Multi-Horizon Career Longevity Prediction for NHL Skaters: Landmark Logistic Regression Reveals Stage-Specific Risk Factors

Elizabeth J. Goldstein[1] and Jessica A. Pearson[1]

ORRO AI GENIUS LLC, North Carolina, USA
lj@orrogenius.com, jp@orrogenius.com

Abstract. We construct a dataset of 5,754 National Hockey League skaters spanning 1979–2024 and apply landmark logistic regression at four career horizons to predict whether a player who has reached season N will play at least one game in season M. The central empirical finding is that the penalty for Russian-born players concentrates at the season 3-to-5 transition (OR = 0.314, $p < 0.0001$), consistent with mid-career migration to the Kontinental Hockey League after entry-level contracts expire.

Keywords: Hockey analytics · Career longevity · Landmark analysis · NHL · Kontinental Hockey League · Competing labor markets

1 Introduction

Predicting how long a professional hockey player will last in the (N)ational (H)ockey (L)eague is a question of practical importance. Prior work on NHL career longevity has typically approached the problem with pooled hazard models or single-outcome regressions that estimate average effects across all career stages [3]. Although these approaches model baseline hazard as a function of tenure (Depken et al. use a fifth-order polynomial), covariate effects—including the Russian coefficient that is central to their paper—are constrained to operate uniformly across tenure. This specification is efficient when the research question concerns average effects, but it cannot detect whether a given factor's influence on exit risk concentrates at specific career transitions or operates uniformly throughout a career.

To address this, we adopt the *landmark analysis* framework from biostatistics [1, 2]. At each of four career horizons, we condition on the player having already reached season N and ask whether he will play at least one game in season M. This analysis shows that a player's risk of departure changes as his career progresses.

A second design choice merits a preliminary discussion. We define career length as the count of distinct NHL seasons with at least one game played, rather than as a consecutive-season streak. Rather than focusing on uninterrupted NHL tenure, we aggregate NHL milestone accumulation. Our choice is well-suited to

studying outside-option effects such as KHL defection, because players may cycle in and out of the NHL.

2 Related Work on NHL Career Longevity

Depken, Ducking, and Groothuis [3] provide the most directly relevant prior analysis of NHL career longevity by nationality. Using a panel of NHL players spanning the 2000–2013 seasons, they fit a pooled discrete-time hazard model in which each player contributes one observation per NHL season and the binary target variable equals one if the player exited the league that season. Covariates include on-ice performance statistics, personal characteristics, and country-of-origin dummies, with tenure entered as a fifth-order polynomial that captures the baseline hazard. The Russian coefficient in this specification is estimated as a single number constrained to apply equally at every tenure, conditional on the polynomial baseline.

Their central findings are twofold: conditional on performance, European-born players exit the league earlier than North American-born players, and Russian-born players in particular exhibit still shorter careers than other Europeans—a gap that widens after the 2004–05 lockout.

A Blinder–Oaxaca nonlinear decomposition in their paper reveals a striking pattern: if the observed exit-rate gap were driven solely by differences in productivity and observable characteristics, Russian-born and other European-born players would be predicted to have *longer*, not shorter, NHL careers than North American-born players—Russians by approximately 5–12% in exit probability and other Europeans by approximately 6%, depending on decomposition technique. The observed shorter-career pattern emerges from the coefficients component of the decomposition: same observable characteristics yield systematically worse outcomes for European-born, and especially Russian-born, players. The authors read this residual gap as evidence of extrinsic push or pull factors beyond productivity. On the push side, shorter careers are consistent with exit discrimination (NHL teams cutting or declining to re-sign Russian players at higher rates than their productivity would predict); on the pull side, they are consistent with players choosing opportunities beyond the NHL such as the KHL.

Christie and Lavoie [4] document the entry-side correlate of this exit-side penalty. Using 2009–2010 NHL draft data, they find that NHL teams do not draft young Russians as early as their talent would otherwise warrant.

Complementing this, Fenn, Gerdes, and Rothstein [5] examine NHL salary determination over 2005–2016 using fixed-effects least squares and quantile regression on two datasets (n = 4,046 and N = 1,717). Fenn et al. fit two specifications: one using lagged career-long performance statistics, and one using statistics from only the player's most recent contract period. In the latter, smaller dataset, they find that Russians command a significant salary premium throughout the lower three-quarters of the salary distribution. They also find an analogous (and weaker) premium for Swedish and Czech defensemen. The authors attribute these patterns to outside-option leverage from other leagues.

3 Data

3.1 Dataset Construction

Player records were collected from the NHL Edge API (`api-web.nhle.com/v1`) [10] via a roster scan across all 40 franchise abbreviations and all seasons from 1979–80 through 2024–25, and a supplemental stats-leader scan to capture players missed by the roster approach.

The final dataset contains 5,754 skaters (goalies excluded) with at least one NHL regular-season game played.

3.2 Variables

Target Variable. For each horizon (N, M), the outcome is binary: $Y = 1$ if the player accumulated $\geq M$ distinct NHL seasons with at least one game played (seasons need not be consecutive).

Predictor Variables. The base specification includes: age at first NHL season, position (forward vs. defense), draft status (binary), birth region, height and weight (NHL roster-listed anthropometrics; season-specific values are available via the API but vary negligibly over an NHL career — mean weight change from season 1 to season 5 is +0.2 lb with SD of 2.0 lb — so we treat them as time-invariant rather than modeling season-specific trajectories), games played (GP) per season up to the conditioning season (GP S1 through GP S4, gated by horizon), and era dummies (decade bins based on draft year, included at horizons where conditioning season < 5).

The enriched specification additionally includes: pick-range bins (1–10, 11–31, 32–93, 94+, with undrafted as reference), number of compatriot teammates in the first season, number of teams in the first three seasons, roster stability (fraction of S1 teammates retained in S2), and first-season roster size. The binary draft indicator is dropped from the enriched specification because it is perfectly collinear with the pick-range bins. Among the enriched features, roster stability is computable for 79.6% of players (players whose S1 team had an identifiable S2 roster); other social features are computable for 99%+ of the sample. The enriched-model complete-case sample retains 95–98% of the base sample at each horizon after all feature filters are applied.

Lockout and Shortened Seasons. The 2004–05 NHL season was cancelled entirely; it contributes zero games to any player's record and is excluded from seasons played for every player whose career spanned it (734 veterans). The four shortened seasons in the cohort window—1994–95 (48-game lockout), 2012–13 (48-game lockout), 2019–20 (70-game COVID), and 2020–21 (56-game COVID)—are treated as full seasons for milestone purposes: they count toward seasons played and toward all landmark outcomes (Y_M), and the raw GP totals are entered into the model without normalization. Approximately 8% of the cohort had at least one of seasons S_1 through S_4 fall in a shortened year. A sensitivity analysis

replacing raw GP with normalized GP ($\text{gp_s}N_{\text{pct}} = \text{gp_s}N_{\text{actual}}/\text{season_max}$, capped at 1.0) shifts test-set AUCs by less than 0.015 at every horizon and leaves the central S3→S5 Russia coefficient essentially unchanged (OR 0.314 → 0.289, $p < 0.001$ in both specifications).

3.3 Train/Test Split

The dataset uses a temporal split: players whose last NHL season was ≤ 2014–15 form the training set; players whose last NHL season fell between 2015–16 and 2023–24 form the test set. Active players with confirmed outcomes (i.e., those who have already reached the target milestone) are included in an expanded test set; active players with unconfirmed outcomes (censored) are excluded. Retired-only robustness checks are reported for all base models. Of the 5,754 skaters in the full dataset, 3,800 are in the training set, 1,012 are in the test set, and 942 are active players with unconfirmed outcomes who are censored and excluded from both splits. The test set has slightly longer average careers than the training set, reflecting the secular trend of NHL career lengthening rather than regime change.

Table 1. Descriptive statistics for the training and test sets.

Variable	Train mean	Train SD	Test mean	Test SD
seasons played	6.4	5.0	6.8	5.1
log(seasons played)	1.47	0.93	1.57	0.92
N	3,800		1,012	

4 Method

4.1 Landmark Logistic Regression

At each horizon (N, M), we fit:

$$\log \frac{P(Y_M = 1 \mid S \geq N)}{1 - P(Y_M = 1 \mid S \geq N)} = \beta_0 + \sum_j \beta_j X_j \tag{1}$$

where S is the number of distinct seasons played and X_j are features observable at season N. The four horizons are:

- **S1→3**: Will this player stick past the roster bubble?
- **S3→5**: Will he become a core player? (second-contract decision)
- **S5→8**: Will this veteran hold up? (UFA decision window)
- **S7→10**: How much runway is left? (aging curve)

At the S7→S10 enriched specification (Table 15), the design matrix is rank-deficient under the full feature set; standard MLE fails to converge, and we fall back to L2-penalized logistic regression via `sklearn`. This fallback is confined to one of the eight primary specifications; full details and the related near-separation issue at S7→S10 base are reported in Appendix E.1.

Prior NHL longevity studies generally report in-sample estimates on pooled panels–we use a temporal train/test split to enable out-of-sample validation.

4.2 Specifications

Two specifications are fit at each horizon. The **base** specification uses age, position, draft status (binary), region dummies, era dummies, height, weight, and season-gated GP history. The **enriched** specification drops the binary draft indicator and adds pick-range bins, compatriot count, team count, roster stability, and roster size. Both specifications are evaluated on the same common sample to enable direct comparison.

5 Results

5.1 Model Performance

Table 2 summarizes the base model performance across all four horizons.

Table 2. Base model performance (expanded test set).

Horizon	N_{train}	N_{test}	Event %	AUC	Brier
S1→3	3,780	1,820	86.3%	0.758	0.190
S3→5	2,669	1,404	88.2%	0.888	0.128
S5→8	1,979	1,006	81.9%	0.840	0.124
S7→10	1,528	749	81.6%	0.813	0.142

All four horizons show useful-to-good discrimination (AUC 0.76–0.89). The S3→5 model is strongest (AUC = 0.888). Calibration is solid across all horizons (Brier $\leq$ 0.19). Retired-only robustness checks show AUC deltas of 0.02–0.05, confirming that the inclusion of active players with confirmed outcomes does not substantially inflate performance estimates.

5.2 How Much Does the Model Add Beyond Age?

To verify the worth of additional variables, we compared the full model against a baseline that knows only the player's age at NHL entry (Table 3).

At only $S1 \rightarrow S3$, the age-baseline performs as well as the base logistic regression.

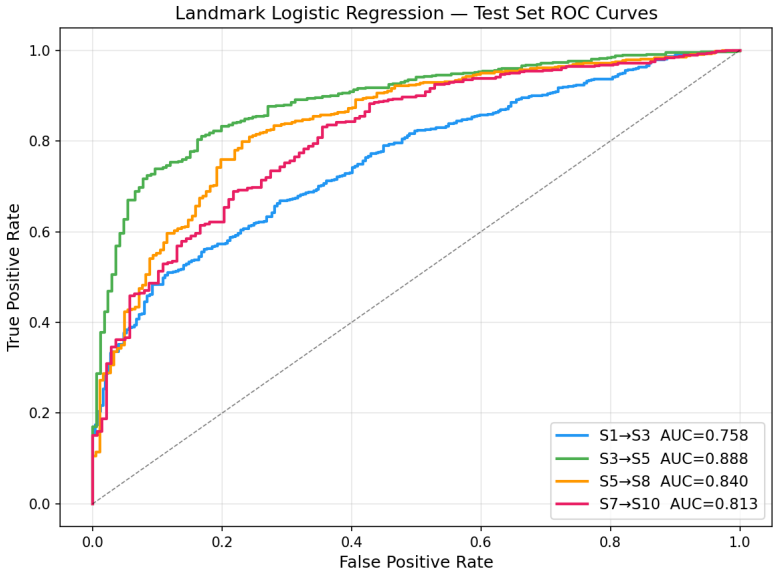

Fig. 1. Receiver operating characteristic (ROC) curves on the test set for the four base-specification landmark models. The $S3{\to}5$ model achieves the highest AUC (0.888), consistent with the second-contract transition being both the most interpretable and the most predictable career inflection point. Diagonal dashed line: random-guess baseline.

5.3 Classification Performance

Beyond AUC (which measures overall ranking ability), Table 4 shows how the base models perform at classifying players using a 50% probability threshold.

At S3→5, the model flags 83.1% of actual departures, with 37.0% precision. Given that only 11.8% of the sample actually departs, identifying 83% of a rare event is operationally useful. For a GM, the model is best at *ranking* players by departure risk (AUC 0.888) rather than making binary yes/no calls.

The headline Russia coefficient is robust to alternative era-control specifications; see Appendix F.

Table 3. Full model vs. age-only baseline (expanded test set).

Horizon	AUC (age only)	AUC (full)	Lift	Verdict
S1→3	0.773	0.758	−0.015	∼Same
S3→5	0.721	0.888	+0.167	Big lift
S5→8	0.646	0.840	+0.194	Huge lift
S7→10	0.692	0.813	+0.121	Strong lift

Table 4. Classification metrics at threshold = 0.50 (base models, expanded test set). "Flag exit" columns show performance at identifying players who will *not* reach the milestone.

Horizon	AUC	Prec.	Recall	F1	Flag exit		Event
					Prec.	Recall	%
S1→3	0.758	92.2%	72.6%	0.812	26.1%	61.0%	86.3%
S3→5	0.888	97.3%	81.0%	0.884	37.0%	83.1%	88.2%
S5→8	0.840	90.8%	88.8%	0.898	54.0%	59.3%	81.9%
S7→10	0.813	88.6%	90.5%	0.896	53.6%	48.6%	81.6%

5.4 The Russia Trajectory

Table 5 presents the Russia coefficient across career stages—the central finding of this paper.

Table 5. Russia birth-region coefficient across career horizons (base model). Reference category: Canadian-born players.

Horizon	Russia OR	p-value	Sig.	Interpretation
S1→3	0.639	0.039	*	36% lower odds
S3→5	0.314	<0.0001	***	69% lower odds
S5→8	0.703	0.255		30% lower odds (n.s.)
S7→10	1.625	0.186		63% higher odds (n.s.)

The penalty concentrates at S3→5: Russian players who reach season 3 have 69% lower odds of reaching season 5 compared to otherwise-similar Canadian players (OR = 0.314, $p < 0.0001$). This is the second-contract window—exactly when KHL offers become attractive. The penalty is more modest at $S1{\rightarrow}3$, not statistically distinguishable from the Canadian baseline at $S5{\rightarrow}8$, and reverses to a non-significant point estimate above one at $S7{\rightarrow}10$.

5.5 Other European and Scandinavian Players

Other European players (a heterogeneous category including Czech, Slovak, German, Swiss, Belarusian, and other Eastern and Central European nationalities) show a similar but more persistent pattern, with significant penalties at three of four horizons (S1→3: OR 0.614, $p = 0.002$; S3→5: OR 0.410, $p < 0.001$; S5→8: OR 0.439, $p < 0.001$; S7→10: OR 0.845, n.s.). Unlike Russia, this penalty does not concentrate at one transition point, suggesting that European leagues (Czech Extraliga, Swiss NL, German DEL) offer career alternatives across the full career arc. Scandinavian players show a very similar pattern, consistent with the SHL and Liiga as viable alternative career paths.

5.6 GP Recency Gradient

At every horizon, the most recent season's games played is the strongest GP predictor. At S5→8, the coefficient for GP S4 (OR = 1.027, $p < 0.0001$) dominates GP S3 (OR = 1.013), GP S2 (OR = 1.006), and GP S1 (OR = 1.003, n.s.). This recency gradient validates the landmark approach: recent performance contains more information about future persistence than distant history.

5.7 S3→5 Coefficient Table

Table 6 presents the full coefficient table for the strongest model.

Table 6. S3→5 base model coefficients (expanded test set, AUC = 0.888). Era dummies included but omitted for space.

Variable	OR	Sig.	Interpretation
Age at NHL entry	0.752	***	Each extra year = 25% lower odds
GP in season 3	1.045	***	Most important GP variable
GP in season 2	1.014	***	Matters, less than S3
GP in season 1	1.004		Minimal once S2/S3 controlled
Weight (lbs)	1.040	***	Heavier = more durable
Height (in.)	0.873	***	Modest negative effect
Russian	0.314	***	69% lower odds vs. Canadian
Scandinavian	0.360	***	65% lower odds
Other European	0.410	***	59% lower odds
American	1.093		Similar to Canadians
Drafted	0.740	.	Selection effect (see text)

The binary drafted indicator is marginally significant at $S3{\to}5$ (OR = 0.740, $p = 0.059$), suggesting that undrafted players who reach season 3 may be positively selected on unobservables relative to drafted players at the same career stage. This pattern is consistent with the enriched specification, where the late-pick bins (picks 32–93 and 94+) rather than the binary draft indicator carry the predictive signal.

5.8 Social Environment Features

The enriched model adds social and environmental features. On the same common sample, the AUC deltas between base and enriched specifications are near zero (ΔAUC: S1→3: −0.022; S3→5: +0.001; S5→8: −0.006; S7→10: −0.002). The social features do not improve discrimination beyond what age, size, GP history, and birth region already capture. One variable is interpretively noteworthy: *roster stability* (fraction of S1 teammates retained in S2) shows large odds ratios at S3→5 (OR = 7.7, $p = 0.0003$) and S5→8 (OR = 10.6, $p < 0.001$). However, these likely reflect organizational quality rather than a causal stability

mechanism—good organizations both retain rosters and develop players well. We treat this as exploratory.

5.9 Robustness Test

The primary robustness test set includes active players with confirmed outcomes (those who have already reached the target milestone and are still playing). These observations are mechanically positive ($Y = 1$), so their inclusion raises the test event rate by 10–13 percentage points relative to a retired-only test set and mechanically raises AUC. We refit each base model with the retired-only test set as a robustness check (Table 7). AUC moves modestly at every horizon, with a maximum shift of $|\Delta\text{AUC}| = 0.047$ at $S1{\to}3$. Even at this horizon, the retired-only AUC of 0.712 remains above the 0.70 "useful model" threshold; at $S3{\to}5$ the AUC shift is only 0.020, leaving the strongest horizon essentially unchanged. The Russia coefficient itself is estimated on training data alone and is therefore unaffected by test-set composition—the retired-only and expanded specifications produce identical coefficients and standard errors by construction.

Table 7. Base-model test-set performance under expanded (includes active-confirmed) and retired-only compositions.

Horizon	AUC_{exp}	AUC_{ret}	ΔAUC	$\text{Event}\%_{\text{exp}}$	$\text{Event}\%_{\text{ret}}$
$S1{\to}3$	0.758	0.712	+0.047	86.3	74.8
$S3{\to}5$	0.888	0.868	+0.020	88.2	77.5
$S5{\to}8$	0.840	0.837	+0.004	81.9	68.2
$S7{\to}10$	0.813	0.783	+0.030	81.6	68.8

6 Discussion

6.1 Analytical Insights

Three findings emerge from this work.

1. The Russia Effect Has a Shape, Not Just a Sign. The Russian penalty concentrates almost entirely at the S3→5 transition, which is exactly the second-contract window when KHL offers become attractive.

2. Simplicity Wins. Age, size, GP history, draft status, and birth region capture nearly all the predictive signal.

3. The Landmark Framework Is the Methodological Contribution. Conditioning on reaching season N before predicting season M is standard in biostatistics but has not generally been applied this way in hockey analytics. It is what makes the Russia trajectory visible and what reveals that risk factors change across career stages. Our landmark framework trades off some statistical efficiency (separate regressions at each horizon) in exchange for stage-specific covariate effects. The pooled discrete-time hazard framework of Depken et al. [3] and similar studies is more efficient when the research question concerns average effects but cannot reveal the stage-specific patterns our approach makes visible.

6.2 Reconciliation with the Existing Literature

Our landmark estimates can be placed on a common scale with the pooled annual-exit logistic regression of Depken, Ducking, and Groothuis [3] by translating their annual hazard rates into the two-season survival metric our model estimates. Their [3] Table 1 reports annual exit rates at tenures 3 through 5 of 0.214, 0.272, and 0.264 for Russian-born players versus 0.104, 0.109, and 0.144 for North Americans. Compounding these into a two-season survival probability from tenure 3 to tenure 5 gives roughly 0.57 for Russians and 0.80 for North Americans, implying a survival odds ratio of approximately 0.34 over that tenure window. Our $S3{\to}5$ landmark estimate is OR $= 0.314$ ($p < 0.0001$, $n = 1{,}404$), closely aligned with the two-season survival implied by their hazard rates at the same career stage. Three aspects of this agreement are worth noting.

First, the alignment is substantive rather than mechanical. Our sample spans 1979–2024 and uses distinct-season accounting; theirs spans 2000–2013 and uses consecutive-season accounting. That the two estimates land in the same neighborhood despite these design differences provides external corroboration of the $S3{\to}5$ Russian penalty rather than a coincidence of specification.

Second, our distinct-season definition makes the landmark estimate conservative relative to a consecutive-season comparison. A Russian player who defects to the KHL for one season and returns counts against the Russian survival rate in Depken et al.'s framing but remains a contributing observation in ours.

Third, and most importantly for the substantive contribution, our landmark decomposition shows that the $S3{\to}5$ severity is localized rather than career-spanning. The same pattern of S3→5 peak severity appears for Scandinavian and Other European players, consistent with the second-contract transition being a structurally important career inflection point for all European-born NHL players.

7 Limitations

Selection Bias. Data collection methods biased the sample toward longer careers. Full resolution would require a complete player registry such as the Hockey Reference player index [11].

Causality. The Russia coefficient trajectory is suggestive of a competing labor market mechanism but is not causal.

Population Drift. The birth-region composition of the test set differs from the training set, reflecting the secular trend of NHL internationalization (Canadian share drops from 61.4% in train to 45.6% in test). Era dummies partially absorb this time trend when included (conditioning seasons < 5), but the Russia coefficient may absorb residual era effects at horizons where era dummies are omitted for identification reasons. Thus, we interpret within-horizon findings cautiously.

8 Conclusion

We have shown that modeling career longevity as a multi-horizon prediction problem reveals dynamics that single-outcome models miss. The central finding—that the Russia penalty concentrates at one specific career transition point rather than operating uniformly—has implications for how teams evaluate international players and structure contracts.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Appendix

A Full Regression Results

This appendix reports full coefficient estimates for all eight primary landmark models: four career horizons ($S1 \rightarrow S3$, $S3 \rightarrow S5$, $S5 \rightarrow S8$, $S7 \rightarrow S10$) crossed with two specifications (base and enriched). All estimates use the expanded test set described in Section 3.2. The reference categories are Canadian-born (for birth region), era for draft years 1985 to 1989 (for era dummies), and undrafted (for pick-range bins). Significance notation: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.10$. Odds ratios OR $= \exp(\beta)$; OR < 1 indicates lower odds of reaching the target milestone relative to the reference category.

Table 8. S1→S3 base specification. $n_{\text{train}} = 3{,}780$, $n_{\text{test}} = 1{,}820$. AUC = 0.758, Brier = 0.190, pseudo-$R^2 = 0.247$.

Variable	β	OR	SE	z	p	Sig
age_at_first_season	−0.381	0.684	0.021	−17.88	0.000	***
pos2_F	0.152	1.165	0.093	1.64	0.100	
drafted	−0.327	0.721	0.113	−2.88	0.004	**
reg_Russia	−0.447	0.639	0.217	−2.06	0.040	*
reg_USA	0.354	1.424	0.116	3.06	0.002	**
reg_Scandinavia	−0.486	0.615	0.160	−3.04	0.002	**
reg_Other Europe	−0.488	0.614	0.157	−3.11	0.002	**
reg_Other	0.265	1.303	0.700	0.38	0.706	
era_1990	0.087	1.091	0.131	0.67	0.505	
era_1995	−0.084	0.920	0.134	−0.63	0.531	
era_2000	−0.262	0.769	0.137	−1.92	0.055	.
era_2005	−0.977	0.376	0.147	−6.65	0.000	***
era_2010	−2.106	0.122	0.315	−6.69	0.000	***
height_listed	−0.140	0.870	0.030	−4.71	0.000	***
weight_listed	0.039	1.039	0.004	9.29	0.000	***
gp_s1_actual	0.045	1.046	0.003	16.88	0.000	***

Table 9. S1→S3 enriched specification. $n_{\text{train}} = 3{,}780$, $n_{\text{test}} = 1{,}820$. AUC = 0.755, Brier = 0.184, pseudo-$R^2 = 0.274$.

Variable	β	OR	SE	z	p	Sig
age_at_first_season	−0.377	0.686	0.023	−16.34	0.000	***
pos2_F	0.155	1.168	0.095	1.64	0.102	
reg_Russia	−1.689	0.185	0.304	−5.55	0.000	***
reg_USA	−0.600	0.549	0.202	−2.97	0.003	**
reg_Scandinavia	−1.783	0.168	0.267	−6.67	0.000	***
reg_Other Europe	−1.778	0.169	0.265	−6.71	0.000	***
reg_Other	−1.134	0.322	0.756	−1.50	0.134	
era_1990	0.145	1.156	0.135	1.08	0.282	
era_1995	−0.147	0.864	0.139	−1.06	0.292	
era_2000	−0.434	0.648	0.144	−3.02	0.003	**
era_2005	−1.157	0.314	0.153	−7.57	0.000	***
era_2010	−2.264	0.104	0.317	−7.14	0.000	***
height_listed	−0.148	0.862	0.031	−4.86	0.000	***
weight_listed	0.041	1.042	0.004	9.65	0.000	***
gp_s1_actual	0.044	1.045	0.003	15.80	0.000	***
pick_1_10	0.208	1.231	0.317	0.66	0.512	
pick_11_31	−0.261	0.770	0.183	−1.43	0.153	
pick_32_93	−0.375	0.687	0.135	−2.77	0.006	**
pick_94_plus	−0.359	0.699	0.124	−2.89	0.004	**
n_compatriot_teammates_early	−0.066	0.936	0.011	−6.19	0.000	***
roster_size_s1	−0.047	0.955	0.011	−4.25	0.000	***

Table 10. S3→S5 base specification. $n_{\text{train}} = 2{,}669$, $n_{\text{test}} = 1{,}404$. AUC = 0.888, Brier = 0.128, pseudo-$R^2 = 0.323$.

Variable	β	OR	SE	z	p	Sig
age_at_first_season	−0.285	0.752	0.029	−9.73	0.000	***
pos2_F	−0.080	0.924	0.123	−0.65	0.518	
drafted	−0.301	0.740	0.159	−1.89	0.059	.
reg_Russia	−1.158	0.314	0.284	−4.07	0.000	***
reg_USA	0.089	1.093	0.154	0.58	0.565	
reg_Scandinavia	−1.022	0.360	0.222	−4.59	0.000	***
reg_Other Europe	−0.891	0.410	0.213	−4.19	0.000	***
reg_Other	−0.521	0.594	0.819	−0.64	0.525	
era_1990	0.373	1.452	0.172	2.17	0.030	*
era_1995	−0.493	0.611	0.167	−2.96	0.003	**
era_2000	−0.439	0.644	0.183	−2.40	0.016	*
era_2005	−0.674	0.509	0.224	−3.01	0.003	**
era_2010	−1.648	0.193	0.954	−1.73	0.084	.
height_listed	−0.136	0.873	0.040	−3.38	0.001	***
weight_listed	0.039	1.040	0.006	6.73	0.000	***
gp_s1_actual	0.004	1.004	0.003	1.31	0.191	
gp_s2_actual	0.014	1.014	0.003	5.33	0.000	***
gp_s3_actual	0.044	1.045	0.003	16.93	0.000	***

Table 11. S3→S5 enriched specification. $n_{\text{train}} = 2{,}565$, $n_{\text{test}} = 1{,}336$. AUC = 0.890, Brier = 0.116, pseudo-$R^2 = 0.342$.

Variable	β	OR	SE	z	p	Sig
age_at_first_season	−0.276	0.759	0.034	−8.13	0.000	***
pos2_F	−0.146	0.864	0.130	−1.12	0.263	
reg_Russia	−2.188	0.112	0.390	−5.62	0.000	***
reg_USA	−0.667	0.513	0.259	−2.58	0.010	**
reg_Scandinavia	−2.021	0.133	0.346	−5.85	0.000	***
reg_Other Europe	−1.971	0.139	0.339	−5.81	0.000	***
reg_Other	−1.600	0.202	0.866	−1.85	0.065	.
era_1990	0.400	1.492	0.185	2.16	0.031	*
era_1995	−0.569	0.566	0.179	−3.17	0.002	**
era_2000	−0.360	0.697	0.199	−1.81	0.071	.
era_2005	−0.607	0.545	0.235	−2.58	0.010	**
era_2010	−1.453	0.234	0.933	−1.56	0.119	
height_listed	−0.145	0.865	0.043	−3.38	0.001	***
weight_listed	0.041	1.041	0.006	6.61	0.000	***
gp_s1_actual	0.003	1.003	0.003	0.89	0.374	
gp_s2_actual	0.016	1.016	0.003	5.42	0.000	***
gp_s3_actual	0.046	1.047	0.003	16.34	0.000	***
pick_1_10	−0.066	0.936	0.343	−0.19	0.848	
pick_11_31	−0.316	0.729	0.234	−1.35	0.177	
pick_32_93	−0.353	0.703	0.189	−1.87	0.062	.
pick_94_plus	−0.323	0.724	0.179	−1.81	0.071	.
n_compatriot_teammates_early	−0.048	0.953	0.013	−3.58	0.000	***
n_teams_early	−0.238	0.788	0.097	−2.46	0.014	*
orig_team_roster_stability_early	2.039	7.684	0.567	3.60	0.000	***
roster_size_s1	0.058	1.060	0.018	3.23	0.001	**

Table 12. S5→S8 base specification. $n_{\text{train}} = 1{,}979$, $n_{\text{test}} = 1{,}006$. AUC = 0.840, Brier = 0.124, pseudo-R^2 = 0.192. Era dummies omitted at this horizon to avoid quasi-separation.

Variable	β	OR	SE	z	p	Sig
age_at_first_season	−0.297	0.743	0.031	−9.62	0.000	***
pos2_F	−0.192	0.826	0.122	−1.57	0.116	
drafted	−0.214	0.808	0.162	−1.32	0.187	
reg_Russia	−0.352	0.703	0.309	−1.14	0.255	
reg_USA	0.047	1.049	0.156	0.30	0.761	
reg_Scandinavia	−0.691	0.501	0.225	−3.07	0.002	**
reg_Other Europe	−0.824	0.439	0.208	−3.96	0.000	***
reg_Other	0.038	1.038	0.881	0.04	0.966	
height_listed	−0.061	0.941	0.040	−1.53	0.125	
weight_listed	0.018	1.018	0.005	3.34	0.001	***
gp_s1_actual	0.003	1.003	0.002	1.44	0.150	
gp_s2_actual	0.006	1.006	0.003	2.45	0.014	*
gp_s3_actual	0.013	1.013	0.003	4.94	0.000	***
gp_s4_actual	0.027	1.027	0.003	10.48	0.000	***

Table 13. S5→S8 enriched specification. $n_{\text{train}} = 1{,}928$, $n_{\text{test}} = 971$. AUC = 0.830, Brier = 0.130, pseudo-R^2 = 0.198. Era dummies omitted at this horizon.

Variable	β	OR	SE	z	p	Sig
age_at_first_season	−0.265	0.767	0.036	−7.41	0.000	***
pos2_F	−0.190	0.827	0.125	−1.52	0.128	
reg_Russia	−0.484	0.617	0.382	−1.27	0.205	
reg_USA	−0.015	0.985	0.230	−0.07	0.947	
reg_Scandinavia	−0.759	0.468	0.314	−2.42	0.015	*
reg_Other Europe	−0.901	0.406	0.304	−2.96	0.003	**
reg_Other	−0.019	0.981	0.916	−0.02	0.983	
height_listed	−0.070	0.932	0.041	−1.72	0.086	.
weight_listed	0.020	1.020	0.006	3.49	0.000	***
gp_s1_actual	0.003	1.003	0.002	1.16	0.247	
gp_s2_actual	0.007	1.007	0.003	2.59	0.010	**
gp_s3_actual	0.012	1.013	0.003	4.56	0.000	***
gp_s4_actual	0.027	1.027	0.003	10.15	0.000	***
pick_1_10	0.096	1.101	0.274	0.35	0.726	
pick_11_31	−0.128	0.879	0.217	−0.59	0.554	
pick_32_93	−0.160	0.852	0.187	−0.86	0.391	
pick_94_plus	−0.261	0.770	0.179	−1.46	0.145	
n_compatriot_teammates_early	−0.001	0.999	0.011	−0.08	0.938	
n_teams_early	−0.066	0.936	0.101	−0.65	0.515	
orig_team_roster_stability_early	2.358	10.574	0.590	4.00	0.000	***
roster_size_s1	0.043	1.044	0.017	2.45	0.014	*

Table 14. S7→S10 base specification. $n_{\text{train}} = 1{,}528$, $n_{\text{test}} = 749$. AUC = 0.813, Brier = 0.142, pseudo-$R^2 = 0.107$. Era dummies omitted; `reg_Other` coefficient omitted from the table owing to near-separation ($\beta = 33.77$, SE on the order of 10^{13}; see Section E).

Variable	β	OR	SE	z	p	Sig
age_at_first_season	−0.326	0.722	0.035	−9.38	0.000	***
pos2_F	−0.234	0.792	0.127	−1.84	0.067	.
drafted	−0.259	0.772	0.174	−1.49	0.137	
reg_Russia	0.485	1.625	0.367	1.32	0.186	
reg_USA	0.000	1.000	0.162	0.00	0.999	
reg_Scandinavia	−0.273	0.761	0.251	−1.09	0.276	
reg_Other Europe	−0.169	0.845	0.225	−0.75	0.455	
height_listed	−0.072	0.930	0.042	−1.72	0.086	.
weight_listed	0.018	1.018	0.006	3.28	0.001	**
gp_s1_actual	0.003	1.003	0.002	1.26	0.207	
gp_s2_actual	0.005	1.005	0.003	1.67	0.095	.
gp_s3_actual	0.010	1.010	0.003	3.41	0.001	***
gp_s4_actual	0.014	1.014	0.003	4.59	0.000	***

Table 15. S7→S10 enriched specification, L2-penalized logistic regression. $n_{\text{train}} = 1{,}492$, $n_{\text{test}} = 729$. AUC = 0.798, Brier = 0.142. Standard errors and p-values are not reported because L2-penalized coefficients do not support standard maximum-likelihood inference; coefficients are shrunk toward zero. Era dummies omitted. See Section E for discussion of the singular design matrix that required L2 fallback at this horizon.

Variable	β	OR
age_at_first_season	−0.278	0.758
pos2_F	−0.219	0.803
reg_Russia	0.391	1.478
reg_USA	−0.008	0.992
reg_Scandinavia	−0.407	0.666
reg_Other Europe	−0.278	0.757
reg_Other	0.198	1.219
height_listed	0.019	1.019
weight_listed	0.014	1.014
gp_s1_actual	0.003	1.003
gp_s2_actual	0.005	1.005
gp_s3_actual	0.010	1.010
gp_s4_actual	0.013	1.013
pick_1_10	−0.205	0.814
pick_11_31	−0.346	0.707
pick_32_93	−0.061	0.941
pick_94_plus	−0.148	0.862
n_compatriot_teammates_early	0.001	1.001
n_teams_early	−0.058	0.944
orig_team_roster_stability_early	0.807	2.241
roster_size_s1	0.015	1.015

B Birth Region Country Codes

The birth-region categorical variable used in all model specifications maps NHL Edge API country codes to six regions as follows:

Region	Country codes
Canada (reference)	CAN
USA	USA
Russia	RUS
Scandinavia	SWE, FIN, NOR, DNK, ISL
Other Europe	CZE, SVK, DEU, AUT, CHE, FRA, LVA, LTU, EST, SVN, HUN, POL, GBR, ITA, BLR, UKR, KAZ, BEL, SRB, HRV, ROU, NLD, IRL, TUR
Other	All remaining nationalities (e.g., Australia, Japan)

Belarus, Ukraine, and Kazakhstan—former Soviet republics whose players also participate in the KHL—are grouped with Other Europe rather than Russia. This keeps the Russia coefficient narrowly defined to RUS-born players and avoids contaminating it with broader post-Soviet KHL dynamics.

C Sample Composition

Table 16 shows the full-sample and train/test breakdown of skaters by birth region.

Table 16. Skater counts by birth region and split. Column percentages ("% of split") give each region's share of that split; row percentages ("% of region") give the train/test/active breakdown within each region. The "Active" column comprises players with unconfirmed outcomes (censored), excluded from both train and test sets.

	Counts				% of split			% of region
Region	Train	Test	Active	Total	Train	Test	Active	Train
Canada (ref.)	2,334	461	387	3,182	61.4	45.6	41.1	73.4
USA	652	236	306	1,194	17.2	23.3	32.5	54.6
Scandinavia	317	144	134	595	8.3	14.2	14.2	53.3
Other Europe	325	109	63	497	8.6	10.8	6.7	65.4
Russia	157	62	50	269	4.1	6.1	5.3	58.4
Other	15	0	2	17	0.4	0.0	0.2	88.2
Total	3,800	1,012	942	5,754	100.0	100.0	100.0	66.0

D Descriptive Statistics

The test set has slightly longer average careers than the training set, reflecting the secular trend of NHL career lengthening rather than regime change.

Table 17. Outcome variable distributions for training and test sets.

Variable	Train mean	Train SD	Test mean	Test SD
career_gp	303.7	357.0	325.3	374.3
log(career_gp)	4.48	2.02	4.59	2.00
seasons_played	6.4	5.0	6.8	5.1
log(seasons_played)	1.47	0.93	1.57	0.92
N	3,800		1,012	

Table 18. Missing-data rates by variable (full dataset, $n = 5{,}754$).

Variable	Missing	%
height_listed	4	0.1
weight_listed	4	0.1
age_at_draft	1,151	20.0
overall_pick	1,151	20.0
draft_round	1,151	20.0
min_early_gp	16	0.3
early_avg_gp	16	0.3

The 20% missing on draft variables (1,151 / 5,754) reflects undrafted players, who are handled in the enriched specification by treating undrafted as the pick-range reference category.

E Robustness Diagnostics

Table 19. High pairwise correlations flagged in the training set ($|r| > 0.7$).

Variable pair	r
draft_round ↔ overall_pick	0.93
min_early_gp ↔ early_avg_gp	0.84
height_listed ↔ weight_listed	0.72

We use `overall_pick` (via pick-range bins) and drop `draft_round`; we use `gp_s1_actual` through `gp_s4_actual` directly and drop the early-career summary variables. Height and weight are both retained; their correlation is not high enough to cause identification problems in logistic regression. The games-played covariates are available for seasons 1 through 4 only; future work should incorporate later-season measures to improve predictive power at the longer horizons.

E.1 Numerical Stability at the S7→S10 Horizon

The S7→S10 horizon combines a small conditioning cohort ($n_{\text{train}} \approx 1{,}500$) with the highest event-rate ceiling in the design and produces the closest-to-singular design matrices in the paper. Two numerical-stability issues arise at this horizon and warrant disclosure.

First, in the S7→S10 *base* specification (Table 14), unpenalized MLE converges overall but the `reg_Other` coefficient exhibits near-separation: the point estimate inflates to $\beta = 33.77$ with a Wald standard error on the order of 10^{13}, reflecting that the `Other` region cell contains effectively no events of the relevant type at this horizon. We omit this single coefficient from the reported table; all other coefficients in the same fit are well-identified and reported with standard maximum-likelihood inference.

Second, in the S7→S10 *enriched* specification (Table 15), the design matrix is rank-deficient under the full enriched feature set and standard MLE fails to converge. For this single horizon-specification combination we fall back to an L2-penalized (ridge) logistic regression via `scikit-learn`. The resulting coefficients are identified but shrunk toward zero by the penalty; standard maximum-likelihood inference (Wald standard errors, z-statistics, p-values) is not valid for penalized estimates and is therefore omitted. Predictive metrics (AUC, Brier) remain valid because they are functions of predicted probabilities rather than parameter-level inference.

The L2 fallback is confined to the S7→S10 enriched cell — the seven other primary specifications (base and enriched at S1→S3, S3→S5, S5→S8, plus base at S7→S10) all converge under unpenalized MLE and are reported with standard inference.

F Era-Control Robustness

The base-spec landmark logits include five-year era dummies indexed by draft year (`era_1985`, `era_1990`, …, `era_2010`, `era_2015+`). Two features of these dummies merit note given the train/test design. First, the `era_2010` dummy is identified on the subset of train-cohort players whose final NHL season fell within five years of their draft, since the train cutoff (last season $\leq$ 2014–15) implies this maximum tenure for that bucket. Second, the `era_2015+` bucket is unrepresented in the training cohort by construction and is therefore absorbed into the reference baseline (`era_1985`) when applied to test-cohort players from later draft years. To verify these features do not drive the headline findings, we refit the $S3 \to S5$ base logit with all era dummies removed; the continuous covariates retained in the model (age at first season, games played in seasons 1–3, height, weight, draft status, and pick-bin indicators) implicitly carry era-related signal. The Russia coefficient is essentially invariant ($\beta = -1.158 \to -1.103$; OR $= 0.314 \to 0.332$; $p < 0.001$ in both); aggregate Brier and recall improve substantially under the no-era specification, while AUC declines by 0.005. We retain the era-dummy specification as primary for consistency with the broader landmark-survival literature.

References

1. van Houwelingen, H.C.: Dynamic prediction by landmarking in event history analysis. Scand. J. Stat. **34**(1), 70–85 (2007)
2. Dafni, U.: Landmark analysis at the 25-year landmark point. Circ. Cardiovasc. Qual. Outcomes **4**(3), 363–371 (2011)
3. Depken, C.A., Ducking, J., Groothuis, P.A.: Career duration in the NHL: pushing and pulling on Europeans? Appl. Econ. **49**(59), 5923–5934 (2017)
4. Christie, T., Lavoie, M.: Entry discrimination in the NHL: evolution and the KHL effect. East. Econ. J. **41**(2), 214–229 (2015)
5. Fenn, A.J., Gerdes, L., Rothstein, S.: Country of origin effects on the average annual values of NHL player contracts. Int. J. Financ. Stud. **7**, 24 (2019)
6. Schuckers, M.: What's an NHL draft pick worth? A value pick chart for the National Hockey League. Technical report, St. Lawrence University and Statistical Sports Consulting (2011) `https://myslu.stlawu.edu/~msch/sports/Schuckers_NHL_Draft.pdf`
7. Shoniker, C.: Estimating player survivability in the NHL. Technical report, Stats Sports Consulting (2017), `https://www.statsportsconsulting.com/wp-content/uploads/Shoniker_Survival-Analysis-in-the-NHL-Draft.pdf`
8. Wang, Y., Csiernik, B., Johnston, K., Baker, J.: Age effects in NHL draftees: a data-driven review of a multi-dimensional concept. Front Sports Act Living **7**, 1593409 (2025)
9. Davis, J., Bransen, L., Devos, L., Jaspers, A., Meert, W., Robberechts, P., Van Haaren, J. and Van Roy, M.: Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned. Machine Learning **113**, 6977-7010 (2024).
10. NHL Edge API, `https://api-web.nhle.com/v1`, last accessed 2026/04/14
11. Hockey Reference, `https://www.hockey-reference.com`, last accessed 2026/04/14