Indiana Pacers Performance Analytics Internship Project: Written Overview

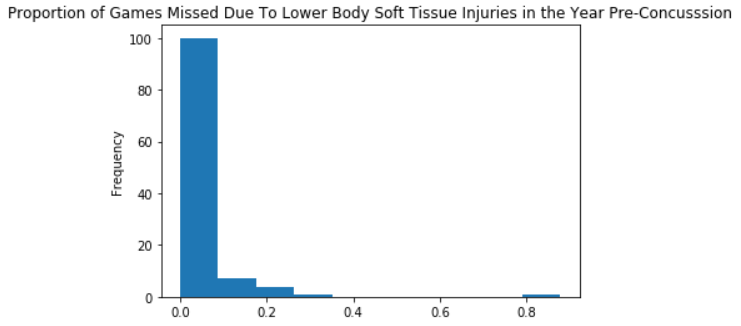
By: Tyler Fuelling

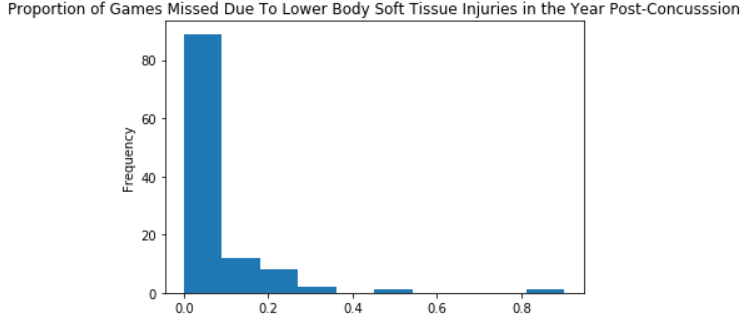
**Synopsis:** The general idea behind this project was to examine the effects of concussions on NBA basketball players, both in terms of their health and in terms of their game performance. The goal was to generate a list of every player that missed a game during the last 10 seasons due to a concussion and then to analyze that dataset. Meaning that my first step for the project was to construct the dataset. I was unable to find one website containing all of the data that I hoped to use for this project, so I used a combination of filtering in both Excel and Python to create a dataset in a Microsoft Excel sheet using data from multiple different websites. After the creation of the data set, I was able to move on to the two main parts of the project: analyzing the relationship between concussions and lower body soft-tissue injuries and the relationship between concussions and a player’s various shooting percentage statistics.

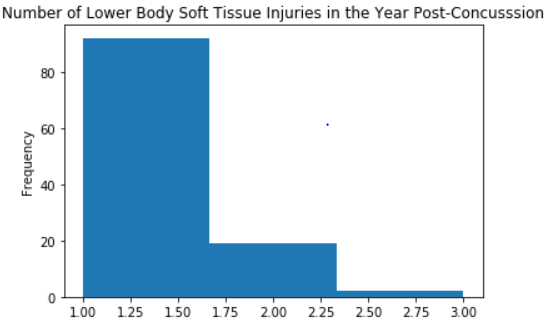
**Concussions and Lower Body Soft-Tissue Injuries:**

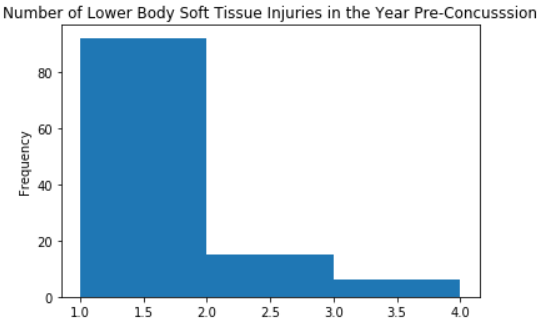
* **Injury Frequency Pre- and Post-Concussion (Lower Body Soft-Tissue Injuries):**

In order to compare injury frequency pre- and post-concussion, I opted to compare the relative frequency of a player suffering a lower body soft-tissue injury in the year (82 games) directly prior to his suffering a concussion to the relative frequency of that same player suffering a lower body soft-tissue injury in the year (82 games) directly following his return to the lineup after sitting out due to the concussion. Using the database on [prosportstransactions.com](https://www.prosportstransactions.com/basketball/Search/Search.php), I was able to determine how many games each of the player’s in my concussion data set missed due to a lower body soft-tissue injury in the 82 games directly preceding their concussion as well as the 82 games directly following their return from the concussion. After that, I was able to use those numbers to calculate a relative frequency a player missing a game for a lower body soft-tissue injury both before and after his concussion. Next, I performed a paired student’s t-test as a method of hypothesis test on the two paired samples. Because the samples were paired, (same player, before and after a concussion), it made sense to perform a paired t-test. However, the test returned a p-value of 0.17702, which is not less than 0.05, and thus not low enough to reject the null hypothesis. After this, I decided to also perform the same hypothesis test on the distributions of number of lower body soft-tissue injuries in the years both pre- and post-concussion. I used [prosportstransactions.com](https://www.prosportstransactions.com/basketball/Search/Search.php) to determine the number of injuries that occurred over the course of 82 games as well. After that, I performed the hypothesis test, which returned a p-value of 0.25158. This is not close to less than 0.05 and thus not low enough to reject the null hypothesis. So, because I was not able to reject the null hypothesis for either of those two hypothesis tests, I was not able to find any evidence that the two distributions were significantly different. As I result, I concluded that there was not a noticeable enough difference in injury frequency pre- and post-concussion to say that concussions definitively had an effect on the frequency of lower body soft-tissue injuries. For further evidence, I have provided histograms of the four distributions mentioned above; and when looking at them, there is not a significant difference between the pre-concussion distributions and the post-concussion distributions.



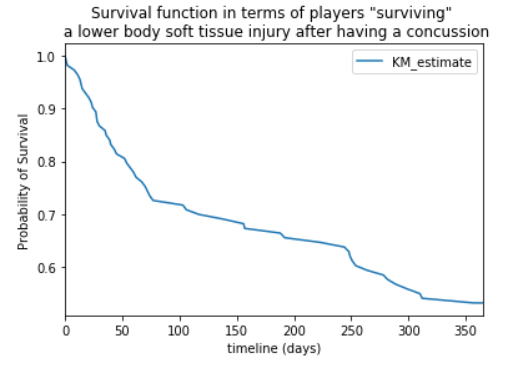




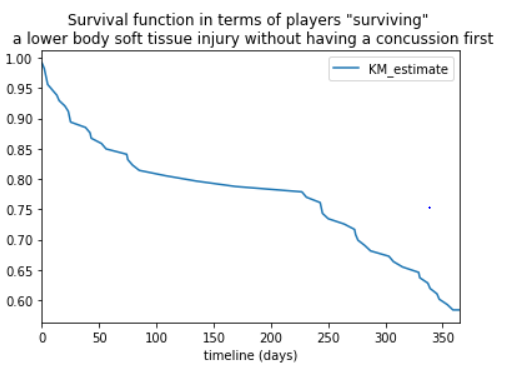


* **Are players more likely to suffer a lower body soft-tissue injury following a concussion?**

My approach to answering this question was to construct a survival function and corresponding survival curve for lower body soft-tissue injuries both pre- and post- concussions. A survival curve is typically used to analyze things like the lifetime of a population, but in this case I decided to use it in such a way where the concussion was the “birth” event, and the suffering of a lower body soft-tissue injury was the “death” event. This way I could see how likely it was for players to last or “survive” certain periods of time after a concussion without sustaining a lower body soft-tissue injury. I was able to obtain the following survival curve for the year following a player suffering a concussion:



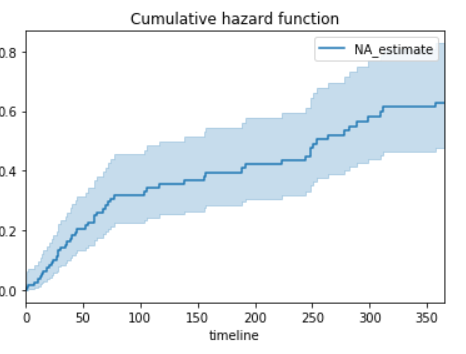
After creating the survival function and curve for the year following a concussion, I needed to have some sort of control to compare it to, in order to determine if it was actually more likely for a player to suffer a lower body soft-tissue injury following a concussion than at other times. So, I created a survival curve with the same premise, except that it would just be for any year rather than specifically the year following a concussion. In order to do this, I used my data for lower body soft-tissue injuries suffered in the year before a player suffers a concussion, and I was able to create the following survival curve:



We can see just by looking at the two survival curves, that NBA players have a higher probability of surviving a certain number of days without a lower body soft tissue when they are not coming off of a concussion. For example, it seems as though an average NBA player would have about an 80% chance to “survive” 200 days without a lower body soft-tissue injury; however, this chance drops to about 65% when it is 200 days directly following his return from a concussion.

In order to further expand on this, I created confidence intervals for the median survival times in each of these two cases. When following a concussion, the 95% confidence interval was [264, inf); but when not following a concussion, the 95% confidence interval was [354, inf). Interpreting this, I can now say with 95% confidence that the median “survival” time for a regular NBA player is greater than or equal to 354 days, while it is a much lower 264 days for NBA players returning to the lineup from a concussion. Based on this data, I would say that it is more likely for an NBA player to suffer a lower body soft-tissue injury following a concussion.

* **Does Literature from other sports that shows that the likelihood of injury increases following a concussion for up to 12 months, with the highest risk for injury occurring in the first three months followed by a depreciated risk between 3-6 months, and then another decrease between 6-12 months, hold true in the NBA?**

My strategy for approaching this question was to use the data that I had to create a cumulative hazard function for the year following an NBA player suffering a concussion, where the “hazard” is a lower body soft-tissue injury. The hazard function is the probability of the death event occurring at time *t*, given that the death event has not occurred until time *t*. Using the same data that was used for the survival curve, I was able to create this cumulative hazard function that displays the hazard of a lower body soft-tissue injury during the year following an NBA player suffering a concussion.

Just by looking at the plot, we can see that the hazard being experienced by the subjects is generally decreasing over time, simply due to the fact that the slope is mostly decreasing over time. However, I want to see if the highest risk for injury (hazard) occurs during the first three months after the concussion, with a decrease from 3-6 months, and then another decrease from 6-12 months; so, I went ahead and calculated the slope of the cumulative hazard function in these three separate regions.

The result was a slope of 0.00321550, in the first three months, a slope of 0.00089963 from months 3 to 6, and a slope of 0.00125132 from months 6 to 12. It is clear that the slope is the highest, and thus the risk/hazard is highest within the first 3 months of the concussion; however, there was actually a larger slope, meaning higher risk, from months 6 to 12 than months 3-6. This of course could be just an error in the data or analysis, but it is something that might prompt more investigation in the future.

**Concussions and Shooting Percentage:**

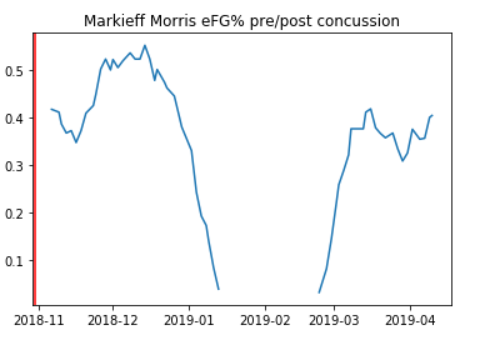
* **When players return from concussion does the FG%, eFG%, 3PT%, and/or FT% decrease, and if so, how long does that decrease last for?**

My strategy for answering this question was to use the idea of rolling means/averages in a little bit of a different way. I decided to keep a rolling average for eFG% across 10 games; however, for FG%, 3PT%, and FT%, I just kept a rolling cumulative percentage across the player’s last ten games. The first step in answering this question was to create a dataset with all of the statistical information I was going to be examining as well as the concussion data that I would need to investigate this question. In order to do this, I got the statistical shooting data from [basketball-reference.com](https://www.basketball-reference.com/) and I obtained more of the concussion data from [spotrac.com/nba/injured-reserve](https://www.spotrac.com/nba/injured-reserve/2018/). Due to time constraints, I made the decision to analyze just the 2018-19 NBA season, assuming that the trends in how concussions affect shooting percentages will remain constant across different seasons; however, that is definitely one area for future expansion and improvement with this project. Once I had the data that I needed, I created a dataframe for each player that contained all of his game logs for the 2018-19 season. Using this, I was able to calculate 10-game rolling averages (or cumulative percentages over the most recent ten games) for the four statistics that I wanted to examine. After this was done, I wanted to compare the results pre- and post-concussion. In order to make the comparison as fair as possible in my opinion, I decided to use the very last rolling mean before a player suffered a concussion (the game right before the concussion occurred) because I felt like that was the most reasonable projection of how he would have continued to play had the concussion not occurred. On the other hand, for the post-concussion value that it will be compared to, I decided to use the rolling mean from 10 games after he returned from a concussion. I chose this because the rolling means were calculated over a period of 10 games and I thought it would be most fair to compare how the player played in his first full “period” back from a concussion in which a full rolling mean could be calculated. This is also another interesting area for possible expansion on this project because it is possible the results could be somewhat different depending on the number of games used to calculate the rolling means.

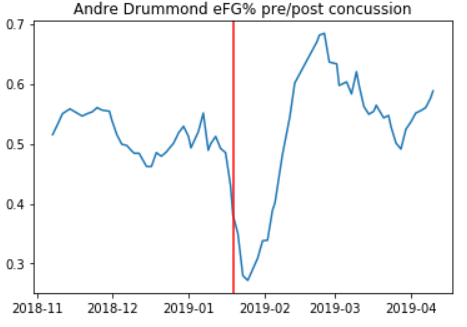
After all of this, I compared the statistics before and after the concussion and found little to no difference. In fact, some of the statistics were actually better in the 10 most recent games after a concussion than they were in the 10 games directly leading up to it. So, I concluded that there was not enough evidence to say that eFG%, FG%, 3P%, or FT% decrease when a player returns from a concussion. Lastly, because there was no noticeable decrease, the second question about how long the decrease lasts for as well as the very last question about whether a possible decrease in shooting performance is correlated to how long a player is out for after suffering the concussion are no longer pertinent.

I will include some of the graphs tracing the rolling means for shooting percentage statistics throughout the year in order to further show that there is no noticeable decrease in the statistics that I was examining from pre-concussion to post-concussion. The vertical red lines on the graphs mark the date at which the player suffered the concussion.

Markieff Morris is an example of a player who was not beneficial to examine because his concussion occurred too early in the season to have reliable data beforehand.

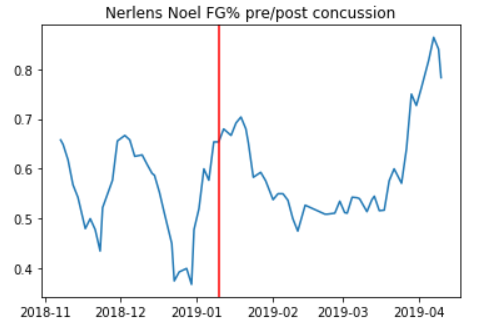


We can see that the rolling mean of Andre Drummond's eFG% drops suddenly after his concussion, but this is because he is sitting out games during that period. Once, he has played 10 games, enough to get an accurate sample for the rolling mean, his eFG% is back to normal.

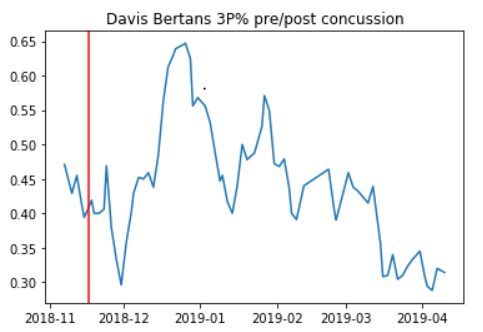


We can see that the rolling mean of Nerlens Noel's FG% stays on the same trend post

concussion as it was on pre-concussion.



We can see that the rolling mean of Davis Bertan's 3P% stays on the same trend post

concussion as it was on pre-concussion.