First-Round Privilege? Regression Discontinuities in the NFL Draft

Adam Finch

**Abstract**

I use a sharp Regression Discontinuity design to estimate the causal impact that the round in which an athlete is selected in the National Football League (NFL) draft has on career longevity and other outcomes. Rational Choice Theory suggests that behavior should not be influenced by sunk cost. This implies that coaches in the NFL should ignore where a player was selected in the draft when deciding which athletes to sign and play. I find no discontinuities in career longevity between athletes selected on either side of the thresholds, indicating that NFL coaches and teams are acting rationally in roster construction and playing time decisions.

**Introduction**

The National Football League (NFL) is the largest and most valuable professional sporting organization in the United States. Recent estimates place the combined value of all 32 franchises near 92 billion dollars (Barrabi 2020). Each year, team owners and managers are tasked with selecting new athletes to represent their organization in a process known as the NFL draft. The NFL draft consists of seven rounds where each team is given one draft pick per round. This paper addresses the causal impact that the round in which an athlete is selected has on their career longevity and other outcomes. In other words, how does the round a player is selected in impact how long their career will last? Outside of career longevity alone, I explore other outcomes such as number of pro bowls a player is selected to, total snap count during a player’s rookie season, and total number of games started within a player’s career.

This is an interesting question because fans, media, and most importantly coaches often label players as “first-rounders”, revealing the higher expectations that franchises have for the players selected earlier in the draft process. First-round draft picks are inherently more valuable than later round draft picks, and NFL franchises invest a lot of energy and money into these early picks. Even years into their career, first-round draft pick selections may be perceived as more talented simply because they possess that coveted label. Regret aversion, sunk cost fallacy, and confirmation bias are all factors that may lead coaches to give more opportunities to early round draft picks. Rational Choice Theory suggests that NFL decision-makers should ignore where a player was drafted when making decisions about future contracts and playing time (Lovett 2006). The draft pick used on a player and the financial compensation given to rookies should both be treated as sunk costs and not factor into future choices. Identifying the causal impact that draft round has on career longevity helps to uncover whether there is a pattern of irrational decision making among NFL coaches and team executives.

Considerable research on the sunk-cost fallacy in sports has already been completed. Keefer has examined the causal impact of NFL draft round on earnings (Keefer 2014), as well as playing time (Keefer 2017). Employing a Regression Discontinuity design, Keefer finds that there is a $240,000 jump in the compensation received by rookies selected at the end of the first round compared to those selected at the beginning of the second round. This confirms the assumption that NFL franchises invest more money into first round picks. Leeds, Leeds, and Motomura (2015) analyzed playing time at the cutoff between rounds of the National Basketball Association (NBA) draft. They found no evidence that playing time is influenced by sunk cost fallacy. However, both of these papers focus on short-term decision making. I focus on the long-term effects that draft round has on career longevity. As such, this research has the potential to add significant value in evaluating how coaches and team-owners make decisions regarding the opportunities that they give to players throughout the entirety of their careers.

**Data**

To complete this analysis, I utilize data from several different sources. Due to the abundance of information and meticulous stat keeping that occurs in a sporting league like the NFL, I was able to construct a fairly complete data set with several interesting variables. Most of the data comes from stathead.com, a website affiliated with the company Pro Football Reference. Stathead allows users to generate large datasets containing valuable information about a player's career, such as statistics on how many games they played in and how long their career lasted (Sports Reference 2022). I also used a helpful dataset on Kaggle, an online community platform for data scientists and economists. This data set contains information on every NFL draft pick selection dating back to the 1980 NFL draft (Gillies 2021). Finally, I took positional data from a publicly available GitHub database and merged these three datasets together to create a master database I could use to run regressions on in STATA (Sharpe 2022).

For my regression analysis, I focus on the first 3 rounds of the NFL draft. This is because of my hypothesis that the discontinuities will be greatest among the early rounds of the draft. The distinction between a first-round pick and a second-round pick is large, while fifth-round picks and sixth-round picks are viewed similarly. Additionally, I limit that dataset to players drafted between 1994 and 2009. This is because in 1994 the NFL expanded and began the 7 round draft process that they have used ever since. 2009 was chosen for the cutoff year because the majority of players who were drafted in 2009 are now retired. These limitations brought the dataset down to 1,555 players for observation.

Key variables include the number of seasons that a player was active in the NFL, the total number of career games a player participated in, and the pick number and round number where a player was selected. Additionally, the dataset includes control variables such as position, the team that drafted a player, and the year that a player was drafted. A table of summary statistics for main variables can be found below. This data is well-suited to answer the focus question because it is thorough, contains many years of data, and has most of the relevant variables needed to find a causal connection between what round an athlete is selected in and their career longevity. A potential weakness with this data set is that it does not include information on injuries. It is possible that a non-random distribution of injuries between first rounders and late round draft picks could cause players chosen early to have longer or shorter careers. Further research could involve a web-scraping project in order to find out which players suffered career ending injuries; excluding these players from the database may help to improve the accuracy of the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Standard Dev.** | **Minimum** | **Maximum** |
| Round Number | 2.03 | 0.82 | 1 | 3 |
| Pick Number | 49.13 | 28.12 | 1 | 103 |
| Number of Seasons Played | 7.46 | 3.81 | 0 | 20 |
| Number of Games Played | 95.59 | 57.12 | 0 | 287 |
| Number of Games Started | 66.29 | 58.72 | 0 | 286 |
| Number of Pro Bowl Appearances | 0.73 | 1.84 | 0 | 14 |

**Table 1: Summary Statistics**

**Identification Strategy**

In order to identify the causal impact that round number has on career longevity, it is important to address selection bias. The NFL draft does not randomly assign where players are drafted; players are selected based on their perceived talent level, among other factors. As such, more talented players get selected earlier in the draft, and less-talented players are typically selected in the later rounds. To combat this selection bias, I use a regression discontinuity design with thresholds at the cutoffs between rounds of the NFL draft. In the estimating equation, Yi is the outcome variable corresponding to the number of seasons played in a player’s career. It also includes two indicator variables that correspond to the two thresholds used (the cutoff between draft rounds). If both of those indicators are equal to zero it means that an observation was drafted in round 1. Xi is a vector of controls showing what position a player is and how many wins the team that drafted them had during the previous season.

Yi = β0 + β1𝟙(Round2) + β2𝟙(Round3) +β3pickNumber + β4Xi + εi

Using pick number as the running variable makes it easy to examine observations near the threshold; players selected just before and the draft round switches. However, although it is customary for each round to have 32 picks, compensatory picks are also sometimes awarded to teams, making it so that some rounds have more than 32 picks. Occasionally, NFL teams are disciplined and their picks are taken away, leading to some rounds with fewer than 32 total picks. Though these cases are rare, instances like these have the potential to muddy the Regression Discontinuity because the threshold would be inaccurate for those years. To combat this problem, and to ensure a strict RD design, I implement a standardized ranking where every pick in the first round is given a decimal value between 1 and 2, every pick in the second round is given a decimal value between 2 and 3, and so on. This ensures that the thresholds remain accurate even in unconventional years where there are more or less than 32 picks in a given round.

Whichever side of the threshold these observations fall can be used to approximate the treatment variable of round number. A few other assumptions are needed to interpret this estimate as causal. The main assumption for a Regression Discontinuity design is that there is no bunching at the threshold. In other words, individuals cannot precisely manipulate which side of the threshold they land on. In the context of this paper, I assume that athletes cannot precisely manipulate whether they are chosen at the end of the first round or the beginning of the second round. While it can be argued that athletes can work harder to manipulate where they end up in the draft, when focusing on observations close to the threshold window this assumption is likely to hold.

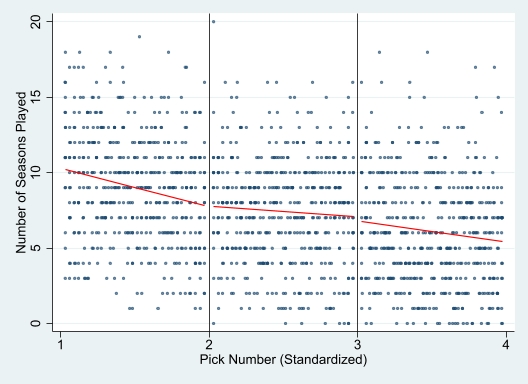
The second assumption for Regression Discontinuity Design is that there is smoothness of covariates across the threshold. In other words, there are no other “jumps at the jump” outside of the treatment -- switching from one round to the next. The process by which the NFL decides the drafting order for teams raises some concerns in regards to this assumption. The first pick in each round is awarded to the team with the worst record in the previous season, and the team that wins the championship is given the last pick in each round. This means that rookies chosen at the end of the round are sent to teams who did very well in the previous season, whereas rookies chosen at the beginning of a new round (just across the threshold) are sent to teams who did very poorly. Despite this potential violation, there are a few reasons to believe that this assumption could still hold. First, NFL teams often engage in trading their draft picks. This means that although the original order is heavily correlated with team success from the previous year, teams that did poorly can trade for the draft picks of teams that did well, and vice versa. Second, NFL rosters change immensely from year to year. Players are traded, contracts expire, and coaches are fired. Free agency is always a period in which teams undergo dramatic transformation. This ensures at least some level of randomness. Additionally, I control for team wins in the previous season in order in the estimating equation, eliminating the risk of this bias impacting the causal estimate.

**Results**

Extensive testing revealed no significant causal impact of the round a player is drafted in and career longevity. As expected, there was a negative relationship between the pick number a player is drafted at and number of seasons played. However, there were no discontinuities or jumps at the thresholds between rounds. A scatterplot between the outcome variable (number of seasons played) and the running variable (composite pick number) shows that although the rate of decline is steeper for observations in the first round of the NFL draft, there is no noticeable jump when switching from the first round to the second round.

Similar results are found when examining other outcome variables, such as the number of Pro Bowls a player is selected to play in, or the total number of games a player starts in their career. In each case, there is a clear negative relationship showing that the further down in the draft a player is selected, the less likely they are to have success in the league. However, there are no discontinuities, implying that perhaps the round number in which a player is drafted is not seen as an important factor to coaches and team owners. These relationships can be visualized

in figures 1 and 2.



**Figure 1: Outcome Variable – Number of Seasons Played**

**Chart, scatter chart

Description automatically generated**

**Figure 2: Outcome Variable – Number of Games Started**

Running a formal regression test confirms the insignificance of these thresholds. In a regression using the number of seasons played as the outcome variable, the coefficients for both variables indicating whether or not an observation was past the threshold had statistically insignificant p-values at even the 10% significance level. The coefficient on pick number was statistically significant in all of the regressions that I performed, confirming the obvious relationship between being chosen later in the draft and less career success for an NFL athlete down the road. Additionally, there were several statistically significant coefficients on various position titles. To interpret these coefficients, it is important to note that the position left out was offensive lineman; therefore, these coefficients are describing how a specific position impacts the outcome variable *in comparison to* offensive lineman. The full regression results for three separate outcome variables is displayed in table 2.

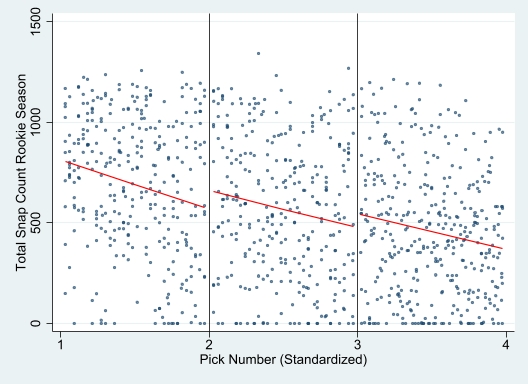
|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| VARIABLES | Number Of Seasons | Games Started | Pro Bowl Appearances |
|  |  |  |  |
| R2indicator | -0.0470 | -4.659 | -0.202 |
|  | (0.442) | (6.576) | (0.227) |
| R3indicator | 0.231 | 0.807 | 0.429 |
|  | (0.810) | (11.74) | (0.399) |
| Pick Number  (Standardized) | -1.625\*\*\*  (0.392) | -28.88\*\*\*  (5.586) | -0.838\*\*\*  (0.192) |
| Team Wins –  Previous Season | -0.00431  (0.0345) | -0.584  (0.477) | -0.00459  (0.0158) |
|  |  |  |  |
| Defensive Back | -1.778\*\*\* | -44.09\*\*\* | -0.825\*\* |
|  | (0.660) | (10.79) | (0.391) |
| Defensive End | -2.180\*\*\* | -50.98\*\*\* | -1.053\*\*\* |
|  | (0.704) | (11.26) | (0.401) |
| Defensive Tackle | -2.476\*\*\* | -50.59\*\*\* | -1.075\*\*\* |
|  | (0.724) | (11.60) | (0.401) |
| Fullback | -1.117 | -46.88\*\*\* | -1.125\*\*\* |
|  | (1.136) | (12.72) | (0.398) |
| Guard | -2.138\*\*\* | -31.03\*\*\* | -0.515 |
|  | (0.747) | (11.96) | (0.450) |
| Kicker | 1.824 | -97.72\*\*\* | -0.711 |
|  | (1.917) | (13.20) | (0.455) |
| Linebacker | -2.126\*\*\* | -41.78\*\*\* | -0.681\* |
|  | (0.682) | (11.09) | (0.403) |
| Nose Tackle | -1.778 | -14.94 | 1.363 |
|  | (2.959) | (41.79) | (1.244) |
| Punter | 3.727 | -82.48\*\*\* | 0.404 |
|  | (3.240) | (12.03) | (0.725) |
| Quarterback | -0.759 | -51.08\*\*\* | -0.572 |
|  | (0.863) | (12.94) | (0.492) |
| Runningback | -2.529\*\*\* | -66.90\*\*\* | -0.798\*\* |
|  | (0.685) | (11.05) | (0.405) |
| Tight End | -0.847 | -32.87\*\*\* | -0.709 |
|  | (0.744) | (11.84) | (0.439) |
| Wide Receiver | -2.464\*\*\* | -57.33\*\*\* | -0.973\*\* |
|  | (0.686) | (11.10) | (0.399) |
| Constant | 13.46\*\*\* | 190.9\*\*\* | 3.606\*\*\* |
|  | (0.840) | (13.24) | (0.496) |
|  |  |  |  |
| Observations | 1,555 | 1,526 | 1,555 |
| R-squared | 0.147 | 0.224 | 0.116 |
|  |  |  |  |
|  |  |  |  |

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 2: Regression Results**

Although the above regressions show that there is not a long-term causal impact between the round in which an athlete is drafted and their career longevity, I wanted to do similar testing to see if there is any effect in the short term. The regression analysis performed above uses NFL draft picks from the years 1994 - 2009; however, similar data is available for more recent years. In 2012, the NFL began tracking snap count data. This means that precise data exists for how many plays a rookie is on the field for during their first season in the league. Figure 3 focuses on data from 2012-2021, where the outcome variable is total snap count, and the running variable remains a standardized ranking of pick number. Once again, no notable disadvantages in opportunities exist for players drafted at the beginning of the 2nd and 3rd rounds compared to the end of the first and second rounds. In fact, a positive jump at the threshold indicates that players going to teams selecting at the beginning of a draft round receive *more* of an opportunity compared to players selected at the end of a round.



**Figure 3: Outcome Variable – Total Snap Count (Rookie Season)**

A few potential mechanisms exist to help explain these results. The simplest explanation is that the round in which an athlete is drafted does not have a causal effect on career longevity or other related outcome variables. This is certainly plausible; NFL coaches often claim that accolades and prior achievements don’t matter at all once training camp starts. Additionally, there are plenty of success stories from late round draft picks or even undrafted free agents making it as successful players in the NFL. In the words of Hall of Fame wide receiver Steve Smith Sr., “making it in the NFL comes down to talent and how a player works and applies himself each day. Every year, former first-rounders get cut or flame out, while one-time underdogs sign new contracts” (Smith Sr, 2018). These results point towards the likelihood that NFL decision-makers are not swayed by sunk cost and are acting within the bounds of Rational Choice Theory. However, it is also possible that inadequate control variables or violated assumptions are the reason for these tests showing that no causal relationship exists. To remedy this, further testing involving team and yearly fixed effects could help to discover data trends that the Regression Discontinuity Design was unable to uncover.

**Conclusion**

Although testing found no causal impact between the round in which an athlete is drafted and their career longevity, further testing may be helpful to continue evaluating this hypothesis. Whether or not a causal effect exists, more exploration regarding this question and similar ones can help to uncover the economics behind sports; specifically, how coaches and team owners make decisions about who they sign onto their teams. In an industry as wealthy as the NFL, answering questions like these could have enormous impacts on team revenue and overall success.

**References**

Barrabi, T. (2000, February 4). What is the NFL worth? Revenue, team values, and other

financial facts. *Fox Business.* https://www.foxbusiness.com/sports/nfl-worth-revenue -team-values

Gillies, K. (2021). *NFL Statistics.* Version 3. Retrieved on November 2, 2022 from

https://www.kaggle.com/datasets/kendallgillies/nflstatistics?resource=download

Keefer, Q. A. W. (2014). Rank-based groupings and decision making: A regression discontinuity

analysis of the NFL draft rounds and rookie compensation. *Journal of Sports Economics*. *17*(7), 748-762. https://doi.org/10.1177/1527002514541448

Keefer, Q. A. W. (2017). The sunk-cost fallacy in the National Football League: salary cap value

and playing time. *Journal of Sports Economics 18*(3), 282-297. https://doi.org/10.1177/152700251557451

Leeds, D., Leeds, M. A., & Motomura, A. (2015). Are sunk costs irrelevant? Evidence from

playing time in the national basketball association. *Economic Inquiry*. *52*(2), 1305-1316.

https://doi.org/10.1111/ecin.12190

Lovett, F. (2006). Rational choice theory and explanation. *Rationality and Society, 18*(2),

237-272. https://doi.org/10.1177/1043463106060155

Sharpe, L. (2022). *Draft Picks*. Version 1. Retrieved on November 2, 2022 from

https://raw.githubusercontent.com/leesharpe/nfldata/master/data/draft\_picks.csv

Smith Sr, S. (2018, September 19). Steve Smith Sr.'s five keys to making it in the NFL. *NFL*

*Network.* https://www.nfl.com/news/steve-smith-sr-s-five-keys-to-making-it-in-the-

nfl-0ap3000000963953

Sports Reference LLC. 2022. *Stathead and Pro-Football Reference Data.* Retrieved on

November 2, 2022 from https://stathead.com/