



- [Search](#)

- [Fanshots](#)
- [Fanposts](#)
- [Best of Outnumbered](#)

Judging Goalies: Should We Include PK Save Percentage?

By [Eric T.](#) [@BSH_EricT](#) on Jan 25 2012, 11:01a [18](#)

[Tweet Share on Twitter \(33\)](#) [Share Share on Facebook \(2\)](#) [Share Share with Flyers friends](#) [18 Comments](#) * [Rec Recommend this Post](#) [3](#)

I've seen a number of discussions lately about the best way to predict future goaltender performance. The analytical community showed long ago that because a goalie doesn't see very many PK shots per year, simple luck doesn't come anywhere near balancing out and a goalie's [PK Sv% bounces around almost completely randomly from year to year](#).

From that, it was natural to infer that the penalty kill just adds noise to our measurement of goalies and that we should focus on even strength save percentage (ES Sv%) instead of total save percentage. This would also presumably remove any unfair advantage a goalie gets in total save percentage by playing for a team that doesn't take many penalties. And so it became a widespread belief that ES Sv% was the best measure of goalie talent.

I personally made that argument just a few days ago, arguing that James Reimer's ES Sv% is a better predictor of his future results than his overall Sv% is. And yet when I went looking for an article that showed this directly, that made the leap from theoretical to empirical, I couldn't find any.

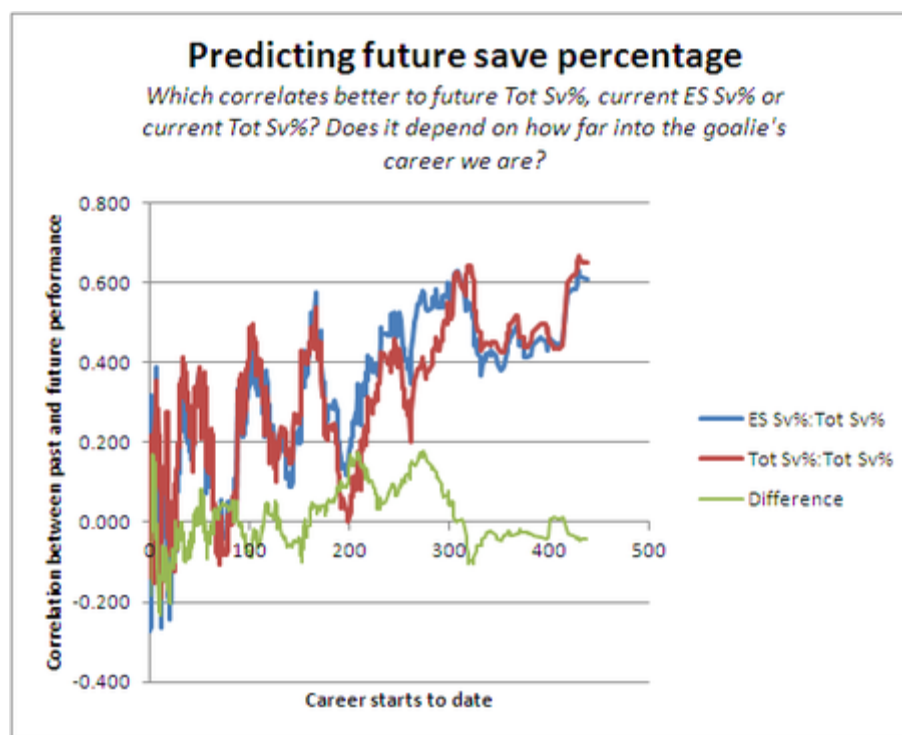
I asked around and most of the people who I talked to thought they'd seen such an article, but nobody could quite put their finger on it. Finally, Kent Wilson of Flames Nation tipped me off to an article by Tom Awad. The article basically asked the question, "if I want to see how a guy's going to do next year, which of this year's numbers should I look at?"

The answer was that whether you are trying to predict a guy's overall performance or just his even strength performance, [you are better off looking](#)

[at his total numbers for this year than his even strength numbers.](#)

This is consistent with the idea that PK Sv% and ES Sv% measure largely the same talent, and that the variability of the PK Sv% comes mostly from the small sample sizes. If that were the case, removing the penalty kill results would be kind of like removing the last five games of each year -- the goalie's performance in the last five games isn't reproducible from year to year, so the data doesn't have much value on its own, but it still helps improve the overall sample size.

I wanted to figure out how much difference this really makes, and whether the answer to that question is dependent on how much of a sample we have to work with. To answer that, I started with the even strength and overall data for every goalie who has played since ES data became available in 1997-98. At the end of each season, I logged each goalie's career totals up to that point and his totals from that date forward. I could then look at how the career numbers predicted the future as a function of how many career starts the player had.



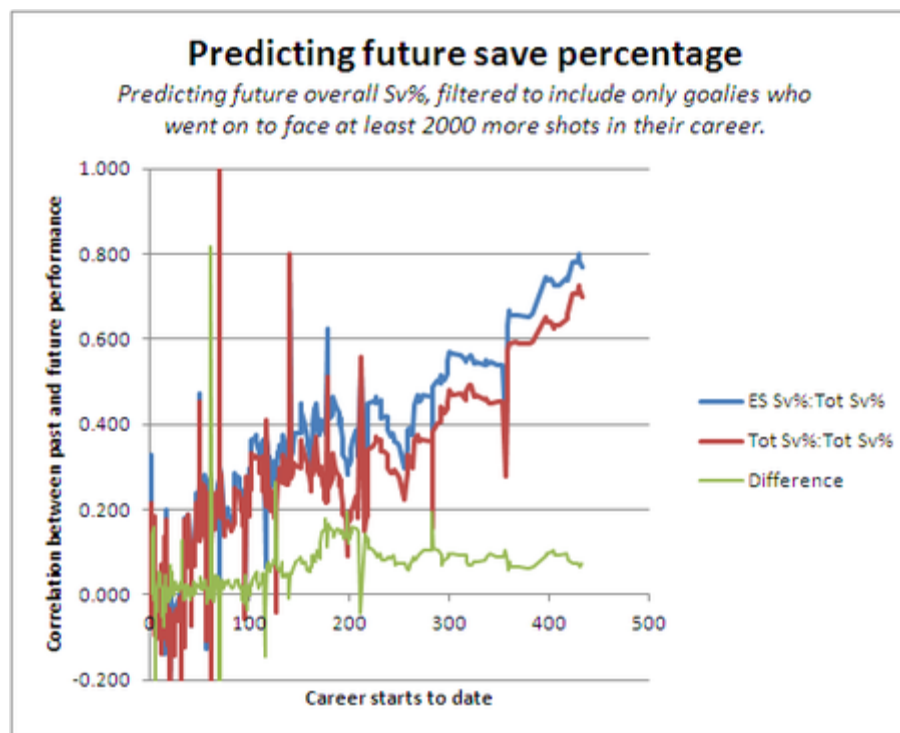
The blue curve represents how well we do at predicting a goalie's future overall Sv% by looking at his current career ES Sv% -- it trends upwards because the more games he has played so far, the more we know his true talent and the better our predictions are. The red curve shows how we do by looking at his current career ES Sv% instead when we make our predictions, and the green curve is the difference between the two.

It turns out that until the goalie has about 150-175 starts, the two measures perform almost identically in predicting a goalie's future -- it doesn't

matter whether you use career Sv% or career ES Sv% early in his career. Once a guy gets up towards 200 starts, ES Sv% does start to look like a better measure (the green curve rises above 0), although the fact that the gap closes again by 300 starts leaves me wondering if this is just a statistical quirk.

The above plot includes all of the goalies who played in the last 14 years, even the ones who didn't play much. We're trying to predict a guy's future save percentage, but if the guy only plays 6 more games, it won't really matter whether he has a 20- or 200-game history for us to look at; we'll probably lose to the randomness over that 6-game sample.

To see whether that kind of noise was what made ES Sv% look like the better predictor over large sample sizes, I filtered the data to only include guys who went on to face at least 2000 more shots and repeated the analysis. Here's what we see in that case:



Now we see a much more steady rise in predictive power as a function of games played. This is partly because we have reduced the noise, but also partly because we have introduced some selection bias: a goalie is only likely to eclipse 300 starts if he plays fairly well, and he is only likely to face 2000 more shots if he continues playing well, so the correlations get pretty strong because the filter has introduced some bias.

However, that bias probably affects both inputs equally, so the difference between the correlations shouldn't be impacted much. And now we see that the noise wasn't *causing* the rise in the green curve at higher sample sizes; it was *obscuring* it. It's probably fair to say that ES Sv% does

appear to be a better predictor than overall Sv% over large sample sizes (100+ games).

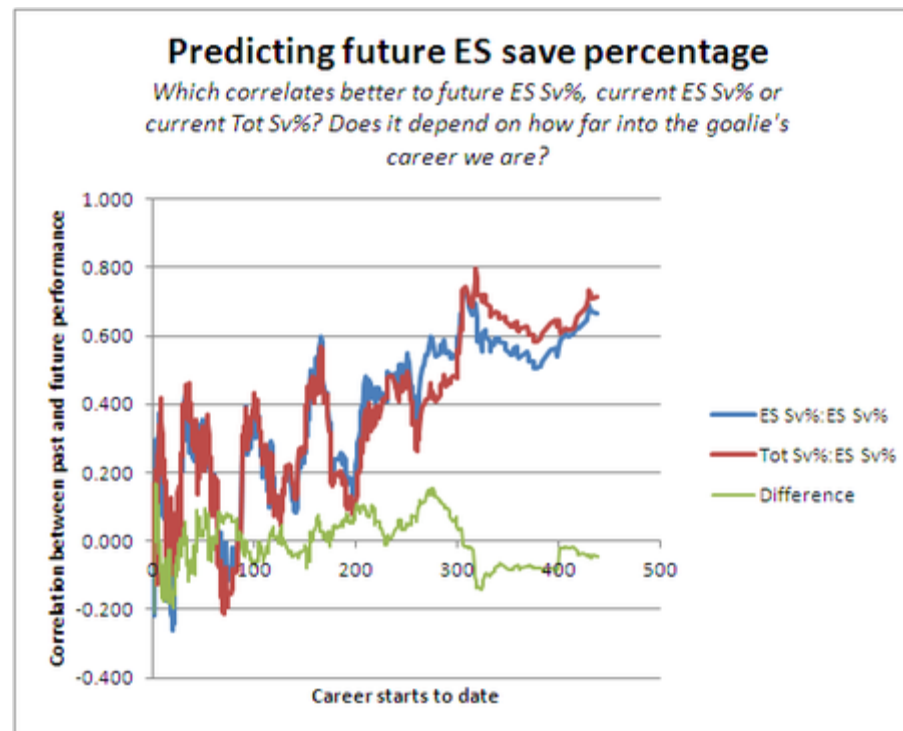
So the overall picture then is that with small sample sizes you want to include all available data, but with large sample sizes you want to focus on the most relevant data. Tom Awad showed that overall save percentage will give the best outcomes if you are using only a single year to make your predictions. Up to about 100-150 games of career numbers, overall save percentage and even strength save percentage perform similarly. And in the long run, after 150+ games, even strength save percentage is the better predictor of a goalie's future success.

★ ★ ★

Statistical post-script

I've chosen to look at the overall Sv% (rather than ES Sv%) as the measure of future performance because I think that's what we're trying to maximize when we pick a goalie -- ES Sv% comes into the conversation because we think it might be a better input, not because we think it's a more important output of the prediction. However, this could conceivably inflate the predictive power of overall Sv%; if a goalie is on a team that consistently takes a lot of penalties, his overall Sv% might be consistently lower than ES Sv% would predict.

As a check, I also looked at how ES Sv% and overall Sv% do when predicting future ES Sv%, and the results were almost exactly the same as when predicting future overall Sv%. So I'm not worried about this as a possible confounding factor. Here's the analogous plot to the first one above:



More from *Broad Street Hockey*

- [Craig Berube waited over a decade to punch Jeremy Roenick in the face](#)
- [Berube needs to ride his best players](#)
- [The Flyers need to double down](#)
- [Flyers vs. Rangers: Game 3 loss was not about 'effort'](#)
- [Gary Bettman discusses NHL expansion, shootouts, concussions and lots more](#)

[Tweet Share on Twitter \(33\)](#) [Share Share on Facebook \(2\)](#) [Share Share with Flyers friends 18 Comments](#) * Rec Recommend this Post 3

Latest News