By The Numbers: OHL: Points, Penalties, and Predictions

A few years back, in 2013, I looked at the effect OHL rookies had on teams' year-end statistics. The study looked at the relationships between various year-end statistics, including number of rookies, penalties, goals for and against, etc. Generally the finding was that, over the eight seasons referenced and for all teams collectively, the proportion of rookies to non-rookies did not tend to correlate with teams' year-end statistics. The study concluded that more detailed analysis using more individual player statistics would be helpful in understanding several apparent anomalies that did suggest, at least for some seasons and at the team level, rookies did have a negative effect on various year-end statistics. (The brief report can be found at DMCmetrix.com.)

I became curious about any effect penalties might have on various year-end statistics, particularly goal scoring. Penalties result in a team being temporarily short-handed so intuitively it would seem the opposition has an advantage. I am sure this has been looked at before because sports and statistics seem inseparable. Anyway, this is my take on it.

Raw data for regular and play off games were manually extracted from statistics posted at hockeydb.com for the ten-year period from 2006/2007 to 2015/2016. Analyses were done using WINKS/Windows KwikStat.

Because hockey is a team sport, despite players with exceptional skills that often seem to carry the team. As such, this study uses team-level, not individual player statistics. Basic statistics are presented, including correlations. [Correlations 101: Correlation does not prove cause and effect, just that there is a relationship between two variables, and that "something" might be going for which further research is needed to determine what underlying factors might exist.]

Ideally I would have liked to have included "fights" as one of the variables; unfortunately I was not able to identify a Website with reliable, consistent fights data.

The types of penalties are not considered in the analyses. For example, a misconduct results in a player receiving a 10-minute penalty, which is included in the year-end statistics, but does not result in his team being short-handed. All year-end penalties in minutes (PIM), as posted to hockeydb.com, are used in calculations.

The Ontario Hockey League (OHL) brought in a 10-fight suspension rule for the 2012/2013 season. The press has reported that this initiative has almost halved the number of fights (Hockey fights cut almost in half since suspension rule, The London Free Press, Wednesday, March 8, 2017). OHL vice-president stated, "An unexpected byproduct of the rule was a significant decrease in the overall fighting in our league." What has been the overall effect of the new rule and has it had an impact on other statistics?

10-Year Summary

Table 1 presents means (averages) for PIMs, GF, and GA for the 10-year study period. Immediately noticeable is that means (averages) are lower for playoff games than regular season games. The noticeable exception is GA. Figures 1 and 2 illustrate this graphically. It is interesting to note that in the season of the suspension rule, play off GF continued a slight drop but went up after that (Figure 1). Also, PIMs showed a slight *decline* for regular play and

a slight *increase* for playoffs (Figure 2). Although the 2012/2013 suspension rule is credited with a decrease in fights, PIMs had been declining since at least 2006/2007, so we may surmise that fights had been declining as well.

Table 1: 10-Year Means for Regular and Play Off Variables for 2006/2007 to 2015/2016

	Regular	Play Offs
Penalties in Minutes per Game (PIMs)	16.1	11.6
Goals For per Game (GF)	3.5	3.1
Goals Against per Game (GA)	3.5	3.7

Table 2: Per Game GF and PIM Pre and Post Suspension Rule

	Regular		Play Offs	
	GF	PIM	GF	PIM
2006/2007 to 2011/2012	3.5	17.8	3.1	16.0
2012/2013 to 2015/2016	3.5	13.7	3.0	13.9
Change	0.0%	-23.0%	-3.0%	-13.0%

A look at Table 2 provides more insight into what was going on before and after the suspension rule. There was no change in mean GF for both regular and play off games. PIM is a different story, which showed declines of 23% and 13%. Separate, season-by-season calculations were done for each of the ten seasons. In 2006/2007 the spread between minimum and maximum PIMs was 34.2 minutes in 2006/2007 but by 2015/2016 that spread had dropped to 16.5 minutes. This is a tightening up around the mean, a smaller standard deviation, which went from 9.2 minutes in 2006/2007 to 5.1 minutes in 2015/2016.

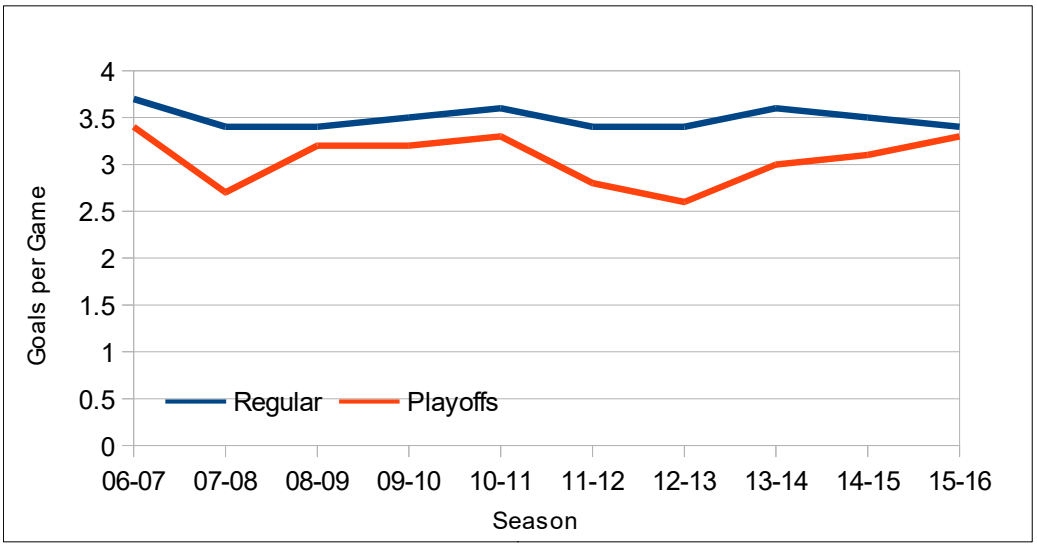

Figure 1: Mean Goals For per Game

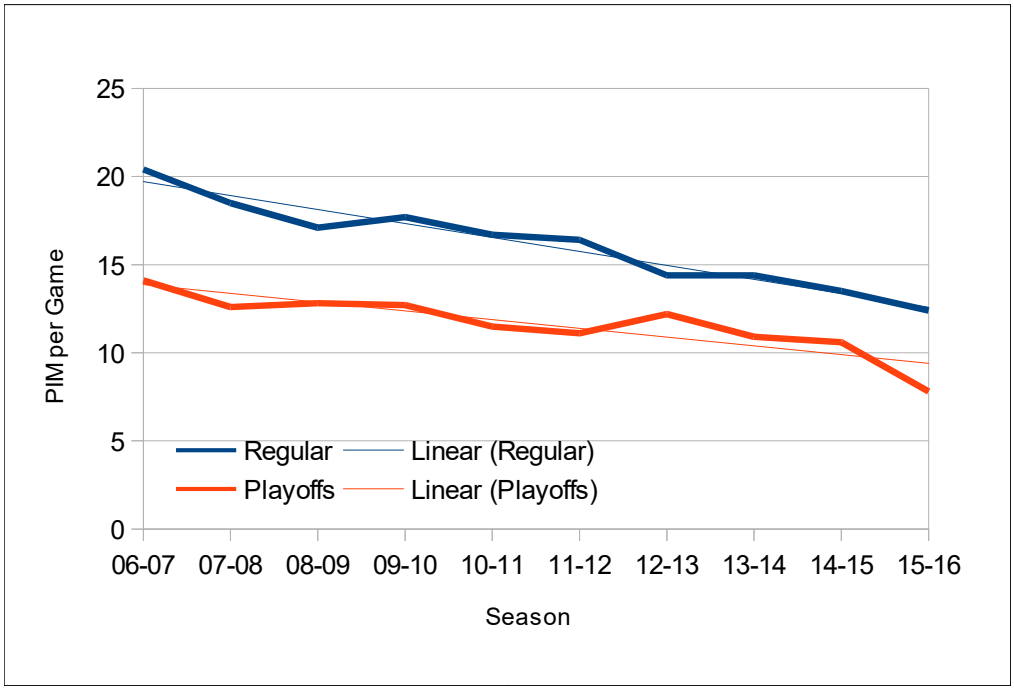

Figure 2: Mean PIMs per Game (Lines marked "linear" are trend lines.)

Correlations

In this section we are going to dig a littler deeper into the raw data using correlations to determine how several variables might be related.

To determine if there is a meaningful relationship between two variables, correlations are used. Do the variables tend to move in the same or opposite direction when one changes. For example, ice cream sales go up as the weather warms or spending less time studying for a test likely results in more mistakes. Calculated correlations are between 0/zero and 1.0: a perfect correlation is r=1.0 and r=.00 shows no relationship. Correlations should be at least 0.400 and higher to show relationships that are "moderate" to "very strong" between two variables. For the statisticians in the crowd, probabilities are shown as p=x.xxx.

<u>Regular Season</u>

For regular season games there are no significant correlations, except GF and GA (r=0.472). Generally, then, PIMs, do not seem to have an effect on GF and GA. Or, stated another way, teams that "earn" higher PIMs, do not seem to be recipients of more GA, nor do they score more often with players on the bench.

Table 4: Regular Season per Game Correlations		
	PIM	Goals Against
Goals For	r= 0.180*	r= 0.472**
Goals Against	r= 0.112	

**p<0.05; **p<0.01.*

Play-offs

Calculations for play off games show no significant correlations. This suggests that PIMs, whether trending higher or lower, do not have an effect on goals: fewer players on the ice are not correlated with more GA for a team, which, intuitively, we might expect.

Table 5: Play Offs per Game Correlations		
	PIM	Goals Against
Goals For	r= -0.015	r= -0.172*
Goals Against	r= 0.158*	

**p<0.05.*

Regular Season and Play-Offs

Here we look at what happened during the regular season and if that pattern is repeated during play offs. For example, do more GF in regular play carry over to the play offs? Table 6 presents regular season/play off correlations.

Table 6: Regular Season and Play Offs per Game Correlations			
		Play Offs	
	PIM	Goals For	Goals Against
Regular Season Goals For	0.062	0.699***	-0.248**
Regular Season Goals Against	0.167*	-0.310***	0.527***
Regular Season PIM	0.547***	0.146	0.016

**p<0.05; **p<0.01; ***p<0.001.*

It appears that how a team performs in the regular season hints at play off performance. GF, GA, and PIM show moderate to strong positive correlations for regular season/play offs. This is not predictive of regular season to play off performance, but 10-year's worth of data does suggest that a team that does well in seasonal play *might* do well in play offs. Correlations are shown graphically below (Figure 3). Taking this to the next step, a regression model is developed and tested to determine if using regular season statistics might be able to "predict" likely play off spots.

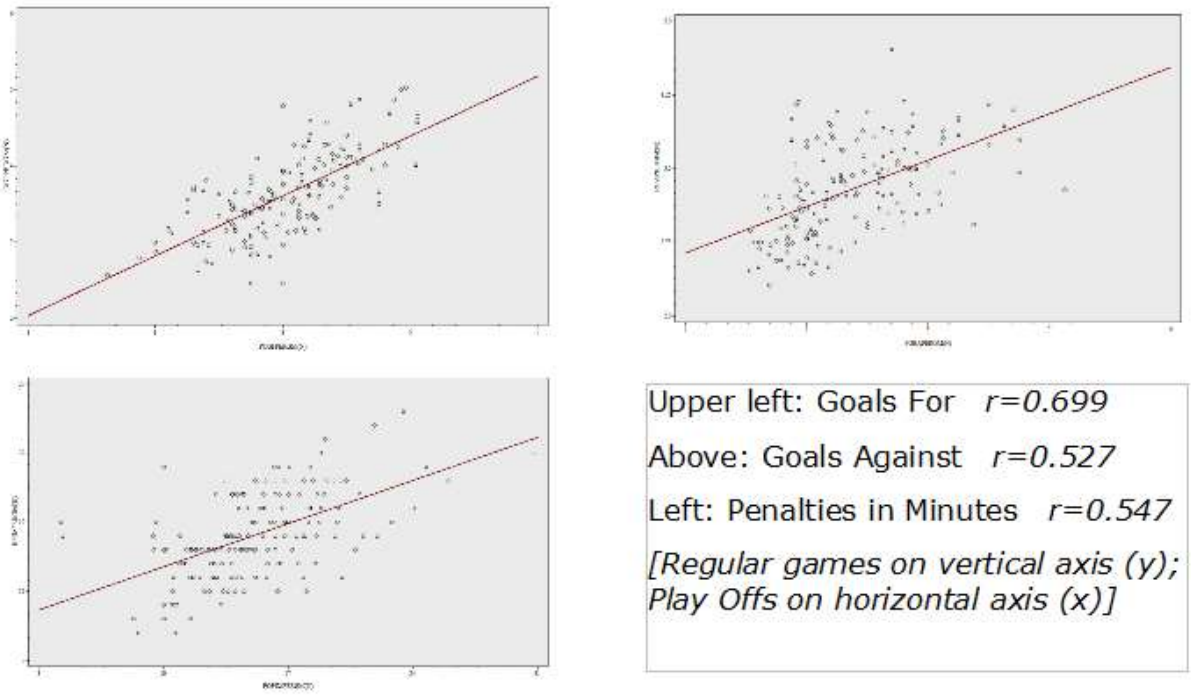

Figure 3: Correlations for Regular Season and Play Offs: per Game GF, GA, and PIM

Regressions

As a purely academic exercise, I developed a regression model to determine which teams are likely to make it to which round. After examining various correlations, I used two year-end statistics for the regular season to "predict" which play off round each team might make it to.

Based on that, the predictions are shown in Table 7. The regression model could not predict the two teams in the finals. I ran the regression equation with the 10 seasons in this study and had some success with predicting play off rounds. The equation was accurate, in predicting play off spots for all 10 seasons, 48.8% of the time, over-estimated 18.8%, and under-estimated 32.5%. The model's accuracy increases to 56.9% by including teams which made the finals within the round 3 group.

The model was then applied to the 2016/2017 season (which was not included in the 10-season data) (Table 7). The model was able to predict, with 60% accuracy, which round each team would make. There are four each for over and under estimations. The model suggests that the energy and coaching developed during a team's regular season performance carries over into play offs. But, because the model is missing some prediction accuracy (albeit better than flipping a coin!), it is missing that "something else" that may be difficult to capture in the data, that something else that is beyond measurement. A mix of art and luck and skill and team-work.

Table 7: Play Off Predictions for 2016/2017 Season		
City	Model Prediction Play Off Round (1)	Actual Play Off Round Achieved (1)
Barrie	0	0
Guelph	0	0
Niagara Falls	0*	1
North Bay	0	0
Sudbury	1	1
Ottawa	1	1
Saginaw	1**	0
Sarnia	1	1
Flint	1	1
Kitchener	1	1
Kingston	1*	2
Hamilton	2**	1
Oshawa	2	2
Peterborough	2*	3
Mississauga	2*	3 (+finals)
Windsor	2**	1
Sault Ste. Marie	2	2
London	3**	2
Owen Sound	3	3
Erie	3	3 (+finals)
(1) Prediction and Actual codes: 0=did not make/no prediction for play offs; 1=Round 1 (1/8th finals); 2=Round 2 (1/4 finals); 3=Round 3 (semi-finals); *under estimated; **over estimation; Source for round achievement: http://www.sportstats.com		

Summary

This study looked at several year-end statistics and set out to develop a basic model to predict teams' play off potential. PIMs are not correlated with either GF or GA. GF and GA *are* correlated for regular season play and were used to develop a basic model to "predict" play off achievement. The model was able to predict likely play off achievement for teams (2016/2017 season play offs) with 60% accuracy.

Further work on the model might increase accuracy but will require analysis at a deeper level than presented in this study.

For more information, visit my Website or contact the author directly.