NHL Shot Statistics, Study of the Effect of Defending Ice Time on Expected Goals

Connor Raney

2024-04-30

Table of contents

1. Introduction

For my final project, I will be studying the topic of hockey data, more specifically, shot data. While there is a wide variety of publicly accessible datasets for hockey, some of the most interesting to me personally is shot data. Modern hockey analytics rely on xGoals (expected goals), a measure of shot quality and goal probability, derived from shot data. Publically accessible datasets have upwards of 150 to 200 variables collected for each shot, and many of these variables are used in the complicated calculations of xGoals. Using a model, the probability of each shot being a goal is calculated using factors such as the distance from the net, angle of the shot, type of shot, events that happened before the shot, location coordinates on the ice, and many more variables that are contained in this dataset.

Just for reference, the mean xGoals per shot measured over the last 2 seasons is ~0.0727 xGoals, and teams average somewhere around 30 shots per game[1] for each team, which works out to around 2 to 3 goals per team every game. With this small value, understanding the small edges and effects that can impact the way we understand the game becomes very important. Even by finding a very small impact that a variable has on the xGoals, teams can change

---

[1]https://www.statmuse.com/nhl/ask/nhl-team-average-shots-on-goal-per-game-2024

the way they structure their play and improve their defensive game by being able to measure something such as the effect that the defending team's ice time has on xGoals. Now, by adding up these probabilities of a team's shots during a game, you can calculate a team's expected goals and essentially measure their offensive performance in addition to their defensive performance by measuring how many expected goals they gave up. Through these statistics, you can see a sort of story about the game and what happened. Did your team get unlucky and score way below expected, or was it just not your night? Or did you get lucky, and did you statistically not 'deserve' to win the game? How well did your goalie perform? A question like this can be answered by taking (expected goals - actual goals) and getting an amount above or below expected to measure a goalie's performance. Each season, the 'best' goalie is typically the one who has the most goals saved above expected, or GSAAx. As you can see, almost all aspects of hockey statistics use the expected model as a frame of reference, which is why I chose to study something I believe influences it.

My research question is: Does the amount of time the defensive team (defensiveTeamAverageTimeOnIce) has collectively been on the ice affect the other (offensive) team's generation of scoring chances, measured by expected goals (xGoals)? My independent variable is defensiveTeamAverageTimeOnIce, which measures across all skaters on the ice at the time of shot taken by the offensive team, how long they have each, on average, been on the ice for. My dependent variable is xGoals, which I previously touched on and explained in detail. These are the main variables I will be focusing on during my analysis, but I will have different treatments and confounders throughout my analysis process with different models.

Of course, it is logical that if players are out on the ice for a very long time, they are more likely to break down defensively and give up better chances to the offensive team, but *how much* does this actually impact the quality of the chances they give up? With all of this context, it is important for us to show a measurable causal relationship rather than just a correlation. As I said before, this relationship can have implications for the ways that teams choose to deploy their players on the ice and how much time coaches choose to have players on the ice for. Having said that, there could be numerous strategic implications, contingent on the impact this relationship has on the chances that teams concede.

2. Data and Hypothesis

The data set that I have chosen to work with is Moneypuck.com's shot data dataset, which contains data with upwards of 1.7 million shots from the 2007-2008 to 2023-2024 seasons. Here is an explanation of the dataset from Moneypuck, "There are 124 attributes for each shot, including everything from the player and goalie involved in the shot to angles, distances, what happened before the shot, and how long players had been on the ice when the shot was taken. Each shot also has model scores for its probability of being a goal (xGoals) as well as other models such as the chance there will be a rebound after the shot, the probability the shot will miss the net, and whether the goalie will freeze the puck after the shot. The data has been collected from several sources, including the NHL and ESPN. A good amount of data cleaning has also been done on the data. Arena-adjusted shot coordinates and distances are

also calculated in the dataset using the strategy War-On-Ice used from the method proposed by Schuckers and Curros.[2]"" As you can see, our data is observational and comes from a variety of sources, compiled into one massive dataset with many variables that can be used for analysis, which is one of the major strengths of this dataset. Moneypuck is also very straight forward with how their xGoals model works, with "factors such as the distance from the net, angle of the shot, type of shot, and what happened before the shot are key factors in the model," and you can find more information about the model on their website.[3]

Overall, our data is very strong, and I believe it will be able to answer our research question with confidence. Our data is on exactly the league, teams, and players that we want to study, has already been cleaned for us, and therefore should be easy to work with. In our data, the outcome is of course the measurement of xGoals (xGoal in the dataset), and the treatment is the defensiveTeamAverageTimeOnIce. These will be the variables for the regression models that I will later run. In addition, I will run a propensity score matching model to further try to measure the effect of defensiveTeamAverageTimeOnIce on xGoals, with the treatment variable being values of defensiveTeamAverageTimeOnIce above the 3rd quartile, or 75th percentile of the data. I will then also be using the covariates xCordAdjusted (absolute value of the x coordinate where the shot was taken), yCordAdjusted (absolute value of the y coordinate), shotType (slap, snap, wrist, etc.), shotRebound (whether the shot was a rebound or not, rebounds have a much higher chance of being a goal than other types of shots), speedFromLastEvent (if events happen very close to a shot being taken, e.g., a pass or a hit/takeaway of the puck, which many times will lead to more dangerous chances), and shotAngleAdjusted (the absolute value of the shot angle, certain angle have better chances of going in).
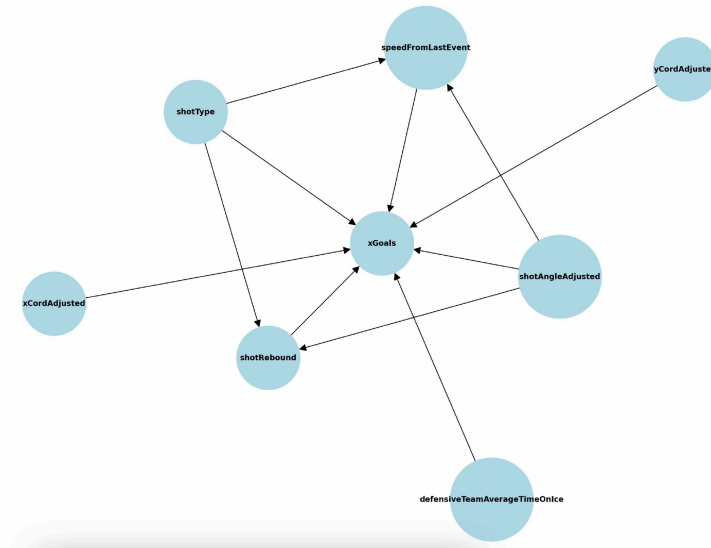

Figure 1: DAG for our X and Y variables, and our confounders & covariates

---

[2]Information and download for the dataset can be found at: https://moneypuck.com/data.htm

[3]Information and download for the xGoals model can be found at: https://moneypuck.com/about.htm

Now, what is my hypothesis here, and what do I think I will find? Well, with years and years of both playing and watching hockey, I think that there will most certainly be a trend here, with more defensive ice time equaling more xGoals in the shots that they will give up to the offensive team. There are not that many sources that have investigated this exact relationship, as it is something more logical in nature, knowing that as your players get more tired from being out on the ice, the risk they have a breakdown, make mistakes, or are just slower is, of course, higher. However, I am curious as to just how big of an impact this will have. While I do not think it will have an earth-shatteringly *huge* impact, I believe that you will see a pretty significant trend in xGoals rising as the amount of time spent on the ice increases. I believe that the causal effect will have somewhere between a 10%-20% effect on the xGoals, after looking at a certain cutoff point, which I will do in my Propensity Score Matching model, with the treatment variable being the 75th percentile of the defensiveTeamAverageTimeOnIce variable. I'll also try to control for some of the main covariates and confounders that I believe will have an effect on the xGoals.

3. Explore the Data

In exploring our data, there were a few things that were very interesting to me. First of all, let's start with the distribution of the xGoals variable over the defendingTeamAverageTimeOnIce:
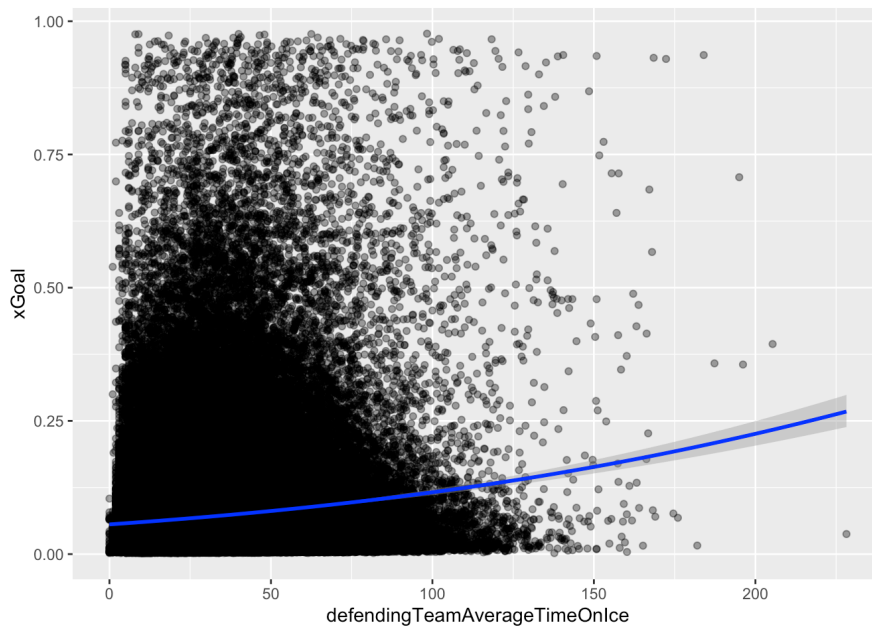

Figure 2: defensiveTeamAverageTimeOnIce over xGoals (all shots)

The shape of the distribution to me was very interesting, as while there is a general upward trend, there are many outliers and small values clustered at low amounts after about 100

seconds of ice time. This general trend of the clusters of data was unexpected, and I am still curious as to why this is the case, but we can still see here that there are many high values as well as ice time increases. Here, this plot shows us a logistic regression trendline with a shallow slope, so while we have a positive relationship, it is somewhat weak in this visualization.

In addition to the distribution, I also found some interesting means and medians of the data, with the average xGoals being 0.07271334 per shot, which was a little lower than I thought. I figured it would be a little closer to 0.1 per shot. Also relating back to the distribution, I expected a more even distribution of xGoals rising over time, but there is an odd pattern of rising and then downward sloping with high outliers, which I did not expect, paired with this interesting mean that was lower than I expected.

Finally, I expected the data to have a variable for the situation, but I realized that it did not, so I had to create this myself. I wanted to see if there was a major difference in the role of defensive ice time when comparing even strength (5v5) play, to non-even strength (5v4 with a power play, 5v3, or 6v5 with a pulled goalie), so I created filters for all three of these, with even strength shots, non-even strength shots, and then all shots. The top number is the intercept, with the bottom number in the table being the coefficient for the effect the defendingTeamAverageTimeOnIce variable had on each situation. This was surprising, as while they were statistically significant, it looks like the non-even strength (most likely powerplay situations) had a much bigger impact on the xGoals than at even strength. While this is the only time that I situationally broke down the data and explored it, it would be interesting to go further into the xGoals calculation with even strength vs. non-even strength situations in another analysis.

DefendingTeamAllSituations	DefendingTeamEvenStrength	DefendingTeamNonEvenStrength
-2.82896***	-2.83252***	-2.44598***
(0.01689)	(0.02136)	(0.02853)
0.00798***	0.00356***	0.00875***
(0.00041)	(0.00056)	(0.00059)

Figure 3: Linear Regressions for Each Situation (with defendingTeamAverageTimeOnIce)

There were also a few things that I observed that needed to be changed before getting into my propensity score model. First of all, the shotType variable needed to be changed to factor, as this was not done before. Also, while the xCordAdjusted variable was in fact adjusted to only the absolute, positive, values, the yCordAdjusted variable was not, so I had to use abs() to make that change.

I also ran descriptive statistics on the numerical variables that I was going to use as confounders and covariates for the Propensity Score Matching model. It was very interesting to see the number of shots being rebounds as low as it was. To see only about 7% of shots being rebounds was surprising to me, as I figured this value would be somewhere closer to 20-30%, especially with the possibility of multiple rebounds coming from one scoring chance or shot.

Besides that, none of the results here really jumped out to me as crazy. Here is the output:

```
$xCordAdjusted
   Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
   0.00   49.00   65.00  61.88   78.00 100.00

$yCordAdjusted
   Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
   0.00    6.00   14.00  15.84   25.00  43.00

$shotRebound
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00000 0.00000 0.00000 0.07156 0.00000 1.00000

$speedFromLastEvent
   Min. 1st Qu.  Median   Mean 3rd Qu.    Max.
  0.000   2.113   4.800  8.467  10.617 193.747

$shotAngleAdjusted
   Min. 1st Qu.  Median   Mean 3rd Qu.   Max.
   0.00   16.39   30.96  33.03   46.40  88.45

$xGoal
     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 0.001537 0.014789 0.036877 0.072713 0.087609 0.976651
```

Figure 4: Descriptive Statistics for Numerical Covariates/ Confounders

4. Our Scenario - Propensity Score Matching

For PSM, I am going to use six other variables to try and account for some of the other effects that contribute towards xGoals: 1. Shot Location (x) –> xCordAdjusted (abs value) 2. Shot Location (y) –> yCordAdjusted (abs value) 3. Shot Type –> shotType 4. Shot Rebound –> shotRebound 5. Speed From Previous Event –> speedFromLastEvent 6. Shot Angle –> shotAngleAdjusted (for simplicity sake we will use the abs value). At first, I was using shotDistance as well, but this was too closely related to the xCord Adjusted, so I got rid of it. These should give us a good baseline to work off of with our propensity score matching model for our analysis. At first, I also had the median defensiveTeamAverageTimeOnIce as the treatment, but the matched set was essentially the same size as the normal dataset, so we did not get any significant results. Now, for this PSM, I am using the 75th percentile as the treatment, which comes out to 44 seconds of defensiveTeamAverageTimeOnIce, and this matched dataset is around half the size of the normal dataset, which includes all shots of every situation. Here is the balance and results of our treatment and control groups before and after matching:

6

	xCordAdjusted	yCordAdjusted	shotType	shotRebound	speedFromLastEvent	shotAngleAdjusted
(Intercept)	61.411***	16.264***	6.330	0.068***	8.737***	33.245***
	(0.045)	(0.027)		(0.001)	(0.026)	(0.051)
treatment	1.837***	−1.680***	−0.086	0.015***	−1.061***	−0.856***
	(0.090)	(0.054)		(0.001)	(0.052)	(0.100)
Num.Obs.	238304	238304	238304	238304	238304	238304
R2	0.002	0.004		0.001	0.002	0.000
R2 Adj.	0.002	0.004		0.001	0.002	0.000
AIC	2081528.1	1841363.0		29961.4	1822386.7	2134197.1
BIC	2081559.2	1841394.2		29992.5	1822417.8	2134228.2
Log.Lik.	−1040761.043	−920678.525		−14977.688	−911190.353	−1067095.531
RMSE	19.08	11.53		0.26	11.08	21.30
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001						

Figure 5: Raw differences between treatment and control group

	xCordAdjusted	yCordAdjusted	shotType	shotRebound	speedFromLastEvent	shotAngleAdjusted
(Intercept)	63.337***	14.488***	6.236	0.084***	7.877***	32.321***
	(0.075)	(0.045)		(0.001)	(0.042)	(0.088)
treatment	−0.089	0.096	0.008	−0.001	−0.201***	0.068
	(0.106)	(0.064)		(0.002)	(0.059)	(0.125)
Num.Obs.	121062	121062	121062	121062	121062	121062
R2	0.000	0.000		0.000	0.000	0.000
R2 Adj.	0.000	0.000		0.000	0.000	0.000
AIC	1050149.1	926600.7		32099.5	908012.9	1088255.5
BIC	1050178.2	926629.8		32128.7	908042.1	1088284.6
Log.Lik.	−525071.567	−463297.346		−16046.769	−454003.473	−544124.757
RMSE	18.51	11.11		0.28	10.29	21.66
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001						

Figure 6: Differences between treatment and control group, Matched Sample

Before matching, we do not have a great balance, as the distribution of covariates is not very similar across the treatment and control groups, with significant p-values for every variable besides shotType, and there are big differences in the coefficients for each covariate. After matching, we have a much better balance between the treatment and control groups, with all the p-values now not significant besides speedFromLastEvent, but with the speedFrom-LastEvent coefficient being much lower than it was before matcbing. Many of the imbalances and differences were taken care of with matching, and overall we have a better balance between our treatment and control groups now having done matching. Now, we can move onto our Propensity Score Matching results. Here are the effects estimated with our original observational data, and then the effects estimated using our PSM sample:

	Naive	Naive + Controls		Effects	Effects + Controls
(Intercept)	0.069***	-0.004*	(Intercept)	0.077***	-0.015***
	(0.000)	(0.002)		(0.000)	(0.003)
treatment	0.016***	0.009***	treatment	0.008***	0.008***
	(0.000)	(0.000)		(0.001)	(0.001)
xCordAdjusted		0.002***	xCordAdjusted		0.002***
		(0.000)			(0.000)
yCordAdjusted		-0.001***	yCordAdjusted		-0.001***
		(0.000)			(0.000)
shotTypeBACK		-0.022***	shotTypeBACK		-0.023***
		(0.002)			(0.003)
shotTypeDEFL		-0.012***	shotTypeDEFL		-0.017***
		(0.002)			(0.003)
shotTypeSLAP		-0.012***	shotTypeSLAP		-0.012***
		(0.002)			(0.003)
shotTypeSNAP		-0.009***	shotTypeSNAP		-0.009**
		(0.002)			(0.003)
shotTypeTIP		-0.018***	shotTypeTIP		-0.023***
		(0.002)			(0.003)
shotTypeWRAP		-0.037***	shotTypeWRAP		-0.039***
		(0.003)			(0.004)
shotTypeWRIST		-0.014***	shotTypeWRIST		-0.013***
		(0.002)			(0.003)
shotRebound		0.110***	shotRebound		0.107***
		(0.001)			(0.001)
speedFromLastEvent		0.001***	speedFromLastEvent		0.001***
		(0.000)			(0.000)
shotAngleAdjusted		-0.001***	shotAngleAdjusted		-0.001***
		(0.000)			(0.000)
Num.Obs.	238304	238304	Num.Obs.	121062	121062
R2	0.005	0.338	R2	0.001	0.301
R2 Adj.	0.005	0.338	R2 Adj.	0.001	0.301
AIC	-410737.7	-507925.5	AIC	-186609.8	-229765.2
BIC	-410706.5	-507769.8	BIC	-186580.7	-229619.6
Log.Lik.	205371.846	253977.772	Log.Lik.	93307.891	114897.576
RMSE	0.10	0.08	RMSE	0.11	0.09

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Naive Models

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Effects w/PSM

Figure 7: Estimate with and without Propensity Score Matching

Here, with our Propensity Score Matching results, we can see a very measurable effect of the treatment variable here, coming out to an 11% increase in xGoals for the treatment group, which in my opinion, is a very significant effect. We can see that the result is also statistically significant (p <.001), which is good. Also, the treatment coefficient decreasing from 0.016 to 0.009 in the naive model with controls, and then remaining stable at 0.008 in the PSM model with and without controls is also a very good sign, meaning that the treatment effect stays robust even across different approaches to modeling and measuring our effect. After the PSM, the covariates also remain significant, and our treatment also remains consistent, showing us that the matching process successfully accounted for confounding variables in an effective manner. Also, in both of our models, the R^2 increases from 0.005 to 0.338 in the naive model with control, and we can see a similar trend with 0.001 to 0.301 in the PSM model with controls. This shows us that with our covariates, the model is a much better fit overall and explains a much larger percentage of the variance in xGoals. Overall, our PSM model was very effective in continuing to show us the trend in xGoals increasing with defensiveTeamAverageTimeOnIce, and our results are consistent across naive models when controlling for covariate effects, as well as with our Propensity Score Matching model.

5. Conclusions

As a reminder, our ultimate question we wanted to answer was: Does the amount of time the defensive team (defensiveTeamAverageTimeOnIce) has collectively been on the ice affect the other (offensive) team's generation of scoring chances, measured by expected goals (xGoals)? When looking at our naive models with effects, in addition to our PSM models with and without effects, paired with our effective matching process, we can clearly see that defensiveTeamAverageTimeOnIce does affect the offensive team's generation of scoring chances measured by expected goals. We can see around an 11% increase, or 0.008 xGoals, in expected goals for the shots treated above the 75th percentile of defensiveTeamAverageTimeOnIce, or 44 seconds, which is a very significant increase when you consider the fact that the average xGoals per shot is only 0.0727. In a sport like hockey where small edges and probabilities matter and add up over time, approaching these results from a management and coaching standpoint is important and shows very directly how you are taking risks by leaving your players out on the ice for very long shifts, 44 seconds and up. The average defensiveTeamAverageTimeOnIce across all shots is ~32 seconds, so a 44 second long shift can be classified as a fairly long shift, being in the 75th percentile. From a tactical standpoint, this reinforces why short shift times, typically around 30 seconds, are preferred and may even prove why more emphasis should be placed on ensuring time on ice is limited to very short periods of time to ensure a more sound defensive game.

Overall, I learned a lot from this process by going through the data on my own and really figuring out how I wanted to measure this effect, as there are, of course, countless different models and ways to do this. I also learned more about the process of propensity score matching, with the importance of having a good subset of data after the matching process, and how much that matters for your end result and analysis using this matching model. The most trouble I

had during the course of this process was finding the size of the dataset that gave me decent results, while of course being able to run on my M1 Mac. While I wish I could have used the whole 1.7 million shot dataset, I was unable to, and I had to resort to only using ~200k shots, with 2 seasons, instead of 10+ seasons as I wanted to do originally. However, if I had access to more computing power and could have conducted the calculations on a larger dataset on a cluster or cloud machine, I believe my confidence in the results would have increased significantly, thereby raising the issue of external validity. We are not sure whether this exact pattern will appear across all of the other ~1.5 million shots, and it is obviously important to use all the data you have access to in order to limit the questions of external validity. When we think about internal validity as well, since there are a ton of confounding variables that can affect the relationship and measurement of xGoals, we did not account for all of them, so we cannot be sure of how much the defensiveTeamAverageTimeOnIce variable is really affecting the xGoals. Similarly, one of the main lessons I took away from this was that you will not always be able to find the *exact* answer that you badly want, and you can only get as close as your models will take you. Although, I would love to further analyze this topic, use some more computing power, refine my models, and get a better answer in the future. Ultimately, I had a great time working on this project, especially since it was on a topic that I am very passionate about. It was very nice to combine the knowledge we've gained this semester with a passion of mine, and to be proud of what we produced at the end of the semester.