# Small but Mighty

## Examining the Utility of Microstatistics in Modeling Ice Hockey

Matthew Palmer

A Senior Thesis submitted in partial fulfillment
of the requirements for graduation
in the Honors Program
Liberty University
Spring 2023

Acceptance of Senior Honors Thesis

This Senior Honors Thesis is accepted in partial fulfillment of the requirements for graduation from the Honors Program of Liberty University.

______________________________
Timothy Van Voorhis, Ph.D.
Thesis Chair

______________________________
Michael Gibson, Ph.D.
Committee Member

______________________________
David Schweitzer, Ph.D.
Assistant Honors Director

______________________________
Date

## Abstract

As research into hockey analytics continues, an increasing number of metrics are being introduced into the knowledge base of the field, creating a need to determine whether various stats are useful or simply add noise to the discussion. This paper examines microstatistics – manually tracked metrics which go beyond the NHL's publicly released stats – both through the lens of meta-analytics (which attempt to objectively assess how useful a metric is) and modeling game probabilities. Results show that while there is certainly room for improvement in understanding and use of microstats in modeling, the metrics overall represent an area of promise for hockey analytics.

# Small but Mighty

## Examining the Utility of Microstatistics in Modeling Ice Hockey

Modern culture seems to fixate on the new and improved wherever possible, and sports analytics is no exception. New metrics appear regularly with various people attempting to find better ways to describe and measure how a sport is being played. In recent years, microstatistics have become increasingly used in the study of ice hockey analytics. However, there has been very little academic research into this area, with most research focusing on more established metrics. This paper will seek to analyze the utility of microstatistics using meta-analytics and examine the use of a selected set of metrics in modeling hockey games. A dataset consisting of manually tracked microstatistics from games in the NHL's 2021-2022 season will be used alongside various sources from both academic and grassroots backgrounds in order to attempt to provide a holistic view of microstatistics' background, use, and utility.

## An Examination of Hockey Analytics

Hockey is often seen as being behind other major sports in the implementation of analytics. This is largely due to the highly volatile and fluid nature of the game, with players coming on and off the ice frequently, rare breaks in play, and very infrequent goals. Early forays into hockey analytics began in earnest in the early 2000's, although they began to come into their own and be accepted into the NHL in the 2010's – primarily through the efforts of bloggers and individual researchers. Initial efforts mainly focused on finding more reliable metrics to measure team and player performance but have since expanded to additionally focus on improving decision making by players, coaches, as well as building models to predict team performance management (Schuckers & Curro, 2013; Stimson, 2016a; Tulsky et al., 2013).

**Hockey Metrics**

Goals and assists form the most basic means of player evaluation in hockey, as they measure direct involvement of players with the production of goals. One notable difference between hockey and most other team sports is the addition of the secondary (or "hockey") assist; this is awarded to a team member who provided the puck to the team member who provided the puck to the goal scorer. However, this does not paint a full picture of a player's contributions or a team's play. To this end, additional stats have been created to better capture the nuance of play. Many of these look at how play progresses as a whole while a player is present on the ice, as over the span of a season the variance in opponent and teammates should isolate the impact a single player has on play. The +/- metric represents the goal differential achieved by a team with a particular player on the ice at even strength and serves as a very rough estimate of how a player impacts the flow of play. Shot differential is also used, but more advanced stats such as Corsi and Fenwick are typically preferred. Fenwick stats add to shots on target those that missed the net, while Corsi stats additionally include blocked shots – representing all shot attempts a team takes. Corsi for and against are often seen as having a lot of utility due their repeatability (Stimson, 2016b). However, these stats do not capture the full picture because not all shots are equal. Shots taken from directly in front of the net with the goalie out of position are intuitively much more dangerous than shots taken from thirty feet away from the net, meaning that simply counting the number of shots a team takes does not necessarily indicate how dangerous that team is offensively. Expected goals models (or xG for short) look to address this gap by using various factors to estimate how dangerous a shot is and assigning an expected value to the shot. These shot values can then be summed to determine how many goals a team should expect to be scoring on the chances that they create over the course of a game, giving a good measure of how dangerous a team really is.

Expected goals models also help to compensate for one of the weaknesses of shot count metrics: attacking mentality. Coaches throughout the NHL coach different systems of play to their teams, resulting in different levels of emphasis being placed on getting pucks on net. While some teams may put the puck on net whenever possible, other teams take a much more measured approach to offense, retaining possession of the puck until a good opportunity to create a dangerous scoring chance presents itself. xG models help to account for these stylistic differences in play by accounting for the relative danger of each shot in addition to the number of shots created overall. However, xG models themselves are subject to differences due to their creators' different design goals. Some models prioritize granularity and attempt to account for as many variables as possible when calculating the value assigned to each shot (e.g., shooter skill, events prior to shot, arena, shot type, etc.) while others focus on avoiding overfitting and instead focus on fewer variables they believe to be more important. This can have a significant impact on the resulting values, as events leading up to a shot can have a significant impact on the probability of scoring. Pre-shot passes have been shown to be especially important when evaluating shot danger, in some cases doubling the value of the shot (Sznajder, 2021). Unfortunately passing data is not widely available to the analytics community due to the high cost of recording individual passes during play. This magnifies the importance of accounting for shot location when evaluating an offense's shot generation.

More recent research in hockey analytics has revealed the importance of controlled zone entries in creating offense. Tulsky et al. (2013) found that when controlling for other factors, the main difference between strong and average attacking teams was how effective they were in the neutral zone rather than the offensive zone. Tulsky's study on entries started a swell of research into how different zone entry strategies affected the flow of play in both offense and defense (garik16, 2015; Stimson, 2016a; Toumi & Lopez, 2019). Most players benefit from carrying the puck into the zone over dumping and chasing (although unskilled players should generally dump the puck as it is still the safer option) (Toumi & Lopez, 2019). Despite the indicated importance of entries, the NHL does not release any public data on player entries, meaning that most public models cannot incorporate transition data. This is especially important when looking at teams which derive most of their offense from transition plays, and teams which can effectively stop the rush are often able to shut down typically potent attacks.

The term "microstatistics" does not have a strict definition within the analytics community, but rather refers to stats which better track the flow of play through individual events as opposed to the publicly available stats provided by the NHL. The vast majority of publicly available microstats come from Corey Sznajder in one way or another, and his data will be used in this paper (Sznajder, 2022). The NHL and teams within it have access to player and puck tracking data which allows for a much clearer view of play, but unfortunately this data is not available to the public. However, the metrics which are available to the public through Sznajder's work are extremely informative when evaluating players and teams. Passing data is important due to the outsized effect of pre-shot movement on shooting percentage. Microstats also paint a clearer picture of how different teams approach the game, allowing viewers to see how a team generates (or fails to generate) offense. At the player level, microstats help to illustrate how a player provides value to the team – whether through transition play, shooting, passing, or defensive work – and can help to explain how different players impact one another when playing together. Importantly, research has shown that various microstatistics are reliable and even have some predictive power on future goals scored (Stimson, 2016b).

**Modeling**

At a basic level, game models attempt to determine the relative strength of different teams and use that along with various other factors to predict the probability of each team winning. The differences between different models tend to arise due to different views of how the sport is best modeled, what is most important when attempting to predict future success, as well as how win probabilities should be determined. Most models fall under the categories of transitive, Poisson, Markovian, or simulation based.

Transitive models are arguably the easiest to evaluate due to their simple design. Models such as the Colley or Massey models have stood the test of time due to their easy implementation along with the steady results they have produced. These models work by assuming that results between teams are loosely transitive and attempt to determine the strength of each team by their performance over the course of a season (Swanson et al., 2018). These strength ratings are then compared against one another to produce a probability of each team winning. In some cases, a transitive model is combined with more complex models, such as the model proposed by Swanson et al. (2018) which uses a modified transitive model based on search engines and implements Corsi stats to enhance the model. These models tend to be simple to implement, but often sacrifice performance for simplicity due to scenarios in which a team may outperform its actual ability and underlying numbers.

Poisson models assume that the scoring of goals can be modeled as a Poisson process and use that fact to approximate how many goals will be scored by each team in a game based on various factors. Likely the largest distinguishing factor between Poisson models is the metric(s) selected to predict goal scoring rates. This can be goals scored, Corsi figures, Fenwick, expected goals, some combination of these, or any other stat the model builder chooses (Buttrey, 2016; Buttrey et al., 2011; Thomas, 2007). Corsi tends to dominate in academic circles due to its predictive power over future goal scoring, but public models tend to rely more on expected goals due to xG accounting for shot danger (Goldman, 2021). Models also differentiate themselves based on how (or whether) they account for various phenomena which occur over the course of play, such as penalties, the goalie being pulled, home advantage, and variations in scoring rates which would appear to violate the assumption of goal scoring rates being modeled as a Poisson process (Buttrey, 2016; Thomas et al., 2013). One approach to the issue of goal scoring rates being lower after goals and at the beginning of periods due to puck drop is to use a hazard function to penalize the goal scoring rate immediately following the start of a period and after a goal (Thomas et al., 2013). Markovian models, while less common, look to address the flawed assumption in Poisson models that goal scoring rates remain constant as score states and other factors vary throughout the game. In truth, goal scoring rates tend to differ based on score differential (Thomas, 2007). These models essentially consist of multiple Poisson models built to model a specific game state in order to improve the accuracy of the model and address changes in scoring rates. Markovian models have also been applied to examine the value of various actions during play, as the nature of Markov chains lends itself to the flowing nature of play in hockey. These applications have helped to show correlation between various actions tracked by microstatistics and team success (Schulte et al., 2017). These models tend to be more computationally intensive, but generally outperform simpler models. However, the requirements for Markovian modeling appear to be violated by hockey due to the aforementioned changes in scoring rates and effects of pre-shot movement on shooting percentage, as this violates the memoryless requirement.

Finally, simulation-based models typically use regressions to determine a team’s offensive and defensive abilities, then simulate a game repeatedly in an attempt to determine each team’s probability of winning. These models are most common in the public sector, but often perform as well as, or better than, academic models and sometimes outperform market models when applied to sports betting (Luszczyszyn, 2022).

## Examining Microstatistic Utility

One of the primary issues faced when working in sports analytics is the tension between providing as much data as is useful while ensuring that the data being provided is tractable for the intended audience – be it a coach, general manager, or the public. This is one reason that some members in the sporting community and members of the public find it difficult to trust analytics, as the sheer quantity of information presented is often overwhelming. To this end, it is crucial to verify that a given metric provides useful information when looking to introduce it to the broader analytics community. For this reason, the metrics being examined in this paper will be evaluated to ensure that they are indeed useful for analyzing hockey and are not needlessly contributing to the mountain of metrics already available today.

### Meta-Analytics

In their 2016 paper, Franks et al. presented a set of so-called “meta-analytics” to address this. The three meta-analytics examine the variance of metrics to determine how efficacious they are in an objective manner in how well they: distinguish between entities, remain stable over time, and how well they provide novel information about a player or team. The first, discrimination, will be examined in detail due to its use in this paper; the other two meta-metrics will be examined briefly due to their important implications in the field of analytics as a whole.

The first of the three, discrimination, measures how well a given metric distinguishes between the entities under consideration (whether players, teams, or leagues) (Franks et al., 2016). This is important for determining how useful metrics are for drawing concrete conclusions about overall performance or playstyle (such as scouting opposing teams and players). If a metric has little to no discriminatory power, this indicates that most of the trends shown in the data are caused more by random variance in the data than any real difference between the players or teams under consideration. The discrimination meta-metric for a given metric is calculated by subtracting the ratio of the mean variance for individual entities and the variance of the entire sample from one. This formulation gives a measure of how much the variance between players or teams differs from the overall sample variance. In metrics which do a good job of distinguishing between entities, the variance for individuals should be considerably smaller than the variance of the entire population, resulting in a high discrimination score.

Stability, as the name would implies, measures how stable a metric remains over time (Franks et al., 2016). This is crucial when attempting to use metrics to make decisions about the future; the consequences of failing to use stable metrics are plain to see in many sports, as general managers will often sign promising young players to contracts based on a strong season that quickly shows itself to be a fluke. Its calculation compares the variation of a metric between seasons to the overall sample variance while controlling for sampling variance. While this metric does appear effective for measuring metric stability, the restriction of a single season's worth of data for this study means that another metric will be needed to measure metric stability. For this purpose, the split-half reliability test will be employed. The split-half reliability test examines how well a metric's average value in a random split of the sample data predicts the remaining data. One drawback of this method is that it does not account for steady changes over time in metrics, but given the nature and structure of the data being used for this study, this should not be a significant problem.

The final meta-metric presented, independence, attempts to quantify the amount of novel information a given metric provides when compared with other available metrics (Franks et al., 2016). This functionally amounts to a parameterization of multicollinearity measures by examining how well various metrics correlate to the metric under consideration. While this certainly represents an interesting avenue for potential future research, the independence meta-metric is outside the scope of this paper.

**Data**

The database used in this study was created from a combination of Corey Sznajder's data collected during the 2021-2022 NHL season and Natural Stat Trick's data from the same season (*Natural Stat Trick*, 2022; Sznajder, 2022). Sznajder's data is recorded manually, resulting in a smaller quantity of games recorded but with a higher degree of granularity. The data set contained 497 games including all 32 NHL franchises. On average, each team had approximately 31 games including in the database, with Arizona (21 games) having the fewest games and

Colorado (42 games) having the most. Game counts for each team can be seen in Table 1. The games recorded for each team were spread throughout the season, making analysis of trends in the data difficult. Because of this, all work with the data was completed without regard to when the game took place within the season. While this does introduce seasonal trends as a confounding variable (as well as causing difficulties regarding in-season trades), attempting to account for in-season trends with inconsistent gaps between recorded games in the relatively small sample size would have been irresponsible. Another important factor for consideration is that the data under consideration needed to be processed at the team level in order to match up with general stats data from the other database. This meant that any injuries to key players or changes in starting goalkeepers could not be modeled. Sznajder’s microstatistics focus on even strength play, meaning that stats needed to be converted to rate stats in order to control for the amount of special teams play teams had in games. This was achieved by converting each metric to estimate its value if the entire 60 minutes of play was at even strength, resulting in their being named “per 60” stats. The Natural Stat Trick dataset covered all regular season games from the 21-22 season and contained detailed counting stats at even strength and special teams situations.

**Table 1**

*Game Count in Training Set by Team*

| Team | Games | Team | Games | Team | Games | Team | Games |
|---|---|---|---|---|---|---|---|
| COL | 42 | WPG | 35 | WSH | 32 | DET | 27 |
| EDM | 39 | NYR | 35 | BOS | 31 | CHI | 26 |
| NYI | 37 | CGY | 34 | NSH | 31 | ANA | 25 |
| TBL | 36 | PIT | 34 | STL | 31 | SJS | 24 |
| MIN | 36 | FLA | 33 | NJD | 29 | BUF | 23 |
| VGK | 36 | LAK | 32 | SEA | 29 | CBJ | 23 |
| TOR | 36 | PHI | 32 | VAN | 28 | MTL | 23 |
| CAR | 35 | DAL | 32 | OTT | 27 | ARI | 21 |

*Note.* Data from Sznajder, C. (2022). *2021-22 Game Sheets*. All Three Zones. Retrieved January 30, 2023, from https://www.allthreezones.com/2021-22-tableau.html

## Methods

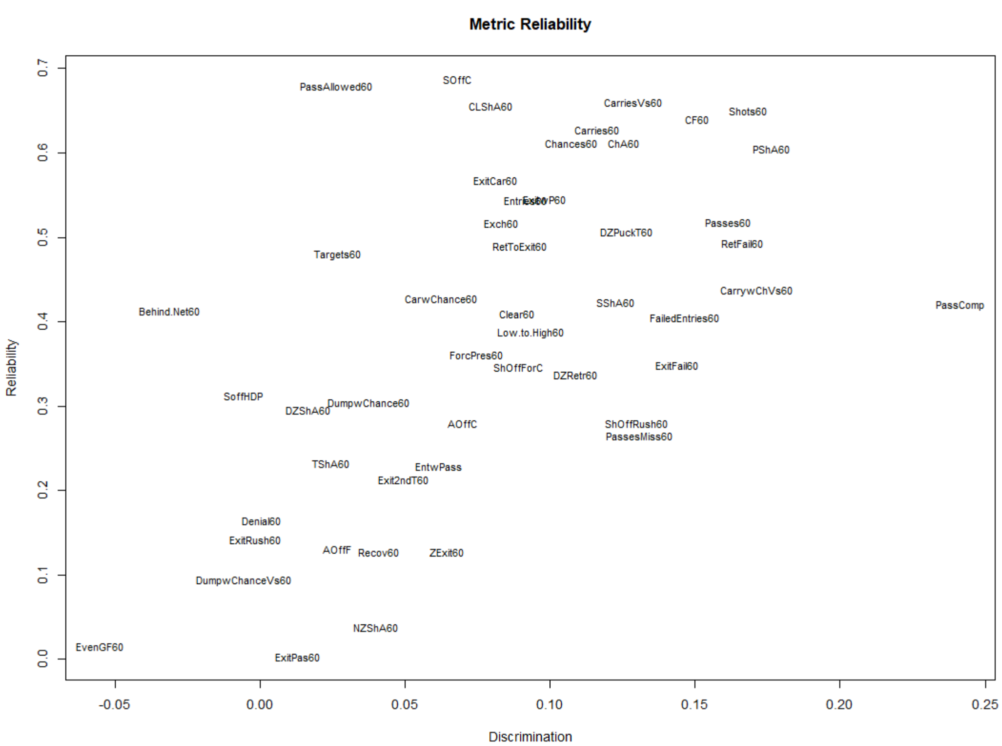
Discrimination scores and the split-half reliability test were used to evaluate the microstatistics under study. All metrics were measured at the team level and were converted to rate metrics to control for the amount time played at even strength. Discrimination scores were found by finding the variance for each team and the league as a whole for each microstat and finding their ratio. Reliability scores were determined using a simple even-odd split. This ensured that the split of games remained semi-random (given the distribution of games across the season already varying the selection of games) while to an extent controlling for trends in stats over the course of the season.

## Results

As can be seen in Figure 1, there appears to be a loose correlation between the reliability of a metric and its discriminatory power. This makes intuitive sense. Metrics which vary significantly from game to game will have large denominators in their discrimination calculations. The result is a much higher bar for differences between teams to reach any given discrimination score. Encouragingly, metrics traditionally seen as reliable (such as Corsi For and Carries per 60) appear to be among the more reliable stats under study – indicating some level of legitimacy in the results. Importantly, Goals For at even strength per 60 looks to be extremely unreliable based on the meta-analysis, which would imply that how good a team is at scoring is not a good measure of how well that team is driving play and is not a good choice of metric to attempt to project future success. While this is a well-known idea, its importance means that it bears repeating. One item of note is that the overall scores appear to be much lower than the results reached in the initial meta-analysis study (Franks et al., 2016). Differences in the data used to calculate the metrics may have had some effect on this, but given that the maximum discrimination score in this dataset was around half the average in the previous study, there appears to be something more at work. The shift from player- to team-based stats is a likely culprit given that this represents the largest difference between the two studies and has some theoretical backing. Due to rules introduced after the lockout in 2005 creating a salary cap, there is a large degree of parity between NHL teams talent-wise. Contrast this with the talent discrepancies between individual players and the gap in discrimination scores begins to make sense. On every NHL team, various players tend to fill various roles – be it star player, grinder, two-way player, etc. – resulting in statistical profiles that tend to line up regardless of team level (Chan et al., 2012). This results in a high degree of discrimination between many player-level stats, as they will tend to differ greatly between various types of players. In contrast, while a strong offensive team in the NHL may average 4 goals per game when compared to a team which averages 3 goals, the game-to-game variance in scores results in a much higher degree of variance. This can be seen in the use of Poisson processes to model goal scoring in the NHL, meaning that the variance for a team averaging 4 goals every 60 minutes will also have a variance of 4 goals, while only differing by 1 goal per 60 when compared to a poorer team.

**Figure 1**

*Discrimination and Reliability Scores for Microstats in Training Dataset*




Overall, most of the metrics rated as being more reliable in the meta-analysis tended to be related to possession in some way, such as carried entries or Corsi. This helps to better contextualize previous research which found that teams' neutral zone play was one of the most important drivers of team success and supports the idea that microstats should be helpful in predicting game results (Tulsky et al., 2013). However, even metrics which did not grade out favorably in the meta-analysis are not useless. While objective utility is important when looking to make decisions about the future and analyzing the past, there is still something to be said for stats which may not have strict utility, but still help to describe the flow of play, such as the play styles of certain teams and players and how they have reached their current positions. As an example, goal scoring is not a very helpful stat for describing the actual performance of a team, but very few people would say that looking at how many goals a team has scored is useless – not least because it aids in primary purpose of professional sports, storytelling. In summary, while all metrics likely contribute something to the discourse in sport, some microstatistics – especially those connected to possession – have objective utility for analyzing and modeling the sport.

## Modeling with Microstatistics

Modeling sports is an extremely difficult process. This is due to the number of players and outside factors at play, including: rest, travel requirements, home advantage, morale, and many others. Despite (or perhaps because of) this, attempting to model and predict the outcome of various sports has been present within the analytics community since its inception. Part of this can certainly be attributed to the pursuit of sports betting, but models are also helpful because they can help to identify what drives team success and what teams should be placing a priority on. Hockey is especially difficult to model when compared to a sport such as baseball due to its continuous nature, rapid pace of play, frequent changes in personnel, and the infrequency of scoring events. While applying microstatistics would seem to risk feeding a model too much data to be able to comprehensively predict future results, the presence of pre-shot movement and entry data should help to fill gaps in the information given to models by current publicly available metrics.

**Design**

The standard for modeling sports such as hockey and soccer with relatively rare scoring has traditionally been to model goal scoring as a Poisson process. While some have made alterations to the basic idea – such as adjusting for score state as a Markovian system – the general premise has traditionally held strong with the exception of simulation models. As this paper seeks to present a proof-of-concept model rather than a full-fledged game prediction, the simpler Poisson model will be used. In order to determine if using microstatistics is better or worse than a more basic model, a control model was built based on goal scoring to control for model design and ensure that any differences in performance are in fact due to the metrics used rather than the model design itself.

The model itself was built by using selected metrics in a linear regression to model each team's goal-scoring and goal-concession rates. The linear model was trained by using mean rates for even and odd splits of each team's games attempting to predict the even strength goalscoring and concession rates in the opposite pool. The resulting model was used to estimate scoring and concession rates over the entire training set, consisting of Sznajder's dataset matched with game data from the Natural Stat Trick dataset. These estimates were combined with penalty rates in a Poisson regression to create a model for estimating goal scoring rates for each team in a game. Once every team's goal scoring had been estimated as Poisson random variables, the difference between each team's goal estimate could be modeled using a Skellam distribution. Calculating the probability of this value being greater than 0 produced the win probability for the desired team.

Akaike Information Criteria (AIC) was used to narrow down the various microstatistics available, predicated on goal scoring and concession. While goal scoring is not a very reliable stat – as was discussed previously – it is the end goal of offense and is thus important to attempt to predict despite the difficulty. While this does run some risk of making the model reliant on correlations in the training set which are not necessarily present in the sport overall, it does a good job of reducing the number of variables under consideration while retaining those that have a strong impact on model efficacy. Variables selected in the AIC process can be seen in Table 2. Goals for were removed from the final models in both cases due to the unstable nature of the metric in addition to the counterintuitive negative estimated parameter it was assigned (implying that past goal scoring predicted less future goal scoring). Opponent failed retrievals were also removed from the goal concession model due to lack of a tractable interaction with defensive performance.

**Table 2**

*List of Variables Recommended for Predicting Goal Scoring and Goal Concession*

| AIC Results | |
|---|---|
| Goal Scoring | Goal Concession |
| Chances/60 | Opp Shots Off Forecheck/60 |
| Exchanges/60 | Exits with Rush/60 |
| Even Strength GF/60 | Opp Failed Retrievals/60 |
| Carried Exits/60 | Primary Shot Assists/60 |
| Shot Attempts/60 | Targets/60 |
| Entry with Pass/60 | Opp Carries Against/60 |
| | Retrieval Fails/60 |
| | Opp Entries/60 |
| | Shots off High Danger Passes/60 |
| | Opp Even Strength GF/60 |
| | Opp Dump with Chance/60 |
| | Passes Allowed/60 |
| | Opp Secondary Shot Assists/60 |
| | Corsi For/60 |
| | Opp Shots off High Danger Pass/60 |
| | Opp Carry with Chance/60 |
| | Opp Denials/60 |

After deciding on the variables for the final model, linear regression was used on the splits of the training set to create linear models for projecting teams' goal scoring and concession rates. These were applied to the full training set to project rates for each team to be used in the test set. These rates (see Table 3) were combined with the rates at which teams reached various special teams situations in a Poisson regression over individual games. Each game was assigned the rates for respective teams in order to attempt to project the number of goals a team would score in an individual game. This regression over the training set gave a Poisson model which could applied over the test set with a Skellam distribution to determine win probabilities.

**Table 3**

*Team Coefficients for Model Parameters*

| Team | Goals60 | Conc60 | PP Min | PK Min | Team | Goals60 | Conc60 | PP Min | PK Min |
|---|---|---|---|---|---|---|---|---|---|
| BOS | 2.839 | 2.071 | 5.046 | 5.244 | VAN | 2.520 | 2.164 | 5.162 | 4.277 |
| BUF | 2.326 | 2.097 | 5.267 | 4.430 | DET | 2.494 | 2.257 | 5.041 | 5.114 |
| CBJ | 2.655 | 2.395 | 4.467 | 5.063 | WPG | 2.220 | 1.787 | 5.756 | 4.989 |
| CAR | 2.906 | 1.932 | 4.450 | 5.897 | ARI | 2.282 | 2.418 | 4.248 | 4.776 |
| NYI | 2.341 | 1.849 | 4.498 | 4.198 | DAL | 2.618 | 2.170 | 4.371 | 4.730 |
| EDM | 2.827 | 2.150 | 4.250 | 5.284 | WSH | 2.672 | 2.119 | 4.566 | 4.428 |
| LAK | 2.506 | 2.095 | 5.352 | 4.430 | MIN | 2.622 | 2.018 | 5.212 | 5.795 |
| PHI | 2.485 | 2.095 | 5.080 | 5.174 | VGK | 2.487 | 2.033 | 4.382 | 3.883 |
| TBL | 2.563 | 2.096 | 5.340 | 5.314 | PIT | 2.774 | 2.017 | 5.780 | 4.431 |
| NSH | 2.540 | 1.766 | 4.548 | 5.748 | NYR | 2.445 | 2.254 | 4.185 | 4.881 |
| COL | 3.292 | 2.354 | 5.916 | 4.348 | TOR | 2.897 | 2.327 | 4.372 | 4.956 |
| CGY | 2.678 | 1.944 | 5.096 | 5.532 | ANA | 2.406 | 2.055 | 4.674 | 4.799 |
| OTT | 2.236 | 1.863 | 4.953 | 4.761 | CHI | 2.257 | 2.208 | 4.973 | 4.171 |
| NJD | 2.489 | 2.327 | 4.610 | 4.064 | MTL | 2.456 | 1.889 | 6.197 | 6.301 |
| STL | 2.911 | 2.122 | 4.431 | 5.209 | FLA | 3.115 | 2.391 | 5.676 | 5.629 |
| SEA | 2.427 | 1.963 | 5.317 | 4.073 | SJS | 2.343 | 1.898 | 3.618 | 5.270 |

These distributions were selected based on past research by others in the field as well as the individual distributions' characteristics. Linear regression was selected to model the mean scoring rates for each team because the data selected to fit the model consisted of each team's mean rates – this should result in a normal distribution being adequate by the central limit theorem. Poisson regression was used to model individual games simply because it is the standard in modeling of low-scoring games such as hockey and soccer. Finally, the Skellam regression is defined as the difference between two Poisson random variables, and is thus well suited for this purpose.

The control model was created using a similar Poisson model only implementing mean goal scoring and concession rates over the entire training set to provide a similarly structured model. This allowed for a more objective comparison to a more basic model without comparing the model to models of other designs. Comparing it to other models would confound any differences in results caused by the utility of metrics used with any effects introduced due to differences in model design.

In addition to the control model, two additional modified models were examined. The first looked to better account for the differences created by special teams play. Each team's power play time in a game was estimated with a linear regression based on how much time teams spent in various special teams situations. Special teams scoring rates were estimated using scoring and concession rates. To better account for how increased power play time cut into even strength play, the time spent in special teams states was subtracted from the initial 60 minutes. In order to accurately incorporate the timing information into the models, the amount of game time spent in each state was divided by 60 (to account for the proportion of game time spent in that state) and its natural logarithm was passed to the Poisson regressions – this ensured that the projected goal scoring rate per 60 would be adjusted for the proportion of the game played in that state. The second model examined included the adjustments for special teams play but also used the ratio of expected even strength goals to even strength goals scored and conceded by each team. The natural logs of these ratios were passed to the even strength Poisson regression in order to attempt to account for the effects of elite shooting and goaltending.

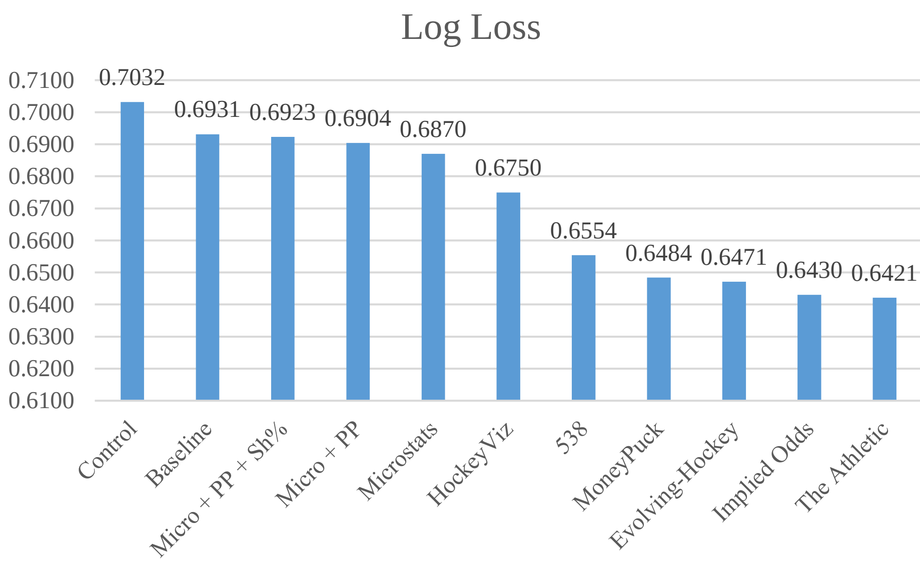
**Results**

Overall, the microstats comprehensively outperformed the control model, indicating that the selected microstatistics are useful for modeling hockey. The base microstats model correctly predicted the game winner 55.21% of the time compared to the control model's 53.62%. The models were also compared by their log-loss. This metric measures how effective a model is in allocating confidence to a winner in addition to predicting the winner by taking the log of the probability left on the table by the model. For example, if a model gave the projected favorite a 56% chance to win and that team were to win, the log-loss would be the natural logarithm of .46: .46 being the gap between the model's confidence in its projected result and the actual result. The mean of all games' log-loss is used to measure how well a model projects a favorite's chances to win. This is an important method of model comparison, as only examining hit rate fails to account for the fact that a strong team with a 60% chance to win will still lose 40% of the time. The microstats model seemed to outperform the control model handily in this area, as the control model's log-loss of .7032 failed to beat the baseline score of .6931 (representing the score if a model gave every team a 50% chance to win), while the microstats model reached .6870. These results suggest that the microstats model was much more effective than the basic goals model.

While the model did perform well against the control, it performed very poorly when compared with popular public models. Dom Luszczyszyn (2022) published a review of performance of his model (The Athletic) after the conclusion of the 2021-2022 season and included the log-loss for his model as well as several other popular public models. As can be seen in Figure 2, the microstats model presented lagged far behind the other models as well as the implied odds taken from the betting market. Luszczyszyn's model also outperformed the microstats model by a large degree in hit rate, with a mark of 64%.

**Figure 2**

*Comparison of Model Log-loss*




There are several important notes to be made regarding the microstats models' performance when compared with these public models. Crucially, these models have been developed and refined over many seasons, allowing them to be better adjusted to produce results. In addition, these models have the benefit of projecting games in chronological order. This allows for model input to weight data from recent games more heavily to account for trends in team performance throughout the season. Many of the top public models are also built using player-level data, allowing for better adjustments for injuries, trades, and starting goalies. One final note is that these models were measured over the entirety of the season, whereas the control and microstats models were judged solely on the games in the test set.

Comparison of the base microstats model to the two additional models yielded interesting results. On the one hand, both enhanced models outperformed the base and control models regarding hit rate, with the power play model reaching 56.81% and the power play and shooting adjusted model reaching 56.44%. Compared to the respective 53.62% and 55.21% of the control and base microstats models, the adjusted models would appear to be superior, however both adjusted models recorded log-losses of over .69, indicating that the base model was much better regarding setting confidence levels. The disparity in results between log-loss and hit rate indicates that while there is merit to the approach behind the adjusted models, more refinements are needed to ensure that they correctly set confidence levels. Overall, while the microstats model was able to outperform the baseline measure and the control model, there is clearly room for improvement in the model.

**Discussion**

The performance of the microstats models illustrates that some microstatistics do have value in modeling, but improvements will need to be made in the modeling methods in order to ensure that the microstats are worth the opportunity for individuals or teams to acquire them, whether by paying for their use or tracking them. Additional research will also be required to determine whether other microstatistics also prove useful for modeling and analysis. The adjusted models show promise for future research and demonstrates ample room for improvement with tuning to account for their deficiencies in setting confidence levels. These models would also benefit from a more stable data supply and a player-level approach. If the order and timing of games could be taken into account, season trends could be accounted for. Player-level analysis also appears to be more effective (Luszczyszyn’s model is an example of this) and would also help to account for injuries and roster changes such as goaltender choice. Finally, it would be helpful to account for the meta-analysis when selecting variables in a more concrete manner, potentially finding a way to weight the AIC towards selecting highly reliable metrics unless there is overwhelming evidence to support including a less reliable metric. This should help to improve the predictive power of the models.

## Conclusion

This paper analyzed the utility of various microstatistics and examined the use of a selected set to model hockey games. Various microstatistics (especially those indicative of neutral zone play) were found to be fairly reliable and did a good job of discriminating between the playstyles of different teams. In addition, the discriminative power of metrics appears to be lessened when examining them at the team-level versus player-level, likely due to the difference between teams not outweighing the inherent variance in the sport of hockey while the differences between different classes of players are much larger. These metrics are also helpful in modeling hockey when compared with basic goals-based models and show promise for creating new models or augmenting existing models in order to improve on currently available projections.

This research opens the door to various opportunities for future research. One obvious application of this research is to sports betting, as, if an adequate model is created, users can potentially make money by beating the market. However, a more useful application is in looking for what drives team success most effectively and directly. This can allow players and teams to focus on developing traits and skills that directly help teams to be more successful. Better player evaluation would also have ramifications for fantasy sports, drafting, and roster building. Oftentimes younger players who post excellent microstatistics numbers but struggle to score will later break out to become superstars – such as budding Devils center Jack Hughes. The success in using these more detailed statistics also suggests that using more detailed metrics may also be successful in researching other sports, such as soccer. Soccer is especially promising given that the rules and flow of play in the two sports are similar as well as both being modeled using Poisson processes. Much work remains to be done to expand knowledge regarding the utility of various emerging metrics, but the area of microstatistics contains promise waiting to be explored.

**References**

Buttrey, S. E. (2016). Beating the market betting on NHL hockey games. *Journal of Quantitative Analysis in Sports*, *12*(2). https://doi.org/10.1515/jqas-2015-0003

Buttrey, S. E., Washburn, A. R., & Price, W. L. (2011). Estimating NHL scoring rates. *Journal of Quantitative Analysis in Sports*, *7*(3). https://doi.org/10.2202/1559-0410.1334

Chan, T. C., Cho, J. A., & Novati, D. C. (2012). Quantifying the contribution of NHL player types to team performance. *Interfaces*, *42*(2). https://doi.org/10.1287/inte.1110.0612

Franks, A. M., D'Amour, A., Cervone, D., & Bornn, L. (2016). Meta-analytics: Tools for understanding the statistical properties of sports metrics. *Journal of Quantitative Analysis in Sports*, *12*(4). https://doi.org/10.1515/jqas-2016-0098

garik16. (2015, November 13). *Individual skill, coaching/systems, or randomness: What is driving open play results in the offensive, defensive and neutral zones?* Hockey Graphs. https://hockey-graphs.com/2015/06/25/individual-skill-coachingsystems-or-randomness-what-is-driving-open-play-results-in-the-offensive-defensive-and-neutral-zones/

Goldman, S. (2021, December 9). *Comparing public expected goal models: How they work and what we should take away from them*. The Athletic. https://theathletic.com/3006028/2021/12/09/comparing-public-expected-goal-models-how-they-work-and-what-we-should-take-away-from-them/

Goldman, S., & O'Connor, C. (2022, June 22). *Explaining the key advanced stats: How NHL analytics measure player and team value*. The Athletic. https://theathletic.com/3377555/2022/06/22/nhl-analytics-player-team-value/

Luszczyszyn, D. (2022, July 5). *The 2021-22 NHL season and what I got right and wrong: Predictions review*. The Athletic. https://theathletic.com/3390081/2022/07/05/2021-22-nhl-predictions-review/

Natural Stat Trick. (2022). *Games* (2021-2022 season) [Data set]. Natural Stat Trick. https://www.naturalstattrick.com/games.php

Schuckers, M., & Curro, J. (2013). *Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events.* statsportsconsulting.com. https://www.statsportsconsulting.com/wp-content/uploads/Schuckers_Curro_MIT_Sloan_THoR.pdf

Schulte, O., Khademi, M., Gholami, S., Zhao, Z., Javan, M., & Desaulniers, P. (2017). A Markov game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, *31*(6), 1735–1757. https://doi.org/10.1007/s10618-017-0496-z

Stimson, R. (2016a, August 3). *Tactalytics: Using data to inform tactical neutral zone decisions*. Hockey Graphs. https://hockey-graphs.com/2016/08/03/tactalytics-using-data-to-inform-tactical-neutral-zone-decisions/

Stimson, R. (2016b, July 11). *Tactalytics: Using data to inform tactical offensive zone decisions*. Hockey Graphs. https://hockey-graphs.com/2016/07/11/tactalytics-using-data-to-inform-tactical-offensive-zone-decisions/

Swanson, N., Koban, D., & Brundage, P. (2018). Predicting the NHL playoffs with PageRank. *Journal of Quantitative Analysis in Sports*, *13*(4). https://doi.org/10.1515/jqas-2017-0005

Sznajder, C. (2021, January 12). *Passing project revisited*. Passing Project Revisited - by Corey S. https://allthreezones.substack.com/p/passing-project-revisited?s=r

Sznajder, C. (2022). *2021-22 Game sheets* [Data set]. All Three Zones. https://www.allthreezones.com/2021-22-tableau.html

Thomas, A. C. (2007). Inter-arrival times of goals in ice hockey. *Journal of Quantitative Analysis in Sports*, *3*(3). https://doi.org/10.2202/1559-0410.1064

Thomas, A. C., Ventura, S. L., Jensen, S. T., & Ma, S. (2013). Competing process hazard function models for player ratings in Ice Hockey. *The Annals of Applied Statistics*, *7*(3). https://doi.org/10.1214/13-aoas646

Toumi, A., & Lopez, M. (2019, September 28). *From grapes and prunes to apples and apples: Using matched methods to estimate optimal zone entry decision-making in the national hockey league*. RPubs. https://rpubs.com/atoumi/zone-entries-nhl

Tulsky, E., Detweiler, G., Spencer, R., & Sznajder, C. (2013). *Using zone entry data to separate offensive, neutral, and defensive zone performance.* hockeyanalytics.com. http://hockeyanalytics.com/Research_files/Using%20Zone%20Entry%20Data%20To%20Separate%20Offensive,%20Neutral,%20And%20Defensive%20Zone%20Performance.pdf