A Statistical Investigation of Data from the NHL Combine

by

Mengyang (Chris) Li

B.Sc., Huazhong University of Science and Technology, 2017

Project Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in the Department of Statistics and Actuarial Science Faculty of Science

© Mengyang (Chris) Li 2019

SIMON FRASER UNIVERSITY
Summer 2019

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.
Approval

Name: Mengyang (Chris) Li
Degree: Master of Science (Statistics)
Title: A Statistical Investigation of Data from the NHL Combine
Examiner Committee: Chair: Jinko Graham
Tim B. Swartz
Harsha Perera
Peter Tingling
Date Defended: May 22, 2019
Abstract

This project seeks to discover useful information from the NHL Combine results by comparing NHL Central Scouting Service rankings, NHL Draft results and measures of player evaluation. Data management is central to this project and we describe the details of handling datasets including the large and proprietary Combine dataset. Many data management decisions are made based on knowledge from the sport of hockey. The investigation of three questions of interest are carried out utilizing modern machine learning techniques such as random forests. Investigation 1 determines whether the Combine serves any purpose in terms of modifying the opinion of Central Scouting. Investigation 2 focuses on which test results of the Combine are important in predicting prospects’ future development. Investigation 3 considers how the Combine results revise Central Scouting’s beliefs.

Keywords: Central Scouting Service, Hockey, NHL Combine, Sports analytics
Dedication

"Nobody remembers number two, boys. Nobody remembers number two."
by Alexandre Daigle (The 1st pick in 1993 NHL Entry Draft)
I would first express my deepest gratitude and appreciation to my senior supervisor Dr. Tim Swartz for his continual support, guidance and patience throughout my two-year journey at Simon Fraser University. Without his encouragement and help, I would not have become a statistics graduate student in sports analytics. I would have remained a soccer fan with a mathematics background.

Thank you to all the members of my examining committee, Dr. Jinko Graham, Dr. Peter Tingling and Dr. Harsha Perera for their valuable suggestions and feedbacks.

I would also like to show my gratitude to Jonathan Wall, Senior Director of Hockey Operations and Analytics, Vancouver Canucks for sharing the proprietary NHL Combine datasets.

I am grateful to all the faculty members who taught me during my graduate studies, including Dr. Brad McNeney, Dr. Boxin Tang, Dr. Rachel Altman, Dr. Jinko Graham, Dr. Tom Loughin and Dr. Tim Swartz. I also wish to thank Charlene Bradbury, Sadika Jungic and Kelly Jay for their kind assistance and help. Special thanks to Ian Berkovitz for his support and help during my time at SFU.

I would also like to thank my fellow students and friends for all the fun we had along the way. I especially thank Lucas Wu and Harry Zhuang; you guys gave me lots of support, help and inspiration.

Last but not least I would like to thank my family, especially my mother and my ex-girlfriend for their love, support and understanding.
# Table of Contents

Approval ii  
Abstract iii  
Dedication iv  
Acknowledgements v  
Table of Contents vi  
List of Tables vii  
List of Figures viii  

1 Introduction 1  

2 Data Management 5  
2.1 NHL Central Scouting ranking results .......................... 5  
2.2 NHL Combine test results ....................................... 6  
2.3 NHL Entry Draft results ......................................... 7  
2.4 Measures of player evaluation .................................... 7  
2.5 Adjustment and combination of datasets ......................... 8  

3 Data Analysis 10  
3.1 Exploratory data analysis ........................................... 10  
3.2 Modelling assumptions ............................................ 14  
3.3 Investigation 1: Regress $k(y_2 - y_1)$ on $X$ .................... 15  
3.4 Investigation 2: Regress $k(y_3 - y_2)$ on $X$ .................... 20  
3.5 Investigation 3: Regress $k(y_3 - y_1)$ on $X$ .................... 23  
3.6 The impact of aerobic and anaerobic capacity .................... 25  

4 Concluding Remarks 29  

Bibliography 31
List of Tables

Table 2.1 An abbreviated list of CSS rankings for 2015 and 2016. . . . . . . . . 6
Table 2.2 An abbreviated list of Entry Draft rankings for 2015 and 2016. . . . . 7
Table 2.3 An abbreviated list of numbers of games played in the first two seasons. 8
Table 3.1 The table of prediction results for Dylan Strome. . . . . . . . . . . . . 26
Table 3.2 The table of prediction results for Jakub Zboril. . . . . . . . . . . . . 27
Table 3.3 The table of prediction results for Anthony Cirelli. . . . . . . . . . . . 27
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Three lines of investigation concerning the Combine data $X$.</td>
<td>3</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>The plot of $y_2$ versus $y_1$ for each year.</td>
<td>11</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>The fitness radar charts for five representative players.</td>
<td>12</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>The plots of $y_3$ versus $y_1$, $y_2$, and $y_2 - y_1$.</td>
<td>13</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>The boxplot of root MSPE for the four different methods using their best parameter settings in Investigation 1.</td>
<td>17</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Mean decrease in RSS of the 20 most important variables in Investigation 1.</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Partial dependence plots for the three most important variables in Investigation 1.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>The boxplot of root MSPE for the four different methods using their best parameter settings in Investigation 2.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>Mean decrease in RSS of the 20 most important variables in Investigation 2.</td>
<td>21</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>Partial dependence plots for the three most important variables in Investigation 2.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>The boxplot of root MSPE for the four different methods using their best parameter settings in Investigation 3.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 3.11</td>
<td>Mean decrease in RSS of the 20 most important variables in Investigation 3.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 3.12</td>
<td>Partial dependence plots for the three most important variables in Investigation 3.</td>
<td>25</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The National Hockey League (NHL) is a professional hockey league in North America, considered to be the premier professional ice hockey league in the world, and one of the major professional sports leagues in the United States and Canada. The championship trophy awarded annually to the NHL playoff winner, the Stanley Cup, is the oldest professional sports trophy in North America.

In 1917, the National Hockey League was formed in Montreal, Quebec, Canada. In its early years it had only four teams—all in Canada, thus the adjective 'National' in the league’s name. The league expanded to the United States in 1924, when the Boston Bruins joined. Currently the NHL comprises 31 teams: 24 in the United States and 7 in Canada. The NHL draws top prospects from all over the world and currently has players from approximately 20 countries. Canadians have historically constituted the majority of the players in the league, with an increasing percentage of American and European players in recent seasons.

As is the case with other major North American sports leagues, the NHL Entry Draft is a highly organized annual selection process developed to prevent the long-term dynasty and hegemony of a single franchise and increase equality and competitiveness in the NHL (Tingling 2017). The NHL Entry Draft has been held in June every year since 1963. The selection order in the Draft is determined by a combination of lottery, regular season standings, and playoff results. By granting the poorer performing teams in the previous season the first selection and the Stanley Cup winner the last choice, the Draft attempts to rebalance the league. Teams that acquire prospects by identifying exceptional talent and acquire veterans by trading draft picks can enrich their team’s potential or current performance.

Unique among North American professional leagues, the NHL operates its own official scouting department as a service to member teams (Tingling 2017). Each year before the Entry Draft, the Central Scouting Service (CSS) invites the top draft eligible hockey players from around the world to participate in the NHL Combine to undergo medical screening, fit-
ness assessment and interviews with NHL teams. The fitness assessment is a compilation of some of the core fitness tests generally employed by NHL team strength coaches. The interviews help teams know the players better in advance of the Draft. Technically, the Combine is not a competition amongst players, but the fitness scores will provide information on a player’s current capacities and assist the NHL teams in their preparation for the NHL Draft.

We have been given access to proprietary datasets that contain the Combine results from 2015 and 2016, involving over 230 players such as Conner McDavid, Jack Eichel and Auston Matthews. The players are comprehensively measured on more than 100 detailed tests. The CSS believes that some of the measurements such as aerobic and anaerobic fitness can provide insight into the projection of a player’s potential. The data were provided to us by Jonathan Wall, Senior Director, Hockey Operations and Analytics of the Vancouver Canucks.

There has been considerable research done on the NHL Draft (Tingling 2017). For example, questions such as these have been addressed in the literature:
- what is the relative value of a draft position?
- do some teams draft better than other teams?
- are the drafting decisions by teams superior to the assessments of CSS?

However, and perhaps due to its proprietary nature, we are unaware of any published research on the results of the NHL combine. One manuscript that does discuss the NHL combine is Schuckers and Argeris (2015). In this paper, the authors suggest that NHL teams use their own internal scouting to develop their own rankings of prospects using the additional information such as the Combine data. Schuckers and Argeris (2015) show that there is a clear benefit to the additional information that team’s possess relative to the CSS, even when accounting for costs.

There are three main lines of investigation associated with the project. First, we investigate whether the Combine serves any purpose. This can be addressed by comparing the Central Scouting rankings (carried out prior to the Combine) with the actual Draft results. Since there is limited new information available to teams prior to the draft, we posit that changes between Central Scouting rankings and Draft rankings ought to be primarily due to the Combine test results. We therefore investigate the effects of the Combine on team selection. Of course, it may be the case that individual teams largely ignore the Central Scouting rankings, and this may be a reason why Central Scouting rankings disagree with the Draft order. We consider this complication in our analysis.

Second, we investigate which test results of the Combine are really important. This can be handled by comparing the Combine results and the actual Draft results with subsequent
measures of player evaluation. Considering that the proprietary datasets are from two separate years (2015 and 2016), we consider statistics such as the cumulative Games Played in players’ first two NHL seasons as measures of player evaluation. To simplify the research and focus on the effects of the Combine, we ignore the team effect which is associated with opportunity and the growth environment. We consider the differences between Draft rankings and the measures of player evaluation and try to explain these differences via the Combine test results. We attempt to determine which test results of the Combine significantly show players’ potential performance and ability.

Third, we investigate how the Combine results revise Central Scouting’s belief. This can be addressed by comparing the Combine results and the Central Scouting rankings with the measures of player evaluation. We posit that changes between Central Scouting rankings and measures of player evaluation ought to be primarily due to the Combine test results. We therefore ignore other possible explanations such as different opportunities for development across teams.

There are four datasets related to this project: (1) NHL Central Scouting rankings, (2) NHL Combine test results, (3) NHL Entry Draft results and (4) measures of player evaluation. The datasets were obtained in the chronological order. Our three lines of investigation are carried out by regression procedures as conveniently summarized in Figure 1.1.

\[
\begin{align*}
\text{Data:} & \quad y_1 = \text{Central Scouting rankings (April 8th, 2015 & April 12th, 2016)} \\
& \quad X = \text{Combine covariates (June 6th, 2015 & June 5th, 2016)} \\
& \quad y_2 = \text{Draft rankings (June 26–27, 2015 & June 24–25, 2016)} \\
& \quad y_3 = \text{player quality evaluation rankings (two years after the Draft)} \\
\text{Investigation 1: Regress } & \quad y_2 - y_1 \text{ on } X \\
\text{Investigation 2: Regress } & \quad y_3 - y_2 \text{ on } X \\
\text{Investigation 3: Regress } & \quad y_3 - y_1 \text{ on } X
\end{align*}
\]

Figure 1.1: Three lines of investigation concerning the Combine data X.

Dataset X (the Combine results) is particularly large and considerable data management was required to prepare the data in a statistically useful format. Data management is a key first step in statistical analysis and is often overlooked. In Chapter 2, we describe the details of data management in this project where many decisions were made. We have attempted to make decisions that are statistically sound and are based on our knowledge from the sport of hockey.
Chapter 3 is the data analysis, where we research the three questions of interest using various methods including modern prediction techniques.

In Chapter 4 we provide concluding remarks.
Chapter 2

Data Management

Our investigation may be considered a big data problem in the sense that we possess various datasets from different sources. Before we dive into the data analysis, the datasets need to be collated, cleaned and converted into a statistically useful format. Datasets from different years need to be combined according to some specific rules. In the data management phase of the project, many decisions are made. We attempt to make sensible decisions which take realities of the sport of hockey into account. Here we describe some details of data management.

2.1 NHL Central Scouting ranking results

The Central Scouting Service ranks prospects according to whether they are from North America or Europe, and whether they are a skater or a goalie. Therefore, CSS produces a segregated set of rankings (North American Skaters, International Skaters, North American Goalies and International Goalies). The four lists make the job easier for Central Scouting but they are limited in terms of where a respective prospect ranks across the lists. And there is not a widely accepted metric to combine the four rankings into a single overall ranking (Schuckers and Argeris, 2015). Considering the fact that the majority of prospects are North American skaters, we utilize the Central Scouting rankings for North American skaters only. Since CSS updates the ranking several times every year, we use the final rank as players’ ranking results. In Table 2.1, we provide an abbreviated list of the CSS rankings for 2015 and 2016 as used in this project.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Year</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Connor McDavid</td>
<td>Pierre-Luc Dubois</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Jack Eichel</td>
<td>Matthew Tkachuk</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Noah Hanifin</td>
<td>Alexander Nylander</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Dylan Strome</td>
<td>Jakob Chychrun</td>
</tr>
</tbody>
</table>

Table 2.1: An abbreviated list of CSS rankings for 2015 and 2016.

2.2 NHL Combine test results

The NHL Combine test results are proprietary and form the most important dataset in the project. The results are from the NHL Scouting Combine 2015 and 2016, collected in the month of June. There are 119 and 114 top prospects that travelled to Buffalo, New York to take part in the fitness assessment. In each year around 210 players are selected in the 7-round Entry Draft but only about half of the prospects attend the Combine. We focus on players in the Combine and adjust the other three datasets so that they only consider the Combine participants.

The original Combine test results are divided into two parts: Fitness Testing and Measurements. Fitness Testing assesses players’ abilities in a series of specific tests such as the standing long jump. And Measurements accurately record players’ body sizes such as Waist circumference. First we combine the two parts together to form a single dataset \( X \) as defined in Figure 1.1. We noticed that there are tiny differences in the variables between 2015 and 2016. Compared to the 2015 Combine, the results of some test items in 2016 are missing and some new items are introduced. We removed the variables that are only available in one year away from the overall dataset. After making the above modification, the 2015 Combine test results and 2016 Combine test results have the same format which allows us to handle them in the same way.

We also simplified the Combine dataset by reducing the number of variables. There were many physical measurements that consisted of a left and right measurement (e.g. length of left forearm in cm and length of right forearm in cm). We do not believe that it is possible for hockey playing ability to be affected by the length of the left forearm but not the right forearm, for example. Therefore, in cases like this with left and right measurements (which we believe to be close in value due to bilateral symmetry of the human body), we averaged the two values to create a single variable. This also has the advantage of minimizing measurements error when there are two independent measurements. The Combine data resulted in two matrices, a \( 119 \times 70 \) matrix in 2015 and a \( 114 \times 70 \) matrix in 2016.
2.3 NHL Entry Draft results

The Entry Draft rank for a player is the place in the draft where he was selected. For example, Connor McDavid was drafted first overall in the 2015 NHL Entry Draft, so he has the rank 1. In Table 2.2, we provide an abbreviated list of the Entry Draft rankings for 2015 and 2016 as used in the project.

<table>
<thead>
<tr>
<th>Rank \ Year</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Connor McDavid</td>
<td>Auston Matthews</td>
</tr>
<tr>
<td>2</td>
<td>Jack Eichel</td>
<td>Patrik Laine</td>
</tr>
<tr>
<td>3</td>
<td>Dylan Strome</td>
<td>Pierre-Luc Dubois</td>
</tr>
<tr>
<td>4</td>
<td>Mitch Marner</td>
<td>Jesse Puljujarvi</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 2.2: An abbreviated list of Entry Draft rankings for 2015 and 2016.

Unfortunately, they are some prospects that have no Draft ranks, even though they had Central Scouting ranks and attended the Combine beforehand. Some of them gave up the Draft for various reasons and some of them were not selected by any team. Therefore these players are removed from the dataset in this project.

2.4 Measures of player evaluation

Measures of player evaluation are statistics that reflect players’ ability and performance in their first two NHL seasons. There are lots of available statistics such as goals, points and games played. Considering that goals and points depend on a player’s position, we choose the number of games played in the first two NHL seasons as the measure of player evaluation. Also, an NHL season is divided into Regular season and Playoffs. The prospects whose team played badly in the Regular season can’t qualify for the Playoffs. This can lead to a different number of total available games for each prospect. Therefore we only collected data from the 82-game regular seasons. That means the highest possible number of games played in prospects’ first two seasons is 164. In Table 2.3, we provide an abbreviated list of the numbers of games played by the prospects as used in the project.
<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Games Played</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Noah Hanifin</td>
<td>160</td>
</tr>
<tr>
<td>2</td>
<td>Jack Eichel</td>
<td>142</td>
</tr>
<tr>
<td>3</td>
<td>Connor McDavid</td>
<td>127</td>
</tr>
<tr>
<td>4</td>
<td>Jakob Chychrun</td>
<td>118</td>
</tr>
<tr>
<td>5</td>
<td>Clayton Keller</td>
<td>85</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>159</td>
<td>Carsen Twarynski</td>
<td>0</td>
</tr>
<tr>
<td>160</td>
<td>Max Zimmer</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.3: An abbreviated list of numbers of games played in the first two seasons.

The number of games played $y_3^*$ is an extremely left-skewed dataset. Only 34.6% of players in the 2015 Draft and 17.1% of players in the 2016 Draft played a single NHL game. And in Table 2.3, only 39 out of 160 prospects played a single NHL game, even though the mean number of games played is 11.69.

Our variable $y_3$ is a ranking of players according to the number of games played $y_3^*$. For players who have played the same number of games, we assign them equal ranks. Therefore, $y_1$, $y_2$ and $y_3$ are measured on the same scale.

2.5 Adjustment and combination of datasets

The next step in the management of datasets is the amalgamation of the four datasets described above into a single dataset in a statistically useful format. We only keep players who have available data in all of four datasets. And the number of players in our final dataset $y_3$ (player quality evaluation) is 160. We have lost 31.3% of the data at this stage, compared to the original Combine dataset $X$.

Considering that some players have been removed, the Central Scouting ranking and Draft ranking are not lists of consecutive positive integers any more in the final dataset. To better compare players’ ranks, we reorder the two rankings to create new lists of consecutive positive integers.

Another decision of adjustment concerns the difference of holistic quality between the 2015 Draft and the 2016 Draft. For example, the No.1 prospect in the 2015 Central Scouting ranking, Conner McDavid, was regarded as the best draft prospect since Sidney Crosby. And the No.1 prospect in the 2016 ranking, Pierre-Luc Dubois, was also an elite prospect but scouts believed that he was not as good as McDavid. So it is not reasonable to simply
assume that the two players with same place of Central Scouting ranking or Draft ranking in different years have exactly the same level of ability. To better compare prospects from different years, we decide to add 1 place for every player in 2016 for both the Central Scouting rankings and the Draft rankings.

The last step of adjustment is handling multicollinearity in the Combine data. So far our Combine data is a $160 \times 70$ matrix. There are some variables that are strongly correlated in the dataset. For example, the Yuhasz Skinfold Test discovers body fat weight and composition. It contains two test results: sum of six site skinfolds and body fat percentage. The body fat percentage is calculated by a linear equation of the sum of six site skinfolds. Multicollinearity could adversely affect our regression analyses and subsequent conclusions. To handle multicollinearity, we remove some variables from the Combine data according to the Variable Inflation Factor (VIF). The VIF for a variable is the ratio of the variance of the estimated coefficient when fitting the full model divided by the variance of the estimated coefficient when fitting on its own (James, Witten, Hastie, and Tibshirani 2013). The smallest possible value for VIF is 1, which indicates the complete absence of collinearity. And a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity. The VIF for the $k$th variable is:

$$
\text{VIF}_k = \frac{1}{1 - R^2_k}.
$$

(2.1)

where $R^2_k$ is the $R^2$ value obtained by regressing the $k$th variable on the remaining variables. To handle multicollinearity of the Combine data, we remove variables with large VIF values until the VIF values of the remaining variables are all less than 10. Our final Combine dataset is a $160 \times 61$ matrix. There are 61 variables remaining for Data Analysis.
Chapter 3

Data Analysis

A sensible, clean and statistically useful dataset has been created in Chapter 2. Our data analysis phase will start with exploratory data analysis (EDA). We will present interesting plots or charts to summarize main characteristics of the dataset. Then, in Section 3.2, we will make basic and concise modelling assumptions according to what we found in the EDA. In Sections 3.3 to 3.5 we will investigate the three main questions of interest.

3.1 Exploratory data analysis

EDA is the first step to find out what the data tell us before formal modelling. We will analyse different variables and try to construct relationships between them.

For every prospect, the Central Scouting rank $y_1$ and Entry Draft rank $y_2$ can be regarded as paired data. A player was evaluated with rank $y_1$ by CSS initially, then participated in the Combine and finally got picked in position $y_2$ in the Draft. We assume that the difference $y_2 - y_1$ is primarily due to the knowledge obtained via the Combine results. If $y_2 - y_1 < 0$, the player’s status has been elevated due to his good performance in the Combine. In contrast, if $y_1 - y_2 > 0$, the player is viewed less favourably following his bad Combine results. Figure 3.1 shows the plots of $y_2$ versus $y_1$ for each year.

The diagonal lines in Figure 3.1 separate all points into two regions. The points above the diagonal represent players whose draft results are depressed with respect to the CSS rankings and the points below the diagonal represent players whose draft results are elevated with respect to the CSS rankings. From Figure 3.1 we can see that prospects at the top level (ranks 1 - 20) are closer to the diagonal than other players. This is reasonable because top prospects are famous and are highly regarded by both CSS and teams before the Draft. There is low probability that the difference $|y_2 - y_1|$ will be large for top prospects. For players in the middle level, CSS may not spend as much time on these players as they do with top players. NHL teams internal scouting may gain insights from the Combine and
form different impression on these middle-level prospects. So players at the middle level have a higher probability to have a large difference $|y_2 - y_1|$. Another explanation is that the players in the "middle of the pack" may be perceived similarly; hence a large difference $|y_2 - y_1|$ may not involve a large difference in player quality.

![Figure 3.1: The plot of $y_2$ versus $y_1$ for each year.](image)

When comparing the two scatterplots in Figure 3.1, the points from the year 2015 are more dispersed than the points from the year 2016. This shows that in 2015, CSS and the NHL teams had larger differences with respect to player evaluation. The larger differences in 2015 might reveal more additional information from the Combine.

Next we consider the Combine data $X$. In Chapter 2 we mentioned that the Combine results contain Fitness Testing and Measurements. At this stage we have little intuition how the body size measurements reflect players’ ability. On the other hand, the test results from Fitness Testing are more interpretable on the spectrum from good to bad. For example, one could imagine that strength (via large measurements) is an advantage and also quickness (via small measurements) is an advantage. We now utilize Fitness Testing measurements of the Combine data to draw radar charts.

Radar charts are attractive but can also be misleading when visualizing data. Yet they are popular in the sports analytics world. For example, see the extensive use of radar charts at www.statsbomb.com. A difficulty with radar charts is that the order of the characteristics as we read each axis on the chart is arbitrary. Changing the positions of the characteristics on the chart can yield totally different plots for the same observation. Also, people tend to interpret largeness via area, and we note that the areas corresponding to radar charts are dependent on the characteristic orderings. Here we use linear combinations to create six...
characteristics (Upper Body Strength, Jump, Aerobic Fitness, Anaerobic Fitness, Agility and Balance) to summarize the Fitness Testing results. We plot radar charts (Figure 3.2) to see if there is a possible relationship between the Fitness Testing results and the difference $|y_2 - y_1|$. 

![Radar Charts](image)

Figure 3.2: The fitness radar charts for five representative players.

The left part of Figure 3.2 shows Fitness Testing results of three players: the red curve is for Brendan Guhle, who has the highest sum of the six characteristics. His difference $y_2 - y_1 = -14$ means his draft result is elevated 14 places with respect to the CSS ranking. The black curve is for Thomas Novak, who has the lowest sum of the six characteristics. His difference $y_2 - y_1 = 31$ means his draft result is depressed 31 places with respect to the CSS ranking. The green curve is for Connor McDavid, who is regarded as the best draft prospect since Sidney Crosby. McDavid received the first rank in both $y_1$ and $y_2$. His difference $y_2 - y_1 = 0$ so we set him as a template for comparison. By comparing the three prospects we can see that Brendan Guhle and Connor McDavid have much better results than Thomas Novak in almost all six characteristics. Based on this very small sample (only three players) of radar plots, there is a slight suggestion that the better (i.e. wider) the radar plot, the greater the chance that a player has an elevated rating after the Combine.

The right part of Figure 3.2 shows radar charts of players with extreme values of difference $y_2 - y_1$: the green curve is still our template Connor McDavid. The orange curve is the mean value of the 10 players who have the smallest difference $y_2 - y_1$. The mean difference $y_2 - y_1$ of these 10 players is $-32.9$, which means they are elevated 32.9 places on average in the draft with respect to the CSS rankings. The blue curve is the mean value of the 10 players who have the largest difference $y_2 - y_1$. The mean difference $y_2 - y_1$ of these 10
players is 31.6, which means their draft results are depressed 31.6 places on average with respect to the CSS rankings. By comparing the orange curve with the blue curve in the right part of Figure 3.2, we can see that the mean value of 10 players with smallest difference is greater than the mean value of 10 players with largest difference in all six characteristics. This again suggests that players whose draft results have been elevated with respect to the CSS rankings have better results from Fitness Testing than players whose draft results have been depressed. It may therefore be the case that Fitness Testing impresses NHL teams.

Next we are interested in the number of games played $y^*_3$. Recall that games played is a measure of player evaluation and that $y_3$ is the ranked variable associated with $y^*_3$. Figure 3.3 shows scatterplots of $y^*_3$ versus three variables ($y_1$, $y_2$ and difference $y_2 - y_1$). The orange curves in Figure 3.3 are cubic smoothing spline curves. Figure 3.3 (a) looks similar to Figure 3.3 (b). It shows that CSS and teams internal scouting have general agreement on player evaluation in prospects’ first two NHL seasons. The smoothing spline curves predict that the number of games played will decrease rapidly from around 100 to 10 in the rank range 1 - 20 of CSS or Draft ranking. The plots essentially indicate that the first 20 players drafted (or ranked 1 - 20 by CSS) have a reasonable chance of playing in the NHL. From that point onward, players have little chance of playing in the NHL in their first two seasons.

![Figure 3.3](image)

(a) $y^*_3$ versus $y_1$  
(b) $y^*_3$ versus $y_2$  
(c) $y^*_3$ versus $y_2 - y_1$

Figure 3.3: The plots of $y^*_3$ versus $y_1$, $y_2$ and $y_2 - y_1$. 
13
Figure 3.3 (c) provides an intriguing relationship between the number of games played \( y_3 \) and the difference \( y_2 - y_1 \). Almost all the non-zero values of \( y_2^* \) are in the interval \((-10, 10)\) of \( y_2 - y_1 \). Players with large differences, either negative or positive, have little chance to make their NHL debut in the first two seasons. The smoothing spline curve suggests that the closer \( y_1 \) and \( y_2 \) are, the more time the player will get on the ice. It shows that the similarity of opinions from CSS and teams may determine players’ prospect for development. Of course, referring back to Figure 3.1, many of these players are players who are thought of highly according to both \( y_1 \) and \( y_2 \).

### 3.2 Modelling assumptions

In this section we make some reasonable assumptions and make important decisions about our model. These assumptions are based on the results we obtained from data management and EDA. In order to take realities of the sport of hockey into account, we need to discuss these assumptions before the investigation of the three main questions of interest.

In Section 3.1 we focused on the difference \( y_2 - y_1 \), but we ignored the magnitude of \( y_1 \) and \( y_2 \). For a player who was ranked number one by CSS, \( y_2 - y_1 = 10 \) means his perceived value dropped greatly by the time of the Draft. But for a player whose CSS rank is 100, \( y_2 - y_1 = 10 \) is less meaningful. Therefore, we can not assume that the meaning of a single place change between \( y_2 \) and \( y_1 \) is equal for every prospect. We now address this issue by introducing a weight \( k \) to the difference \( y_2 - y_1 \).

We employ Michael Schuckers’ Draft Pick Value Chart (Schuckers 2016). The chart utilizes ice time for prospects drafted from 2003 to 2008 at each pick in their first seven NHL seasons to assess the value of each draft place. Also, the Draft Pick Value is scaled and rounded so that the first pick has value 1000 (Schuckers 2016). The value decreases rapidly from 1000 to 179 in the draft rank range 1 - 25 and drops slowly from 179 to 30 in the draft rank range 25 - 210. Based on the curve of the Draft Pick Value, we believe that it is a reasonable measurement that reflects the fact that decisions regarding top prospects are more meaningful than others. To set the weight factor \( k = 1 \) for the first pick, we define \( k \) for the \( i \)th draft pick as:

\[
k_{\text{(the i-th pick)}} = \frac{\text{Draft Pick Value for the i-th pick}}{1000}.
\]  \(3.1\)

The observed values of \( k \) in our dataset are within \((0.034, 1)\). We will regress \( k(y_2 - y_1) \) on the Combine data \( X \) in Section 3.3.
Next, we discuss the assumptions of normality and linearity for our models. Recall that $y_1$, $y_2$ and $y_3$ are ranks. The quantity $y_i - y_j$ is discrete and therefore the normality assumption used when regressing $k(y_i - y_j)$ on $X$ in a linear model may not be reasonable. We prefer to utilize modern machine learning methods such as random forests rather than models based on normality. An advantage of machine learning techniques is that they also do not require a linear relationship between the response $k(y_i - y_j)$ and $X$.

Last, as we mentioned in Chapter 1, we will investigate the effects of the Combine results on $k(y_i - y_j)$. We have ignored any additional information or factors that could affect $y_i - y_j$. But there indeed exists some additional factors. For example, NHL teams have internal scouts who may have different standards of player evaluation. Also, the internal scouts need to make different rankings from the CSS to show teams that they are creating real value. Both issues above could affect the results of the Draft $y_2$. In our investigations, we assume that the Combine is the only factor that affects $y_i - y_j$.

### 3.3 Investigation 1: Regress $k(y_2 - y_1)$ on $X$

In this Section we regress $k(y_2 - y_1)$ on the Combine results $X$ to investigate whether the Combine data has an impact on modifying opinions between the CSS rankings and the Draft rankings. Our first step is to fit the model using different modern machine learning techniques. Then we will choose the best model by comparing the model accuracy of the different methods. Lastly we will analyse the best model to find out if some test results from the Combine impact the difference $k(y_2 - y_1)$.

The modern regression methods that we will use in the investigations are: Regression Trees, Neural Nets, Random Forests and Gradient Boosting. All of these four methods are common techniques. The last three methods need lots of tuning where different parameter combinations will result in different models.

Regression Trees fit a piecewise-constant model to a feature space that has been partitioned into subsets. Partitioning takes place recursively. A split is done on one variable at a time to maximally reduce the residual sum of squares (RSS). The trees keep splitting until some stopping criterion is achieved. The final result consists of $M$ subsets. Within each subset, we use the sample mean of the data for prediction. The resulting fit of Regression Trees is discrete and coarse.

Neural Nets are considered one of the premier tools for 'automated' prediction. The most widely used version of Neural Nets is called the single hidden layer back-propagation
network (Hastie, Tibshirani and Friedman 2009). Each hidden node of the hidden layer is a function of a linear combination of inputs $X$. The final prediction of $Y$ is a function of a linear combination of hidden nodes. There are two main parameters that need tuning: the number of nodes and weight decay. Weight decay is a factor less than 1 which prevents the weights from growing too large. In Investigation 1, we try all combinations of $(1, 3, 5, 8, 10, 15, 20)$ hidden nodes and $(0, 0.001, 0.01, 0.1, 1)$ weight decay.

Random Forests are a substantial modification of bagging (Hastie, Tibshirani and Friedman 2009). Bagging fits the same regression tree many times to a bootstrap-sampled version of training data. On the base of bagging, Random Forests randomly select a subsample of explanatory variables at each split node of the trees. The predicted value is the mean from the terminal nodes in each tree in the forest. There are two main parameters that need tuning: the number of variables selected in each split and the minimum size of terminal nodes. Considering that the Combine data $X$ has 61 variables, we try all combinations of $(1, 3, 5, 15, 20, 30, 41, 61)$ number of variables selected in each split and $(2, 3, 5, 6, 8, 10, 15, 20)$ minimum size of terminal nodes.

Gradient Boosting produces a prediction model in the form of an ensemble of decision trees. It combines weak learners into a single strong learner iteratively using alternative loss functions. There are three main parameters that need tuning: the maximum depth of each tree, the learning rate and the fraction of the training set of observations randomly selected to propose the next tree in the iteration. In Investigation 1, we try all combinations of $(1, 2, 4, 6)$ maximum depth of each tree, $(0.001, 0.01, 0.1)$ learning rate and $(0.5, 0.6, 0.75, 0.9)$ fraction.

In order to choose the best model, we will split the data randomly into two parts: a training set that will be used to build the models and accounts for around 75% of the full dataset, and a test set that will be used to assess the model accuracy and accounts for around 25% of the full dataset. And we will repeat the random split 20 times and record the predicted value of the models on the test set each time. The model accuracy will be measured by the mean squared prediction error (MSPE). MSPE is calculated as follows:

$$\text{MSPE} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}$$

where $\hat{y}_i$ is the predicted value and $y_i$ is the observed value for response $i$ in the test set. We first determine the best parameter combination for each modern regression method according to MSPE. Then we will choose the best overall method by comparing the boxplots of root-MSPE.
Figure 3.4 shows the boxplots of root-MSPE for the four different methods using their best parameter settings in Investigation 1. We can see that compared to Neural Nets, Regression Trees, Random Forests and Gradient Boosting have better model accuracy. It's hard to choose the best method among Regression Trees, Random Forests and Gradient Boosting from the boxplot as they have similar results. Random Forests have the smallest mean MSPE but Regression Trees have the smallest median. Gradient Boosting is also good but it requires much more computing time in tuning than other methods. Considering the fact that Random Forests are a substantial modification of single Regression Trees, we choose Random Forests as our best model for the subsequent analysis. The best parameter combination for Random Forests is obtained by setting the number of variables selected in each split to 1 and the minimum size of terminal nodes to 8.

![Boxplot of root MSPE](image)

Figure 3.4: The boxplot of root MSPE for the four different methods using their best parameter settings in Investigation 1.

Our next goal is to find out if some variables of $X$ are important with respect to the difference $k(y_2 - y_1)$. Random Forests provide several variable importance measurements. We prefer to use the mean decrease in RSS (Hastie, Tibshirani and Friedman 2009). Random Forests are tree-based ensemble learning methods. At every parent node of a tree, if a candidate variable gets chosen, the split reduces RSS by some amount. The "important" variables will make relatively large changes when they are chosen and the "poor" variables will not change RSS much. We can measure how much each variable contributes to the reduction in RSS. The average reduction across all trees is the mean decrease in RSS. Figure 3.5 shows the mean decrease in RSS of the 20 most important variables in the Combine data $X$ on $k(y_2 - y_1)$. 
The three most important variables are the Calf Circumference, the length from Hind Foot to Great Toe (i.e. Foot Length) and the Standing Long Jumping Length. From Figure 3.5 we can see that there is a gap between Calf Circumference and Foot Length. And there is also a gap between Standing Long Jumping Length and the fourth most important variable Pro Agility Results. We focus on the three most important variables.

The Calf Circumference measures players’ calf muscle mass. And the Standing Long Jumping Length reflects the explosive power of the legs. Considering the fact that the three most important variables are all related to human’s lower limb, we think that the strength of lower limb might be more important with respect to the difference \( k(y_2 - y_1) \) than the strength of other muscles such as arms. For hockey players, the lower limb strength is likely to affect speed and stability on the ice. Figure 3.6 shows the partial dependence plots for the three most important variables.
A partial dependence plot shows the marginal effect of a variable on the predicted value of response (Friedman 2001). It can show whether the relationship between the response and a variable is linear, monotone or more complex. Recall that the negative difference $k(y_2 - y_1)$ represents a player has been elevated with respect to the CSS rankings. The partial dependence plot of Calf Circumference shows that the predicted difference $k(y_2 - y_1)$ decreases greatly when the Calf Circumference increases from 39 centimeters to 40 centimeters. This means that players with large Calf Circumference (greater than 40 centimeters) are elevated with respect to the CSS rankings.

In contrast, the predicted difference $k(y_2 - y_1)$ increases when the Foot Length increases from 26 centimeters to 28 centimeters. This means players with small feet (smaller than 26 centimeters) are elevated more than players with large feet. The result is surprising because in common sense, players with larger feet are more likely to be taller, stronger and more dominant on the ice. But in today’s NHL, speed dominates. No longer can teams do well with a slow roster. Players with smaller feet may be more fast and agile than players with larger feet. That might explain the partial dependence plot of Foot Length.

For Standing Long Jump, the predicted difference $k(y_2 - y_1)$ decreases greatly when the Standing Long Jump Length increases from 105 inches to 115 inches. It shows that players with better Standing Long Jump results are elevated with respect to the CSS rankings. The result makes sense to us because the explosive power of the legs is an important attribute for hockey players. The NHL teams would be impressed by players with strong explosive power of the legs and pick them earlier than other players.

For us, it seems unlikely that a team would modify the collective opinion (i.e. Central Scouting) about a hockey player based on Calf Circumference or Foot Length measurements. However, perhaps it is possible that the overall activities at the Combine provide an impression regarding a player’s explosive capabilities. We believe that it is plausible that such impressions could influence teams. We would like to discuss this observation with NHL
teams.

3.4 Investigation 2: Regress $k(y_3 - y_2)$ on $X$

In this Section we regress $k(y_3 - y_2)$ on the Combine results $X$ to investigate which test results of the Combine are important by comparing the Draft results with subsequent player quality evaluation rankings. We use the same procedure as we did in Investigation 1. The same four modern regression methods (Regression Trees, Random Forests, Neural Nets and Gradient Boosting) are used. Considering the fact that the response $k(y_3 - y_2)$ is different from the response in Investigation 1 and the Combine data $X$ remains the same, all of the tested parameter combinations are the same as in Investigation 1.

Figure 3.7 shows the boxplots of root-MSPE for the four different methods using their best parameter settings in Investigation 2. We can see that the results are similar to the results of Investigation 1. Compared to Neural Nets, Regression Trees, Random Forests and Gradient Boosting have better model accuracy according to root MSPE. Neural Nets are one of the premier tools for machine learning but the single hidden layer network we used might not be suitable for our datasets. For the remaining three methods, we find that the best parameter combination of Gradient Boosting uses only one tree (i.e. one iteration) to fit the model. The resulting fit of a single tree is discrete, random and not suitable for explaining variable importance. Therefore, we choose Random Forests as our best model for the subsequent analysis in Investigation 2. The best parameter combination for Random Forests is obtained by setting the number of variables selected in each split to 1 and the minimum size of terminal nodes to 20.
Our next goal is to find out if some variables of $X$ are important with respect to the difference $k(y_3 - y_2)$. We still use the mean decrease in RSS of the Random Forests resulting fit. Figure 3.8 shows the mean decrease in RSS of the 20 most important variables in the Combine data $X$ on $k(y_3 - y_2)$.

![Figure 3.7: The boxplot of root MSPE for the four different methods using their best parameter settings in Investigation 2.](image)

![Figure 3.8: Mean decrease in RSS of the 20 most important variables in Investigation 2.](image)

The three most important variables are the amount of Oxygen Utilized in the VO2max Test, the Quadriceps Circumference and the Peak Power Output in the Wingate Cycle.
Ergometer (WCE) Test. The VO2max Test assesses players’ aerobic fitness by measuring the amount of oxygen utilized during maximal cycle ergometer exercise. In contrast, the WCE Test assesses players’ anaerobic fitness by measuring the peak power output during short-term high intensity exercise (Gledhill and Jamnik 2007).

We observe that both aerobic and anaerobic fitness are important with respect to the difference \( k(y_3 - y_2) \). Anaerobic fitness is critical to hockey players because of the many rapid spurts of energy that are involved (Gledhill and Jamnik 2007). Aerobic fitness is also important on reflecting player’s endurance or stamina. But the fact that the mean decrease in RSS of aerobic fitness (VO2max) is larger than the mean decrease in RSS of anaerobic (WCE) may be surprising. In common sense, hockey is more of an anaerobic activity because of the short periods of maximum effort (i.e. line changes occur roughly every 45 seconds).

Figure 3.9 shows the partial dependence plots for the three most important variables. Recall that similar to \( k(y_2 - y_1) \), the negative difference \( k(y_3 - y_2) \) represents a player has been elevated with respect to the Draft rankings. The partial dependence plot of aerobic fitness shows that the predicted difference \( k(y_3 - y_2) \) decreases greatly and rapidly when the amount of oxygen utilized increases from 3.8 liters per minute to 4.2 liters per minute. This means that players with strong aerobic fitness are predicted to have better development in their first two NHL seasons than other players with respect to the Draft rankings.

![Partial Dependence on Aerobic Fitness](image)

![Partial Dependence on Quad Circumference](image)

![Partial Dependence on Anaerobic Fitness](image)

Figure 3.9: Partial dependence plots for the three most important variables in Investigation 2.

The partial dependence plot of quadriceps circumference shows that players with moderate strong quadriceps are predicted have better development in their first two NHL seasons than other players with respect to the Draft rankings. The players with thin quadriceps (smaller than 55 centimeters) are not predicted to have good development. Quadricep strength is also related to explosiveness (Burr, Jamnik, Baker, Macpherson, Gledhill and McGuire 2008) and therefore this observation is in keeping with the results from Investig-
tion 1.

For anaerobic fitness, the predicted difference $k(y_3 - y_2)$ decreases when the peak power output increases from 1200 Watts to 1250 Watts. The players with strong anaerobic fitness (between 1400 Watts and 1600 Watts) are predicted to have better development in their first two NHL seasons than other players with respect to the Draft rankings. The result makes sense to us because players with stronger anaerobic fitness have the ability to perform in bursts which is required during short line shifts.

3.5 Investigation 3: Regress $k(y_3 - y_1)$ on $X$

In this Section we regress $k(y_3 - y_1)$ on the Combine results $X$ to investigate which test results of the Combine are important to revise Central Scouting’s belief. Recall that $y_3$ is a measure of subsequent player evaluation. We use the same procedure as we did in Investigation 1 and 2. The same four modern regression methods (Regression Trees, Random Forrests, Neural Nets and Gradient Boosting) are used. All of the tested parameter combinations are the same as Investigation 1 and 2.

Figure 3.10 shows the boxplots of root-MSPE for the four different methods using their best parameter settings in Investigation 3. We can see that the results are similar to the results of Investigation 1 and 2. The model accuracy of Neural Nets are worse than the model accuracy of the other three methods. Among the remaining three methods, Gradient Boosting has the smallest mean MSPE but uses only one tree (i.e. one iteration) to fit the model in its best parameter combination. To better analyse variable importance, we again choose Random Forests as the best model in Investigation 3. The best parameter combination for Random Forests is obtained by setting the number of variables selected in each split to 1 and the minimum size of terminal nodes to 8.
Figure 3.10: The boxplot of root MSPE for the four different methods using their best parameter settings in Investigation 3.

Figure 3.11 shows the mean decrease in RSS of the 20 most important variables from the Combine data $X$ on $k(y_3 - y_1)$. The three most important variables are the same as Investigation 2: the Peak Power Output in the WCE Test, the Quadriceps Circumference and the amount of Oxygen Utilized in the VO2max Test. The only difference is that anaerobic fitness (WCE) is more important than aerobic fitness (VO2max) according to the mean decrease in RSS.

Figure 3.11: Mean decrease in RSS of the 20 most important variables in Investigation 3.
The similar results of Investigation 3 and Investigation 2 might be explained by the similarity of the CSS rank $y_1$ and the Draft rank $y_2$. The NHL teams make their picks ($y_2$) in reference to the CSS rankings ($y_1$). That causes $k(y_3 - y_2)$ to be strongly correlated with $k(y_3 - y_1)$. On the other hand, the similar results shows that both aerobic and anaerobic fitness are important not only with respect to players’ development in their first two NHL seasons, but also with respect to Central Scouting’s belief.

The partial dependence plots of Investigation 3 (Figure 3.12) are also similar to that of Investigation 2. For anaerobic fitness, the predicted difference $k(y_3 - y_1)$ decreases when the peak power output increases from 1200 Watts to 1550 Watts. We ignore the sudden increase which is caused by outliers in the right tail of the curve. The players with stronger anaerobic fitness are predicted to have better development in their first two NHL seasons than other players with respect to the CSS rankings.

![Partial Dependence Plots](image)

Figure 3.12: Partial dependence plots for the three most important variables in Investigation 3.

The partial dependence plot of quadriceps circumference shows that stronger quadriceps suggest better development with respect to the CSS rankings. For aerobic fitness, the predicted difference $k(y_3 - y_1)$ decreases rapidly when the amount of oxygen utilized increases from 3.8 liters per minute to 4.2 liters per minute.

The relative importance between aerobic and anaerobic fitness differ between Investigation 2 and Investigation 3. As we mentioned in Investigation 2, in common sense, hockey is more of an anaerobic activity. So the results of Investigation 3 make more sense to us. Anaerobic fitness is more important than aerobic fitness as a revision to Central Scouting’s beliefs.

### 3.6 The impact of aerobic and anaerobic capacity

Recall that in Investigation 2 and Investigation 3 we observed that both aerobic and anaerobic fitness are important with respect to the differences $k(y_3 - y_2)$ and $k(y_3 - y_1)$. These
investigations concerned the difference in performance based on the first two NHL seasons ($y_3$) with previous determinations of player quality ($y_1$ and $y_2$). In this Section, we investigate the degree to which aerobic and anaerobic measurements affect the differences $k(y_3 - y_2)$ and $k(y_3 - y_1)$.

In Investigation 2, the best model is based on Random Forests. With the original Combine data $X$, we can obtain the predicted value of $k(y_3 - y_2)$ for a given player. Then we can obtain the predicted value of $y_3$ since the values of $k$ and $y_2$ are known for each player. The predicted values of $y_3$ do not correspond to ranks because the predicted values are means from the terminal nodes in each tree in the forest. To handle the non-integer values, we give each player a rank $y_{3p}$ according to the predicted value of $y_3$. The new ranking $y_{3p}$ is the predicted player evaluation ranking.

Next we choose some representative players who have weak aerobic or anaerobic ability, change their test results regarding aerobic and anaerobic fitness and repeat the prediction procedure to see how $y_{3p}$ changes. For aerobic fitness, the first quartile of the amount of Oxygen Utilized is 4.4 liters per minute. The median is 4.7 liters per minute and the third quartile is 5.1 liters per minute. For anaerobic fitness, the first quartile of the Peak Power Output is 1275 Watts. The median is 1367 Watts and the third quartile is 1477 Watts. We interpret the first quartiles as bad values, and the third quartiles as good values.

The first player we are interested in is Dylan Strome. He has weak aerobic fitness (4.1 liters per minute) and weak anaerobic fitness (1188 Watts). His Draft ranking $y_2$ is 3 and his player quality evaluation ranking $y_3$ is 16. Table 3.1 shows the prediction results for Strome.

<table>
<thead>
<tr>
<th>Player 1 – Dylan Strome</th>
<th>Original results</th>
<th>Stronger Aerobic Fitness &amp; Original Anaerobic Fitness</th>
<th>Original Aerobic Fitness &amp; Stronger Anaerobic Fitness</th>
<th>Stronger Aerobic Fitness &amp; Stronger Anaerobic Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic Fitness (liters per minute)</td>
<td>4.1</td>
<td>5.1</td>
<td>4.1</td>
<td>5.1</td>
</tr>
<tr>
<td>Anaerobic Fitness (Watts)</td>
<td>1188</td>
<td>1188</td>
<td>1477</td>
<td>1477</td>
</tr>
<tr>
<td>Draft ranking ($y_2$)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Predicted value ($y_{3p}$)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3.1: The table of prediction results for Dylan Strome.

We can see that based on the original Combine results, $y_{3p} = 5$ for Dylan Strome under all variations of aerobic and anaerobic fitness. This is not greatly surprising since Strome was viewed as a top prospect where a couple of test values were not likely to change his evaluation.
The second player we are interested in is Jakub Zboril. He also has weak aerobic fitness (3.8 liters per minute) and weak anaerobic fitness (1246 Watts). His Draft ranking $y_2$ is 11 and his player quality evaluation ranking $y_3$ is 28 (The 28th rank of $y_3$ in Draft year 2015 means this player has never played a single NHL game in his first two seasons). Table 3.2 shows the prediction results for Zboril.

<table>
<thead>
<tr>
<th>Player 2 – Jakub Zboril</th>
<th>Original results</th>
<th>Stronger Aerobic Fitness &amp; Original Anaerobic Fitness</th>
<th>Original Aerobic Fitness &amp; Stronger Anaerobic Fitness</th>
<th>Stronger Aerobic Fitness &amp; Stronger Anaerobic Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic Fitness</td>
<td>3.8</td>
<td>5.1</td>
<td>3.8</td>
<td>5.1</td>
</tr>
<tr>
<td>(liters per minute)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaerobic Fitness</td>
<td>1246</td>
<td>1246</td>
<td>1477</td>
<td>1477</td>
</tr>
<tr>
<td>(Watts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draft ranking ($y_2$)</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Predicted value ($y_{3p}$)</td>
<td>15</td>
<td>13</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3.2: The table of prediction results for Jakub Zboril.

We can see that based on the original Combine results, the predicted ranking $y_{3p}$ for Zboril is 15. But under variations of stronger aerobic or anaerobic fitness (or both), Zboril’s predicted player evaluation ranking $y_{3p}$ is elevated. Considering the fact that Zboril never played a single NHL game in his first two seasons, we suggest that with stronger aerobic or anaerobic capacity (or both), Zboril might have received a chance to make an NHL debut in his first two seasons.

The third player we are interested in is Anthony Cirelli. He has weak aerobic fitness (4.1 liters per minute) and extremely weak anaerobic fitness (921 Watts). His Draft ranking $y_2$ is 54 and his player quality evaluation ranking $y_3$ is 28. He also never played a single NHL game in his first two seasons. Table 3.3 shows the prediction results for Cirelli.

<table>
<thead>
<tr>
<th>Player 3 – Anthony Cirelli</th>
<th>Original results</th>
<th>Stronger Aerobic Fitness &amp; Original Anaerobic Fitness</th>
<th>Original Aerobic Fitness &amp; Stronger Anaerobic Fitness</th>
<th>Stronger Aerobic Fitness &amp; Stronger Anaerobic Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic Fitness</td>
<td>4.1</td>
<td>5.1</td>
<td>4.1</td>
<td>5.1</td>
</tr>
<tr>
<td>(liters per minute)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaerobic Fitness</td>
<td>921</td>
<td>921</td>
<td>1477</td>
<td>1477</td>
</tr>
<tr>
<td>(Watts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draft ranking ($y_2$)</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>Predicted value ($y_{3p}$)</td>
<td>60</td>
<td>58</td>
<td>58</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 3.3: The table of prediction results for Anthony Cirelli.

We can see that based on the original Combine results, the predicted ranking $y_{3p}$ for Cirelli is 60. But under variations of stronger aerobic or anaerobic fitness (or both), Cirelli’s predicted player evaluation ranking $y_{3p}$ is elevated to 58. Unfortunately, the 58th rank is
not good enough for Cirelli to make an NHL debut in the first two seasons. In fact, he made his debut in his third season and played all of 82 regular-season games in his fourth season.

The prediction results of the three representative players suggest that players with stronger aerobic or anaerobic capacity may have better development than predicted under low aerobic and anaerobic levels. However, the improvement is modest at best.
Chapter 4

Concluding Remarks

This project is the first statistical analysis that investigates the effects of the NHL Combine data on the CSS rankings, the Draft rankings and player quality evaluation rankings. We utilize some modern machine learning techniques to assess important test results of the Combine in the three main lines of investigation.

In Investigation 1, we find out that the strength of the lower limb is important with respect to the difference $k(y_2 - y_1)$. Especially, NHL teams appear impressed by players with strong explosive power of the legs and pick them earlier than other players.

The results of Investigation 2 and 3 are similar. Both aerobic and anaerobic fitness are important components of player development in their first two NHL seasons with respect to the CSS or the Draft rankings. But the two Investigations give opposite results on which of the two fitness measurements is more important. We need additional information to investigate this in future research.

Is there a take-away message in this project that may lead to a competitive advantage for the Vancouver Canucks and other NHL teams? First, we suggest that Investigation 3 was the most important as it concerned the difference between collective opinion (i.e. Central Scouting $y_1$) and actual hockey performance ($y_3$). In Investigation 3, there was an indication that high aerobic and anaerobic test results are predictors of elite hockey performance. Finally, in Section 3.6, the performance effects of aerobic and anaerobic results were assessed. Unfortunately, these effects were minor. However, we might suggest that in choosing between two mid-draft prospects who are regarded similarly, an NHL team may benefit by choosing the player with higher aerobic and anaerobic measurements.

All of the three Investigations are based on datasets collected over two years. We would have benefitted from more years of data.
The NHL Combine focuses on players’ physical fitness. But the psychological fitness can also influence players’ performance on the ice and their future development. Psychological fitness has been investigated and proved to be important in other sporting activities (Johnson, Stimpson and Clark 2012). It may be useful to extend the Combine by taking psychological fitness into consideration.
Bibliography


