

How To Use Estimated Breeding Values to Genetically Improve Dog Guides

Presented at a Meeting of the "Original Group"
September 11-12, 2003
at
The Seeing Eye, Inc.
P.O. Box 375
Morristown, New Jersey 07963

Eldin A. Leighton, Ph.D.
The Jane H. Booker Director of Canine Genetics

Abstract

Estimated breeding values describe the relative genetic value of each member in a breeding population. They can also be used as a basis for objectively choosing which animals are the best candidates to select for breeding to produce the next generation of offspring. Livestock and plant breeders have used estimated breeding value techniques for decades to obtain genetic improvement in their breeding lines, but guide dog breeders have generally not embraced these techniques. This paper introduces techniques for computing estimated breeding values, using an example from a behavior test score. A simple example is worked through to completion, and results from this example are used to illustrate how estimated breeding values can be used as the basis for making selection decisions that will genetically improve a population over several generations of selection. The techniques described herein can also adjust for or take into account the effects of non-genetic factors, such as age, contemporary groups, or sex, that often contribute to total variation observed in a population. Issues related to computing estimated breeding values on a regular basis are discussed along with some ideas about how members of the Original Group could work together more closely to share both genetic information and germplasm among schools.

Introduction

Estimated breeding values (EBV's) are a statistical prediction of the relative genetic merit of animals in a breeding population. They are used to rank available candidates for selection as replacement breeders, so that the "best" ones can be saved for breeding while the others are not allowed to reproduce. When this principle is consistently applied across several generations of selection, genetic change will occur if the traits under selection are to some degree genetically influenced.

For the past five decades, livestock breeders have used EBV's to improve their selection decisions. They have increased volume, butterfat content, and protein in milk from dairy cows. They have substantially improved the production of lean meat from beef animals by genetically increasing weight at a year of age and by increasing the amount of muscle tissue. In pigs, they have genetically reduced the amount of fat in and around the muscle tissue while at the same time selecting for increased muscle mass and improved feed efficiency.

More recently, EBV's have been used by The Seeing Eye¹ to improve the accuracy of selection decisions among dogs. In German Shepherd Dogs, Labrador Retrievers and Golden Retrievers, the selection emphasis has been to genetically improve the ability of dogs to be

trained for work as guides while at the same time selecting to genetically improve hip quality and other aspects of general health.

Theoretical Background

The physical assessment of a dog's behavior is an assessment of how the dog performs under a standardized test environment. Phenotype is what can be seen and measured with respect to the physical trait. In addition to a behavior phenotype, a dog has many other phenotypes such as hip quality, elbow quality, mature weight and height, and conformation. Each of these phenotypes is conceptually composed of two parts: a genetic component and a non-genetic component. The genetic component was passed down to each individual dog by its parents when a sperm and an egg united to form a new zygote. The non-genetic component includes the environment in which the dog was raised and all other aspects of each dog's own existence. In many cases, the non-genetic component can be further decomposed into parts attributable to "fixed" effects (like sex, year of birth, management conditions, and age when measured) and all other effects which cannot be measured. Very simply, then, an observation on a single animal can be conceptualized as²:

$$\text{Phenotypic observation} = \text{Environmental effects} + \text{Genetic effects} + \text{Residual effects}$$

or, in terms of a statistical model, this can be written as:

$$y_{ij} = \mu_i + g_i + e_{ij}$$

Equation 1

where

y_{ij} = the j'th record observed for the i'th animal

μ_i = the identifiable, non-random (fixed) environmental effects such as management group, sex, or year of birth of the i'th animal,

g_i = the sum of the additive (g_a)_i, dominance (g_d)_i, and epistatic (g_e)_i genetic values of the genotype of the i'th animal, and

e_{ij} = the sum of random environmental effects from the j'th record of the i'th animal .

Now, by substituting g_a , g_d , and g_e for g_i , and regrouping terms in (1), the statistical model becomes:

$$\begin{aligned} y_{ij} &= \mu_i + (g_a)_i + (g_d)_i + (g_e)_i + e_{ij} \\ &= \mu_i + (g_a)_i + e_{ij}^* \end{aligned}$$

Equation 2

where y_{ij} and μ_i are defined as above, but e_{ij}^* is now equal to $(g_d)_i + (g_e)_i + e_{ij}$. Equation 2 is the basic equation used to calculate EBV's, but an implicit assumption in applying equation 2 is that μ_i is known without error. In fact, this is almost never true, so the data in hand are often used to simultaneously estimate μ_i and $(g_a)_i$.

The method used to estimate both μ_i and $(g_a)_i$ simultaneously from the same data is named best linear unbiased prediction (BLUP) in the literature³. When BLUP techniques are used, EBV's are obtained by a process that simultaneously adjusts for differences attributable to

identifiable fixed effects and accounts for all genetic relationships through the pedigree of the population.

Implicit in applying BLUP technology is that one has in-hand an estimate of heritability for each trait being considered. If two or more traits are being considered jointly, then one must also have estimates of the genetic and environmental covariances among the traits.

In the scientific literature, one will also find the term "mixed model equations" used to describe the BLUP equations routinely used to obtain EBV's. They are termed "mixed" because the equations can contain terms for both fixed and random effects. Fixed effects are terms which describe attributes of an animal that are unchanging, such as its sex or the year in which it was born. Random effects are terms that describe attributes of an animal that are estimated with some degree of error, such as the animal's breeding value (g_{ai} from equation 2, above). In matrix notation, the basic linear model can be written as:

$$Y = X\beta + Zu + e$$

Equation 3

where

Y = the vector of observed values,

β = the vector of fixed effects,

u = the vector of random genetic effects (the EBV's),

e = the vector of residual errors, and

X and Z are design matrices that relate elements in β and u, respectively, to elements in Y.

Now, leaving out many intermediate steps in the derivation, it is stated without proof (but clearly shown in the literature) that the mixed model equations for estimating breeding values for a single trait or characteristic are:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \beta \\ u \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}$$

Equation 4

where

A^{-1} = matrix inverse of A, the numerator relationship matrix, which mathematically describes all genetic relationships among animals in the pedigree, and

$\alpha = (1 - h^2) / h^2$, where h^2 is heritability of the trait under analysis. Heritability simply describes the amount of total variation in the trait arising from additive genetic differences among animals.⁴

The powerful ability of BLUP to obtain estimates of breeding values using all genetic relationships is embodied in the numerator relationship matrix, A, in equation 4, above. Examining some of the properties of this matrix will illustrate some of its features. For the sake of illustration, consider the 9 dogs in the pedigree given in Table 1.

Table 1. An example pedigree.

Animal	ID	Sire ID	Dam ID	Sex
	1	Unknown	Unknown	M
	2	Unknown	Unknown	F
	3	1	Unknown	M
	4	1	Unknown	F
	5	Unknown	Unknown	M
	6	2	3	M
	7	2	3	F
	8	5	4	M
	9	5	4	F

Now, the numerator relationship matrix among these 9 animals can be written down following well-defined rules.⁵ To begin applying these rules, individual elements in the matrix must be identified. Usual convention uses subscripts to number the rows and columns and to use a lower case a_{ij} to identify each individual cell. Thus, a_{11} cell refers to the intersection of row 1 with column 1 and refers to the single cell for animal 1. Cells with equal subscripts for i and j are said to lie on the lead diagonal of the matrix. Cells with subscripts where $i < j$ are said to lie above the lead diagonal, and cells with subscripts where $j < i$ are said to lie below the lead diagonal. Another feature of the numerator relationship matrix is that it is symmetric, which means that if it were folded along the lead diagonal, each cell in the upper-half of the matrix would fold on top of a cell in the lower-half of the matrix with exactly the same value. Said another way, symmetry implies that $a_{ij} = a_{ji}$. Finally, it must be pointed out that animals in the pedigree need to be ordered such that parents precede their offspring in the list, and for convenience here, animals are identified by numbering them from 1 to n . Now for the rules:

For $i = 1$ to n ,

For $j = 1$ to i ,

1. If both parents (s and d) of animal i are known, then
if $j < i$, $a_{ji} = a_{ij} = 0.5(a_{js} + a_{jd})$,
if $j = i$, $a_{ii} = 1 + 0.5(a_{sd})$, for $j = i$
2. If only one parent (s) is known and assumed unrelated to the mate
if $j < i$, $a_{ji} = a_{ij} = 0.5(a_{js})$,
if $j = i$, $a_{ii} = 1 + 0.5(a_{sd})$
3. If both parents are unknown and are assumed unrelated, then
if $j < i$, $a_{ji} = a_{ij} = 0$
if $j = i$, $a_{ii} = 1$

Applying these rules, the numerator relationship matrix among the animals defined in Table 1 is shown in Table 2.

Table 2. Numerator relationship matrix among the 9 animals shown in Table 1.

Animal ID's									
Animal ID	1	2	3	4	5	6	7	8	9
1	1.0	0.0	0.5	0.5	0.0	0.25	0.25	0.25	0.25
2	0.0	1.0	0.0	0.0	0.0	0.5	0.5	0.0	0.0
3	0.5	0.0	1.0	0.25	0.0	0.5	0.5	0.125	0.125
4	0.5	0.0	0.25	1.0	0.0	0.125	0.125	0.5	0.5
5	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.5	0.5
6	0.25	0.5	0.5	0.125	0.0	1.0	0.5	0.0625	0.0625
7	0.25	0.5	0.5	0.125	0.0	0.5	1.0	0.0625	0.0625
8	0.25	0.0	0.125	0.5	0.5	0.0625	0.0625	1.0	0.5
9	0.25	0.0	0.125	0.50	0.5	0.0625	0.0625	0.5	1.0

Clearly, the large number of non-zero off-diagonal elements indicates that there is substantial genetic information shared among these animals. BLUP techniques, through the mixed-model equations, take into account this shared information. Equally as important is the concept that if phenotypes have been measured on only some animals, breeding values can still be estimated for all their relatives because a portion of the population's germplasm is shared among the relatives. In this example, if phenotypes are only measured on animals 6, 7, 8, and 9, EBV's can still be calculated for animals 1-5 based on their relationships with animals 6-9.

The astute reader has probably noticed that the matrix inverse of the numerator relationship matrix is required in equation 4, but the rules given above are for calculating A , not A^{-1} . For nearly 20 years, the animal breeding community knew that mixed-model equations offered a dramatic improvement in the technology for estimating relative genetic differences among animals, but only in cases where the number of observations were relatively small could solutions be obtained. Clearly, standard rules for inverting A could have been applied, but the stumbling block was how to compute A^{-1} when 10,000 or more animals needed to be included in the analysis. In two landmark papers^{6,7} in the mid-1970's, rules were published for directly calculating the inverse of A without first having to form A itself. With the appearance of these papers, it suddenly became feasible to write computer software capable of simultaneously calculating EBV's for literally hundreds of thousands of animals in one analysis. Not only was it now possible to calculate EBV's on one trait for many animals, it soon became possible to simultaneously calculate EBV's for several genetically related traits in one comprehensive analysis. Several software packages now exist to calculate EBV's using multiple-trait techniques that can utilize different statistical models for each trait for a virtually unlimited number of animals.

Statistical models

The mixed-model equations have the flexibility to simultaneously adjust for differences attributable to fixed effects and to estimate breeding values. One's choice of the statistical model used to partition variation in observed phenotypes is crucial to obtaining a fair genetic comparison among animals. Especially important in defining the model for a particular trait is the set of fixed effects one chooses. This is such an important component of the analysis procedure, that it warrants a closer look.

Depending upon the trait being examined, sex of the animal may, or may not, be important. In many measures of physical stature, there clearly are differences between males and females in most species. Dogs are no different. For example, male German Shepherd Dogs

in The Seeing Eye's colony are on average about 10 pounds heavier at a year of age than the females ($P < 0.05$). When hip quality has been assessed using the PennHip score, the females, on average, have higher values than the males ($P < 0.05$). On the other hand, when hip quality has been assessed using a qualitative score assigned to radiographs in the hip-extended view, the sex effect is non-significant ($P > 0.05$). On a trait-by-trait basis, one should assess whether or not sex of the animal contributes to variation observed in the phenotype.

If measurement of the phenotype is subject to a value-judgment interpretation, then it may be important to include in the model a term to identify the person who made the physical interpretation. Certainly, this can be an important factor contributing to variation in the evaluation of either hip or elbow quality where the qualitative assessment of a radiologist must be relied upon to obtain the quality grade. There will be variation among evaluators, and if that variation is not accounted for by the statistical model, then its overall contribution to total variation will be assigned to the "residual or error" term in the model.

Age of an animal when evaluated is often another important source of variation in phenotypic values. Age can be statistically accommodated in the model in one of two ways. First, age could be included as a simple covariate, which is appropriate if the effect of increasing age is essentially linear across the range of ages observed among animals with measurements. Second, a group of age classes can be defined, and each animal can be classified as belonging to one of c mutually exclusive age classes. This approach has the added benefit of accommodating a non-linear (or curving) relationship between increasing age and the trait under analysis.

For some phenotypes, the scale of measurement or the local environment in which the animal lives can change with time, especially over longer periods of time, say years. When this occurs, it is often possible to account for these changes in scale by defining a contemporary group of animals as all animals evaluated under the same, or similar, conditions in a shorter time interval. At The Seeing Eye, two classes of contemporary groups are routinely used to account for possible changes of scale in scores used to describe dogs. For example, all dogs are evaluated on their ability to be trained as guides using a 9-point scale (1= least desirable; 9=most desirable). For each dog, the date when the score was given is recorded, and a contemporary group of dogs is defined as all dogs of a given breed which received their "trainability" score in the same calendar quarter. Thus all German Shepherd Dogs evaluated in January, February, or March 2001 comprise the dogs in that particular contemporary group. Although it is possible, it is very unlikely that a dramatic shift in the definition attached to a particular trainability score on the 9-point occurred in this 3-month time period. Over several years, however, it is certain that definitions associated with values on the scoring scale changed. A similar definition of contemporary groups is also used in The Seeing Eye's procedure for calculating EBV's for hip quality as assessed by a qualitative hip score. In this case, the date used for assigning a dog to a specific contemporary group is the date when radiographs were taken, rather than the date when the trainability score was applied. In many cases, these two dates will occur in the same calendar quarter, but it is not a necessary requirement for the EBV calculation process that they do so.

There are, no doubt, other identifiable non-genetic factors that could help explain variation observed in other traits. It is important that appropriate statistical tests be completed to assess the relative importance of these putative effects. Those that are statistically significant should be included in the statistical models used to partition variation in a trait into its genetic and non-genetic components. EBV's that are calculated using the statistical model that explains the largest proportion of total observed variation will most accurately rank animals available for selection.

Computing and Using EBV's

Estimated breeding values are calculated by forming and solving the mixed-model equations using an appropriate statistical model. As an example to motivate discussion in the remainder of this paper, consider the set of 9 animals shown in the pedigree in Table 1. Assume that an in-for-training behavior test was administered to animals 6 through 9, and the test is scored using a 5-point scale. Their scores are 2, 4, 3, and 5, respectively. Also, assume that heritability of this score is 20%, and for the sake of simplicity in this example, the statistical model will be limited to include only a single fixed effect for contemporary group of the dog. These data are summarized in Table 3.

Table 3. Example data.

Animal ID	Sire ID	Dam ID	Sex	Contemporary Group	In-for-Training Test Score
1	Unknown	Unknown	M		
2	Unknown	Unknown	F		
3	1	Unknown	M		
4	1	Unknown	F		
5	Unknown	Unknown	M		
6	2	3	M	1	2
7	2	3	F	2	4
8	5	4	M	1	3
9	5	4	F	2	5

All information is now available to calculate EBV's for the 9 animals in this pedigree, so the pieces of the mixed-model equations can be constructed. Again, the basic mixed linear model can be written as:

$$y = X\beta + Zu + e$$

The individual elements of each matrix are:

$$\begin{bmatrix} y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \end{bmatrix} + \begin{bmatrix} e_6 \\ e_7 \\ e_8 \\ e_9 \end{bmatrix}$$

where

y_6 through y_9 are the observed in-for-training test scores for animals 6 through 9,
 b_1 = fixed effect for contemporary group 1,
 b_2 = fixed effect for contemporary group 2,
 a_1 through a_9 are the estimated breeding values for animals 1 through 9, and
 e_6 through e_9 are the residual errors for animals 6 through 9.

Next, A^{-1} is needed. It is:

$$\begin{bmatrix} 1.6667 & 0.0000 & -0.6667 & -0.6667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 2.0000 & 1.0000 & 0.0000 & 0.0000 & -1.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.6667 & 1.0000 & 2.3333 & 0.0000 & 0.0000 & -1.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.6667 & 0.0000 & 0.0000 & 2.3333 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & -1.0000 & -1.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & 2.0000 & 0.0000 & 0.0000 & 0.0000 & -1.0000 & -1.0000 \\ 0.0000 & -1.0000 & -1.0000 & 0.0000 & 0.0000 & 2.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -1.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 & 2.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & -1.0000 & -1.0000 & 0.0000 & 0.0000 & 2.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & -1.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 & 2.0000 & 2.0000 \end{bmatrix}$$

Finally, α , which is a function of heritability, must be calculated. It is:

$$\alpha = (1 - 0.2) / 0.2 = 4.$$

Just to refresh the path down which this discussion is heading, recall that symbolically, the left-hand side of the mixed-model equations is represented as:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix}$$

and when these matrix operations have all been completed, the 11 x 11 results is:

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 6.6667 & 0 & -2.6667 & -2.6667 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 4 & 0 & 0 & -4 & -4 & 0 & 0 \\ 0 & 0 & -2.6667 & 4 & 9.3333 & 0 & 0 & -4 & -4 & 0 & 0 \\ 0 & 0 & -2.6667 & 0 & 0 & 9.3333 & 4 & 0 & 0 & -4 & -4 \\ 0 & 0 & 0 & 0 & 0 & 4 & 8 & 0 & 0 & -4 & -4 \\ 1 & 0 & 0 & -4 & -4 & 0 & 0 & 9 & 0 & 0 & 0 \\ 0 & 1 & 0 & -4 & -4 & 0 & 0 & 0 & 9 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -4 & -4 & 0 & 0 & 9 & 0 \\ 0 & 1 & 0 & 0 & 0 & -4 & -4 & 0 & 0 & 0 & 9 \end{bmatrix}$$

The right-hand side of the mixed-model equations is symbolically:

$$\begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}$$

and when these matrix operations have been completed, the result is:

$$\begin{bmatrix} 5 \\ 9 \\ 0 \\ 0 \\ 0 \\ 2 \\ 5 \\ 3 \\ 4 \end{bmatrix}$$

Solving these equations is now possible, and the resulting solution vector is:

$$\begin{bmatrix} 2.50 \\ 4.50 \\ 0.00 \\ -0.09 \\ -0.07 \\ 0.07 \\ 0.09 \\ -0.13 \\ -0.13 \\ 0.13 \\ 0.13 \end{bmatrix} = \begin{bmatrix} \text{Contemporary group 1 average} \\ \text{Contemporary group 2 average} \\ \text{EBV 1} \\ \text{EBV 2} \\ \text{EBV 3} \\ \text{EBV 4} \\ \text{EBV 5} \\ \text{EBV 6} \\ \text{EBV 7} \\ \text{EBV 8} \\ \text{EBV 9} \end{bmatrix}$$

Notice that the dogs in contemporary group 1 have an average score of 2.5 while those in contemporary group 2 have an average score of 4.5.

How to use EBV's to make selection decisions

EBV's describe the germplasm available for selection. In some cases, animals with the highest EBV's are more desirable. In other cases, animals with the lowest EBV's are deemed to be genetically better, while in yet other cases, animals with mid-range EBV's are the best ones for breeding. It is very important to understand this concept: EBV's simply describe the germplasm. EBV's do NOT indicate which animals are the best ones to select for breeding.

The decision about which animals to select for breeding must be driven by the goals one wishes to reach. In the case of genetic diseases, the goal is to eliminate the disease, but in the case of performance traits or traits of physical size, extremes in either direction may not be the most desirable. Assuming that one wants to improve the in-for-training behavior test score and had in hand the EBV's calculated above as a description of the germplasm available for selection, how could the EBV's be used for making selection decisions? The first step is the

separate the EBV's into two groups by sex, and the second step is to order the EBV's from the most positive to the most negative. The rankings by sex are shown in Table 4, below.

Table 4. Rank ordering of EBV's by sex.

Male ID	Male EBV	Female ID	Female EBV
8	0.13	9	0.13
5	0.09	4	0.07
1	0.00	2	-0.09
3	-0.07	7	-0.13
6	-0.13		

Since the objective in this example is to improve the behavior test score in this population, the dogs most desirable for selection are those with the highest EBV's. Several interesting points stand out in Table 4. First, genetically the best male had only a phenotypic score 3, which is only mid-range on the 1-5 scale of this example. This is reasonable, however, when one considers the large effect attributable to contemporary groups in this example, and clearly a score of 3 for dog 8 is above the mean of 2.5 for all members of contemporary group 1. Second, genetically the worst female (number 7) had a phenotypic score of 4, which is only one level below the highest quality score. Third, the second best female, based on EBV's, didn't even have a score in the dataset, so without the EBV's, there was no basis for deciding whether she was genetically superior or inferior with respect to the behavior test score.

In this example, EBV's have been calculated for only one trait. In the real world, breeders will want to know the EBV's for each dog on a set of traits, but having a large set of EBV's available on a group of dogs can lead to much confusion about which ones are the best candidates for selection. To make the selection process even more objective, a set of weights or coefficients of importance, some of which can be zero if no attention is to be given to a particular trait, can be defined for the set of EBV's. For a particular dog, each of its EBV's is multiplied by the coefficient of importance for that trait and all these products are summed together for that dog to obtain one overall index value. This overall index value can now be used as the basis for ranking all dogs available for selection on the basis of several traits, and the best dogs from the top of the list can be readily identified. If this process is applied over several generations of selection, genetic change will occur.

Multiple-trait models

The model illustrated above is known as a single-trait model because it only deals with one trait. A general extension of this methodology leads to multiple-trait models. In these models, a set of traits can be analyzed simultaneously, which then makes it possible to obtain estimates of genetic covariances among related traits. Genetic covariances occur when a set of genes affect more than one phenotype. For example, there is a non-zero genetic covariance between mature weight and mature height of dogs. Between the distraction index and an OFA-type score, both of which are measures of hip quality, there exists a genetic covariance, and a multiple-trait model can take into account this covariance in the process of calculating EBV's.

The Seeing Eye currently uses a multiple-trait approach to simultaneously obtain EBV's on four phenotypic traits of every dog. These four traits are: (1.) distraction index, a measure of hip quality; (2.) qualitative hip score; (3.) trainability score; and (4.) mature weight.

Software available for calculating EBV's

In recent years, several computer programs designed for calculating EBV's have been written. Most of these packages are a by-product of on-going research being conducted in various university and government laboratories around the world. Four of these packages are discussed below.

[MTDFREML](#), which is an acronym for multiple-trait derivative-free restricted maximum likelihood, was written by scientists at the USDA Meat Animal Research Center, Clay Center, Nebraska. It was primarily written as a tool for estimating genetic and environmental variance and covariance components, but as a by-product of these calculations, it will also calculate EBV's. [MTDFREML](#) is written in FORTRAN and must be compiled to use. There is a set of FORTRAN parameter statements that can be user modified to accommodate different modeling needs. These parameters control options such as the number of random covariates that can be included in the model, the total number of animals, and the total number of traits that can be simultaneously fit.

[Wombat](#) is similar to [MTDFREML](#). It was written in Fortran by [Dr. Karin Meyer](#), University of New England, Armidale, NSW, Australia, who continues to maintain and improve it. [Wombat](#) is very flexible in the statistical models it can accommodate, and it estimates variances, covariances, and heritability for each trait included in the analysis. In addition, it will calculate EBV's.

[MTGSAM](#), which stands for multiple-trait Gibbs Sampling, is another set of FORTRAN computer programs for estimating heritability and calculating EBV's. These programs were written by [Dr. Curt Van Tassell](#), an animal breeder now employed by the U.S. Department of Agriculture. The user interface for these programs is very similar to the user interface for [MTDFREML](#), but the internal workings of the calculation routines are quite different. [MTGSAM](#) can be used to obtain estimates of variance and covariance components and heritability. It will also solve the mixed-model equations to obtain EBVs. These programs need to be compiled by a FORTRAN compiler on the target machine.

[MATVEC](#) is a computer program for doing linear algebra and data analysis. It was written by Dr. Tianlin Wang, who received inspiration and motivation from Dr. Rohan Fernando, both of whom are animal breeders. This program is written in C++ and is supplied as a binary executable, which can be [downloaded](#) from the web-site of [Dr. Steve Kachmann](#).

[MATVEC](#) can fit many kinds of statistical models and is very general in the types of problems it can address. To realize that generality, one must learn its commands and know exactly which commands are appropriate for fitting a particular model. [MATVEC](#) is NOT intended for use by the novice or faint-of-heart. Just because the program successfully completes a particular set of calculations is no guarantee that the model was appropriate or the fit of the model is adequate. Burden rests with the user to know that an appropriate model has been specified and that the fit of the model obtained by [MATVEC](#) is adequate for the data under analysis.

How often should EBV's be updated?

EBV's are strictly a function of the data available for estimating them. As new data are collected and EBV's are updated, old values will change. This motivates the need to discuss how often EBV's need to be re-calculated.

At The Seeing Eye, EBV's are updated every night, so that staff beginning to work with them each morning have benefit of the most recent data added to the record keeping system

the previous day. Updating them that often may be impractical for other dog guide schools. The frequency with which updated EBV's are calculated depends a great deal on the set of traits for which EBV's are sought. If the traits are mostly health related, then updating once or twice annually may be sufficient, but if the traits include measures of working performance, then the updates may need to be more frequent.

Educating dog guide breeding program managers to use EBV's

Education of domestic livestock producers on the use of EBV's for genetically improving domestic livestock produced for food was largely accomplished by the Cooperative Extension Service, an agency of the U.S. Department of Agriculture. The Cooperative Extension Service has a long and successful history of teaching advanced production concepts to U.S. based farmers and ranchers. If teaching these concepts to dog guide breeders is important, then maybe this is a role that could be undertaken by the International Federation

One of the most successful tools for teaching beef cattle breeders how to use EBV's was the computer cow game. In its first rendition, each "cow" was represented by a computer punch card. While playing the game, each breeder could choose which "bulls" to use for breeding and then could choose how those "bulls" were to be mated with the available "cows". Typical of the real-world, a certain percentage of cows in every herd would fail to conceive, calves would die, and other calamities would strike. Out of all this mayhem and information, however, emerged new calves with new records that became the basis for a new round of EBV's. Selection decisions could be made and the process repeated. Cattle breeders enjoyed competing with each other to see who could produce the greatest genetic improvement in their herd in five generations of selection. Such a tool, specifically written for breeding guide dogs and implemented on the world-wide web, would be an especially effective technique for teaching animal breeding concepts, especially EBV's.

An "Across-Schools" Genetic Analysis

Beef producers realized that EBV's were essential to their economic survival when they learned that a breed-wide analysis had been done that took into account differences attributable to contemporary groups within a ranch and also took into account differences among farms and ranches where the cattle had been raised. It was possible to do this type of nationwide analysis because there was considerable movement of germplasm from one ranch to another through the sale of young bulls. In addition, breed associations designated a group of bulls each year to be reference sires. Under this reference sire scheme, records on calves raised on a particular ranch would only be included in the nationwide analysis if a minimum number of calves in a contemporary group had been sired by a bull designated as a reference sire. Most of these calves from reference sires were conceived by artificial insemination (AI), so this approach only became feasible as the use of AI became more widespread. With frozen semen and transcervical insemination now feasible in dogs, it is now possible to extend the beef cattle model to the breeding of dog guides.

An across-schools genetic analysis of dog guide breeding stock would make it possible to predict how the offspring of a male from school A would actually work in school B. It will only be possible, however, to do an across-schools genetic analysis when males from school A have been used to produce offspring in school B and when males from school B have been used to produce offspring in school A. Several males from both schools would need to be exchanged over several years before the genetic ties between schools would support an across schools analysis.

One of the first questions likely to be asked is why should schools even consider sharing germplasm? The short-term answer is that there is no compelling reason to share germplasm if a school is satisfied with the quality of puppies being produced, but the long-term answer to that question is just the opposite. A small breeding population within a single school can likely remain genetically isolated for only two or three decades, at most. In that length of time, inbreeding will increase, and as it increases beyond 35-40%, the likelihood increases that it will adversely impact reproduction. By the time a problem is recognized, the damage will have been done, and there will be no choice but to introduce outside germplasm.

Sharing germplasm will reduce the rate of increase in inbreeding

The need to introduce outside germplasm into dog guide breeding programs can be largely mitigated by increasing the effective population size. The rate at which inbreeding will increase in a population can be approximated by the formula⁸:

$$\Delta F \approx \frac{1}{8N_M} + \frac{1}{8N_F}$$

From this formula, it is easy to see that by increasing the number of males used to sire new offspring will slow down the rate of increase in the average inbreeding coefficient. At The Seeing Eye, the number of males used as sires in each successive generation has been increased by limiting the number of matings for every male. For several years, males were limited to making 15 matings in total, but more recently, that number has been reduced to 10 matings for most males. This means that Seeing Eye males are relatively young when they have completed their terms of service as studs, so many of them have undergone a career change and continue their service as guides. While these studs are in active breeding status, it would be feasible for The Seeing Eye to increase the quantity of frozen semen placed in storage, and some of this semen could be shared with other schools around the world. If semen sharing were routinely done for 5-10 years, then the effective population size for schools participating in the semen sharing scheme would be the worldwide breeding colony instead of their own, individually much smaller closed populations.

Reference sires can be used to create genetic ties among schools

Members of the Original Group who want to participate in a semen sharing scheme for the purpose of obtaining an across-schools genetic analysis could begin the process now by nominating two studs per breed per year to be used as reference sires by other schools. A protocol would need to be worked out and approved for handling the logistics of such a scheme, but the long-term payoff to participants could be the difference between reproductive survival of their breeding colonies and extinction.

One concern sometimes mentioned by individual schools when they consider participating in an across-schools genetic analysis scheme is how would the analysis accommodate different approaches to measuring similar phenotypes. For example, hip quality at The Seeing Eye is assessed by both the PennHIP method and a qualitative hip score that is similar to, but not done by the OFA. Other schools use the OFA, while others have internally developed scoring schemes. If genetic ties exist among populations being analyzed, then there will be some sires in the database that have offspring at school A evaluated by their scheme and other offspring at school B evaluated by that school's scheme. Using a multi-trait approach to the analysis, all dogs available for selection and their parents and more removed ancestors can be assessed under the evaluation schemes of either school by considering the hip quality traits to be separate but correlated. Sires that could be used across schools would thus have

EBV's estimated under both phenotype evaluation schemes and each school could assess the merits of breeding stock from either school using their own method for phenotype evaluation.

Summary and conclusions

EBV's are an important tool for genetically describing individuals in a population. EBV's can be used as a component in the process of making selection decisions, and when EBV's have been calculated using best linear unbiased prediction techniques, their use will likely improve the accuracy of selection decisions. Improved accuracy would be greatest for traits that are low to moderately heritable, say when heritability is below 0.4.

Advances in computing algorithms and computer software to implement these algorithms now make it possible to simultaneously calculate EBV's on several traits measured on thousands of dogs.

Reference sires can be used to quickly create genetic ties among otherwise isolated dog guide breeding populations. Their use among member schools of the Original Group would facilitate completing an across-schools genetic analysis for several key traits. With this information in hand, each school could objectively evaluate the germplasm available from other schools when they are considering using outside germplasm in their own colony.

References

- ¹ Leighton, EA. Genetics of canine hip dysplasia. *J Am Vet Med Assoc* 1997; 210: 1474-1479.
- ² Mrode, RA. *Linear models for prediction of animal breeding values*. Wallingford, UK: CAB International, 1996; 1.
- ³ Henderson, CR. Sire evaluation and genetic trends. In: *Proc Animal Breeding and Genetics Symposium in Honor of Dr. J.L. Lush*. ASAS and ADSA, Champaign, Illinois, 1973; 10-41.
- ⁴ Falconer, DS and TFS MacKay. *Introduction to quantitative genetics*. 4th ed. Essex, UK: Addison Wesley Longman Ltd, 1996: 122-123.
- ⁵ Mrode, RA. *Ibid*, 24.
- ⁶ Henderson, CR. A simple method for computing the inverse of a numerator relationship matrix used in predicting of breeding values. *Biometrics* 1976; 32: 69-83.
- ⁷ Quaas, RL. Computing the diagonal elements of a large numerator relationship matrix. *Biometrics* 1976; 32: 949-953.
- ⁸ Falconer, DS and TFS MacKay. *Ibid*. 67.