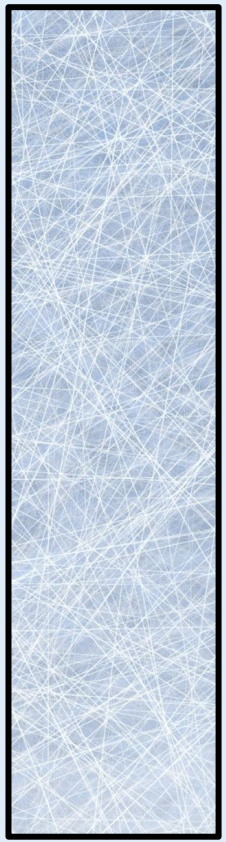
Survival Analysis of NHL Prospect Timelines

Namita Nandakumar
Wharton School, University of Pennsylvania
The Athletic Philly, Hockey Graphs
@nnstats

Wharton Sports Business Summit 2017



What do we know about hockey?

- Well, we know some stuff about what players do after they make it to the NHL.
- We know a bit about the variables that affect teams' draft decisions.
- We know almost nothing about the factors that drive what happens in between.

Draft → ¯_(ツ)_/¯ → NHL

The NHL Draft: An Overview

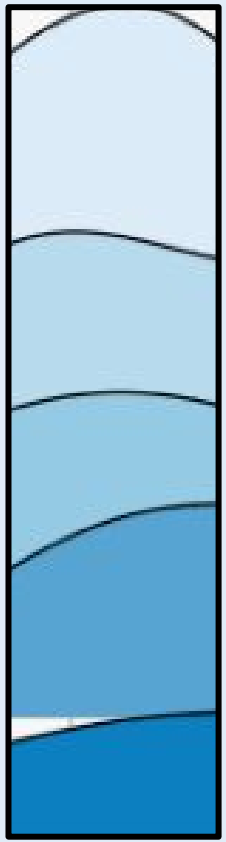
- There were 30 (now 31!) NHL teams that are allotted a pick per round for 7 rounds.
- Players are eligible to be drafted at age 18.
- Prospects are drafted anywhere from Canadian junior leagues to European pro leagues.
- They usually take “*a few years*” to “*make it*” to the NHL.

What's the existing literature?

- I couldn't find much, so I wrote an article for The Athletic Philly.
- I tried to answer 2 questions:

  How long does it take for different types of prospects to make it to the NHL?

  How does this prospect timeline relate to the **value** they eventually create for their NHL teams?

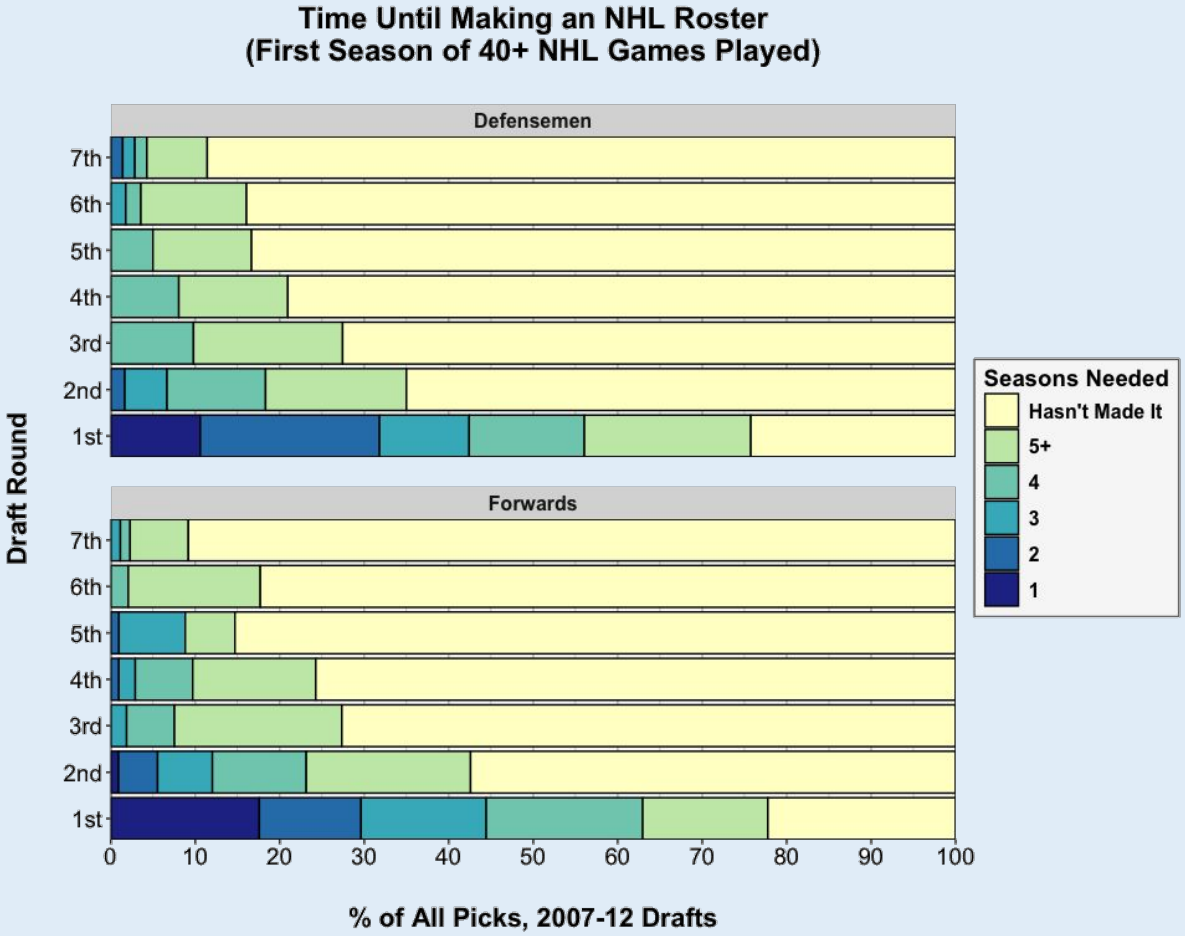


My Original Methodology

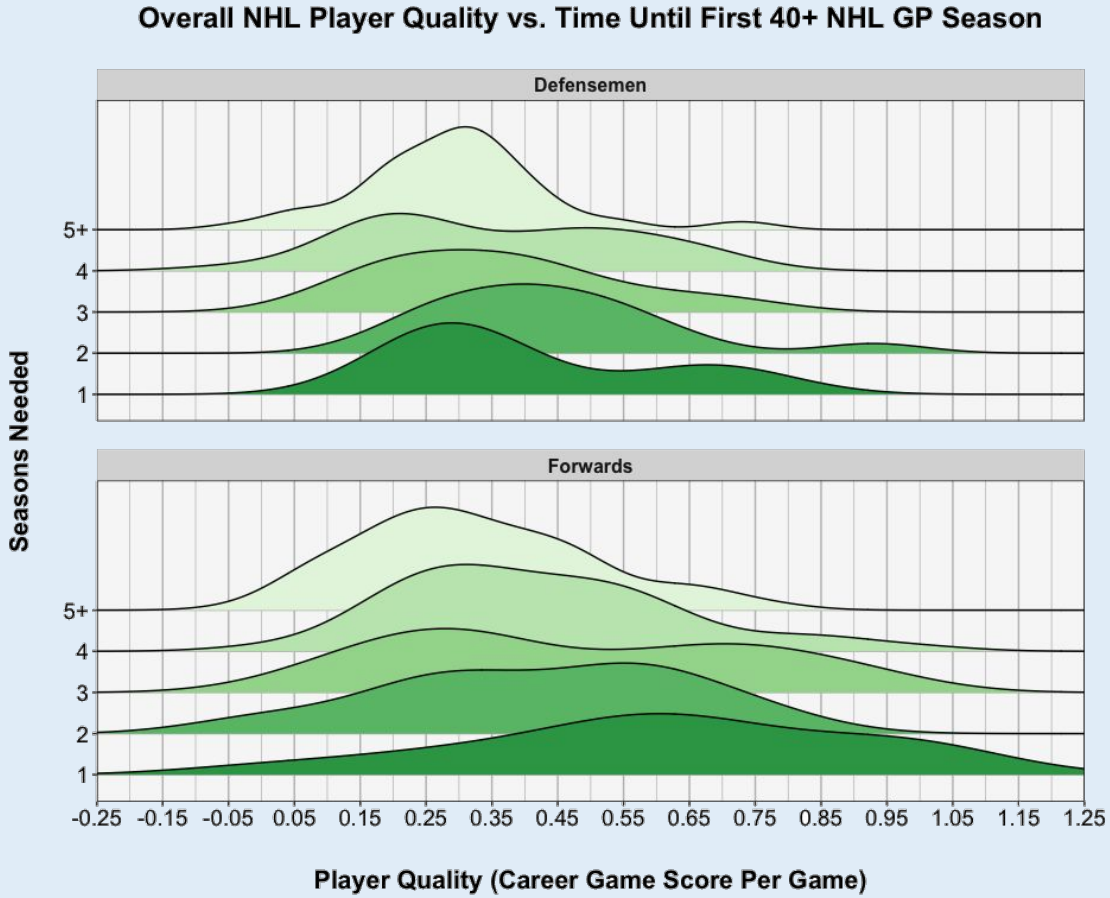
- Classify “making an NHL roster” as 40+ NHL games played in a single season*.
- Look at the distribution of prospect timelines stratified by draft round and position for the ‘07-12 drafts.
- Test for a statistically significant relationship between time until making a roster and NHL impact.

* All of these analyses are looking exclusively at skaters.

Stratification of Draft Round and Position



NHL Value Distributions by Timeline

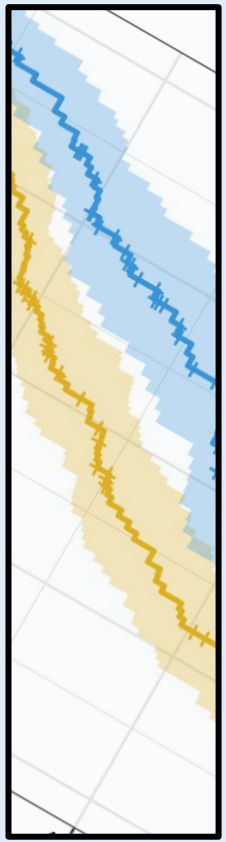


Game Score courtesy of Dom Luszczyszyn

What are some problems with my approach?

A month ago,
I thought it was a
pretty good article.

- *Limiting the Data*: I excluded recently drafted players and binned longer timelines.
- *Loss of Granularity*: We don't know when, within a season, these prospects played.
- *Arbitrary Cutoffs*: 40 games? Who cares about 40 games?
- *Undefined Effects*: We know that variables like position and draft round have effects, but what are they?



Survival Analysis

- Often used to answer questions in fields such as biostatistics and marketing.

  How long do patients live after treatment? How long do customers go before trying our products?

  How long do prospects develop before making the NHL?

- Good for dealing with right-censored data, like most recent draftees.
- Can estimate the effects of covariates like draft position and size.
- Usually a tradeoff between imposing very few assumptions vs. ease of interpreting and predicting outcomes.

What does the data look like?

1st Game Benchmark:

time	status	year	round	overall	team	player
1	1	2015	1	1	EDM	CONNOR.MCDAVID
1	1	2015	1	2	BUF	JACK.EICHEL
84	1	2015	1	3	ARI	DYLAN.STROME
83	1	2015	1	4	TOR	MITCHELL.MARNER
1	1	2015	1	5	CAR	NOAH.HANIFIN
82	1	2015	1	6	NJD	PAVEL.ZACHA
83	1	2015	1	7	PHI	IVAN.PROVOROV

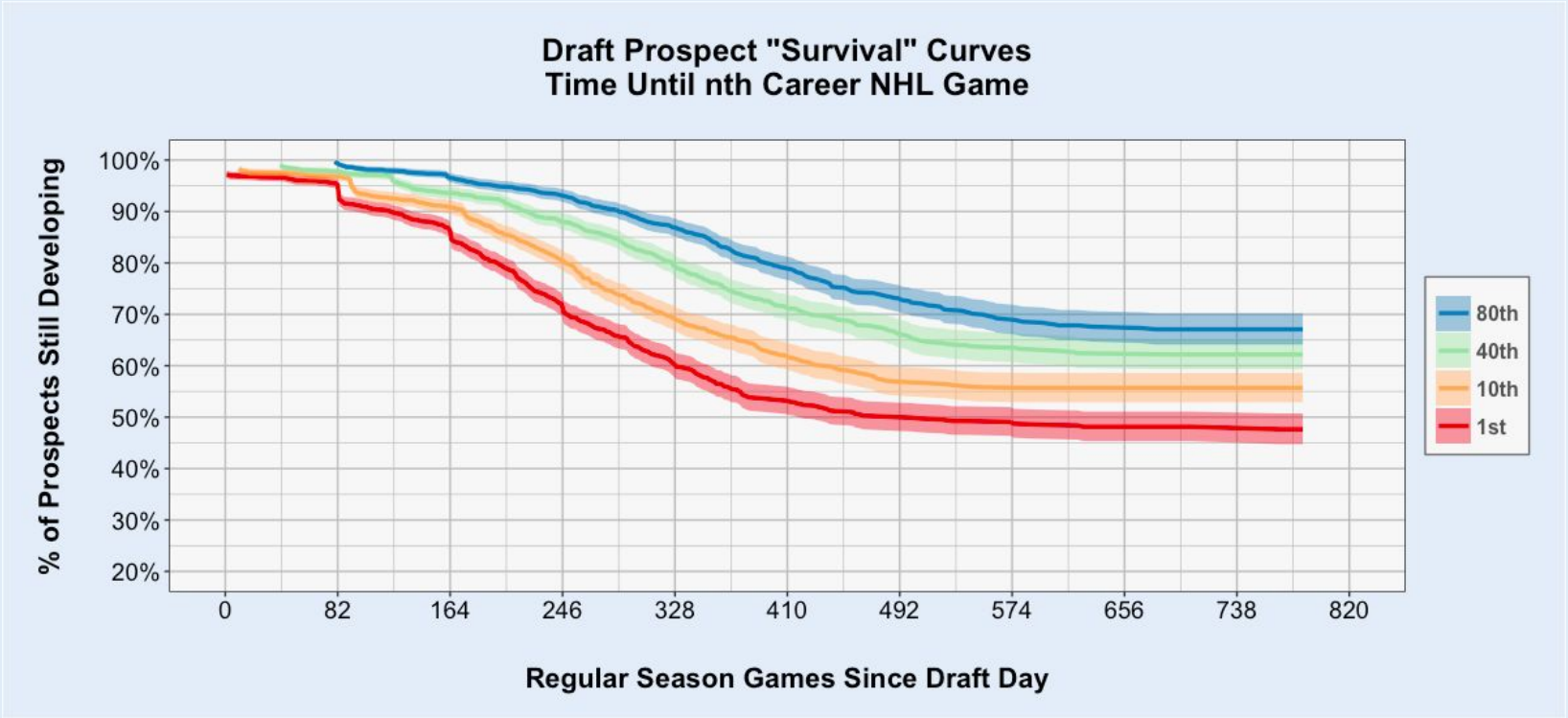
80th Game Benchmark:

time	status	year	round	overall	team	player
117	1	2015	1	1	EDM	CONNOR.MCDAVID
81	1	2015	1	2	BUF	JACK.EICHEL
164	0	2015	1	3	ARI	DYLAN.STROME
164	0	2015	1	4	TOR	MITCHELL.MARNER
83	1	2015	1	5	CAR	NOAH.HANIFIN
164	0	2015	1	6	NJD	PAVEL.ZACHA
162	1	2015	1	7	PHI	IVAN.PROVOROV

I decided to evaluate time until 1st, 10th, 40th, and 80th career games.

Time: regular season games since draft day

Status: 1 if entry into the NHL was observed at that time,
0 if it hasn't happened by the end of the 2016-17 regular season

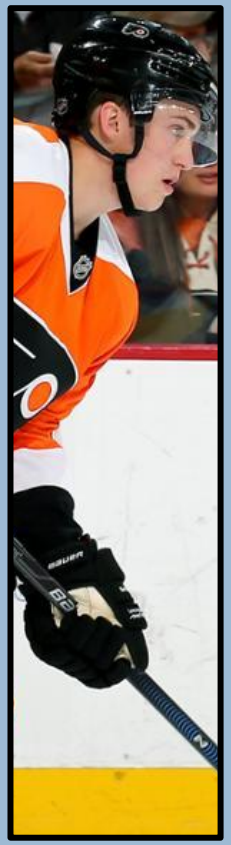


Kaplan-Meier Curves: All Skaters

Cox Proportional Hazards Model

In 30 seconds.

- Semi-parametric.
- Can estimate the multiplicative effects of covariates.
- (Relative) ease of interpretation.
- Using a baseline hazard estimator (Breslow), we can compute “survival” curves for individual players.



Before We Discuss Covariates...

Remember that the process of prospect entry into the NHL is governed by two distinct features:

- Player quality + performance at lower levels.
- Team needs + preferences.

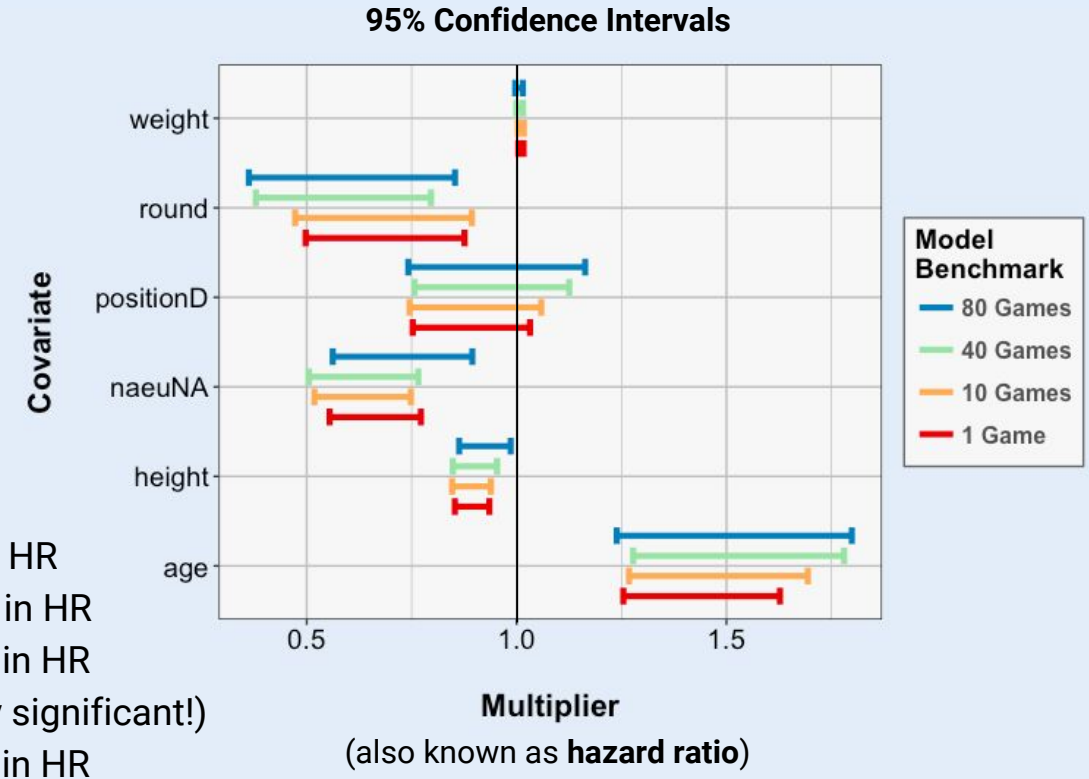
The answer to the question "*Why is this covariate value associated with prospects making it to the NHL earlier?*" can really be a mix of two answers:

- The variable is associated with better quality players.
- The variable is associated with players that teams like and/or feel that they particularly need.

Covariate Effects

hazard rate *(HR) =*
P(entering the league at time t given that you haven't by t-1)

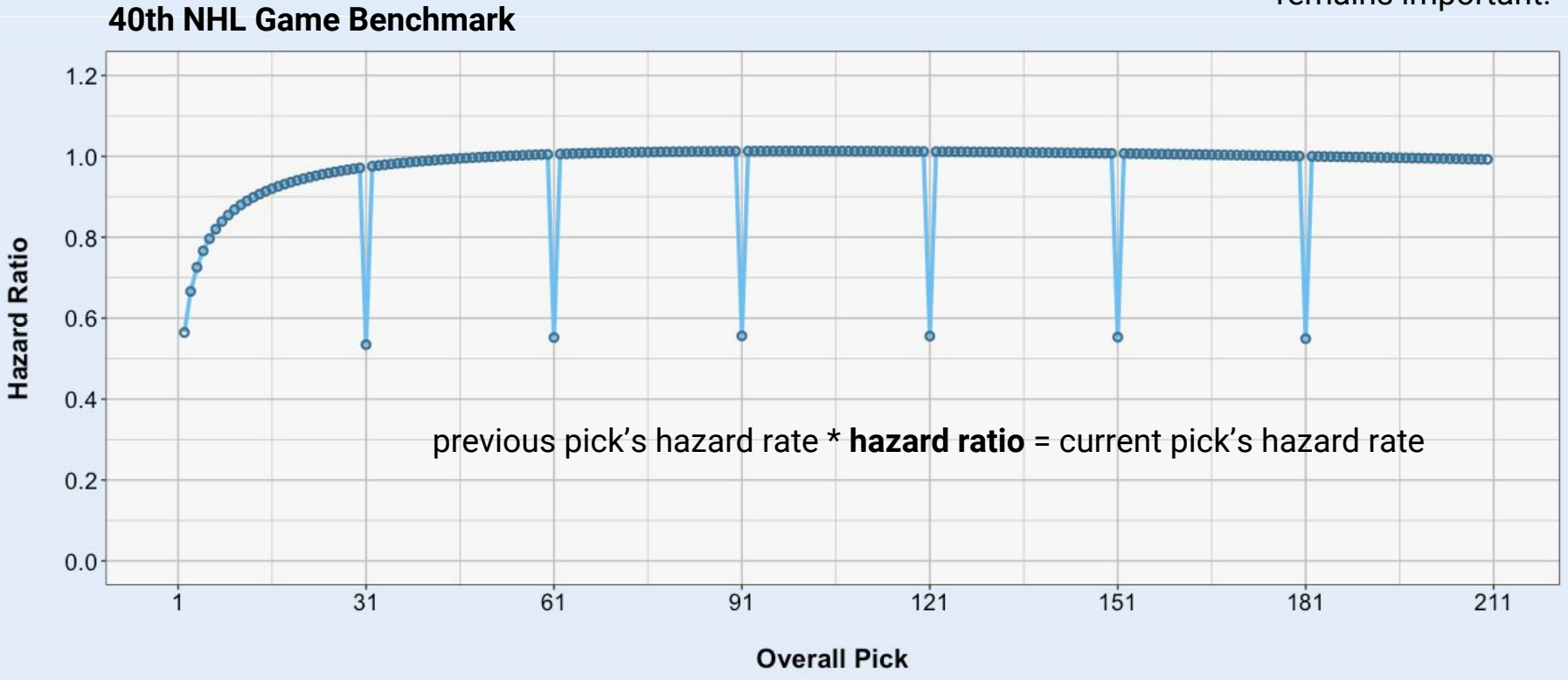
- + 1 pound heavier = ~1% increase in HR
- + 1 draft round = ~45% decrease in HR
- defensemen = ~10% decrease in HR
  (not statistically significant!)
- North American = ~35% decrease in HR
- + 1 inch taller = ~10% decrease in HR
- + 1 year older = ~47% increase in HR

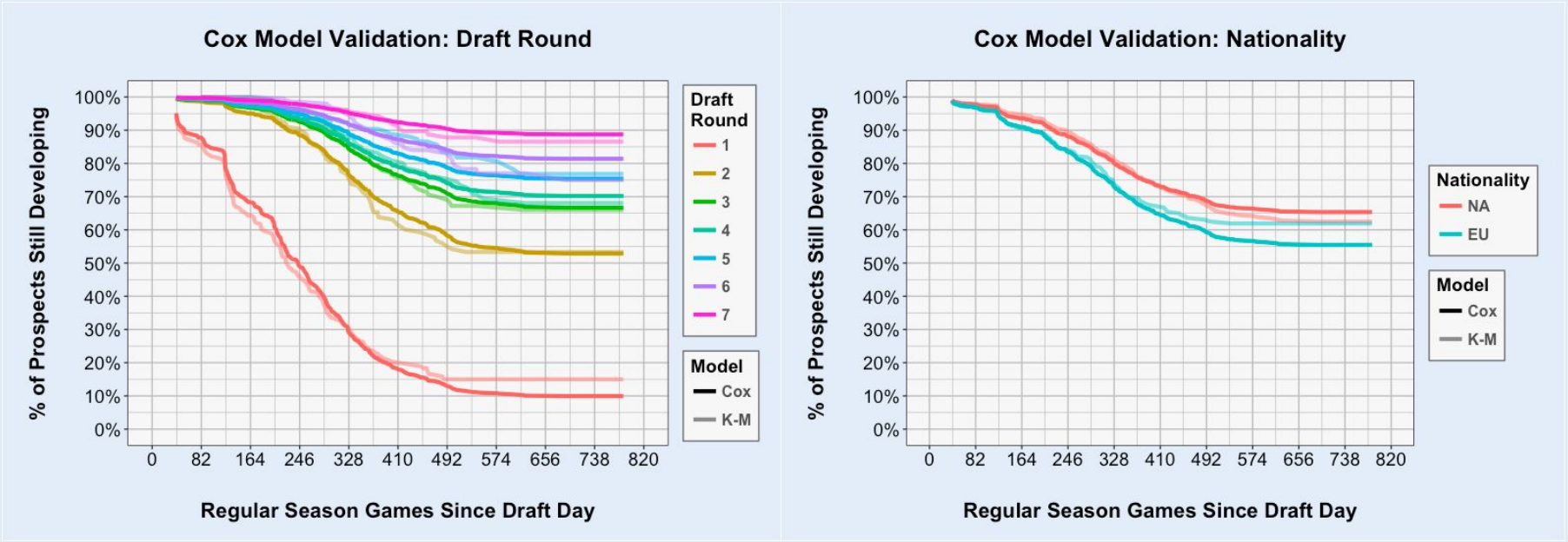


(also known as **hazard ratio**)

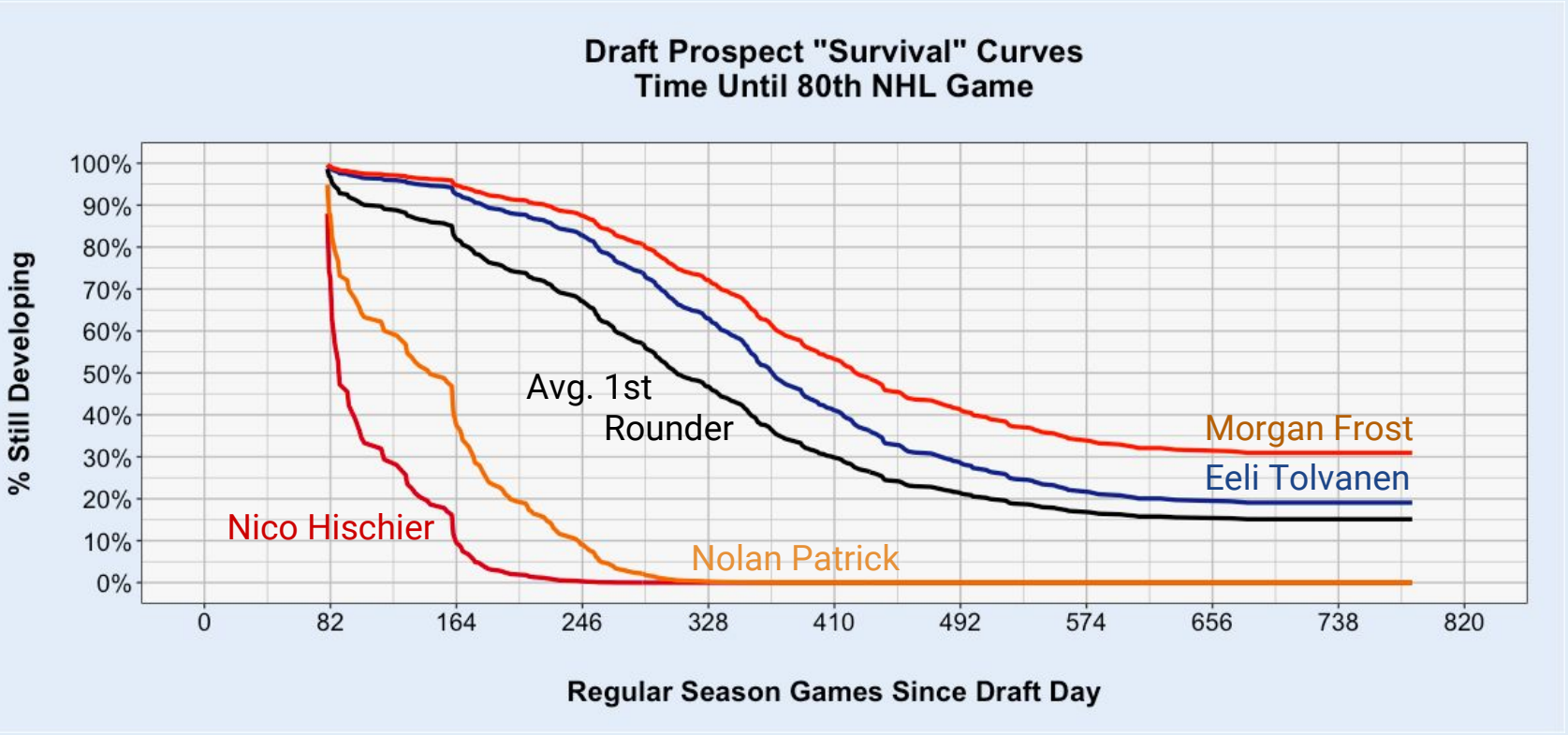
Effects of Draft Round + Pick

The effect between picks dissipates quickly, but the effect between rounds remains important!





Graphical Validation of Cox Model: 40th NHL Game Benchmark



Prospect Projections: 2017 Draft

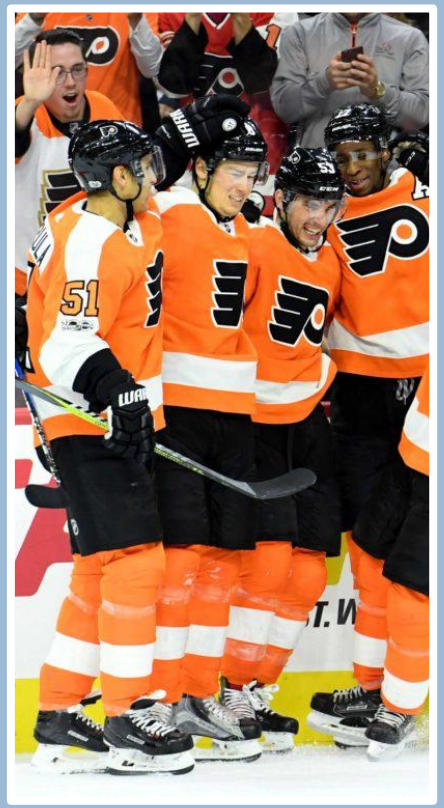
Takeaways

- Age, size, draft round and pick # have significant impacts on prospect development time.
- The anecdotal rule of longer timelines for defensemen may be overemphasized.
- Draft round gives us more information than overall pick # alone, perhaps due to prospect hierarchies within specific teams.

Additional Questions

For the future.

- Should we include additional covariates (ex. junior point production)?
- How do these NHL timeline estimates relate to eventual NHL performance?
- After a prospect makes a roster, is staying in the NHL a time-varying Markov chain?
- Which teams over- and under-season their prospects to a significant degree?



Thank you!

To all of you for listening, but in particular, to:

- Prof. Shane Jensen (Wharton Statistics) and Elliot Oblander (Wharton) for the analytical advice.
- Manny Perry (corsica.hockey) for providing me with NHL game data.

I'll be sharing slides and extensions of this work on Twitter, @nnstats.

Appendix: Cox PH Model Output

1st Game Benchmark

```
                          coef  exp(coef)   se(coef)       z Pr(>|z|)
height              -1.131e-01  8.930e-01  2.348e-02  -4.818 1.45e-06 ***
weight               1.032e-02  1.010e+00  3.124e-03   3.303 0.000958 ***
I(overall^(0.5)) -1.295e+00  2.740e-01  1.235e-01 -10.483  < 2e-16 ***
overall              9.535e-02  1.100e+00  1.457e-02   6.544 5.99e-11 ***
I(overall^(2))      -1.314e-04  9.999e-01  2.901e-05  -4.531 5.86e-06 ***
positionD           -1.267e-01  8.810e-01  8.053e-02  -1.573 0.115764
age                  3.568e-01  1.429e+00  6.643e-02   5.371 7.83e-08 ***
naeuNA              -4.240e-01  6.544e-01  8.437e-02  -5.026 5.02e-07 ***
round               -4.154e-01  6.601e-01  1.445e-01  -2.875 0.004037 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                    exp(coef) exp(-coef) lower .95 upper .95
height                 0.8930     1.1198    0.8529    0.9351
weight                 1.0104     0.9897    1.0042    1.0166
I(overall^(0.5))       0.2740     3.6493    0.2151    0.3491
overall                1.1000     0.9091    1.0691    1.1319
I(overall^(2))         0.9999     1.0001    0.9998    0.9999
positionD              0.8810     1.1350    0.7524    1.0317
age                    1.4287     0.6999    1.2543    1.6274
naeuNA                 0.6544     1.5281    0.5547    0.7721
round                  0.6601     1.5150    0.4973    0.8761

Concordance= 0.789  (se = 0.011 )
Rsquare= 0.363   (max possible= 0.996 )
```

10th Game Benchmark

```
                          coef  exp(coef)   se(coef)       z Pr(>|z|)
height              -1.149e-01  8.915e-01  2.644e-02  -4.345 1.39e-05 ***
weight               1.016e-02  1.010e+00  3.534e-03   2.876  0.00402 **
I(overall^(0.5)) -1.547e+00  2.129e-01  1.373e-01 -11.265  < 2e-16 ***
overall              1.211e-01  1.129e+00  1.638e-02   7.395 1.41e-13 ***
I(overall^(2))      -1.864e-04  9.998e-01  3.286e-05  -5.671 1.42e-08 ***
positionD           -1.187e-01  8.881e-01  8.966e-02  -1.323  0.18569
age                  3.822e-01  1.465e+00  7.391e-02   5.171 2.33e-07 ***
naeuNA              -4.737e-01  6.227e-01  9.366e-02  -5.058 4.24e-07 ***
round               -4.312e-01  6.497e-01  1.626e-01  -2.652  0.00801 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                    exp(coef) exp(-coef) lower .95 upper .95
height                 0.8915     1.1217    0.8464    0.9389
weight                 1.0102     0.9899    1.0032    1.0172
I(overall^(0.5))       0.2129     4.6960    0.1627    0.2787
overall                1.1288     0.8859    1.0931    1.1656
I(overall^(2))         0.9998     1.0002    0.9997    0.9999
positionD              0.8881     1.1260    0.7450    1.0587
age                    1.4655     0.6824    1.2679    1.6939
naeuNA                 0.6227     1.6060    0.5183    0.7481
round                  0.6497     1.5391    0.4724    0.8936

Concordance= 0.794  (se = 0.013 )
Rsquare= 0.327   (max possible= 0.988 )
```

Appendix: Cox PH Model Output

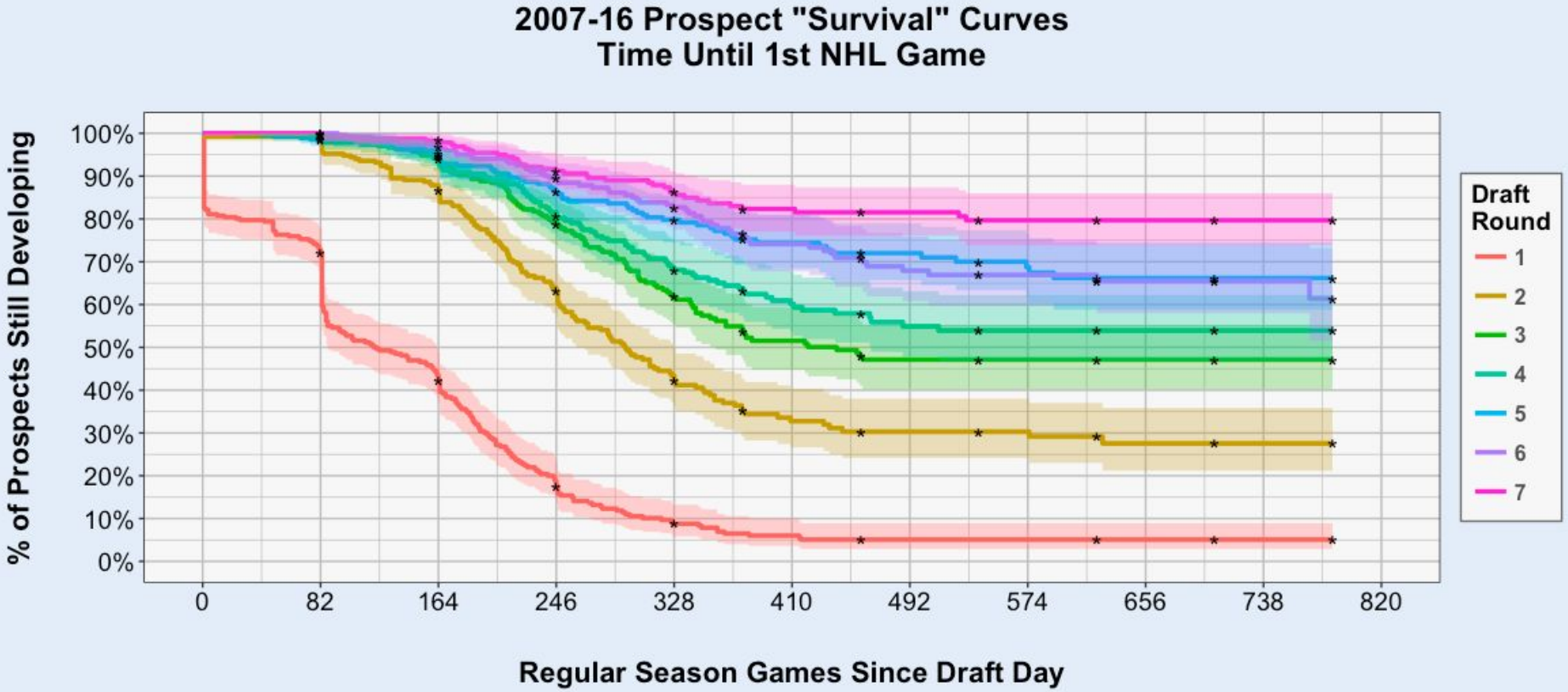
40th Game Benchmark

```
                      coef  exp(coef)   se(coef)       z Pr(>|z|)
height          -0.1060682  0.8993633  0.0299011  -3.547 0.000389 ***
weight           0.0074141  1.0074416  0.0040374   1.836 0.066308 .
I(overall^(0.5)) -1.7211444  0.1788614  0.1519049 -11.330  < 2e-16 ***
overall           0.1420136  1.1525923  0.0186304   7.623 2.49e-14 ***
I(overall^(2))   -0.0002150  0.9997850  0.0000374  -5.749 8.98e-09 ***
positionD        -0.0801400  0.9229871  0.1011928  -0.792 0.428388
age               0.4109134  1.5081947  0.0846442   4.855 1.21e-06 ***
naeuNA           -0.4749296  0.6219288  0.1061682  -4.473 7.70e-06 ***
round            -0.5993628  0.5491615  0.1894869  -3.163 0.001561 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                 exp(coef) exp(-coef) lower .95 upper .95
height              0.8994     1.1119    0.8482    0.9536
weight              1.0074     0.9926    0.9995    1.0154
I(overall^(0.5))    0.1789     5.5909    0.1328    0.2409
overall             1.1526     0.8676    1.1113    1.1955
I(overall^(2))      0.9998     1.0002    0.9997    0.9999
positionD           0.9230     1.0834    0.7569    1.1255
age                 1.5082     0.6630    1.2776    1.7804
naeuNA              0.6219     1.6079    0.5051    0.7658
round               0.5492     1.8210    0.3788    0.7961

Concordance= 0.808  (se = 0.014 )
Rsquare= 0.297   (max possible= 0.969 )
```

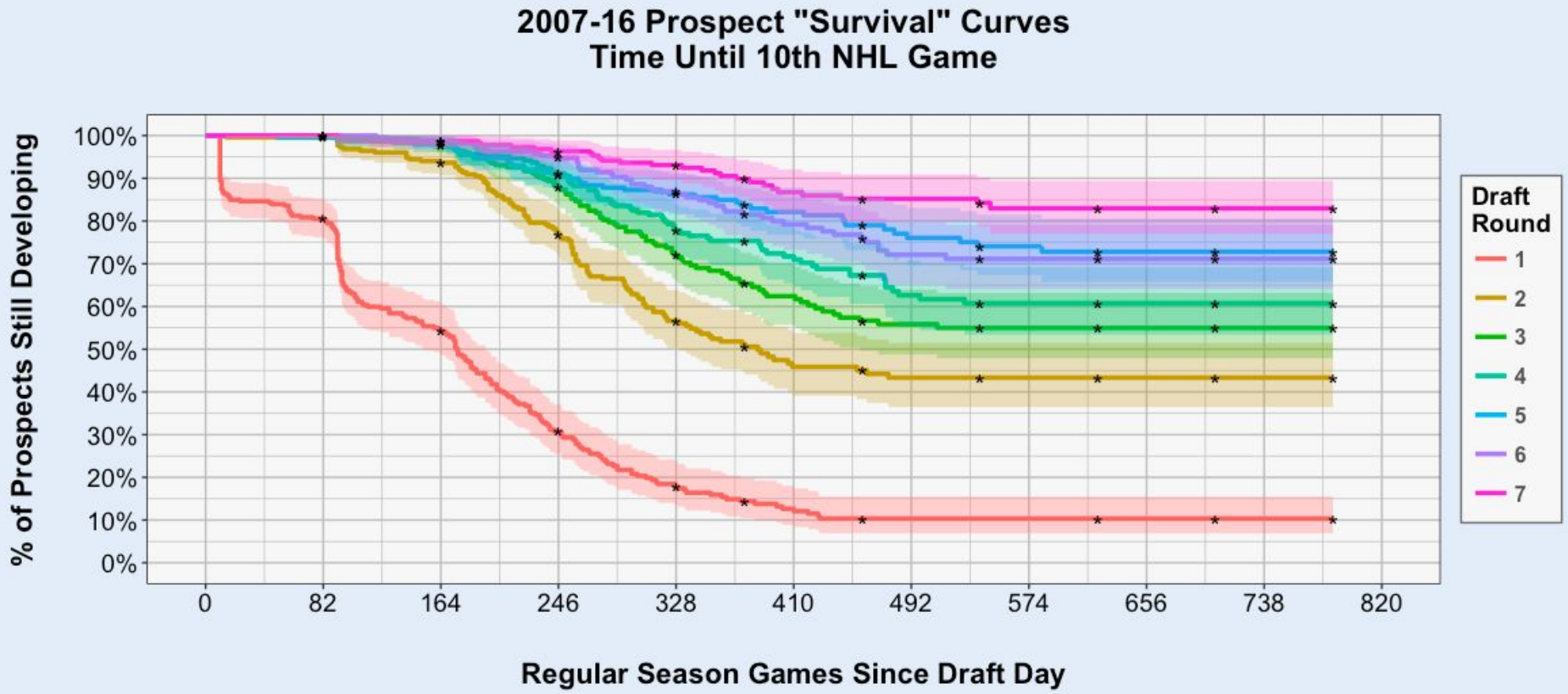
80th Game Benchmark

```
                      coef  exp(coef)   se(coef)       z Pr(>|z|)
height          -8.069e-02  9.225e-01  3.390e-02  -2.380  0.01732 *
weight           5.765e-03  1.006e+00  4.612e-03   1.250  0.21134
I(overall^(0.5)) -1.713e+00  1.803e-01  1.685e-01 -10.167  < 2e-16 ***
overall           1.433e-01  1.154e+00  2.125e-02   6.740 1.58e-11 ***
I(overall^(2))   -2.241e-04  9.998e-01  4.288e-05  -5.226 1.73e-07 ***
positionD        -7.355e-02  9.291e-01  1.147e-01  -0.641  0.52152
age               4.006e-01  1.493e+00  9.514e-02   4.211 2.55e-05 ***
naeuNA           -3.439e-01  7.090e-01  1.186e-01  -2.900  0.00373 **
round            -5.871e-01  5.559e-01  2.186e-01  -2.686  0.00724 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                 exp(coef) exp(-coef) lower .95 upper .95
height              0.9225     1.0840    0.8632    0.9859
weight              1.0058     0.9943    0.9967    1.0149
I(overall^(0.5))    0.1803     5.5468    0.1296    0.2508
overall             1.1540     0.8665    1.1069    1.2031
I(overall^(2))      0.9998     1.0002    0.9997    0.9999
positionD           0.9291     1.0763    0.7420    1.1634
age                 1.4927     0.6699    1.2388    1.7987
naeuNA              0.7090     1.4105    0.5620    0.8945
round               0.5559     1.7988    0.3622    0.8533

Concordance= 0.816  (se = 0.016 )
Rsquare= 0.253   (max possible= 0.933 )
```

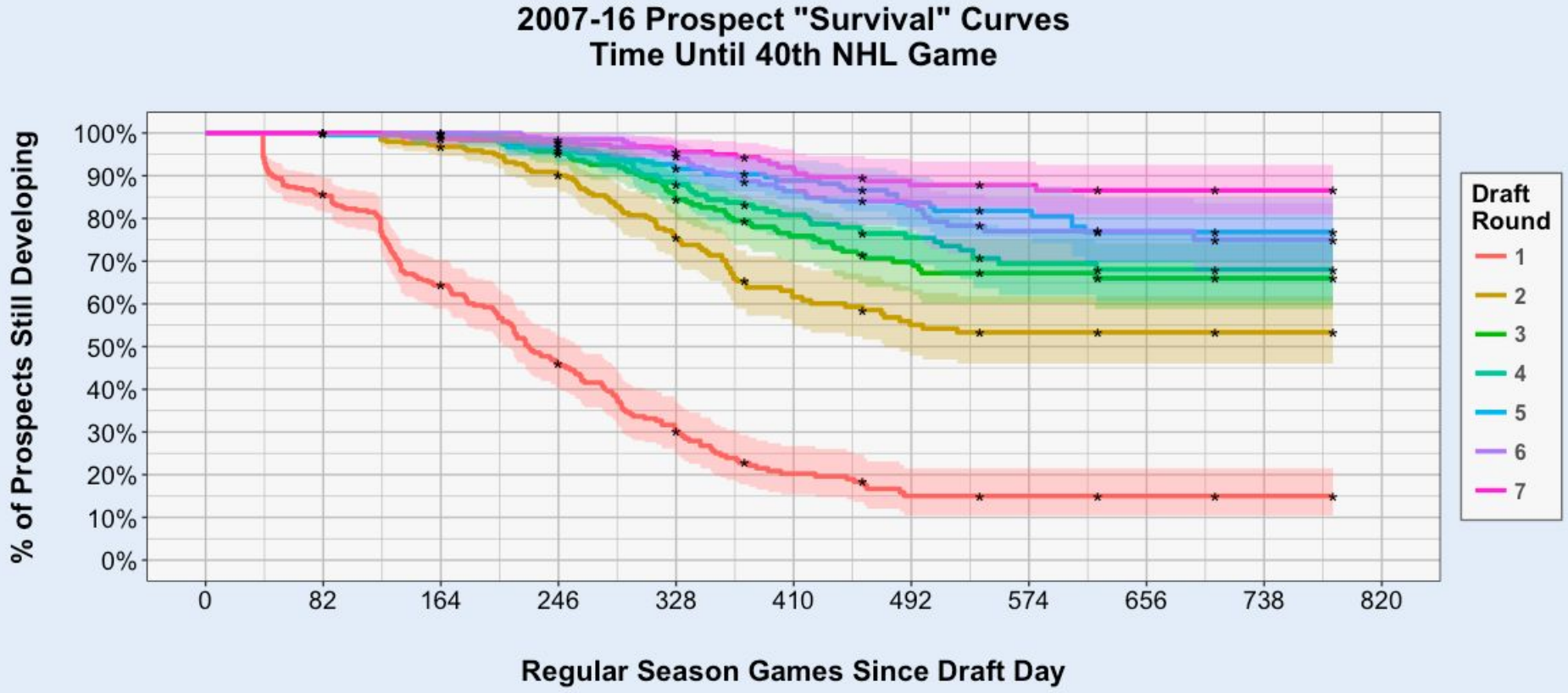
Appendix: Just a Ton of Kaplan-Meier Curves



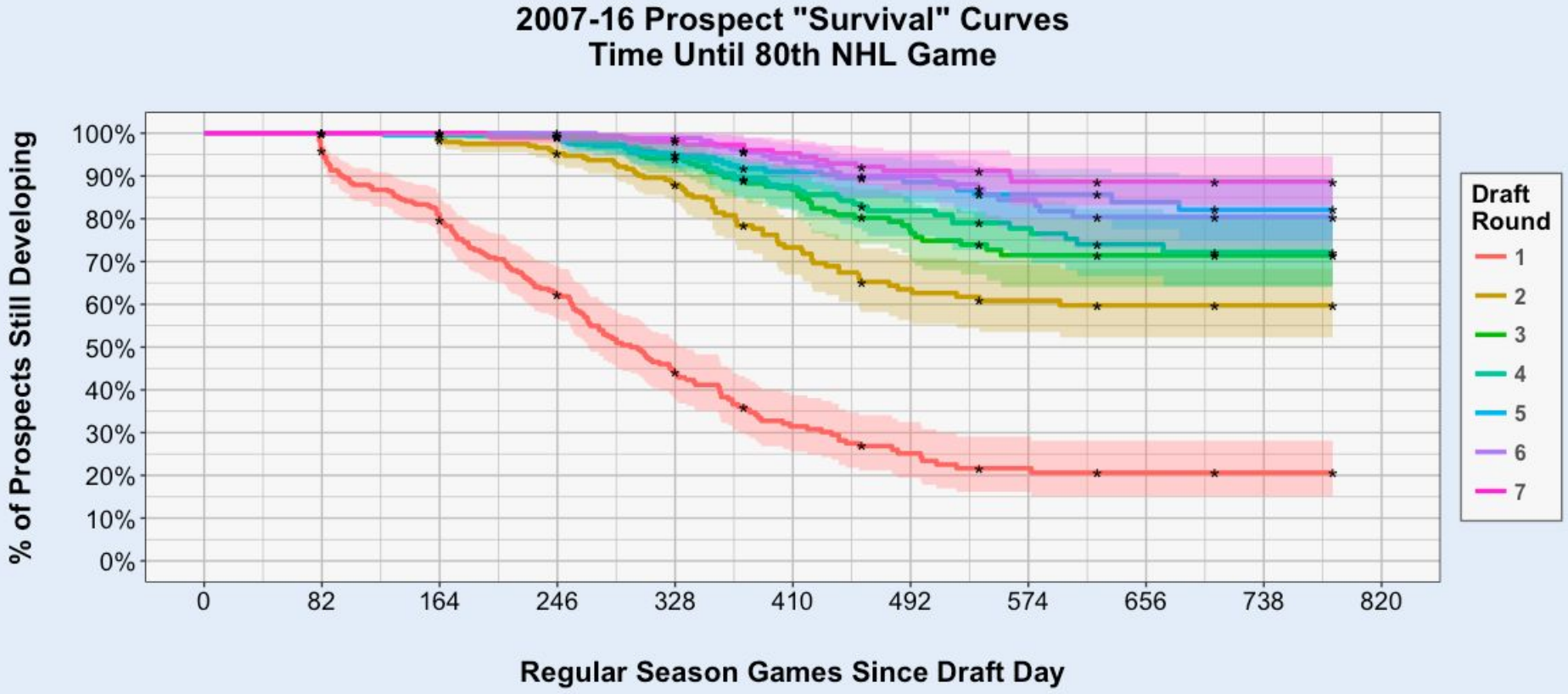
Draft Round: 10 Games



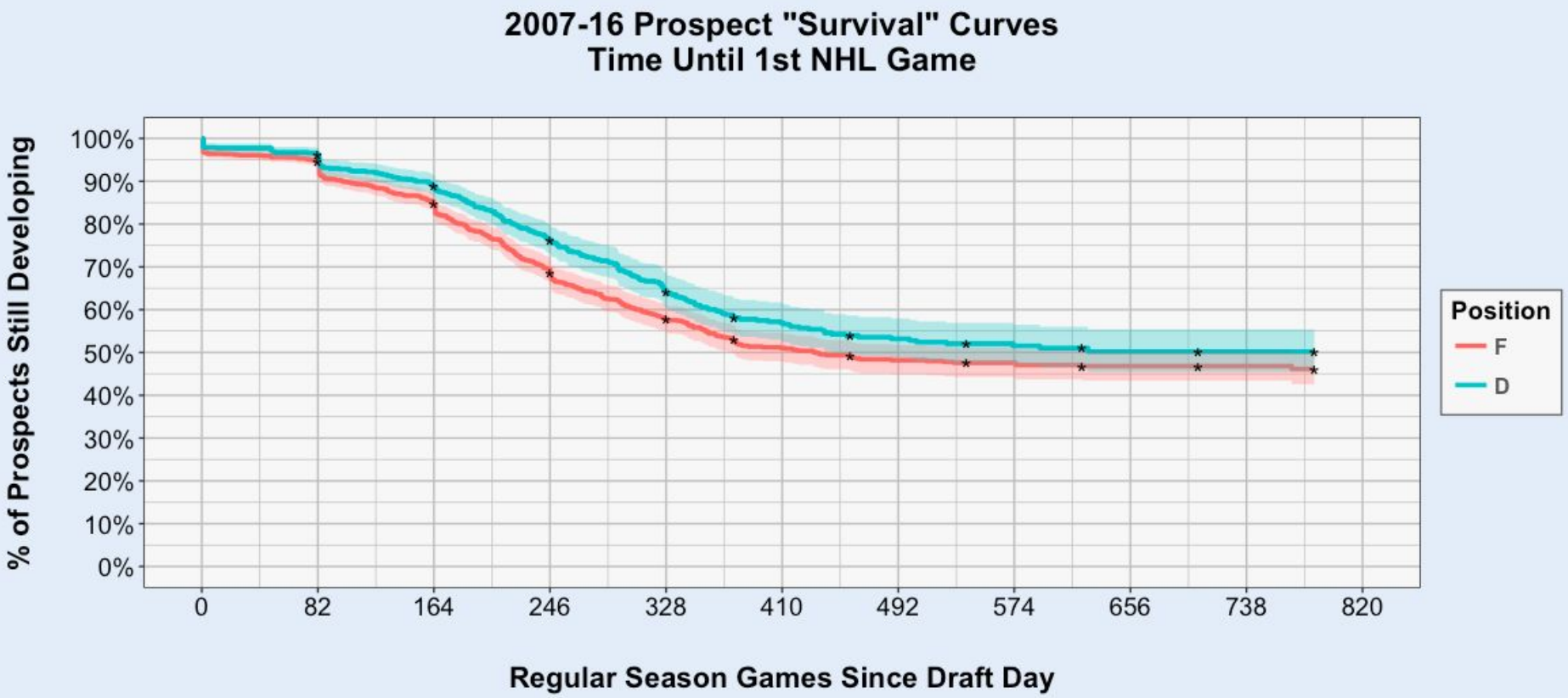
Draft Round: 40 Games



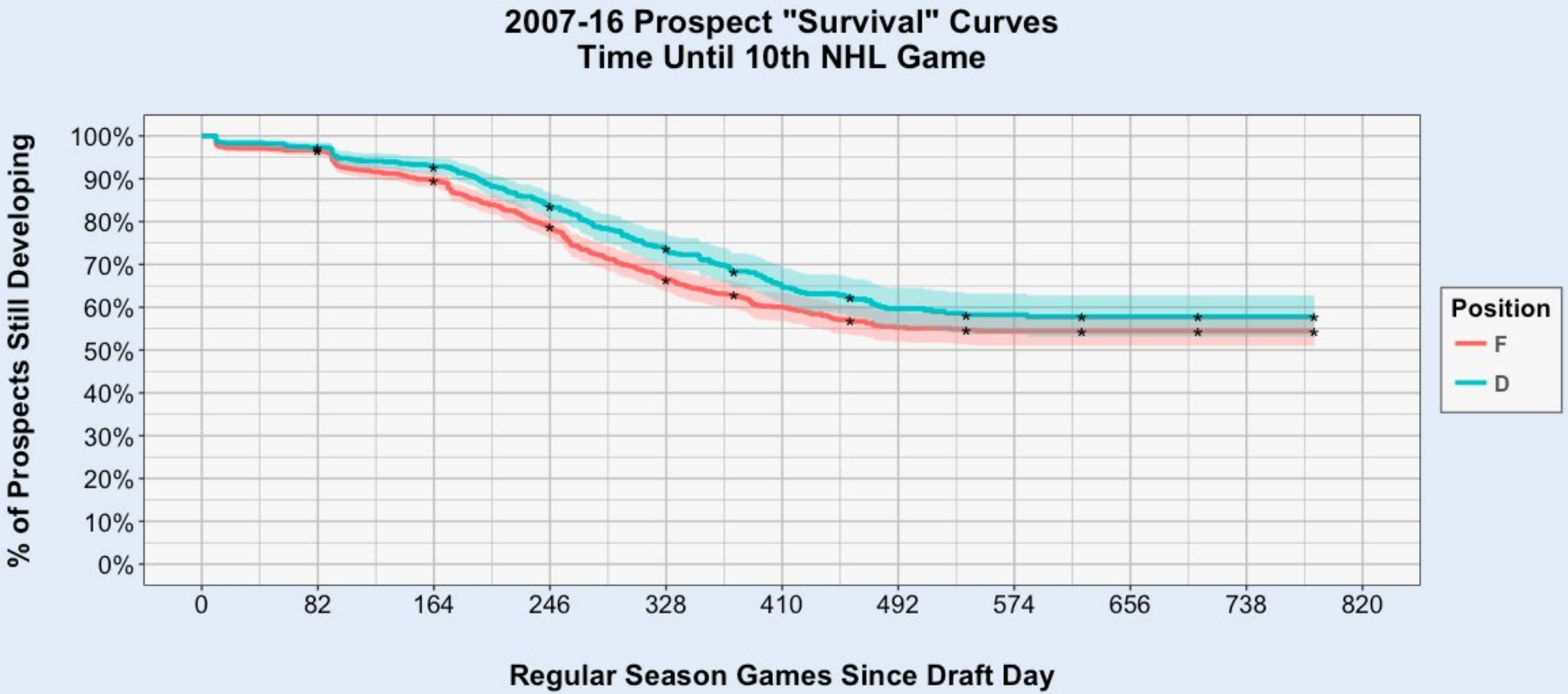
Draft Round: 80 Games



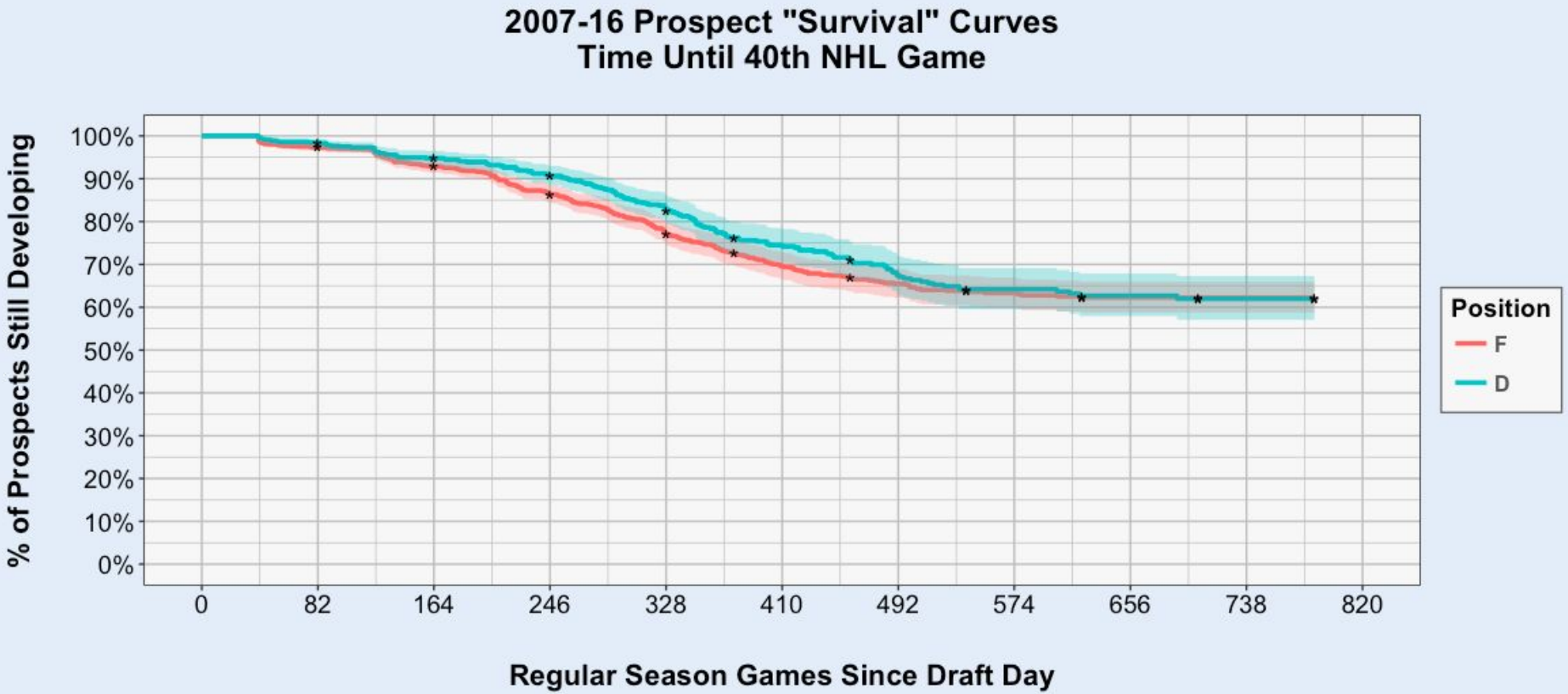
Position: 1 Game



Position: 10 Games



Position: 40 Games



Position: 80 Games