December 2020

# The Presence of the Relative Age Effect in Professional Sports

CeCe Jordan Hensley
*Clemson University*, cece.hensley@yahoo.com

# THE PRESENCE OF THE RELATIVE AGE EFFECT IN PROFESSIONAL SPORTS

---

A Thesis
Presented to
the Graduate School of
Clemson University

---

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
Economics

---

by
CeCe J. Hensley
December 2020

---

Accepted by:
Dr. Frederick Andrew Hanssen, Committee Chair
Dr. Raymond Sauer
Dr. Scott R. Templeton

## ABSTRACT

This paper adds to the literature on the relative age effect in athletic, academic, and professional pursuits. It estimates the effects of month of birth on professional athletes in the NHL, NBA, and MLB regarding their performance and "making it" to the professional level. I used data from each of the sports and compared their month of birth distributions to what we would expect from the general population and analyzed the relationship between player's month of birth and their performance statistics.

## ACKNOWLEDGMENTS

I want to express my deepest gratitude to Professor Andrew Hanssen for his consistent encouragement and guidance. Without his insightful advice and dedicated support, the goal of this project would never have been realized. Next, I would like to thank Professor Scott Templeton and Professor Raymond Sauer for serving on my thesis committee. The Clemson University Economics department has provided me with a plethora of knowledge, relationships, and opportunities that I will cherish forever. Finally, I am indebted to my family and friends for all of their support, encouragement, and unconditional love during this time.

## TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

Figure Page

## I. INTRODUCTION

There is an increasing amount of empirical evidence that suggests relatively older children have a physical and developmental advantage over younger children within the same age group. The consequences that result from these differences in age are described as the relative age effect. This paper investigates whether a relative-age effect impacts the selection of professional athletes and their level of performance at the professional level.

In summary, the relative age effect implies that the oldest children in an age group will be the most successful as a result of their physical and developmental edge. In youth sports, this translates to encouragement from coaches and parents that leads to an increase in self-confidence, opportunities to play against better opponents, and access to better coaching and training (Kiikka, 2017). The oldest children continue to benefit from this cycle as they age, which means that hypothetically; they are more likely to make it to the professional level.

Malcolm Gladwell popularized the relative age effect and its implications for professional hockey in his novel, *Outliers*. Before Gladwell, there was significant research surrounding the relative age effect in adolescent sports, all of which suggest that older children or children born closest to the eligibility cutoff date are at an advantage.[1] Gladwell discusses the theoretical relationship between month of birth and the success of Canadian hockey players. He concluded that the relative age effect continues to have implications on hockey players long after adolescence, and into their professional careers.

[1] See Musch & Hay (1999); Stracciolini (2016); Buchheit (2013); Baker & Logan (2007)

Since *Outliers* was released, there has been a substantial increase in research exploring and attempting to explain the relative age effect on athletes in the long term. This research spans across multiple professional leagues.[2] The results of these studies often vary between sports. The relative age effect has attracted many academics in fields outside of athletics. These researchers have found significant evidence that it plays a role in the psychological and intellectual development of adolescents, sometimes impacting them for the rest of their lives.[3]

This paper investigates three professional sports leagues; the National Hockey League (NHL), the National Basketball Association (NBA), and Major League Baseball (MLB) in an attempt to uncover evidence of a relationship between month of birth and having a professional career. I also examine player statistics from within each league to determine if the relative age effect impacts athletic performance once at the professional level. My study is one of the first to analyze the effects of the relative age effect in MLB and the NBA.

The investigation starts by examining the structure of each sport's youth leagues. In general, all three youth leagues have birthdate-related eligibility requirements, but their windows of eligibility and starting dates are all different. Understanding the structure of the youth leagues is important for a multitude of reasons. First, it is within the youth leagues that the relative age effect is "created". Players are separated based on age and skill. The level of structure and application of age requirements depends on the sport.

[2] See Wattie and Baker (2007); Fumarco (2016)
[3] See Thompson, Barnsley and Dyck (1999); Bassok (2013); Du, Gao, and Levi (2012)

Second, I speculate that the difference in starting date and eligibility cut off to determine the strength of the relative age effect, and how much of an impact it will have in the long term.

The paper's analysis begins with professional hockey. The methodology starts with a description of the sample of players. The sample consists of 1485 right-wingers who played professional hockey in the NHL at some point between the years of 1917 and 2019. I focus only on right-wing forwards to maintain consistency and clarity. When analyzing the sample, I found that the distribution of the players' months of birth is significantly and economically different than that of the general population. In this case, the athletes tend to be born within the first three months of the year. This is indicative of the presence of the relative age effect in the NHL.

After the analysis of the NHL, this study provides the same type of analysis on a sample of professional basketball players. The basketball sample consists of the top 200 point scorers in the NBA between the years of 1946 and 2019. [4] The results indicate that the basketball players are not typically born within a specific time, rather their birth months are distributed normally. I find that the birth months' distribution of the sample is not significantly or economically different than that of the general population. Among the three leagues, the basketball sample provided the weakest and most insignificant evidence of the presence of the relative age effect and a relationship between month of birth with the likelihood of making it as a professional player.[5]

[4] The statistics are only for regular-season games and do not include playoffs, or outside league participation.

[5] Although we do not know if only using the top players will impact our results, we do not expect it will. The results should not change, but there is always the possibility that they could.

The final birth month analysis comes from a sample of professional baseball players in the MLB, excluding pitchers. The sample consists of 15468 professional baseball players born between the years of 1832 and 1998. The sample includes each player's batting statistics, so pitchers are not included. In this case, the distribution of monthly birth rates mirrors the general population fairly closely, with the exception that slightly more professional baseball players are born in August. The results of the baseball players' analysis are less convincing of a relationship between birth month and making into the MLB, as compared to the NHL.

After the analysis of birth month on making it to the professional level, the paper investigates the relationship between month of birth and various game statistics for each sample to determine if there is evidence of the relative age effect on player performance. The analysis of the relationship between month of birth and level of professional performance begins with the hockey sample. The statistics used are total points, total games played, and total goals scored. These statistics are useful because they are consistently reported and provide insight about one of the main skills of right-wing forwards, scoring. The results of the tests provide little evidence that once the players have made it into the NHL, their relative age or month of birth continues to contribute to their level of performance. Next, the paper analyzes the basketball sample using points, field goals, and games played followed by the baseball sample using on-base percentage, slugging percentage, and games played. All three samples show weak evidence in favor of the relative age effect and even suggest a reversal. Players born in the borderline months seem to perform worse than in other months.

One of the reasons youth hockey leagues have been so well researched is because as compared to most youth sports leagues their age requirements are strict and the structure is consistent through all local leagues and age groups. I expect to find results similar to previous research on professional hockey from my sample of right-wingers in the NHL. To broaden the application of the relative age effect in sports, I use samples from baseball and basketball. This allows me to test the implications of different cut-off dates and weaker league structures. The results from these samples will contribute to the growing research on the relative age effect.

The results imply the presence of the relative age effect in hockey and by contrast a much weaker effect in basketball and baseball. The weaker results may reflect the fact that these sports do not have as much league by league structure, suggesting that cut-off dates may reinforce the relative age effect. Finally, the paper summarizes the findings and discusses their broader implications.

## II. Literature on the Relative age Effect

The relative age effect is essentially a domino effect that creates a performance gap among adolescents. It can be observed in both academic settings and athletic settings. The central implication is that the oldest children in their grades, or on their teams will be more successful because of their age. Why does the relative age effect exist? There is lots of speculation, and it is impossible to determine the exact reasons. In the context of sports, most researchers speculate it is “because of their physical and emotional maturity relative to their peers" (Stracciolini, et al., 2016).[6] The relative age effect and its consequences essentially follow a four-step loop. The initial difference in maturity causes them to stand out on the field, which leads to encouragement from coaches and parents. This encouragement breeds player confidence, and then they hypothetically get more playtime and bigger roles on the team. In the next season, the players with more skills are picked for better teams, who play better opponents and encourage growth in their skillset. The physical effects of the athletes get them specialized attention and treatment. This positive affirmation can result in more exponential skill growth versus the less physically endowed players on the team. Another possible consequense is that the skill growth of smaller or younger players is hindered without the same treatment and can lead them to stop playing the sport entirely. This cycle is just a hypothetical representation of why the relative age effect exists in adolescents.

"For decades, sports scientists have known that the month of birth may affect an athlete's likelihood of reaching the professional level because of relative age effects" (Baker & Logan, 2007). It has been well documented that the presence of the relative age effect has major implications in adolescent and minor league hockey.[7] These studies find that the older players in the leagues are more likely to progress faster and ultimately achieve athletic success because of the relative age effect. Within the last decade, there has been an increase in the study of the relative age effect in professional hockey. The universal finding is that hockey players born within the first half of the year are more

[6] The Relative Age Effect on Youth Sports Injuries
[7] See Baker (2001); Stracciolini (2016)

likely to be drafted into the NHL.[8] In general, there is a disproportionate percentage of hockey players born in the first quarter of the year. This is because the initial maturity of the players affords them more opportunities to train and succeed when they are minors. This translates to more of the older children being suitable for the NHL draft. There are a few studies that imply that after the draft, the relative age effect disappears, or even reverses.[9] This is an interesting phenomenon because it implies that once at the professional level; the younger players perform at a higher level than their older counterparts. This paper attempts to isolate the effects on right-wing forwards and expand the research regarding performance.

"To date, relative age research has reported significant and substantial achievement differences within the confines of athletic and academic pursuits" (Thompson, Barnsley, & Battle, 2004). Beyond the scope of athletics, the relative age effect has been found to have an impact on academic success, emotional and social development, and future career success. There is a trend in academics called "redshirting" where parents will wait to enroll their children born in the months closest to the cutoff date in kindergarten (Bassok, 2013). The goal is to counteract the consequences of being younger and less mature than their peers. These consequences include lower overall achievement and standardized test scores. The National Center for Education Statistics (NCES) has found evidence to suggest that redshirted children have higher reading ability than their peers (National Center For Education Statistics, 2009).

8 See Baker & Logan (2007); Wattie, Baker Cobley & Montelpare (2007); Kikka (2017)
9 Gibbs, Jarvis & Dunfur (2012); Baker (2001); Fumarco, Gibbs, Jarvis, & Rossi (2017)

There is also evidence suggesting that a child's relative age affects their emotional development. "It is suggested that the higher incidence of youth suicide in the group of relatively younger school children may have resulted from poorer school performance, which in turn led to lowered confidence and self-esteem" (Thompson, Barnsley, & Dyck, 1999). There are important personal and social consequences that relative age research can and should neutralize, particularly regarding the livelihood and safety of children.

Most of the research beyond adolescents have been conducted in the scope of professional sports. However, there is one study that looks at the relationship between the relative age effect and career success as a corporate CEO. Du et al. attribute the relative age effect "to school admissions grouping together children with age differences up to one year, with children born in June and July disadvantaged throughout life by being younger than their classmates born in other months" (Du, Gao, & Levi, 2012). The results suggest that the relative-age effect has a long-lasting influence on career success.

The implications of this paper's research can be applied to much more than professional sports. It will offer insight into the presence of the relative age effect in many aspects of human life.

## III. Structure of the youth leagues

There is very little similarity in the way that youth leagues are organized across sports. Not only do the seasons of play differ, but each generally has its own structure, specific age requirements, the involvement of an overarching league, and level of strictness. It is important to understand the differences between each sport and the structure of each youth leagues because these differences could be contributing factors in the relative strength or lack of a relationship between month of birth and professional success later on.

### A. Hockey

In hockey, each division is separated by the calendar year, meaning all children born within the same year, starting on January 1st will play in the same league. Hockey has been an ideal candidate for analyzing the relative age effect because the window of eligibility is fairly small (only one year), whereas other sports, like basketball, usually have a larger window of eligibility. Youth hockey leagues typically hold tryouts about three months before they start practicing, but the actual season starts in September and concludes in mid-March. The most commonly used cut-off date for youth hockey is January 1st because each league is divided based on year of birth. Each league is divided based on age and level of competitiveness starting at the initiation/mini-mite level for five and six-year-olds.[10] At the initiation level, the qualification is 6U, which means the maximum age a player can be during the season is six years old. Within each age group, there are levels of competition that players are divided into based on their level of skill (Levels of Minor (youth) hockey, 2015). This being said, a player born in January of 2001 is going to be older and more mature than a player born in December of the same year and is more apt to be placed in a league with a higher level of competition. The

[10] Mini-mite is the first level in the United States Youth Hockey Leagues, and Initiation is the first level in Canadian leagues.

structure of youth hockey leagues does not change much until players are around fifteen years old and can be placed in leagues with older players.

### B. Basketball

According to USA basketball guidelines, the “playing year” for purposes of eligibility is between September 1st and August 31st. The general cut-off date is on or before August 31st. In youth basketball, there is a transition year. Once players are fourteen years old or entering their eighth-grade year, the eligibility window switches to a grade-based approach. This means that each league is separated by grade level so that all of the players are playing against peers within the same grade. Before this transition, or until the players reach the age of thirteen players; leagues are segmented by calendar year. This approach is similar to the structure established in youth hockey. Before the transition year, all players within a league are born within twelve months of each other. The difference between youth basketball and youth hockey is that youth hockey never switches to a grade-based structure. Youth basketball dismantles the age-requirement structure, and the requirements become laxer. For example, “if a player in the 8th grade turns 16 prior to the beginning of the season, they are moved to the next division” (USAB, 2019). This means that if a player turns sixteen during the season, they are still eligible to play against a fourteen-year-old because they are in the same grade. If this were the case in youth hockey, the sixteen-year-old player would only be eligible to play against other sixteen-year-olds; regardless of the grade, they are in. The transition to grade-based age requirements means that the stricter calendar league structure is not maintained for as long as it is during youth hockey.

### C. Baseball

There are no definitive eligibility requirements, or cut-off date in youth baseball because each league is structured differently depending on the overhead corporation or overall league organization. Although there are some differences between each organization's age-requirements, most seem to abide by the same loosely similar structure. In general, youth baseball leagues are divided by age groups and do not consistently follow a calendar year or grade-based approach; unlike in basketball and hockey. To simplify the analysis, we will be using the cut-off date set following Little League Baseball guidelines, which states that "the Age Determination Date for a Little League Baseball player is the actual age of a child on August 31st of the current year." Since the leagues are divided by age groups, as long as a player is within the range of ages on August 31st of the year, they are eligible to play. For example, Tee-ball leagues encompass players between the ages of four and seven. As long as the player turns four or does not turn eight on or before August 31st of the season, they are eligible to play in the tee-ball league. Youth baseball leagues are never divided by year of birth alone as far as I have found, and generally, children with higher levels of skill will move to older or better leagues, regardless of their age. Youth baseball is structured more based on skill and less on age, even in the beginning. Youth league baseball has the laxest age restrictions and league structure of all three sports in question.

## IV. MAKING IT TO THE PROFESSIONAL LEVEL

### A. In the NHL

#### 1. Data Set Description

To investigate the possible presence of a relative- age effect among professional hockey players, I collected birth-date information for right-wing forward hockey players in the NHL between 1907 and 2019.[11] The NHL provides names, birthdate information, and all-time individual statistics for each player during the regular season.[12] The statistics are accumulated and updated over each player's entire career, so once they retire or leave the NHL the statistics are final. In total, I identified 1485 right-wing players to be used in the sample; each player is an observation. I chose to use the right-wing forward position because they are generally the primary point scorers on the team, even compared to the rest of the offensive players (Hockey Monkey, 2020). Points and goals are a straightforward measure of skill, so right-wing forwards were the ideal candidate. If I were to re-do this analysis, I would probably expand the sample to include centers, and left-wing forwards because they also contribute to point-scoring.

#### 2. Birth Month Analysis

Simple descriptive statistics and comparisons without statistical interference help to identify initial patterns in the data, which makes it easier to understand. Table 1 shows

[11] A right-wing forward, in the game of ice hockey, is a forward position of a player whose primary zone of play on the ice is along the outer playing area. They typically work by flanking the center forward. Their primary job is scoring goals and assisting other offensive players.

[12] Statistics for playoff games and any out of season participation are not included in the NHL's records. All of the players included in the sample have played in at least one game after they were drafted, so players who have never appeared in the game and have no relevant statistics are not included.

hockey birth rates by month, by count and by proportion. As can be seen, nearly eleven percent of right-wings in the sample were born in January and ten percent were born each month from February through April. By contrast, less than seven percent were born in November and December and less than eight percent were born in any month after June. We see an unusually high number of hockey players born within the first three months of the year.

To see how the sample compares to the general population, I compare the monthly levels of both samples. The population distribution used for the comparison is shown in Table 1 of the appendix.[13] Figure 1 illustrates the monthly mean comparisons between the general population and the sample of hockey players as a bar graph. This figure shows each sample's monthly distribution side by side and makes it easy to see the stark differences between the two. Again, it is easy to see that the distribution of the hockey samples' birth months is much higher between the months of January through April, as compared to the rest of the year, and contrary to the general population.

We see that the differences are quite large. The first four months, in particular, are much higher than the average seven and a half to eight and a half percent we would expect. These initial findings are consistent with prior research suggesting that more hockey players are born in the first six months of the year.[14] The difference in birth month among right-wingers is not simply a reflection of the fact that more children

[13] We collect the information on monthly births of the US population during 2009–2015 from the annual Vital Statistics of the United States, Natality Series, Volume I, published by the Centers for Disease Control and Prevention (http://www.cdc.gov/nchs/products/vsus.htm).
[14] See Gladwell (2011); Deaner, Lowen & Cobley (2013); Wattie, Baker, Cobley &Montelpare (2007)

generally are born in the early months of the year – the average for every month of the year is between seven and a half and eight and a half percent of the sample.

The monthly differences are large, but are they statistically significant? To find out, I conduct chi-squared "goodness of fit" test and a t-test for the mean monthly comparisons. I begin with the chi-squared test to analyze the overall distribution. The chi-square statistic is a measure of how far the "observed counts" (actual number of hockey players born each month) are from the "expected counts" (how many would be born each month based on the distribution of the actual population). The goal of a chi-squared test is to determine if the distributions of the two populations' birth months are the same.[15]

For the chi-squared test, we use the actual proportions of the US population based on birthrate data compiled by the CDC between the years of 2009 and 2015 to determine the expected frequencies. The assumption for this test is that the general population is normally distributed.[16] As shown in Table 2, the chi-squared statistic is 63.87, which is large because of the discrepancies between the observed frequencies (actual number of NHL right-wings born each month) and expected frequencies (how many we expect to be born based on the in the general population distribution). Our statistic is greater than the critical value of 19.68, which leads us to reject the null hypothesis that the distribution of the birth months of NHL right-wing players is the same as the distribution of birth months in the general population.[17] This provides evidence that the frequencies of NHL

15 The two populations being the hockey players and the general population.
16 Regardless of the underlying distribution, the results are largely the same.
17 The results of the chi-squared test are significant even at the 1% significance level since the p-value is only around .0000017.

right-wing's birth months are significantly different and that the month of birth frequency is not drawn from the same underlying distribution as the general population. These results are also economically significant because we are using proportions from the actual population.

Next, I turn to my two-tailed t-test to determine if the mean monthly distributions of the NHL right-wing players could have happened by chance. While the chi-squared test was for the overall distribution, the t-test focuses on the month to month differences. Table 3 shows the results of this test. The t-test compares the mean monthly number of hockey players to the mean monthly number of babies born in the general population to see if the mean number of hockey players born each month could have happened by chance, and how significant the differences in the means are. The two-tailed test regresses the dummy variables for each month on the null of the general population's monthly distribution. In all cases except for in May and June the null is rejected at least at the 5% level, some are even more significant.[18] When we reject the null, it means that there is statistical evidence to suggest that the number of players born in that month cannot be attributed to chance, and there is an outside variable involved. The p-values are listed to show the levels of significance. The smaller the p-value is the more evidence there is to suggest that the mean monthly number of players born is statistically different from the mean monthly number of babies born in the general population and is not due to chance.

[18] The differences in these months are too small to be significant.

In short, these statistical tests support what casual inspection of Figure 1 indicted: All of the findings support the prior research and current hypothesis that right-wing NHL players are more likely to be born within the first six months of the year.

### B. In the NBA

As of now, there has been minimal research done on the professional basketball frontier, but the National Collegiate Athletic Association or the NCAA has done initial research that alludes that there is no substantial effect on college basketball players, or college athletes in general (NCAA, 2012). According to “The Relative Age Effect in Under 18 basketball: Effects on performance according to playing position”, there are some effects on older forwards and centers. The smaller younger players were at a slight disadvantage, but the other positions were not significantly affected by age. I couldn’t find any other data that supports or refutes their research. Based solely on the presence of the relative age effect in other aspects of adolescence; we may observe minimal effects. Considering that most youth basketball leagues switch to grade-level division fairly early, thus decreasing the amount of impact a child’s relative age has, I do not expect to see any long-term effects. The only research on the relative age effect in youth basketball suggests that forwards and centers are likely to be impacted by the relative age effect.

#### 1. Data set Description

To investigate the possible presence of a relative- age effect among professional basketball players I collected birth-date information for the top performers in the NBA between 1946 and 2019. Based on the NBA’s database, I first identify the names of the players and then search for their birthdates. I identify the birth- date information of the top 200 point scorers in the NBA. The game statistics on the players are updated over the course of their entire careers until they retire or leave the NBA. If I were to re-do this analysis, I would include a larger randomly selected sample of players, to prevent bias. The reason I did not do this in the first place is because of time and resource constraints. Again, I chose to use points because they are a straightforward measure of skill and can be compared across sports relatively easily.

### 2. Birth Month Analysis

The analysis of the basketball sample follows the same set-up as the hockey sample. Table 4 shows basketball birth rates by month, by count and by proportion. As can be seen, ten and a half percent of players in the sample were born in February and ten percent were born in March. By contrast, less than seven percent were born in October and December. The rest of the months average between seven and nine percent. We do not see an unusual number of basketball players born within the months nearest to the perceived August cut-off, unlike in the hockey sample.

To see how the sample compares to the general population, I compare the monthly means of the two. We use the same general population distribution in Table 1 of the appendix for the comparison.[19] Figure 2 shows the comparison in the form of a bar graph. We can see that the differences are generally quite small, except for in the months of February, March, October, and December. It seems that the differences for the majority of the year are quite small since the bars look relatively close together, in

[19] We collect the information on monthly births of the US population during 2009–2015 from the annual Vital Statistics of the United States, Natality Series, Volume I, published by the Centers for Disease Control and Prevention (http://www.cdc.gov/nchs/products/vsus.htm).

contrast to the hockey sample. The figure shows that the birth month distribution among professional basketball players seems to reflect the fact that more children in the general population are born in the late months of the year. Since there has not been much research on this front, there is not much to compare to, but it is interesting to see that the distribution does not seem to reflect evidence of the relative age effect.[20]

Considering that the differences between the sample and the general population look quite small, I do not expect that they are statistically significant. To find out, I conduct the same type of chi-squared test and monthly mean t-test as I did for the hockey sample. For the chi-squared test, I use the same underlying distributions to determine the expected frequencies. The only difference between this and the initial test is that the observed frequencies (the actual number of NBA players born each month) come from the basketball sample.

For the chi-squared test, we again use the actual proportions of the US population based on birthrate data compiled by the CDC between the years of 2009 and 2015 to calculate the expected frequencies.[21] As shown in Table 5, the chi-squared statistic is 5.43 which is smaller than the critical value of 19.68. [22] This leads us to fail to reject the null that the distribution of birth months of the top NBA players is the same as the distribution of birth months in the general population. According to these results, the observed frequency (actual NBA players born each month) and expected frequencies (how many would be born based on the general population distribution) are not statistically different.

[20] I do not control for any unobservable factors like height, birth region, public/private school upbringing, family backgrounds, or college experience.
[21] We are still assuming that the general population is normally distributed.
[22] The p-value of this test is close to one, which also leads us to fail to reject the hypothesis.

Thus, we cannot conclude that the distribution for the NBA players' months of birth is not from the same underlying distribution as the general population.

It is easy to see that the monthly distribution of the basketball sample differs less from the general population than the hockey sample, so the results of the chi-squared test makes sense. Regardless of the assumed underlying distribution, the results are largely the same.[23] The chi-squared test does not provide evidence that the samples do not have the same distribution.

Next, I turn to my two-tailed t-test to determine if the mean monthly distributions of the NBA players could have happened by chance. Table 6 shows the results of this test. The t-test compares the mean monthly number of basketball players to the mean monthly number of babies born in the general population to see if the mean number of basketball players born each month could have happened by chance, and how significant the differences in the means are. The test is to see if the null distribution (of the general population) is representative of the sample. The two-tailed test regresses the dummy variables for each month on the null of the general population's monthly distribution. In all cases, we fail to reject the null at any reasonable level of significance.[24] All of the p-values in the test are quite large. When we fail to reject the null, it means there is no significant evidence to suggest that the number born in that month cannot be attributed to chance, meaning there is no sign that an outside variable is involved. The p-values are listed to show the levels of significance. The larger the p-value the less evidence to

[23] This holds true for all three samples.
[24] The differences in these months are too small to be significant.

suggest that the mean monthly number of players born is statistically different from the mean monthly number of babies born in the general population (against the null). In summary, any difference between the mean monthly number of NBA players and the mean monthly number of babies born in the general population can be attributed to chance.

Based on the common cut-off date for youth basketball leagues being August 31st, and accounting for the relative age effect; it would make sense that a majority of players would be born in September, October, and November. The results of both tests of statistical significance show the contrary. In short, the statistical tests support what casual inspection of Figure 2 indicted: the findings refute the current hypothesis that the top point scorers in the NBA are more likely to be born in the three months following the August cut-off date.

### C. In the MLB

#### 1. Data set Description

To investigate the possible presence of a relative- age effect among professional baseball hitters I collect birth-date information on all batters in the MLB between 1871 and 2019. Sean Lahman's Baseball Database provided each player's month of birth and seasonal performance statistics.[25] I average each player's seasonal statistics to determine their lifetime averages. Players were only excluded if they were pitchers, or if there was a lack of birth month information available. Just like the other two samples, the game

[25] The Lahman Database — a free relational database of individual and team statistics that covers the game back to 1871 which is found at seanlahman.com.

statistics on the players are updated over the course of their entire careers, until they leave the MLB. Again, I do not control for outside variables in these regressions.[26] The average of all of the seasonal statistics gives the lifetime statistics so that the results can be compared to the other samples which already used lifetime statistics. The statistics I chose are slugging percentage, on-base percentage, and total games played. All of the statistics I chose are commonly used to measure batter skill in professional baseball. They essentially provide information on whether or not the players are productive at-bat. I also included games played because it is one of the only statistics present in all of the samples, which allows for between sports comparison. If I were to continue researching, I would expand to pitcher and out-field performance.

### 2. Birth Month Analysis

The analysis of the baseball sample follows the same set-up as the hockey and basketball samples. Table 7 shows baseball birth rates by month, by count and by proportion. As can be seen, August and October have the highest number of players at around ten percent each. By contrast, the rest of the months are fairly evenly distributed between seven and nine percent. We do see that a slight majority of baseball players born in January, August, and October as compared to the rest of the distribution.

To see how the baseball sample compares to the general population, I compare the monthly means of both samples. We use the same general population distribution in

[26] I ran separate regressions to control for the formalization of the little league separating the sample into players born before and after 1980. The results were not statistically different from the ones when the sample was combined. This means that changes in the league structure do not have a significant impact on the relationship between month of birth and player performance in professional baseball players in the MLB.

Table 1 of the appendix for the comparison.[27] Figure 3 illustrates the comparison of the general population's monthly means and the sample of baseball players' monthly means as a bar graph. We can see that the differences are quite small, in most cases only a fraction of a percentage. June is the only exception with a difference of 1.01%. The differences for the majority of the year are even smaller than within the basketball sample, so the bars look relatively close together. This means that the distribution of professional baseball players' months of birth seems to reflect the fact that more babies in the general population are born within the second half of the year. There has not been much research within professional baseball specifically, so there is not much to compare to, but it is interesting to note that the distribution seems to reflect the general population and does not seem to provide evidence of the relative age effect.

Considering that the differences between the sample and the general population look quite small, I do not expect that they are statistically significant. To find out, I conduct the same type of chi-squared test and monthly mean t-test as I did for the hockey sample and the basketball sample. The only difference between this chi-squared test and the initial test is that the observed frequencies (the actual number of MLB players born each month) come from the baseball sample.

For the chi-squared test, we again use the actual proportions of the US population based on birthrate data compiled by the CDC between the years of 2009 and 2015 to calculate the expected frequencies.[28] As shown in Table 8, the chi-squared statistic is 6.13

[27] We collect the information on monthly births of the US population during 2009–2015 from the annual Vital Statistics of the United States, Natality Series, Volume I, published by the Centers for Disease Control and Prevention (http://www.cdc.gov/nchs/products/vsus.htm).
[28] We are still assuming that the general population is normally distributed.

which is smaller than the critical value of 19.68. [29] This leads us to fail to reject the null that the distribution of birth months of the MLB players is the same as the distribution in the general population. According to these results, the observed frequency (actual MLB players born each month) and expected frequencies (how many would be born each month if all months are distributed as they are in the general population) are not statistically different. Thus, we cannot conclude that the month of birth distributions of the two populations are different.

The monthly differences between the general population and baseball are much smaller than between the general population and hockey, so the results of the chi-squared test makes sense.

Next, I turn to my two-tailed t-test to determine if the mean monthly distributions of the MLB players could have happened by chance. Table 9 shows the results of this test. The t-test compares the mean monthly number of baseball players to the mean monthly number of babies born in the general population to see if the mean number of baseball players born each month could have happened by chance, and how significant the differences in the means are. The test is to see if the null distribution (of the general population) is representative of the sample. The two-tailed test regresses the dummy variables for each month on the null of the general population's monthly distribution. In January, May, June, July, August, October, and November we reject the null.[30] The p-values of the other six months are quite large, so we fail to reject the null at any

[29] The p-value of this test is close to one, which also leads us to fail to reject the hypothesis.
[30] At the 5% significance level, some even at the 1% level.

reasonable level of significance. The p-values are listed to show the levels of significance. When we reject the null, it means that there is statistical evidence to suggest that the number of MLB players born in that month cannot be attributed to chance, and there is an outside variable involved. When we fail to reject the null, it means the number born in that month can be attributed to chance, and there is no outside variable involved. We reject the null in both October and November, so there is evidence that the mean number of players born in each is statistically different than what we would expect based on the normal population's birth rates, and it is not due to chance.

Based on the common cut-off date for youth baseball leagues being August 31st, to account for the relative age effect; it would make sense that a majority of the MLB players would be born in September, October, and November. The results of both tests of statistical significance show the contrary.

In short, the tests of statistical significance support what casual inspection of Figure 3 indicated: that MLB players are more likely to be born within the second six months of the year.

## V. PERFORMANCE AT THE PROFESSIONAL LEVEL

### A. In the NHL

The relative age effect appears to affect making it into the actual NHL. Does it affect performance for players who have made it? I do not expect so -- When it comes to playing professional hockey, there is always going to be some slight variation in player skill levels, but in general, players should be on about the same level. If the relative age effect does have a long-term effect on their skill, then players born nearest to the cutoff should be notably better than those born in later months.

### 1. Data

The main measures of performance that I will be using to determine the presence of the relative age effect are overall goals scored, assists, and the total number of games played. I chose these measures because right-wing forwards are generally on the offensive, so arguably the most important skill to possess is the ability to score points and assist other players in scoring points for the team.[31] Additionally, I chose to incorporate the number of games played to see if there is any relationship between how often they are on the ice and because it can be used to compare across sports. While I cannot speak to the accuracy of the reports, points, goals, and games played were consistently available for each of the 1485 players in the sample without any scrubbing.

### 2. Empirical Evidence

The regressions are used to determine if, and how much month of birth affects right-winger skill level once they have been drafted into the NHL. Since scoring is arguably the main goal of an offensive player it is important to know if the relative age effect impacts their ability to score once they are playing professionally. Table 10 shows the results of these regressions. All three regressions use the same birth month data and assign each month a dummy variable, but the dependent variables (in this case the player

[31] Additional individual statistics like the types of goals, faceoff statistics, and game-winning percentages, were also tested but did not offer any new information. They are available in a separate appendix available to the reader.

statistics) are changed for each regression. All of the regressions seek to answer how being born in month “x” impacts performance statistic “y”.

The results in the first column of Table 10 are from the regression of the dummy variables for each month on total goals scored. The R-squared of the regression is .014, meaning only 1.4% of the variation in goals scored can be explained by the month of birth. It also implies that the data show no evidence of a significant trend. The coefficients for all of the months except April and January were statistically insignificant at any reasonable level. January’s coefficient suggests that being born in the first month of the year leads to a 22.12 unit decrease in the number of goals scored. April’s coefficient suggests that being born in the fourth month of the year leads to a 31.68 unit decrease in the number of goals scored.[32] The regression shows that it is not advantageous to be born in January once playing at the professional level. This is surprising because significantly more NHL right-wingers are born in January. This suggests that the relative age effect is beneficial leading up to playing professionally, but that is could be detrimental (for goal scoring) once players are actually at the professional level. There is no statistically significant evidence that suggests being born in the first six months of the year leads to an increase in the number of goals scored.

The second column of Table 10 shows the results from the regression of the dummy variables for each month on total assists. The R-squared of the regression is .013, which suggests that the month of birth can only account for 1.3% of the variation in total assists and shows almost no evidence of a trend. The coefficients for all of the months

[32] April is statistically significant at the 5% level and January is statistically significant at the 20% level

except for January, April, and September were statistically insignificant at any reasonable level. January's coefficient suggests that being born in the first month of the year leads to a 29.94 unit decrease in the number of goals scored. April's coefficient suggests that being born in the 4th month of the year leads to a 39.44 unit decrease in the number of assists. September's coefficient suggests that being born in the ninth month of the year leads to a 24.02 unit decrease in the total number of assists.[33] This is again surprising because it suggests that being born in January leads to fewer career assists. The model shows no significant evidence that being born in the first six months of the year leads to an increase in the number of assists. This regression also suggests that there is no consistent relationship between month of birth and the number of assists scored by right-wingers in professional hockey.

The third column shows the results from the final regression of the dummy variables performed on the dependent variable of total games played. The R-squared of the regression is .008, which means that month of birth can only account for .8% of the variation in total games played and that the data shows almost no evidence of a trend. The coefficients for all of the months except for July were statistically insignificant at any reasonable level, and even still the P-value for July was around .18 so it is not highly informative. July's coefficient suggests that being born in the seventh month of the year leads to a 62 unit decrease in the number of games played. The model shows no significant evidence that being born in the first six months of the year leads to an increase in the number of games played.

[33] April is statistically significant at the 5% level and January is statistically significant at the 20% level

When it comes to goals and assists, right-wingers born in January score less than players born in other months of the year. If these findings supported the presence of the relative age effect, it would mean that the further from the cutoff an athlete is born, the more negatively their game statistics would be impacted, but this does not seem to be the case. The consensus is that once players have reached the level of skill required to play professionally (or even before that), they will have outgrown the relative age effect and closed the hypothetical performance gap. For the majority of players, there is no significant relationship between month of birth and game performance, but it seems that players born in January score fewer goals and make fewer assists than in other months. Our results even suggest that once at the professional level, right-wing players can overcome the relative age effect and there are no long-term effects on performance. The results of these regressions imply a reversal of the relative age effect for those born in January, just like in "The Rise of the Underdog? The Relative Age Effect Reversal Among Canadian-Born NHL Hockey Players: A Reply to Nolan and Howell".

The initial consequences of the relative age effect likely keep more of the "older" players in the game for longer, which is why it seems like most of the right-wings were born nearest to the cutoff.  If older players do benefit from the relative age effect, then it is more likely that they will continue to pursue the sport, even if they don't intend on going pro; thus the distribution of applicants will favor the first half of the year simply because there are more of them still in the game.

The overall conclusion is that for professional right-wing hockey players is that being born in the first six months of the year is advantageous in the time leading up to recruitment, but once the players have been drafted; there are no lasting implications on skill level from the relative age effect, except for players born in January. This appears reasonable: The expectation is that the league hires players up to the point that the inframarginal player born in any month is equal in ability to the inframarginal player born in any other month – if not, there is money being left on the table.

One of the broader implications of the data is that it can be applied to more than just professional hockey. Players in sports that use the calendar year approach to their cutoff date are more likely to be impacted by relative age effect than sports that align with academic calendars. This idea is based on the fact at hand: the relative age effect is typically stronger in sports with smaller age windows and has more pronounced effects on player talent earlier in their careers. This early career boost encourages them to continue playing thus, a larger majority will attempt to play professionally. If there are more right-wing players born in January attempting to play professionally, then obviously the number of players who make it professionally will reflect this instance. This is just one possible explanation of how the relative age effect affects right-wing hockey players in the NHL.

### B. In the NBA

The relative age effect does not appear to have affect making it into the NBA. But, does it affect performance for players who have made it? I do not expect so -- When it comes to playing professional basketball, there is always going to be some slight variation in player skill levels, but in general, players should be on about the same level. Especially since there is a lack of evidence showing that the relative age effect exists up to the time of recruitment. If the relative age effect does have a long-term effect on their skills, then players born nearest to the cutoff should be notably better than those born in later months.

### 1. Data

To determine if the relative age effect has any impact on professional basketball player performance, I performed multiple regressions on the sample, mirroring the regressions on the hockey sample. For the basketball sample, I chose three performance statistics: games played, points, and field goals made. The field goal statistic records the number of baskets scored on any shot or tap other than a free throw, it is worth two or three points depending on the distance of the attempt from the basket. The points statistic records the number of points scored via free throws and field goals. I chose these statistics because they are relatively straightforward, and measure one of the major skills in basketball.

### 2. Empirical Evidence

The regressions are used to determine if, and how much month of birth affects a player's skill level once they have been drafted into the NBA. Since scoring is arguably the main goal of an offensive player it is important to know if the relative age effect impacts their ability to score once they are playing professionally. Table 11 is a presentation of the results from the performance regressions. All three regressions use the same birth month data and assign each month a dummy variable, but the dependent variables (in this case the player statistics) are changed for each regression. All of the regressions seek to answer how being born in month “x” impacts performance statistic “y”.

The results in the first column of Table 11 are from the regression of the dummy variables for each month on total games played. The R-squared of the regression is .032, meaning month of birth can only be attributed to 3.2% of the variation in games played. It also implies that the data show no evidence of a significant trend. The coefficients for all of the months except for July and August were statistically insignificant at any reasonable level. July’s coefficient suggests that being born in the seventh month of the year leads to a 127 unit increase in the number of games played. August’s coefficient suggests that being born in the eighth month of the year leads to a 99 unit increase in the number of games played.[34] Based on these results it is beneficial to be born right before the cut-off date of August 31st, not within the months directly after. The model shows no significant evidence that being born within the first six months after the cut-off date leads to an increase in the number of games played. This initial regression suggests that there is no significant relationship between month of birth and the number of games played by professional basketball players in the NBA. One interesting concept to note is that the months right before the cut-off date (July and August) suggest an increase in the number of games, which does not support the presence of the relative age effect (it implies the opposite).

The results in the second column are from the regression of the dummy variables for each month on total points scored. The R-squared of the regression is .038, which

[34] July is statistically significant at the 10% level and August is statistically significant at the 20% level

suggests that month of birth only accounts for 3.8% of the variation in total points scored and that the data shows no evidence of a trend. The coefficients for all of the months except for October were statistically insignificant at any reasonable level. Even still, the P-value for October was almost .2 so it is not highly informative. October's coefficient suggests that being born in the tenth month of the year leads to a 2658 unit decrease in the total number of points scored. Considering that October is one of the "cut-off months" and has a significantly negative coefficient; we do not have evidence to suggest that being born within the first six months after the cut-off date leads to an increase in the number of points scored by the top 200 NBA players. Our evidence seems to suggest the opposite. This initial regression suggests that there is no significant relationship between most months of birth and points scored by professional basketball players in the NBA.

The results in the third column are from the regression of the dummy variables for each month on total field goals made. The R-squared of the regression is .038, which suggests that month of birth can only be attributed to 3.8% of the variation in field goals made and suggests that the data shows no evidence of a trend. Again, the coefficients for all of the months except for October were statistically insignificant at any reasonable level. The P-value for October was almost .14, so it is still not highly informative. October's coefficient suggests that being born in the tenth month of the year leads to a 1103.027 unit decrease in the total number of field goals made. This implies that being born in October leads to fewer field goals versus players born in other months. This means there is no relative age effect. The model shows no significant evidence that being born within the first six months after the cut-off date leads to an increase in the number of field goals made by the top 200 NBA players.  This initial regression suggests that there is no significant relationship between most months of birth and field goals made by professional basketball players in the NBA.

The results of the regressions on the performance of NBA players are similar to the results of the NHL sample in that neither implies the presence of the relative age effect once playing at the professional level. They suggest there could be some type of reversal because the player statistics for the months nearest the August cut-off date were seemingly the most negatively impacted.

### C. In the MLB

The relative age effect does not appear to affect making it into the MLB. But, does it affect performance for players who have made it?  I do not expect so -- When it comes to playing professional baseball, there is always going to be some slight variation in player skill levels, but in general, players should be on about the same level. Especially since there is a lack of evidence showing that the relative age effect exists up to the time of recruitment. Also, unlike in the other two sports baseball players are generally recruited into minor leagues to improve their skills and set themselves apart before they can play professionally. This step could be a factor that helps "even the playing field" and dismantle the relative age effect. If the relative age effect does have a long-term effect on their skills, then players born nearest to the cutoff should be notably better than those born in later months.

## 1. Data

To determine if the relative age effect has any impact on professional baseball player batting performance, I performed multiple regressions on the sample, mirroring the performance regressions on the hockey and basketball samples. For the baseball sample, I chose three performance statistics: slugging percentage, on-base percentage, and games played. Slugging percentage represents the total number of bases a player records per at-bat. The statistic applies different weights to singles, doubles, triples, and home runs. On-base percentage (OBP) refers to how frequently a batter reaches base per plate appearance. Times on base include hits, walks, and hit-by-pitches, but do not include errors, times reached on a fielder's choice, or a dropped third strike. Both of these statistics are commonly used to evaluate batter performance. The final statistic is games played, in general, the better a player is the more appearances they will make in any sport. I chose these statistics because they are relatively straightforward and measure the major skills of hitters in baseball.

## 2. Empirical Evidence

The regressions are used to determine if, and how much month of birth affects a player's skill level once they have been drafted into the MLB. Since scoring is arguably the main goal of batters it is important to know if the relative age effect impacts their ability to score once they are playing professionally. Table 12 is a presentation of the results from the three-separate t-tests. All three tests use the same birth month data and assign each month a dummy variable, but the dependent variables (in this case the player statistics) are changed for each regression. All of the regressions seek to answer how being born in month “x” impacts performance statistic “y”.

The results in the first column are from the regression of the dummy variables for each month with on-base percentage as the dependent variable. The R-squared of the regression is .001, meaning month of birth only accounts for .1% of the variability in on-base percentage. It also implies that the data show no evidence of a significant trend. The coefficients for all of the months except for June were statistically insignificant at any reasonable level. June’s coefficient suggests that being born in the sixth month of the year leads to a .9% decrease in on-base percentage.[35] Since June is one of the borderline months, these results lead us to believe that there is no relative age effect on on-base percentage for hitters in the MLB. Based on these results there is no significant evidence that being born in a particular month will benefit a player’s on-base percentage. This initial regression suggests that there is no significant evidence of a positive relationship between month of birth and on-base percentage of professional baseball players in the MLB.

The results in the second column are from the regression of the dummy variables for each month on slugging percentage. The R-squared of the regression is .001, meaning month of birth only accounts for .1% of the variability in slugging percentage. It also suggests that the data shows no evidence of a trend. The coefficients for September and October show some of the smallest decreases in slugging percentage and are both statistically significant at the 1% level. The June coefficient has the smallest decrease in

[35] June is statistically significant at the 10% level.

slugging percentage at only 4%. While all of the months have negative coefficients, there is some evidence that being born in September or October makes players better off. For example, being born in September leads to a 1.3% decrease in slugging percentage on average versus a 2.4% decrease on average in June. The model shows no significant evidence that being born within the first six months of the cut-off date leads to an increase in the slugging percentage of professional baseball players in the MLB. The relative age effect does not seem to positively impact professional baseball player slugging percentage, it seems that being born in the months right after the cut-off date contributes negatively to the slugging percentage.

The results in the third column are from the regression of the dummy variables for each month on total games played. The R-squared of the regression is .001, which suggests that month of birth only accounts for .1% of the variation in games played and that the data shows almost no evidence of a trend. The coefficients for all of the months except for January and September were statistically insignificant at any reasonable level. January's coefficient suggests that being born in the first month of the year leads to a 33 unit decrease in the number of games played. September's coefficient suggests that being born in the ninth month of the year leads to a 30 unit decrease in the number of games played.[36] Again, September is one of the borderline months, and it shows a significant negative impact on total games played, which refutes the presence of the relative age effect. The model shows no significant evidence that being born within the first six

[36] January is statistically significant at the 15% level and September is statistically significant at the 10% level

months of the cut-off date leads to an increase in the number of games played. This initial regression suggests that there is no significant relationship between most months of birth and number of games played by professional baseball players in the MLB.

The results of the regressions on the performance of MLB players are similar to the results of the NHL sample and the NBA sample in that neither implies the presence of the relative age effect on performance once playing at the professional level. The results of the regressions suggest that being born right after the cut off could negatively impact player statistics at the professional level.

## VI. CONCLUSION

The relative age effect has long term implications on right-wing hockey players until they reach the professional level. Players born in the first six months of the year are more likely to play professionally. Once hockey players reach the professional level, the benefits of the relative age effect dissipate and even diminish the performance. The same cannot be said for professional basketball players in the NBA, or professional baseball players in the MLB. We found no evidence that the players born in the months following the cutoff dates are more likely to play professional basketball or professional baseball. When it comes to skills and performance, there are statistically significant findings for all three sports that players born in the immediate months following the cut-off dates, experience negative effects regarding their performance. This implies a reversal of the relative age effect or a negative relationship with skill and month of birth. Without further research, there is no way to prove that the relative age effect has long-term or statistically significant effects on future performance as a professional athlete. As of now, it seems the relative age effect does not have long term implications on professional athletes. This is due to the fact that players tend to catch up with each other, in terms of performance, if they choose to pursue professional hockey, basketball, or baseball. Overall, it is likely that the relative age effect can only be observed in a select group of professional sports. One of the options for future research would be to look specifically at youth sports instead of professional sports and assess the effects of the relative age effect in terms of different windows of eligibility/ cut-off dates. Identifying the implications of the relative age effect in youth sports, rather than professional sports, would be an advantageous opportunity for future research. This type of study would allow researchers to reconsider windows of eligibility and cut-off dates with the best interests of children in mind.

There is no way to know the exact reason that some sports seem to be impacted by the relative age effect and others are not, but we can speculate. One possibility I suspect is the difference between the cut-off dates and the sizes of the windows of eligibility in each of the sports. Hockey leagues rely on the calendar year to determine their cut-off dates and only allow for 365 days for league eligibility. This means that they solely rely on age as a factor to determine team or level of play. By this standard, youth hockey leagues are reinforcing the relative age effect. This is because particularly at the beginning of their careers the biggest difference in the players is based on their physical edge. This small difference creates a performance gap that persists over time. It is important to note that youth hockey leagues are divided based on age and level of competition, so the structure inherently benefits children with a physical edge and places them in more competitive/ better leagues. On the contrary, youth basketball leagues use both age and academic cut-off standards that allow for a twenty-four-month inclusion window. The twenty-four-month window helps explain why the NBA month of birth distribution mirrors the general population. This means that if a parent decides to "redshirt" their child there is a chance that they will be playing against a child that is up to twenty-four months younger than them. This large window increases the so-called "season" of birth and decreases the concentration of players that can benefit from the relative age effect. Essentially, if the players are born within the first twelve months of the eligibility period or qualify by their school guidelines, they will benefit from the relative age effect. This makes it much more difficult to narrow down a specific month because the range of inclusion is higher, thus more players benefit. Even still, a player born at the beginning of the twenty-four-month eligibility period should hypothetically benefit most from the relative age effect, but according to our results, they do not. The baseball sample is the most difficult to analyze because of the differences in their league structures. Since the age group structure is so loose, hypothetically we shouldn't see much of a relative age effect since there is no basis to compare players who are close in age. If league structure does play a substantial role in implementing the relative age effect, then there should be the smallest instances in baseball.

It is important to note that the relative age effect is not the only reason that some athletes excel in their sport. Athletic success comes from a multitude of factors and cannot simply be contributed to what month a baby is born. The relative age effect is more prevalent in sports where (at least in adolescence) maturation and size is seen as beneficial and determines whether someone “makes the cut” (ie. hockey, football, and rugby). Once a player has reached the level of skill needed to be selected by a professional team (in any sport) it is more than likely that their talent and dedication to the sport is on the same level as all of the other players in the same running. In some cases, our data shows that these players are better off.

This study could be improved by using larger samples, different positions, and a better way to control for the cut-off date in the baseball sample specifically. Beyond sports, the results of this study help us to understand the relative age effect regarding academic advantage, CEO advantage, youth suicide statistics, and social/behavioral development in adolescents. Further research is needed to fully understand the relative age effect and how it impacts society.

# VII. BIBLIOGRAPHY

Baker, J. (2001). Genes and Training for Atheletic Performance Revisited. *Sportscience*, 1.

Baker, J., & Logan, A. J. (2007). Developmental contexts and sporting success: birth date and birthplace effects in national hockey league draftees 2000–2005. *Br J Sports Med*, 515-517.

Bassok, D. (2013). “Academic Redshirting” in Kindergarten: Prevalence, Patterns & Implications. *Educational Evaluation and Policy Analysis*, 1-18.

Buchheit, M., Mendez-Villanueva, A., Mayer, N., Julien, H., Marles, A., Bosquet, L., . . . Lambert, P. (2013). Locomotor Performance in Highly-Trained Young Soccer Players: Does Body Size Always Matter? *International Journal of Sprots Medicine*, 1-12.

CDC. (2017). Provisional monthly and 12-month ending number of live births, deaths, and infant deaths and rates: United States.

Deaner, R. O., Lowen, A., & Cobley, S. (2013, February 27). Born at the Wrong Time: Selection Bias in the NHL Draft. *PLOS*, 1-5.

Du, Q., Gao, H., & Levi, M. D. (2012, August 14). The relative-age effect and career success: Evidence from Corporate CEOs. *Elsevier*, 37-49.

Fumarco, L. R. (2016, April 23). Long-term Relative Age Effect: Evidence from Italian Football. *VOX*, 1.

Fumarco, L., Gibbs, B., Jarvis , J., & Rossi, G. (2017). The relative age effect reversal among the national hockey league elite. *PLoS One*, 8.

Gibbs, B. G., Jarvis, J., & Dunfur, M. J. (2012). The Rise of the Underdog? The Relative Age Effect Reversal Among Canadian-Born NHL Hockey Players: A Reply to Nolan and Howell. *International Review for the Sociology of Sport*, 644-649.

Gladwell, M. (2011). *Outliers: The Story of Success.* New York: Back Bay Books.

Hockey Monkey. (2020). *Ice Hockey Positions: Skills, Roles & Responsibilities Explained*. Retrieved from https://www.hockeymonkey.com/learn/hockey-positions#

Ibáñez, S. J., Mazo, A., Nascimento, J., & García-Rubio, J. (2018). The Relative Age Effect in under-18 basketball: Effects on performance according to playing position. *PLoS ONE*, 1-11.

Kiikka, D. (2017, March 31). *Birth Month is a Big Factor in Atheletic Performance*. Retrieved from danielkiika.com: https://danielkiikka.com/performance/birth-month-athletic-performance/

Knoema. (2017, March 16). *Which are the Busiest Months for Births?* Retrieved from Knoema: https://knoema.com/infographics/fjsoxbc/which-are-the-busiest-months-for-births

Lahman, S. (2020). *Baseball Archive*. Retrieved from SeanLahman.com: http://www.seanlahman.com/baseball-archive/

*Levels of Minor (youth) hockey*. (2015). Retrieved from The Hockey Fanatic: http://www.thehockeyfanatic.com/hockey-parents/levels-of-minor-hockey/

Little League. (2020). *League Age of Determination*. Retrieved from littleleague.org.

Musch, J., & Hay, R. (1999). The Relative Age Effect in Soccer: Cross-Cultural vidence for a Systematic Discrimination Against Children Born Late in the Competition Year. *Sociology of Sport Journal, 16*(1), 54-64.

National Center For Education Statistics. (2009, June). Is He Ready for K? *Working Mother, 32*(6), 88.

NCAA. (2012). The Birthday Effect in College Atheletes. *Extra Point*. Retrieved from ncaa.org.

Stracciolini, A., Friedman, H. L., Casciano, R., Howell, D., Sugimoto, D., & Micheli, L. J. (2016). The Relative Age Effect on Youth Sports Injuries. *Medicine and Science in Sports and Exercise*, 13-22.

Thompson, A. H., Barnsley, R. H., & Battle, J. (2004). *Educational Research, 46*(3), 313-320.

Thompson, A. H., Barnsley, R. H., & Dyck, R. J. (1999). A New Factor in Youth Suicide: The Relative Age Effect. *Can J Psychiatry*, 82-85.

USAB. (2019). *Youth Basketball Guidelines*. Retrieved from USA Basketball: https://www.usab.com/youth/development/youth-basketball-guidelines.aspx

Wattie, N., Baker, J., Cobley, S., & Montelpare, W. (2007). A Historical Examination of Relative Age Effects in Canadian Hockey Players. *International Journal of Sport Psychology*, 178-186.

**Table 1: Distribution of NHL Players Month of Birth**

| Month | # of Hockey Players | Hockey Birth Rate |
|---|---|---|
| January | 163 | 10.97643 |
| February | 145 | 9.76431 |
| March | 151 | 10.16835 |
| April | 151 | 10.16835 |
| May | 123 | 8.282828 |
| June | 123 | 8.282828 |
| July | 114 | 7.676768 |
| August | 107 | 7.205387 |
| September | 110 | 7.407407 |
| October | 105 | 7.070707 |
| November | 98 | 6.599327 |
| December | 95 | 6.397306 |
| sum | 1485 | |

## Table 2: Chi-Squared test for Hockey Sample

| Month | Observed | Expected % | Expected | Residual | (obs-exp)^2 | Chi squared |
|---|---|---|---|---|---|---|
| January | 163 | 0.081433574 | 120.9288573 | 42.07114271 | 1769.981049 | 14.63654821 |
| February | 145 | 0.075537769 | 112.1735868 | 32.82641321 | 1077.573404 | 9.606302471 |
| March | 151 | 0.082827128 | 122.9982849 | 28.00171514 | 784.096051 | 6.374853535 |
| April | 151 | 0.079968556 | 118.7533052 | 32.24669478 | 1039.849324 | 8.756382166 |
| May | 123 | 0.083112985 | 123.4227828 | -0.42278282 | 0.178745313 | 0.001448236 |
| June | 123 | 0.083148717 | 123.4758451 | -0.475845065 | 0.226428526 | 0.001833788 |
| July | 114 | 0.088186951 | 130.9576217 | -16.95762167 | 287.5609326 | 2.195831972 |
| August | 107 | 0.089187451 | 132.4433645 | -25.44336454 | 647.3647991 | 4.887861323 |
| September | 110 | 0.087150718 | 129.4188166 | -19.41881655 | 377.0904362 | 2.913721871 |
| October | 105 | 0.085221182 | 126.5534553 | -21.5534553 | 464.5514353 | 3.67079219 |
| November | 98 | 0.080361609 | 119.3369899 | -21.33698992 | 455.267139 | 3.814970859 |
| December | 95 | 0.08386336 | 124.53709 | -29.53708997 | 872.4396841 | 7.005460657 |
| | | | | | | 63.86600727 |

**This test assumes normal distribution.**

**The critical value for comparison is 19.68.**

**Table 3: T-test for Mean Monthly Differences Between the Hockey Sample and General Population**

| Month | Hockey Birth Rate (1) | Gen pop Birth Rate (2) | Test of Differences (1)-(2) | P-Value | Test Statistic |
|---|---|---|---|---|---|
| January | 10.97643 | 8.143357 | 2.833073 | 0.0005 | 3.4913 |
| | | | | | (0.0081) |
| February | 9.76431 | 7.553777 | 2.210533 | 0.0042 | 2.8688 |
| | | | | | (0.0077) |
| March | 10.16835 | 8.282713 | 1.885637 | 0.0164 | 2.4035 |
| | | | | | (0.0078) |
| April | 10.16835 | 7.996856 | 2.171494 | 0.0057 | 2.7678 |
| | | | | | (0.0078) |
| May | 8.282828 | 8.311299 | -0.028471 | 0.9683 | -0.0398 |
| | | | | | (0.0072) |
| June | 8.282828 | 8.314872 | -0.032044 | 0.9643 | -0.749 |
| | | | | | (0.0072) |
| July | 7.676768 | 8.818695 | -1.141927 | 0.0987 | -1.6524 |
| | | | | | (0.0069) |
| August | 7.205387 | 8.918745 | -1.713358 | 0.0264 | -2.5526 |
| | | | | | (0.0061) |
| September | 7.407407 | 8.715072 | -1.307665 | 0.0546 | -1.9235 |
| | | | | | (0.0068) |
| October | 7.070707 | 8.522118 | -1.451411 | 0.0293 | -2.1812 |
| | | | | | (0.0067) |
| November | 6.599327 | 8.036161 | -1.436834 | 0.0259 | -2.7344 |
| | | | | | (0.0064) |
| December | 6.397306 | 8.386336 | -1.98903 | 0.0018 | -3.1312 |
| | | | | | (0.0064) |

**Table 4: Distribution of NBA Players Month of Birth**

| Month | # of Basketball Players | Basketball Birth Rate |
|---|---|---|
| January | 14 | 7 |
| February | 21 | 10.5 |
| March | 20 | 10 |
| April | 14 | 7 |
| May | 18 | 9 |
| June | 18 | 9 |
| July | 16 | 8 |
| August | 19 | 9.5 |
| September | 17 | 8.5 |
| October | 14 | 7 |
| November | 16 | 8 |
| December | 13 | 6.5 |
| sum | 200 | |

**Table 5: Chi-Squared test for Basketball Sample**

| Month | Observed | Expected % | Expected | Residual | (obs-exp)^2 | Chi squared |
|---|---|---|---|---|---|---|
| January | 14 | 0.08143357 | 16.2867148 | -2.286714786 | 5.22906451 | 0.321063184 |
| February | 21 | 0.07553777 | 15.1075538 | 5.892446223 | 34.7209225 | 2.298249141 |
| March | 20 | 0.08282713 | 16.5654256 | 3.43457443 | 11.7963015 | 0.712103741 |
| April | 14 | 0.07996856 | 15.9937111 | -1.993711141 | 3.97488411 | 0.248527942 |
| May | 18 | 0.08311299 | 16.622597 | 1.377402987 | 1.89723899 | 0.114136136 |
| June | 18 | 0.08314872 | 16.6297434 | 1.370256557 | 1.87760303 | 0.112906314 |
| July | 16 | 0.08818695 | 17.6373901 | -1.637390124 | 2.68104642 | 0.152009248 |
| August | 19 | 0.08918745 | 17.8374902 | 1.162509826 | 1.3514291 | 0.075763411 |
| September | 17 | 0.08715072 | 17.4301436 | -0.430143643 | 0.18502355 | 0.010615148 |
| October | 14 | 0.08522118 | 17.0442364 | -3.044236404 | 9.26737528 | 0.543724874 |
| November | 16 | 0.08036161 | 16.0723219 | -0.072321875 | 0.00523045 | 0.000325432 |
| December | 13 | 0.08386336 | 16.7726721 | -3.77267205 | 14.2330544 | 0.848585983 |
| Sum | 200 | | | | | 5.438010554 |

**This test assumes normal distribution.**

**The critical value for comparison is 19.68.**

**Table 6: T-test for Mean Monthly Differences between Basketball Sample and General Population**

| Month | B-ball Birth Rate (1) | Gen pop Birth Rate (2) | Test of Differences (1)-(2) | P-Value | Test Statistic |
|---|---|---|---|---|---|
| January | 7 | 8.143357 | 1.143357 | 0.528 | -0.6321 (0.018) |
| February | 10.5 | 7.553777 | 2.94622 | 0.1767 | 1.3558 (0.022) |
| March | 10 | 8.282713 | 1.71729 | 0.4203 | 0.8075 (0.021) |
| April | 7 | 7.996856 | 0.996856 | 0.5822 | -0.5511 (0.018) |
| May | 9 | 8.311299 | 0.6887 | 0.7346 | 0.3395 (0.020) |
| June | 9 | 8.314872 | 0.68513 | 0.7359 | 0.3377 (0.020) |
| July | 8 | 8.818695 | 0.818695 | 0.6708 | -0.4257 (0.019) |
| August | 9.5 | 8.918745 | 0.58125 | 0.78 | 0.2796 (0.021) |
| September | 8.5 | 8.715072 | 0.215072 | 0.9135 | -0.1088 (0.0198) |
| October | 7 | 8.522118 | 1.522118 | 0.401 | -0.8416 (0.0181) |
| November | 8 | 8.036161 | 0.036161 | 0.985 | -0.0188 (0.019) |
| December | 6.5 | 8.386336 | 1.886336 | 0.2817 | -1.0794 (0.018) |

**Table 7: Distribution of MLB Players Month of Birth**

| Month | # Baseball Players | Baseball Birth Rate |
|---|---|---|
| January | 1346 | 8.7 |
| February | 1195 | 7.7 |
| March | 1268 | 8.2 |
| April | 1195 | 7.7 |
| May | 1195 | 7.7 |
| June | 1131 | 7.3 |
| July | 1255 | 8.1 |
| August | 1490 | 9.6 |
| September | 1347 | 8.7 |
| October | 1453 | 9.4 |
| November | 1322 | 8.5 |
| December | 1271 | 8.2 |
| Sum | 15468 | |

**Table 8: Chi-Squared test for Baseball Sample**

| Month | Observed | Expected % | Expected | Residual | (obs-exp)^2 | Chi squared |
|---|---|---|---|---|---|---|
| January | 1346 | 0.081433574 | 1259.614522 | 86.38547845 | 1769.981049 | 1.405176758 |
| February | 1195 | 0.075537769 | 1168.418209 | 26.5817909 | 1077.573404 | 0.922249752 |
| March | 1268 | 0.082827128 | 1281.170014 | -13.17001358 | 784.096051 | 0.612015613 |
| April | 1195 | 0.079968556 | 1236.95362 | -41.95361967 | 1039.849324 | 0.840653447 |
| May | 1195 | 0.083112985 | 1285.591653 | -90.59165297 | 0.178745313 | 0.000139037 |
| June | 1131 | 0.083148717 | 1286.144358 | -155.1443579 | 0.226428526 | 0.000176052 |
| July | 1255 | 0.088186951 | 1364.075752 | -109.0757522 | 287.5609326 | 0.21081009 |
| August | 1490 | 0.089187451 | 1379.55149 | 110.44851 | 647.3647991 | 0.469257439 |
| September | 1347 | 0.087150718 | 1348.047309 | -1.047309369 | 377.0904362 | 0.279730862 |
| October | 1453 | 0.085221182 | 1318.201243 | 134.7987565 | 464.5514353 | 0.352413137 |
| November | 1322 | 0.080361609 | 1243.033374 | 78.96662617 | 455.267139 | 0.36625496 |
| December | 1271 | 0.08386336 | 1297.198456 | -26.19845637 | 872.4396841 | 0.672556832 |
| | 15468 | | | | | 6.13143398 |

**This test assumes normal distribution.**

**The critical value for comparison is 19.68.**

**Table 9: T-test for Monthly Differences between MLB and General Population**

| Month | Baseball Birth Rate (1) | Gen pop Birth Rate (2) | Test of Differences (1)-(2) | P-Value | Test Statistic |
|---|---|---|---|---|---|
| January | 8.7 | 8.143357 | 0.556643 | 0.014 | 2.646 (0.00227) |
| February | 7.7 | 7.553777 | 0.146223 | 0.423 | 0.8 (0.00215) |
| March | 8.2 | 8.282713 | -0.082713 | 0.7 | -0.316 (0.00221) |
| April | 7.7 | 7.996856 | -0.296856 | 0.206 | -1.263 (0.00215) |
| May | 7.7 | 8.311299 | -0.611299 | 0.006 | -2.728 (0.00215) |
| June | 7.3 | 8.314872 | -1.014872 | 0.0 | -4.792 (0.00209) |
| July | 8.1 | 8.818695 | -0.718695 | 0.001 | -3.212 (0.0022) |
| August | 9.6 | 8.918745 | 0.681255 | 0.003 | 3.01 (0.00237) |
| September | 8.7 | 8.715072 | -0.015072 | 0.976 | -0.03 (0.00227) |
| October | 9.4 | 8.522118 | 0.877882 | 0.0 | 3.715 (0.00235) |
| November | 8.5 | 8.036161 | 0.463839 | 0.023 | 2.271 (0.00225) |
| December | 8.2 | 8.386336 | -0.186336 | 0.443 | -0.767 (0.00221) |

**Table 10: Hockey Performance Regression**

| Variable | Goals | Assists | Games Played |
|---|---|---|---|
| January | -22.12141*** | -29.941** | -33.027 |
| | (14.17053) | (17.333) | (41.974) |
| February | 5.480581 | 5.526 | 29.655 |
| | (14.49078) | (17.725) | (43.945) |
| March | -14.4336 | -20.437 | 40.469 |
| | (14.37637) | (18.585) | (43.598) |
| April | -31.67863* | -39.443* | -37.041 |
| | (14.37637) | (18.585) | (43.598) |
| May | -4.20113 | -10.467 | 1.394 |
| | (14.99498) | (18.341) | (45.474) |
| June | -7.526316 | -8.288 | 16.824 |
| | (14.99498) | (18.341) | (45.474) |
| July | 7.315789 | 4.582 | 62.067*** |
| | (15.25074) | (18.654) | (46.25) |
| August | -16.42351 | -18.519 | -28.488 |
| | (15.47583) | (18.929) | (46.933) |
| September | -17.28086 | -24.019*** | 5.697 |
| | (15.37626) | (18.808) | (46.631) |
| October | 12.53083 | 10.631 | 5.798 |
| | (15.54499) | (19.014) | (47.142) |
| November | -8.393663 | -14.221 | -0.307 |
| | (15.8065) | (19.334) | (47.935) |
| | | | |
| R2 | 0.014 | 0.013 | 0.008 |
| # obs | 1485 | 1485 | 1485 |

*=stat significant at the 5% level

**=stat significant at the 10% level

***=stat significant at the 20% level

**Table 11: Basketball Performance regressions**

| Variable | Games Played | Points | Field Goals Made |
|---|---|---|---|
| January | 86.093 | -102.203 | 49.83 |
| | (78.451) | (1868.497) | (717.353) |
| February | 35.736 | -1410.322 | -490.766 |
| | (71.88) | (1712.003) | (657.271) |
| March | 99.358 | -1050.996 | -449.685 |
| | (72.564) | (1728.29) | (663.524) |
| April | 58.022 | 167.154 | 148.973 |
| | (78.451) | (1868.497) | (717.353) |
| May | 63.974 | -637.235 | -284.274 |
| | (74.135) | (1765.709) | (677.891) |
| June | 58.752 | -661.902 | -271.44 |
| | (74.135) | (1765.709) | (677.891) |
| July | 127.308* | 210.279 | 114.99 |
| | (76.053) | (1811.397) | (695.431) |
| August | 99.202*** | 1090.312 | 259.563 |
| | (73.312) | (1746.115) | (670.368) |
| September | 58.602 | -1066.67 | -495.443 |
| | (75.044) | (1787.355) | (686.201) |
| October | 26.165 | -2658.132*** | -1103.027** |
| | (78.451) | (1868.497) | (717.353) |
| November | 10.245 | -1453.132 | -549.385 |
| | (76.053) | (1811.397) | (695.431) |
| | | | |
| R2 | 0.032 | 0.038 | 0.038 |
| # obs | 200 | 200 | 200 |

*=stat significant at the 10% level

**=stat significant at 15% level

***=stat significant at 20% level

**Table 12: Baseball Performance Regression**

| Variable | On Base Percentage | Slugging Percentage | Games Played |
|---|---|---|---|
| January | -0.001 | -0.016** | -33.821** |
| | (0.005) | (0.008) | (19.243) |
| February | -0.006 | -0.019* | -5.945 |
| | (0.005) | (0.009) | (19.824) |
| March | -0.006 | -0.016** | 5.492 |
| | (0.005) | (0.009) | (19.528) |
| April | 0.002 | -0.004 | 12.748 |
| | (0.005) | (0.009) | (19.824) |
| May | -0.002 | -0.012 | -16.612 |
| | (0.005) | (0.009) | (19.824) |
| June | -0.009** | -0.024* | 9.634 |
| | (0.005) | (0.009) | (20.111) |
| July | -0.004 | -0.016** | -5.572 |
| | (0.005) | (0.009) | (19.579) |
| August | -0.005 | -0.01 | -4.413 |
| | (0.005) | (0.008) | (18.768) |
| September | -0.004 | -0.013*** | -30.155*** |
| | (0.005) | (0.008) | (19.239) |
| October | 0.001 | -0.012*** | 4.981 |
| | (0.005) | (0.008) | (18.895) |
| November | -0.006 | -0.021* | 7.558 |
| | (0.005) | (0.009) | (19.327) |
| | | | |
| R2 | 0.001 | 0.001 | 0.001 |
| # obs | 15468 | 15468 | 15468 |

*=stat significant at the 5% level

**=stat significant at 10% level

***=stat significant at 15% level

**Figure 1: Monthly Comparison of NHL sample to the General Population**

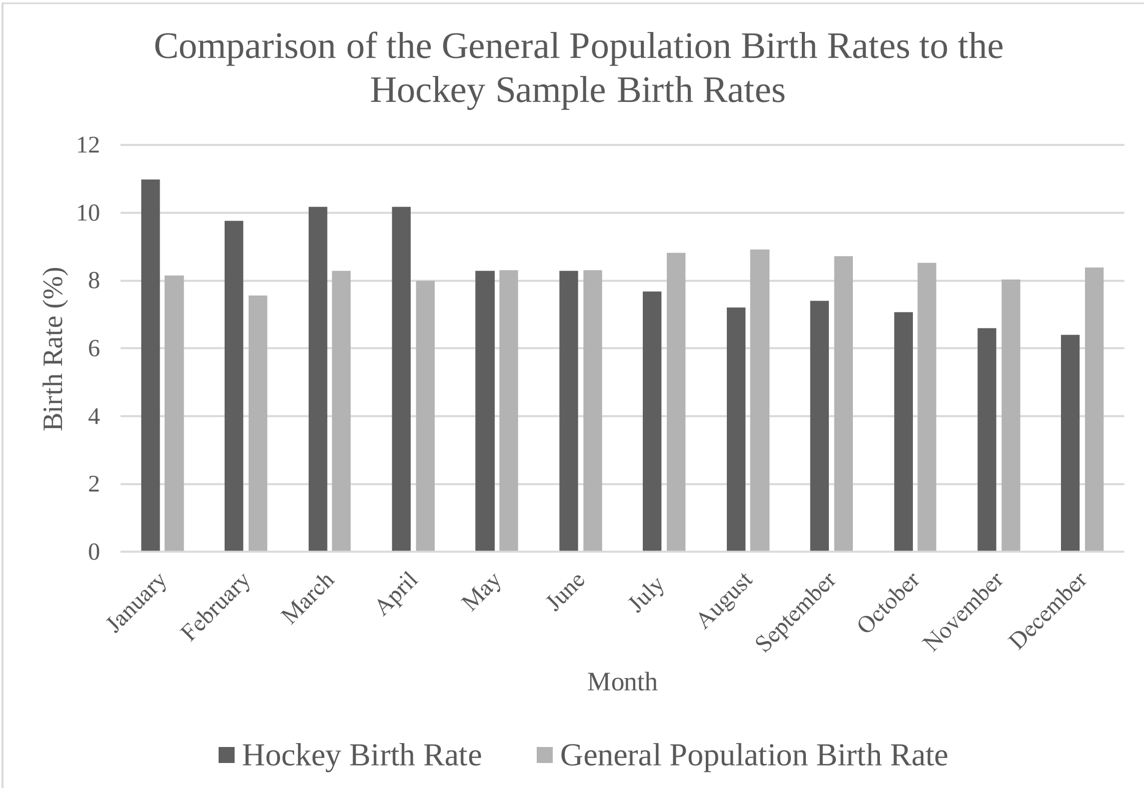


**Figure 2: Monthly Comparison of NBA sample to the General Population**

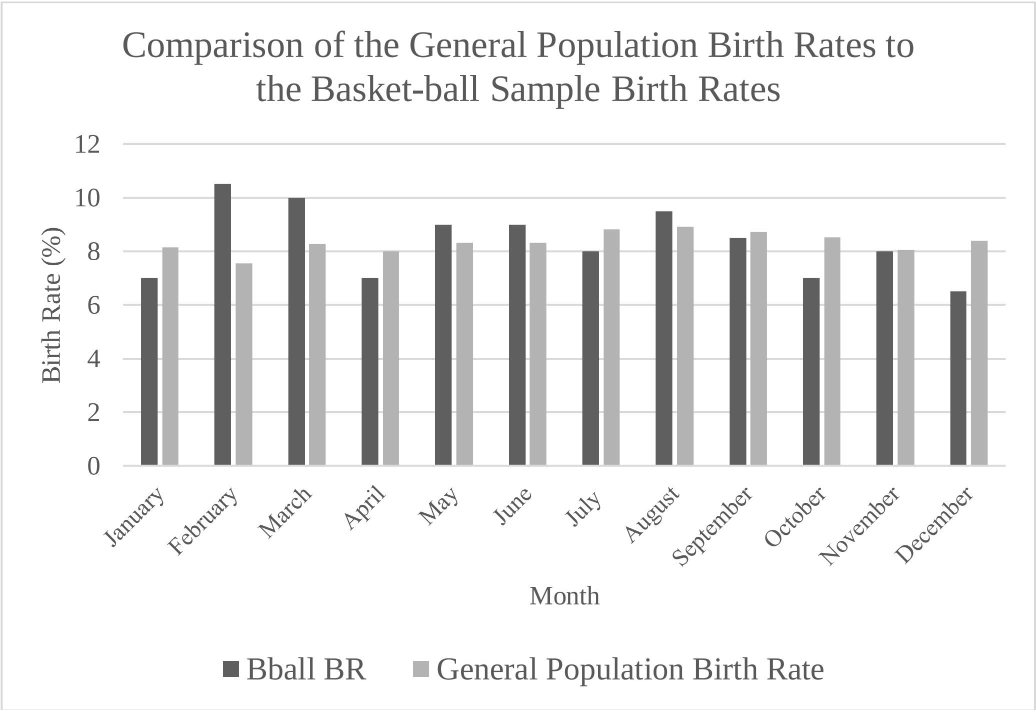
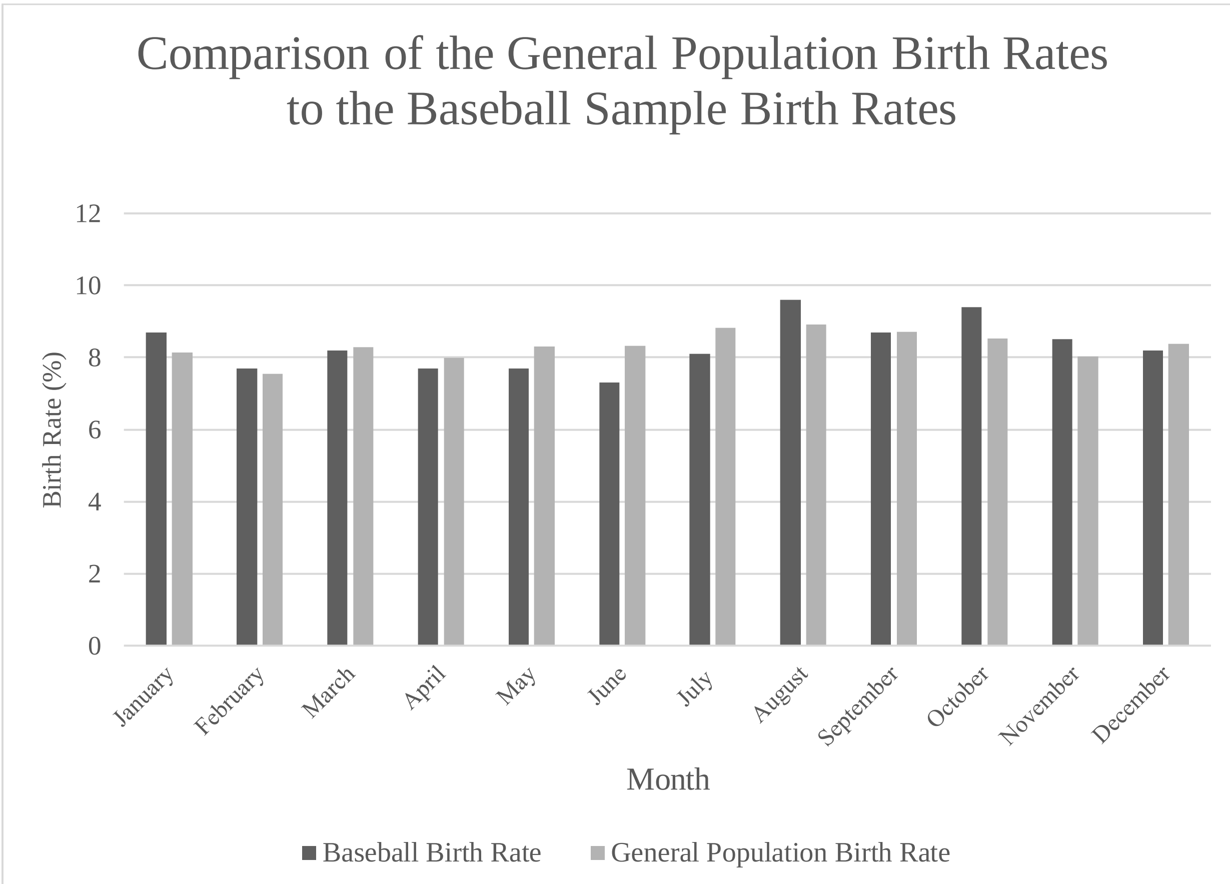


**Figure 3: Monthly Comparison of MLB sample to the General Population**



## VIII. APPENDIX

**Table 1: Distribution of the General Populations Month of Birth**

| Month | General Population Births | General Population Birth Rate |
|---|---|---|
| January | 2279000 | 8.143357 |
| February | 2114000 | 7.553777 |
| March | 2318000 | 8.282713 |
| April | 2238000 | 7.996856 |
| May | 2326000 | 8.311299 |
| June | 2327000 | 8.314872 |
| July | 2468000 | 8.818695 |
| August | 2496000 | 8.918745 |
| September | 2439000 | 8.715072 |
| October | 2385000 | 8.522118 |
| November | 2249000 | 8.036161 |
| December | 2347000 | 8.386336 |
| sum | 27986000 | |