2018

The Effect of Injuries on Player and Team Performance: An Empirical Analysis of the Production Function in the National Hockey League

Ryan Warnock
Skidmore College, rwarnock@skidmore.edu

The Effect of Injuries on Player and Team Performance: An Empirical Analysis of the Production Function in the National Hockey League

By

Ryan Warnock

A Thesis Submitted to

Department of Economics

Skidmore College

In Partial Fulfillment of the Requirement for the B.A Degree

Thesis Advisor: Qi Ge

May 1, 2018

Abstract

This paper analyzes the effect that injuries have on team performance in the National Hockey League (NHL) and on the production of the individual athlete. By using player level statistics and team level performance metrics from the 2013/2014 through the 2016/2017 NHL seasons, my analysis adds to the current production function literature in sports economics by incorporating injury data to put forth a more comprehensive production frontier. My results suggest that there is a statistically significant negative effect on both team performance, and on individual production when players are injured. This paper begins by employing a probit regression model to identify the most significant contributors to injury in professional hockey, followed by an interrogation of individual level performance, and finally a comprehensive team level analysis, which shows that not only do injuries negatively affect a team and individual performance, but there is also a magnified effect when a player is more valuable to an organization. The results suggest shorter seasons, and more evenly distributed playing time as ways to protect players and ensure more success for teams.

Table of Contents:

I. Introduction:

Injuries are a common occurrence in professional sports. They occur frequently depending upon the sport and are an inevitable component of being an elite athlete. In professional hockey, injury rates rank amongst the highest given the high speed and high intensity nature of the sport. Dealing with injuries poses an immense challenge to coaches, owners, and league officials in the NHL and other professional sports leagues. Coaches and managers must adjust game plans and strategies according to the inputs that are healthy and available. Injuries not only disrupt team chemistry but they also limit a team's on ice production. The issue of player safety and injury goes beyond a fundamental moral obligation that coaches have to keep players safe; but as this essay will attempt to show by analyzing the production function in professional hockey, teams that struggle to stay healthy also rarely reach the playoffs or have postseason success.

The existing literature surrounding production in professional sports is quite expansive. Rottenberg (1956) and Scully (1974) were pioneers in the application of the production function in Major League Baseball. Their works introduced the idea that a player's talents and skills can be directly linked to a team's success. Krautmann (1999) built off of the aforementioned papers to incorporate a wider, more expansive set of variables to more accurately determine what factors contribute to a team's production of wins in the MLB. Following Rottenberg (1956), Scully (1974), and Krautmann (1999), was Zak et al. (1990). There, the authors adopted a similar production function framework to the National Basketball Association to determine that shooting ability was the most significant indicator of a team's production of wins. Berri (1999) attempted to look at production at a more individual level in the NBA, by attempting to isolate each player's marginal product. His analysis introduced variables that controlled for factors such as home court advantage and incorporated other ball handling metrics to determine who was the most valuable player in the 1997/1998 NBA season. A similar methodology was applied to Premiere League soccer by Carmichael et al. (2001) who was in agreement with Zak et al. (1990) and Berri (1999) that production in professional sports can be categorized into three groups: scoring, defense, and ball handling. Empirical analyses focused on professional hockey is the sparsest among the other North American pro-sports leagues. The main focus of analysis in the NHL is primarily concerned with discriminatory hiring and compensation. Jones and Walsh (1988), Jones et al. (1999) and Kahane (2005) are the most notable contributors to the NHL literature with regard to discrimination against players of French-Canadian descent. Their works have translated the production function framework to fit the NHL and the dynamics of hockey, but as is the case with all of the previously mentioned literature, there is no incorporation of player injury data.

That is where my study will attempt to fill the gaps. My analysis will offer a comprehensive production function framework for the NHL that incorporates player injury data. The main question that this essay will attempt to answer is to what degree is a team's success limited when players are hurt? And on a related note, I will also explore what factors contribute to injury and the ways in which injury affects a player's individual performance. To do so, I have collected 2,954 distinct player level observations and 120 team level performance summaries across four NHL seasons from 2013/2014 to 2016/2017. Through the construction of three different models that touch on the different levels of production in the National Hockey League, this essay will fill in the gaps that exist in sports economics production function literature by investigating the NHL with the addition of injury data. As the results will show, there is a distinct relationship between the number of injuries that a team incurs and the success the team has throughout the season. This will be accomplished by first isolating the factors that contribute to injury from a medical and exercise science perspective. My results are in agreement with exercise science professionals that the factors that are most likely to contribute to sports injury probability are, competition hours, unique game characteristics, age, and, behavior/playing style, I was able to translate that to the National Hockey League. By doing so, the results have shown that the player's age, and ice competition time are the most significant contributors to the injury rate in professional hockey players.

Once I have grounded the medical literature in the National Hockey League, I then assessed how injuries affect performance at the individual player level. By adopting basic frameworks from the existing literature, I was able to create a more comprehensive production function for the National Hockey League. This was accomplished by recognizing that in most team sports, production and success at the player level depends on three factors: scoring ability, defensive skill, and the ability to obtain and retain possession of the ball or puck. This framework was adopted from the work of Zak et al. (1990) and Berri (1999) in the NBA, as well as Carmichael et al. (2001) in Premiere League Soccer. However, as previously mentioned, the absence of injury data was a glaring limitation of these models. That is where my study attempts to fill the gaps; by introducing a player level dummy variable, and the number of games the player misses, I was able to analyze how a player’s performance suffers when he is injured. It was revealed that when it comes to the injury data, the health variables retain all of the expected signs with varying levels of significance. The player level model also incorporates two performance based metrics that attempt to shed light on the fundamental differences between injured players. More specifically, a scoring-based variable and salary-derived figure will provide a performance based injury measure to represent the value of a players missed contribution to a team. These figures are based on the premise that some players are more important to a team than others, thus his injury is more detrimental. These weighted variables will have more pertinence at the team level, however they do show that a there is a strong negative relationship between an injured player's value and his on ice success.

At the team level, this study again is one of the first analyses to incorporate injury level data to better understand what really leads to team success. The team level fixed effects model shows that player health variables retain the negative signs that would be expected. Furthermore, it was shown that the number of injuries a team suffers has a statistically significant negative impact on a team's success. As previously mentioned, similar performance based variables will be incorporated, summed at the team level, to again show how some players are intrinsically more important to a team. Coefficients from the other independent variables showed that goal differential had the most significant effect of a team's production of league points, followed by goalie play and variables related to penalty minutes and power play opportunities.

By introducing the injury data to the production function models, and establishing a player injury format, this study accomplished the goal of analyzing the effects of injuries on individual and team performance. All the results offered compelling evidence that can help owners and coaches better manage their players and guide league rule changes for officials.

The remainder of this paper will be organized in the following format: In Section 2 and 3, I will provide a review of the literature relating to the production function from a traditional labor perspective, and then in the context of sports economics. Section 4, will then look at more advanced methodologies and comment on the evolution of the production function in pro-sports. Section 5 will focus on literature surrounding the NHL in particular and address the gaps I will seek to fill. Section 6 looks more closely at inactivity, absenteeism, and injury to set the groundwork for my analysis of injury data and establish the pertinence in the NHL. Section 7 introduces a clear summary of my contribution. Section 8 and 9 will comment on my methodology, data, and results. And finally, Section 10 provides some concluding remarks.

II. Basic Production Function Framework:

This section of the paper will first introduce the production functions from a basic labor economics perspective, and then shift to establish the pertinence in sports economics where the production frontier can be applied to professional sports teams. Equation 1 offers a basic production function for an output of Y with a given a combination, or vector, of inputs x, and a metric u to display inefficiencies:

$$Y = F(x) * u \tag{1}$$

This basic framework, explained by Zak et al. (1990), has applications across most industries, firms, and sectors. Theoretically, for a given number of inputs, x, the maximum output, $F(x)$ will differ from the output produced, Y, by a factor of u, which is restricted between 0 and 1 (Zak et al., 1990). Production frontier models are critical for firm owners and team executives alike to ensure that they are optimizing the available resources, and minimizing inefficiencies. Similar to the production of any firm or organization, a professional sports team relies on the interdependence of inputs (i.e. players, and units of talent) in order to produce a given level of output (wins).

In order to estimate a production function, inputs and outputs must be measurable. While this critical requirement is not limited to sports, professional leagues tally thorough performance statistics and extensive outcome metrics, which makes the task of analyzing inputs and outputs easier. When it comes to sports and professional leagues, there are several distinctions that must be made that will help conceptualize the fundamental differences that exist in sports leagues when compared to traditional firms and markets.

In professional sports leagues, individual teams produce a joint product with competing teams in the form of wins. Thus, a league's output is fixed by the league's schedule (Noll, 2003). From the input side of the function, team owners have limited labor pools to choose from and must also balance roster restrictions when considering labor and talent inputs (Noll, 2003). These restrictions, among others factors, make professional sports, and the corresponding production functions an incredibly fascinating topic for research.

III. The Evolution of the Production Function in Sports:

Now that the output function, production efficiency, and the intricacies of the sports market for wins have been properly established, we can now focus on the empirical work. Understanding the evolution of the production function will be important for my analysis because it will show that not only has the previous literature failed to produce a truly comprehensive production function for the NHL, but it has also omitted a critical aspect of professional sports: injuries. Nevertheless, Table 1 Summarizes the production function literature that will be examined more closely in the subsequent sections.

Table 1: Summary of Empirical Studies

	Sport/ League	Measure of Output	Input Factors Positively Affecting Production	Input Factors Negatively Affecting Production	Model Specifications
Rottenberg (1956)	MLB	Wins	Slugging percentage, batting average	N/A	N/A
Scully (1974)	MLB	Win Percentage	Slugging Percentage, Strike to Walk ratio	N/A	OLS Regression
Krautmann (1999)	MLB	Win Percentage	Slugging Average, number of at bats per year	N/A	OLS Regression
Depken (2000)	MLB	Win Percentage	Higher wage levels	Higher wage disparity	OLS Regression
Zak et al. (1990)	NBA	Ratio of final scores	shooting percentage, offensive and defensive rebounds, and steals	Personal fouls and turnovers	Cobb-Douglas/OLS Regression
Berri (1999)	NBA	Number of wins	Rebounds, avoidance of turnovers,	Personal fouls and	OLS Regression

			and shooting efficiency	turnovers	
Carmichael et al. (2001)	Prem. Lg Football (Soccer)	Ratio of points won to possible points	Defense (esp. goalie play), and ball handling	Poor passing	OLS Regression
Jones and Walsh (1988)	NHL	Salary	Points per game, penalty minutes, goals allowed	N/A	OLS Regression
Jones et. al (1999)	NHL	Salary	Game Experience, All-Stardom, size	N/A	OLS Regression
Kahane (2005)	NHL	Ratio of points won to possible points	Coaches win percentage, playing history	Organizational structure of team	Stochastic Frontier
Stewart et al. (1992)	NHL	Win Percentage	Moderate fighting and violence	N/A	OLS Regression

The production function in the context of professional sports was first introduced by Simon Rottenberg in the late 1950's. Rottenberg (1956) looked at Major League Baseball (MLB) and recognized that: "A baseball team, like any other firm, produces its product by combining factors of production." Rottenberg goes on to explore the baseball player's labor market more closely, namely the implications of the MLB's "Reserve Clause" which restricted the movements of players in the mid-late 1900s. This exploration into the compensation of laborers (players) in professional sports proved to be the stepping stone for additional studies in the years to come.

Following Rottenberg, Scully (1974), was primarily concerned with estimating the marginal revenue product (MRP) of a player, and the degrees to which the MLB Reserve Clause resulted in exploitation of players. Even though Scully was not focused on a team's production, this application of the production function still applies, and is necessary to show the evolution in this line of work. The econometric model utilized by Scully is shown below in Equation 2:

$$W_{it}=f(X_{it}) \tag{2}$$

In this model, Scully figured that a team's output, in this case winning percentage (W_{it}), for team *i,* in season *t,* is a function of the inputs of talent each player provides. Thus, Equation 2 theorizes that the combination of all the player's talents, X_{it}, is what produces wins in Major League Baseball. The model that Scully employed estimated X_{it} as a function of a player's slugging average and strike to walk ratio, two performance metrics that are clear indicators of talent in professional baseball. According to Scully, these performance metrics should be directly correlated with the wins a team produces, and more importantly the revenue a team earns from each win.

Krautmann (1999) took a second look at Scully's estimates, and criticized his method for determining the marginal revenue product of baseball players by analyzing team revenues. Krautmann considers slugging average and number of at bats per year (with an embedded dummy variable for shortstops and catchers who have primarily defensive roles on the team) as inputs. All of these factors are believed to positively affect a team's output, which in this case is measured by winning percentage.

Depken (2000) took the MLB production function a step further. He investigated wage disparity in the MLB and the effects it had on team productivity. Depken utilized a 1985-1998 panel data set and also measured output as a win percentage. His findings were in line with expectations: higher wage levels and lower wage disparity had significantly positive effects on team performance and productivity.

As the previous reviews have shown, the early literature surrounding production in professional team sports have some shortcomings. For one, the early production literature utilized basic performance metrics which fall short in capturing the complex nature of professional sports. Another critique is that these attempts were predicated on answering questions regarding monopolistic exploitation and other similar issues surrounding wage. Another striking shortcoming is that these studies were primarily focused on professional baseball instead of other popular North American sports. For many years, professional baseball was the prime subject for the econometric study of the production function for the following reasons: most baseball statistics have inherently low degrees of input interdependence (i.e. pitching and hitting), meaning that the talent input measures could easily be identified and further developed in order to isolate an individual's impact on production (Leeds and Von Allmen, 2011). However, as the next section will show, with the introduction of more complex models, and the development of more sophisticated game level statistics, economists were able to apply the production function to a more robust set of leagues, sports, and countries.

IV. Other Sports, More Advanced Models, and Alternative Methods:

Applying the production framework to sports outside of professional baseball started with an early attempt by Zak et al. (1990). Zak et al. looked at the relationships between many talent variables, and wins in the National Basketball Association (NBA). The authors took into account a variety of input talent measures that resulted in the production of a given output, which in the case of Zak et al. (1990), was measured by the ratio of the final scores. Inputs included: 1) shooting statistics, such as field goal and free throw percentages, as well as offensive rebounds, 2) defensive statistics, incorporated defensive rebounds, fouls, blocks and steals, and 3) miscellaneous variables such as a turnover ratio, and a dummy variable for home and away. One key model distinction that the authors made, unlike some of their predecessors, was that the input and output variables were measured in the form of a ratio against their opponent. This was done to achieve comparative measures, as opposed to absolute measures and also because ratios serve the purpose of demonstrating quality and the competitiveness of a given game or contest (Zak et al., 1990). Their study was able to capture the four most critical aspects of basketball: shooting, rebounding, ball handling, and defense. Their findings revealed that shooting ability was the most important factor in ensuring higher production of points, followed by offensive and defensive rebounds, as well as steals. Personal fouls and turnovers proved to reduce output from a player's perspective, while blocked shots and the variable for home-court advantage coefficient came out insignificant. The Zak et al. model will certainly assist in identifying how production can be categorized and measured. This paper will help in translating a similar framework to the National Hockey League. Another noteworthy aspect of the Zak et al. model is the previously mentioned utilization of ratios. The analyses I conduct will follow a comparable variable format to accomplish a similar goal. However, one of the major criticism of Zak et al. (1990), is the omission of injury data, and also the heavy value that was placed on shooting in the study. This overemphasis on shooting was also critiqued in a paper by Berri (1999) who questioned the functional form of the Zak et al. model. Berri's analysis attempted to answer how to measure an individual's impact when participating in a team sport; the methodology was similar with a few exceptions. Berri noted that under Zak et al., many of the inputs are dependent on the accumulation of other statistics. Take for example the value of shooting skill, which in the previous model is heavily dependent on ball possession (you cannot shoot without the ball). Thus, a team that rebounds poorly, or has limited number of steals will therefore be less productive under their model. But that is not always the case. From a player perspective, different skill sets are valued differently across teams and organizations, therefore the Zak et al. model makes "identical player" comparisons problematic according to Berri. The model that Berri employs addresses this shortcoming, and is grounded in the hypothesis that a player's statistical value is, and should be modeled, completely independent of whatever teammates he has on a respective team. To achieve this, Berri's model translated the player performance narrative offered by Zak et al. which is that shooting, rebounding, ball handling, and defense are the determinants of performance into the following hypothesis: Offensive production, that is a team's ability to score, is a function of how efficiently the team acquires the ball, how effectively the team handles the ball and values possession, and finally, the team's ability to convert additional possessions into points. Defensive production, is simply the inverse of these same factors (Berri, 1999). Berri's results showed that the factor with the largest marginal impact on the number of wins in the National Basketball Association was offensive rebounds, followed by ball handling variables (turnovers), and three point field goals. Berri also determined that a player's shooting percentage is more important than points scored, and that assists and personal fouls have relatively little importance. Berri left separation of a player's innate ability, experience, coaching, and team chemistry as avenues for further research. Some of the key takeaways from this Berri analysis is that the categories and formats of team and player performance variables are flexible. Furthermore, the Berri (1999) and Zak et al. (1990) analyses have shown that in professional basketball, performance statistics can be summarized into three basic categories: scoring, defending, and ball possession/avoidance of turnovers. This framework will have applications beyond basketball and will be translated to accommodate the innate differences in the National Hockey League. One shortcoming for these more contemporary analyses is still the omission of basic injury data to quantify how absence due to injury impacts player and team level performance. This omission will be remedied in my analysis by incorporating injury statistics and adding onto the previously mentioned three categories of performance with a "player health" category.

Thus far we have seen the production function evolve from the MLB to the NBA, as more advanced metrics and methodologies have become better suited to capture the intricacies of fast paced team sports with many interdependent inputs. Carmichael et al. (2001) decided to look at the production function in the context of English Premier League Football. Carmichael et al. took a more Scully-based approach with Premier League Football. The authors utilized match-play statistics for the 1997 to 1998 season for each of the 20 clubs. The authors decided to specify output in terms of league performance, which is determined by the total number of match points accumulated; that is the total number of points achieved during the season expressed as a percentage of the maximum number of points possible. In other words, the 20 clubs in the Premier League play in a double round robin style tournament (each club plays the others twice, once at the home team's stadium and once at their opponent's). League points are accumulated in the following way: teams receive 3 points for a victory, 1 point for a draw, and 0 points for a loss. In this context, and in order to achieve the highest output, a Premier League team must be both proficient on offense, and also in defending against opponents scoring attempts. Therefore, total league points will be positively affected by goals scored (GS), which are a function of shot attempts and ball possession, and the inverse will be true for goals conceded (GC) which will account for defensive weaknesses and the attacking performance by the opposition. Carmichael et al. (2001) concluded that defensive prowess is crucial to success, most notably goalie play. On the offensive side, ball handling, especially passing efficiency, is most important in achieving a higher shooting percentage which is significantly more important to goal generation than any other metric. In general, the authors decided that in football: "What matters is its quality in all its manifestations, which is itself related to players' skills and their team-working relationships, constrained by those of their opponents" (Carmichael et al., 2001). An important takeaway from Carmichael et al. is the relative agreement in the ways in which performance variables are categorized across sports. This cross-sport consensus is encouraging especially since the low scoring nature of soccer more closely resembles the scoring patterns in the National Hockey League. Therefore, the Carmichael et al. frameworks will certainly have some strong influence on the model that will be employed in my empirical analysis. Understanding the narrative of competition in the NBA and the Premier League will help me formulate a similar format for competition in the NHL. This hockey adaptation will retain many of the same competition variables that have been shown to be fundamental aspects of professional sports.

As the previous section has shown, the evolution of the production function involves many sports and methodologies that build off each other to create a more accurate way to represent the ways in which teams and players produce various forms of output. However, one common theme throughout has been the lack of analysis involving the National Hockey League, and also the omission of injury data. The next section will begin to introduce the literature that looks at the National Hockey League more specifically.

V. Production Function in the National Hockey League:

Before we address the existing literature concerning the NHL and production, a brief explanation as to why professional hockey is worth studying in the context of my research question is necessary. The National Hockey League is a prime subject for the analysis of injury on individual and team production for many reasons. Professional hockey is one of the fastest and most dangerous team sports played around the world. Aggressive contact between players, boards, and the puck occur frequently as players skate at speeds of 30 mph or faster (Ornon et al. 2011). Additionally, intense ice time exposure and varying playing styles adds different degrees and mechanisms to injury. Furthermore, the availability of hockey injury data and statistics are also reasons why I choose to study the NHL. When merged, the two groups of data form the basis of an entirely comprehensive data set that will be used to explain how production is limited when players are injured and by doing so, we can begin to understand the multitude of factors that affect performance.

This section of the literature review will now shift towards the National Hockey League where the empirical analysis and applications of the production function is relatively sparse compared to the other North American sports leagues. The majority of the analysis in the NHL has been centered mostly upon discriminatory hiring and compensation. Nevertheless, the papers here will shed light on the variables that will be the primary indicators and contributors to a team's success.

Jones and Walsh (1988) explored instances of discriminatory hiring and compensation for the 1977-1978 NHL season. The aim of their study was to link skill, salary, and discrimination for English players and players of French-Canadian descent. They found that in most cases, skills, as measured by points per game for offensive players, penalty minutes for defenders, and career goals allowed for goalies, were the most significant determinants for player salary. Jones et al. (1999) also assessed wage discrimination in the NHL by looking at player data from the 1989-1990 season. Their model accounted for various player characteristics such as veteran status, All-Star selections, a "goon" attribute for those who are referred to as "enforcers", and a "trophy" variable to introduce the number of honors a player has received into the model. The results of their study suggested points per game, game experience, All-Stardom, and player size all had positive effects on production, and thus salaries.

Kahane (2005) also investigated the production process in the National Hockey League to reveal discriminatory hiring practices. Kahane's analysis measured output as proportion of total league points won versus total possible points. The author applied a stochastic frontier model with additional error terms built in in order to account for certain firm-level factors. Inputs for production according to Kahane came in two forms: organizational inputs, and individual inputs. Organizational specific inputs included two factors: 1) the role of upper management, whose primary job is to hire the right coach and provide training facilities (among other things), and 2) coaching ability, where the coaching staff should be primarily concerned with overseeing training, and optimally combining players in order to ensure efficient on-ice production. Inputs for individual performances differed from some of the previously mentioned studies under the Kahane model. Rather than using traditional averaged career measures and statistics as metrics for inputs, Kahane decided to use a measure that would be more representative of certain intangibles that players bring to an organization. These intangibles include leadership and mentoring abilities that aren't necessarily represented in career statistics, but will certainly have a positive effect on team production (Kahane, 2005). Therefore, Kahane decided to use an alternative approach and utilize team payroll as a measure of input. The reason for using the payroll method is as follows: first, it addressed the problem of omitting inputs that are hard to measure, but that management will likely recognize value for and compensate accordingly. The payroll method also helped lessen the effect of teammate complementarities that are captured in team-average performance measures. The author's results showed that coaching quality, as measured by career win percentage and playing history, had significantly positive coefficients that supported the hypothesis that talented coaches who have previously played for their current organization are more adept at creating greater output for given inputs. Conversely, certain organizational structures, namely franchise age and syndicate ownership had slightly significant negative effects on league points captured. Kahane's analysis provides some interesting variation to the production function. For one, the use of the proportion of total league points won versus total possible points as the output measure is an attractive option for my analysis. Also known as point share, this methodology simply divides the number of points won by a team (2 for a win, 1 for a tie, 0 for loss), by the total possible number of points (which equals 164 in an 82 game season). This point share measure sheds light on league competitiveness and contributes a more accurate figure to show a team's overall league success. However, Kahane's study, and his choice to use salary instead of standard Scully-style input variables, does not align with the aim of this study which will be looking directly at on ice performance and the effect of injuries on these variables.

Jones and Walsh (1988), Jones et al. (1999) and Kahane (2005) have accomplished relatively the same task: They offered an analysis of the degrees of wage discrimination in the NHL against French Canadians. However, their studies provide some insight into the types of performance variables that can be employed in NHL level regressions. Jones and Walsh, and Jones et al., contributed some interesting variables that are unique to the hockey and the NHL. As previously mentioned, these variables include a "Trophy" figure which signifies the number of times a player earned all-star honors or won individual game level awards. This is an interesting variable to consider given its ability to measure some intangibles that certain players have over others that are not directly related to scoring, including leadership, toughness, and morale. Kahane (2005) also puts forth different form of output that can be used in my analysis to contribute to a more comprehensive production function. Nevertheless, Jones and Walsh (1988), Jones et al. (1999) both failed to directly relate how these player statistics contribute to a team's production of wins. Furthermore, as previously mentioned, Kahane's use of salary as an input determinant does not align with the aims of this study. However, for the purpose of my study, Kahane's use of league point share as output, in tandem with Jones and Walsh, and Jones et al. and their input measures with will certainly contribute to the methodology of my study. By incorporating the categories of variables and performance that were introduced in the Berri, Zak et al., and Carmichael analyses, I will be able to construct a production function that will capture the true nature of professional hockey and league point production while also addressing the effects of injuries.

VI. Productivity, Inactivity, and Injuries:

The previous sections have covered how the production function has evolved into sport economics and the numerous variations and applications for the models. However, as previously mentioned, one critical lapse in the literature is the absence of injury data to report how production is missed or limited when players get hurt. Man Games Lost (MGLs), that is the number of games a player does not participate in when he is injured, has significant effects on team success across all sports. When players miss games, the effect on a team can be three-fold. For one, injuries to players disrupt team cohesion. Gammage et al. (2001) discussed cohesion in team sports. There, the authors discovered a significantly positive cohesion-performance relationship across all sports. Perceptions and norms of productivity intuitively decrease in the absence of players and teammates. The second consideration surrounding injury and absence stems from Jones et al. (1999) who noted, “All-Stardom” is a critical predictor a team’s on ice success. That being said, the absence of a player who is a significant contributor (an All-Star) creates more disruption to a team and lessens the likelihood of winning. Finally, players who are away from the game for extended periods of time due to injury will certainly experience decreased performance and production upon return or over the course of a season. We can draw upon Ge and Lopez (2016) to shed light onto the implications of a player's absence from playing in the NHL. In the case of the aforementioned paper, absence from the game comes in the form of a temporary lockout, not injury. Nevertheless, Ge and Lopez (2016) investigated the production of players who decided to stay in North America during the 2012-2013 NHL lockouts compared to some players who opted to continue to play in the very competitive European hockey leagues. Their results showed slightly higher goal-scoring rates for the latter group of players who continued to play in Europe. Additional studies have looked at work stoppage and worker productivity. For example, Herrmann and Rockoff (2012), analyzed the impact of absenteeism on productivity in the field of education. Their analysis investigated how teacher absenteeism impacted achievement of their students. Their paper found statistically significant negative impacts on student achievement across certain subjects in some circumstances. Mas (2008) looked at product quality following periods of labor strife. Their results also indicated negative relationships between labor absenteeism and product quality.

These studies are related to the work that I am attempting to accomplish because they introduce the possibility of decreased production when workers are inactive or absent, and the corresponding negative impacts both within sports and in other industries and sectors. These tangential studies have detailed the effects of absenteeism and inactivity on productivity, production, and performance, however, as this next section will show, sports injuries are fundamentally different. This distinction is important because the previous studies have failed to recognize the possibility of decreased production at the firm level when certain inputs are inactive. This can come in the form of missing the team's best player in critical moments during the season or when other players become injured while their replacements fail to make up for that player's regular production. Teams who experience higher injury rates also sometimes rely on other players to "pick up the slack" as others rehab to return to play. Coaches and managers are then faced with decisions surrounding which players to bring up from lower level "farm teams" which are used to train and prepare players for the NHL. But before these distinct sports related scenarios can be discussed further, we must first introduce medical literature that will begin to explain some of the intricacies and contributing factors related directly to athlete injury.

Saragiotto et al. (2014) offers some insights. Their analysis aimed to investigate the opinions of medical professionals familiar with elite athletes to determine the main factors that contribute to athlete injury occurrence across various sports. Based upon several structured interviews with physicians and physical therapists, the authors were able to determine that "movements inherent to the sport" (i.e. jumping in basketball, hitting in football, or heading the ball in soccer) are the most significant and common cause of injury according to the medical experts. Additionally, the training and playing load imposed on a player was also an indicator that overuse was a strong contributor to injury. The study also found that there was some degree to which behavioral and cognitive features influenced injury. The authors explained how this behavioral aspect connects the ways in which a player acts, thinks, and reacts and the rate at which they might get hurt. This discussion introduces the idea that playing style may explain the likelihood of injury, citing that athletes with aggressive or competitive personalities might be more prone to injuries.

In a study conducted with 400 athletes from different sports, Watson (1993) found that age had an important role in the severity and occurrence of injuries across sports. Watson also reported that unlike endurance sports, fast paced and high intensity sports will lead to increased injury rates. This is in line with Saragiotto et al. (2014) who also mentions unique game characteristics as reason for injury.

Parkkari et al. (2001) offers some of the more applicable research with regards to my analysis. The authors here categorized risk factors associated with injuries in the following way. For one, they recognized that for athletes, the injury rate increases with the frequency of violent contacts within the sport. Additionally, the evidence has shown that competition hours (as opposed to training hours) have higher injury rates per hour of activity. And finally, as in line with Saragiotto et al. (2014) and Watson (1993), type, frequency, intensity and duration of training and competition, play a large role in the nature, severity, and occurrence of injuries. Parkkari et al. (2001) also cites excessive height, weight, muscle weakness, and other biomechanical elements as injury influencing factors. The final factor that influences injury according to these authors are psychological factors which include motivation and risk taking. These factors closely resemble those mentioned by Saragiotto et al. (2014) who refers to them as behavioral and cognitive features of athletes.

In summary, the medical and exercise science literature can be summarized in the following ways. Most importantly, the factors that are most likely to contribute to sports injury probability and severity are 1) competition hours, 2) unique game characteristics, 3) age, and 4) behavior/playing style. These categories of injury factors will help make sense of the ways in which players in the National Hockey League experience injury.

VII. My Contribution:

My study will employ the production frontier, with the addition of injury data, to model a comprehensive function to truly understand the value of player contribution that is missed during injury. I will first analyze the factors that contribute to injury and then apply a basic production function framework to the individual players, and then to the separate teams within the NHL. This new framework will have applications beyond professional hockey and will help owners and rule makers when making personal and regulation changes.

Based upon the combination of fundamental labor economic analysis and the introduction of the exercise science literature, I can expect to find three things in my analysis. For one, by using basic statistical modeling I anticipate to find a positive relationship between a player's injury susceptibility and his age, playing style, competition exposure, and also incidences of contact within the game of hockey. Secondly, by applying the production function to each individual player, I can hope to calculate a measure for each player's missed production when he gets injured. It will also be important to show how a certain player's ability and value to the team is reflected in this measure. I expect to find a negative relationship between the duration of a player's injury and his output on the ice. Furthermore, the aforementioned performance based measure will likely show a negative relationship exists between the player's value and his production. In other words, this figure will show how an injured player's missed production is magnified if he is a critical member of the team. Finally, at a more firm and team level, we can combine the previously mentioned expectations to see how injury occurrence and severity, as well as identifying which inputs are limited due to injury, and the negative effect they have on a team's success. I expect to see a negative relationship between team success and the number of injuries and also the number of games lost due to injury. Through the development of these methodologies, this essay will offer a comprehensive production function format that incorporates injuries, which will mirror the existing production function literature, the studies regarding inactivity, and also integrate the critical aspect of sport injury related works.

VIII. Data and Methodology:

The following sections will begin to detail the data, models, and results that will help further explore and understand what factors lead to on ice success. To begin, there are three distinct data sets that will be utilized in this essay: 1) Individual player level injury data, 2) Player level performance statistics and 3) Team level metrics. The three datasets include full performance and injury data from the 2013/2014 season though the most recent 2016/2017 NHL season. The Player Injury (INJ) dataset and the individual Player Level (PL) dataset have two unique sources. The majority of the statistics for players were collected from

HockeyReference.com. The second, and slightly more critical source to this paper, is HocketAbstract.com, where NHL experts and fans have compiled very thorough and complete player performance statistics, including in-depth injury information. By merging the two sources at the player and season level, the INJ and PL dataset yield 2,954 distinct observations of National Hockey League players from the four seasons. At the Team Level (TL), the data set is composed of quantitative performance statistics for all 30 NHL teams from the 2013/2014 season through the 2016/2017 season yielding data set of 120 observations. The majority of the data at the team level was also compiled from HockeyReference.com, while injury figures were reported as the aggregate of the PL data points, sorting by team and year.

Player Injury Susceptibility:

The player injury dataset is composed of various different measures that will help set the groundwork for this paper. It will begin to examine which factors contribute to injury based upon the aforementioned medical and exercise science literature. The variables that will be used in the player injury regression are detailed below in Table 2.

Table 2: Player Injury (INJ) Variable Definitions

VARIABLE NAME	DESCRIPTION
INJ	Injury Dummy Variable for whether the player was injured or not throughout the season
HITS	Total Hits. Number of hits delivered and received by a player
BLKS	The number of shots that were blocked by a player
TOI/GP	Average time on ice per game
STYLE	Playing Style. Function of Penalty minutes, hits, majors, blocked shots and fights.
AGE	The age of the player
POS	Position. Vector of positions (Winger, Center, Defenseman)

Here, and in accordance with the exercise science literature, the player injury variables can be divided into four categories that explain how each contributes to injury susceptibility. The first category is *contact factors* and will represent the "unique game characteristics" that were mentioned in the exercise science literature. Contact factors include hits that a player delivers and receives (HITS), as well as the shots that are blocked by the player (BLKS). The second factor that contributes to injury is *exposure*. Exposure factors are most clearly represented by a player's time on ice (TOI_GP) measured in minutes per game. Choosing to use ice time per game instead of the total number of minutes will make analysis and interpretation easier. Playing style was another significant indicator of injury according to the medical literature. Therefore, a *playing style* variable will help account for the risk taking and motivational qualities of players that may lead to an increased rate of injury. The playing style variable (STYLE) will be calculated by adding the factors that best portray a player's motivational and behavioral qualities. These variables include the player's penalty minutes, the hits the player delivers, the number of major penalties committed by the player (5 minute major penalties are the result of very dangerous and overly aggressive plays-- akin to the red card in soccer), the number of blocked shots tallied by the player, and the number of fights he engaged in. This summation will give us an index of playing style and aggressiveness. The final injury contributing factor is *aging.* The aging factor is simply a function of the player's age (AGE)[1]. The player's position (POS) will also be used as a control variable. The variable POS is a vector of three different positions in hockey (excluding goalie): Winger, Center, and Defenseman. Summary Statistics for the Player Injury model are available in Table 3 below.

---

[1] Originally incorporated AGE^2 but the results were not significant, indicating no non-linear trend in the likelihood of injury with respect of age.

Table 3: Player Injury (INJ) Summary Statistics

VARIABLES	(1) Mean	(2) S.D	(3) Min	(4) Max	Expected Effect on Injury
Injured	0.631	0.483	0	1	N/A
Hits	161.4	87.30	2	543	(+)
Blocked Shots	47.02	41.69	0	283	(+)
Time on Ice (per game)	15.80	4.191	4.006	29.40	(+)
Playing Style	163.8	99.35	3	622	(+)
Age	26.76	4.506	18	44	(+)

In the NHL, between 2013/2014 and 2016/2017, there were 1,864 injuries to players, resulting in 23,296 MGLs for teams. The 2013/2014 season had the most number of injuries, with 473 and 6,130 MGLs. In 2014/2015, the NHL experienced the most number of MGLs at 6,137 but only 466 injuries, tied for second fewest with the 2015/2016 season. The 2015/2016 season also had the fewest number of man games lost (5,476). The 2016/2017 season had the lowest number of injuries (459) and only 5,553 MGLs. As shown in Table 3, over 60% (63.1%) of NHL players were injured at least once at some point during the four seasons. As mentioned earlier, it is this high injury rate associated with the NHL that makes this league a perfect organization to study in an attempt to understand how injuries affect performance. It was also found that the average number of hits received and delivered by a player is 161.4, plus the roughly 47 shots a player blocks, equals over 200 contact factors in a season for a player. These contact factors were incurred during an average ice time of roughly 15 minutes per game with the average age of a NHL player being approximately 27 years old. Finally, playing style, the index of aggressiveness ranges from 3 to 622 with an average playing style rating of 163.8.

A basic probit regression model will be employed here to analyze how the previously mentioned contact, exposure, style, and aging factors contribute to the injury susceptibility of player *i* in each of the four seasons. Equation 3 shows the methodology of this preliminary regression that uses robust standard errors to correct for heteroscedasticity.

$$INJ_i = \Phi(\beta_1 HITS_i + \beta_2 BLKS_i + \beta_3 TOI\_GP_i + \beta_4 STYLE_i + \beta_5 AGE_i + \beta_6 POS_i + \in_i) \qquad \textbf{(3)}$$

The model described in Equation 3 will be ran for each season individually. This method will help isolate differences across seasons and treat each year and player as an isolated observation. Another adjustment that will be made in this model is the exclusion of roughly 600 NHL players across the 4 seasons who played in less than 10 games. By removing these players, the model will more accurately identify the injury factors that influence the players who have significant contributions to the team. The players who play in fewer than 10 games also lack the full exposure to the regular stress of competition. The players who compete consistently throughout the season are exposed to the injury factors more regularly than those who rarely get on the ice. Players who play in less than 10 games accumulate roughly 40 minutes of ice time and less than half a point (.4 points) in production in each of the four seasons. For these reasons, players who competed in less than 10 games will be removed.

The main limitation in my player injury model lies within the playing style variable (STYLE). While the medical literature has certainly emphasized the impact of behavior on injury, this factor is hard to quantify. There may be slightly more effective ways at measuring the behavioral qualities of a player, however I believe that penalty minutes, hits delivered, major penalties, blocked shots, and fights are the best indicators for a player's aggressiveness (or lack thereof). I hypothesize that exposure time and contact factors, namely hits, will have the most significant effects on a player's health. I anticipate playing style to have some mixed results given the construction of the variable. Age will also likely have some significance, but that is to be expected given the nature of professional sports which favors the young athletes.

Player Level Production:

The variables that have been selected for the player level analysis are slightly different from the injury susceptibility dataset. Here, individual player statistics will be analyzed to show which factors contribute to, and take away from, a player's on ice production. The variables that will be used in the PL analysis are detailed in Table 4 below.

Table 4: Player Level (PL) Variable Definitions

VARIABLE NAME	DESCRIPTION
PLSMNS	Plus/Minus ratio
SH_PCT	Shooting Percentage
FO_PCT	Face-off percentage
TKA	Takeaways. The number of times the player steals the puck from the opposition
GVA	Giveaways. The number of times a player turns the puck over to the other team
HITSF	Hits For. The number of hits a player delivers
MGL	Number of games missed due to injury
MISSPROD	Missed Production. An estimate of the missed production a player does not contribute due to injury
CHIP	Cap Hit weighted by the ratio of games missed to total games (82)
AGE	The age of the player
AGE2	The age of the player, squared
POS	Position. Vector of positions (Winger, Center, Defenseman)

At the PL, and in accordance with Zak et al. (1990), Berri (1999) and Carmichael et al. (2001), this model stipulates that a professional athlete's contribution to a team depends on three factors: 1) scoring ability, 2) defensive skill, 3) ability to obtain and retain possession. In line with my contribution, this study will include a fourth category to this framework, *player health*. This injury related variable is described by Man Games Lost (MGL), the number of games lost due to injury. The model will use the Plus/Minus ratio (PLSMNS) as the measure of output to provide an all-inclusive metric for on ice contribution. The plus/minus ratio can be summarized in the following way: A player is awarded a "plus 1" every time he is on the ice when his team scores a goal (either at even-strength or shorthanded). Similarly, the player receives a "minus 1" every time he is on the ice when the opposing team scores. This measure provides a good indicator of the player's offensive production, but also his defensive contribution as well. The independent variables can be divided into the following categories: Scoring (SH_PCT), defense (HITSF and TKA), puck possession (GVA and FO_PCT), and player health (MGL). Similar to Krautmann (1999), the player's position (POS) will also be controlled for to account for the different purposes and roles of players on the ice. The player's age will be controlled for with Age and Age^2. The final addition to this player level model addresses the idea that some players are more valuable to an organization than others, thus injuries have intrinsic value based upon the player's contribution to the team. This essay will propose two methods of weighting the length of a player's injury to portray these differences in value. The first method relies on the following assumption: a player has an average per game production of points, also known as points per game or PPG. This variable is simply a players total points tallied (Goal + Assists) divided by the number of games the player participated in. If a player gets injured, we can then multiply his PPG by his MGLs to predict the amount of point production a player is not producing. For example let's look at Connor McDavid, the star center for the Edmonton Oilers. McDavid was the 1st overall pick in the 2015 NHL Draft and has played in over 200 games scoring 256 points over his three year career. In his first year in the NHL (2015/2016), the young star broke his collarbone causing Edmonton's best player to miss 37 games. At the time of his injury, McDavid had scored 48 points in only 45 games equaling an impressive 1.07 PPG. By multiplying McDavid's MGLs by his PPG, we can estimate that the Oilers missed out on roughly

40 points during McDavid's absence. Thus, McDavid's missed production (MISSPROD) was -39.47 in 2015/2016. Introducing this form of performance interpretation allows us to investigate the magnitude to which an injury to one player is different from an injury to another, less productive teammate. However there is a limitation to this method. By simply stating that a player's missed production to a team is a function of his PPG and MGLs is slight over simplification of league dynamics. Some players tend to have much larger contributions to a team outside of scoring. To address this potential inaccuracy, and to provide a more inclusive weighting method for injury, this analysis will also include a salary weighted variable, also known as CHIP[2]. The underlying unit of analysis in the CHIP measure is a player's salary and corresponding "cap hit". A player's cap hit is the total compensation a player will receive in salary over the life of a given contract, divided by the number of years it is effective. With this figure, NHLInjuryViz.com mentions how the best way to incorporate injury is to simply multiply the cap hit by the ratio of games missed to games in a season. In other words, CHIP is calculated in the following way:

$$\text{CHIP} = \text{Cap Hit} \times (\text{MGL}/82) \quad \textbf{(4)}$$

Utilizing objective salary weighted figures, CHIP will provide us with a unit of measure that captures the intrinsic value of a player's injury and absence. This figure accounts for player attributes and skills outside of scoring ability (i.e. defense, leadership, hustle, and grit). However, this method also carries some limitations. The main limitation being that salaries and the corresponding cap hit can sometimes be blurred by a team's ability to pay their players. In addition, this measure can potentially be skewed by large fee-agency deals where players are over- or sometimes under-compensated. While neither of these performance based metrics are

[2] This method was created and originally presented by NHLInjuryViz.com (http://nhlinjuryviz.blogspot.com/p/index-page.html)

perfect, MISSPROD and CHIP provide a performance based injury measure to represent the differences in value when players get hurt.

Summary statistics for the variables included in this model are summarized in Table 5 below. As shown, players on average missed 12.50 games in each of the 4 seasons. From a scoring perspective, players tallied an average of 20 points throughout the league and across the four seasons while shooting at roughly 7% efficiency. Faceoff percentages across the league were around 44%. Takeaways and giveaways in the NHL were 26 and 21 respectively. Another defensive measure, hits delivered, averaged out to 80 hits in a season. The first value based measure, missed production, revealed that on average, a team misses out on 2.75 points when a player get injured. The CHIP metric, which attempts to put a monetary value on the missed ice time, showed that when players get hurt, organizations are forfeiting $415,638 of player investment in the form of salary.

Table 5: Player Level (PL) Summary Statistics

VARIABLES	(1) Mean	(2) S.D	(3) Min	(4) Max	Expected Effect on Plus/Minus
Plus/Minus	-0.327	10.58	-39	39	N/A
Shooting %	0.0697	0.0466	0	0.333	(+)
Face-off %	0.435	0.214	0	1	(+)
Takeaways	21.81	16.11	0	128	(+)
Giveaways	26.65	19.51	0	153	(-)
Hits Delivered	80.12	56.75	0	382	(+)
Man Games Lost	12.50	12.58	1	71	(-)
Missed Production	2.755	5.079	0	76.47	(-)
CHIP	415,638	592,611	0	6.738e+06	(-)
Age	26.76	4.506	18	44	(+)

The player level regression will be applied to each of the four seasons of data independently, this will show how these results vary across multiple years and if there are any patterns. The model will be calculated for each player *i* in all four seasons. Again, given the aim of this portion of the study is to examine how injuries affect individual performance, players who have competed in fewer than 10 games will also be removed. This will isolate the players who are routinely exposed to the traumas and stress of game level competition. Equation 5 shows the PL regression model that will utilize the four performance based categories, and also the weighted variables to account for differences in player value. Equation 5 also uses robust standard errors to correct for heteroscedasticity.

$$\begin{aligned} PLSMNS_i &= \beta_0 + \beta_1 SH\_PCT_i + \beta_2 FO\_PCT_i + \beta_3 TKA_i + \beta_4 GVA_i + \beta_5 HITSF_i \\ &+ \beta_6 MGL_i + \beta_7 MISSPROD_i + \beta_8 CHIP_i + \beta_9 AGE_i + \beta_{10} AGE2_i + \beta_{11} POS_i + \in_i \end{aligned} \tag{5}$$

There are some potential limitations in my player level model. For one, the plus/minus ratio (PLSMNS) might put a slight over emphasis on shooting and scoring. What's more is that the plus/minus ratio includes power play and shorthanded goals scored and conceded. These inclusions may affect those players who are skilled penalty killers negatively and those who are power play specialists positively. Nevertheless, given the nature of hockey, the plus/minus ratio best encapsulates each player's contribution while he is on the ice. The exclusion of the 600 players who only played in 9 games or fewer does in fact decreases the sample size, but as previously mentioned, this is done to ensure a more accurate fit for the model.

I hypothesize that with regard to the player health variables, MGLs will have a significantly negative effect on the player's Plus/Minus ratio based off the findings in the labor economic literature. I anticipate missed production and CHIP to retain their expected negative signs and offer insights into the magnitude of which player value is related to on ice performance. I also expect to see strong significance for the other independent variables, namely shooting percentage, takeaways, and giveaways.

Team Level Performance:

The TL analysis will be more straight forward and offer insights into the quantifiable contributions of certain game factors, including player injury, on team performance. The variables that will be included in the three team level production regressions are presented in Table 6 below.

Table 6: Team Level (TL) Variable Definitions

VARIABLE NAME	DESCRIPTION
PTS	Team Points. Total team points accumulated by a team (2 points for a win, 1 point for a tie)
PTS_PCT	Team Point Share. Points earned out of possible points.
WIN_PCT	Win Percentage. Number of wins divided by total games played
GOAL_DIFF	The difference between the goals scored by a team less those produced by the opposition
SOG_DIFF	The difference between the shots on goal accumulated by a team less those produced by the opposition
FOW_DIFF	Face off win Differential. Difference between number of face off wins and losses
PIM_DIFF	Penalty Minute Differential. Difference between penalty minutes accumulated less those produced by the opposition.
PPG_DIFF	Power Play (PP) Goal Differential. Difference between PP goals scored and conceded
SV_PCT	Save Percentage. Average save percentage of a team's goaltender(s)
HITS	Total number of hits delivered by a team
SH_PCT	Average shooting percentage of the team
INJ	Number of Injuries. Total number of injured players in a season
MGL	Man Game Lost. Aggregate number of games missed by all players
MISSPROD	Missed Production. Total number of missed production from all players
CHIP	Cap Hit weighted measure, summed for each team (measured in millions of dollars)

The output measures chosen for this team level analysis include league points, league point share, and win percentage. The reason for choosing three output measures is as follows: Traditionally, wins and win percentages have been common measures of a professional sports team's output and production. These output measures are commonly utilized in the NBA and MLB where the win percentages determine league standing and playoff contention. However, in the National Hockey League, rankings are determined by Team League Points (PTS), where a given team is awarded two points for a win, one point for a tie, and no points for a loss. This unique distinction to the National Hockey League justifies using Team League Points as the main output measure. Furthermore, Kahane's (2005) utilization of League Point Share (PTS_PCT) provides an alternative measure to consider, and it will also be used in my analysis as a dependent variable. Finally, this study will also include win percentage as an output measure to complete the comprehensive analysis and investigation this study seeks to provide.

Similar to the Player Level analysis, the independent variables at the Team Level can be separated into the following categories: Scoring, Goaltending, Puck possession, Defending, and Player health. Player health variables, INJ and MGL, are simply the aggregate of the player level data set. One distinction that needs to be made is with regard to the format of the other independent variables, (i.e. percentages, ratios, or value), this essay will utilize a variety of variable formats that will based upon that specific performance metric. Shot on goal differential (SOG_DIFF), power play goal differential (PPG_DIFF) and goal differential (GOAL_DIFF) will be used to report on the team's scoring ability relative to another team, a similar technique was used by Zak et al. (1990). The reason for choosing differentials is that the method provides comparative measures, as opposed to absolute measures, and also because differentials demonstrate quality. By investigating differentials, the measures will shed light on the competitiveness of a given hockey game. Face-off win differential (FOW_DIFF), and penalty minute differential (PIM_DIFF) will be used to assess puck possession capability and introduce extra man opportunities into the model. Save percentage (SV_PCT) and hits (HITS) will be left in their original format due to availability of data. This model will also incorporate the two value based measures for injury, MISSPROD and CHIP. These variables are simply the aggregate of the individual player level data set with the only exception being that CHIP will be measured in millions of dollars. This Team Level regression will include a time dependent season variable (SEASON_NUM) and a team specific fixed effects (TEAM_CODE) that will help control for autocorrelation across seasons and teams. Equation 6 shows the TL regression model that will explain the production of league points of team *j* in season *t*. A similar regression will be run using the other previously mentioned dependent variables: league point share and win percentage. Equation 6 will employ robust standard errors to correct for heteroscedasticity.

$$PTS_{jt} = \beta_0 + \beta_1 GOAL\_DIFF_{jt} + \beta_2 SOG\_DIFF_{jt} + \beta_3 FOW\_DIFF_{jt} + \beta_4 PIM\_DIFF_{jt} + \beta_5 PPG\_DIFF_{jt} + \beta_6 SV\_PCT_{jt} + \beta_7 HITS_{jt} + \beta_8 INJ_{jt} + \beta_9 MGL_{jt} + \beta_{10} MISSPROD_{jt} + \beta_{11} CHIP\_ADJ_{jt} + \delta_j + \mathrm{T}_t + \in_{jt} \quad \textbf{(6)}$$

The variables included to calculate the team level fixed effects model are summarized in Table 6 below. As shown, the average team experienced 24.12 injuries in each of the four seasons. It was also determined that teams saw players miss an average of over 230 games per season. Teams accumulated an average of 92 league points, a 56% share of total points, and an average win percentage of .500. Looking more closely at the value based measures, teams forfeited over $8 million in salary expenditures when their players were hurt. Additionally, these teams left 93.04 points unproduced while players were recovering from their injuries.

Table 6: Team Level (TL) Summary Statistics

VARIABLES	(1) Mean	(2) S.D	(3) Min	(4) Max	Expected Effect on Output
League Points	91.81	14.65	48	120	N/A
League Points (Share)	0.560	0.0894	0.293	0.732	N/A
Win %	0.500	0.0930	0.256	0.683	N/A
Goal Differential	0	38.87	-113	84	(+)
Shots on Goal Differential	0	262.9	-981	564	(+)
Face-off win Differential	0	199.8	-524	455	(+)
Penalty Minute Differential	0	69.76	-182	134	(-/+)
Power Play Goal Differential	0	17.92	-65	21	(+)
Save %	0.895	0.0145	0.860	0.927	(+)
Hits	1,813	351.9	1,034	2,672	(+)
Injuries	24.12	5.827	11	37	(-)

Man Games Lost	236.30	83.20	51	464	(-)
Missed Production	93.04	35.77	28.73	189.80	(-)
CHIP (in millions)	8.375	3.139	1.901	17.88	(-)

At the team level, there may be some slight limitations. For one, even though the variable categories are certain (i.e. Scoring, Goaltending, Puck Possession, Defending, and Injuries), choosing the hockey metrics to represent them is a difficult task. To combat this, I have chosen to use a variety of different variable formats (i.e. differentials and percentages) to represent the variable in the most appropriate way possible. Another limitation of the team level model is the formulation of the MISSPROD and CHIP variables. As previously mentioned, there are certainly some shortcomings to these methods. However, they incorporate a performance based injury measure to represent the differences in value when players get hurt. Finally, the exclusion of the 600 players who only played in 9 games or fewer decreases the sample size, yet again, this is done to ensure a more accurate fit for the model.

I hypothesize that the number of injures that a team suffers will have a strong negative effect on team success across all three output measures. I anticipate a negative, yet less significant, relationship to exist with MGLs and also the two performance based variables. I believe that CHIP will provide a lower and more conservative reflection of the value of foregone contribution, while missed production will likely be slightly higher.

IX. Results

The marginal effects of the Player Injury susceptibility probit regression are reported in Table 7 below. The probability model revealed that almost all variables retained their expected positive relationship to injury with the exception of blocked shots (BLKS).

Table 7: Injury Susceptibility Regression Results

VARIABLES	(1) 2013/2014	(2) 2014/2015	(3) 2015/2016	(4) 2016/2017
Hits	5.91e-05	0.000847	0.00116**	0.000219
	(0.000530)	(0.000600)	(0.000590)	(0.000566)
Blocked Shots	-0.00170*	-0.00217**	-0.00295***	-0.000201
	(0.000895)	(0.000934)	(0.000926)	(0.000936)
Time on Ice (per game)	0.0185***	0.00874*	0.0183***	0.00364
	(0.00599)	(0.00667)	(0.00654)	(0.00654)
Playing Style	0.000686	0.000897	0.000435	0.000185
	(0.000518)	(0.000638)	(0.000615)	(0.000572)
Age	0.0199***	0.0180***	0.0103**	0.00482*
	(0.00421)	(0.00415)	(0.00430)	(0.00425)
Control for Position	Yes	Yes	Yes	Yes
Observations	752	730	736	736

Dependent variables here for all specifications are probability of injury.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As reported above, the most significant contributor to a player's likelihood of being injured is the aging factor. In the 2013/2014 and 2014/2015 season, aging was a statistically significant contributor to injury at the 1% level. In 2015/2016 and 2016/2017, age was significant at the 5% and 10% level respectively. It was shown that in the four seasons, every additional year in age resulted in roughly a 1.3% increase in the probability of injury. This increased risk factor was highest in 2013/2014 where a one year increase in age resulted in a 1.9% increased probability of injury, and lowest in 2016/2017 which had only a 0.5% increase in injury susceptibility. The TOI/GP exposure factor also had significantly positive effects on the player's injury susceptibility across three of the four seasons. Additional exposure factors, that is a 1 minute increase in number of ice minutes per game, contributed to a roughly 2% increase in injury probability in the 2013/2014 and 2015/2016 season with significance at the 1% level. In 2014/2015, an additional ice minute increased injury susceptibility by .9% with 10% significance. Contact factors, hits and blocked shots, had mixed results. The results showed that every additional hit that a player either delivers or receives resulted in an increase of 0.006% to .11% of injury probability across the four seasons. Hits were significantly significant at the 5% level in 2015/2016 where an additional hit resulted in a .10% increase in injury susceptibility. The 2013/2014, 2014/2015, and 2016/2017 seasons retained their positive relationship with injury but lacked significance. To give some context, the average player experiences roughly 160 hits in a season, therefore at the end of a season, the player's likelihood of being injured had increased by approximately 8%. Blocked shots had significant *negative* effects on injury susceptibility. While these results were not expected, one likely explanation is that players who block the most shots are trained to position themselves in ways that protect them from injury. Playing style proved to have no significance across the four seasons. However, the results retained their expected positive sign, and have shown that a one unit increase in playing style (more aggressive playing) can be attributed to a 0.02% to 0.09% increase in injury risk. Again, to put this in perspective, the average playing style was 163, which can increase injury probability by 9% by the end of the season. However, the STYLE variable does have room for improvement with regard to the formulation. By weighting the variables differently (i.e. giving more weight to major penalties) rather than simply summing the values, there may be an increased fit to the model. Furthermore, there may be additional metrics that could better represent playing style. Nevertheless, my results are in agreement with Watson (1993) and Parkkari (2001) who also found that aging, high intensity exposure, and competition hours are the most significant contributors to injury. Saragiotto et al (2014) attributed increased injury rates to sport specific movements and behavioral features (hits and blocked shots). My results however, put less emphasis on these factors, and more on aging and exposure.

The player level regression also yielded satisfying results. The findings are detailed further in Table 8.

Table 8: Player Level Production Results

VARIABLES	(1) 2013/2014	(2) 2014/2015	(3) 2015/2016	(4) 2016/2017
Shooting %	52.85***	66.86***	47.33***	20.70***
	(14.16)	(14.99)	(12.65)	(24.54)
Face-off %	0.311	-2.043	2.840	-0.475
	(2.568)	(3.500)	(2.461)	(2.444)
Takeaways	0.0941	0.113**	0.0613	0.0573
	(0.0691)	(0.0533)	(0.0458)	(0.0639)
Giveaways	-0.107	-0.00785	-0.0406	-0.0134
	(0.0778)	(0.0648)	(0.0492)	(0.0502)
Hits Delivered	0.00134	0.000867	0.000500	0.00165
	(0.0113)	(0.00863)	(0.00988)	(0.0110)
Man Games Lost	-0.126*	-0.0613	-0.0739	-0.0684
	(0.0729)	(0.0874)	(0.0597)	(0.0650)
Missed Production	-0.505**	-0.243	-0.273*	-0.105
	(0.178)	(0.182)	(0.139)	(0.133)
CHIP	-3.18e-06*	-1.16e-06	-1.82e-06	-3.94e-07
	(1.82e-06)	(1.53e-06)	(1.31e-06)	(1.29e-06)
Age	0.460	2.364*	-2.391**	-0.247
	(1.244)	(1.397)	(1.139)	(1.329)
Age^2	-0.00574	-0.0419*	0.0462**	0.00491
	(0.0216)	(0.0246)	(0.0201)	(0.0232)
Constant	-11.90	-35.93*	27.77*	-6.543
	(18.00)	(19.73)	(16.02)	(18.45)
Control for Position	Yes	Yes	Yes	Yes
Observations	329	318	318	305
R-squared	0.110	0.137	0.116	0.118

Dependent variables here for all specifications is the Plus/Minus ratio.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results of the player level regression showed that Man Games Lost retained the negative expected sign and was significant at the 10% level in 2013/2014. More specifically, every additional game lost for a player lead to a decrease in his plus/minus ratio by an average of 0.05. This result adds some validity to the hypothesis that players who are injured for an extended period of time miss out more on the production of points while also being unable to contribute to team defense more so than those who are injured for less time. These findings in particular are in line with the literature surrounding inactivity and productivity, namely Ge and Lopez (2016) and Herrmann and Rockoff (2012). The coefficients for MGL were insignificant for the other three seasons. The performance based variables used to weight the injuries, missed production and CHIP, retained their expected signs. Missed production, showed slightly significant results in 2013/2014 and 2015/2016. In 2014/2015 and 2016/2017 the coefficients were also negative, but insignificant. Interpretations of the missed production variable are not very intuitive due to the construction, but it can be analyzed in the following way: as a player misses more games, he contributes less to the team's production. This missed production is represented in an average 0.25 unit decrease in his plus/minus ratio.

The coefficients of CHIP, the salary based measure, also showed that there is a slightly significant negative relationship between the weighted value of the players MGLs and his performance. More specifically, the average CHIP for a given player is roughly $415,000, which is value of the time the player spent recovering from an injury. Therefore, his CHIP value, can be attributed to a .78 point decrease in his plus/minus ratio. Slightly higher than the approximate loss put forth in the MISSPROD variable.

The coefficients of the other independent variables in the model show that a high shooting percentage was the best indicator of a favorable plus/minus ratio as well as slightly significant positive effect for defensive takeaways, mostly in line with the sports economics literature. Shooting percentages proved to be significant at the 1% level in all four seasons and can be interpreted in the following way: a reasonable increase in shooting percentage, a 0.01% increase in shooting efficiency, can contribute to .2 to .6 point higher plus/minus ratio. Defensive skill, as measured by takeaways and hits delivered, retained their positive signs, but lacked significance. It was surprising to see a slightly negative sign associated with faceoff percentages in 2014/2015 and 2016/2017. One possible explanation is that a face-off win does not necessarily result in more scoring chances. Nevertheless, the coefficient was insignificant for faceoff percentages. The other puck possession variable, giveaways, retained the negative sign but also lacked significance. These results reflected the findings of Zak et al. (1990), Carmichael et al. (2001) and Berri (1999), namely that shooting efficiency, and defense were critical to success in their respective sports as well. The age control variables also deserve a brief discussion: the variation that is reported with respect to age, most notably the mixed signs across seasons and the disagreement between Age and Age^2 show that there is slight degrees of heterogeneity in the model with respect to aging. For example, in 2013/2014, Age had a positive effect on plus/minus and Age^2 reported a negative relationship; the inverse was true in the 2015/2016 season.

The results of the Team Level fixed effects regression are presented in Table 9 below. The model employed three different output measures: league points, league point share, and win percentage.

Table 9: Team Level Production Results

VARIABLES	(1) League Points	(2) League Points (Share)	(3) Win %
Goal Differential	0.300***	0.00183***	0.00185***
	(0.0370)	(0.000227)	(0.000237)
Shots on Goal Differential	-0.00335	-2.03e-05	-1.76e-05
	(0.00251)	(1.54e-05)	(1.93e-05)
Face-off win Differential	-0.00113	-6.75e-06	-1.23e-05
	(0.00201)	(1.23e-05)	(1.57e-05)
Penalty Minute Differential	0.0116*	7.15e-05*	4.92e-05
	(0.00589)	(3.60e-05)	(4.83e-05)
Power Play Goal Differential	0.0290	0.000178	-0.000132
	(0.0522)	(0.000318)	(0.000390)
Save %	109.4	0.672	0.784*
	(67.35)	(0.412)	(0.465)

Hits	0.000485	2.59e-06	-3.81e-06
	(0.00210)	(1.29e-05)	(1.43e-05)
Injuries	-0.335**	-0.00201**	-0.00248**
	(0.141)	(0.000867)	(0.00101)
Man Games Lost	-0.00641	-3.97e-05	-2.57e-05
	(0.00918)	(5.61e-05)	(7.24e-05)
Missed Production	-0.0424*	-0.000258*	-0.000324*
	(0.0253)	(0.000155)	(0.000190)
CHIP (in millions)	-0.0137*	-7.63e-05	-0.00214
	(0.269)	(0.00164)	(0.00198)
Constant	4.626	0.0239	-0.126
	(61.36)	(0.375)	(0.423)
Season FE	Yes	Yes	Yes
Team FE	Yes	Yes	Yes
Observations	120	120	120
R-squared	0.968	0.967	0.959

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

At the team level, the results confirmed the hypothesis that the number of injuries a team incurs has negative effects on performance. It was shown that every additional injury that a team suffers, contributes to a 0.33 point decrease in league points, a .20% decrease in league point share and win percentage all with 5% significance. Man games lost retained the negative sign and revealed that every additional game that a player misses due to injury results in a 0.006 point decrease in league points, and a 0.004% and 0.003% decrease in point share and win percentage respectively. To add some context to these player health variables, if the average team suffered 24 injuries and 236 MGLs, league point production is decreased by approximately 8 points, and a 5% decrease in point share and win percentage due to the number of injuries alone. Furthermore, a team can expect to miss 1.5 league points due to the number of MGLs. Going deeper, a player's average MGLs due to injury is 13, therefore an injury to an average player can reduce team performance by 0.09 league points, multiplied by the average number of injuries for a team (24), we can assume that an average team with injuries to average players will lose out on

2 league points due to injury every season. Let it be noted that this analysis is based off of averages, and does not incorporate the value of players as previously discussed, meaning that this number will be higher (or lower) based upon the intrinsic value of the players that are injured. The results for CHIP and MISSPROD attempt to solve this limitation. The CHIP variable showed that when a team forfeits $1 million dollars due to player injury, there is a 0.01 point decrease in league points, a 0.007% decrease in point share, and a 0.2% decrease in win percentage. Keep in mind that the average team forfeits roughly $8 million in salary when players get hurt, therefore, at the end of the season, a team will experience a 0.11 point reduction and 5% decrease in point share as well as a 1.6% lower win percentage. This impact is in essence a magnifying effect on the team's already decreased performance.

The additional independent results show that most significant determinants of success in the NHL is goal differential followed by goalie play and the accumulation of penalty minutes and power play opportunities. Surprisingly, the model reported very low levels of significance for the other variables. The negative sign associated with faceoff percentages is again attributed to the fact that a face-off win does not necessarily result in more scoring chances and could depend on the zone in which the team is facing off in.

X. Conclusion

In conclusion, the results from the three models conducted for this paper show that player injury has a significantly negative effect on a NHL team's league performance, as well as on individual player output. The most significant factors that contribute to injury have also been identified as age, and game exposure time.

Tables 9 displays the results and methodology of the team level regressions based on various factors that contribute to, and take away from a team's on ice success. The model shows that every three additional injuries that a team suffers contributes to a one unit decrease in league points (a decrease of similar magnitudes was also found in league point share, and win percentage). Interpretation of the man games lost variable revealed that on average, a team will miss out on half a league point every season because of the prolonged absence of players. By attempting to incorporate a performance based variable to weight each injury, this model stipulates that injuries to players who are more valuable to a team, have a more significant negative impact on a team's overall performance. More specifically, as CHIP increases (the salary based value of player absence), a team's league performance suffers drastically. In some cases, injuries to the more valuable players (i.e. those with an above average CHIP value), could have a dramatically negative impact. For example, Steven Stamkos' injury in 2016/2017, sidelined the elite goal scorer for 65 games. Across the four seasons, his CHIP value of $6,737,805, ranked the highest. According to the team level model, his absence can be linked to a half point decrease in team league point production alone. The missed production variable tells a similar story, where Stamkos' failed to contribute an estimated 76 points, leading to a 3 point reduction in league points. Let it be noted that by choosing an upper level player like Stamkos to conduct this form of analysis, the calculations will report a higher than average predicted impact compared to a less skilled athlete. However, I use the more talented player to expose the differences in player value, and to show how teams are affected when highly valued players are hurt.

The team level regression was supported by an individual player analysis, which also reported a slightly significant negative relationship between a player's on ice production and the number of games missed due to injury. Table 8 presents the findings and shows that every additional game missed could potentially decrease the player's plus/minus ratio by 0.08. If this per game figure is multiplied by the average MGLs (13), we can assume that an average injured player will have a plus/minus ratio that is 1 unit lower than a player who remained healthy throughout the course of the season. The performance based variables used to weight the injuries show that this decrease can be magnified in players who have a higher value, proxied by their salary and predicted missed production. To continue with the Steven Stamkos example, the Center would have had a plus/minus ratio that could have been between 3 (CHIP) and 7 (MISSPROD) points higher had he not been injured.

The first model this paper analyzed was focused on the factors that contribute to injury in professional hockey. The probit regression showed that a player's age and his ice time are the most significant contributors to higher injury rates. In fact, a one year increase in age contributes to roughly a 1.5% increase in injury risk. Let it be noted that there was no non-linear trend in the likelihood of injury with respect of age. Additional ice exposure minutes also resulted in an average increase in injury risk by 1.20%, where a five minute increase in ice time per game could result in a 6% increase in injury likelihood.

Based off my finding and analysis, I urge coaches to protect their players by playing more athletes from the bench and making sure that the best players are getting the rest that they need. Coaches should find ways to use player more strategically to ensure that the players are not over-exposing themselves to the unavoidable high intensity factors associated with professional hockey. I also recommend that league officials explore the possibility of a shorter seasons to protect players from over exposure, however, this will likely be hard to implement due to the decrease in team revenues associated with the shortened season. Yet another policy recommendation that could be implemented is an injury prevention program that will help protect older and aging players who have begun to deteriorate over the span of their careers. This could come in the form of making additional physical therapy and treatment options available, or by instructing coaches and managers to closely monitor those who are slightly older. Avenues for further research include an improved formulation of the playing style variable within the player injury model to better represent the motivational and behavioral aspects of a hockey player. Additionally, the performance based injury variables deserve further improvement and analysis to fully understand how a player's value is related to decreased production. Moving away from hockey, this comprehensive production model should have applications in other professional leagues. It will be important to see the implications of injury on performance in other high intensity sports such as rugby and American football, as well as lower impact sports such as soccer and basketball.

References

Berri, D. J. (1999). Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics, 20*(8), 411-427.

Carmichael, F., Thomas, D., & Ward, R. (2001). Production and efficiency in Association Football. *Journal of Sports Economics, 2*(3), 228-243.

Depken, Craig A.,,II. (2000). Wage disparity and team productivity: Evidence from Major League Baseball. *Economics Letters, 67*(1), 87-92.

Gammage, K. L., Carron, A. V., & Estabrooks, P. A. (2001). Team cohesion and individual productivity: The influence of the norm for productivity and the identifiability of individual effort. *Small Group Research, 32*(1), 3-18.

Ge, Q., & Lopez, M. J. (2016). Lockouts and player productivity: Evidence from the National Hockey League. *Journal of Sports Economics, 17*(5), 427-452.

Herrmann, M. A., & Rockoff, J. E. (2012). Worker absence and productivity: Evidence from teaching. *Journal of Labor Economics, 30*(4), 749-782.

Jones, J. C., & Walsh, W. D. (1988). Salary determination in the National Hockey League: The effects of skills, franchise characteristics, and discrimination. *Industrial and Labor Relations Review, 41*(4), 592-604.

Jones, J. C., Nadeau, S., & Walsh, W. D. (1999). Ethnicity, productivity and salary: Player compensation and discrimination in the National Hockey League. *Applied Economics,* Kahane, L. H. (2005). Production efficiency and discriminatory hiring practices in the National Hockey League: A stochastic frontier approach. *Review of Industrial Organization, 27*(1), 47-71.

Krautmann, A. C. (1999). What's wrong with Scully-estimates of a player's marginal revenue product. *Economic Inquiry, 37*(2), 369-381.

Leeds, Michael (Michael A), & Von Allmen, P. (2011). *Economics of sports* (4th ed. ed.). Boston: Addison Wesley.

Mas, A. (2008). Labour Unrest and the Quality of Production: Evidence from the construction equipment resale market. *Review of Economic Studies, 75*(1), 229-258.

Noll, R. G. (2003). The organization of sports leagues. *Oxford Review of Economic Policy, 19*(4), 530-551.

Ornon G, Fritschy D, Ziltener J*, et al,* Professional ice hockey injuries: a 4 years prospective study *British Journal of Sports Medicine* 2011;45:366.

Parkkari, J., Kujala, U. M., & Kannus, P. (2001). Is it possible to prevent sports injuries? Review of controlled clinical trials and recommendations for future work. *Sports Medicine (Auckland, N.Z.), 31*(14).

Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy, 64*, 242-258.

Saragiotto, B. T., Di Pierro, C., & Lopes, A. D. (2014). Risk factors and injury prevention in elite athletes: a descriptive study of the opinions of physical therapists, doctors and trainers. *Brazilian Journal of Physical Therapy*, *18*(2), 137–143.

Scully, G. W. (1974). Pay and performance in Major League Baseball. *American Economic Review, 64*(6), 915-930.

Stewart, K. G., Ferguson, D. G., & Jones, J. C. H. (1992). On violence in professional team sport as the endogenous result of profit maximization. *Atlantic Economic Journal, 20*(4), 55-64.

Watson, A. W. (1993). Incidence and nature of sports injuries in Ireland. Analysis of four types of sport. *The American Journal of Sports Medicine, 21*(1), 137-143.

Zak, T. A., Huang, C. J., & Siegfried, J. J. (1990). Production efficiency: The case of Professional Basketball. In B. L. Goff, & R. D. Tollison (Eds.), (pp. 103-117) Texas A&M University Economics Series, no. 11; College Station:; Texas A&M University Press.