

# Thoroughbred Breeding Applications

by Dave Astar

Years ago I learned to utilize data to understand variation. This ability allowed me to minimize emotion and dismiss unsubstantiated opinions, so I could focus on facts when making business decisions. As a result, I was lucky enough to retire from an executive management career in my mid fifties and turn my attention to the thoroughbred racing and breeding industry.

I found that so much had been written about thoroughbred breeding that it took me several months to simply catch up on the literature. Breeding theories were like chain drug stores. They were similar but kept popping up all over the place. Many published breeding theories seemed reasonable but rarely did I find them to be statistically valid. Again, I found that the best way to eliminate opinions, false science, and honest to goodness BS, was to go to the data.

The general nature of breeding produces unfathomable variation and when small sets of data are evaluated, discernable patterns are difficult to recognize. Genetic combinations are random and thousands of data points need to be analyzed to truly understand what does or doesn't work in the breeding shed. Importantly, those who understand statistical variation know that many published breeding theories may indeed be true, but may never be proven. For people like me, a theory that cannot be proven is a waste of time and only diverts my attention from higher priorities associated with learning how to apply proven theories.

To offer an example of a breeding theory that seems reasonable but to my knowledge has never been statistically proven, I researched the "type to type" breeding theory. This popular theory has existed for many years and with some variation from author to author, essentially consists of breeding mares to sires who have similar physical traits. If these traits match, the resultant foal will have a higher theoretical probability of success on the racetrack.

While the "type to type" breeding theory is very popular and may be valid, has anyone proven the theory? Better yet, has anyone really designed a controlled definition of physical characteristics, gathered the racing history of thousands of foals from predefined samples and then compared type to type versus non-type to type breeding results? I know I can not statistically design such an experiment, isolate all other competing variables and make the thousands of resultant foal observations required to prove this theory.

If a breeder at some time in history had racing success, and claimed to utilize a type to type breeding pattern with as many as 100 foals, the theory remains unproven statistically. Even if that breeder is a breeding legend and produced physical details of the broodmare and sire similarities that produced these same 100 foals, the theory remains a statistical myth because there were not enough observations to prove the theory, and other variables were not isolated.

Human history is laden with false beliefs based on unproven theories, and always will be. In the medical field for example, people were bled to treat a fever for thousands of years. Medical practitioners consistently observed a pattern of people saying they felt better and more relaxed after the treatment. These observations proved the theory in their minds and these same medical practitioners were some of the most learned people of their time. As we now know, the bleeding theory was only recently proven as false, considering the entire record of human history. Simply put, fascination with the visible was the great problem and will likely remain so in the future!

Rather than focus on unproven theories, opinions or myths, I have always found that successful people focus on concepts that have, or can, be proven. In racing, the historical data is statistically relevant. Horse racing success is definable and not simply a matter of subjective opinion. Success in racing is a matter of factual history. As a result, all I have to care about to prove certain breeding theories is how fast the parents and foals of those parents were, and at what level they were able to win races. With that information alone, I can prove certain breeding theories and discover how to statistically apply them, without ever seeing a horse.

After reviewing innumerable breeding theories, I found that there are only a few can be statistically proven and people who are likely smarter than me, are already utilizing methods to apply them. For example, the "best to the best" breeding theory consists of breeding the best sires with the best dams. This theory has existed for many years and simply put, is valid. The correlation between parental racing performance and foal racing performance is statistically obvious.

My analysis of varying foal subsets of stakes winning parents versus non- stakes winning parents, demonstrated an obvious divergence in foal performance. I completed best to best studies utilizing both stakes winning statistical groupings and groupings by speed rating. By doing so, I could use imperfect but nevertheless normalized measures to understand variation. I found that parental speed ratings do not directly dictate the speed ratings of their offspring. This was expected because foal speed ratings naturally regress to the mean average of all foals. Nevertheless, higher speed rated parents produce higher speed rated foals than lower speed rated horses produce. For example, utilizing one popular speed rating system, foals with parental average speed ratings over 100, produce foals well under 100, but their average rating was still a good 10 to 15 lengths faster than foals whose parents averaged speed ratings near 80.

Breeding the "best to the best" will indeed create a higher probability of breeding success and amazingly this knowledge alone can be helpful. While no one can exactly predict that one specific foal with parents that averaged a 100 speed rating will be better than a foal by parents with a 90 average, if all other elements are equal, your probability of success with the faster parented foal is much higher. The amazing thing to me is that so many people don't seem to understand this basic breeding fact. They continually breed unraced or proven slow mares, to fair or good sires. They remind me of people who think they can hit major league home runs with little league bats. Certainly, lightning may strike but should breeders really bet on that?

Another theory that can be proven is called "nicking" in the equine business. This breeding theory essentially focuses on sire lines within the pedigrees of two parents. When certain sire lines have been crossed with others, those crosses produce consistently performing foals. While it takes more data to prove nicking than it does to prove the best to the best breeding theory, nicking is statistically relevant and a sound basis for predicting probabilities of success.

The ability to predict probabilities does not assure success, unless a very high number of probable events are observed. In other words, I can tell you that the most probable outcome when rolling two dice is a seven. Seven should come up 6 times out of 36 rolls. That doesn't mean that it will but this 16.7% probability is unquestionable considering the use of two standard six sided dice. Even though seven is the most probable single throw outcome, it is even more probable that you will **not** throw a seven in three consecutive throws! Nevertheless, if you have to bet on a roll of the dice you are just plain silly if you do not bet on seven coming up. This probability example applies directly to predicting foal success. In other words, any theory that is proven to predict a more successful outcome will not assure the success of any given outcome, but that is the way to bet.

To help make a pedigree breeding analysis simple, I will reflect on how both the best to best and nicking breeding theories impacted a breeding I planned with my stallion and one of my broodmares. I bred a solid broodmare, Sugar Hills Miss, to my new stallion Stormy Business. Sugar Hills Miss was a stakes winner with a top end 89 speed rating. Stormy Business, our stallion, was a multiple stakes winner with a top end speed rating of 105.

In Minnesota, these speed ratings and stakes winning performances would clearly be considered the best of the best. For example, the recent 2010 Minnesota yearling sale had 52 yearlings in the original sale catalog. The Jockey Club has defined "Black Type" pedigree conditions as an objective "best" standard when pedigrees are illustrated. Black Type is earned by horses based upon their finish in high purse stakes races.

Only 4 of 52 yearlings listed in the 2010 Minnesota yearling sale catalog were produced by both a **CAPITAL LETTER Bold Black Type Stakes Winning Stallion and Broodmare**. The other 92% of the yearlings in the sale were produced by either stallions or broodmares that could not be quantified as "best to best" breeding, using the Jockey Club black type definition. Since Sugar Hills Miss and Stormy Business both fall into the Jockey Club capital black type definition, and less than 10% of yearling foals seem to fall into that category in Minnesota, my best to best criteria was obviously satisfied for a horse bred in Minnesota.

To evaluate nicking, I turned my attention to the sire lines imbedded in both the Sugar Hills Miss and Stormy Business pedigrees. After reviewing historical data, I extracted foal information that matched sires in the stallion's and broodmare's pedigree. That foal information was then weighted to value the influence of recent sires more heavily than 3rd or 4th generation sires. For example, crosses with the sire/father of Sugar Hills Miss are more credible than crosses to a 3rd generation sire/great grandfather.

I also study matches with the sons of sires and have found that sons of sires foal information is important in resultant foal probabilities. For example, my stallion is a grandson of the great Storm Cat

who once commanded an amazing \$500,000 sire fee. Storm Cat's graded stakes winning son Storm Creek is the sire of my stallion. Storm Cat's sons have produced great runners and by not only evaluating data from foals sired by Storm Cat and Storm Creek, but also the foal data from all sons of Storm Cat, I can better predict the success probability of a foal. I can prove that sons of sire data is relevant when evaluating resultant foals, but I do not know why. It may be because only certain sires have the ability to transmit siring greatness to their sons.

As I indicated, the greatest influences on any breeding combination can be found in recent generations. The following chart illustrates how the most recent Stormy Business sire influences have matched up with the Sugar Hills Miss sire, Mt. Livermore. The key Stormy Business sire influences come from his father Storm Creek, sons of Storm Cat, and his grandfathers Storm Cat and Colonial Affair. When these influential Stormy Business sires, listed in the first column of this chart, were bred with Mt. Livermore daughters like Sugar Hills Miss, here are the results:

	<u>Foals</u>	<u>Runners</u>	<u>Winners</u>	<u>2 YO</u> <u>Winners</u>	<u>Stakes</u> <u>Winners</u>
Storm Creek	4	3	2	1	0
Sons of Storm Cat	114	87	62	22	5
Storm Cat	8	7	6	3	0
Colonial Affair	1	1	1	0	0
TOTAL:	127	98	71	26	5

These crosses show that 77% of named foals have run and 56% have won. Furthermore, 20% won a two year old race and 3.9% eventually became stakes winners. These are not opinions or theories but simple facts about how effective the foals of Mt. Livermore daughters have been when Storm Cat, Storm Creek, Colonial Affair and Sons of Storm Cat have sired those foals. (There was no data on sons of Colonial Affair or that data would have been included in this first level analysis.)

Now we move to the next generation of Sugar Hills Miss and look at how the broodmares sired by her grandfathers have matched up with the Stormy Business sire influences. The Sugar Hills Miss grandfathers are Blushing Groom and Within Hail.

	<u>Foals</u>	<u>Runners</u>	<u>Winners</u>	<u>2 YO</u> <u>Winners</u>	<u>Stakes</u> <u>Winners</u>
Storm Creek	1	1	1	1	0
Sons of Storm Cat	20	17	10	2	1
Storm Cat	31	25	20	6	4
Colonial Affair	5	3	2	0	1
TOTAL:	57	46	33	9	6

Blushing Groom and Within Hail sired broodmares have produced 57 foals with 81% of those foals having run and 58% having won. Furthermore, 16% of the foals have won as two year olds and 10.5% have won stakes races.

I repeat this same analysis process going down through the generations of both the sire and dam. Each generation is then weighted based upon my studies of influencing credibility. The final weighted Stormy Business/Sugar Hills Miss nicking foal expectation follows:

	<b>Probability</b>
The Foal Will Run	77.59%
The Foal Will Win	55.63%
The Foal Will Win At Two	19.59%
The Foal Will Win A Stakes Race	4.84%

These probabilities alone mean little. To effectively understand probabilities, normative data is needed. The following chart illustrates the normative data and the foal's comparative probability:

	<b>Foal Probability</b>	<b>Average Nationally</b>	<b>Top 1% Sires</b>	<b>Foal Compared To Nat'l Avg</b>	<b>Foal Compared To Top Sires</b>
The Foal Will Run	77.6%	71.0%	84.5%	9.3%	-8.2%
The Foal Will Win	55.6%	48.1%	65.2%	15.6%	-14.7%
The Foal Will Win At Two	19.6%	11.6%	18.0%	69.0%	8.9%
The Foal Will Win A Stakes Race	4.8%	3.7%	8.6%	29.7%	-44.2%

Why does this nicking stuff work? Frankly, I don't know but more importantly, I don't care. I'm sure that there are genetic combinations that produce some pattern over time related to physical racing abilities. What I do know however is that a predictable statistical pattern exists when certain sire lines are crossed.

To quote a fellow horseman, "nicking probabilities are just a bunch of made up numbers and good for nothin". This attitude doesn't surprise me and as a long time executive in the white collar business world, I learned that people constantly relied on personal experiences, what they described as instinct and individual data points to form opinions. This human tendency to rely on experience and personal anecdotes is easy and common. After all, if you bet on seven and you throw something else 5 times in a row, why keep betting on seven? Unfortunately, those who never understand and accept statistical science will always be taken advantage of by others, and likely spend a portion of their life wondering why others have such good luck and theirs seems so bad.

For me, pedigree analysis information is valuable and helps me distinguish between low and high probabilities of success. For example, the nicking probabilities suggest that a Stormy Business/Sugar Hills

Miss mating will create a foal that has a 29.6% higher probability of winning a stakes race than the national average. I also recognize that the national average stud fee far exceeds the Stormy Business stud fee. On the down side the data also suggests that a Stormy Business/Sugar Hills Miss foal will have a 44.2% lower probability of winning a stakes race than the foals sired by the top 1% of sires. Of course, the top 1% of sires command six figure stud fees and see only the very best broodmares in the country.

The one place where handfuls of data can be impactful in predicting foal performance is when the broodmare has previously produced foals. Direct foal breeding results are highly credible. In other words, if a sire/mare combination has produced a successful foal, a repeat breeding offers a solid probability of success. There is also a high correlation between sibling performance when siblings have similar recent generation sire lines.

Sugar Hills Miss had a foal by a son of Storm Cat. That foal, Sugar Storm, won twice recently, ran a 90 speed rating and just became stakes placed running second in a stakes race. Sugar Hills Miss has had three other foals that have also raced, all have won and one was a stakes winner. With Sugar Hills Miss going 4 for 4 with winners, one being stakes placed, one being a stakes winner, and a son of Storm Cat sire breeding producing a stakes placed 90 speed rated horse, the probability of a Storm Cat grandson crossing with her to produce a successful foal are quite high. Once the actual broodmare foal history probabilities are combined with the nicking analysis and the best to best analysis, the probability of a Sugar Hills Miss/Stormy Business foal winning a stakes race jumps from the previously illustrated nicking 4.8% probability to 6.4%, 70% higher than the national average.

Buying decisions associated with purchasing weanlings, yearlings and two year olds require conformation analysis to determine if the physical characteristics and biomechanics of the horse will allow it to stand up to the rigors of both training and racing. In other words, the probability analysis I represented in this paper assumes all things as being equal. Even if the pedigree data allows me to predict a 10% stakes winning probability for a foal, the 10% probability is irrelevant if the horse is string hocked, toes out badly, is incorrectly top heavy, short legged or has any other number of significant physical deformities which override the pedigree analysis. Nevertheless, two equally well conformed horses can be differentiated, **to a significant extent**, by properly evaluating the previously described pedigree elements.

Over the last 6 years our stable results could be described as lucky. Nevertheless, with horses still in the middle of their racing careers, over 17% of the horses we have purchased at public auctions have won stakes races. 1 out of every 4 two year olds purchased from two year old in training sales have won stakes races, and 100% of those two year olds have won. Considering the average price we paid for our horses was a mere fraction of the national average, and nationally only 3.7% of foals ever win a stakes race, we are pretty happy with our results which seem to support our statistical method for determining foal success.

We will start to see how these pedigree applications impact our new breeding program when our Stormy Business two year olds begin their racing careers in 2011.