**Intro/Background:**

In professional sports, determining when an injured player should return to the team is often a difficult decision that involves conflicting interests. If the player is skilled, it is likely that the team will have a strong desire for the player to return as quickly as possible. This pressure from the team may lead to a premature return to action by the player, putting them in a compromised position physically. On the other hand, sports is a career to the player, so the player may want more time for rest and recovery than is necessary in an attempt to nullify the risk of re-injury and secure their financial future.

**Problem Definition:**

In an attempt to lessen the divide between the conflicting interests of the team and the player, this project will formulate two models: one that predicts player performance post-injury and one that predicts whether or not a player will re-aggravate a previous injury. Assuming quality predictive models are created, this would allow both the team and the player to make more educated decisions regarding the amount of treatment needed for certain injuries.

**Methods:**

Presently, we have data on all of the injuries that have occured in the NBA from 2010 to 2020. Additionally, we have access to player statistics over the last decade. Cleaning and organizing this data will likely be the most involved and critical part of the model-making process because for each injury we have to combine the injury data (time missed, injury type, etc.) with the relevant player statistics before and after the injury. We plan to categorize the data by injury in order to keep variables that are not accounted for in our models as consistent as possible. For example, we will be ignoring injury treatment methods. These treatment methods will be much more similar for two instances of the same injury as opposed to two different kinds of injuries. So, by categorizing by injury, it is likely that injury treatment methods will have a diminished influence on our models.

As a tentative plan, we may do some general regression modeling to determine which statistics are most relevant to each of the two models. To do this for the re-injury model would need to make a more general regression model that measures the relation between a certain statistic, such as age, and time between injuries (time between the most recent previous injury and the current). After determining the most relevant input parameters, we will use a portion of our data to train a neural network that, given a particular injury and the relevant input parameters, outputs the predicted performance of the player post-injury (for example, the model may predict a 5% scoring decrease post-injury). Additionally, the neural network will output whether or not the player will himself in the same manner within some time frame. As an alternative, we may move away from the discrete will-reinjure/will-not-reinjure output to an output that estimates how long the player will play without re-injuring himself.

**Potential Results/General Discussion:**

Looking at the data of injuries in the NBA over the past decade, we can stumble across multiple findings. Our models could lead us to identify the worst type of injuries - having the greatest negative effect on a player’s performance - as well as the likelihood that a player will reaggravate an injury considering multiple factors such as recovery time and game load after returning from injury. A study by Sports Health looked at injuries in the NBA over a 17 year period and concluded that “Patellofemoral inflammation [was] the most significant problem in terms of days lost in competition” and that ankle sprains were the most common injury. Our own models could come across similar findings and find correlations between the type of injury and effect on the player’s performance. Also, similar to studies done by Journal of Athletic Training, our predictive models can find how the chance of getting injured decreases

**References:**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8128995/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107769/>