Predicting Advancement from Midget to Junior Ice Hockey Leagues: Individual Performance and Team Skill

By Sam Faier and Akhil Sehgal

Using analytical performance data provided by CKM Sports Management, we identify the properties that predict the advancement of hockey players from the midget to junior level, looking at both individual performance statistics and the team effect of the top 100 ranked midget players that play in leagues across Western Canada. Using a linear probability model, we predict that an additional point per game increases the probability of advancement by 33.9%, while the number of games played and player position (forward or defence) are insignificant in explaining advancement. However, defensive players benefit more than forwards from an additional point per game. We also devise two measures of team skill, to control for the team effect. When we add these variables to our individual model, they return insignificant results, reinforcing the importance of individual performance statistics.

I. Introduction

Ice hockey is one of the most popular sports in Canada, both in viewership and participation. The sport is synonymous with Canadian culture, and one would be hard-pressed to find a Canadian who does not have at least a passing interest in their home team. Many young people across the country enjoy the sport and have been playing it from a young age. In fact, 43% of players in the NHL are Canadian—the largest group in the league, despite a majority of the teams being based in American cities (Szporer, 2021). Thus, an important subject for skilled young Canadian hockey players, as well as hockey fans in general, is the progression path for players to move up the ranks and reach higher levels of professional hockey. Many factors influence player progression throughout the various stages of their hockey career, and studies have been done on some of these factors, revealing the importance of individual player performance statistics. Much of this literature focuses on the highest level of progression, namely the NHL drafts. This paper contributes to the existing literature on player progression in professional hockey by examining the individual performance factors that predict the advancement of players in the earlier stages of their careers, from midget to junior level hockey leagues. Additionally, we are interested in measuring the skill of a player's team, in order to determine whether this has an effect on our predictions.

To answer this question, we look at a dataset consisting of the top 100 midget and junior players across several leagues in Western Canada. We create a variable that identifies whether a player advances from the midget to junior dataset. We then perform a linear probability model regression of this variable on a player's performance, represented by their points per game, as well as a position variable and a variable for the number of games played by that player during that season. We find that one additional point per game on average increases the probability of advancement by 34%, and this result is significant at the 1% level. We also split points per game into goals per game and assists per game, which return near identical results. We examine this result for forwards and defencemen separately, finding that defensemen actually see a higher coefficient on points per game, with one additional unit leading to an increase of 41% in the probability of advancement. We speculate that this is due to the increased competition for defencemen in being chosen to advance, causing players in that role with higher points per game to stand out.

We then create two variables to measure the skill of a player's team. The first is a "leave-out" average of the team's points per game over the three years of the dataset. This variable is positively and significantly correlated with the dependent variable, but when we add it to our individual model, this effect disappears, leaving individual performance nearly unchanged as the main predictive factor. The second measure is the number of players on a team, excluding the individual player, that appear in the dataset. Since this dataset represents the top 100 players from various midget leagues, we argue that a higher number of players on a team in the dataset corresponds to a more highly skilled team. This variable is not significantly correlated with advancement for an individual player and does not significantly alter the results of our individual model when included. For robustness, we add team and year fixed effects to our individual model, which also do not significantly alter the coefficient on points per game. We perform a probit regression of our individual and team models as well, and similarly find points per game to be the central predictive factor, though with a smaller marginal effect. Ultimately, we find that individual performance as measured by points per game is the most important factor in predicting advancement from midget to junior leagues, and this result remains unchanged when we include measures of team skill.

II. Background

The path to professional hockey starts much earlier than many realize. There are generally two routes for aspiring hockey professionals to reach the highest levels of competition: major junior leagues, and college teams. This paper will focus on the major junior leagues in Western Canada. Specifically, we are interested in the progression of players from midget to junior leagues. Players in the midget leagues are between the ages of 15 and 17, while junior leagues begin at age 16 until age 20 (Hockey Canada n.d.). The midget leagues in our dataset include the Alberta Midget Hockey League, Canadian Sport School Hockey League, and BC Hockey Major Midget League. The junior leagues are the Alberta Junior Hockey League, Western Hockey League, and British Columbia Hockey League.

Major junior leagues draft players from lower-level junior leagues, as well as outstanding players from major midget leagues. The difference in skill from major midget to junior is significant: in our dataset, only 122 out of 775 players in major midget leagues between 2016-2018 appeared in major junior leagues in the following years (2017-2019). Only the most skilled players from major midget are selected to advance to major junior, and from there only the most skilled are selected to advance further. Thus, the question arises: how do teams determine player skill?

There appears to be an obvious answer for this: on-ice performance. Research by He (2020) studied the advancement of forwards from midget to junior leagues, using the same dataset as this paper from previous years. He tested the effects of various player statistics of forwards on advancement to major junior leagues and found that the significant determining factor was points. He focuses on performance totals (total points, total goals, and total assists) rather than per game performance metrics and finds that the effect of total points on advancement increases the higher a player's total points.

Many studies in the past have also focused on individual performance, but there are factors that affect performance which are difficult to estimate. Riley (2017) states that for players in the NHL, "individual player performance is highly dependent on linemate performance" (4). Similarly, Idson and Kahane (2000) argue that a player's team has an effect on that player's compensation not only through the team's financial position, but also indirectly through increases in player performance, stemming from complementarity increases to productivity.

Though these findings come from analysis of NHL data, it is reasonable to assume productivity increases from team complementarity may have an effect on player performance across different levels of competition.

III. Data Description and Sample Construction

The dataset that we obtained from our community partner, CKM Sports Management, is an unbalanced panel that includes individual performance statistics, player information, and team information for the top 100 ranked midget level players from 2016-2018 across the three midget leagues, and the top 100 ranked junior level players from 2017-2019 across the three junior leagues. The player statistics given in the dataset include rank, goals, assists, points, points per game, penalty infraction minutes, and games played. The dataset also includes player name, position, team, year, and league.

The original dataset contained 2027 observations and was split into six sections, divided by year and level (junior or midget). However, for our purposes, we combined the midget players from 2016-2018 into one data frame and the junior players from 2017-2019 in another separate data frame. The midget data frame becomes our main data frame, and all subsequent changes are made to this set of data. In our main data frame, there were a few players that played on multiple teams in a given year. For these players, we simply took the totals of their performance statistics across all teams for that year as a single observation. Table 1 describes the dependent and independent variables that we use in our models.

Table 1: Description of Variables

Variable	Description
LevelUp	Binary variable, indicating whether a player advanced from midget to junior
Points per Game	Total points per games played (Points/Games Played) for an observation
Goals per Game	Total goals divided by games played for an observation
Assists per Game	Total assists divided by games played for an observation
Games Played	Total number of games played for an observation
PIMS	Penalty infraction minutes, total time spent in the penalty box
Forward	Binary variable, indicating whether a player is a forward (Centre, Rightwing, Left Wing)
Defence	Binary variable, indicating whether a player is a defencemen
TeamCount	Number of players that appear in a dataset for a given team
TeamYearCount	Number of players that appear in a dataset for a given team, categorized by year
Team Points per Game	Aggregate sum of points per game for all players on a given team

Note: This table presents the main variables and their descriptions.

We first create explanatory variables that capture position and per game metrics for performance. Since position was listed in the same column as name, we extracted these positions and created a binary variable, denoting whether a player was a defencemen or a forward (listed as either center, rightwing, leftwing, or forward in the dataset). For example, if a player is a forward, he would have a 1 for Forward and 0 for Defence. Another binary variable (TwoPosition) was created for any player that was listed as both a forward and defencemen. We also transform two variables to consider per game averages. We divide goals and assists by games played for each observation in order to capture these variables on a per game basis. For example, we calculate goals per game by dividing total goals by total number of games for each observation. We transform assists to assists per game utilizing the same method (total assists/total games played for each observation).

We then create our dependent variable, which is a binary variable that indicates whether an observation in the midget data frame also appears in the junior data frame. In the context of our model, LevelUp represents the probability of a midget level player advancing to the junior level. By matching player names from midget to junior, we assign a value of 1 in the midget data frame if a player also appears in the Junior data frame, and a 0 if not. This means that if a player has advanced, we would see 1 in LevelUp for that observation and a 0 if the midget player has not advanced. This way, we avoid values of 1 that would occur after a player has already advanced, which would otherwise have been a problem for our model. Furthermore, if a player appeared in multiple years, we only assigned a value of 1 for their final year and 0 for the others. In doing so, we avoid assigning a 1 for observations in years prior to advancement. Table 2 presents the summary statistics for our key variables.

Table 2: Summary Statistics for Main Variables

Variable	Minimum	Median	Maximum	Mean	Std. Dev.
LevelUp	0	0	1	0.14	0.34
Points per Game	0.41	0.87	2.74	0.95	0.37
Goals per Game	0	0.35	1.44	0.38	0.2
Assists per Game	0.08	0.52	1.81	0.56	0.24
Games Played	10	34	40	33.55	4.44
Forward	0	1	1	0.83	0.38
Defence	0	0	1	0.16	0.37
TeamCount	3	23	39	22.96	8.82
TeamYearCount	1	9	16	8.5	3.33
Team Points per Game	2.06	21.59	42.39	22.12	9.67

Note: This table presents the summary statistics for the main variables of the midget dataset.

In our main dataset, with these transformations, we have a total of 900 observations. 775 of these observations represent a unique player, as some players had been at the midget level for a few years prior to advancing. Out of these 775 unique observations, 122 midget players advanced to the junior level. We include all positions under the assumption that there are different variables (points per game, games played, etc.) that capture the contribution of different positions. Since defencemen contribute to the success of the team and their contributions can be captured through individual performance metrics, we find it imperative to include them in our predictive model. In our robustness checks, we run a model without defencemen and with only defencemen to confirm our suspicions.

In figure 1, we see that most players in the dataset average between 0.6 and 1.0 points per game. In figure 2, we compare the points per game for players that did not advance to players that did. Players that level up average 1.29 points per game, while players that do not advance, average 0.90 points per game. In a preliminary analysis, we note a substantial difference in performance between those that advanced and the players that did not. These differences extend to other performance indicators such as goals per game and assists per game.

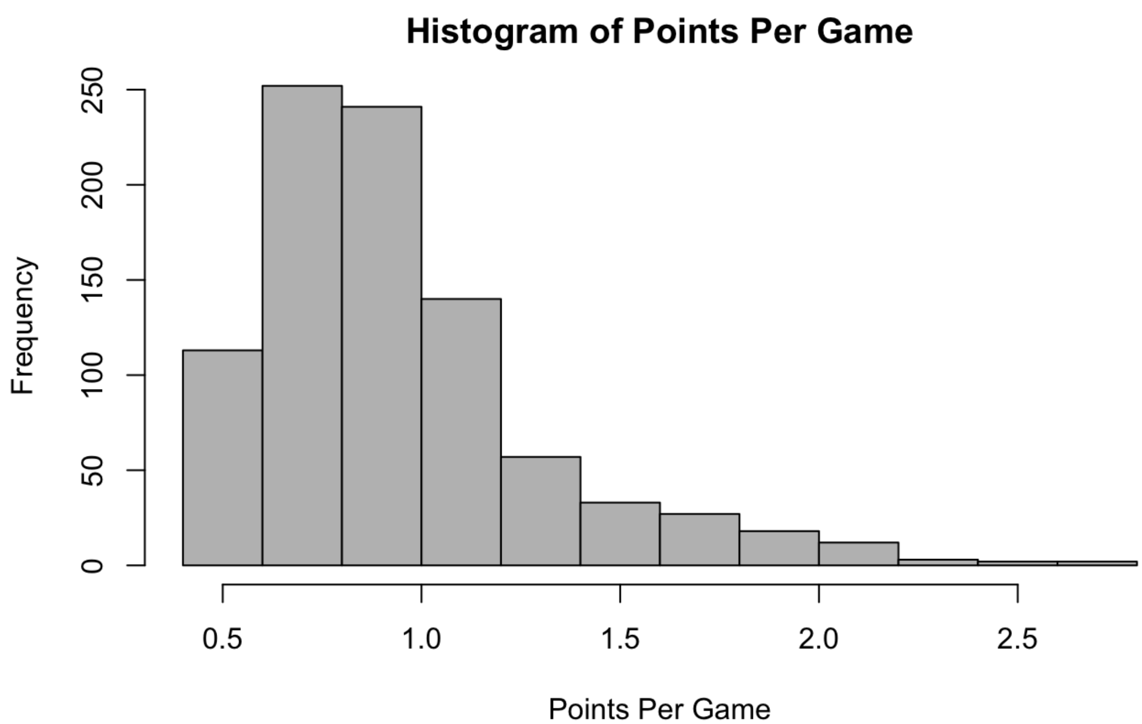

Figure 1. Histogram of Points per Game for observations in dataset.

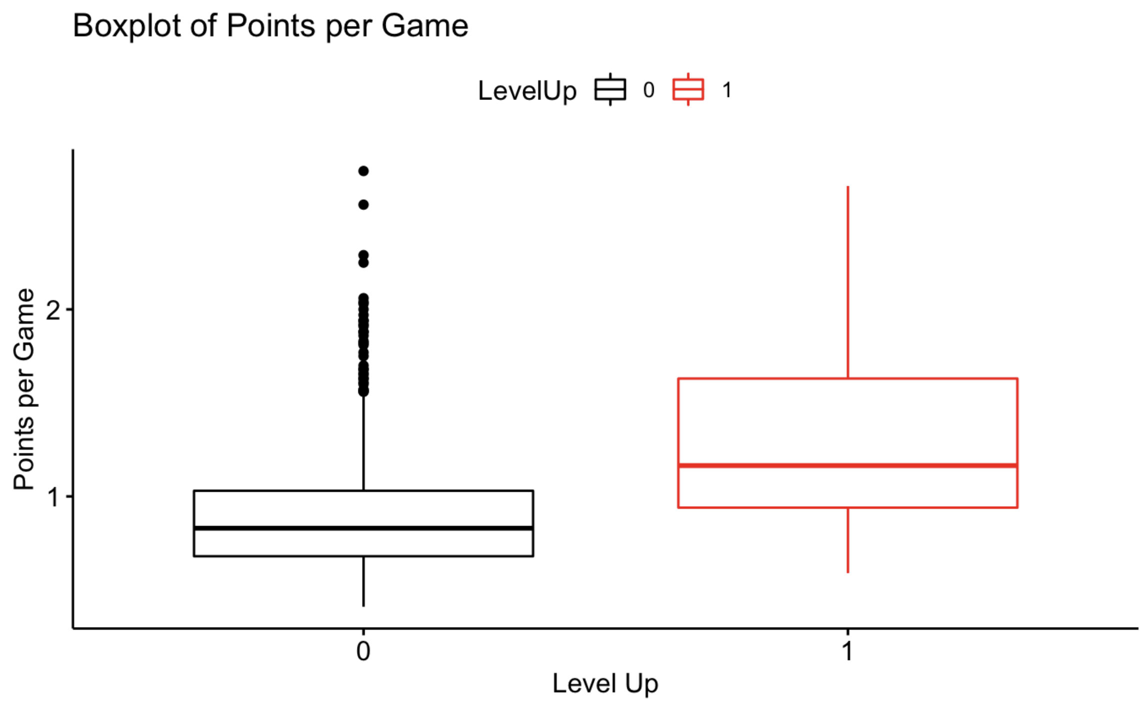

Figure 2. Boxplot comparing the top 100 ranked midget players that do not appear in the junior data frame (LevelUp = 0) to players that appear in the data frame (LevelUp = 1).

As for team-related changes to the dataset, we create 3 new variables: The first two are TeamCount and TeamYearCount. TeamCount represents the number of players that appear in the dataset for a given team, while TeamYearCount does the same, but also categorizes by year. The third variable is TeamPTS.GM, an aggregate sum of points per game for all players on a given team. For these team variables, players on multiple teams are counted once for each team they played on, to accurately measure the number of players and performance statistics of all the players who played on each team. We use these variables to create measures of team effectiveness, which we will explain in the methodology section.

IV. Methodology and Empirical Model

A. Individual Model

The empirical model that we use to estimate a prediction on the probability of advancing from midget to junior is a linear probability model that has the dependent variable of LevelUp and 3 independent variables. This model is shown by the following specification:

$$LevelUp_i = \beta_0 + \beta_1 Pointspergame_i + \beta_2 GamesPlayed_i + \beta_3 Forward_i + \varepsilon_i \qquad (1)$$

where LevelUp is a dummy variable indicating probability of advancing from midget to junior level, points per game is the average points scored per game played for an observation, games played is the total number of games played for an observation, and forward is a dummy variable, indicating whether a player is a forward (forward =1). β_0 is the constant and ε_i is the residual.

B. Individual Model Selection

Since our dependent variable is binary and our model is predictive, we use a linear probability model. Prior to selecting this model, we tested a logit model where we took the log of the dependent variable; however, we saw no substantial changes in the estimation of our explanatory variables. Using a logit model would complicate our results with no additional benefit. We also consider a probit model that tests both our individual and team variables in our section on robustness. However, we use a linear probability model to explore the properties that explain advancement. When regressing LevelUp (outcome variable) on Points per Game (explanatory variable), and plotting our results, we can see in figure 3 that most of the predicted values (as shown by the red circles) are between 0.2 and 0.8, with no values over 1.0 and only a few below 0.0, further justifying the decision on our model. Using a linear probability model, we assume a linear relationship between the outcome and explanatory variables. Since we are using a linear probability model, we use heteroskedasticity-robust standard errors.

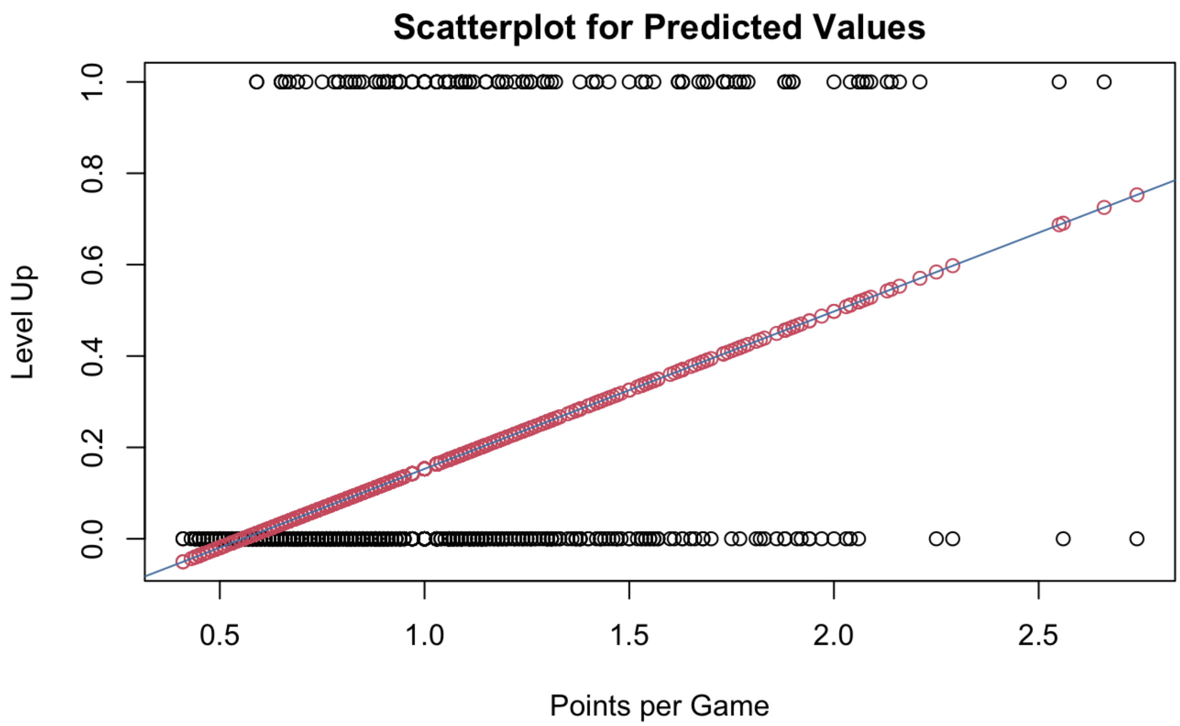

Figure 3. Scatter Plot plotting the predictive values of LevelUp given a player's points per game. The black circles represent whether players appeared in the junior dataset given their points per game. The red circles represent the predicted values for each observation. The blue line represents results from LevelUp regressed on Points per Game.

Raw performance (total goals, total assists, and total points) is prone to multicollinearity as players that play more games have an additional performance advantage. We focus on per game performance (points per game, goals per game, assists per game) rather than raw totals. To determine which per game performance metric to use (points per game or goals per game and assist per game), we ran a regression with all three variables as shown in column 1 of table 3. We would expect each variable to be significant, and for Goals per Game and Assists per Game to have a positive relationship with LevelUp; however, this was not the case. The reason we cannot use all three variables is that points is equal to goals plus assists, leading to multicollinearity. To avoid multicollinearity, we must choose between points per game or goals per game and assists per game. To make this decision, we compared the two. In table 3, LPM1 (LevelUp regressed on Points per Game) and LMP2 (LevelUp regressed on Assists per Game plus Goals per Game) both produce significant results. LPM1 has a higher adjusted R squared; however, R squared is not meaningful in this context. We then look to testing the sensitivity and specificity of LPM1 and LPM2. The sensitivity is the exact same for both LPM1 and LPM2, and the specificity for LPM1 is 9.8% and 10.7% for LPM2. Since both LPM1 and LPM2 perform similarly, we choose the simpler model, points per game, as it has one less variable. However, in our robustness checks, we also test goals per game and assists per game with our final model.

Before settling on points per game, we consider the log of points per game. Figures 2 and 3 show that points per game is right-skewed, leading us to consider the log version of the variable. As shown in column 4 of table 3, the co-efficient on log points per game is large, significant, and positive. Since taking the log of points per game does not significantly alter the effect, we simply use points per game in our model.

Table 3: Predicted Probability of Advancing from Midget to Junior based on Individual Performance Statistics

	(1)	(2)	(3)	(4)
	LPM0	LPM1	LPM2	LPM3
	LevelUp	LevelUp	LevelUp	LevelUp
Points per Game	3.229	0.344***		
	(3.594)	(0.039)		
Log(Points per Game)				0.346***
				(0.036)
Goals per Game	-2.951		0.275***	
	(3.59)		(0.073)	
Assists per Game	-2.833		0.396***	
	(3.594)		(0.058)	
Constant	-0.197***	-0.192***	-0.194***	0.176***
	(0.032)	(0.033)	(0.032)	(0.014)
Observations	900	900	900	900
R Squared	0.135	0.135	0.135	0.129

Standard Error in Parenthesis
***p<0.01, **p<0.05, *p<0.1
Note: This table presents the regression results of the linear probability model using individual performance statistics with robust standard errors.

To determine which other variables to include, we utilized a backwards stepwise regression, starting with all variables and eliminating variables based on whether the variable fit our model. We immediately removed rank as it does not constitute an individual performance metric but rather the outcome of performance. We then removed penalty infraction minutes as it is not an accurate metric for determining performance. Before finalizing our model, we explored different combinations of variables, including interaction terms, the standardization of variables, and other transformations. For example, we expected forwards to have more points than defencemen, leading us to interact Points per Game with Forward. However, the results were not helpful as we realized that while forwards average more goals, defencemen average more assists. Since both goals and assists are worth one point, the difference in points between forwards and defencemen is relatively low. Specifically, forwards average 0.17 additional points per game in comparison to defencemen. Overall, we find that regular linear variables work best in the context of our models.

In our final model, we include Points per Game as it is large and statistically significant in explaining the outcome variable. We also include Forward due to the abundance of forwards in our dataset and to estimate the effect of position. Lastly, Games Played is added to our specification to take into account the vast differences in the number of games played across players, across leagues.

C. Team Model

Once we have chosen our final model predicting advancement using individual performance statistics, we devise two separate methods to measure overall team skill, which we use to control for the team effect on a player's probability of advancement. The first of these measures is a "leave-out" average of a team's points per game, defined as:

$$LeaveOutPTS.GM_{in} = \frac{TeamPTS.GM_n - PTS.GM_i}{TeamCount_n - 1}$$

For a given player i, this variable measures the average points per game of all top 100 players on that player's team n, excluding the player in question. This variable is meant to capture the average skill of the team's other top players, to see if a team having highly skilled top players affects an individual player's chances of advancing from midget to junior. Ideally, we would split the teams by year to create this variable, which would more accurately estimate the skill of a player's team during the year they were eligible to level up. However, due to the limitations of the dataset, we are unable to do this. This is because the number of players on each team during each year that appear in the dataset is too low, with most teams having fewer than five players, and some having only one—resulting in a denominator of zero. Therefore, we instead measure this variable over the three-year period of the dataset, from 2016 to 2018. Once we have created this variable, we first regress LeaveOutPTS.GM on LevelUp:

$$LevelUp_{int} = \beta_0 + \beta_1 LeaveOutPTS.GM_{nt} + \varepsilon_i$$

The coefficient on LeaveOutPTS.GM shows a positive, statistically significant result (0.252, significant at the 1% level).

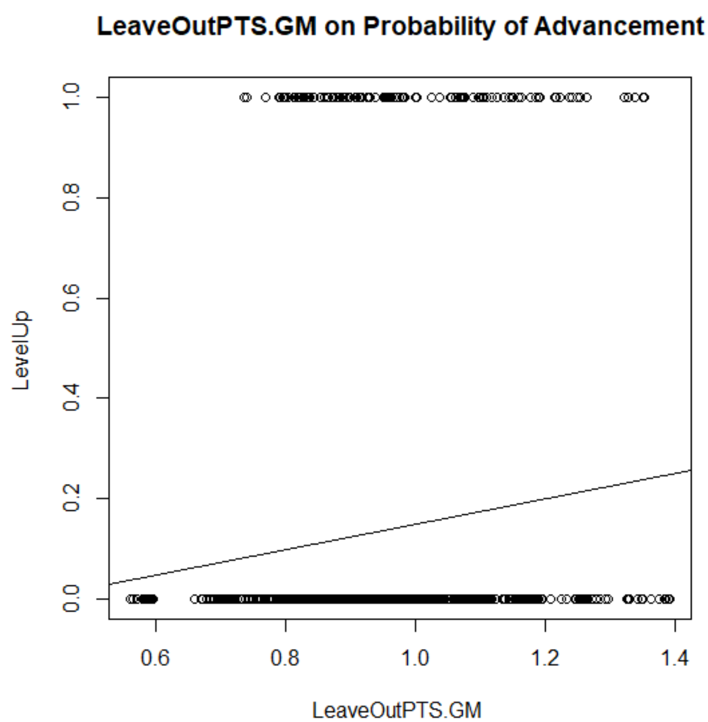

Figure 2. Scatterplot describing the relationship between LeaveOutPTS.GM and LevelUp.

We then perform our final individual model regression with LeaveOutPTS.GM included, estimated as:

$$LevelUp_{in} = \beta_0 + \beta_1 Points\ per\ game_i + \beta_2 Games\ Played_i + \beta_3 Forward + \beta_4 LeaveOutPTS.GM_{in} \quad (2)$$

Our second method to control for the team effect is to create a new variable, LeaveOutPlayers:

$$LeaveOutPlayers_{nt} = TeamYearCount_{nt} - 1$$

For any player on a given team n, LeaveOutPlayers represents the number of other players that appear in the dataset for that player's team during year t. Because the dataset represents the top 100 players from three midget leagues, we argue that the number of players in the dataset can be used to measure the overall skill of a team. In other words, we assume that the more top 100 ranked players on a team, the better that team.

Once again, we regress our new team measure on LevelUp:

$$LevelUp_{int}=\beta_0+\beta_1 LeaveOutPlayers_{nt}+\varepsilon_i$$

The coefficient on LeaveOutPlayers is positive, but very small and not statistically significant.

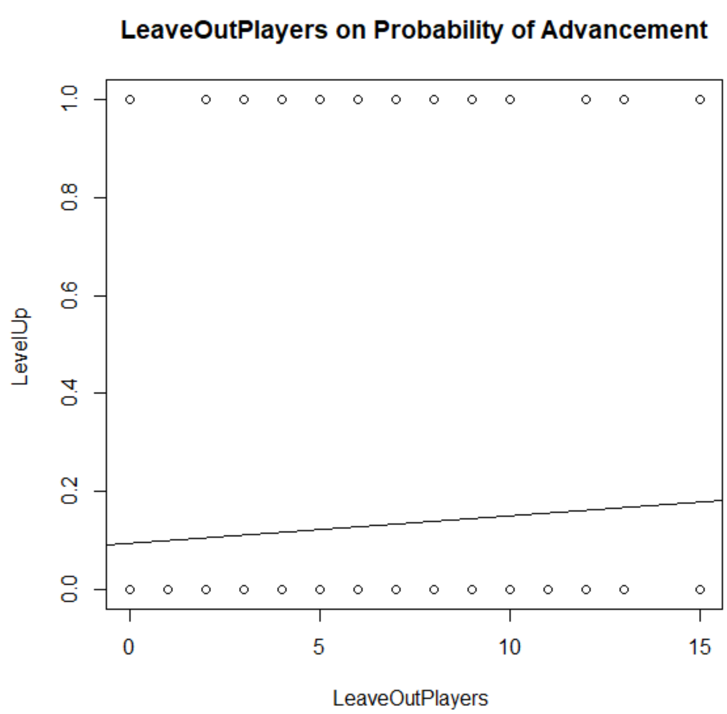

Figure 2. Scatterplot describing the relationship between LeaveOutPlayers and LevelUp.

We then perform our final individual model, once again including the relevant team variable:

$$LevelUp_{int}=\beta_0+\beta_1 Points\ per\ game_i+\beta_2 Games\ Played_i+\beta_3 Games\ Played_i+\beta_4 LeaveOutPlayers_{nt}+\varepsilon_i \quad (3)$$

V. Results

Table 4 shows the results of the linear probability model. In column 1, we see the results from the individual model. The coefficient on points per game is positive and large. This model predicts that higher points per game increases the probability of advancement. In particular, a one-point increase in points per game is associated with a 33.9% increase in the probability of advancing from midget to junior. The estimated coefficient of points per game is significantly different from 0 at the 1% level when using robust standard errors. Column 1 also shows that both games played and the dummy variable for forward are negative; however, they are not statistically significant.

Table 4: Advancing from Midget to Junior based on Individual Performance Statistics and Team Measures

	(1) LPM Individual LevelUp	(2) LPM Team 1 LevelUp	(3) LPM Team 2 LevelUp
Points per Game	0.339***	0.345***	0.337***
	(0.041)	(0.042)	(0.042)
Games Played	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.004)
Forward	-0.036	-0.039	-0.035
	(0.028)	(0.027)	(0.035)
LeaveOutPTS.GM		-0.052	
		(0.070)	
LeaveOutPlayers			0.001
			(0.003)
Constant	-0.024	0.035	-0.027
	(0.109)	(0.132)	(0.107)
Observations	900	900	900
R Squared	0.136	0.136	0.136

Standard Error in Parentheses
***p<0.01, **p<0.05, *p<0.1

Note: This table, represents results from our team and individual models, using robust standard errors.

Once we have established our individual model, we move on to the team component. We first perform our chosen linear probability model with LeaveOutPTS.GM included. The results are presented in Column 2 of Table 4. When we include the variables from the individual model, the coefficient on LeaveOutPTS.GM becomes negative, and is no longer significant at any level. Instead, Points per Game remains strongly significant with a positive value of 34.5%, similar to the individual model. The other variables from the individual model remain negative and statistically insignificant.

Next, we use our second measure of overall team skill, LeaveOutPlayers, by adding it to our chosen individual model. The results can be seen in column 3 of Table 4. These results are consistent with those of the individual performance model, as well as our leave-out average model, with Points per Game being the only significant variable at 33.7%. LeaveOutPlayers remains positive and insignificant, while Forward and Games Played remain negative and insignificant.

VI. Discussion

A. Individual Discussion

In our individual model, we look at the properties that determine the probability of advancing to the next level. We find that a unit change in points per game raises the probability of advancing from the midget to junior level by 33.9%. An explanation for this result is that higher per game performances as measured by points per game send a positive message to junior level programs that midget players can continue to play at a higher competitive level. In the context of our model, points per game is the only significant factor in determining advancement. The variable for games played and forward are insignificant, indicating that the two variables do not explain advancement to the next level.

Points per Game

Since points is simply the sum of goals and assists, points per game represents performance on a per game basis. As expected, an additional point per game is associated with a greater chance of "levelling up". In other words, the better one's individual performance, the higher the probability of advancement. As shown in figure 1, very few players average more than 2 points per game. In the context of the model, players that average more than 2 points per game have a 50% chance of advancement. In our dataset, only 2 players average more than 2.5 points per game; however, these players have more than a 65% chance of levelling up, whereas those players that average 1 point per game have approximately a 20% chance of advancement. This result illustrates the importance of points per game in predicting individual success, as defined by advancement from midget to junior. Figure 2 also shows predicted occurrences that are below 0, which is an issue for our model. To confirm our results, we test with a probit model as a robustness check.

Games Played

In our linear probability model, column 1 of table 4 shows that games played has a statistically insignificant effect on LevelUp. Meaning, the number of games played by an individual player does not impact the probability of advancing to the next level. We think that this might be due to most players playing a similar number of games. Standard deviation measures variation for a variable and a low standard deviation shows a value that is close to the mean. The variable Games Played has a fairly low standard deviation (0.03). Additionally, good players (higher points per game) might receive more playing time (higher games played), leading to multicollinearity. Multicollinearity increases standard error, possibly reducing significance.

Forward

The effect of whether a player is a forward or not on LevelUp produces a result that is not significantly different from 0. Meaning, being a forward is not a significant factor in explaining advancement. This confirms our prediction that the position of a player has no predictive value on advancement to the junior level. Since all players contribute to the success of a team, a position is inadequate in describing advancement.

However, out of 122 players that advanced in our dataset, only 16 were defencemen. Since defencemen are less likely to score goals, points per game may not describe their performance. To further test robustness across positions, we run our model with only forwards and only defencemen. In column 1 of table 5, we see that with only forwards, points per game is still statistically significant at the 1% level with a similar coefficient (33.4%) to our original model. In column 2, we see that with only defencemen, points per game is statistically significant at the 1% level. However, we see a counter-intuitive result, where a one-point increase in points per game is associated with a 41% increase in probability of advancement. We suspected omitted variable bias and ran the model again with the inclusion of penalty infraction minutes. In column 3, we see that penalty infraction minutes and games played are insignificant, while the point estimate for points per game is large, positive, and significant.

Table 5: Predicted Probabilities of Advancing from Midget to Junior based on Individual Performance and Position

	(1)	(2)	(3)
	LPM Forward LevelUp	LPM Defence LevelUp	LPM Defence LevelUp
Points per Game	0.335***	0.413***	0.414**
	(0.042)	(0.155)	(0.161)
Games Played	-0.005	0.002	0.003
	(0.003)	(0.006)	(0.003)
PIMS			-0.001
			(0.001)
Constant	-0.019	-0.297	-0.279
	(0.117)	(0.301)	(0.325)
Observations	744	143	143
R Squared	0.147	0.07	0.068

Standard Error in Parentheses
***p<0.01, **p<0.05, *p<0.1
Note: This table presents regression results for the linear probability model using only forwards, or only defencemen, with robust standard errors.

The relationship between points per game and advancement is summarized in figure 4, where the black line represents LevelUp regressed on Points per Game for forwards, and the red line for defencemen. From this figure, we see that the effect of points per game on advancement is greater for defencemen as the red line is much steeper. Defencemen average less points per game; however, an increase in points per game leads to a greater increase in probability of advancement for defencemen.

Only a few defencemen advanced in the three years of our dataset. An explanation for this result is that a higher points per game might be a positive signal to junior teams during the selection process as it allows the defencemen to differentiate themselves from competition. Additional study is required to explore this result.

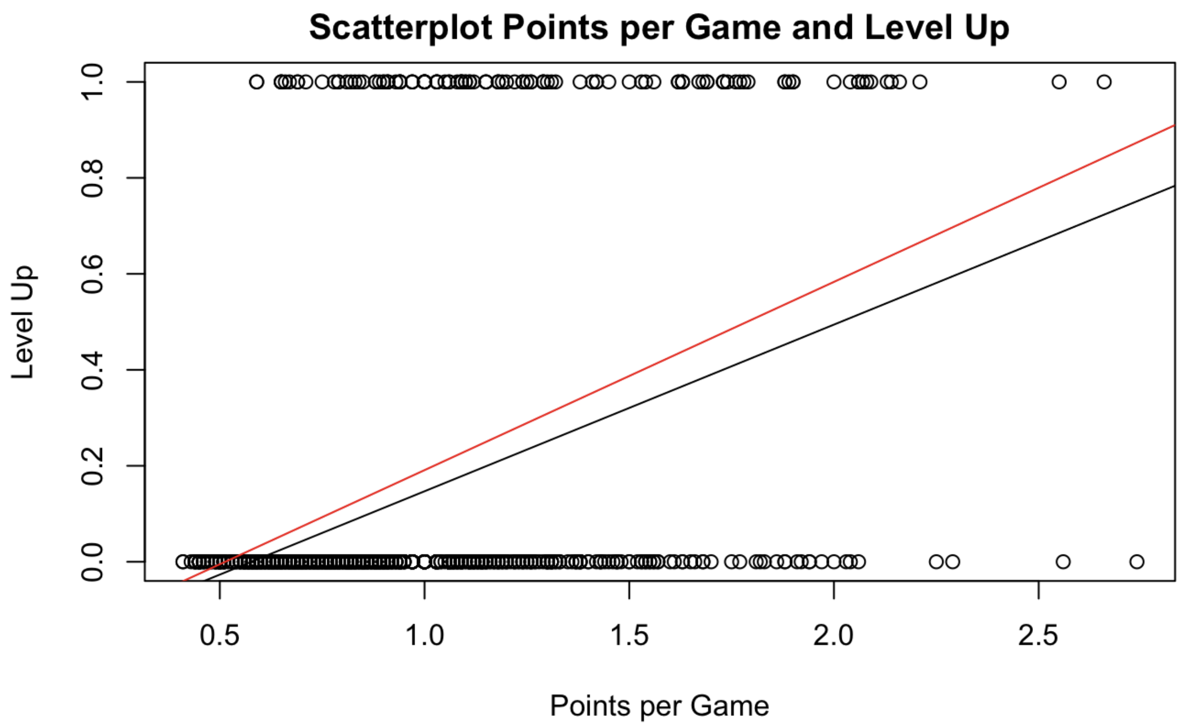

Figure 4. Scatterplot plotting LevelUp regressed on Points per Game for only forwards (black line) and only defencemen (red line).

Alternate Specifications:

Instead of goals per game and assists per game, we decided to use points per game to avoid multicollinearity as points are the sum of goals plus assists. To test the alternate specification, we replace points per game with goals per game and assists per game. This model is shown by the following specification:

$$LevelUp_i = \beta_0 + \beta_1 Goalspergame_i + \beta_2 Assistspergame_i + \beta_3 GamesPlayed_i + \beta_4 Forward_i + \varepsilon_i \quad (4)$$

As shown in column 1 of table 5, goals and assists per game are properties that determine advancement from midget to junior. Even with an alternate specification, advancement seems to be linked to per game individual performance metrics (goals per game plus assists per game). The relationship between the variables also appears to be similar as it is positive, large, and significant at the 1% level, confirming the importance of individual performance statistics as a property describing advancement from midget to junior.

Table 6: Predicted Probability of Advancing based on Goals and Assists per Game

	(1) LPM Alternate LevelUp
Goals per Game	0.279***
	(0.085)
Assists per Game	0.378***
	(0.061)
Games Played	-0.004
	(0.003)
Forward	-0.022
	(0.031)
Constant	0.030
	(0.107)
Observations	900
R Squared	0.137

Standard Error in Parenthesis
***p<0.01, **p<0.05, *p<0.1
Note: This table presents regression results using an alternate specification. We use robust standard errors.

B. Team Discussion

Team Points per Game

LeaveOutPTS.GM, our measure for the average points per game of a team's other top players excluding player *i*, is positively and significantly correlated with LevelUp in a single explanatory variable linear probability model, with one additional point per game among a team's top players on average being associated with a 25% increase in the probability of another top 100 player on that team's advancement to junior leagues. However, once we include the individual performance statistics from our individual model, this effect entirely disappears, returning a negative and insignificant coefficient on LeaveOutPTS.GM as seen in column 2 of table 4. Points per game remains strongly significant, with a similar coefficient of 34.5%. From this result, we conclude that our measure of team points per game has no significant effect on an individual player's advancement, and it doesn't change the significance of individual performance.

Number of Top 100 Players

Our measure of the number of top 100 players on a team excluding player *i*, LeaveOutPlayers, is not significantly correlated with the probability of advancement. It is unsurprising, then, that it does not significantly alter the results of the individual model. As seen in column 3 of table 4, LeaveOutPlayers retains its positive and statistically insignificant coefficient value, while Points per Game remains the dominant explanatory variable, with a coefficient of 33.7%

Effect of Team Skill

Overall, neither of our measures of team skill returned significant results. Rather, the most important factor remains our performance measure of Points per Game. It is also important to note that the value of the coefficient on Points per Game remained almost identical in all three models, only ranging from 33.7% to 34.5%. This finding implies that neither of the two team variables represent significant omitted variables in the individual model, since they do not take away any predictive power from Points per Game. Ultimately, using our measures of team skill, our results appear to indicate that a player's team does not have a strong impact when predicting whether that player will appear in the Junior dataset, and does not affect the strong, positive correlation of individual performance as measured by Points per Game. To verify these results, we perform a number of robustness checks.

VII. Robustness

A. Fixed Effects

We also compare the results of our central model with Team and Year fixed effects. Team fixed effects give us another measure of the team effect, to compare with our main models. Year fixed effects are included to account for potential omitted variable bias, as some years may have seen more players drafted than others. As presented in Table 7, The main results do not change, and the Team fixed effects column in particular is consistent with our team and individual models, returning a coefficient value of 35.32% on Points per Game. For the Year and Team + Year models, this value actually increases slightly to 38.1%, while remaining strongly significant at the 1% level. Interestingly, our position dummy, Forward, becomes significant at the 10% level in only the Team + Year fixed effects model, while maintaining its negative coefficient across all three. Overall, the fixed effect models appear to be consistent with our main findings, while pointing to some potential omitted variable bias with respect to the significance of the position variable.

Table 7: Advancing from Midget to Junior: Individual Model with Team and Year Fixed Effects

	(1)	(2)	(3)	(4)
	LPM Individual	Team FE	Year FE	Team & Year FE
	LevelUp	LevelUp	LevelUp	LevelUp
Points per Game	0.339***	0.354***	0.364***	0.381***
	(0.041)	(0.044)	(0.041)	(0.044)
Games Played	-0.004	-0.002	-0.001	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Forward	-0.036	-0.045	-0.04	-0.051*
	(0.028)	(0.029)	(0.027)	(0.029)
Fixed Effects:				
Team	NO	YES	NO	YES
Year	NO	NO	YES	YES
Constant	-0.024	-0.066	-0.064	-0.071
	(0.109)	(0.165)	(0.107)	(0.165)
Observations	900	900	900	900
R Squared	0.136	0.133	0.17	0.167

Standard Error in Parentheses

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

Note: This table represents results from our individual model with team and year fixed effects, using robust standard errors.

B. Alternate Model: Probit

As shown in Figure 1, there are a small number of occurrences in the individual model of LevelUp on Points per Game where the predicted values are below 0. To account for this, in our final robustness check, we perform a probit model regression of our three main models: Individual model (1), Team LeaveOutPTS.GM model (2), and Team LeaveOutPlayers model (3). Since Probit is non-linear, we cannot directly compare the coefficients. However, Points per game remains positive and extremely significant and the team variables and position remain insignificant, while Games Played becomes significant at the 5% level. Once again, the Points per Game variable remains the most important factor, maintaining its significance at the 1% level for all three models. From here, we evaluate the marginal effects. Table 9 shows the marginal effects of just the individual Probit model. As shown in table 9, a one-point increase in points per game is associated with a 24.8% increase in points per game. The coefficient for Forward remains negative and insignificant. Games Played is negative and significant at the 5% level. The marginal effects on these variables do not change significantly when including the team measures. Overall, these results confirm that points per game is a significant determinant in predicting advancement; however, in a non-linear model, it has a smaller effect.

Table 8: Marginal Effects of Probit Model using Individual Variables

	dF/dx	Std. Err.	z	P>|z|
Points per Game	0.248***	0.028	8.806	0.000
Games Played	-0.005**	0.002	-2.121	0.034
Forward	-0.037	0.033	-1.137	0.256

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

Note: The table represents the marginal effects of the Probit model with variables from the individual model.

C. Limitations

In the previous section, we discuss and confirm the robustness of our findings; however, our data construction, empirical model, and topic of study indicates several limitations. Firstly, we do not control for eligibility constraints such as age. Midget players cannot advance to the junior level until the age of 16. It is recommended that a player remain at the midget level until the age of 17. Our dataset did not include age related information. Future research should incorporate age-related eligibility restrictions as this would lead to more robust results. Additionally, our junior dataset only includes the top 100 ranked players across several Western Canada junior leagues. Because of this, a midget player may advance and not appear in our dataset. There are also various characteristics that are excluded from our model. There are factors, we cannot control for: individual characteristics such as overall effectiveness, physical attributes, mental capacity, upbringing, and ambition. This may lead to omitted variable bias. To improve this model, we could include plus/minus statistics, physical fitness metrics, and background information on a player's environment.

Lastly, there are limitations to our team metrics due to the nature of the dataset. Because this dataset does not include the entire team rosters, but rather the top 100 players from different leagues, the information on overall team performance is incomplete. Because of this, our leave-out team points per game variable measures not the overall average points per game of players on a team, but instead the average points per game of only the team's top players. While intuitively a team having highly skilled top players may affect an individual player's probability of advancement, the other players on the team also contribute to a team's overall skill level, which this dataset does not capture.

Additionally, because not all players on each team are included in this dataset, the number of players on each team during a given year that appear in the data is fairly low. This makes it difficult to group the data by team and year, which would more accurately represent the skill of teams during the year individual players were eligible to advance to junior leagues. Our LeaveOutPTS.GM is not separated by year due to this low number of observations, and we are also unable to accurately measure Team*Year fixed effects because of this. However, we argue that LeaveOutPlayers, which measures the number of top 100 players per team per year, provides a fairly accurate representation of team skill by year, and it returns almost identical results to our other team measures when included in our final model, which bolsters their validity.

VIII. Conclusion

Current research recognizes the importance of individual performance statistics in predicting player success. However, it remains unclear whether team performance influences individual success. Additionally, there is a lack of research on the individual and team related properties that determine success for midget level hockey players in Western Canada. We use a linear probability model to explore the effect of individual performance (points per game), position, and the number of games played on advancement from the midget to junior level. We also test for the team effect using measures

Although limitations exist, namely the lack of eligibility information and the risk of omitted variable bias, we demonstrate the robustness of our results with an additional specification that captures individual performance, various team skill and team effect models, and an alternate model: probit.

To summarize these results:

1. An additional point per game increases the probability of advancing from midget to junior level by 33.9% when using robust standard errors. Points are defined as goals plus assists. If we replace points per game with goals per game and assists per game, we still see significant results, indicating the importance of individual performance statistics.
2. Points per game is a property that determines advancement, regardless of position. Testing the model with only defencemen indicates that an additional point per game has a greater effect on advancement for defencemen than forwards. A unit change in points per game is associated with an 41% increase in probability of advancement for defencemen. Additional study is required to rule out omitted variable bias.

3. All three team measures, including team fixed effects, left points per game as the main significant variable, maintaining a positive and significant coefficient with a value between 34-38%. Both of our measures of team skill did not return significant results, leading to the conclusion that a player's performance is the most important factor regardless of their team.

We have demonstrated that individual performance statistics (points per game or goals per game and assists per game) is an important factor in predicting advancement from the midget to junior level. Regardless of position, team, or numbers of games played, players that have higher individual performance statistics are more likely to advance to the next level. Our model allows us to predict whether a player advances given their individual performance statistics. We find that players that wish to advance should focus on individual properties that they can control: how many points they score per game. We find that a higher points per game sends a positive signal to junior teams, providing evidence that a player can contribute at a higher level. Although we expect the importance of individual performance statistics to hold, it is necessary to explore other player characteristics such as upbringing, physical traits, experience, and mental capacity. In doing so, we may uncover omitted variable bias or additional significant properties. For further investigation, we suggest including eligibility years, quantifying physical attributes, and taking a closer look at defensive players. Additionally, to further examine how a player's team may affect their chances of advancement, we recommend using a dataset that includes the full roster of players on each team, to improve the accuracy of the team performance metrics, as well as increase the number of observations per team per year.

References

He, E. (2020). Top Ranking Hockey Player Performance Statistics in Predicting Top Ranking Appearances Between Minor League Hockey Divisions. *University of British Columbia*.

Hockey Canada. (n.d.). Canadian development model. Retrieved from https://www.hockeycanada.ca/en-ca/hockey-programs/players/essentials/downloads.

Idson, T., & Kahane, L. (2000). Team effects on compensation: An application to salary determination in the national hockey league. *Economic Inquiry, 38*(2), 345-357.

Riley, S. N. (2017). Investigating the multivariate nature of NHL player performance with structural equation modeling. *PloS One, 12*(9), e0184346-e0184346.

Szporer, R. (2021). Current NHL Players by Country: On Top of Their Game and the World. Retrieved from https://thehockeywriters.com/current-nhl-players-by-country/