Accuracy of EBV's

Kim Bunter, Bruce Tier and David Johnston

Introduction

In previous versions of PIGBLUP, breeders had no knowledge of the accuracy with which specific estimated breeding values (EBV's) were calculated. This was generally of no concern. Replacement breeding stock were selected based on EBV's knowing that young animals will generally have EBV's with similar accuracies. Consequently, even if accuracy was an issue for the breeder, there would be little room to descriminate between young animals based on accuracy.

However, the development of an across herd genetic evaluation system has resulted in software to routinely calculate accuracies for EBV's. This is because individuals compared across herds may range in age and the amount of information available to predict their EBV's, which will be reflected in their accuracies. Consequently, breeders wishing to reduce their risk of importing a genetically inferior animal from another herd could make use of information provided by accuracies of EBV's to help limit this risk. Calculation of accuracies will subsequently be incorporated into PIGBLUP V4.01 as an additional feature. What accuracies tell you, and how they may be used in your breeding program, will be discussed in this document.

What Does Accuracy Mean?

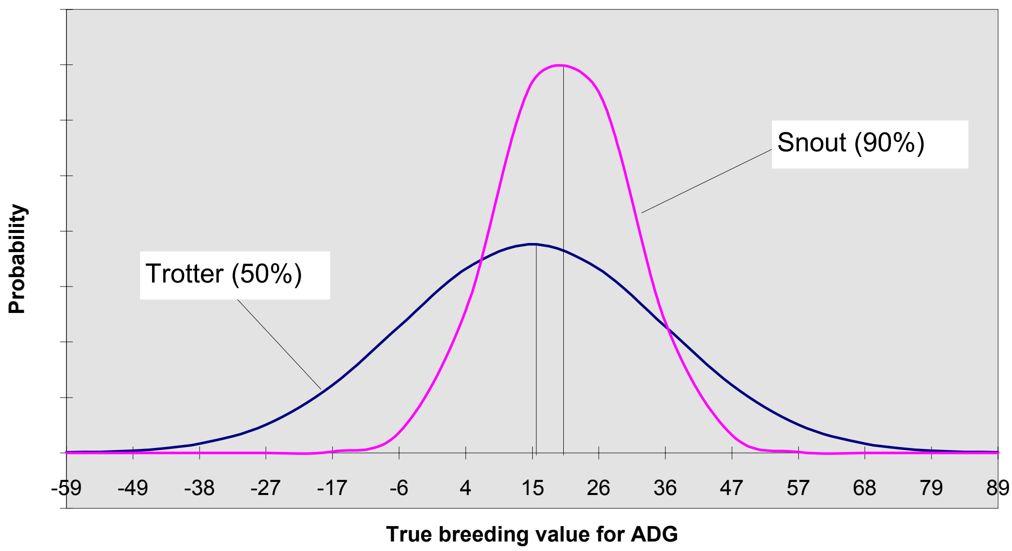
An EBV is a comparative measure of an animal's genetic merit; how good (or bad) this animal is relative to another included in the analysis. The *accuracy* of an EBV is a measure of the quantity, and to some extent the quality, of information used to estimate each animal's EBV. Given that we can never know an animal's true breeding value, accuracy gives us some idea of how the estimated breeding value may differ from an animal's true breeding value. It is a confidence measure, indicating how likely an EBV is to change as new information is used, and by how much it might change. However, accuracy does not indicate in which direction an EBV may change when more data is added - it is just as likely to go up as down.

Figure 1 illustrates 2 boars (Snout and Trotter) whose estimated breeding values for ADG are +15 and +19. However, Snout's EBV has an accuracy of 90%, whereas Trotter's EBV is estimated with an accuracy of 50%. Consequently, Snout's EBV is likely to be closer to his true breeding value than Trotter's, and is expected to change less with the use of additional information. A highly likely range for Snout's true breeding value (± 1 standard deviation) is between 8 and 30 *grams/day*, whereas Trotter's equivalent range is between -6 to 36 *grams/day*. The extreme possible EBV's shown for both boars are only likely with a very low probability.

If EBV's were 100% accurate, you would expect that the EBV calculated = true breeding value. **NB. Regardless of the accuracy, an EBV is still the best estimate of an animal's**

breeding value given the available information. To some extent, EBV's are automatically adjusted for their accuracy. For example, extreme EBV's are less likely for individuals with little available information.

Figure 1: Probability of true breeding value for ADG when boars differ in both EBV and accuracy.



What Affects Accuracy?

Accuracy of any particular EBV is affected by a number of factors. It may not always be intuitively obvious why certain animals have the accuracies that they are calculated to have! Consider the following points.

1. The heritability of a trait

Heritability of the traits we observe affects the value of any single record for predicting EBV's. Consequently, heritability also affects the accuracy with which EBV's are estimated. In Table 1, accuracies arising from varying amounts of information contributing to EBV's are illustrated. Higher accuracies are always associated with traits of higher heritability for a given amount of information.

In addition, heritability affects the relative value of information from different sources for predicting EBV's. For lowly heritable traits, the relative value of information from progeny compared with the animals own observation is higher. For example, if seven full-sibs of an individual are recorded for a trait (h2 = 0.1; no litter effect) the EBV for this unrecorded individual is estimated with a higher accuracy than if the animal had its own record. However, this is not the case when the heritability is increased to 0.25 (see Table 1). At higher heritabilities, information from the records of relatives is comparatively less valuable for estimating a breeding value than the animal's own record.

Table 1: Approximate accuracies (%) where source of data differs, for two different heritabilities, and in the absence ($c^2 = 0$) or presence ($c^2 > 0$) of common litter effects

Information available	Heritability		
	0.1	0.25	
	c^2=0	c^2=0	c^2=0.15
Individual performance record only	32	50	50
Individual + 7 full-sibs(FS)	43	61	59
7 full-sibs	32	46	41
Individual + 7FS + 32 half-sibs(HS)	48	64	62
7 full-sibs + 32 half-sibs	40	51	47
8 progeny records	35	48	41
32 progeny records	59	74	67
Individual + 32 progeny records(P)	63	78	73
Individual + 7FS + 32HS + 32P	68	81	77
Individual + Dam (or Sire)	35	53	53
Individual + D + S + 7FS + 32HS + 32P	69	81	78

Table 1 also illustrates the likely range in accuracies for common traits. For example, sires would need several daughters recorded for number born alive (NBA) to have a relatively accurate EBV for this trait. This accumulation of data is also associated with a substantial time lag between the sires selection and accurate knowledge of his EBV for NBA. On the other hand, individuals recorded for average daily gain (ADG) or backfat (BF) can achieve quite high accuracies for their EBV's if their litter mates and parents, and/or paternal half sibs are recorded. The presence of common litter effects has a relatively small effect on accuracies overall.

2. The number of contemporaries

Contemporaries are animals which are reared and performance tested under the same conditions. The larger a group of contemporaries, the more useful is the performance data which is obtained. This is because the number of head to head contrasts between individuals are increased with larger group sizes. However, within a group of contemporaries, related individuals will reduce the value of some observations.

At the extremes, if only a single animal is present within a contemporary group, its records will be of no use for estimating its own or parental EBV's (who is it compared with?). Consequently, its EBV's are the average of the parental EBV's, with its accuracy determined by parental accuracies. Further, if a contemporary group represents progeny from only one sire, their records will not contribute to the sire's EBV (which sire is he compared with?), but are still useful for estimating their own and their dam's EBV's. The accuracy of a sire's EBV will not be improved by more progeny if these are not compared with progeny from other sires.

The value of a sire's (or dam's) progeny in a contemporary group is given by the general formula:

$$\text{Effective progeny number (EPN)} = n(N-n)/N$$

where n is the number of sire's progeny in the group, and N is the total size of the group. The EPN is the effective number of direct comparisons between a sire's progeny and those of another sire. EPNs for dams are generally much lower than those for sires, reflecting that the majority of progeny information is provided through male paths. If n=N, there are no effective comparisons of progeny from different sires (dams). As noted above, EPNs are also reduced by relatives compared in the same group.

3. Information from other observed traits

Information from one trait may contribute to the EBV of another trait if the two traits are genetically correlated. Increasing trait heritability or the genetic correlation between traits will increase the value of information from one trait for predicting merit in the other. If both traits are recorded, accuracies for both traits will be improved.

4. Information from repeated records

Repeated performance of the same trait will allow more accurate estimation of EBV's. How valuable these additional records are depends on the traits repeatability. If repeatability is low (as for NBA), then additional records will continue to significantly improve the accuracy of EBV's for this trait. For highly repeatable traits, there is little benefit to increasing the number of records taken. In Table 2, it can be seen that when repeatability equals 0.17 (as assumed in PIGBLUP), use of additional records per sow for this trait will improve the accuracy of her EBV for NBA. However, by record 7, little is added to accuracy by including additional records in the analysis. If the repeatability were to be significantly higher for this trait (eg. r=0.5), the upper limit to useful additional records is reduced.

Table 2: The influence of repeated records on accuracy (%) of EBV's for NBA, illustrated for different repeatabilities

Number of records	Repeatability = 0.17	Repeatability = 0.50
1	32	32
2	42	37
3	48	39
4	53	40
5	56	41
6	59	41
7	61	42
8	62	42
9	64	42

10	65	43

5. **Accuracy of the parents' EBV's**

In the absence of records for an individual or its descendants, parental accuracies will determine the accuracy of the EBV for the individual. In this case, the individual's EBV will be the average of its parent's EBV's. However, the accuracy of the EBV is not a simple average, and for independent parents is given by:

$$Accuracy_{prog} = \frac{\sqrt{acc_{sire}^2 + acc_{dam}^2}}{2}$$

As parental accuracies already account for trait heritability, predicted accuracy of unrecorded progeny's EBV's are independent of trait heritability (Table 3).

Table 3: Approximate accuracies (%) based on parental accuracies

Information available	**Heritability**	
	0.1	**0.25**
Sire (70%) and Dam (40%)	40	40
Sire (90%) and Dam (70%)	57	57

6. **The number of progeny recorded**

As the number of progeny recorded increases, the accuracy of a parent's EBV increases (refer Table 1). How much parental accuracy increases depends on trait heritability. The value of information from progeny relative to the contribution from an animals own record increases as heritability decreases.

7. **The number of other descendants recorded**

Similar to point 5, records from other descendants (eg. grand-offspring) may contribute to the accuracy of individual EBV's. For animals without their own record and with few progeny recorded, this source of information may be important.

When you consider the above points, and given the variety of sources of information which contribute to accuracy, it is possible for:

- animals to have the same traits recorded but different accuracies
- animals to have similar numbers of progeny but different accuracies
- animals to have different numbers of progeny but similar accuracies
- a particular amount of information (eg. own record) to contribute more or less to accuracy of an EBV depending on how much information was already known.

In addition, it is important to remember that the accuracy of EBV's calculated is simply a function of the total amount of effective information supplied to the analysis. Accuracies give no indication of the integrity of the information. For example, failure to record management groups or use of incorrect records (which can result in biased EBV's), would not be detectable as a lower accuracy for EBV's affected by this incorrect data.

How will accuracy be calculated in PIGBLUP V4.01?

Calculating true accuracies is computationally very demanding, requiring the inversion of large matrices. Consequently, accuracies reported from various genetic evaluation systems are generally approximations of true accuracies. In PIGBLUP V4.01, approximate accuracies of EBV's will be calculated for all animals (boars, sows and gilts) for each ticked trait, and reported as percentages ranging from 0-99%. Approximate accuracies are computed for EBV's using an algorithm which makes use of the sources of information outlined above. For each individual each source of information is converted to a measure of effective progeny numbers, added together, and then converted to an accuracy using the formula:

$$\text{Accuracy} = \sqrt{EPN / (EPN + \lambda)}$$

where: EPN is the sum of EPNs from all available information and $\lambda = (4 - h^2) / h^2$. This procedure results in accuracies which are highly correlated with true accuracies, and considerably less demanding of computing time relative to alternative methods.

How to Use Accuracies

1. Within a herd

Information on accuracies provides a tool to help manage risk. However, if you are a breeder, in general the primary consideration when selecting breeding stock should be their EBV's, **NOT** their accuracies.

As noted previously, EBV's estimated with high accuracies are most likely for animals with many relatives (especially progeny) recorded. On the other hand, young animals with limited information tend to have lower and similar accuracies, but are expected to be of better genetic merit on average than the previous generation (if you selection program is effective of course!). Overlooking these young animals in favour of those with higher accuracies would mean the selection of fewer superior animals. The result may be a reduction in possible response to selection if much emphasis is placed on individual accuracies in selection decisions. In addition, such a policy may favour young animals which already have a substantial number of relatives recorded.

In Table 4, results from an example population of 200 animals spanning five generations are presented. Selection of parents each generation was at random, and generations overlap. The trait simulated had a heritability of 0.29, so an animal with it's own (effective) record would have an EBV with an accuracy of 0.54 in the absence of other information.

Table 4: The number of individuals present in two accuracy classes in different generations

Generation	Recorded?	Number	Accuracy<0.55	Accuracy>0.55 (% of Number)
1	No	30	26	4 (13%)
2	No	54	51	3 (6%)
2<generation<=3	Yes	73	26	47 (65%)
3<generation<=4	Yes	42	27	15 (36%)
4<generation	No	1	1	0 (0%)
	Total	200	131	69 (35%)

If a breeder decided to confine selection decisions to those animals whose accuracy was greater than 0.55 (ie EBV's obtained using more information than the animals own record would provide) EBV's may only be compared between relatively few individuals. For example, in the last two generations, only 35% of the total number of animals had accuracies for their EBV's above 0.55. For accuracies greater than 0.65, only 3 individuals (1.5%) met this criterion, and these were already parents. Further, some of the best individuals were in the lower accuracy group. This example population should serve to illustrate what happens in general if you consider accuracy when making selection decisions - you limit opportunities for comparisons, and may end up choosing inferior animals to satisfy your aversion to risk. Remember - the EBV is just as likely to go UP as down with less than 100% accuracies.

Consequently, for selection within a herd, little consideration should be given to individual accuracies in general - replacement breeding stock should be selected on EBV. For smaller breeders it remains difficult to manage risk effectively. For larger breeders who are concerned about the risk of using unproven youngstock extensively, increasing the number of boars/gilts used as parents is an option (ie. Don't put all your eggs in one basket). This is because the accuracy of the mean breeding value of the group selected will be improved relative to individual accuracies. For example, three young boars have EBV's for NBA (accuracies in brackets) of +1.0(25%), +0.8(30%) and +0.5(50%). Which to select - high EBV but low accuracy versus moderate EBV with higher accuracy? If you have the capacity to use all three boars then the mean EBV of the group is 0.77, the accuracy of this group mean EBV is 84% from theory. However, even if the young boars are unrelated, the fact that they contribute to data in the same analysis will probably result in a lower accuracy of the group mean than predicted. The magnitude of this reduction will depend on how the accuracies of these boars were related.

2. Across herds

I don't want to dwell on this here! However, different strategies for using accuracies of EBV's may be suitable where selection (or purchase!) decisions occur across different herds.

For breeders wishing to source outside breeding stock, the biggest risk (apart from disease) is that animals purchased for breeding from other herds are inferior to their own, thereby reducing the effectiveness of their breeding program. The use of *across* herd EBV's along with individual accuracies may help to reduce the risk associated with introducing outside animals. For example, the average performance of progeny from high accuracy boars may be closer to that expected than for low accuracy boars. However, once again, high accuracy animals are likely those that have already been extensively used as parents. For younger animals with lower accuracies, purchasing more than one for use as a sire/dam may help reduce risk. However, this is often only an option for larger herds.

Accuracies may also be useful for other purchasers of breeding stock who are interested in meeting certain criteria (eg. market specifications). However, in order to do this effectively, the purchaser needs to know how the EBV for a boar from herd 1, for example, translates to progeny performance in his (her) own herd. This may not be possible without some trial and error initially.

Summary

The routine reporting of accuracies for EBV's was prompted by the development of across herd genetic evaluation procedures. In this scenario, it was important to give breeders the ability to discriminate between animals both in EBV and in how much information contributed to that EBV (herds differed vastly in their contributions of data to the analysis). This information could then be used to lessen the risk of importing inferior animals between herds which participate in an across herd analysis.

An offshoot of this development will be the provision of accuracies for EBV's in PIGBLUP V4.01. However, in general information on accuracies is of less benefit for within herd selection decisions. This is because breeding replacements are usually chosen from youngstock whose EBV's will be estimated with similar accuracies. Further, considering accuracy in selection decisions may result in reductions to response through exclusion of younger animals from the selection process which have less information contributing to their EBV's.