Product Recall Notice for ‘Shot Quality’

Data quality problems with the measurement of the quality of a hockey team’s shots taken and allowed

Copyright Alan Ryder, 2007
Hockey Analytics
www.HockeyAnalytics.com
Introduction

In 2004 I outlined a method for the measurement of “Shot Quality”¹. This paper presents a cautionary note on the calculation of shot quality.

Several individuals have enhanced my approach, or expressed the outcome differently, but the fundamental method to determine shot quality remains unchanged:

1. For each shot taken/allowed, collect the data concerning circumstances (e.g. shot distance, shot type, situation, rebound, etc) and outcome (goal, save).

2. Build a model of goal probabilities that relies on the circumstance. Current best practice is that of Ken Krzywicki involving binary logistic regression techniques². I now use a slight variation on that theme that ensures that the model reproduces the actual number of goals for each shot type and situation.

3. For each shot apply the model to the shot data to determine its goal probability.

4. Expected Goals:  $\text{EG} = \text{the sum of the goal probabilities for each shot for a given team.}$

5. Neutralize the variation in shots on goal by calculating
   Normalized Expected Goals:  $\text{NEG} = \text{EG} \times \text{Shots}/ \text{League Average Shots For/Against}.$

6. Shot Quality:  $\text{SQ} = \text{NEG} / \text{GA}<$
   (\text{GA}< = \text{League Average Goals For/Against}).

I have called the result $\text{SQA}$ for shots against and $\text{SQF}$ for shots for. I normally talk about a team’s shot overall quality but one can determine shot quality any subset of its total shots (e.g. individual goaltenders, individual shooters, short handed, Tuesday nights, etc.).

The way to interpret an SQA of 1.050 is that the quality of shots allowed by the team is such that it would result in 5% more goals than a team with an SQA of 1.000, all other things being equal. An SQF of 1.05 means that the quality of shots taken by the team is such that it would result in 5% more goals than a team with an SQF of 1.000, all other things being equal.

Goal Prevention

SQA gives us a measure of shot quality that is independent of the number of shots on goal and independent of goaltending.

Before shot quality measurement was conceived goal prevention looked like a simple model:

\[ GA = \text{SOG} \times (1 - \text{SV}) \]

We knew that Shots on Goal (SOG) was a defensive responsibility, so we have historically attributed it to the skaters on a team. We knew that Save Percentage (SV) was principally a goaltender statistic so we have historically attributed it to goaltenders. But shot quality gives us a better model:

\[ GA = \text{SQA} \times \text{SOG} \times (1 - \text{SQNSV}) \]

where SQA is the Shot Quality Allowed Index and SQNSV is the Shot Quality Neutral Save Percentage.

To calculate SQNSV observe that the two models both give us goals against. That means:

\[ \text{SOG} \times (1 - \text{SV}) = \text{SQA} \times \text{SOG} \times (1 - \text{SQNSV}), \]

or
\[ (1 - \text{SV}) = \text{SQA} \times (1 - \text{SQNSV}), \]

which means
\[ \text{SQNSV} = \frac{1 - (1 - \text{SV})}{\text{SQA}} \]

In this model we attribute both SQA and SOG to the defense and SQNSV to goaltending. Clearly SQNSV is a better measure of the goaltender’s contribution to team success than is SV. It represents the save percentage one would expect with no variation in shot quality from team to team.

As SQA and SOG are attributed to defense, it sometimes makes sense to combine them as SQNSOG (Shot Quality Neutral Shots on Goal), a kind of ‘wind chill’ calculation for goaltenders.

Goal Scoring

The same logic applies to our understanding of offense:

\[ \text{GF} = \text{SQF} \times \text{SOG} \times \text{SQNSH} \]

where
\[ \text{SQNSH} = 1 - \frac{\text{SH}}{\text{SQF}} \]

but the interpretation is different. On offense SOG describes the ‘Decision’ to shoot, SQF describes the ‘Circumstances’ of the shot and SQNSH (Shot Quality Neutral Shooting Percentage) describes the ‘Execution’ of the shot attempt.
SQF is a strange beast. My research has indicated that GF is highly correlated to SOG and SQNSH but nearly uncorrelated with SQF. The only way this makes sense to me is that, on offense, the circumstances of a shot are much less impactful than execution. Note that this is not the case with SQA, which explains a fair bit of goal prevention.

The Problem

I have been mainly interested in SQA. Since the time I first started looking at shot quality the New Jersey Devils have lead the league. Always. And this has concerned me. Greatly.

The problem, you see, is the way that the data is collected. In my Shot Quality paper I made the following observation:

The source of these game summaries is the NHL’s Real Time Scoring System (RTSS). RTSS scorers have a tough job to do, recording each on ice “event” and player ice time. When it comes to a shot, the scorer records the shooter, the distance and the shot type by tapping several times on a screen. The time is recorded by the system based on one of these taps. Distance is captured by a tap on a screen resembling the rink. The system calculates the distance.

All of this happens pretty quickly. The highest priority is the shooter, as this data does get summarized and published. In the heat of battle, it is easy to get the time, shot type and distance wrong. The database clearly has embedded errors. There are shots that are impossibly close together in time. There are “wrap” shots from 60 feet. There are “tip in” shots from 60 feet. There are likely to be other coding errors (slap shots coded as wrist shots). It is easy to imagine that the record of distance is off, at least by small amounts. It is easy to imagine that the record of distance is off, at least by small amounts. It is easy to imagine that two different scorers would give us two different records of the same event.

I have been worried that there is a systemic bias in the data. Random errors don’t concern me. They even out over large volumes of data. I seriously doubt that the RTSS scorers bias the shot data in favour of the home team. But I do think that it is a serious possibility that the scoring in certain rinks has a bias towards longer or shorter shots, the most dominant factor in a shot quality model. And I set out to investigate that possibility.

The Study

In statistics ‘paired’ data is a powerful thing. Conduct an experiment and measure its results. Change one variable and repeat the experiment and, viola, you have paired data.

Thanks to Ken Krzywicki’s tireless efforts his 2006-07 (‘2007’) shot data set captures the home and road team. If we study shot quality for home and road games separately we have paired data and should be able to identify any systemic bias in the measurement of shot quality.
SQA

To the right is a summary of the home and road SQAs of NHL teams in the 2007 season. I have sorted the data from the highest to lowest differential.

And sure enough we have a smoking gun. On the road the New York Rangers look like an average defensive team with an SQA factor of .982. But at home the Blueshirts look awful with an SQA of 1.314. It is difficult to believe that the Rangers allow shots that are 33% more dangerous at home. It is more likely that RTSS scorers in the Big Apple have a certain myopia that causes them to record shot distances too short.

Without the help of statistical inference techniques it is difficult to know whether this is a meaningful deviation.

A crude test is to see if any of the data lies outside of two standard deviations of the mean. The standard deviation of this difference data is 0.094. This means that the Rangers data is very suspicious. It also means that the data from the left coast of Florida is potentially problematic as well.

Upon further inspection we can see that the standard deviation of road SQAs is 0.043 whereas the standard deviation of home SQAs is 0.093. This is further evidence of a reporting bias.

When one studies measurement errors in an unbiased measurement process the errors have a normal distribution. So the next test we would want to apply is a test of normality to the difference column. The Anderson-Darling test of normality tells us that this data has about a 7% chance of being normal. This is not conclusive but suggests that there is something non-random in the home-road differentials.
SQF

To the right is a summary of the home and road SQFs of NHL teams in the 2007 season. I have again sorted the data from the highest to lowest differential.

And there is the same smoking gun. On the road the Rangers look like a below average offensive team with an SQF factor of .969. Yet at home New York looks superhuman with an SQF of 1.377. It is very difficult to believe that the Rangers take shots that are 41% more dangerous at home. And, again, it is more likely that RTSS scorers in the Big Apple have myopia (recording shot distances too short).

San Jose ranked third on the SQA list and second here. Carolina was second on the SQA list and fourth here. Tampa Bay was last on the SQA list and third last here. Buffalo flipped this profile (third last on this list and last in SQA). The correlation of the rankings is quite high.

The standard deviation of the difference data is 0.107. This means that the Rangers data is very suspicious. The standard deviation of road SQFs is 0.064 whereas the standard deviation of home SQFs is 0.106. The evidence of a reporting bias continues to mount. The Anderson-Darling test of normality, applied to the difference column, tells us that this data has about a 0.2% chance of being normal, the New York outlier being the principal problem. The case is closed.

Combining the Data

Each of the analyses above looks only at one end of the rink. If we add the home-road differences for SQA and SQF we can get a sense of the systemic bias.

That data is presented below and it looks systemically biased to the naked eye. The Anderson-Darling test of normality tells us that this data has about a 0.9% chance of being normal. Notice that the signs of the first two columns match almost uniformly in the top and bottom 10 teams. This is a vivid picture of reporting bias.

There is clearly a large problem here. There is good chance that reporting of up to two thirds of the teams is materially off in its measurement of shot distance or type.
The principal driver of a shot quality model is shot distance (a shorter shot having a greater chance at becoming a goal). Let’s have a look at shot distance (both home and visiting teams) for certain reporting cities:

<table>
<thead>
<tr>
<th>Reporting City</th>
<th>Even Handed</th>
<th>Power Play</th>
<th>Short Handed</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY Rangers</td>
<td>25.9</td>
<td>24.0</td>
<td>30.5</td>
<td>25.6</td>
</tr>
<tr>
<td>San Jose</td>
<td>30.8</td>
<td>28.4</td>
<td>43.3</td>
<td>30.8</td>
</tr>
<tr>
<td>Carolina</td>
<td>32.2</td>
<td>33.4</td>
<td>33.0</td>
<td>32.5</td>
</tr>
<tr>
<td>St. Louis</td>
<td>30.6</td>
<td>29.8</td>
<td>38.8</td>
<td>30.7</td>
</tr>
<tr>
<td>League Average</td>
<td>34.1</td>
<td>32.8</td>
<td>43.9</td>
<td>34.2</td>
</tr>
<tr>
<td>Dallas</td>
<td>36.5</td>
<td>35.8</td>
<td>39.1</td>
<td>36.5</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>35.7</td>
<td>35.6</td>
<td>38.0</td>
<td>35.7</td>
</tr>
<tr>
<td>Buffalo</td>
<td>34.8</td>
<td>33.5</td>
<td>45.6</td>
<td>35.0</td>
</tr>
<tr>
<td>Tampa Bay</td>
<td>33.6</td>
<td>33.4</td>
<td>35.4</td>
<td>33.6</td>
</tr>
</tbody>
</table>

It is easy to see why the shot quality reported out of Madison Square Garden is so high.

The second largest influence on shot quality is situation. A team that both takes and draws a considerable number of penalties will record shot quality, at both ends of the rink, which is higher than average (and vice versa). It is unlikely that situation is incorrectly reported in the RTSS data. And we have no way of verifying this information in any case.

The last driver of shot quality is shot type. For a shot of a given distance a robust shot quality model will assign differing goal probabilities to shot types tip-in, slap, snap, wrist, backhand and wrap (listed in declining order of danger). I think that wrist and snap shots are easily interchanged. A wrap shot is really just a circumstance of a wrist or snap shot and could easily be misreported. I suspect that tip-ins are more aggressively reported by some scorers and less so by others. I have not investigated the potential for misreporting shot type.

Shot distance certainly seems part of the problem of reporting bias. Shot type may also be a source of error. For the moment the source of the errors does not really matter.
The Implications

There are two standard uses of shot quality models.

The first is to properly balance our view of the relative impact of defense and goaltending. Without adjustment for the reporting effects we would conclude that goaltenders in places like Manhattan, Calgary and San Jose are hard done by their woeful defense and we would make an incorrect adjustment to get to their shot quality neutral save percentages in an attempt to compensate. Likewise we would conclude that goaltenders in places like Tampa Bay and Los Angeles are benefiting from superior defense, and dial down their SQNSVs when in fact they are not so fortunate. Below is an indication of the potential impact of this on selected goalies, using road SQA’s rather than overall SQAs:

<table>
<thead>
<tr>
<th>Goaltender</th>
<th>Team</th>
<th>Unadjusted</th>
<th>SQNSV based on overall SQA</th>
<th>SQNSV based on road SQA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Henrik Lundqvist</td>
<td>NYR</td>
<td>.917</td>
<td>.927</td>
<td>.915</td>
</tr>
<tr>
<td>Evgeni Nabakov</td>
<td>SJS</td>
<td>.914</td>
<td>.914</td>
<td>.907</td>
</tr>
<tr>
<td>Cam Ward</td>
<td>CAR</td>
<td>.897</td>
<td>.902</td>
<td>.893</td>
</tr>
<tr>
<td>Ryan Miller</td>
<td>BUF</td>
<td>.911</td>
<td>.902</td>
<td>.906</td>
</tr>
<tr>
<td>Johan Holmqvist</td>
<td>TBL</td>
<td>.893</td>
<td>.888</td>
<td>.896</td>
</tr>
</tbody>
</table>

Before this work I was getting ready to crown Lundqvist as the unheralded king of the crease …

The other frequent use is to assess expected versus actual offense. Based on the SQF reporting discrepancies we would have excessive offensive expectations for teams such as the Rangers, Sharks, Blues and Hurricanes and understated expectations of teams such as the Sabres, Stars and Lightning.

The Fix

I always felt that there was substantial variation in the reporting of shots, especially distance. But I took comfort in one of the laws of large numbers – errors average out.

But we clearly have an issue of RTSS scorer bias. The problem appears to be brutal at Madison Square Garden but is clearly non-trivial elsewhere. The NHL needs a serious look at the consistency of this process.

Clearly the use of ‘overall’ SQ factors, heavily polluted by ‘rink effects’, has the potential to mislead. One obvious solution is the use of road factors only (as I have done above). This has the benefit of bringing back into play the comfort that ‘errors average out’ in a large, diverse data set. Road SQA and SQF factors for 2007 are detailed above.

The problem with the simple solution is best illustrated by example. If we use road factors only, the Rangers are never influenced by Madison Square Bias. But the Islanders, Devils, Flyers and Penguins each play about 10% of their road games in Manhattan. Each of these teams has a dose of Madison Square Bias in their data,
overstating goal probabilities at both ends. These teams have more of this than other
teams in the East and, of course, way more than the teams of the West.

Using road SQ factors is much better than using overall factors. But the best solution is
to solve for rink-neutral SQA and SQF factors. This involves solving a system of
simultaneous equations.

**Shot Quality Recall**

It’s not really a recall. Shot quality is not broken. Just don’t use it without understanding
it. Use the road factors!

Shot quality is a very powerful tool. Like any tool the user needs to understand (a) how
to use it and (b) its limitations. My method for assessing shot quality relies on the
underlying data. There is considerable potential here for “garbage in, garbage out”. But,
properly used, it remains a powerful tool.

And the Devils really are (almost) that good.