By Sean McGrath
Breed Improvement Coordinator for the Canadian Simmental Association

Where do EPDs Come From – From Breeding to Birth of an EPD

There is a common misconception among many beef cattle breeders, that EPDs are pulled from the air or based on some imaginary information. In fact, nothing could be further from the truth. In this article we will attempt to explain (and in some cases vastly oversimplify) where EPDs are derived from. Having a basic understanding of the process of development and some of the science behind an EPD can help us to appreciate how they work and that they do work. This understanding also demonstrates why it is important to provide data that is of as high a quality as we can collect.

The conception stages of an EPD should revolve around the concept of Economically Relevant Traits (ERT). An ERT is a trait that directly impacts the profitability of, in our case, a beef operation. This means the EPD should express genetic differences in traits that either have a direct impact on either costs or income. For example, if we market calves at weaning by the pound, then weaning weight is an ERT, since it produces income.

We have EPD for traits that are not ERT and usually these are what we call indicator traits. These are often historic artifacts because we had the data to develop them before the technology to calculate the relevant ERT was available. Birth weight is a good example here. The technology to calculate genetic differences in birth weigh was available before the technology to calculate what really costs money (Calving Ease) was available. Or in other cases we use them because they are easier to measure than the trait that is the actual ERT. For example, it is easier to measure mature cow size than feed intake of a cow on pasture.

Decisions on which EPD to pursue are usually based on collaboration between researchers and producers (often with breed associations acting as the middle man). For example, producers might indicate that they are concerned about, or need to work on knowledge of carcass characteristics in their cattle to better serve their customer base. The association may work to help them collect and pool carcass and ultrasound data and work with researchers to develop and deploy carcass EPDs for various traits such as marbling or eye muscle area. Data collection is really, just a first step. A good current example would be the work of IGS to collect feet and leg scores from breeders.

Secondarily, researchers and producers may identify areas of concern that are more difficult to measure or are novel traits. In this case research herds, such as those at Universities, Research Stations or Co-operator Herds may be used to build datasets. Probably a good recent example of this may be work done on PAP scoring for resistance to altitude sickness.

Once there is a data set established on the trait of interest, researchers will go to work understanding the data and what it means. This is anchored in biology and an understanding of how natural processes workand also in what expression of the trait will inform commercial selection decisions. A simple example could be weaning weight expressed in relative pounds of difference between offspring.

It is a big simplification, but any new analysis involves understanding the trait and how to express it, and understanding what the data is telling us about the genetics behind that trait. You may hear terms like, “estimation of variance components” or “heritability estimates.” These are basically scientific terms for understanding data and how it is spread out across the population. We will use an oversimplified example of heritability to illustrate the point. Heritability can be thought of as a way of understanding how much variation in phenotypes for a trait, is due to the genetics for that trait. For traits with a high heritability, each phenotypic record contains a lot of information about the genetics involved. Traits with a low heritability are generally subject to a lot of environmental influence and so each record may contain slightly less genetic information.

Heritability is calculated using data. Again, in a vast oversimplification, we look at expected vs actual differences in the data and calculate heritability. There is a lot more to calculating heritability than what is shown, but this basic example of calculating a realized heritability for birth weight will hopefully shed some light on the idea behind the process.

If we have a population with an average birth weight of 80 pounds and we selected parents from that population with a birth weight of 70 pounds. We would expect the offspring of those parents to be 70 pounds or 10 pounds less than the average if birth weight was 100% due to genetic factors. Let’s say in our data set the average birth weight of the offspring was 77 pounds. The expected variation was a 10 pound decrease, but the realized difference was only 3 pounds. The realized heritability is the difference between the original population and the offspring of the selected parents divided by the difference between the average of the original population and the average of the selected parents.

$$\text{Realized Heritability} = \frac{\text{Average BW} - \text{Actual BW}}{\text{Average BW} - \text{Expected BW}} = \frac{80 - 77}{80 - 70} = \frac{3}{10} = 0.30$$

Again, this is a very simplified example and does not show all of the processes behind the scenes, but it provides the general idea of some of the math behind determining genetic contribution to a trait and developing an EPD.When done on a large scale (several thousand records) and with clean data we can get a very good idea of what portion of observed differences are explained by the additive genetic component. These are our “variance components” mentioned earlier.

Correlations are also determined from data and an understanding of biology. Correlations do not “cause” a genetic result, but they do allow us to use data from other associated traits to improve our predictive capability. One of the best examples here is birth weight and calving ease. Heavier birth weights are associated with more difficult births, but they are not the only cause of difficult calving. Calf shape and maternal environment also play major roles. These traits are all “associated” with calving ease. By using information such as birth weight and calving difficulty scores, we can do a much better job of figuring out genetics for calving ease.

Once the background biology is done, and the math complete, an EPD can be calculated. This is the basic process of assigning contemporary groups and examining how the genetics of animals play out across groups with the same management and environmental history, and applying our new variance components to the data.

Part of the process of EPD development involves taking the models developed in the earlier stages and computing EPD on cattle, then comparing the actual performance of the cattle (data records) to see if the predictions are actually producing the expected results. It also includes the litmus test of providing trial results to breeders and seeing if they match up with what they are seeing on the ground and submitting the EPDs to others in industry and the scientific community to scrutinize the results and ask questions. Generally this process takes several iterations, and may even involve going back into the steps of understanding the data and the trait we are working on.

While this article has several glaring oversimplifications, this also serves to illustrate the point that as the science evolves and we gain better understanding of how a cow works and how she interacts with her environment we can garner more genetic information out of our existing phenotype and pedigree data. This is one of the biggest drivers behind changes in EPD over time. Our evaluation, software, hardware and even our data collection improves and thus our EPD can be made more predictive as well. Just like a calf that is born, the science of genetic evaluation continues to grow and evolve. Although outside the scope of this article, a good recent example of this would be the inclusion of DNA or genomic data into genetic evaluation. DNA technology has rapidly changed and enhanced our ability to calculate accurate genetic predictions and make better comparisons of potential seedstock.

A lot of work goes on in the background before an EPD ever sees the light of day. From assessing the traits of interest and ensuring they are in fact economically important, to collecting good data sets on traits of interest, to examining the relevant biology and evaluating the data to determine the genetic components of a trait, running a model and comparing expected results with actual on the ground performance, an EPD is based on science and is the most powerful selection tool available in the context of a targeted breeding objective.