

James A. Brander*, Edward J. Egan and Louisa Yeung

Estimating the effects of age on NHL player performance

Abstract: Using NHL data for the 1997–1998 through 2011–2012 seasons, we examine the effect of age on scoring performance and plus-minus for NHL skaters (non-goalties) and on save percentage for goaltenders. We emphasize fixed-effects regression methods that estimate a representative age-performance trajectory. We also use a method based on the best performances over time, a method based on the age distribution of NHL players, and a “naïve” specification that does not correct for selection bias. In addition we estimate individual age-performance relationships to obtain a distribution of peak ages. All methods provide similar results (with small but understandable differences) except the naïve specification, which yields implausible results, indicating that correcting for selection bias is very important. Our best estimate of the scoring peak age is between 27 and 28 for forwards and between 28 and 29 for defencemen. Both forwards and defencemen exhibit near-peak performance over a wide range, going from about 24 to 32 and 24 to 34, respectively. Goaltenders display little systematic performance variation over most of the age range from the early 20s to late 30s.

Keywords: age and performance; fixed-effects regression; hockey; NHL; selection bias.

***Corresponding author: James A. Brander**, Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC, Canada V6T 1Z2, e-mail: james.brander@sauder.ubc.ca
Edward J. Egan: National Bureau of Economic Research (NBER), 1050 Massachusetts Avenue, Cambridge, MA 02138, USA
Louisa Yeung: Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC, V6T 1Z2 Canada

1 Introduction

The effect of age on performance in sports is a subject of longstanding interest that has attracted considerable academic attention. The age-performance relationship in major league baseball (MLB) has been extensively

studied, starting with the work of sabermetric¹ pioneer Bill James (1982) and including considerable academic research, such as Albert (1999, 2002, 2009), Fair (2008), and Bradbury (2009). Individual timed sports such as running, cycling, swimming and triathlon have also been extensively studied, as described in a recent review article by Lepers, Knechtle and Stapley (2013), and many other sports have received at least some academic attention, particularly golf, as in Tiruneh (2010).

The primary objective of this paper is to provide an assessment of the age-performance relationship in the National Hockey League (NHL), including estimating the age of peak performance, referred to as the “peak age.” We assess and compare the age-performance relationship for the three different position categories – forwards, defencemen, and goaltenders. We emphasize fixed-effects regression methods, but we also compare the resulting estimates with information from participation data, with what we call the “elite performance” method, and with a “naïve” method that does not correct for selection bias. We also use a method based on the estimation of individual age-performance functions. Therefore, in addition to providing what we believe is the most reliable assessment of age effects on performance in the NHL currently available, we also seek to provide a methodological contribution to the literature on age effects in sports by comparing a broad set of relevant methods.

There is a conventional wisdom about the age-performance relationship in the NHL that can be inferred from many NHL “blogs,” bulletin boards, and from the general hockey media. See, for example, Chen (2010), who quotes an NHL general manager as saying that a 27-year-old forward has “his best hockey years ahead of him.” Chen goes on to do his own assessment based on the careers of 11 star players and concludes that “peak performance does seem to follow the general notion of 27–32” as the peak period. Many other blogs provide similar comments. However, as originally pointed out by James (1982) and emphasized by many others, assessing the age-performance relationship is susceptible to the problem of selection bias. Although this problem is well-known, correcting for it effectively is not easy. However, there is much to be gained from a careful analysis of the data using formal statistical methods that seek to explicitly correct for selection bias.

¹ Sabermetrics refers to the statistical analysis of baseball. The term was originally coined by Bill James and is derived from the acronym SABR, which stands for the Society for American Baseball Research.

The selection bias problem can be understood by considering the effect of looking at how players of different ages perform in a given season. It would, for example, be very misleading to take the difference between the average NHL performance of 19 year olds and the average performance of 26 year olds in a given year as a measure of how much players improve between the ages of 19 and 26. Relatively few 19 year olds play in the NHL as only the very best 19 year olds get significant NHL playing time. Therefore, this comparison based on cross-sectional variation in performance creates bias by selecting only the very best 19 year olds to compare with the full range of ability levels at age 26 (the modal age). This selection bias leads to a considerable understatement of the gains in performance that occur in a player's early 20s. Similarly, only the very best players play into their late 30s, so comparing their performance with the full range of players at age 26 understates the extent of age-related decline. We refer to the method of simply comparing the performance of NHL players of different ages without correcting for selection bias as the "naïve" method of estimating the age-performance relationship.

A method of correcting for selection bias used by James (1982) and Albert (2002) for baseball, and by many others for a variety of sports, is what we call the *elite player* method. This method selects a small number of elite players with long careers and tracks how their performance varies with age. If each of these players reaches his or her peak performance at age 28, for example, then we might reasonably conclude that 28 is the general age of peak performance. The logic of this method is that estimates of age-performance relationships are based on comparisons over time for a given player (within-player comparisons) rather than on comparisons across players.

Modern panel data fixed-effect regression methods are a generalization of this elite player approach, except that all players in the data are used, greatly increasing the reliability of the results as a guide to typical player performance. The effects of age on performance are based entirely on within-player comparisons – how each player's performance evolves over time. It is possible to use both parametric and nonparametric methods. A parametric method assumes a particular functional form for the dependence of performance on age – often called the "aging function" – and then estimates the parameters of this function. A non-parametric method does not assume a specific functional form and instead estimates the predicted performance level at each age numerically – imposing no particular constraint on the form of this relationship. Parametric methods have the advantage of requiring less data to obtain statistically significant results because of

the maintained assumption that we know the functional form of the relationship. Non-parametric methods have the advantage of being largely free from errors created by misspecification of functional form.

We use parametric methods based on the quadratic and cubic functional forms that are typically used for assessing age-performance relationships. We also use what might be called a non-parametric method using only categorical age variables (in addition to player-specific fixed effects) as regressors. As pointed out by a reviewer, however, much of the emphasis in non-parametric estimation is on smoothing algorithms that we do not use here. To avoid overstatement, we therefore we refer to this method as "regression with categorical age variables." Using both approaches is useful as the comparison of the two methods is instructive. Moreover, information learned from a categorical variable approach can be used to inform how to proceed in parametric analysis, improving the confidence we can place in parametric methods.

Another method that is sometimes used to assess the effect of age on performance is the *elite performance method* which compares the best performances at different age levels. See, for example, Berthelot et al. (2013), who apply the elite performance method to several different sporting activities.² We use an elite performance method that considers the top 10 players for each position in each age category. This method corrects for the primary selection bias problem, and it provides additional information of interest. Specifically, it assesses the age-performance relationship for the very best players, which might differ from the average age-performance relationship estimated by regression methods using the full sample of NHL players.

Our fixed-effect regressions assume an underlying dependence of performance on age (aging function) that is common across players playing a given position. Differences between players are assumed to arise from player fixed effects (some players are better than others) and from the (individual and age-specific) error term in the regression. We estimate the underlying aging function and use it to infer an estimated age of peak performance for each position. However, as emphasized by Albert (1999, 2009) for baseball, there is variation across players in the underlying aging function: Some players peak earlier than others and some decline more quickly with age than others. Therefore, as suggested by a reviewer, we also use a method that estimates player-specific aging functions.

² There is also related work by Kovalchik and Stefani (2013) that uses performances of Olympic medalists over time to assess time-related improvements in athletic performance.

This method yields very similar estimates for the average or typical age of peak performance as the other non-naïve methods, and also provides interesting information about how the peak age varies across players.

Throughout this paper we refer to player “performance.” For NHL “skaters” (forwards and defencemen) we use points scored and plus-minus. We acknowledge that these measures, while of great interest, have well-known limitations. Using points scored (goals plus assists) has the limitation that it considers only the offensive half of the game even though preventing goals is just as important as scoring. This measure is particularly incomplete for defencemen, as preventing goals is their primary function. The plus-minus statistic is the number of even strength and shorthanded goals scored by a player’s team while that player is on the ice minus the number of even strength and shorthanded goals scored by the opposing team while that player on the ice. Thus plus-minus attempts to incorporate both offensive and defensive contributions. However, it is subject to the limitation that it depends in large part on the quality of other players on the ice – both teammates and opponents. For goaltenders we use save percentage – the ratio of saves to shots on goal, which is widely accepted as a measure of performance.

Considerable effort has gone into developing more accurate measures of player performance than points scored and plus-minus. See, in particular, Gramacy, Jensen and Taddy (2013), Macdonald (2011), Vollman (2013), and the various metrics described at www.behindthenet.ca and elsewhere. However, we focus here on scoring and plus-minus for a variety of reasons. First, these metrics remain by far the most transparent and widely used measures of player performance. Second, extensive data on these measures is readily available. And they certainly are closely related to performance, even if not perfect measures.

Furthermore, as described more fully in Section 4, we believe that our implementation of fixed-effects regression largely corrects for deficiencies in points scored and plus-minus as performance measures because this method looks at within-player comparisons. As long as overall performance for a given player over time is correlated with scoring and/or plus-minus for that player, as seems likely, then our estimates of the age profile of scoring and plus-minus would also apply to overall performance. In any case, we would suggest that the effects of age on scoring and plus-minus (and on save percentage) are of interest in themselves, even if these metrics are not the best possible measures of performance.

The remainder of this paper is organized as follows. Section 2 describes the data. Section 3 presents evidence

of performance based on participation data. Section 4 describes our fixed-effects regression methodology and presents our main results. Section 5 provides comparative results based on the elite performance method and the naïve method and assesses the importance of selection bias in NHL data. Section 6 deals with individual variation in the age-performance relationship and Section 7 provides a discussion of how our results relate to the sports science literature on age-related variations in basic physiological functioning. Section 8 contains concluding remarks.

2 Data

Our primary data source is the widely used publicly available data set provided by the NHL at www.nhl.com. We use data from the 1997–1998 season through the 2011–2012 season. As the 2004–2005 season was lost due to a labor dispute, this comprises 14 years of data. We judged this to be a long enough period to contain sufficient time series information while being sufficiently recent to be reflective of the current situation. In the NHL, as in other professional sports leagues, there are long run variations in scoring and other variables that make comparing players across different eras difficult.³ The years covered in our analysis should be comparable.

We define age to be a player’s age on January 1 – approximately mid-season. For some purposes we treat age as a continuous variable, in which case we use the player’s exact age on January 1.⁴ For example, a player who turned 27 on November 26, exactly 36 days (10% of a year) before January 1, would be counted as 27.1 – that player’s age on January 1. For some purposes we round off age to the nearest whole number as of January 1 to create convenient age categories: 26 year olds, 27 year olds, etc.

In assessing scoring, it is important to decide whether to use points scored in a season, points per game played, or points per minute played. We focus on points per game as the primary measure of scoring performance. Points per

³ One very thoughtful effort to use statistical methods to compare players across eras in a variety of sports, including in the NHL, is provided by Berry, Reese and Larkey (1999).

⁴ Addona and Yates (2010) provide a very interesting paper investigating the effects of birth month on NHL performance. They investigate the hypothesis that players born early in the year have an advantage due to being grouped with children several months younger than themselves throughout childhood and therefore exhibit better relative performance and get more attention, better coaching, etc.

game is preferable to points per season because a large source of variation in total points per season is due to variation in games played, and the most important source of variation in games played is injury. We take the view that we should correct for time lost due to injury (or for other reasons) by using points per game.

Similarly, one important source of variation in points per game is minutes played. Players who play more minutes will have more opportunities to get points. We can correct for this effect by using points per minute played. However, as pointed out by a reviewer, the ability to play a large number of minutes at a high level is an important performance characteristic. Furthermore, coaches choose to give players who are playing better more playing time. In fact playing time is sometimes used as a measure of performance. Therefore, we use points per game as the primary performance measure. However, we also report some results using points per minute. Using points per game instead of points per minute implies a slightly later peak age (about a year later for forwards and less than a year later for defencemen). This effect reflects the fact that younger players tend to play fewer minutes and is consistent with the sports science literature (such as Shultz and Curnow 1988) indicating that endurance does not peak until the late 20s.

There is also some question as to whether we should adjust plus-minus on the basis of games played (or minutes played). We have done the analysis with and without such adjustments for plus-minus and find that it makes little difference. This is not surprising as no bias is created in plus-minus by varying the number of games played. Although the expected value of points scored is increasing in games played, the expected value of aggregate plus-minus is always nearly zero.⁵ The variance of plus-minus does increase with games played, however. As the season progresses we get larger positive and negative values. A player who plays only 20 or 30 games, for example, would typically have a smaller absolute value of plus-minus than a player who plays a full season of 82 games. By using a player's plus-minus for all games played in a season (total plus-minus), our regression methods effectively put more weight on players who play more games. We view this as a desirable property. As there

is no offsetting bias to be concerned about, we report results only for total plus-minus in this paper.

We do a certain amount of data-cleaning, including dropping age categories with very small sample sizes. Specifically, we drop players under 19 and over 40 from the data. Cleaning the data in this way has very little effect on the results, but it drops age categories for which the results might be dominated by one or two idiosyncratic observations and for which the sample sizes are so small that performance estimates for that age group have large standard errors and are therefore very imprecise. For example, there is only one defenceman under the age of 19 in the full data set, along with only four 41 year olds, three 42 year olds, etc. Sample sizes for ages 19–40 are sufficient to draw meaningful statistical inferences for most purposes.

We also drop player-years in which a player played <20 NHL games on the grounds that performance assessments based on that small a sample of games adds more noise than useful information to the analysis. There are, for example, cases where a young player called up late in the season to get some experience plays in only two or three games but is fortunate enough to get a couple of points. Such a player could lead a team in “points per game” or “points per minute.” It seems clear that such observations should be dropped from the data set. This small amount of data-cleaning has very little effect on the results. After this data-cleaning process we have 2033 players and a total of 9901 player years.

We look only at regular season performance. Playoffs have the disadvantage that only about half the players in the league play in the playoffs in any given year and most of those players play only a small number of games. And of course the quality of the opponents varies by round, creating significant difficulties in interpreting performance. Table 1 provides summary data on player

Table 1 Descriptive statistics for player age and performance by position.

	25th percentile	median	75th percentile	average	standard dev.
Forwards					
Age	24.3	27.2	30.8	27.7	4.4
PPG	0.23	0.40	0.63	0.45	0.28
Plus-minus	-7.0	-1.0	5.0	-0.63	10.5
Defence					
Age	24.7	27.7	31.3	28.1	4.5
PPG	0.14	0.23	0.37	0.28	0.17
Plus-minus	-7.0	0.0	7.0	0.48	11.5
Goalies					
Age	26.2	29.0	32.6	29.4	4.3
Save%	90.0	90.7	91.6	90.7	0.012

⁵ With even-strength play (also called five-on-five play) every goal results in an addition of 1 for all five skaters on the scoring team and a subtraction of 1 for all skaters on the other team, so the average is always zero. However, with shorthanded goals the four players on the penalty kill that scores the shorthanded goal record +1 while all five players on the powerplay scored against record -1. No pluses or minuses are given for powerplay goals. Therefore the average plus-minus is slightly negative as more players get a negative than get a positive for shorthanded goals. Given the very small relative number of shorthanded goals, this effect is very small.

age and performance. The points data is shown as points per game (PPG). A full season is 82 games so PPG can be translated to a full season by multiplying by 82, although most players do not play all 82 games due to injury and for various other reasons. The plus-minus numbers shown are full season totals.

Even these summary statistics provide interesting information, including differences between positions. Defencemen are slightly older than forwards at all percentiles. And goaltenders are older still. It is also worth emphasizing that scoring by defencemen, while less important than for forwards, is still a very important aspect of performance. The average PPG number for defencemen is 0.28, which is about 62% of the average PPG for forwards.

Our data is set up as panel data. The unit of observation is the individual player and that player is tracked through time. The panel is unbalanced as different players are in the league for different time periods. This dataset provides the opportunity to apply a variety of statistical techniques. We use six different methods in total. To keep the logical development of our exposition clear, we start with the participation method, then discuss categorical and parametric fixed-effects regression methods. We then consider the elite performance method and the naïve method for the purposes of comparison and to assess the relative importance of selection bias. Finally we introduce a method that focuses on individual variation in the age-performance relationship.

3 Player participation as an indicator of performance

The most common ages in the data should indicate the ages of peak performance. Many players who play during these most common years are not good enough to play in the NHL when younger or when older, but just manage to make the NHL when playing at their peak performance levels. Therefore, the age categories when we observe these players in the NHL should be the age of peak performance. Higher quality players would play during their peak years *and* for significant parts of their non-peak years. And only the very best players would be in the NHL for age categories a long way from peak performance. Therefore this *participation method* would use the relative frequency of different player ages as an indicator of performance at different ages.

The player participation method is very easy to implement as it does not use performance measures at all, apart from the binary indicator of whether the player is good

enough to play in the NHL. All it does is to identify, by position, the number of players of each age in the NHL. The relative frequencies for forwards and defencemen are shown in Figure 1 and for goalies in Figure 2.

The age distributions for the three positional categories are similar, but there are noticeable differences, with forwards being the youngest group, defencemen in the middle, and goaltenders the oldest. The modal or most common age for forwards is 25, but all 4 years in the range 24 through 27 are very similar. The peak frequency for defencemen is a year later, at age 26, in the middle of a 3 year peak range covering ages 25 through 27. The age distributions of both forwards and defencemen are skewed to the right. Goaltenders have their peak frequency at age 28 and a clear 4-year peak period that runs from 26 through 29. Therefore, if we use peak participation as an indicator of peak performance we would predict peak performance for forwards at age 25, for defencemen at age 26, and for goalies at age 28 with near peak performance for 1 or 2 years on either side.

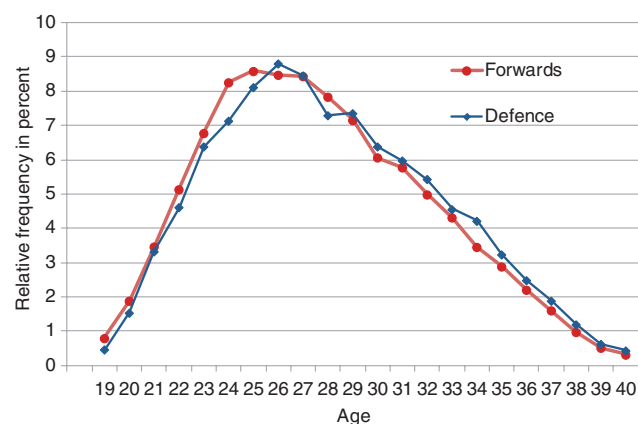


Figure 1 NHL player age distributions for forwards and defence – 1997–2012.

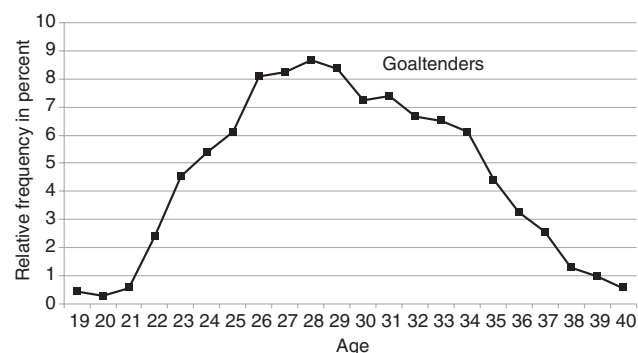


Figure 2 The NHL player age distribution for goaltenders – 1997–2012.

This simple analysis based on Figures 1 and 2 contains useful information but is incomplete in several ways. First, the precise peak for each position is determined by the marginal players – players just barely good enough to make the NHL for only a few years. As we show later, higher quality players tend to peak later so focusing on marginal players understates the typical peak age. Second, some players drop out due to injury rather than declining underlying ability. In addition, many marginal players leave the NHL voluntarily once they reach their late 20s rather than bounce back and forth between the NHL and the minor leagues even though they might still be good enough to continue in this marginal role.⁶

These sources of bias imply that the participation method would understate the age of peak performance. Furthermore, it is unlikely that participation frequencies would track the relative level of performance for the full range of relevant ages, even if it does pick out something close to the age of peak performance. However, keeping these points in mind, the participation method is informative. It is instructive to compare inferences drawn from participation data with the results obtained from regression methods to see if, as we expect, those estimates imply slightly older peak ages.

4 Fixed-effects regression estimates of the age-performance relationship

A fixed-effects regression is based on a regression specification of the following type:

$$y_{it} = f(x_{it}) + u_i + e_{it} \quad (1)$$

where y_{it} is the performance of player i at time t , x_{it} is player age at time t , $f(x_{it})$ is some function of player age (and may incorporate a constant term), u_i is a fixed effect for player i , and e_{it} is a random error applying to player i at time t .

We refer to $f(x)$ as the aging function, as it shows the effect of age on performance. The full age-performance

trajectory for a given player incorporates this aging function along with the fixed effect u_i and the trajectory of random errors. In this section we assume that $f(x)$ is common across players, with variation across players being captured by the fixed effects and by the error term. (Individual aging functions are considered in Section 6.) If we specify the functional form of $f(x)$ (often taken to be quadratic or cubic), then the resulting regression is parametric. If we assume that the form of $f(x)$ is unknown and we seek to construct $f(x)$ numerically using an estimation procedure, the approach is nonparametric.

The logic of the fixed-effects approach is that estimates of the effect of player age are determined by looking at changes over time for individual players. Suppose player A scores 30 points at age 22 and 40 points at age 26, while player B scores 60 points at age 22 and 70 points at age 26. Using a fixed-effects estimator we would infer that 26 year olds tend to score about 10 points more than 22 year olds. The fact that player B scores more as a 22 year old than player A scores as a 26 year old would not matter. The fixed-effects estimator is sometimes called a *within* estimator because it is based on variation within each unit. The fact that some players do not play in the NHL early in their careers and that some retire early even when they are good enough to continue playing does not bias the analysis in any way. Thus, the fixed-effects estimator is precisely what we want to use to solve the selection bias problem discussed in the introduction.

Fixed-effects estimators are widely used in general, although we have found only a few examples in sports. Applications of fixed-effects estimators to sports include Fair (2008) for baseball, Arkes (2010) for basketball, and Broadie and Rendleman Jr (2013) for golf. In all three cases, fixed-effects are used to identify and control for individual player ability, as in this paper. It is also possible to use a *random effects* approach, which incorporates cross-sectional as well as within-player variation over time to estimate the age-performance relationship. Computationally, the random effects estimator is a weighted average of the fixed-effects estimator, which is unbiased in our case, and an estimator based on cross-sectional variation, which is biased in our case. It therefore seems clear that the fixed-effects estimator is preferred in our context. However, there is relatively little difference in the results if random effects estimation is used, reflecting the fact that our data imply putting little weight on the cross-sectional aspect of variation by age. We therefore report only the fixed-effects panel regressions.

One paper that uses random effects to estimate the effect of age on performance in a sport is Bradbury (2009)

⁶ As of 2013–2014 a typical marginal player who will divide a season between an NHL team and its minor league American Hockey League (AHL) affiliate or “farm team” has a “two-way contract” that might pay \$500,000 to \$900,000 per year on a pro-rated basis for games played in the NHL, but perhaps only \$70,000 or \$80,000 per year on a pro-rated basis for games played in AHL. Salary data can be found on the website capgeek.com maintained by Matthew Wuest.

for baseball. He reports the random effects results rather than the fixed-effects results on the grounds that when the fixed-effects estimator was used in combination with a correction for possible serial correlation in the errors, the results were implausible. We note that computational problems can arise when estimating fixed-effects models with a correction for serial correlation. However, we have tried the serial correlation correction without difficulty in our case and find very similar results to those we report. We actually use a more general correction that deals with both possible serial correlation and heteroscedasticity by using the cluster option in STATA 11.

The fixed-effects estimator should correct for the major selection bias problems. In addition it mitigates other limitations in our performance measures. Consider, for example, points scored as a measure of performance. As previously noted, scoring looks only at offensive performance and neglects defensive performance. However, our assessment of the effect of age on performance is based on each player's changes in performance over time. Therefore, as long as a player's scoring performance is positively correlated with that player's overall performance, as is likely in most cases, focusing on changes in scoring will also provide an indication of changes in overall performance. The relationship between points scored and plus-minus is in fact strong enough to imply a positive correlation between improvements in scoring and improvements in defensive play.

The fixed-effects approach also mitigates the problem that plus-minus performance measures are affected by the quality of other players – teammates and opponents. Players on weak teams tend to have negative plus-minus numbers while players on good teams tend to have positive plus minus numbers. However, if a player on a bad team improves from -10 to -5 that increase of 5 counts just as much an increment of 5 for a player on a good team who improves from 5 to 10. More generally, if the net quality of other players on the ice (teammate quality minus opponent quality) is uncorrelated with age, then leaving this net quality out of the regression would not bias the estimated effect of age. As a check, we have estimated some specifications with team goal differential as a control variable and find no virtually no effect on estimated peak ages.

4.1 Fixed-effects regression with categorical age variables

Parametric methods have the disadvantage that the estimates and the significance levels are subject to the assumption that the functional form $f(x)$ in equation (1) is known and correctly specified. An incorrect functional

form can lead to significant biases, as well as other problems. Nonparametric estimation has the advantage that the form of the regression function is determined by the data. However, much more data is required in order to achieve statistical significance than with parametric methods, precisely because the data must determine the form of the regression model. Semi-parametric methods are also possible and include an unknown functional form f over some explanatory variables and a specified functional form over other explanatory variables.

Our first fixed-effects method is based on categorical age variables. Thus we do not specify any particular form for $f(x)$ but simply assume that it varies with age in a way to be determined by the data. It could be argued that specifying linear fixed effects makes our approach semi-parametric. Our approach could also be called a nonparametric model with fixed effects.⁷ We refer to the method simply as fixed-effects regression with categorical age variables. We use the panel data fixed-effects procedures available in the statistical program STATA 11 and would recommend the textbook by Wooldridge (2008) as a good source on panel data methods.

For our categorical regressions we continue to treat age in 1-year intervals as we did to estimate participation frequencies (19 year olds, 20 year olds, etc.). We then estimate the expected performance for each age level by regressing a measure of performance (such as points per game) on age-specific dummy variables, while incorporating player-specific fixed effects by using the STATA fixed-effects estimator. We treat 19 year olds as the base category, so the estimated coefficient for each age shows the expected difference at that age relative to age 19. For each age category, the sum of the coefficient on the age-specific dummy variable and the constant (which provides the coefficient for 19 year olds) then provides the expected performance at that age for an average NHL player (i.e., for $u_i=0$).

We carry out this exercise for forwards, defencemen and goalies separately. The results for forwards are shown in Table 2. The implied full season total for points can be obtained by multiplying points per game by 82. Thus, for example, the representative point total for a 19-year-old forward who played all 82 games in a season would be $(82)(0.162)=13.3$, much less than the implied 82 game total of a representative 28 year old, which is $(82)(0.511)=41.9$. Points per 60 min played are also shown. (A team's best forwards typically play about 20 min per game or slightly more, while the marginal players might play only 5 to 10 min in a game.) Plus-minus numbers are also shown.

⁷ See, for example, Henderson, Carroll and Li (2008).

Table 2 Age-performance relationships for NHL forwards.

Age	Pts. Per Game	Std. Error	Pts. per 60 min.	Std. Error	Plus-Minus	Std. Error
19	0.162	0.032	1.18	0.082	-7.48	1.76
20	0.290	0.029	1.49	0.080	-3.68	1.81
21	0.352	0.032	1.64	0.084	-1.16	1.82
22	0.396	0.032	1.70	0.082	-0.25	1.75
23	0.442	0.033	1.79	0.084	1.27	1.81
24	0.480	0.033	1.85	0.083	1.40	1.86
25	0.487	0.033	1.83	0.085	1.12	1.84
26	0.489	0.033	1.81	0.085	0.22	1.84
27	0.507	0.034	1.84	0.088	1.07	1.86
28	0.511	0.035	1.84	0.090	0.38	1.88
29	0.507	0.034	1.83	0.088	0.07	1.87
30	0.486	0.034	1.75	0.089	-1.09	1.86
31	0.477	0.035	1.76	0.089	-1.47	1.90
32	0.450	0.034	1.70	0.088	-1.90	1.89
33	0.436	0.036	1.66	0.092	-1.39	1.93
34	0.415	0.036	1.62	0.092	-2.68	1.96
35	0.373	0.037	1.50	0.095	-3.28	2.02
36	0.328	0.038	1.42	0.098	-4.55	2.07
37	0.311	0.039	1.39	0.101	-5.40	2.17
38	0.250	0.039	1.24	0.106	-11.90	2.40
39	0.192	0.043	1.16	0.113	-10.64	2.39
40	0.165	0.061	1.21	0.154	-10.66	3.71

Peak and near-peak performances are shown in bold.

For forwards, peak performance in points per game is reached at age 28, but the 8 years from age 24 through 31 (in bold) exhibit similar scoring performance. If we focus on points per minute played, we implicitly adjust for the fact that younger players tend to play fewer minutes. This adjustment reduces the estimated age of peak performance. However, there is very little variation in points per 60 min over a fairly long period of near-peak performance – covering 9 years from age 23 through 31. Standard errors are shown to the right of the point estimates.

The initial plus-minus at age 19 is not very good, but forwards improve consistently for the 5-year period from 19 through 23. At age 23 they enter a 3-year peak period, then have some decline in the late 20s and a more significant decline through their 30s. The analysis implies that the average forward would have a large negative plus-minus in his late 30s and at age 40. Of course average players are no longer playing at age 39 or 40. The actual players still playing at those ages are much better than average players and can still compete effectively. Therefore, the actual plus-minus scores of the players still playing at that age are much better than shown in Table 2. Even so, such players are not nearly as good in their late 30s as they were 8 or 9 years earlier and their decline in performance contributes to the estimates in Table 2. Plus-minus peaks earlier than points per game. Table 3 shows the results for defencemen and goaltenders.

Table 3 Age-performance relationships for NHL defencemen and goaltenders.

Age	Defencemen						Goaltenders	
	Pts. Per Game	Stand. Error	Pts. per 60 min.	Stand. Error	Plus-Minus	Stand. Error	Save %	Stand. Error
19	0.135	0.030	0.63	0.076	-5.12	2.96	89.32	0.54
20	0.219	0.032	0.79	0.082	0.21	2.93	90.20	1.03
21	0.205	0.028	0.72	0.076	-0.84	2.98	90.08	0.82
22	0.243	0.030	0.79	0.077	-0.99	3.13	90.91	0.66
23	0.269	0.031	0.83	0.078	1.21	3.16	91.02	0.56
24	0.272	0.031	0.83	0.078	0.35	3.04	90.60	0.56
25	0.275	0.031	0.82	0.077	1.58	3.01	90.99	0.58
26	0.296	0.031	0.86	0.078	1.20	3.03	90.97	0.55
27	0.293	0.031	0.84	0.077	1.73	3.08	90.89	0.56
28	0.300	0.032	0.87	0.079	0.04	3.08	91.01	0.56
29	0.305	0.033	0.87	0.081	1.67	3.12	90.81	0.54
30	0.304	0.033	0.87	0.082	0.59	3.15	90.76	0.57
31	0.297	0.033	0.84	0.081	1.33	3.12	90.64	0.57
32	0.298	0.034	0.86	0.083	0.49	3.16	90.46	0.58
33	0.296	0.034	0.85	0.084	-0.77	3.22	90.64	0.58
34	0.270	0.035	0.81	0.085	-1.91	3.24	90.70	0.58
35	0.247	0.034	0.77	0.086	0.83	3.32	90.36	0.60
36	0.212	0.034	0.70	0.085	-1.08	3.28	90.10	0.63
37	0.227	0.036	0.77	0.091	-1.27	3.57	90.33	0.65
38	0.183	0.039	0.71	0.101	-0.78	3.66	89.92	0.73
39	0.205	0.041	0.79	0.103	-1.92	4.14	90.40	0.70
40	0.153	0.042	0.66	0.104	-4.32	5.49	89.61	0.64

Peak and near-peak performances are shown in bold.

Defencemen enter their scoring peak later than forwards – at about 26. The highest level of scoring per game occurs at age 29, which is also the best year for plus-minus. Once again, however, there is a fairly long period of near-peak performance within which year to year differences are small for scoring and somewhat erratic for plus-minus. Putting scoring and plus-minus information together suggests a period of peak performance lasting from about age 26 through 32. Goaltenders exhibit little age-related change in performance. There is some indication that very young goaltenders (19 year olds) and goaltenders at the upper limit (age 40) operate below peak performance, but the period in between is remarkably stable.

The inference for the age of peak performance implied by these fixed-effect regressions with categorical age variables has some similarity to the participation method suggested by Figure 1 depicting NHL participation rates by age. However, the regression-based analysis yields a later peak (using scoring per game as the performance measure) and a longer period of near-peak performance. Thus participation rates understate the peak age and overstate age-related decline, as is consistent with the biases inherent in the participation method discussed earlier. Retirement due to injury or for voluntary reasons rather than due to declining

performance is a particular concern with participation rates that is not shared by fixed-effects regression methods.

Figure 3 illustrates a number of important aspects of the analysis. The figure shows the age-scoring relationship for points per game played for forwards and defencemen with 90% confidence intervals.

The overall pattern of the age-scoring relationships is statistically significant, but the year-to-year differences between any two adjacent years are not statistically significant except near the limits of the age range. Furthermore, the actual age-specific estimates are somewhat erratic. We do not really believe that expected scoring per game for defencemen would fall from 0.219 to 0.205 as the player ages from age 20 to 21 and then rise back to 0.243 at age 22. We do not have enough data to separately estimate scoring for each age category with a high degree of precision, as is a common problem with such methods. This up and down pattern could be smoothed using more sophisticated nonparametric measures based on kernel density estimation, but we do not have enough data to use such methods effectively. An alternative method of smoothing is to fit a trend-line to the estimates shown in Figure 3.

A casual look at Figure 3 suggests that a quadratic trend line would fit the data well, although the scoring performance for forwards shows some skewness to the right – forwards improve more quickly than they decline. Thus a cubic function might be more appropriate. However, if we are going to implicitly smooth the results in this way, it makes sense to estimate a parametric model directly, which we do in the next section.

To save space we do not show the corresponding diagrams for plus-minus or for save percentage (for goaltenders). The plus-minus estimates are, as noted earlier, similar in pattern to the scoring relationships but are more erratic, especially for defencemen. The goaltender save percentage exhibits no meaningful age-based trajectory.

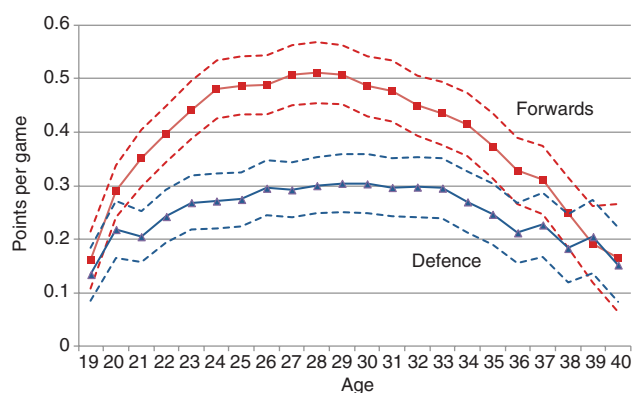


Figure 3 The effects of age on scoring using categorical regression with fixed effects.

4.2 Parametric regression

Parametric methods have the advantage of requiring less data and of imposing plausible smoothness or regularity on the data. In addition, such methods often provide a closed form algebraic representation of the relationship that can be used for a variety of purposes, including providing a compact summary of the age-performance relationship. We can think of imposing a functional form as a way of incorporating (and taking advantage of) prior information. In this case we have a lot of prior information about athletic performance implying that it first improves with age and then declines and is single-peaked, and there is considerable prior information that performance over time in many areas is well-approximated by a quadratic function.

In fact, the quadratic form fits the data well. However, quadratic functional forms imply symmetry – performance declines in later years at the same rate as it improves in earlier years. There is evidence from our fixed-effect regressions with categorical age variables that rates of improvement and decline differ, at least for forwards. Therefore we do not wish to impose symmetry on the aging function. A cubic functional form allows for skewness and nests the quadratic form as a special case. The statistical significance of the cubic term provides a test of whether the skewness is significant. We estimate cubic and quadratic specifications for both forwards and defencemen. In these regressions we treat age as a continuous measure (such as 26.1) rather than relying on integer approximations. We do not report results for goaltenders as there is no discernible age-related pattern of performance for goaltenders.

For a cubic specification we estimate an equation of the form

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 x_{it}^2 + \beta_3 x_{it}^3 + u_i + e_{it} \quad (2)$$

where β_0 , β_1 , β_2 , and β_3 are regression coefficients and, as in equation (1), y_{it} is the performance of player i at time t , x_{it} is player age at time t , u_i is a fixed effect for player i , and e_{it} is a random error for player i at time t . To obtain a quadratic specification we constrain β_3 to be zero.

As each player has a fixed effect – a personal constant – and there is an additional common constant, β_0 , there are “too many” constants and some normalization is required. Rather than simply dropping β_0 we adopt the more common convention of requiring the average player-specific fixed effect to be zero. The estimate of β_0 is the estimated common or average fixed component of performance and the estimated player-specific fixed effect shows the deviation from this mean. Therefore, if we use

the coefficients reported in Table 4 for a given age, assuming the player-specific fixed effect to be zero, we obtain the predicted average performance for that age. Table 4 shows the major results for forwards for points per game, points per minute (actually points per 60 min played) and plus-minus.

In the quadratic regression in column (1) explaining points per game, both age and age squared are highly significant with the expected signs. An optimal age of 28.1 is implied. However, performance for forwards has a small but statistically significant skew to the right as improvement occurs more quickly than decline. Thus the cubic term in column (2) is statistically significant at the 0.01 level. The implied age of optimal scoring performance is 27.6 for this specification, so we take 27.6 as our best estimate of the peak age for scoring by forwards. The 90% confidence intervals for the peak ages are fairly tight. However, the estimated aging function implies a long period of near peak performance as forwards would be within 90% of peak performance for ages 24 to 32.

The estimated peak age is determined from the coefficients of the estimated quadratic or cubic function and is therefore a function of these estimated coefficients. As the coefficient estimators are random variables, so is the peak age estimator. However, its closed form distribution is not easily determined. We therefore obtain estimated standard errors using a Monte Carlo method in the form of the bootstrapping option available in STATA 11.

The third column contains a cubic specification explaining points per 60 min played. This regression implies an earlier peak age for forwards (by 1 year), reflecting the fact that younger forwards typically play fewer minutes per game than others. This is essentially the same result as is shown in Table 2 for the fixed-effect regressions. As for plus-minus performance, the cubic term is not statistically significant, suggesting that the quadratic specification is preferred. The implied optimal age for plus-minus is 26.2. We also ran a version of the regression including team goal differential as a control variable. This control variable is highly significant, but age and age squared remain highly significant and the implied peak age is unchanged at 26.2. Table 5 shows the age-performance regressions for defencemen.

For defencemen, a cubic form adds no explanatory power and all of the age coefficients lose their statistical significance. The improvement and decline of defencemen is very symmetric and slower than for forwards. Also, as suggested by our regressions with categorical age variables and by participation rates, defencemen peak on the order of 1 year later than forwards for all measures. And defencemen stay close to peak performance longer. In the

Table 4 Age-performance regressions for NHL forwards.

Variables	(1)	(2)	(3)	(4)	(5)
Age	0.161*** (0.0098)	0.373*** (0.065)	0.798*** (0.177)	3.449*** (0.526)	8.496** (3.74)
Age squared	-0.0029*** (0.00017)	-0.0103*** (0.0022)	-0.0233*** (0.0061)	-0.0658*** (0.009)	-0.243* (0.131)
Age cubed		0.000086*** (0.000026)	0.00021*** (0.000070)		0.00204 (0.0015)
Constant	-1.75*** (0.14)	-3.72*** (0.613)	-6.790*** (1.67)	-44.38*** (7.42)	-91.28** (35.2)
Obs.	6017	6017	6017	6017	6017
R-squared	0.75	0.75	0.73	0.33	0.33
Peak age	28.1 (0.211)	27.6 (0.257)	26.6 (0.308)	26.2 (0.486)	25.9 (0.505)
90% conf. int. for peak age	27.8–28.5	27.1–28.0	26.1–27.1	25.4–27.0	25.1– 6.7

***, **, and * indicate the coefficient is significantly different from zero at the 1%, 5% or 10% levels of statistical significance, respectively. Standard errors are in parentheses. The standard errors for the optimal age are obtained using the bootstrap option in STATA 11. Preferred estimates for the peak age are shown in bold.

Table 5 Age-performance regressions for NHL defencemen.

	(1)	(2)	(3)	(4)	(5)
Variables	pts. per game	pts. per game	pts. per 60 min	plus-minus	plus-minus
Age	0.078*** (0.0091)	0.0894 (0.058)	0.104*** (0.218)	1.866** (0.737)	7.517 (5.065)
Age squared	−0.00134*** (0.00015)	−0.00176 (0.002)	−0.0018*** (0.00037)	−0.0336*** (0.126)	−0.231 (0.176)
Age cubed		3.74e−06 (0.00002)			0.00225 (0.0020)
Constant	−0.818*** (0.135)	−0.93 (0.55)	−0.62* (0.318)	−24.75** (10.7)	−77.74 (47.9)
Obs.	3179	3179	3179	3179	3179
R-squared	0.72	0.72	0.58	0.283	0.284
Peak Age	28.9 (0.491)	28.9 (0.59)	28.7 (0.68)	27.8 (1.80)	26.8 (1.14)
90% conf. int.	28.1–29.8	27.9–29.8	27.5–29.8	24.8–30.7	24.9–28.6

***, **, and * indicate the coefficient is significantly different from zero at the 1%, 5% or 10% levels of statistical significance, respectively. Standard errors are in parentheses. The standard errors for the optimal age are obtained using the bootstrap option in STATA 11. Preferred estimates for the peak age are shown in bold.

preferred (quadratic) specification, defencemen reach 90% of peak performance about age 24 (as do forwards) but stay within 90% longer – up to about age 34 (compared with 32 for forwards).

We do not report age-performance regressions for goaltenders, as there is no meaningful effect of age on performance except at the extremes of the age range. There is a lot of individual variation across goaltenders. Some do well early and fade in their 30s. Many others do better in their 30s than in their 20s. However, on average, if we compare the early 20s, late 20s, early 30s and late 30s, there is very little systematic difference in performance.

5 The elite performance method, the naïve method and selection bias

5.1 The elite performance method

This method focuses on the best performances at each level. Before providing formal analysis, we show the top

10 performances in the data for scoring, plus-minus, and save percentage.

Tables 6 and 7 illustrate the very best performances in the data. They demonstrate that elite performances can occur over a wide age range. Among the top scoring performances for forwards are Sidney Crosby at age 19 and Mario Lemieux at age 35. For defencemen the top 10 performances include Erik Karlsson at 22 and Nicklas Lidstrom at 38. For plus-minus, which includes both forwards and defencemen, we have Chris Pronger (a defenceman) at age 23 and Chris Chelios (also a defenceman) at age 38. Still, for scoring by forwards and for plus-minus, there is a clustering in the mid-to-late 20s for these elite performances. Scoring performance for the best defencemen often peaks late, however, with six of the top 10 performances coming after age 30. For goalies there is no discernible clustering. Half of these top 10 performances occur over the age of 30 – and well into the 30s, while goalies of 24 and 26 are also in the top 10.

A more formal approach to using the elite performance method is to take the top 10 performances for each integer age level between 19 and 40 and then take the average of

Table 6 Top 10 scoring performances (PPG) for forwards and defencemen.

Rank	Forward	Age	Pts. per Game	Rank	Defenceman	Age	Pts. Per Game
1	Mario Lemieux	35	1.77	1	Mike Green	23	1.07
2	Sidney Crosby	24	1.68	2	Mike Green	24	1.01
3	Sidney Crosby	23	1.61	3	Nicklas Lidstrom	36	1.00
4	Jaromir Jagr	27	1.57	4	Brian Leetch	33	0.96
5	Joe Thornton	27	1.54	5	Erik Karlsson	22	0.96
6	Jaromir Jagr	28	1.52	6	Bryan McCabe	31	0.93
7	Sidney Crosby	19	1.52	7	Chris Pronger	26	0.92
8	Alex Ovechkin	24	1.51	8	Nicklas Lidstrom	38	0.92
9	Jaromir Jagr	34	1.50	9	Al MacInnis	37	0.92
10	Jaromir Jagr	29	1.49	10	Sergei Zubov	35	0.91

Table 7 Top 10 plus-minus and goaltending performances.

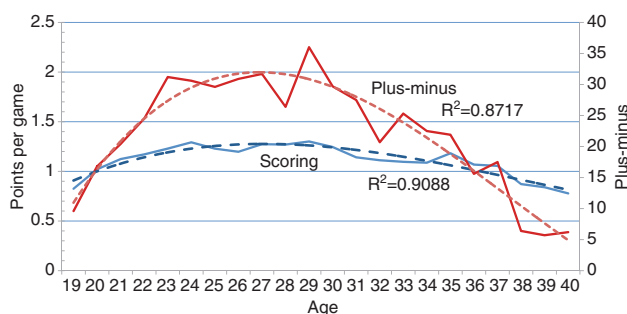
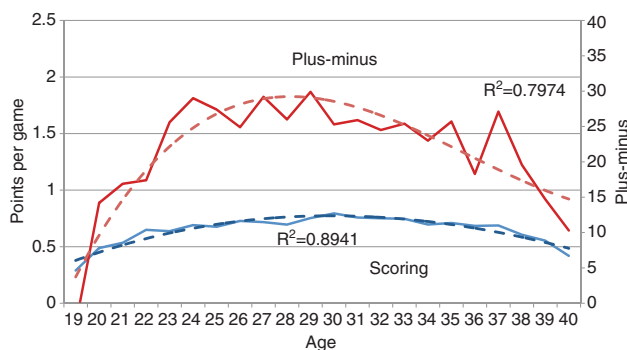
Rank	Player	Age	Plus-Minus	Rank	Goaltender	Age	Save %
1	Peter Forsberg	29	52	1	Brian Elliott	27	0.940
2	Chris Pronger	25	52	2	Tim Thomas	37	0.938
3	Milan Hejduk	27	52	3	Dominik Hasek	34	0.937
4	Jeff Schultz	24	50	4	Cory Schneider	26	0.937
5	Chris Chelios	38	48	5	Dwayne Roloson	34	0.933
6	Chris Pronger	23	47	6	Miikka Kiprusoff	27	0.933
7	Alex Ovechkin	24	45	7	Tim Thomas	35	0.933
8	Joe Sakic	31	45	8	Marty Turco	27	0.932
9	Patrik Elias	25	45	9	Dominik Hasek	33	0.932
10	Thomas Vanek	23	47	10	David Aebischer	24	0.931

each of those age group's top 10 performances. Figure 4 provides the results in a diagram for forwards for both scoring (points per game) and plus-minus. Figure 5 shows the results for defencemen. Trend-lines fitted to these averages are also shown. Cubic trend-lines are used as they fit the data slightly better than quadratic trend-lines.

The peak performance for elite forwards occurs over the 27–29 age range for scoring and over the 23–30 range for plus-minus. For elite defencemen the peak scoring age is the 29–33 range and plus-minus shows little systematic change over the range 23–33. These elite performance results indicate a slightly later range of peak performance

than fixed-effects regression methods using the full sample. Also, elite performances are slightly more skewed – with a longer period of near peak performance. We interpret the results to mean that elite players improve faster initially, continue to improve for slightly longer and experience slower age-related decline. They do not experience a major drop-off in performance until their late 30s.

We do not provide a diagram for goaltenders, in large part because there is almost no variation to see. In Table 8 we show the average save percentage for the top 10 goalies of each age. Only ages 22 through 37 are covered as the other age categories have fewer than 10 goaltenders each. Elite goaltender performance from age 23 to age 35 is remarkably consistent and exhibits little variation.

**Figure 4** Elite (Top 10) scoring and plus-minus performance average for forwards by age.**Figure 5** Elite (Top 10) scoring and plus-minus performance average for defencemen by age.

5.2 The Naïve Method and Selection Bias

The naïve method simply calculates average performance, by position, for each age category in the data. For

Table 8 Elite (top 10) save percentages for goaltenders by age.

Age	Save %
22	0.913
23	0.920
24	0.920
25	0.925
26	0.925
27	0.925
28	0.924
29	0.923
30	0.923
31	0.922
32	0.922
33	0.921
34	0.923
35	0.922
36	0.917
37	0.916

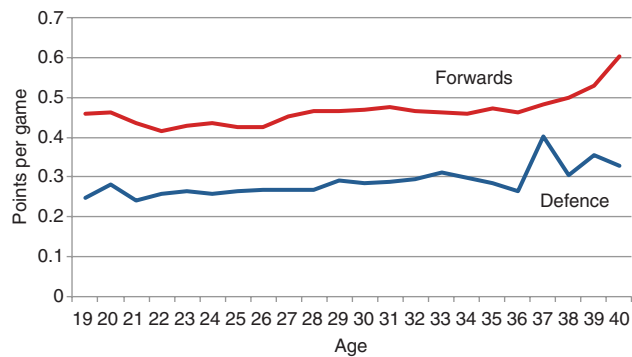


Figure 6 NHL scoring averages by age – naïve model.

example, there are 98 forwards of age 19 in the data. The naïve estimate of the relevant performance for a 19-year-old is the average performance (points per game, points per minute, or plus-minus) taken over these 98 observations. We determine the average performance for each age category in the same way. In the case of 26 year olds, there are 1134 forwards in the data, so the average is taken over a much larger set. The average performance by age is then the estimated age-performance relationship. Figure 6 shows the result of using the naïve model to assess the age performance relationship. We show only points per game to save space. The pattern for plus-minus is very similar.

Figure 6 shows no discernible single-peaked pattern relating performance to age using the naïve method. An extension of the naïve method would be to fit an estimated regression line to these age-average performance combinations. However, Figure 6 makes it clear that selection bias overwhelms the underlying age performance relationship using this approach. As noted in the introduction, this method suffers from selection bias. Only the very best 19 or 20 year olds and 39 or 40 year olds are compared with a full range of players at ages 25 and 26, overstating the relative performance of the youngest and oldest age categories and thereby understating the gains due to age in the early 20s and the decline due to age once players pass their early 30s.

If anything, the naïve method implies that scoring performance increases gradually with age up to the maximum age of 40. The highest scoring average is in fact achieved by 40 year-old forwards. Over the entire 14 year period covered by the data there are only nineteen 40 year-old players in the sample, including players such as Mario Lemieux, Jaromir Jagr, Mark Messier, Teemu Selanne, Mike Modano, and other players of that level – the greatest scoring stars in the game over this period. At age 40 only the very best scorers remain in the game, creating a very substantial selection bias.

This selection operates at two levels. Coaches and managers will select only the best players of that age to play and, in addition, the players themselves will exercise self-selection. Many players who could play effectively at age 40 choose not to – possibly because of fear of injury, possibly because they do not want to go the effort of staying in shape (which gets increasingly difficult with age) and possibly because they want to spend time with their families or pursue other objectives. At age 40 playing is worthwhile only for the most gifted athletes – those who can command high enough salaries to make playing worthwhile and who are still good enough to make the game enjoyable to play. This self-selection effect likely explains why older players actually score more than younger players (rather than just being about the same). The fixed-effects regression methods we use in section 4 correct for both kinds of selection bias.

Comparing the naïve model with the results from the fixed-effects regression methods, or with the participation information, suggests that any analysis of the age-performance relationship that fails to correct for selection bias is essentially meaningless.

6 Player-specific aging functions

As indicated by equation (1) we have assumed a common underlying aging function, $f(x)$, across players as the basis for our fixed-effects regressions. In this specification, differences between players are assumed to be captured by player-specific fixed effects (some players are better than others) and by random errors. However, as emphasized by Albert (1999, 2009) for baseball, individual aging trajectories vary across players – some players peak earlier than others.

We interpret our earlier results only as providing results for a representative or typical player. We would say, for example, that forwards typically peak in scoring performance between 27 and 28, but recognize that some forwards will peak sooner than others. However, it is still important to consider player-specific aging functions for two reasons. First, it is valuable as a robustness check to assess whether allowing for individual aging functions would lead to results that are inconsistent with or at least different from our fixed-effect regression results. Second, it is of interest to ask how much variability in the age-performance relationship is present across players.

In this section we use a method suggested by a reviewer that allows for variation across players in the age-performance relationship. The basic method is to

estimate an aging function for each player [using points per game (PPG) as the performance measure]. Thus, for each player i we estimate

$$\text{PPG}_{it} = f_i(x_{it}) + e_{it} \quad (3)$$

Thus player i has his individual f_i aging function, and there is no fixed effect in the regression. To save space we report results only for a quadratic functional form for $f(x)$, which can be written as $f_i(x_{it}) = \beta_{0i} + \beta_{1i}x_{it} + \beta_{2i}x_{it}^2 + e_{it}$.⁸ Thus each player has his own coefficients and hence his own aging trajectory. From the estimated quadratic function for player i we then obtain his estimated peak age. From the distribution of peak ages we can determine the mean to compare with the peak age estimated using fixed-effects regressions and other methods. We can also assess the extent of variation across players in the age of peak performance.

This method is a simplified version of the method used in Albert (1999), Albert (2009), and Lavieri et al. (2012). In those papers, individual aging functions are jointly estimated for all subjects in the data under the assumption that the regression parameters of the individual aging functions are drawn from a particular common prior distribution. Imposing this prior distribution has the effect of smoothing individual trajectory estimates toward the average trajectory while incurring the cost of imposing more external structure on the estimation process. The simpler version we use is sufficient to provide a good indication of the extent of individual variation and the effect of allowing for individual variation on estimated peak ages.

The primary difficulty with this method is that for many players we have only a few observations – not enough to estimate a quadratic function with much reliability. For example, there are some players who first played in 2010–2011 and then played again in 2011–2012 – the final year of our sample period. As we have only two observations for this player we cannot meaningfully estimate an individual quadratic age-performance relationship for him. STATA will provide an estimate – defaulting to a straight line relationship that “fits” the two points perfectly. If the player improves over this period the implied optimal age is then infinite.

⁸ As Albert (1999) points out, quadratic age-performance provide a parsimonious and effective way of capturing individual trajectories sufficiently to estimate peak age functions even though they have the disadvantage of imposing symmetry – the assumption that players are assumed to improve at the same rate as they decline. See Albert (2009) for an application using the more flexible spline functions to capture individual trajectories.

We clearly need some minimum number of years for each player that will allow meaningful estimation of a quadratic function. Requiring more years provides more reliable individual estimates but reduces the number of eligible players. We find that 5 years balances this trade-off effectively and there is little change in the results if we use 4 years or 6.

With a minimum time requirement of 5 years, player trajectories for most players can be reasonably estimated. However, there are some players whose trajectories cause a problem. There are several ways to handle such trajectories and they are few enough in number that our results are not significantly affected regardless of the method we use. In addition, as a check, we use the approach suggested in Lavieri et al. (2012) of comparing estimated peak ages with the actual ages at which players achieve their peak performance. Most are within 2 years in one direction or the other.

There are three general classes of problem cases. One such class arises when the estimated quadratic function is convex rather than concave, implying a peak age that approaches infinity. A closely related problem arises if the estimated trajectory is concave and increasing but is close enough to being linear that a player's implied peak age is absurdly large – perhaps in his 50s or 60s. The third possible problem is that the estimated trajectory may be concave but decreasing, implying an absurdly young peak age. For all three of these types of problem we use the player's actual age of peak performance instead of the estimated age of peak performance. (Alternatively, we can just drop all these cases, which yields very similar results.)

Figure 7 shows two typical player trajectories for points per game. The dots show the actual performances and the solid lines show the estimated quadratic performance trajectories. Brendan Morrison (red dots) was a very good player who broke into the league at age 23,

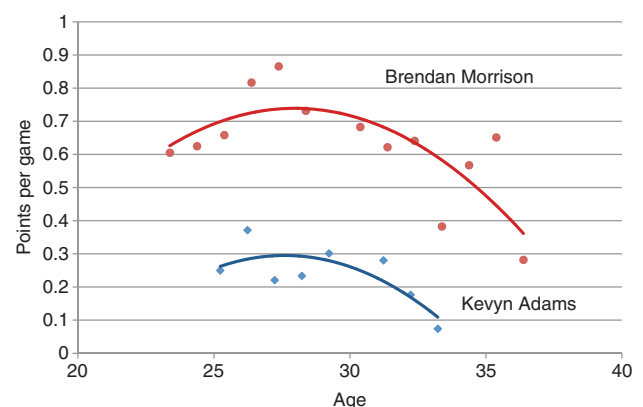


Figure 7 Typical performance trajectories.

improved over the next few years, had his best year at age 27 (when he was 27.4 as of January 1) and then declined. His estimated peak age is 27.97. Kevyn Adams (blue) was a good player but not as successful as Morrison, entering the league later, retiring earlier, and scoring less. His estimated peak age is, however, very similar at 27.60.

Figure 8 illustrates two types of trajectory that cause a problem in estimating a peak age. Eric Dazé (blue) was a young player at the start of our sample period (22.5 as of Jan. 1, 1998) who improved steadily over the next few years. However, he had continuing back problems, with three back surgeries in 5 years. He played a full year at age 27.5 but his worsening back problems allowed him to play only a small number of games the following season – not enough to be in our data set for that season – and he retired shortly after. His trajectory is only slightly concave, leading to an implausible implied peak age of 54.36. We therefore use the age of observed peak performance (26.50) instead of the predicted peak age.

The other player is Marek Svatos (red), regarded as a future star during his rookie season at age 23.5 (as of January 1). His second season was disappointing, partly attributed to nagging groin injuries, but his performance continued to be disappointing and he was dropped from being a “top 6” forward (top two lines) to being in the “bottom 6” (3rd and 4th lines) with correspondingly reduced ice time. He retired from the league after his 5th season. His estimated trajectory is slightly convex so there is no interior peak age. There are also a very few players who, like Svatos, have a downward sloping trajectory, but whose estimated trajectories are slightly concave instead of slightly convex, which implies a finite but absurdly young peak age. For both these types of trajectory – convex estimated aging functions or absurdly young peak ages – we use the age of actual peak performance in the peak age distribution analysis.

A reader might ask why we do not use the full distribution of actual peak ages instead of the estimated peak

ages. In fact, the distribution of actual peak ages is not very different from the distribution of estimated peak ages. However, using estimated ages is better for two reasons. First, the actual peak age uses just one data point for each player. To the extent that one particularly good year (a “career year”) is partly the result of good luck (a positive error), this apparent peak age can be misleading. The *estimated* peak age is based on a player’s entire career trajectory, which is at least 5 years and is often 10 years or more. Thus the estimated peak ages are based on much more information and are less prone to random errors.

This advantage of using estimated peak age is illustrated by Kevyn Adams in Figure 7. It is unlikely that Adams’ true ability rose sharply from age 25 to 26 then fell sharply at age 27, then rose again at ages 28 and 29. It is much more likely that his underlying ability followed a path close to the smooth estimated trajectory and that deviations from this trajectory were the results of random events – a few lucky or unlucky bounces, playing with better or worse teammates, changes in how the coach used him, etc.

Second, using estimated ages allows the use of more players. One problem with actual ages is that players in the data might not yet have reached their peak. For example, a player who enters the league at age 20 and plays 5 years – through age 24 – up to the end of our sample period (2011–2012) will typically not yet have reached his peak. In our data his apparent peak age will necessarily be 24 or less. Similarly, someone who was past his peak in the first year of our sample period (1997–1998) will have an apparent peak age that is older than his true peak age. To guard against this we would have to drop a large number of players. Using estimated trajectories has the advantage of allowing us to “predict” a peak age that has not yet been achieved or that was achieved before the sample period began.

Table 9 summarizes the main characteristics of the distribution of peak scoring ages for forwards and defencemen and of peak save percentage performance for goalies.

Table 9 shows that the average ages of peak performance obtained when allowing for individual aging functions are very similar to those obtained using fixed-effect

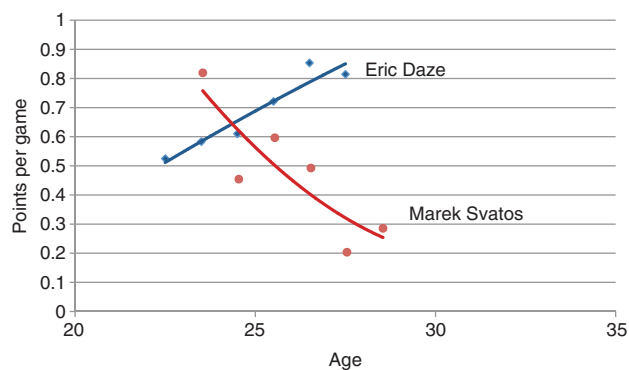


Figure 8 Atypical performance trajectories.

Table 9 Age of peak performance based on individual aging functions bootstrapped standard errors for the means are shown in parentheses.

	Number	25th pc	Median	Mean	75th pc	St. Dev.
Forwards	566	25.2	27.6	27.7 (0.143)	29.9	3.32
Defence	309	25.6	28.0	28.2 (0.254)	31.0	3.88
Goalies	71	26.0	28.6	29.0 (0.444)	31.7	3.89

regression methods. The results reinforce the finding that forwards peak in points per game between 27 and 28. For defencemen this individual player method provides an estimated mean peak age of 28.2 rather than the estimate of 28.9 obtained with our fixed-effects quadratic regression.

This approach indicates a difference in the trajectories of forwards and defencemen that is broadly consistent with our earlier analysis. There are small differences between these positional categories at the 25th percentile and at the median, and the differences open up more at the 75th percentile, capturing the fact that quite a few defencemen peak in their 30s, while that pattern is much less common for forwards. In Figure 9 we show the distributions of estimated peak ages for forwards and defencemen.

Estimated individual player trajectories also allow us to revisit the question of whether better players peak later. We consider each player's peak scoring performance,

estimated peak age, and position. The estimated regression is $\text{PeakPPG} = 0.251 + 0.0062(\text{Peak Age}) + 0.252(\text{Forward})$, where Forward is a dummy variable that takes on value 1 for forwards. The t -statistic for Peak Age is 2.45, indicating statistical significance at better than the 0.05 level. Thus peak age has a small but positive and statistically significant effect on maximum performance. (We obtain essentially the same result using a player's average scoring performance over his career as the dependent variable instead.)

This finding is also consistent with the results from the participation method (Section 3). The participation method identifies the peak age using the performance of marginal players. If the weakest NHL players (i.e., the marginal players) peak earliest, this is a partial explanation for the younger peak ages suggested by the participation method than by the regression methods.

There are many additional things that could be done to investigate individual variation.⁹ However, our analysis is sufficient to show that allowing for individual aging functions yields similar estimated peak ages to our fixed-effect regression methods. To the extent that the methods do differ, the fixed-effect regressions have the advantage of being able to use more data. The alternative of estimating individual player functions, as done in this section, requires enough observations for each player to estimate such a function – and we settled on 5 years as the best number. This 5 year requirement forces us to drop just over half the players in the data and we go from analysis using 2033 players down to analysis using just 946. In principle, a player who plays for only 2 or 3 years provides relevant information regarding how much a player improves from 1 year to the next. The fixed-effects methods allow us to include all such information in the analysis.

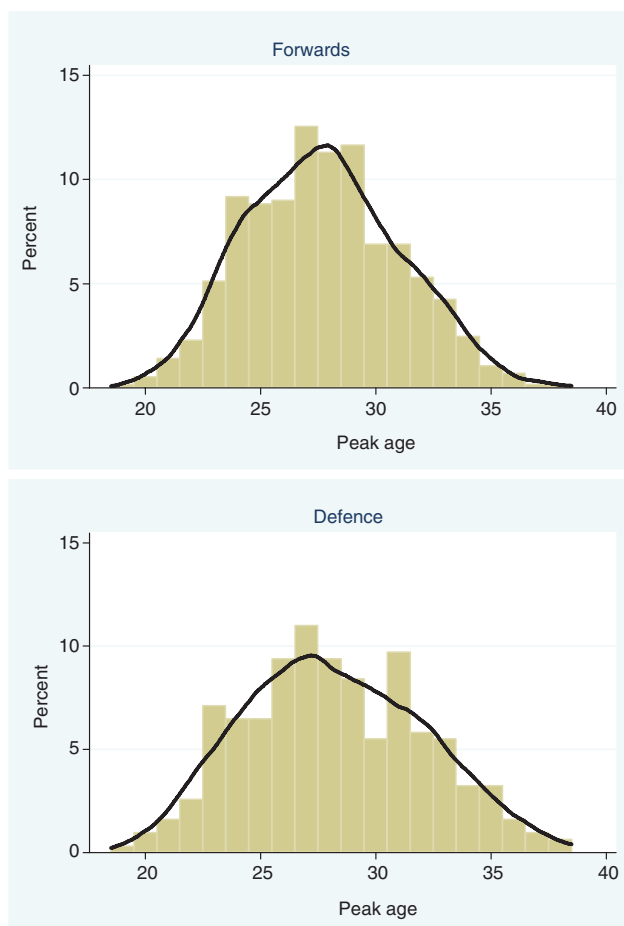


Figure 9 Histograms of estimated scoring peak ages for forwards and defencemen. (The dark line is the estimated kernel density for the peak age distribution.).

7 Explaining patterns of performance

In general, hockey players in all three positional categories reach something close to peak performance by about age 23 or 24 and reach their actual peak in the late 20s. A typical forward begins a significant decline in his early 30s. Defencemen maintain near-peak performance somewhat longer – until their mid-30s, and goaltenders do not show a clear decline until their late 30s.

⁹ Among other things we confirmed that the individual aging functions estimated in this section exhibit a similar curvature to the common or representative aging function estimated in Section 5.

These patterns are consistent with earlier work on the relationship between physiological and mental factors and the development of basic sports skills. Schulz et al. (1994) emphasize four categories of underlying factors: physical development, experience, motivation, and “wear and tear.” The science underlying physical development has been much studied. As noted in Schulz and Curnow (1988), skills related to reaction time and to speed and explosive power of muscle movement peak in the early to mid-20s. Such skills are very important in hockey. In particular, speed and explosive power of muscle movement largely determine skating speed and the ability to shoot the puck with high velocity. However, endurance and skill at complex physical tasks peak later – in the late 20s or early 30s. Such tasks are also very important in hockey. In particular the ability to play at high intensity for extended periods (endurance) is very important, as is the ability to be a “playmaker” – requiring a high level of physical coordination and an understanding of patterns as they develop in the game.

Experience and wear and tear and also very important in hockey. As players become more experienced their positional play improves and their ability to anticipate improves, implying improvement with age. But wear and tear works in the opposite direction. By the time a typical player reaches his early 30s he has accumulated a number of wear and tear injuries that slow skating speed or reduce other aspects of performance. As for motivation, we are not aware of any evidence to suggest that motivation varies systematically with age over the relevant period – up to the late 30s.

We expect an increase in physiological capacity and experience to the mid-20s, after which some physical capabilities (such as speed and quickness) start to decline while others (such as stamina) continue to improve, and experience continues to accumulate. This pattern allows for a peak in the late 20s and gradual decline afterwards. The effects of injury, wear, and tear would be relatively minor (for most players) throughout the 20s, and accumulate in the 30s, contributing to a decline in overall performance during the 30s.

The differences among the positions can also be readily understood. Probably the most important single skill for a forward is skating ability, requiring explosive speed, the ability to make quick powerful turns, and the ability to stay balanced and “strong” on his skates. Skating ability is much less important for defencemen, who rely more on pattern recognition and anticipation – which improve with experience – to play a good positional game and to make good passes. It is widely understood that playing forward (especially wing) is “simpler” than

playing defence. For this reason defencemen can play forward if necessary and very often do, but the opposite transition is almost never observed. In light of these differences between the positions it is not surprising that forwards peak slightly earlier and that defencemen maintain near peak performance later.

Goaltending is somewhat different from the other positions with respect to the required skill mix. Goaltending does not depend on skating speed or on strength or on the ability to shoot or pass a puck. And goalies are less likely to suffer from wear and tear injuries. The two main skills are reaction time, which peaks early, and good anticipation and positional play, which peak late. The net effect of these offsetting factors can apparently produce peak performance any time between the early 20s and late 30s with no particular concentration around any one age.

8 Concluding remarks

The primary objective of this paper is to assess the relationship between age and performance among NHL players. We have placed considerable emphasis on solving the classic selection bias problem and we believe that we have obtained a more reliable estimate of the age performance relationship than has previously been obtained for NHL players.

Several of our results are consistent with what might be described as “conventional wisdom” and some are not. Results that fall in the “no surprise” category would include the estimated age of peak performance for both forwards and defencemen. Specifically, we believe that our best estimates are those obtained from the regression models, and we believe that the best scoring metric is points per game played. For forwards, a cubic specification is the best fit to the data when using a parametric fixed-effects approach and implies a peak age of 27.6. The individual player model implies a nearly identical average peak age of 27.7. For points per minute, on the other hand, the peak age is about a year earlier, reflecting the fact that younger players have less endurance and play fewer minutes per game. The optimal age for plus-minus for forwards is obtained from a quadratic specification and is 26.2 – earlier than for scoring.

For defencemen the best estimate using fixed-effect regressions for the peak age for both scoring and plus-minus is obtained from a quadratic specification and is 28.9 for scoring and 27.8 for plus-minus. The individual player model indicates an average scoring peak age of 28.2

These parametric estimates are consistent with the estimates obtained from fixed-effects regressions with categorical age variables. They are also consistent with participation data if we make an allowance for the small amount of selection bias in participation data due to injury and voluntary retirement and for the finding that better players tend to peak later.

Therefore, we feel that the analysis strongly suggests peak scoring performance between ages 27 and 28 for forwards and between ages 28 and 29 for defencemen. This would not come as a surprise to people familiar with NHL hockey – fans, players, coaches, journalists, or researchers. The numbers are perhaps on the early side of the conventional wisdom, but not far off, and the finding that defencemen peak later than forwards would also not be surprising.

There are however some surprises in the data. One surprise is that goaltenders exhibit very little systematic variation in performance by age. Admittedly, sample size is an issue. After all, in a given season a typical team will have 14 or 15 forwards that meet the 20 game threshold for inclusion in the data, along with seven or eight defencemen, but normally only two goaltenders and sometimes only one. Overall there are 1246 forwards in the data, 635 defencemen, and only 152 goalies. Still, if there is a consistent pattern of performance relative to age, it should show up with 152 goalies observed over their NHL careers, at least using parametric regression methods. The fact is that goaltenders can perform well (or poorly) at any age within a broad range. This fact is particularly striking for elite goaltenders.

Another slight surprise is the finding that the peak age for plus-minus occurs before the peak age for scoring (as measured by points per game). The difference is not large – about a year and half for forwards and about a year for defencemen in our preferred specifications – but it is interesting. We believe the likeliest explanation is selection bias. Veteran players are more likely than young players to be assigned the task of playing against the other team's top players. Thus, other things equal (such as ability) they tend to have worse plus-minus numbers. Such an effect would reduce the apparent peak age for plus-minus. However we cannot rule out the possibility that plus-minus performance does in fact peak before scoring performance.

A striking feature of our analysis is the estimated length of the period of near-peak performance. Our sense is that the conventional wisdom exaggerates both the time taken for player development, particularly for forwards, and the importance of age-related decline. Using points per game we estimate that forwards reach

near-peak performance (defined as 90% of peak performance) fairly early – about age 24 – and maintain peak or near-peak performance until about 32. For defencemen our parametric regressions indicate a similar starting point for near peak performance that extends to about 34. We get a more complete picture of defencemen when we consider player-specific aging functions. Looking at Figure 9 we see that a substantial number of defencemen have a late peak, even though many others have an early peak.

The effect of age on year to year variation in performance is relatively modest. It is much more important to be a good player at age 24 or 34 than to be an average player at the prime age of 28 or 29. Much of the talk about “player development” and “upside” is misplaced. Thus the many fans, coaches, and journalists who talk about “expected improvement” for 25 or 26 year old forwards are probably guilty of wishful thinking. Forwards who do not develop into consistent scorers by age 23 or 24 in most cases never will.

A quadratic approximation for the aging functions works well for both forwards and defencemen at both the aggregate level and at the individual level, implying near symmetry in improvement and decline. However, for forwards the aggregate aging function is slightly better approximated by a cubic function, implying some skewness, with an improvement rate in the early years that is steeper than the rate of decline in later years.

An additional significant contribution of our analysis is the comparison of elite players – using the elite performance method – with the overall or average pattern of NHL performance obtained using fixed-effects regression or the individual player model. Elite players appear to peak later and have a longer peak period. This is not just a matter of elite players being better and therefore being able to play effectively for a longer period. They also have a slower rate of age-related decline relative to their individual peak performance.

Our analysis of individual aging functions is consistent with the later peak of elite players. We find a significant positive relationship between peak age and performance: better players peak later. The individual model analysis shows that there is significant variation across players regarding when they reach their peak. However, this fact does not seem to impart any bias to our earlier estimates of the representative or typical peak age: between 27 and 28 for forwards and between 28 and 29 for defencemen.

By comparing the results with and without corrections for selection bias we provide an indication of the importance of selection bias in assessing age-performance

relationships and find that such problems are likely to be very serious. However, once appropriate corrections for selection bias are made, the resulting estimates of the age-performance relationship appear to be reliable and robust.

Acknowledgments: We are grateful to Jim Albert (the editor), an associate editor, and three reviewers for very helpful comments. We also thank Tom Davidoff, Keith Head, Jim Jamieson, John Ries, and Joanna Zhu for valuable discussions regarding this paper.

References

- Addona, Vittorio, and Philip A. Yates. 2010. "A Closer Look at the Relative Age Effect in the National Hockey League." *Journal of Quantitative Analysis in Sports* 6(4), Article 9:1–17.
- Albert, Jim. 1999. "Comment." *Journal of the American Statistical Association* 94(447):677–680.
- Albert, Jim. 2002. "Smoothing Career Trajectories of Baseball Hitters." Unpublished manuscript, Bowling Green State University, at bayes.bgsu.edu/papers/career_trajectory.pdf.
- Albert, Jim. 2009. "Is Roger Clemens' WHIP Trajectory Unusual?" *Chance* 22:8–20.
- Arkes, Jeremy. 2010. "Revisiting the Hot Hand Theory with Free Throw Data in a Multivariate Framework." *Journal of Quantitative Analysis in Sports* 6(1):Article 2.
- Berthelot, Geoffroy, Stéphane Len, Philippe Hellard, Muriel Tafflet, Marion Guillaume, Jean-Claude Vollmer, Bruno Gager, Laurent Quinquis, Andy Marc and Jean-François Toussaint. 2013. "Exponential growth combined with exponential decline explains lifetime performance evolution in individual and human species." *Age* 34:1001–1009.
- Berry, Scott, C. Shane Reese, and Patrick D. Larkey. 1999. "Bridging Different Eras in Sports." *Journal of the American Statistical Association* 94(447):661–676.
- Bradbury, J. C. 2009. "Peak Athletic Performance and Ageing: Evidence from Baseball." *Journal of Sports Science* 27(6):599–610.
- Broadie, Mark, and Richard J. Rendleman Jr. 2013. "Are the official world golf rankings biased?" *Journal of Quantitative Analysis in Sports* 9(2):127–140.
- Chen, Mike. 2010. "When is an NHL player's prime age?" SB Nation Special Feature at fromtherink.sbnprivate.com/2010/7/16/1572579/when-is-an-nhl-players-prime-age.
- Fair, Ray C. 2008. "Estimated Age Effects in Baseball." *Journal of Quantitative Analysis in Sports* 4(1):Article 1:1–39.
- Gramacy, Robert B., Shane T. Jensen, and Matt Taddy. 2013. "Estimating player contribution in hockey with regularized logistic regression." *Journal of Quantitative Analysis in Sports* 9(1):97–111.
- Henderson, Daniel J., Raymond J. Carroll and L. Qi. 2008. "Nonparametric estimation and testing of fixed effects panel data models." *Journal of Econometrics* 144:257–275.
- James, Bill. 1982. *The Bill James Baseball Abstract*. New York: Ballantine Books.
- Kovalchik, Stephanie Ann, and Ray Stefani. 2013. "Longitudinal analyses of Olympic athletics and swimming events find no gender gap in performance improvement." *Journal of Quantitative Analysis in Sports* 9(1):15–24.
- Lavieri, M., M. Puterman, S. Tyldesley, and W. Morris. 2012. "When to Treat Prostate Cancer Patients Based on their PSA Dynamics." *IIE Transactions in Health Care Systems* 2:62–77.
- Lepers, Romuald, Beat Knechtle, and Paul J. Stapley. 2013. "Trends in Triathlon Performance: Effects of Sex and Age." *Sports Medicine*, <http://link.springer.com/article/10.1007/s40279-013-0067-4>, published online June 2013, forthcoming in print.
- Macdonald, Brian. 2011. "A Regression-Based Adjusted Plus-Minus Statistic for NHL Players." *Journal of Quantitative Analysis in Sports* 7(3):Article 4, 1–29.
- NHL.com. 2013: The official website of the National Hockey League, www.nhl.com.
- Schulz, Richard, and Christine Curnow. 1988. "Peak Performance and Age Among Superathletes: Track and Field, Swimming, Baseball, Tennis, and Golf." *Journal of Gerontology: Psychological Sciences* 43(5):113–120.
- Schulz, Richard, Donald Musa, James Staszewski, and Robert S. Siegler. 1994. "The Relationship Between Age and Major League Baseball Performance: Implications for Development." *Psychology and Aging* 9(2):274–286.
- Tiruneh, Gizachew. 2010. "Age and Winning Professional Golf Tournaments." *Journal of Quantitative Analysis in Sports* 6(1): Article 5.
- Vollman, Rob. 2013. Rob Vollman's Hockey Abstract, www.hockey-abstract.com.
- Wooldridge, Jeffrey M. 2008. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.
- Wright, Vonda J., and Brett C. Perricelli. 2008. "Age-Related Rates of Decline in Performance Among Elite Senior Athletes." *Journal of the American Journal of Sports Medicine* 36(3):443–450.
- Wuest, Matthew. 2013. CapGeek.com.