

Does the NFL Combine Really Matter

Paul Park

Under the Supervision of Prof. David Aldous

University of California at Berkeley

August 2016

Abstract

The NFL Combine measures a variety of physical abilities. This research questions the overall usefulness of the event by focusing on both current and past players of QB, WR, and RB positions, who have performed at NFL Scouting Combine from 2000 to 2010. Statistical methods using the idea of multilinear regression, random forest and hypothesis testing on correlations between one's Combine measurements and draft order, and as well as his NFL performance are used to mark and order the importance of the Combine drills in terms of how well they predict one's career. The study reveals that an outstanding performance at the NFL Combine is not sufficient to guarantee one's success as a professional player. However, certain traits seem to indicate a positive attitude that is usually necessary to be drafted into the NFL. What the Combine data could imply about one's potential also depends on his position; wide receivers and running backs are highly regarded for their speed, which means faster players are more likely to be drafted, while quarterbacks perform quite independently from all the Combine measurements. To conclude, the study offers suggestions on how one's raw athleticism measured at the Combine could be effectively utilized to assess the chance of one's being drafted by encouraging players in different positions to focus on different drills that could best demonstrate their potential based on statistical findings.

Introduction

As college football players prepare for the NFL combine at the start of the year, many rumors and speculations go around regarding who will be this year's number 1 pick at the NFL draft. A group of people including coaches, scouts, and general managers from the professional football teams all around the nation attend this mini-Olympic to see who has the most potential to complete the task given as a NFL player. With this level of media attention, players are almost forced to perform well at the Combine, and it is a known fact that there are various prep courses that insist they can improve one's performance. However, some might disagree and claim that the Combine results do not matter as much as they seem, especially for those who have already proven to be top prospects and know that there are teams out there that would want them. With all these matters in hand, the most important question everyone wants to ask is whether or not the Combine result is a good predictor of a player's success.

There have been a few studies that attempted to find the correlation between the Combine measurements and the draft status. There were two studies that successfully produced a statistically significant outcome that showed a strong relation between the draft order and the Combine performance. The study conducted by McGee and Burkett derived a regression equation using step-wise regression with the draft order as the response variable and 11 different Combine measurements including height weight as the explanatory variables. It further concluded that the coefficient of determination, namely R^2 , was significantly high, for positions including WR, RB, DB, and QB indicating players in these positions were more likely to be drafted if they had excellent Combine measurements. On the other hand, players under positions OL, DL, and LB did not have statistically significant R^2 between their draft order and Combine measurements. The study notes that these positions require qualities that are not assessed in the Combine. For example, instinct to quickly read the opponent's offense and being able to react to it is the primary role of LB, and there is no way to directly measure this ability by the drills that the Combine offers. Further analysis was conducted using ANOVA to detect the major difference in the Combine performance between the players who were drafted in the first two rounds and the players who were drafted in the last two rounds. It turned out that the significant differences were in the vertical jump, broad jump, and 3-cone drills. Although the study admits the draft order could be quite accurately predicted, it also throws a bit of skepticism how well the Combine measurements transfer to players' NFL career, especially with the presence of some

qualities that seem to be constantly overlooked in the NFL Combine (6-10).

Kuzmits and Adams, whose study also focused on the relation between Combine performance and draft order, concluded a very similar result to that of the aforementioned study, by providing the correlation coefficients between numerous values and the Combine measurements. These values include the draft order, salaries at different years into the league, number of games played at different years into the league, and also other measures. All of these are done separately by the positions. However, the study argues that few of the correlations are statistically significant, the degree that is defined as having a p-value less than 5%. The ones that did show a level of significance were dominantly the correlations between the draft order and the Combine performance. Quarterbacks were highlighted by a strong correlation between their measurements at vertical, broad jumps and their draft orders. Wide receivers and running backs were marked by a strong correlation between their fast speed and their draft orders. For quarterbacks and wide receivers, only 4 out of 80 correlations showed some kind of significance, which is almost consistent with a random chance model. This indicates that other than for the draft order, the Combine measurements failed harshly as an effective predictor for players' career (1721-5).

For running backs, however, the sprint time seemed to be a fairly accurate predictor for not only the draft order, but also the players' success as professional football players, as noted by the fact that faster players tended to earn more money, play more games, and have higher average yards. This can be a rational interpretation, since running backs are heavily influenced by their ability to outrun the opponents, and that along with some other qualities alone can make them dominate the game even at the professional level. Not surprisingly, the other seven Combine measurements used in this study did not prove to be a useful predictor for the players' career, even for running backs. The study concludes by questioning the overall effectiveness of the Combine performance as a reference for the NFL performance. It argues that the combine only measure one's pure athleticism but not his ability to play the game at a higher level. Moreover, it touches the issues with the presence of the prep courses and that might be a reasonable explanation for this matter since players who lack the qualities of NFL players might still be drafted if they took the courses and improved enormously at performing the Combine drills, assuming the relation between the draft order and the Combine measurements is indeed statistically significant (1721-7).

While the previously mentioned studies paid particular attention to how much the Combine

performance affect the chance of one's being drafted to the NFL, this study wants to further examine the co-relationship, if there is any, between the Combine results, draft order, and the NFL performance. Furthermore, the study conducted by McGee and Burkett only used the Combine results from 2000 and highlighted how significant Combine measurements were in predicting the draft order. However, some skepticism naturally rises about the correctness of their analysis since they have such small dataset, and especially for quarterbacks, there are only about ten of them who were eventually drafted and it is already unreasonable to derive a regression line for a sample of that size. The study of Kuzmitz and Adams is more intuitively reliable. However, their method of primarily relying on Pearson's test of correlation coefficient for paired samples might not be totally wrong, but with a fairly large sample size, the chances are that they would still be finding significant p-values even though the correlation itself barely exists. In this study with a even larger sample size, we will be looking into more systematic statistical techniques such as multi-linear regression and random forest.

Methods

Subjects Out of all the college football players who attended the NFL Scouting Combine from 2000 to 2010, players who played in QB, WR, and RB positions are specifically chosen. This choice of years was made so that the data include players who, on average, have roughly five or more years of NFL experience. It is worthwhile to note that many players who are drafted in the NFL do not make it into the final roster, and also many of them who do make it to the roster lose their job soon simply due to their inability to play in the professional level or due to injuries. Some players even take a break for a year and come back after an injury. This is the primary reason why this study does not include any of the recently drafted players, so the data are, in some sense, stabilized across different players, rather than consisting of statistics of players who have only played for a couple of years. The data of 212 quarterbacks, 489 wide receivers, and 298 running backs are used in this study (approximately 60% of them were eventually drafted).

Data The Combine results and draft orders were collected from <http://www.pro-football-reference.com>. The website is known to be trustworthy with all their data being almost perfectly accurate. The data include the following Combine measurements. Data related to players' NFL performance were collected from www.footballdb.com, another NFL database website. All data

were extracted and manipulated mostly using the 'XML' package in a well-known software 'R,' where further statistical analysis was performed as well.

40-Yard Dash The 40-yard dash drill is designed to test one's acceleration and speed. Electronic device is used to measure time to one-hundredth of a second. The athlete sprints from the starting line and runs straight as fast as he could until he passes the completion mark placed at 40-yard.

Vertical Jump The vertical jump drill measures one's leg strength and explosiveness. The athlete positions himself under the vanes of Vertec, then jumps straight up by first lowering his body and then pushing against the ground using only his leg strength. While up in the air, the athlete hits the highest vane possible with one of his hands. The measurement is the height of the highest vane subtracted by his standing reach.

225 Pounds Bench Press This drill is a good measure of one's upper body strength. The athlete completes as many bench press repetitions as possible.

Standing Broad Jump This drill measures one's lower body strength. The athlete positions himself behind the marked line, then once he is set, jumps horizontally off from both feet. The distance jumped is measured from the start line and to the point of heel contact or to the nearest body part.

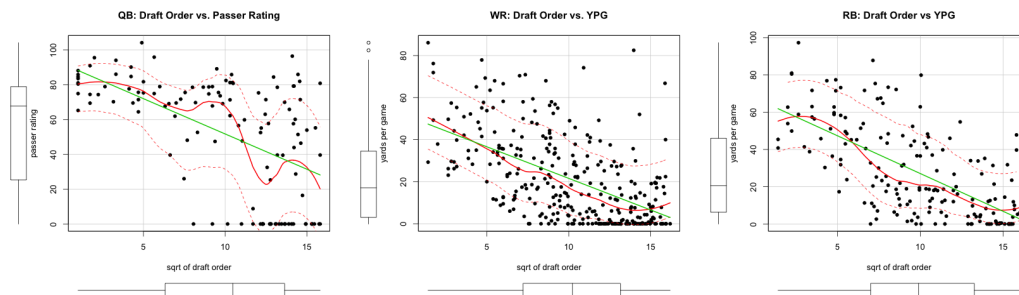
3-Cone The 3-cone drill measures one's agility and ability to change directions. The cones placed in the formation of an upside-down 'L.' The athlete starts in three-point stance at cone 1 and runs straight ahead for 5 yards to reach cone 2 and then immediately returns to cone 1. Without stopping, he changes direction and sprints and corners cone 2, and then move right away to cone 3, which is place 5 yards lateral to cone 2. He approaches the inside of cone 3, goes around it, returns and corners cone 2, and then completes the drill by sprinting past cone 1.

20-Yard Shuttle This drill is designed to measure one's speed and body control. The athlete starts in a three-point stance sprints 5 yards in one direction, quickly changes direction and runs 10 yards in the other direction, and then sprints back to the starting point.

Draft Order A draft order of a player is a specific number assigned to each player. Someone with a lower number of draft order is considered to be more valuable than others with higher numbers. Numbers can vary from 1 to 256, and those who are not drafted are automatically labeled as 257 for the purpose of this study.

Results

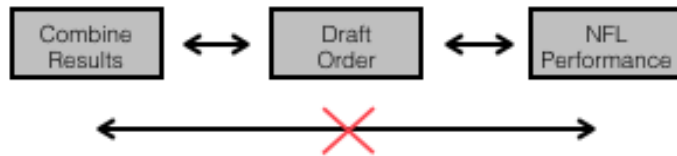
For all three positions, data analysis was divided into two parts. The first part was finding a correlation between draft order and NFL performance. Then the second part was to find a relationship between various Combine measurements and draft order. A simple plot for each position shows a great deal of correlation between draft order and NFL performance as shown below:



For these plots, only players who were drafted were included. Those who are lying on the horizontal axis represent players who did not make it to the final roster, which means having zero NFL performance measure. Speaking of performance measure, widely known ‘passer rating’ was used to assess the effectiveness of quarterbacks, while ‘yards per game’ was used for both wide receivers and running backs for their contribution to the game. Note that for these plots, we took the square root of draft order (on the x-axis) in a purpose of emphasizing the difference between lower numbers. For example, the difference between the 1st overall pick and the 20th over pick should be more noteworthy than that of the 200th and 220th picks. The computed correlations for these plots were below -0.5 for quarterbacks and wide receivers and below -0.6 for running backs, with extremely significant p-values. Also, another remark to be made is that the majority of the early round picks, especially quarterbacks, tend to have successful careers in the league, as opposed to the late round picks who tend to perform much worse on average.

Now that the relationship between draft order and NFL performance for all three positions is established, we now attempt to examine how the Combine results and the draft order are tied together. The reason for having draft order as a stepping stone between the two end points is

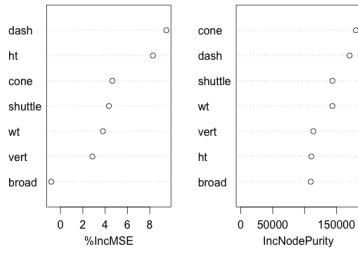
because when the Combine results were directly tested for association with NFL performance, they showed very little of it. One simple way to explain this is that when we have a correlation of 0.5 for A&B and also for B&C, we would probably end up with a correlation that is far less than 0.5 for A&C. Below figure is the visualization to compensate for not being entirely clear on describing the concept.



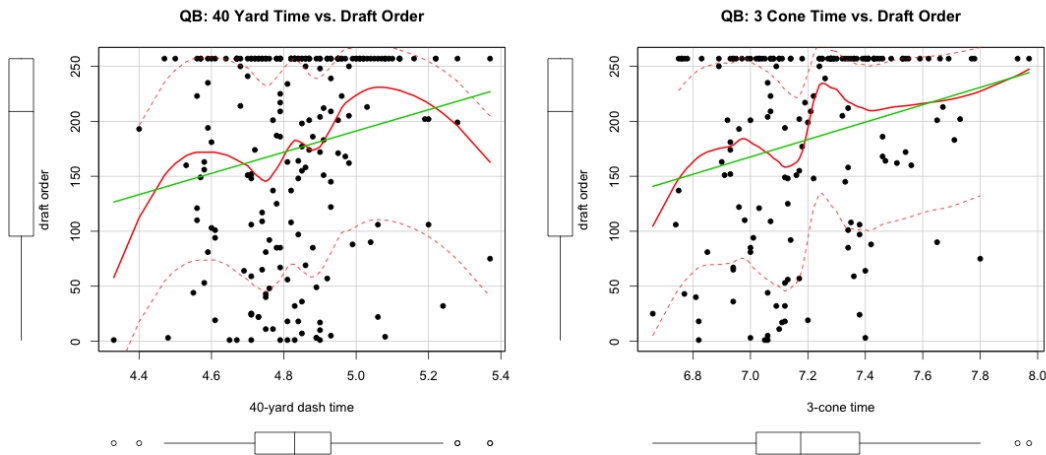
In our case, if we can show that having good Combine results have positive effects improving one's chance of being drafted early or being drafted at all, then we could safely assume that the effects are being carried over to one's NFL performance, since we have already made it clear of the relationship between draft order and NFL performance. Of course, they are not two completely linear processes and a good portion of information would be lost or deemed unnecessary when the two processes are combined, but think of it as a chain reaction where one good thing leads to another. While we are at it, addressing the other weakness of this study now should be appropriate. There are quite a bit of missing values for Combine measurements for players in general (except for 40-yard dash time; everyone has height, weight, and 40-yard measured), and since linearity between them is unlikely to be present, using any sorts of statistical imputation might be misleading. So in all of the statistical analyses that follow, missing values have been excluded from the process, which probably have affected the results slightly. This would need to be taken into consideration in interpreting the results.

Now that all the confessions are made, let's jump right in. We will start with quarterbacks. Despite of being the smallest sample size, they have the widest variations for the majority of the measurements. It might be because they do not really have to run fast or jump higher to be considered high potential. In this situation with overly nonlinear dataset, random forest, a machine learning technique, was used immediately, without much hesitation. Note that 225 pounds bench press drill was not included in this process because of too many missing values. Below figure is the result of random forest process showing the order of importance in terms of % increase in mean squared error and increase in node purity when the dependent variables (Combine measures) were tested for association with the response variable (draft order):

Quarterbacks: Random Forest Result



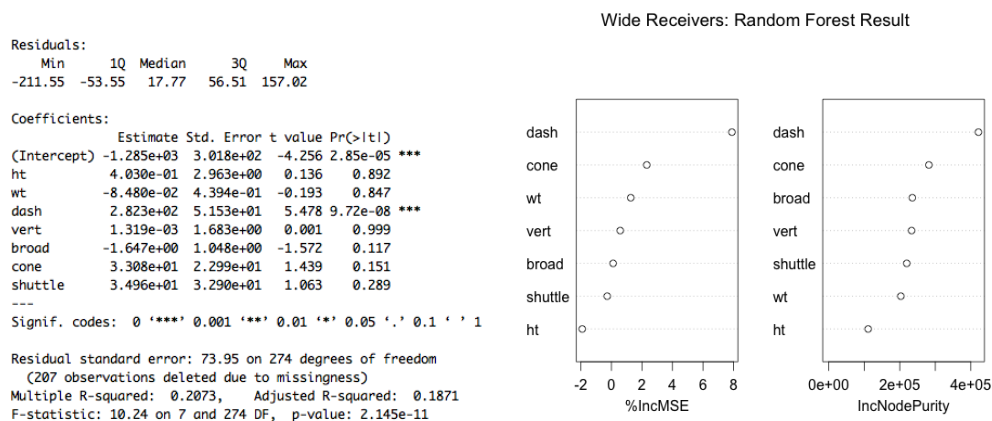
A brief explanation of what these terms mean would be suitable. Increase in mean squared error(MSE) is computed by taking the difference of the MSE of the original model and that of the pseudo model where the values of the corresponding dependent variable are replaced by random, yet realistic (permuted over the values of the entire dataset), numbers. Observe that the original MSE is always smaller. On the other hand, increase in node purity is computed by summing up the amount of decrease in nude impurity (residual sum of squares for decision trees) at each split. However, the fact that %Var explained in this model is less than 5% and no visible correlations showed in the scatter plots points to the conclusion that it is essentially not possible to build a predictive model for draft order solely based on Combine measures. For reference, the below figure consists of some of the scatter plots for the variables of high importance according to the random forest model.



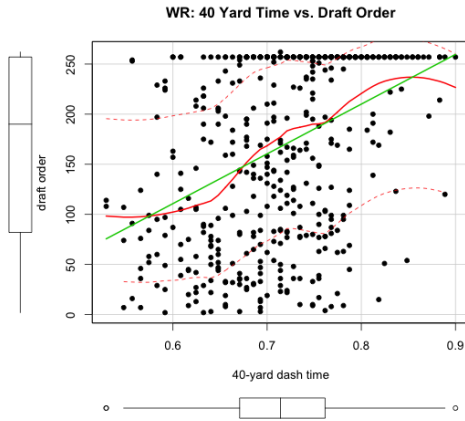
Numerous dots on the horizontal line just above the 250 mark represent players who were not drafted. Note that just how fast they ran 40 yards or 3 cone drill was not an important factor for being drafted. This is reasonable considering their roles in the game. However, it is still a little surprising to see doing well on the Combine rarely has any effect in increasing one's chance of being drafted. After all, what is being measured in the Combine is one's pure athleticism which should never be a bad thing to have for an athlete. This might be due to the

fact that many of the top college quarterbacks already know they are being drafted and have no exciting reason to perform well. Also, there are position specific drills for every position, and most of the qualities NFL quarterbacks ought to carry are best demonstrated by those drills that can actually test one's ability to throw the ball. Also, other important factors including hand size and arm length are not taken into account in the data this study collected, so we might have failed to grasp some of the elements that could have proven to be difference makers for having a better draft status. For example, this year's number one overall pick Jared Goff only had average measures for general Combine drills but he was known to do extremely well on the quarterback drills. Like we mentioned above, early round picks in the quarterback category tend to enjoy more successful career compared to others, so something must be there that the NFL scouts see to gauge one's potential; it is just that general drills do not matter much for them.

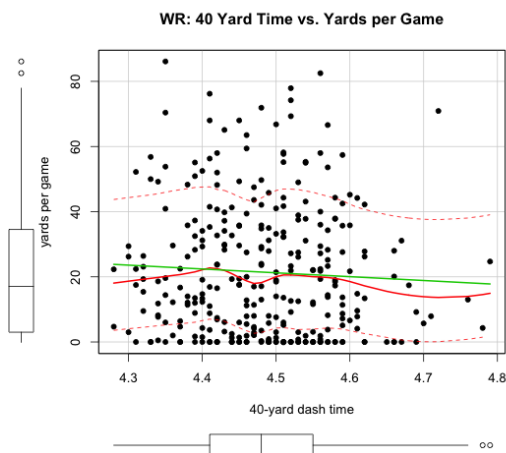
For wide receivers, let's first take a look at the result of an attempt to fit a multilinear regression model to the data, and also a random forest result of order of importance. 225 bench press drill was again excluded for the same reason as above.



We are looking at a mediocre model indeed. The adjusted R^2 is a little too low to convey any significance. A simple look at the scatter plots between the Combine drills and draft order reveals that this is largely due to nonlinearity of data. However, it is almost impossible to ignore from the above figure just how 40-yard dash variable in this model is behaving; it is strongly suggestive of something. The random forest process also gives a very similar suggestion of what 40-yard dash is liable of. Here is the scatter plot of 40-yard dash time and draft order. We first subtracted 4 and took the square root on the dash time to emphasize the difference between the faster players.



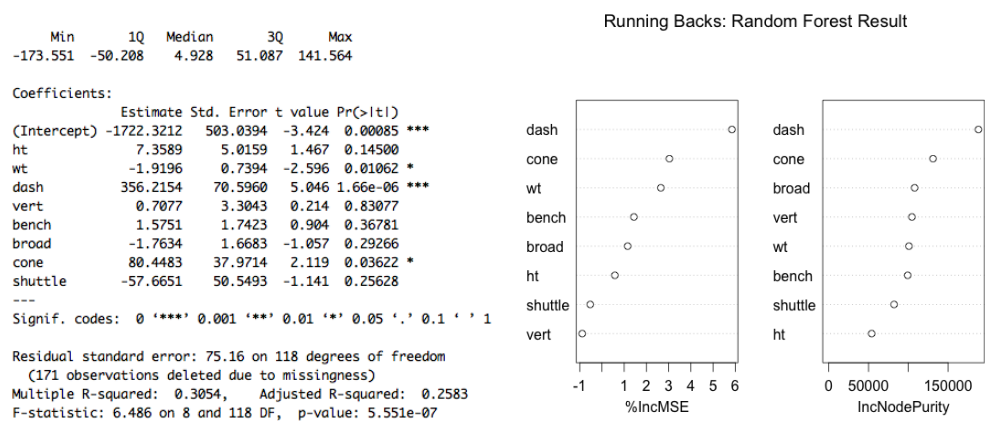
All those dots lined just above the 250 mark represent players who were not drafted. It seems fairly clear that 40-yard dash was an extremely important factor for draft decision. Further examination tells us that about 93% of players who were one standard deviation above the mean, which is equivalent to having run faster than 4.41, were eventually drafted, while only 25% of those who were one standard deviation below the mean, meaning they ran slower than 4.62 for 40 yards, were drafted. The mean time for quarterbacks was about 4.83. Even though all the other measures seem to be poor predictors, 40-yard dash alone could provide surprisingly trustworthy predictions of one's chance of ever being drafted if we look at it as a simple classification problem. That being said, even this single magical drill is probably not so adequate for predicting the exact draft order, since there is no obvious pattern showing in the plot if we disregard the dots for the undrafted. What this further suggests is that it would not be of any good at telling one's success in the NFL either. Here is a scatter plot consisting only of drafted players' 40 yard dash time and their career statistics.



It is safe to assume that there exists no correlation. Compared to the plot at the beginning of the result section with draft order on the horizontal axis, this plot is of no use. A short summary of this long story can be made. Acing the 40-yard dash will almost guarantee a

prospect's entrance to the NFL, but that ends about right there; once in the NFL, it certainly takes other things to succeed in the league. A plausible, yet an obvious, explanation for how early round picks perform better in the NFL is this: the scouts start looking for players long before the actual Combine, probably since they are in high school for elite players, and they seem to be impressively successful in telling whose talent will still prevail even at the professional level.

Let's now consider the last group of players, running backs. The analysis shall follow a similar procedure for that of wide receivers. Here are the summaries of multilinear regression and random forest processes.



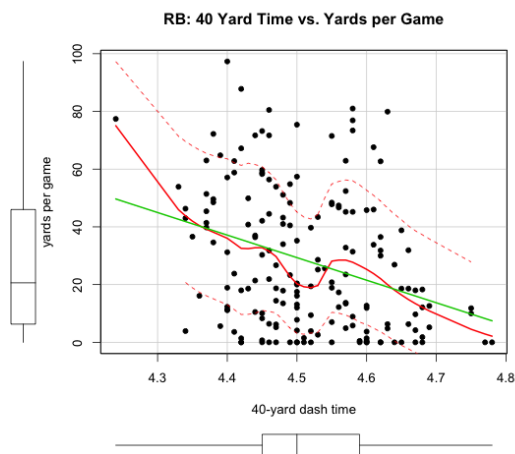
Although the linear model is also not too convincing, observe the consistency of the importance of variables between the two different processes. Other than 40-yard dash, which again is seemingly significant, 3-cone and weight also appear to possess a bit of information, even though the magnitude of it is quite weak relative to that of 40-yard dash. For verification, let's first look at their scatter plots.



The plot of 40-yard dash shows a similar trend to that of wide receiver's, while the other two plots show weaker trends. The correlations for these three are approximately 0.47, 0.26, and -0.17, respectively, all with highly significant p-values. Nevertheless, because the linear model itself is rather poor, these variables should be best when used for classification purposes.

85% of players who were one standard deviation above the mean in the 40-yard dash were selected to play in the NFL while only 30% of those who were one standard deviation below the mean were given the same privilege. However, when a similar method was applied to the other two variables, the differences in the percentage of the drafted and the undrafted were not significant, differing only by from 20% to 30%. This was somewhat expected from the random forest summary since dash is the only drill that stands out from the rest, while cone and wt are not too far away from the ones below them.

Yet, there is another remark to make in the 40-yard dash plot. It seems as if there remains a correlation even after disregarding all the dots on the 257 mark. This could mean that running backs who have run relatively faster in the 40-yard dash drill have higher chance of being not only being drafted, but being drafted early, which leads us to believe that on average, they will have better NFL career. So here is a scatter plot of all the drafted players' 40-yard dash time and career yards per game.



This correlation is noteworthy considering that the analogous plot for wide receivers had little correlation. Overall, running backs are more diversely affected by the Combine measurements in their draft status than wide receivers. However, 40-yard dash is the one drill that could make a difference. Even in the NFL, the ability to run fast for 40 yards seems to be a defining factor for dominating the game. For wide receivers, running fast did not simply translate to NFL success, but it appears that the notion of fast equals better is true to a certain degree for running backs.

Discussion

It definitely comes as a surprise that most of the Combine results lack any prediction power for neither draft order nor NFL performance. Raw athleticism may not be as important as people make it to be at the professional level. Frankly, every player who has made it to the NFL probably has enough of athletic potential to become an elite player. After that, work ethic, motivation, and attitude are what it takes to really dazzle. Moreover, players will all those values are quite likely to stand out even in their college years, and one's true talent should really reveal itself in a football field rather than at the Combine, especially when there are many other factors that can test one's mental strength such as weather issues and crowd noise. Since they are already considered top prospects long before the actual draft, that clarifies why many players who do not perform exceptionally well in the Combine still get drafted early and proceed to have decorated professional career. There will be special cases of course. We were able to verify one of them which showed that those who are in running back position show a fairly reasonable linear relationship between their speed, power and NFL performance. Extended research utilizing more data including position specific Combine drills and college statistics could deliver more accurate prediction of NFL performance for all positions. As for building a statistical model from data used in this study, it was actively disappointing, and often times, a mixture of various techniques had to be conducted just to make sure that we were not making up stories. Only after more research has been conducted better conclusions can be made.

References

Kuzmits, F, and Adams, A. The NFL Combine: Does it predict performance in the National Football League? *Journal of Strength and Conditioning Research*. 22(6): 1721-7, 2008.

McGee, K, and Burkett L. The National Football League Combine: A Reliable Predictor of Draft Status? *Journal of Strength and Conditioning Research*. 17(1): 6-11, 2013.