BY THE NUMBERS

by André Garcia, senior geneticist

Accuracy Matters

Insights into the confidence of genetic predictions.

> *"Ultimately, even in the age of genomic selection, phenotypes and progeny data are the drivers of accuracy in our genetic evaluations."* — ANDRÉ GARCIA

When discussing selection tools, people typically think of the expected progeny differences (EPD) first. However, upon closer examination of the data, another important value often accompanies the EPD: the accuracy.

Accuracy is a statistical value representing the amount of information behind an EPD. It ranges from 0 to 1, with zero reflecting the least accuracy and 1 signifying the highest accuracy. This metric is typically located directly below the published EPD, providing insights into the confidence and precision of the genetic predictions.

CED Acc % Prog	BW Acc % Prog	WW Acc % Prog	YW Acc % Prog
+9	+1.1	+68	+117
.30	.48	.39	.34
30%	40%	25%	40%

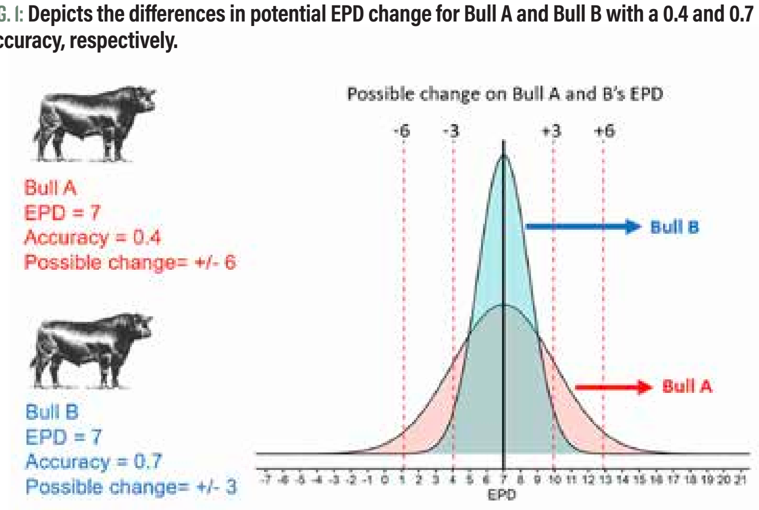
That accuracy number is derived from the same model and data used to estimate the EPD, and it can be used as an indicator of risk for possible changes in the animal's individual EPD as more information gets added to the evaluation.

Using the accuracy values, we can calculate a possible change distribution, which is the potential amount of future change in an EPD prediction. Possible change is also described as the measure of expected change or potential deviation between the EPD prediction and the "true" progeny difference of that animal. The more accurate the EPD, the smaller the range of possible change. Figure 1 depicts the differences in potential EPD change for Bull A and Bull B with a 0.4 and 0.7 accuracy, respectively. With the lower accuracy bull, Bull A, you have a wider bell-shaped curve, representing the larger possible change distribution of +/-6. Whereas, the higher accuracy bull, Bull B, has a narrower bell-shaped curve, indicating a smaller possible change distribution of +/-3.

The American Angus Association publishes a possible change table, which represents the possible change distribution for each EPD at a given level of accuracy. This table is derived from the statistical properties of the genetic evaluation model and allows us to create a confidence interval for each EPD. For each given accuracy level, about two-thirds of the time, an animal should have a "true" progeny difference within the range of the EPD plus or minus the possible change value. While this range is the most probable amount of change to expect, it is possible to fall outside of this range of one standard deviation, especially on young, unproven animals.

From a breeder's perspective, the accuracy value and the possible change table can serve as a risk mitigation tool. It allows some insight into how EPDs might change in the

FIG. 1: Depicts the differences in potential EPD change for Bull A and Bull B with a 0.4 and 0.7 accuracy, respectively.



Courtesy of Jorge Hidalgo, Assistant professor at University of Georgia future. It is important to note that the changes observed follow a normal (bell-shaped) distribution, and changes can be positive or negative.

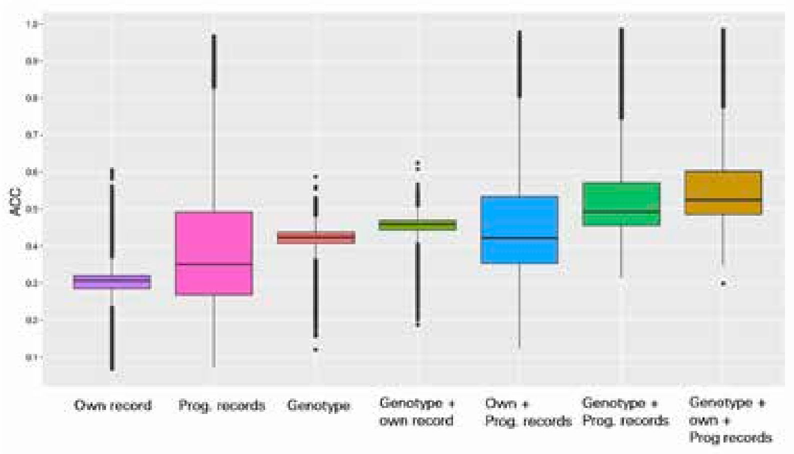
What are the sources of information that influence the amount of accuracy an EPD has?

Because accuracies are calculated from the same models as the genetic evaluations, the sources of information affecting accuracies are also the same: the phenotypic data (own and progeny), pedigree information (from parents and progeny), and if an animal is genotyped or not. Additionally, different factors contribute to the accuracy of a given trait, such as the heritability of the trait (the higher the heritability, the higher the overall accuracy); the statistical model; and the overall amount of phenotypic data in the genetic evaluation for that particular trait.

The more information available for a given animal, the more accurate the EPD prediction will be. In a genetic evaluation there are animals in many stages of data recording. For instance, we have young selection candidates that start off with parental average only, until their birth weight or weaning weight gets reported. Then a genotype gets collected, and finally progeny records are added. This translates to animals having varying levels of accuracy across the population.

With the advent of genomic selection, many young selection candidates are now genotyped, and they get an increase in accuracy earlier without having to wait for all their progeny to report a phenotype into the evaluation. This has allowed breeders to make more accurate decisions earlier in the selection process, which ultimately decreases the generation interval and increases the rate of genetic gain. However, even with a genotype added to the evaluation, the accuracy of that individual will not reach the same level as an individual with thousands of progeny data reported.

FIG. 2: Accuracy distribution with different levels of information available.



In Figure 2, we can see that the more information we add from the different sources, the more accurate the prediction gets.

Keep in mind that all those levels have an average accuracy, but also a distribution around that average. For instance, a bull born in the 1990s may not have been genotyped, but has thousands of progeny weaning weight records reported and will have a more accurate EPD compared to a young, genotyped animal that does not have progeny data yet.

Another interesting takeaway from Figure 2 is to get to the highest levels of accuracy, individual records — but mainly progeny records — are necessary. Ultimately, even in the age of genomic selection, phenotypes and progeny data are the drivers of accuracy in our genetic evaluations.

Accuracy in genetic selection

While accuracy does not rank animals, it can be a helpful risk mitigation tool for selection decisions, and breeders always must balance the risk and reward of those decisions. For instance, the easiest way to avoid risk is to use proven animals only, but it comes with a potentially lower reward because of slower genetic gain through slower generational turnover. The opposite strategy is to try to maximize the rate of genetic gain by using only young, unproven animals for faster generation turnover; but this strategy comes with bigger risks.

The bottom line

Fundamentally, the EPD is our best tool to rank animals based on genetic merit. However, the process of selection involves an inherent element of risk management. To navigate this challenge, leveraging EPD accuracy and possible change tables becomes quite helpful. These tools provide valuable guidance, offering insights into the potential risks in the decision-making process, thereby enhancing our ability to make informed and strategic selections. AJ

agarcia@angus.org