values (EBV's or  $\hat{A}$ ).

Breeding values are average effects of genes that are transmitted by a parent to an offspring.

Genetic change due to selection based on EBV will be lower than if selection had been on true breeding value. The relative response is proportional to accuracy of EBV, and accuracy is between 0 and 1.

The most obvious piece of phenotypic information we can use to estimate an animal's breeding value is the animal's own phenotype. But we can also use information from relatives, such as the sire, the dam, siblings and progeny. Commercial genetic evaluation systems produce EBVs for each animal for all traits of relevance. Such an evaluation is based on a statistical procedure leading to Best Linear Unbiased Prediction (BLUP) of breeding values. Inclusion of information from relatives is automatically taken care of in the BLUP method, provided such information is available in the database through the knowledge of pedigree. We will discuss the BLUP procedure in greater detail in later topics.

In this topic we will discuss the principle of estimation of breeding value. First we present how phenotypic information is turned into an EBV. The heritability of the trait is important here. Second, we discuss the accuracy of EBV, how it is calculated, and why it is important. Basically, the higher the accuracy, the higher the response to selection based on EBV. Closely related to accuracy are the variance of the EBV and the Prediction Error Variance of EBV. The variance of the EBV indicates how much difference in EBV values we can expect. This is relevant if we want to predict the EBV of a selected group of individuals. The

Prediction Error Variance gives information about the uncertainty of an EBV. It is used to predict how much an EBV can still change when more information becomes available. The higher the accuracy of an index, the smaller the change when more information comes in. This change has great practical importance, e.g. when determining the price differences between top breeding animals based on EBV.

## 2.1 Principle of estimating breeding values

### Components of variation

Breeding Values are estimated based on phenotypic differences between animals. It is based on the notion that part of these phenotypic differences is due to genetic components. Quantitative genetic theory is used here, and one of the key principles is that phenotypic differences occur due to genetic as well as non-genetic differences. Differences are measured as variance, and therefore, phenotypic variance is the sum of genetic and non-genetic variance. Not all genetic differences between individuals are passed on to progeny.

We are only interested in additive genetic variance as

- it is the variance in breeding value, i.e. that part of the genetic differences that are passed on to progeny
- it can be more easily estimated than total genetic variance, e.g. differences between progeny groups of different sires are a reflection of additive genetic differences.

The non-additive genetic effects are due to dominance effects as well as (some) epistatic effects. These effects are not unimportant in real life, e.g. they could explain why some sires make particularly good combinations with certain cows (farmers call this 'nicking'), but these effects are hard to estimate (because the same sire is not often used to the same cow), therefore non-additive effects are hard to utilise in breeding. Non-genetic effects are therefore put in the basket of 'residual effects' together with non-genetic effects due to environmental differences. Therefore

Phenotypes are made up of breeding values and residual effects:

$$P = A + E$$

We use "E" for 'environmental effect, although this is actually a residual effect, also containing non-additive genetic effects

Phenotypic differences are due to differences in breeding values plus differences in residual effects

$$VP = VA + VE$$

where: VP = Phenotypic Variance  
 VA = Additive Genetic Variance  
 VE = Residual (or Environmental) Variance

Table 2.1 gives an illustration of these concepts, by considering the differences between 5 rams for fleece weight

**Table 2.1 Performance of 5 rams and underlying true effects.**

| Observed Phenotype (Fleece Weight) | P Phenotypic Deviation | A Additive Genetic Value | E Residual Effect |
|------------------------------------|------------------------|--------------------------|-------------------|
| 4.4                                | 0.4                    | 0.2                      | 0.2               |
| 4.2                                | 0.2                    | -0.1                     | 0.3               |
| 4.05                               | 0.05                   | 0.25                     | -0.2              |
| 3.75                               | -0.25                  | -0.05                    | -0.2              |
| 3.6                                | -0.4                   | -0.3                     | -0.1              |
| Standard Deviation                 | 0.326                  | 0.226                    | 0.235             |
| Variance                           | 0.106                  | 0.051                    | 0.055             |
| Mean                               | 0                      | 0                        | 0                 |

Table 2.1 is hypothetical, as in real life we are not able to observe A and E, but it illustrates a number of things:

- Phenotypic differences are due to differences in both breeding value (A) and residual effects (E)
- Differences in A and E are unrelated, i.e. animals with positive A do not have more chance to have a positive (or negative) E. In the example, the correlation between A and E is 0
- Differences in E are generally larger than differences in A. The standard deviation is larger and consequently the variance is larger. In fact, the heritability in the example is  $0.06/0.11 = 0.48$ .
- Animals with the best phenotype do not necessarily have the best breeding value, but...
- Selecting animals on phenotype will provide animals with on average positive breeding values (there is a correlation between P and A)

It is common that differences in E are larger than differences in A. A measure for difference is variance. Variance is formally calculated as the sum of the squared deviations. The mean values for P, A and E are zero. The variances and standard deviations in the example of Table 2.1 are:

|              |                |
|--------------|----------------|
| $VP = 0.106$ | $SD_P = 0.326$ |
| $VA = 0.051$ | $SD_A = 0.226$ |
| $VE = 0.055$ | $SD_E = 0.233$ |

Hence, about 50% of the observed phenotypic differences are due to breeding value (A). This figure is called *heritability*.

$$\text{Heritability} = \text{Additive genetic variance} / \text{Phenotypic variance} = VA / VP$$

In our example, the exact heritability is  $0.051 / 0.106 = 0.482$ .

It is important in quantitative genetics to have a good understanding of distributions. Taking Table 2.1 as an example: the phenotypic SD = 0.3. In a larger population we would expect the extremes to be about 3 SD deviating from the mean, so ranging from about -1 to +1 (or between 3kg and 5 kg in trait measurement). Note 1: the example is only a small population so the extremes are closer to the mean. If we sampled many animals, the range of values would be roughly between 3.0 and 5.5 Kg). The variance in true breeding value is smaller. If many were sampled we would expect values between -0.7 and +0.7

## Estimating breeding value

Without the knowledge of true genotypes, the only information that can be used to estimate breeding value is the phenotype. For this purpose we use phenotypic differences between animals, or more specifically, phenotypic deviations. The breeding value is now estimated as a proportion of this deviation, being the proportion of total variation that is due to variation in breeding value. This proportion is equal to heritability, i.e. in our example; the proportion is equal to 0.48. Hence, when the information used is an animal's own phenotypic deviation, the breeding value is estimated as:

$$EBV = h^2 \cdot P$$

Where  $h^2$  is the symbol for heritability and P is the phenotypic deviation, i.e. how much does the individual's performance deviate from the mean (Note, this is not the same as phenotypic standard deviation!). The heritability is a population parameter, i.e. the value is constant. It can only vary between traits or between the same trait in distinctly different environments or different populations (breeds).

The reason why we use  $h^2$  rather than h is statistical. The correlation between breeding value and phenotype is equal to h. The proportion of variation explained by breeding value is  $h^2$ . This notation is equivalent to statistical modeling, where we use  $r^2$  for the proportion of variance explained by the model, and r for the correlation between observed values and predicted values.

In the example (Table 2.1) the correlations are:  $\text{corr}(A,E) = 0.00$ ;  $\text{corr}(A,P) = 0.69$  and  $\text{corr}(E,P) = 0.72$ . The latter two are equal to h and to  $\sqrt{1-h^2}$ .

Notice that we don't use the absolute phenotypic value, but we use it as a deviation from a mean, since we are only interested in differences between animals (variation). The larger the heritability of

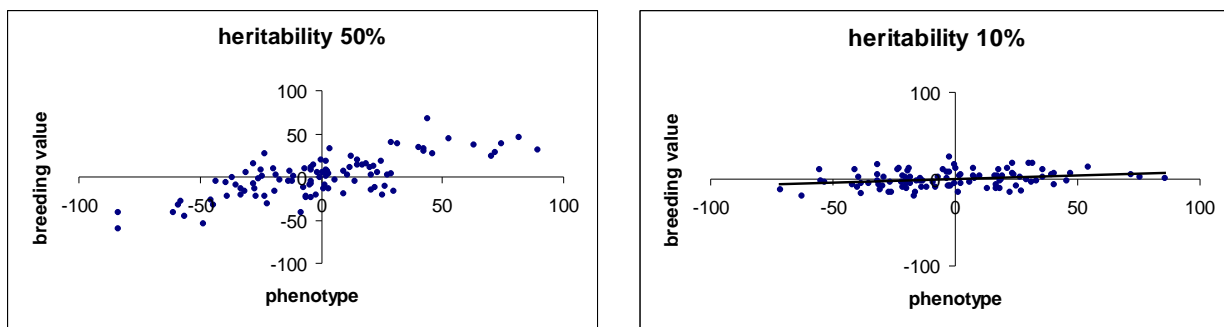
a trait, the more we believe that observed phenotypic differences are due to breeding value.

The principle of breeding value estimation is based on regression. We want to know differences in breeding value based on observed differences in phenotype. Regression is illustrated in Figure 2.1. If we regress breeding values on the phenotypic observations, the slope of the regression line tells us how much difference we have in breeding values per unit of difference in phenotype. This slope is equal to the heritability. This can be derived from quantitative genetic theory, since the slope of a regression line is:

$$b_{xy} = \frac{\text{cov}(x, y)}{\text{var}(y)} \text{ which is now equal to } \frac{\text{cov}(P, A)}{\text{var}(P)} = \frac{\text{var}(A)}{\text{var}(P)} = h^2$$

recalling that  $\text{cov}(P, A) = \text{cov}(A+E, A) = \text{cov}(A, A) = \text{var}(A)$ . This is a theoretical derivation, as in practice we can not observe true breeding values (A). However, we can estimate heritability (see in later lectures) and therefore, with the knowledge of  $h^2$  and phenotypes, we can estimate breeding values.

**Figure 2.1 The relationship between breeding value and phenotype, depending on heritability. The higher the heritability, the larger the part of phenotypic difference that is attributed to breeding value.**



Example

In Table 2.1, the best ram had a fleece weight of 4.4 kg. The mean of his contemporaries is 4.0 kg. The heritability of fleece weight is 0.48. What is the ram's EBV?

$$EBV = h^2 \cdot P = 0.48 \cdot (+0.4) = + 0.19 \text{ kg}$$

**Table 2.2 Performance of 5 rams, underlying true effects and EBVs.**

| Observed Phenotype (Fleece Weight) | P     | A     | E     | EBV   |
|------------------------------------|-------|-------|-------|-------|
| 4.4                                | 0.4   | 0.20  | 0.20  | 0.19  |
| 4.2                                | 0.2   | -0.10 | 0.30  | 0.10  |
| 4.05                               | 0.05  | 0.25  | -0.20 | 0.02  |
| 3.75                               | -0.25 | -0.05 | -0.20 | -0.12 |
| 3.6                                | -0.4  | -0.3  | -0.1  | -0.19 |

*P = Phenotypic Deviation; A = Additive Genetic effect (Breeding Value), E = Environmental Effect*

Again, Table 2.2 illustrates some important aspects of EBV

- The ranking based on EBV is not exactly the same as the ranking based on true breeding value (A), but.....
- Animals with the highest EBV have on average a higher true breeding value

- In fact, when only using animals' phenotype, the ranking on EBV is the same as ranking on P. This will change when we also use P from relatives, or when P gets adjusted for non-genetic effects such as flock, sex, birth type, age. Therefore the correlation between EBV and A is also equal to  $h$  (square root of heritability). This is the expected accuracy for an EBV when using as information the phenotype of the individual. Hence, the accuracy is 0.69.
- The variation in EBV is lower than the variation in A. See later in this lecture,
- There is a prediction error on each EBV (being the difference between A and EBV). The estimation of EBV might seem very poor for individual animals, but as a criterion for achieving genetic change the EBV is the most efficient, as will be discussed later in this topic. Also realise that with more information, especially information on progeny, the EBV will be closer to A.

In reality, we cannot see A, and therefore we will not know prediction error. The expectation of prediction error is zero (meaning that they are zero – on average). The size of the prediction error depends on the accuracy (see later in this topic). With more information available the prediction error becomes smaller, and the EBV will slowly move toward the true breeding value. With an accuracy of 1, the prediction error will be zero.

*An important thing to note is that the animal with the best EBV has no more chance to have a negative prediction error than the animal with the worst EBV, i.e. each animal has just as much chance to change upward than downward if more information becomes available.*

## Correcting for fixed effects

Estimation of breeding values has two aspects to it. This first is about deriving regression coefficients, i.e. how much about a piece of information do we attribute to breeding value. This will be discussed in more detail in the next topic. The other aspect is that breeding values should be based on fair comparisons between individuals, i.e. they should not be biased by the fact that some animals had more chance to realise a good phenotype than others. Systematic effects that affect phenotypes are called *fixed* effects, e.g. the effect of the animals' sex, its herd (flock) or management group, the season it was born in, whether it was born as a single or twin, etc. For those fixed effects that are observable we can do a correction. This is in contrast to the *random* environmental effects "E" which we cannot observe, and cannot correct for.

Correction for fixed effects will be discussed in more detail when we discuss linear models and BLUP. However, the principle is that for a fair comparison among animals, we need to consider their observed phenotype as a deviation of an expected mean, i.e. as a deviation of a contemporary group mean.

A simple example can illustrate this concept.

| Ram  | Phenotype | Flock Mean | P    | EBV ( $h^2 = 0.50$ ) |
|------|-----------|------------|------|----------------------|
| Bert | 4.8       | 4.2        | +0.6 | +0.3                 |
| John | 4.5       | 3.8        | +0.7 | +0.35                |

So although Bert is a heavier cutting ram, its EBV is lower. His heavier fleece weight was mainly because of being in a good flock. Correcting for fixed effects makes sense, and is the main reason why selecting on EBV is more sensible than selecting on phenotype. However, at the same time this creates a lot of confusion as breeders might wonder why their best looking animals do not have the best EBV.

Note that it is possible that Bert's flock used better rams, and in that case the correction would be unfair. However, in the BLUP procedure, there is a joint correction for flocks and rams used in these flocks, and BLUP would take such a thing into account (this was one of the main reasons why BLUP was introduced in dairy cattle evaluation in the 1970s).

Besides correcting for fixed class effects, as above, we can also correct for continuous effects such as time since last shearing. In the example below, we use a correction for fleece weight of 0.4 kg/month (assume that we have first estimated from data that fleece growth per month is 0.4 kg). Each animal is adjusted to a 12 month fleece weight, by adjusting their fleece weight according to time since last shearing. For example, Frank was shorn 11 months ago, and so his phenotype of

4.5 kg is only equivalent to 11 months of growth. We expect he would cut 0.4 kg heavier if he had 12 months of growth. Therefore, he gets 0.4 kg extra credit, and we put his corrected fleece weight at 4.9 kg. Ben however hasn't been shorn for 13 months and so his phenotype of 4.6 kg has the equivalent of 1 month's extra wool growth (0.4 kg). Thus 0.4 kg is removed from his fleece weight to give his corrected phenotype.

| Ram   | Phenotype | Mths since last shorn | Corrected Phenotype <sup>1</sup> | P    | EBV ( $h^2 = 0.50$ ) |
|-------|-----------|-----------------------|----------------------------------|------|----------------------|
| Frank | 4.5       | 11                    | 4.9                              | +0.9 | +0.45                |
| Ben   | 4.6       | 13                    | 4.2                              | +0.2 | +0.05                |

<sup>1</sup>assuming that fleece growth per month is 0.4 kg.

## 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), - see Topic 3- 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. Table 2.3 shows examples of accuracy. It illustrates that:

- Accuracy is higher when more information is used, e.g. from 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 2.3 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 (see Topic 3). Here we only give a simple example for the derivation of accuracy of an EBV based on own performance:

<for reference only>

$$EBV = I = h^2P$$

giving  $r_{IA}$  = is the correlation between estimated and true breeding value is  $\text{corr}(h^2P, A)$ .

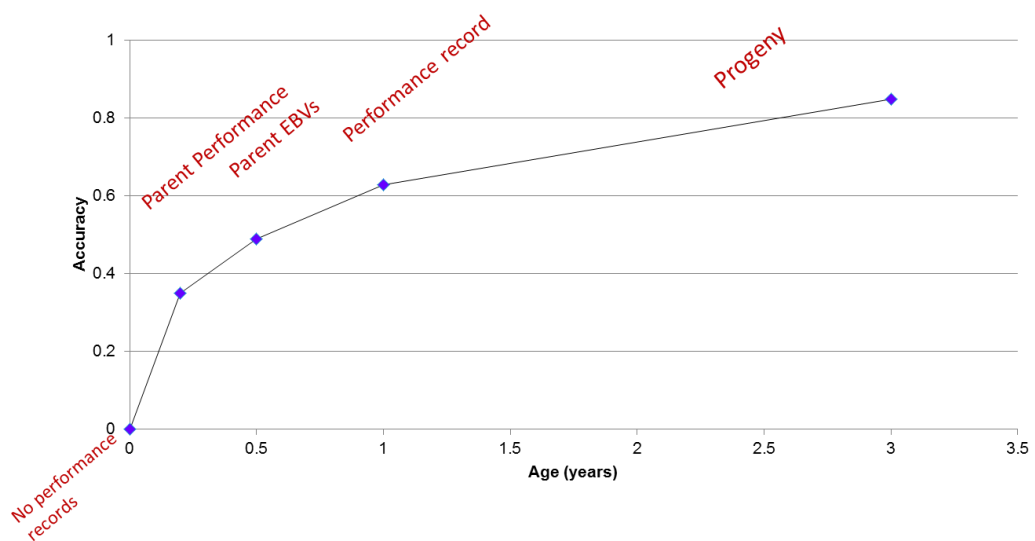
$$= \frac{Cov(h^2P, A)}{\sqrt{V(h^2P)VA}} = \frac{h^2V_A}{\sqrt{h^4V_P V_A}} = \frac{h^2V_A}{\sqrt{h^2V_A V_A}} = \frac{h^2}{\sqrt{h^2}} = h$$

If the heritability is higher, EBV's based on own performance records become more accurate. You can check in the example of Table 2.2 that the correlation between EBV and A is indeed 0.69

It is good to realize that the accuracy of an animal's EBV increases as it gets older, as more and more information will become available, see the example below (assumed  $h^2 = 0.25$ ). We can also see in Table 3a how a genomic test that can be done early in an animals life, adds more to the accuracy when there is not much other information available.

**Table 2.3a Accuracies of EBV depending on source of information used.**

| info used          | heritability 10% |          | heritability 30% |          |
|--------------------|------------------|----------|------------------|----------|
|                    | no genomics      | genomics | no genomics      | genomics |
| DNA test only      | 0                | 0.22     | 0                | 0.39     |
| Parents records    | 0.22             | 0.31     | 0.39             | 0.51     |
| ....+ 20 half sibs | 0.35             | 0.40     | 0.49             | 0.58     |
| .....+own info     | 0.45             | 0.48     | 0.66             | 0.69     |
| .....+20 progeny   | 0.66             | 0.67     | 0.84             | 0.85     |
| .....+100 progeny  | 0.86             | 0.86     | 0.95             | 0.95     |



# 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

In general, the variance of EBVs can be written as depending on accuracy as well as additive genetic variance:

$$\text{Var}(\text{EBV}) = r_{IA}^2 V_A$$

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).

Therefore:  $\sigma_{\text{EBV}} = r_{IA} \sigma_A$  is the standard deviation of EBVs

It is useful to consider the following

If  $r_{IA} = 0$  then  $\text{Var}(\text{EBV}) = 0$ : accuracy is zero, there is no information and all EBVs have the same value (=0)

If  $r_{IA} = 1$  then  $\text{Var}(\text{EBV}) = V_A$ : Accuracy is one is a perfect accuracy (the correlation between EBV and BV is one); 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 young 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:

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

[ ... as  $h^2 = V_A / VP$   $VP = V_A / h^2$  ]

In Table 2.2, the heritability is equal to 0.48, and the accuracy based on own performance EBV is  $\sqrt{0.48} = 0.69$ .

We expect  $\text{Var}(\text{EBV}) = 0.48 V_A$  and  $\text{SD}_{\text{EBV}} = 0.69 \sigma_A$ .

Remember that  $\sigma_A = 0.226$  and  $V_A = 0.051$ .

Using the numbers from Table 2.2 we obtain  $\text{Var}(\text{EBV}) = 0.024$  and  $\sigma_{\text{EBV}} = 0.156$ .



## Prediction Error Variance

The Prediction Error Variance (PEV) gives insight into the amount of error, and therefore the distribution of the true breeding value – see Figure 2.2. For this distribution we use the standard deviation, which is the Standard Error of Prediction (SEP). The SEP of a breeding value is important as it can be used when assessing a change in an EBV when more information becomes available – how much can they still change?

The prediction error variance (PEV) depends again on accuracy and is calculated as:

$$V(A - EBV) = (1 - r_{IA}^2)V_A$$

And the Standard Error of Prediction (SEP) is  $\sqrt{PEV}$

$$= \sqrt{(1 - r_{IA}^2)V_A}$$

Table 2.4. Example with phenotypes, underlying effects A and E, EBV and Prediction error (PE) and their variances.

| Observed Phenotype<br>(Fleece Weight) | P     | A     | E     | EBV   | PE    |
|---------------------------------------|-------|-------|-------|-------|-------|
| 4.4                                   | 0.4   | 0.20  | 0.20  | 0.19  | -0.01 |
| 4.2                                   | 0.2   | -0.10 | 0.30  | 0.10  | 0.20  |
| 4.05                                  | 0.05  | 0.25  | -0.20 | 0.02  | -0.23 |
| 3.75                                  | -0.25 | -0.05 | -0.20 | -0.12 | -0.07 |
| 3.6                                   | -0.4  | -0.3  | -0.1  | -0.19 | 0.11  |
| Standard Deviation                    | 0.326 | 0.226 | 0.235 | 0.156 | 0.163 |
| Variance                              | 0.106 | 0.051 | 0.055 | 0.024 | 0.027 |
| Mean                                  | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Check

$$r_{IA} = 0.69$$

$$SD_{EBV} = r_{IA} \sigma_A = 0.69 * 0.226 = 0.156$$

and  $SEP = \sqrt{(1 - r_{IA}^2)V_A} = 0.163$

It is useful to consider the following:

- when there is no information, and accuracy is 0: all EBV's will be 0 and the variance of the prediction error is equal to  $V_A$ .
- when there is full information, the EBV will be equal to the true BV and the variance of the prediction error will be 0.
- PEV becomes smaller with higher accuracy
- $Var(EBV) + PEV$  is equal to  $V_A$ :  $r_{IA}^2 V_A + (1 - r_{IA}^2)V_A = V_A$ .

The prediction error of an EBV is important as it gives us an indication of the difference between the TBV and the EBV. This is important for example to answer questions like: how much could an EBV still change if we obtain more information on the animal. Changes in EBV's are not good for the industry's confidence in the genetic evaluation system. However, we have to realise that an EBV is never exact, unless the accuracy is 100%. We expect the TBV to be the same as the EBV, but there is a certain probability that it will be a bit different. The probability distribution of the TBV, given an EBV looks like Figure 2.2.

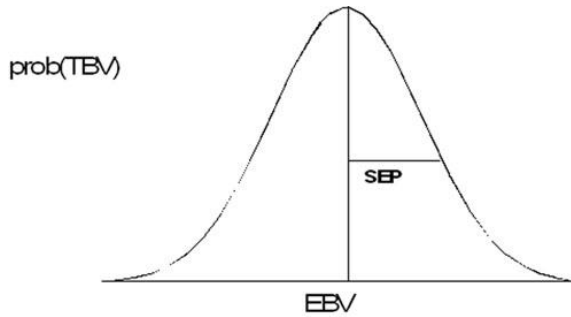


Figure 2.2 The distribution of true breeding value (TBV), given the estimated breeding value (EBV). The SD of the distribution is equal to the standard error of prediction (SEP).

Take again the example where the best ram had an EBV for fleece weight of +0.19.

$$V_A = 0.051 \text{ kg}^2; \sigma_A = 0.226 \text{ kg}; r_{IA} = 0.69 \text{ and } r_{IA}^2 = 0.48.$$

$$\text{Then the SEP} = \sqrt{(1-0.48) \cdot 0.226} = 0.16$$

If we would take a 95% confidence interval for the TBV we would take  $EBV \pm 1.96 \times SEP$

TBV is  $+0.19 \pm 1.96 \cdot 0.16$  is a range from -0.13 to 0.51.

We see that even with a high heritability (and with  $r_{IA} = 0.69$ ), the TBV can deviate still quite a bit from the EBV and it is not surprising that EBVs still change as more information becomes known about an animal. Such information can come from relatives such as the sire, the dam, half and full sibs or progeny (and possibly even more distant relatives). Using information from progeny is the only way to obtain EBV's with a very high accuracy. For example, dairy bulls often have thousands of progeny, and the accuracy of their EBV is close to 1. Such EBV's are not expected to change anymore in subsequent genetic evaluations. In fact, if EBV's are changing more than expected based on the PEV, this could be a sign that something is wrong with the evaluation procedure, e.g. a wrong model is being used. However, Table 2.4 and Figure 2.3 below show that even accurate EBVs can still change.

Table 2.4 gives confidence intervals about the best ram's breeding value for different accuracies. It illustrates that, even with high accuracy, confidence intervals are still quite large. However, one always needs to take Figure 2.2 in mind, showing that the highest probability about TBV is around the EBV.

**Table 2.5. 95% Confidence interval of TBV of the best ram (EBV = +0.19) for different accuracies ( $\sigma_A = 0.226$ ).**

| Accuracy                       | SEP  | lower thresh. | upper thresh. |
|--------------------------------|------|---------------|---------------|
| 0.51 (parent average)          | 0.19 | -0.19         | 0.57          |
| 0.67 (PA + Own Performance)    | 0.17 | -0.14         | 0.52          |
| 0.85 (PA + OP + 20 progeny)    | 0.12 | -0.04         | 0.42          |
| 0.95 (PA + OP + 100 progeny)   | 0.07 | 0.05          | 0.33          |
| 0.99 (PA + OP + 600 progeny)   | 0.03 | 0.13          | 0.25          |
| 0.999 (PA + OP + 6000 progeny) | 0.01 | 0.17          | 0.21          |

Table 2.4 might suggest that it is very useful to have a high accuracy. Also Figure 2.3 shows that we can only have a very high confidence about EBV when accuracies are extremely high.

However, it is wrong to conclude that only breeding animals with the highest accuracy should be used. In the next section we will see that response to selection depends linearly on accuracy. In optimising breeding programs, where we need to balance accuracy versus generation interval, it is often efficient to select younger animals and reduce generation interval while accepting lower selection accuracy.

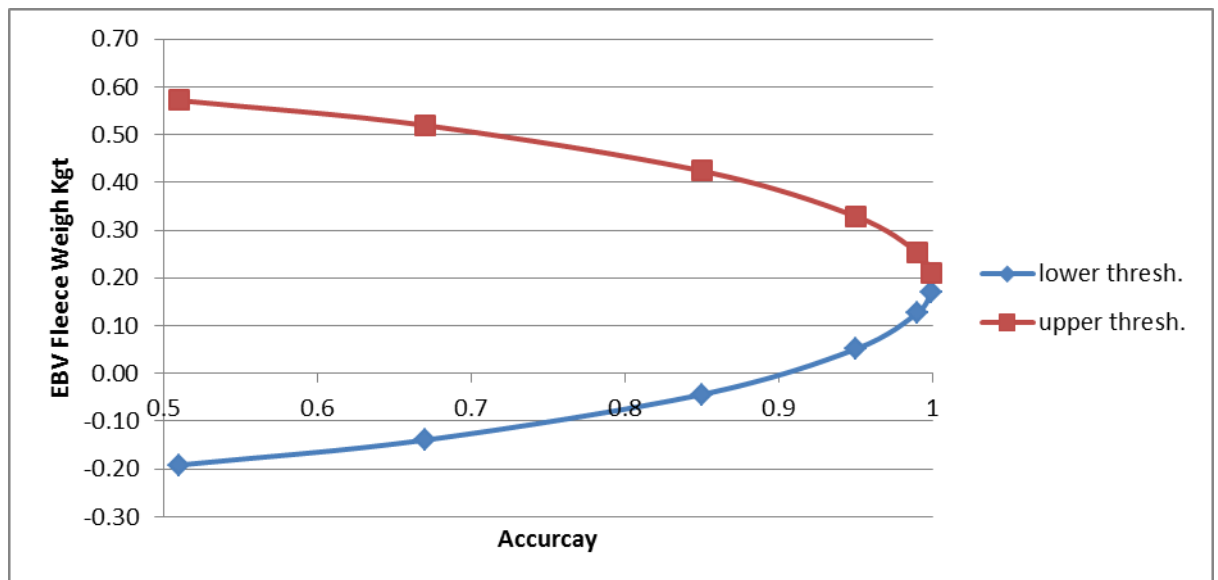


Fig 2.3. 95% Confidence Interval of True Breeding Value of best ram ( $EBV = +0.19$ ) for different accuracies ( $\sigma_A = 0.226$ ).

## Response to selection

Ultimately the main issue about EBVs is the response we may expect from selecting on them.

The expected value of a selected group of animals - when selecting on EBV:

Expected average EBV:  $i \cdot \sigma_{EBV}$

Expected average true BV:  $i \cdot r_{IA} \cdot \sigma_A$

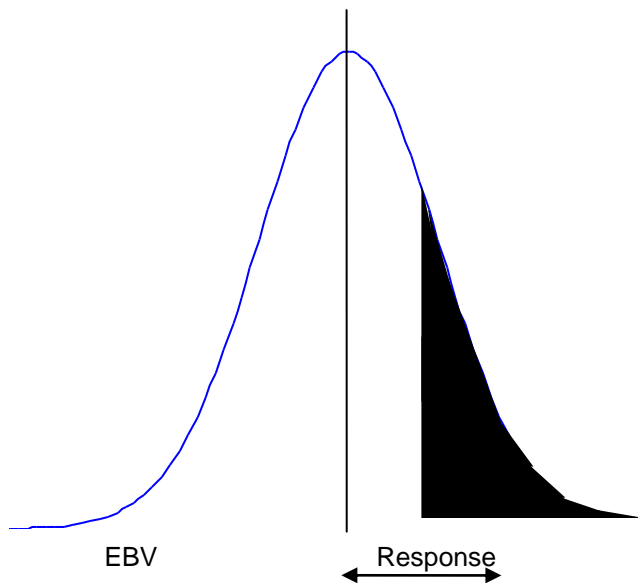
Because the expected value of an EBV is equal to the true BV, see Figure 2.2.

The expected breeding value of a selected group is equal to *selection response*.

Note that selection response depends directly (linearly) on accuracy.

The response is equal to the selection intensity multiplied by the SD of the EBV.

$$R = i \cdot \sigma_{EBV}$$



More generally:

$$\text{Response} = \underset{\text{Intensity}}{i} \quad * \quad \underset{\text{Accuracy}}{r_{IA}} \quad * \quad \underset{\text{Genetic SD}}{\sigma_A}$$

Often there is more information available on the selection candidates of one sex, and the accuracy of EBV's may differ between sexes. Also, the selection intensity will differ. Furthermore, we are interested in a response *per year* rather than per generation. A more appropriate formula to predict selection response is therefore:

Response per year

$$R_{yr} = \frac{i_m \sigma_{EBVm} + i_f \sigma_{EBVf}}{L_m + L_f} = \frac{i_m r_{IAm} + i_f r_{IAf}}{L_m + L_f} \sigma_A$$

where subscripts m and f refer to males and females.

Note that with a lot of information on each animal,  $\sigma_{EBV}$  increases and so response increases. In other words, the response to selection is directly linked to the accuracy of EBV. It makes sense therefore, to increase the accuracy of EBV by including relatives' information. This is particularly important if we select on traits with low heritability, since selection on own phenotype only (mass selection) is not very accurate in that case. Also, the use of family information can be very useful for traits which can be measured on one sex only, or they are measured very late in (or even after!) life (e.g. longevity, carcass traits).

### *Predicted progeny performance based on parental EBV*

Expected breeding value of offspring:

$$EBV_{\text{offspring}} = \frac{1}{2}EBV_{\text{sire}} + \frac{1}{2}EBV_{\text{dam}}$$

Note that the expected phenotypic deviation and the expected genotypic value are equal to  $EBV_{\text{offspring}}$ , as progeny dominance deviation and environmental deviation are unknown and have 'expectation' zero. The term expectation means the average value of many possible outcomes of this event, of course one particular outcome has a good chance to deviate somewhat around that expected value.

Sometimes it is stated that the heritability of an EBV is equal to 1.

This depends on the definition of heritability. The relevant definition in the context of selection response is:

"Proportion of parental superiority (in EBV) transmitted to progeny"

This is equal to the regression of true breeding value on EBV (how much difference do we expect between progeny for a certain difference of EBV) (Figure 2.4).

$$b_{A,EBV} = \frac{\text{cov}(A, EBV)}{\text{var}(EBV)} = \frac{r_{IA}VA}{r_{IA}VA} = 1$$

A selected animal is expected to pass half of this EBV superiority on to its progeny independent of the accuracy of that EBV.

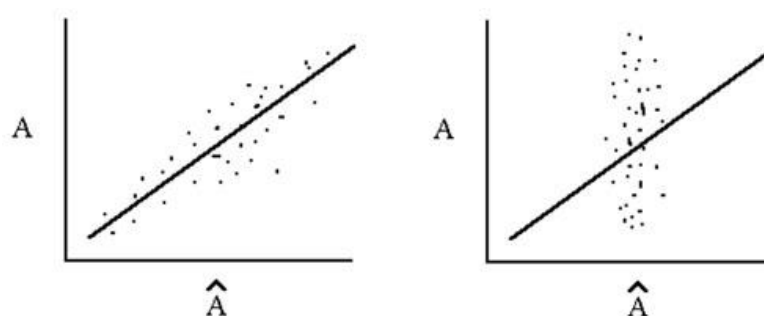


Figure 2.4 Regression of true breeding value  $A$  on EBV with high accuracy (left graph) and low accuracy (right graph) is (should be) equal to 1 in both cases.

Note that  $b_{A,EBV}$  (the slope) is the same for high and low accuracy. The variance of inaccurate EBV's is very low, and therefore the selection superiority based on inaccurate EBV's will not be very high.

An interesting problem is the following. Suppose that two rams have the same EBV, however, ram A has an accuracy of 95% (based on a progeny test) whereas ram B has an accuracy of 50% (based on parent average). Which ram should be selected?

Most people would vote for animal A. However, both animals have the same expected value for their progeny. The range around this expected value is higher for animal B. However, progeny have just as much chance to be better than their expected value than to be worse. Therefore, whether you choose A or B depends on your attitude towards risk. A breeder that is interested in breeding the very best ram might be more interested in animal B, as he has more chance that his best son will be high. A commercial producer might be more interested in reducing risk and go for animal A.

It might seem that EBVs are not of much value, as the confidence intervals about any prediction based on the EBV seems so large. However, again you should be reminded that ultimately, selection response depends linearly on selection accuracy. Table 2.5 illustrates a small simulation, where 10 rams are ranked on their EBV based on parent average. It shows their actual realised fleece weight as well as true BV and EBV based on own performance. For individual cases, there seem to be huge discrepancies. However, when selecting the top 50% (best 5), we see indeed that selection response depends on selection accuracy, but even inaccurate EBVs provide a worthwhile response (Table 2.6). Predicted response (remember how this can be predicted?) and realized response are not exactly the same as for a small sample, the effect of individual outliers is relatively large.

**Table 2.6 EBV based on parent average (EBV\_PA), realised phenotype, EBV based on own performance (EBV\_OP) and true breeding value for 10 rams for fleece weight.**

| EBV_PA | Phenotype | EBV_OP | True BV |
|--------|-----------|--------|---------|
| 0.34   | 4.14      | 0.07   | 0.22    |
| 0.29   | 4.51      | 0.25   | 0.56    |
| 0.26   | 4.04      | 0.02   | 0.32    |
| 0.23   | 4.20      | 0.10   | 0.00    |
| 0.18   | 3.54      | -0.23  | 0.17    |
| 0.14   | 4.43      | 0.22   | 0.21    |
| -0.02  | 4.67      | 0.34   | 0.79    |
| -0.02  | 4.47      | 0.24   | -0.08   |
| -0.23  | 3.14      | -0.43  | -0.38   |
| -0.40  | 3.91      | -0.05  | -0.02   |

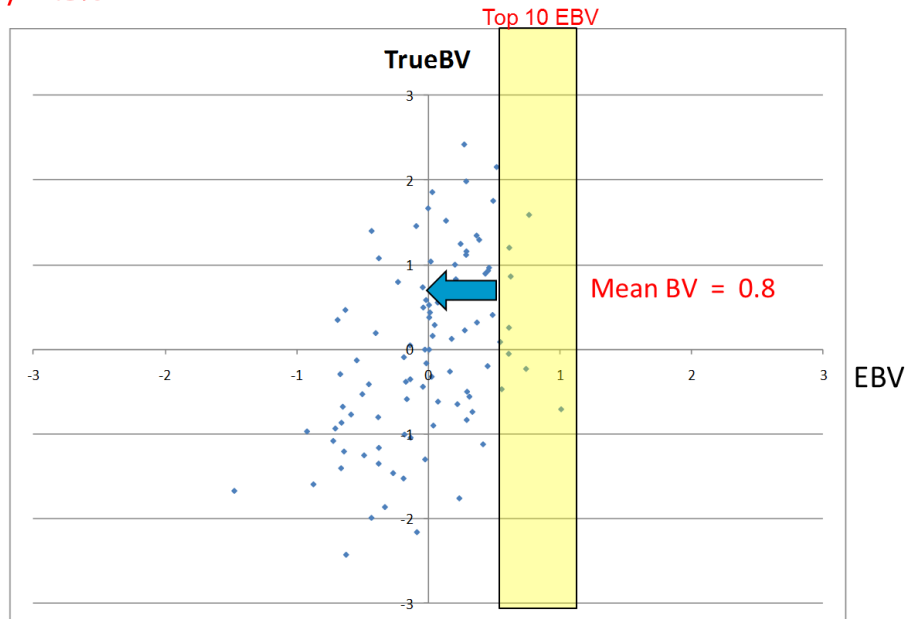
**Table 2.7 Selection response based on EBV based on parent average (EBV\_PA), EBV based on own performance (EBV\_OP) or true breeding value (TBV) for top 50% of 10 rams for fleece weight ( $\sigma_A = 0.5$ ;  $h^2=0.5$ ).**

| Selection On | Accuracy | Predicted Response <sup>1</sup> | Realised Response |
|--------------|----------|---------------------------------|-------------------|
| EBV_PA       | 0.45     | +0.13                           | +0.25             |
| EBV_OP       | 0.71     | +0.20                           | +0.30             |
| TBV          | 1.00     | +0.28                           | +0.42             |

<sup>1</sup> Response is calculated as the average TBV of the top 50% when ranking is based on each of the selection criteria.

A graphical representation of the relationship between accuracy and response is given in Figure 2.5

Accuracy = 45%



Accuracy = 90%

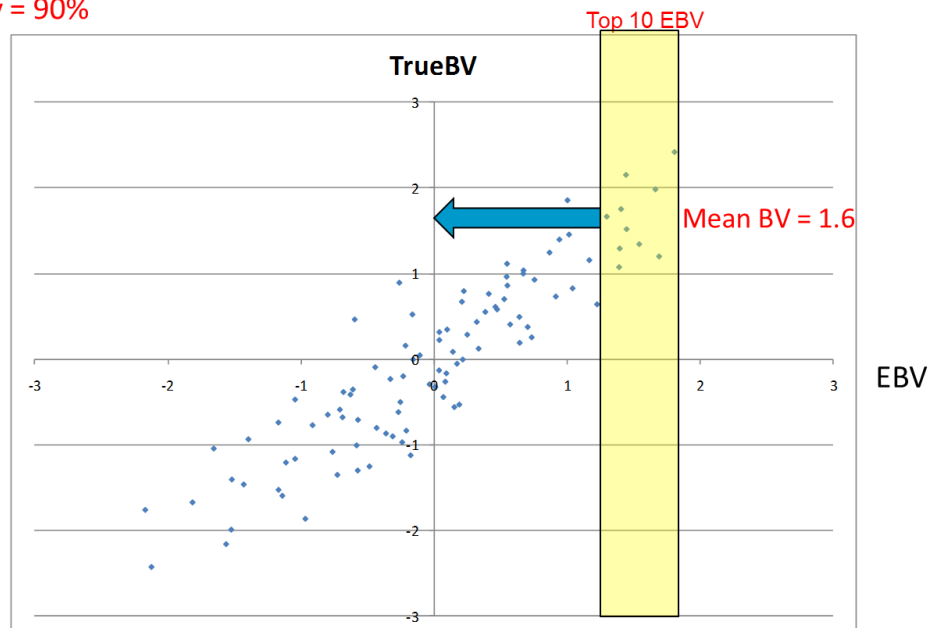


Figure 2.5. Relationship between accuracy and selection response for low and high accurate EBV. Selection response is the average true breeding value (along y-axis) of the top 10% individuals selected on Estimated breeding value (EBV x-axis). Note also that the variance in TBV is the same, but the variance in EBV increases with accuracy.

---

## Summary

Estimated Breeding Value - EBV is an estimate of the true breeding value, based on phenotypic differences between animals. EBVs can be estimated based on own performance, but also on performance of relatives. EBV is estimated as a regression of breeding value of phenotypic differences. The higher the heritability, the more of these differences are attributed to EBV.

The accuracy of EBV is a key property. It is an indication of the selection efficiency. The accuracy can range generally from 35-99%, with higher values for traits that have a higher heritability and when more information is used, e.g. more relatives' information. To achieve accuracies above 70 %, information on progeny is required. The accuracy is used to predict the variation among EBV, the selection differential, and the prediction error.

Best Linear Unbiased Prediction (BLUP) is the name of a method that is used worldwide to give estimated breeding values (EBV's) for commercially important traits. BLUP uses all available information to estimate an animal's EBV and corrects for fixed effects. Further topics will provide more detail on BLUP and focus on the principles to understand the role of BLUP EBVs in animal breeding programs.

---

## References

Simm, G. 2000, *Genetic improvement of cattle and sheep*, Farming Press, Miller Freeman, UK.



## Glossary of symbols and terms

|                  |  |   |
|------------------|--|---|
| $r_{IA}$         | Accuracy of EBV                        | The correlation between true and estimated breeding value   |
| $V_A$            | Additive genetic variance <sup>1</sup> | The variance in a trait due to the combined effects of genes with additive action                                       |
| $h^2$            | Heritability                           | Proportion of phenotypic difference that is transmitted to offspring  |
| EBV or $\hat{A}$ | Estimated Breeding Value               | Prediction of the additive genetic merit of an animal   |
|                  | Fixed effects                          | Systematic effects that affect the phenotype of an animal eg. its sex or its herd or flock                              |
| $V_P$            | Phenotypic variance <sup>1</sup>       | Variance in a trait due to the combined effects of genes and the environment  |
| $\sigma_P$       | Phenotypic Standard Deviation          | Degree of differences in phenotype  |
|                  | Random effects                         | The effects that impact on the phenotype of an animal that cannot be observed or corrected for eg environmental effects |
|                  | Regression <sup>1</sup>                | A procedure that measures the direction and strength of an association between two characters                           |
|                  | Residual variance <sup>1</sup>         | Variance in a trait due to non-genetic effects  |
| TBV or $A$       | True Breeding Value                    | The additive genetic merit of an animal. Cannot be directly, only estimated (EBV)                                       |
| PEV              | Prediction Error Variance              | Standard Deviation of the error on an estimated breeding value  |
| Var()            | Variance                               | Amount of variation about a variable  |
| Cov()            | Covariance                             | Amount of difference in one variable that can be predicted from variation in another.                                   |

<sup>1</sup> Glossary terms taken from Simm (2000).