# Data Analysis and the NHL Entry Draft: A Review

Michael Schuckers

St. Lawrence Univ., NY Statistical Sports Consulting, LLC schuckers@stlawu.edu @SchuckersM @EmpiricalSports

May 28, 2019

Where?



# NHL Entry Draft

- Annually at the end of June (Vancouver this year, 6/21&6/22)
- Method by which newly eligible players are allocated to NHL teams
- Eligible players: 18 years old on or before September 15 and not older than 20 years old before December 31 (Worldwide)
- Each team (of 31) starts with a pick in each of 7 rounds
- Team with the worst record has best chance at first pick in first round (lottery among non-playoff teams)
- After 1st round, picks are made in order of reg. season finish
- Draft picks are tradable assets

# Plan for this talk

- Motivating Example
- How to Value Draft Picks
- Which Players to Draft
- Other Draft-ish Work
- Possible Areas of Research

The 2011 NHL Draft, first six picks

• EDM: Ryan Nugent-Hopkins (F)

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- NJD: Adam Larsson (D)

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- Since (Section 1998) Strome (F)

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- NYI: Ryan Strome (F)
- OTT: Mika Zibanejad (F)

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- NYI: Ryan Strome (F)
- OTT: Mika Zibanejad (F)

The 2011 NHL Draft, first six picks

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- NYI: Ryan Strome (F)
- OTT: Mika Zibanejad (F)
- WPG: ????

Who should WPG select? Consensus Available (in order):

The 2011 NHL Draft, first six picks

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- NYI: Ryan Strome (F)
- OTT: Mika Zibanejad (F)
- WPG: ????

Who should WPG select? Consensus Available (in order):

The 2011 NHL Draft, first six picks

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- INYI: Ryan Strome (F)
- OTT: Mika Zibanejad (F)
- WPG: ????

Who should WPG select? Consensus Available (in order): Sean Couturier (F), Dougie Hamilton (D), Ryan Murphy (D), Sven Baertschi (F), Nathan Beaulieu (D), Duncan Siemens (D), Mark Scheifele (F)

The 2011 NHL Draft, first six picks

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- INYI: Ryan Strome (F)
- OTT: Mika Zibanejad (F)
- WPG: Mark Scheifele (F)

Who should WPG select? Consensus Available (in order): Sean Couturier (F), Dougie Hamilton (D), Ryan Murphy (D), Sven Baertschi (F), Nathan Beaulieu (D), Duncan Siemens (D), Mark Scheifele (F)

- EDM: Ryan Nugent-Hopkins (F)
- COL: Gabriel Landeskog (F)
- FLA: Jonathan Huberdeau (F)
- INJD: Adam Larsson (D)
- Since (Section 1998) Strome (Section 1998) Strong (Section 1998
- OTT: Mika Zibanejad (F)
- WPG: Mark Scheifele (F)
- PHL: Sean Couturier (F)
- BOS: Dougie Hamilton (D)
- MIN: Jonas Brodin (D)
- COL: Duncan Siemens (D)

• Consensus that Scheifele pick was a surprise

- Consensus that Scheifele pick was a surprise
- Some analytics about Couturier >> Scheifele (return to this)

- Consensus that Scheifele pick was a surprise
- Some analytics about Couturier >> Scheifele (return to this)
- Post draft: Couturier had 77, 46 GP (Games Played) 1st 2 yrs

- Consensus that Scheifele pick was a surprise
- Some analytics about Couturier >> Scheifele (return to this)
- Post draft: Couturier had 77, 46 GP (Games Played) 1st 2 yrs
- Post draft: Scheifele had 7, 4 GP 1st 2 yrs

- Consensus that Scheifele pick was a surprise
- Some analytics about Couturier >> Scheifele (return to this)
- Post draft: Couturier had 77, 46 GP (Games Played) 1st 2 yrs
- Post draft: Scheifele had 7, 4 GP 1st 2 yrs
- Today after 8 seasons: Couturier has 578 GP, 343 Pts, 134 Goals

- Consensus that Scheifele pick was a surprise
- Some analytics about Couturier >> Scheifele (return to this)
- Post draft: Couturier had 77, 46 GP (Games Played) 1st 2 yrs
- Post draft: Scheifele had 7, 4 GP 1st 2 yrs
- Today after 8 seasons: Couturier has 578 GP, 343 Pts, 134 Goals
- Today after 8 seasons: Scheifele has 448 GP, 371 Pts, 151 Goals

Illustrate some themes:

• Evaluation of players is hard

- Evaluation of players is hard
- and it takes time (players develop at different rates from 18yo)

- Evaluation of players is hard
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple

- Evaluation of players is hard
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric

- Evaluation of players is hard
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric
- In short, noisy and sparse data with slow feedback and

- Evaluation of players is hard
- and it takes time (players develop at different rates from 18yo)
- Metrics for assessment are not simple
- nor single agreed response metric
- In short, noisy and sparse data with slow feedback and
- (to teams) results incredibly important

#### What is a draft pick worth

Line of reasoning originally from NFL:

- Assign value/currency to each pick
- Allows for evaluating trading of picks
- Based upon historical trades
- Assigned 3000 to first pick, why?
- Created by Dallas Cowboys coach Jimmy Johnson
- "Draft Pick Value Chart" (DPVC)

#### Mathematically

For a trade trying to find some  $\hat{f}$  that we can use to such that:

$$\sum_{i\in\mathcal{I}}\hat{f}(p_i)pprox\sum_{j\in\mathcal{J}}\hat{f}(p_j)$$

where  $p_i$ ,  $p_i$  are draft pick selections for first and second team.

For example, if a team wanted to move from the 10th pick to the 5th pick, they would have to trade the 10th and the 106th pick so that the value on both sides was roughly the same. E.g.

$$\hat{f}(5)pprox\hat{f}(10)+\hat{f}(106)$$

#### Alternative DPVC

Alternatives begin in NFL

- Massey and Thaler (2006, 2013) second contracts
- Stuart (2012) player performance, career approximate value through 5 years
- Schuckers, JQAS (2012) player performance, games played

#### DPVC - Chase Stuart



#### source:

http://www.footballperspective.com/creating-a-draft-value-chart-part-ii/

Published versions:

- Awad (2010?), Goals Versus Threshold (GVT)
- Schuckers (2011), 1988-1997,LOESS, GP
- Tulsky (2012), log regression/exponential decay, Trade Values
- Sprigings (2014), Points Shares, log regression
- Burtch (2015), *F's only*, Pts/G\*Prob( > x GP)
- Schuckers (2016), 2003-09 monotonic nonparametric, 1st 7 years TOI

And others, e.g. Josh Flynn (CBJ-internal), arbitrary 1-10 scale each player.



Draft Pick Value Chart

source:

http://statsportsconsulting.com/2016/05/26/1658/

Issues with making a DPVC

• Response metric

- Response metric
- Functional form, parametric or non-parametric

- Response metric
- Functional form, parametric or non-parametric
- Career totals or subset (first 5 yrs, 7yrs, career?)

- Response metric
- Functional form, parametric or non-parametric
- Career totals or subset (first 5 yrs, 7yrs, career?)
- Data time window (2003-2010, 1970-2013,?)

- Response metric
- Functional form, parametric or non-parametric
- Career totals or subset (first 5 yrs, 7yrs, career?)
- Data time window (2003-2010, 1970-2013,?)
- Across positions or adjustments (eg. TOI for D)

# If you look at statistics and point to a column and say, 'We're drafting this guy' — have fun. I hope you're in my division. - Brian Burke, MIT Sloan Sports Analytics Conference, 2013

Source:

http:

//www.thestar.com/sports/leafs/2013/03/01/former\_toronto\_maple\_leafs\_gm\_brian\_burke\_speaks\_out\_a\_conference

Desjardins (2004-), NHL League Equivalencies (NHLe),

- Estimate  $\hat{\rho}_{\ell}$  where  $\hat{Y}_{NHL,t} = \hat{\rho}_{\ell} Y_{\ell,t-1}$
- ratio estimator
- for each league,  $\ell$ , e.g. SEL, OHL, QMJHL, NCAA
- where Y is Points generally
- t is year, t-1 previous year
- can be adjusted further by TOI
- allows a measure of league quality
- need amount of data

Extended by Rob Vollman, added by age  $\hat{\rho}_{\ell,age}$ 

#### NHLe from 2014-15

Updated Translation Factors, as of 2014-15 NHL season .80 Kontinental Hockey League (up .02) .60 Swedish Hockey League (up .05) .47 American Hockey League (up .02) .44 Western Collegiate Hockey Association (defunct, up .02) .41 National Collegiate Hockey Conference (new league) .40 Switzerland NLA (up .04) .37 Hockey-East (up .04) .35 Big-10 (new league) .32 Ontario Hockey League (up .02) .32 Central Collegiate Hockey Association (defunct) (up .02) .29 Finland SM-Liiga (down .01) .27 Western Hockey League (up .01) .26 Quebec Major Junior Hockey League (no change) .23 Eatern College Athletic Conference (down .02)

source:

http://www.hockeyabstract.com/thoughts/updatedtranslationfactors

#### coda on Scheifele

Desjardins is from Winnipeg

Hosted the most popular hockey analytics blog at the time

Panned the selection of Scheifele (over Couturier) at the time and wrote about it

Though he would argue, IMO, that is was still a bad bed, regardless of the outcome.

#### coda on Scheifele



"I know that people have come up with all kinds of reasons why Scheifele isn't really a 4-1 longshot to be better than Couturier, but I'm skeptical - this isn't too far away from hitting on 18 at the blackjack table on a hunch." - Gabe Desjardins aka Hawerchuk source: https://www.arcticicehockey.com/2011/7/22/2287979/scheifele-vs-couturier-redux

Lawrence and Weissbock (2015), Prospect Cohort Success

- Nearest neighbors approach
- Inputs: League, Age, Points, Height
- Generated comparable players (cohort)
- Pct Success (> 200 GP) among cohort
- Out of sample testing outperformed teams

Hired by Florida Panthers

Schuckers (2016), Draft By Numbers

- GAM approach, log link (Poisson)
- Response TOI or GP (both)
- Inputs:
  - Scouting via Central Scouting Rankings
  - Previous season performance PPG or SVPct
  - Demographics: Height, Weight, Position
  - League from previous season + interactions
- Out of sample rank correlation outperform NHL teams

Seppa, Schuckers and Rovito (2017), Text Mining to Improve Player Evaluation

- Used text of scouting reports on players
- Sentiment of players
- Small subset of players, only forwards
- Response: AHL performance (G+1A)/60
- Out Of Sample improvement on NHLe
- Promising but limited

Seppa and Rovito now work at NHL league office in NYC

Speak (2017), Draft Expected Value,

- NHLc xPS
- NHLc is probability of > 200 NHL GP
- NHLc modelled via binary classifier
- xPS is expected point shares (hockey-reference)
- xPS is regression(?) on similar metrics to Schuckers (2016)

Schulte, Liu & Li (2018), Model Trees for Exceptional Player Identification

- Decision Trees
- Response: boolean GP > 0(zero)
- Inputs like Schuckers (2016) ...
- ... plus more draft year predictors
- Outperforms out of sample NHL draft order on Spearman corr.

#### Other Draft-ish Work of Note

- Nandakumar
  - What Does It Mean to Draft Perfectly
  - Survival Analysis of NHL Prospect Timelines
  - Evaluating Overage NHL Prospects and the Teams That Draft Them
- Hohl: SEAL-Adjusted Scoring and why it matters for prospects (SEAL = Secondary Assist and League)
- Sprigings: NHL Draft Probability Tool
- Schuckers & Argeris: Estimating ROI for NHL Team Scouting
- @MannyElk, cryptic work on draft with NHLe for every league (Bayesian?) or other ML
- Tulsky eTOI, estimated TOI for leagues that don't record TOI
- White and Schuckers, Shiny App for NHL Draft http://shiny.stlawu.edu:3838/hockey/NHLDraftApp/

And more online ...

©M. Schuckers (SLU)

#### **Future Directions**

Some possibilities

- Women's Hockey
- NHL Combine(noun) Data (happening now in BUF)
- More ML, Stat Learning, Clustering, SVM, N-Nets
- Functional Data Analysis Career Trajectories
- Modelling (implicit) value of players for draft trades
- Discrimination Drafting and Development
- Update Fyffe's Central Scouting Integrator
- Adjusting for team %in% league quality
- Zero-Inflated Poisson Regression

#### OTTHAC19

#### SAVE THE DATE 5<sup>th</sup> Ottawa Hockey Analytics Conference #OTTHAC19 November 15-16, 2019 Carleton U, Ottawa, ON

#### The End

# Thanks schuckers@stlawu.edu

#### Resources

- Data: www.Eliteprospects.com
- Data: www.hockey-reference.com
- Articles: https://metahockey.com/publications/articles

#### 2011 Consensus draft Links

https://www.eliteprospects.com/draft/nhl-entry-draft/2011 http://www.nhl.com/ice/page.htm?id=69706 http://www.nhl.com/ice/page.htm?id=69703 http://www.nhl.com/ice/page.htm?id=69704 https://www.draftsite.com/nhl/mock-draft/2011/s/ https:

//edmontonjournal.com/sports/hockey/nhl/cult-of-hockey/ 2011-nhl-entry-draft-handicapping-the-top-30

Draft Pick Value Charts: https: //myslu.stlawu.edu/~msch/sports/Schuckers\_NHL\_Draft.pdf https://www.broadstreethockey.com/2013/4/25/4262594/ nhl-draft-pick-value-trading-up https://www.sportsnet.ca/hockey/nhl/ analyzing-value-nhl-draft-picks/ https://repository.upenn.edu/cgi/viewcontent.cgi?article= 1029&context=joseph\_wharton\_scholars http://statsportsconsulting.com/2016/05/26/1658/ http://donttellmeaboutheart.blogspot.com/2014/11/ nhl-draft-pick-value-chart.html

Who to Draft http://www.behindthenet.ca/projecting\_to\_nhl.php http://www.behindthenet.ca/blog/labels/Equivalencies.html http://hockeyanalytics.com/Research\_files/League\_Equivalencies.pdf https://canucksarmy.com/2015/05/26/draft-analytics-unveiling-the-prospect-cohort-success-model/ https://canucksarmy.com/2015/05/18/draft-theory-on-risk-and-reward-54e5db85-f1c1-4291-9268-830220dff34a/ http://statsportsconsulting.com/2016/05/26/1658/ http://www.hockeyabstrat.com/thoughts/updatedtranslationfactors

https://oilersnation.com/2015/01/23/development-and-the-200-game-mark/

#### More Who to Draft http://statsportsconsulting.com/main/wp-content/uploads/ TextMiningScoutingNHLDraftAnalyticsFeb2017.pdf https://www.arcticicehockey.com/2011/7/20/2284326/ on-post-junior-success-or-couturier-vs-scheifele https://www.arcticicehockey.com/2011/7/22/2287979/ scheifele-vs-couturier-redux

- http://prospect-stats.com/blog/DEV
- https://arxiv.org/pdf/1802.08765.pdf

Additional Links https://hockey-graphs.com/2017/03/28/ exploiting-variance-in-the-nhl-draft/ https://repository.upenn.edu/joseph\_wharton\_scholars/26/ https://hockey-graphs.com/2016/06/15/ seal-adjusted-scoring-and-why-it-matters-for-prospects/ https://hockey-graphs.com/2016/06/08/ nhl-draft-probability-tool/#more-9010