Does the Relative Age Effect Exist in Elite Sport? An Analysis of Olympic Competition

Kathryn McGhee Wingfield

Thesis submitted to the Graduate Faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
in
Human Nutrition, Foods and Exercise

Committee Members:

Jay H. Williams
Elena Serrano
Kathy Hosig

May 5, 2017
Blacksburg, Virginia

Keywords: sport, youth, Olympics, development, exercise, talent identification

Does the Relative Age Effect Exist in Elite Sport? An Analysis of Olympic Competition

Kathryn M. Wingfield

ABSTRACT

Studies have concluded that youth sports programs have a bias selection process in identifying player talent. Athletes that are identified as talented are more likely to be born in the first three months after the eligibility cut-off for a program's particular age group. This is referred to as the relative age effect (RAE) and has been identified in many youth sports. However, it is not known if the RAE carries over into elite, adult competition. The purpose of this study was to determine if the RAE exists in Olympic competition and to compare the RAE between genders, team vs individual sports, weight class vs non-weight class sports, and medalists vs non-medalists. Data on Olympians competing in the 2012 London and 2016 Rio de Janeiro Olympics were gathered from publicly available databases. Lorenz curves were constructed and Gini coefficients calculated to detect unexpected distributions of birth months. In addition, linear regression was used to determine a directional distribution. A negative Gini coefficient and a statistically significant negative slope of the birth month distribution suggested the existence of a RAE. The results showed that there was a RAE in Olympic competition. For all athletes, the Gini coefficient was -0.0324 and the slope of -0.0014 fraction of athletes born per month. Within specific sports, the RAE varies considerably with some showing a positive RAE. Further, the RAE in Olympic athletes does not seem to be influenced by gender, type of team or success of the athlete.

Does the Relative Age Effect Exist in Elite Sport? An Analysis of Olympic Competition

Kathryn M. Wingfield

GENERAL AUDIENCE ABSTRACT

Many children and youth participate in sport programs around the world. Research studies have concluded that youth sports programs have a biased selection process in identifying player talent. Athletes identified as talented are more likely to be born in the first three months after the cut-off date for a program's age group. This is known as the relative age effect (RAE). As RAE is known in youth sports, it is not known if the RAE carries over into elite, adult competition. The purpose of this study was to determine if a RAE exists in Olympic competition and to compare the RAE between genders, team vs individual sports, weight class vs non-weight class sports and medalists vs non-medalists. Data on Olympians competing in the 2012 London and 2016 Rio de Janeiro Olympic games were gathered from publicly available databases. From the data, the athletes were sorted based on birthdate, height, weight, gender, sport and medals earned. Sports were then classified as team or individual and as weight class or non-weight class. The expected distribution was 8.3% per month, since this represents an equal number of athletes across all months. Linear regression was used to determine the direction. A negative slope of the birth month distribution suggested the existence of a RAE. A RAE appears to exist within Olympic level competition. Furthermore, RAE gets smaller but still exists as athletes move from the youth to the adult level. Within specific sports, the RAE varies considerably with some not showing a RAE.

ACKNOWLEDGEMENTS

Wow! What a year! This past year has been one of the craziest and best rides on this roller coaster that we all call life!

First, thank you to the good Lord above for all the blessings of life. I never knew that obtaining a Master's degree in one year was attainable, but with God, all things are possible. (Matthew 19:26)

Second, thank you to my parents and sister. I would not be half the person I am today without your support and encouragement along the way. You all have shown me what hard work and dedication resembles and for that, I am truly thankful.

Next, I would like to thank my advisor, Dr. Jay Williams. You have been nothing short of a blessing to me. You picked me up at my lowest point and believed in me when I did not believe in myself. You have taught me to persevere through tough obstacles that life will throw. The lessons that I have learned from you in and out of the classroom are unexplainable and the memories of this experience will last a lifetime.

Lastly, I would like to thank my committee members, Dr. Elena Serrano and Dr. Kathy Hosig. Both of you were mentors for me during this process and I am grateful that you were on my committee. I am thrilled that I had the opportunity to get to know both of you and work on this project together. Your input and guidance helped make this project successful.

As my time at Virginia Tech is ending, I am not only taking away memories, but also an exceptional education. During my graduate career, I gained a purpose and that is something that cannot be taught in a classroom. Thank you Virginia Tech! Now, I know what it truly means to be a Hokie.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

It is well known that many children or youth participate in youth sport programs around the world. The purposes of these programs are to increase physical activity and have a positive impact on youth development. Sports organizations strive to make youth sports programs fair for everyone. To have different youth leagues, these groups are known to split up young athletes by age categories to hopefully ensure that players receive age-appropriate training and competition in order to maximize opportunity for success (Helsen, 2005). These organizations categorize youth by age group that spans over a one to two-year period with the players born between January 1 and December 31 (Williams, 2010). Using this style of sorting, it should be known that participants born earlier in the year may be as much as twelve months older than their peers born later in that same year. For the younger players, not only can their peers be older by age, but also developmentally and maturity wise which is an important factor in the athletic regime. As a general hypothesis, the athletes that are most successful within their sport are usually the older participants. Malina et al. (2005) concluded that their results are consistent with this general hypothesis that male athletes who are biologically more mature are more successful in soccer as they grow through adolescence.

The discovery of this phenomenon dates back to about thirty-three years ago, Grondin, Deshaies and Nault in 1984 and Barnsley, Thompson and Barnsley in 1985 first discussed the possible relationship among athletes' birthdate and sport participation that is now known as the "relative age effect" (RAE) (Musch et al. 2001). The RAE suggests that more successful athletes are more likely to be born in the first three months after the eligibility cut-off for a particular age group in sports. (http://playerwelfare.worldrugby.org/) In 1984, Grondin et al. first identified a skewed representation in young hockey players and in the National Hockey League. This research showed an overrepresentation of athletes born in the first three months of the year and an underrepresentation of athletes born in the last three months. Recent research has concluded that RAE is prevalent among youth sports such as soccer, baseball, volleyball, basketball, ice hockey, tennis and many other team and individual sporting events (Helsen, 2005; Williams, 2009; Malina, 2005; Malina, 2000; Arrieta, 2016; Edgar, 2005; Campos, 2016; Boucher, 1994). The authors, Grondin et al. (1984) and Barnsley et al. (1988) also stated that this skewed relationship with birth date could be caused by the cutoff date to determine the age grouping of minor hockey. These authors conducted a follow up study and found that RAE was increased in elite youth teams. The authors' results displayed that over thirty percent of the top tier players were born in the first half of the year. In addition, less than fifteen percent of the top tier players were born in the fourth quarter, October through December (Barnsley et al., 1988).

In many youth sports, the coaches have the authority to choose their players. This is especially true in team sports. With this concept in mind, coaches are more prone to select young athletes based on subjective evaluation of performance rather than on objective measures. Some examples of characteristics that influence subjective player selection are height, weight, physical maturity and relative age. These parameters may influence a coach's decision but are not necessarily reflective of the player's talent, skill or potential for development. On the other hand, objective measures of performance are often overlooked.

For example, how well someone performs a task, interaction with teammates and tactical knowledge may be viewed as less important. Subjective evaluations may have a negative impact on the younger players because they may be noticeably smaller and not as physically developed as their older teammates despite having more skill and knowledge of the sport. The comparative analysis by Malina et al. (2005) suggests that soccer suffers from systematically excluding late maturing boys and favors early maturing as the chronological age group and sport specialization increase.

A consequence to subjective team picking may lead to an increased number of younger athletes not participating in sport programs. This concern may forever affect their ability to compete in athletics later in life. To determine how the RAE may impact athletes, Helsen et al. (1998) conducted a study to quantify the influence of relative age on success and dropout in male soccer players in Belgium. This study concluded that players born late in the selection year tended to drop out as early as twelve years old. The authors compared four groups ranging from professional players, youth players selected on the national team, youth players that transferred to a first division youth team and regular youth leagues. In Belgium, the selection year is from August to July. This study found an overrepresentation of young athletes and professional players born in August, September and October and an underrepresentation for May, June and July. The data suggests that relatively younger players drop out at a higher rate than older.

In addition, a study completed by Williams (2010) on the RAE in youth soccer in the FIFA U17 World Cup competitions showed the distribution of birth months was highly skewed towards the first month of the year. This study suggested that as U17 players are approaching physical maturity, RAE was consistent with the idea that relatively younger players drop out and may no longer be available for selection. Furthermore, Williams (2010) went on to identify unintended consequences that are a result of this RAE. One example noted was that many younger players may possess potential for developing into outstanding athletes, but drop out before their potential is known. Another case is that younger players are constantly faced with a natural bias that will never be able to overcome. These athletes may decide to participate in other sports or activities that do not consider subjective measures. In all, Williams (2010) clarifies that the effects of RAE are negative in that potential high-level players may be lost and/or relatively younger players may not develop to their full potential. This is also pointed out by Jimenez et al. (2008) who states that the current identification process allows for age bias and results in wasted potential. With this in mind, coaches that select players based on age and talent should be encouraged to also take the athlete's potential into perspective.

Statement of the Problem

Given that the RAE is an established phenomenon in many youth sports, one must wonder if it persists into higher level, adult competition and is consistent across team and individual sports as well as sports where physical size is more or less important for success. Olympic competition is a high level, senior sporting event that involves a vast array of sports, teams and athletes. Using Olympic competition, the existence and diversity of the RAE could be examined. In addition, this source of information could identify influences on the RAE such as player gender, type of sport and success. To address these questions, data sets will be obtained from the 2012 London and 2016 Rio de Janeiro

Summer Olympic games that will then be analyzed to determine if RAE continues into this elite, senior level of sport.

Research Questions

The questions addressed in this investigation are:

- Does the RAE exist in Summer Olympic competition?
- Are there differences in the RAE between male and female athletes?
- Is the RAE present in both medalists and non-medalists?
- Does the RAE differ between team and individual sport competition?

Limitations and Delimitations

This study is delimited by the following constraints imposed by the investigator:

- Only data from the 2012 and 2016 Summer Olympic Games are included.
- Several specific sports were collapsed into single groupings (e.g., swimming and diving were defined as "aquatics").
- The variables used to describe the RAE provide a reasonable estimate of the phenomenon.

This study was limited by the following constraints beyond the investigators control.

- The birth dates and other data included in the databases are accurate.
- All athletes participating in the 2012 and 2016 Summer Olympics are included in the databases.

CHAPTER 2

LITERATURE REVIEW

Relative Age Effect in Sport

A review of the literature shows that that the RAE persists in today's athletic world. Table 1 summarizes multiple studies examining the RAE across a number of sports and age groups. As can be seen, The RAE is most often seen in youth sports, particularly team sports such as soccer, ice hockey and volleyball. In individual sports such as gymnastics and boxing, it is less prominent. Although there are few studies on sports at the adult and professional level, it appears that the RAE is less often observed. Nevertheless, the summary shown in Table 1 indicates the existence of a RAE with more players being born in the early part of the completive year versus late.

One example of the RAE is highlighted by Diaz del Campo et al. (2010) studying youth soccer players from Spain. This study has a sample of different youth ages over three different experience level elite soccer teams from different seasons with a total of 4,211 subjects. The expected value was the Spanish population born within 1986-1997 interval. The researchers analyzed the data chronologically according to birthdates. To compare birthdates, January and December were the first and last months of the age-group year. Birth months and quarters were calculated based off the data collected and the Spanish population. Chi-square tests and the Kruskal-Wallis tests were utilized to determine if there was an unequal month distribution for the youth athletes selected. These tests also determined if the results varied across the teams analyzed. Next, the Mann-Whitney U-test was used to identify which groups demonstrated a RAE by comparing the athletes to the expected population. The researchers found that the birth-date distribution of the soccer

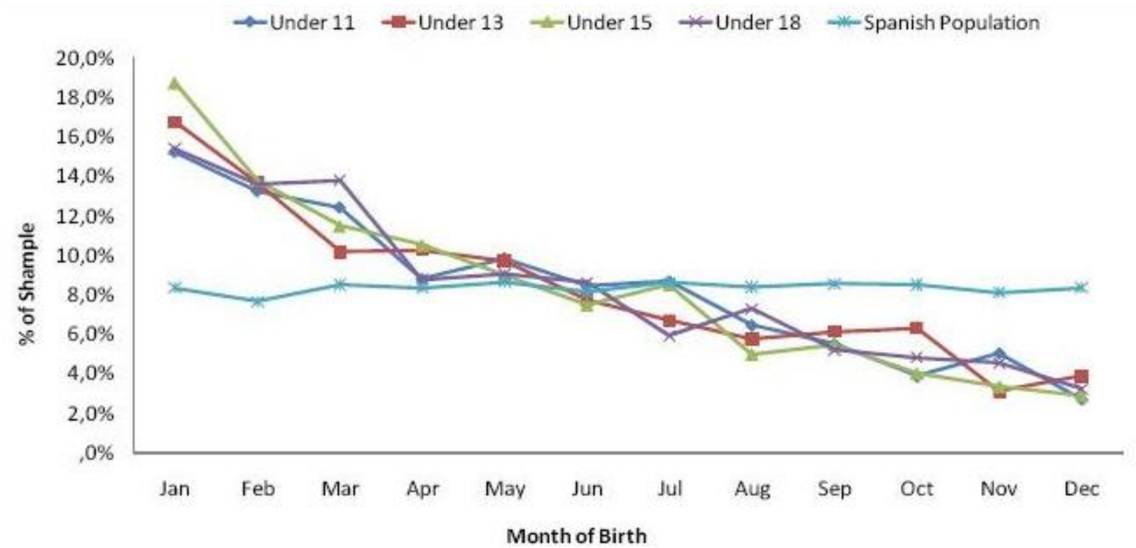
Table 1. Summary of Studies Examining the RAE in Youth and Adult Sports.

SPORT	AGE	GENDER	RAE	STUDIES
Soccer	Youth	Male	Yes	Williams, 2010; Helsen et al, 2000; Souza et al , 2015; Hirose, 2009; Barnsley et al., 1992; Helsen et al., 1998; Brewer et a.l, 1995; Musch, 1998; Vincent et al., 2006; Musch & Grondin, 2001; Vincent et al., 2004; Baxter-Jones et al., 1995; Baumler, 1996; Del Campo et al., 2010; Helsen et al., 2005
Soccer	Professional	Male	Yes	Cobley et al., 2008; Verhulst, 1992; Dudink, 1994; Vaeyens et al., 2004;
Soccer	Youth	Female	Yes	Vincent et al, 2006; Delorme et al., 2009;
Soccer	Elite Junior	Male	Yes	Baxter-Jones et al, 1995;Brewer et al., 1992
Soccer	Professional	Male	No	Baumler, 1996; Ford et al., 2011
Chess	Youth	Female	Yes	Breznik et al., 2016; Helsen et al., 2014
Chess	Youth	Male	Yes	Breznik et al., 2016; Helsen et al., 2014
Volleyball	Youth	Male	Yes	Campos et al., 2016; Grondin et al., 1984
Volleyball	Youth	Female	Yes, very weak	Grondin et al., 1984
Basketball	Youth	Male	Yes	Arrieta et al., 2016
Basketball	Youth	Female	No	Arrieta et al., 2016
Basketball	Professional	Female	No	Werneck et al., 2016;
Basketball	Professional	Male	No	Werneck et al., 2016;
Judo	Professional	Male	Yes	Albuquerque et al., 2015;
Judo	Professional	Female	Yes	Albuquerque et al., 2015;
Boxing	Amateur	Female	No	Delorme et al., 2014;
Boxing	Amateur	Male	No	Delorme et al., 2014;
Ice Hockey	Youth	Male	Yes	Sherar et al., 2007; Boucher et al., 1994
Ice Hockey	Elite Junior	Male	Yes	Barnsley et al., 1988;
Ice Hockey	Professional	Male	No	Ford et al., 2011

Swimming	Elite Junior	Male	Depends on age and categories	Baxter-Jones et al., 1995;
Taekwondo	Elite	Female	No	Albuquerque et al., 2012
Taekwondo	Elite	Male	No	Albuquerque et al., 2012
Tennis	Elite Junior	Male	Yes	Baxter-Jones et al., 1995; Dundink, 1994; Edgar & O'Donoghue, 2005
Alpine Skiing	Elite	Female	No	Bjerke et al., 2016
Alpine Skiing	Elite	Male	No	Bjerke et a.l, 2016
Gymnastics	Elite Junior	Male	No	Baxter-Jones, et al., 1995
American Football	Professional	Male	No	Stanaway & Hines, 1995;
Australian Football League	Professional	Male	No	Ford et al., 2011
Referees	Obtain certification	Male	No	Delorme et al., 2011
Baseball	Youth	Male	No	Thompson et al., 1992
Baseball	Professional	Male	No	Grondin et al., 2000;
Baseball	Professional	Male	Yes	Grondin et al., 2000; Thompson et al., 1991; Stanaway et al., 1995
Athletics - Sprinting	Youth	Male	Yes	Romann & Cobley, 2015
Taekwondo	Olympic	Male	No	Albuquerque et al., 2012
Taekwondo	Olympic	Female	No	Albuquerque et al., 2012
Alpine Skiing	Elite	Male	Yes – speed events, No-technical events	Bjerke et al., 2016
Alpine Skiing	Elite	Female	No – all events	Bjerke et al., 2016
Olympic Sports	Olympians	Male	Yes	O'Neill et al., 2016
Olympic Sports	Olympians	Female	Yes	O'Neill et al., 2016

players differed from the Spanish Population. In conclusion, the birth-date distribution of players decreased from January to December, with the biggest variations in comparison to the Spanish population occurring at the start of the year (January, February and March) and the end of the year (September, October, November and December). In addition, by means of the Kruskal-Wallis Test, it was noted that distributions of births groups significantly differed amongst them. The results of this article revealed there was a biased distribution of the young athletes that displayed RAE as seen in Figure 1.

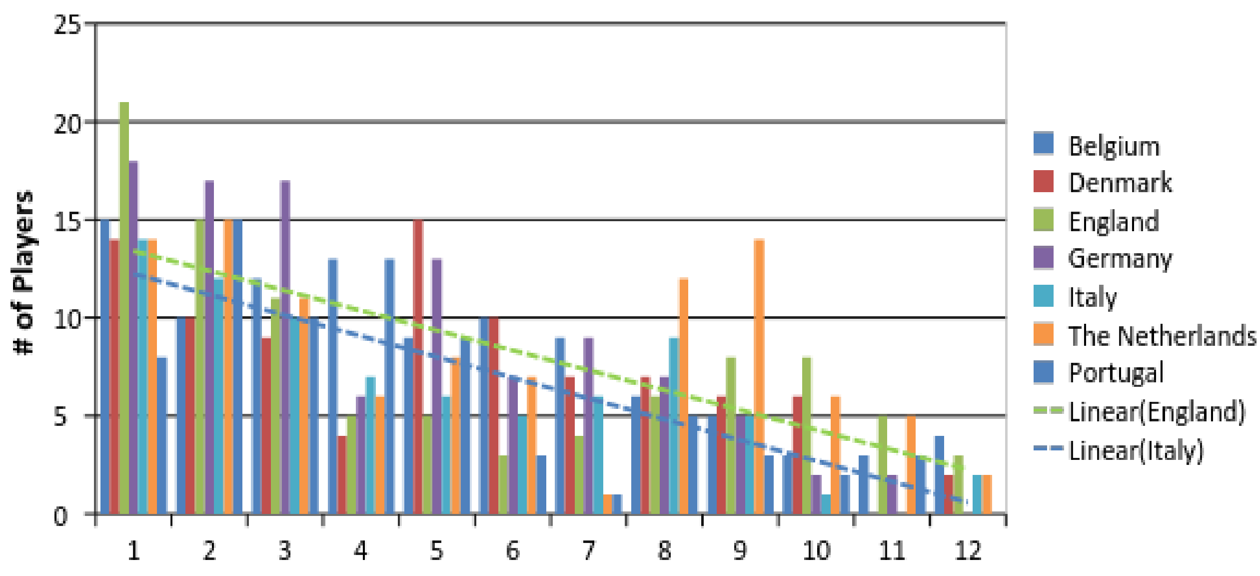
Figure 1. Age Distribution of Youth Soccer Players in Spain. Graph obtained from Diaz del Campo et al. (2010)



Helsen et al. (2005) conducted a study to determine RAE in youth soccer across Europe. The researchers analyzed birth date distributions of youth soccer teams that participated in the Union des Associations Europeennes de Football (UAEF), the European governing soccer body. The countries represented in this study were Belgium, Denmark, England, France, Germany, Italy, The Netherlands, Portugal, Spain and Sweden for a total of 2,175 subjects. The birth dates were then broken up into under-12 (U-12), U-14 ,U-15, U-16, U-17, U-18, U-21, women's U-18 age groups and the Meridian Cup, which is the championship organized by the UEFA. The only data that could be obtained from France, Spain and Germany were the players in groups U-16 and U-18 participating in the UEFA competitions. The age samples were grouped by birth month for their appropriate category starting with January and ending with December. The authors stated that Kolmogorov-

Smirnov one-sample tests were used to assess differences between the observed and the expected birth date distributions. In addition, Helsen et al. (2005) utilized regression analysis to examine the relationship between the number of players per age category and the corresponding birth month. As seen in figure 2, the birth date distributions for U-15, U-16, U-17 and U-18 represent the ten European countries national teams. Significant results were found for Belgium, Denmark, England, France, Germany, Italy, The Netherlands, Portugal, Spain and Sweden. In the UEFA tournament, significant results were shown in the U-16, U-18 and the Meridian Cup teams. The results were not significant for the men's U-21 or the women's U-18 team. In addition, the U-12 and U-14 teams both had significance in that athlete's birth dates were overrepresented in the first three months of the year and underrepresented in the last three months of the year. In conclusion, this research found significant results to conclude a relative age effect in many of the youth age groups that participated in this study.

Figure 2. Birth Date Distribution of the U15, U16, U17 and U18 Selections per Country (except France, Spain and Sweden) Data taken from tables presented in cited paper (Helsen et al., 2005).



A third study examined the relative age effect and performance in the U16, U18 and U20 European basketball championships. Arrieta et al. (2016) had 2395 male and female athletes that participated in this study. To determine a relative age effect, the athletes were divided into four quarters by their birth date. More specifically, quarter one represented players born on January 1 through March 31, quarter two April 1 through June 30, quarter three July 1 through September 31 and quarter four October 1 through December 31. Next, the relative age effect was taken into perspective. This was calculated by "the youngest player of the tournament received a value of zero in each cohort. Each player received the value (relative age) of the number of days that had elapsed from the date of birth of the youngest player to his/her birth date. This means that the oldest players obtained the highest values and the youngest the lowest." (Helsen et al. 2005). The results of this study found a significant result, displaying RAE, in both male and female U16, U18 and male U20 teams. The female U20 team did not result in a significant difference. This group showed a more even birth date distribution. The results can be seen in figures three and four.

Figure 3. Percentage of Female European Basketball Players Divided by Birthdates into Quarters (Graph created from data table in cited paper, Arrieta et al., 2016).

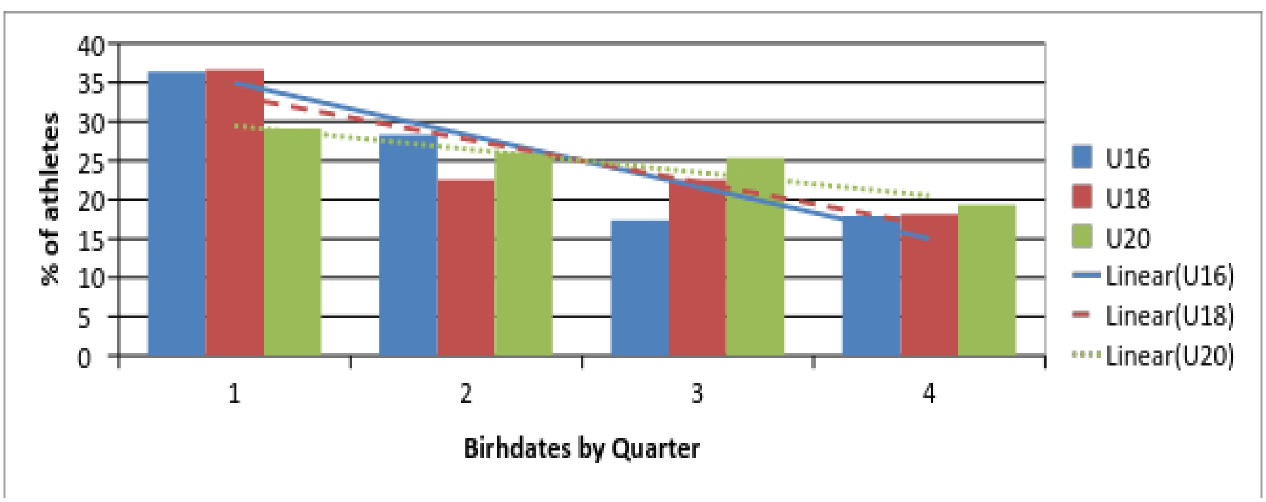
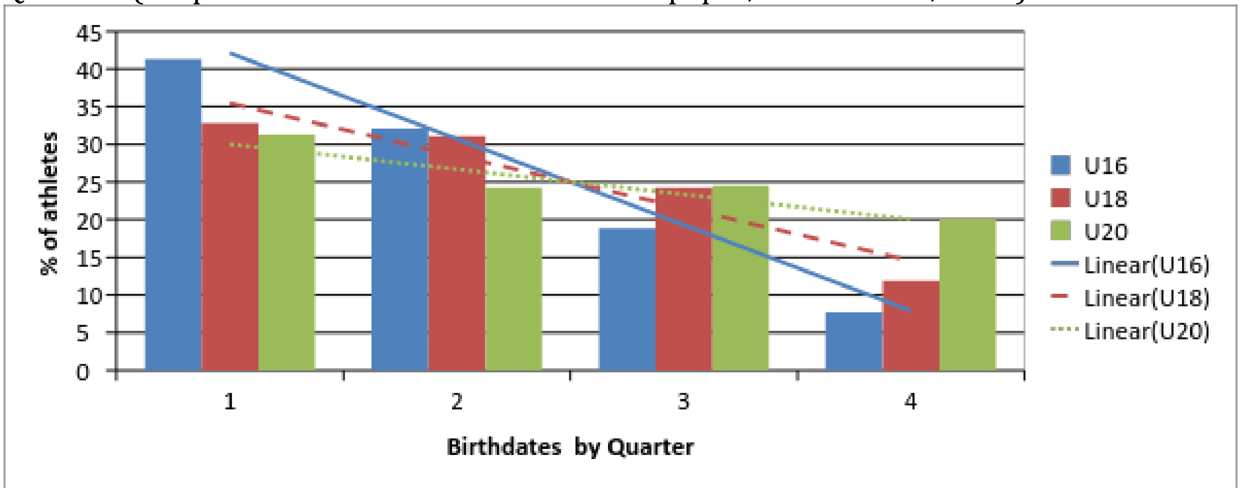


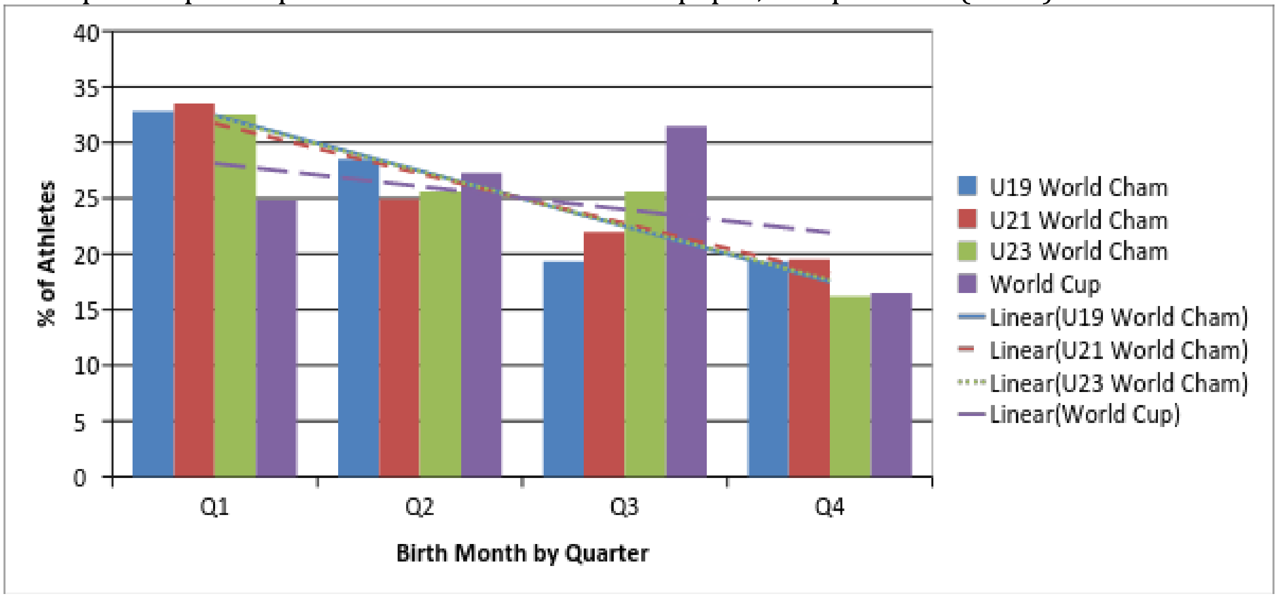
Figure 4. Percentage of Male European Basketball Players Divided by Birthdates into Quarters (Graph created from data table in cited paper, Arrieta et al., 2016).



Aside from youth team sports, this phenomenon is also being researched in weight class sports, which are more of an individual sport than a team competition. Delorme (2014) investigated whether weight categories prevent young athletes from being exposed to a relative age effect. This study utilized birth dates of 708 female and 5,440 male French amateur boxers during one year time period. The data were obtained from the National Institute of Statistics and Economics Studies. The boxers were separated by birthdate into four age categories 12–13, 14– 15, 16–17 and 18–18+. While 12–13 female boxers (n = 19) were excluded from this study because of a small sample size. This study used the quarter method for distributing the birth dates in each category, for a refresher, quarter one represents January 1 through March 31. Odds ratio was used to determine the effect size if any RAEs occurred. Odds ratio and 95% confidence intervals were calculated by comparing each birth quarter to quarter four (October through December). This study resulted in uneven distributions, but no significant evidence of a relative age effect in any of the age categories for the French amateur female and male boxers. In addition, the 18-18+ boxers, female and male boxers, displayed an inverse relationship of RAE, meaning there were more boxers being represented at the end of the calendar year than at the beginning.

Campos et al. (2016) investigated the relative age effect in male volleyball world championships. The objective was to examine the existence of a relative age effect in male volleyball to identify this influence in four different age categories. Data was obtained from four men's volleyball championships (U19, U21, and U23) and Men's World Cup in 2015, for a total of 1,105 athletes. All of the data utilized was obtained from the official website of the International Volleyball Federation (FIVB). The cut off date for the FIVB was January 1. The athletes were categorized by birth date into birth quarters. Quarter 1 was January through March and so forth through the year ending with quarter four, which represented October through December. This study identified RAE when a significant difference was found between expected number of players born during a quarter and the observed number of players that were born within the quarters. In addition, a chi-square goodness-of-fit test to determine whether the observed distribution differed from the expected values. This study resulted in significant findings displaying RAE in the U19, U21 and U23 teams, but was not present in the world cup volleyball championships as shown in figure five.

Figure 5. Percentage of Volleyball Athletes at Different Age Group Volleyball Championships. Graph created from data in cited paper, Campos et al. (2016).

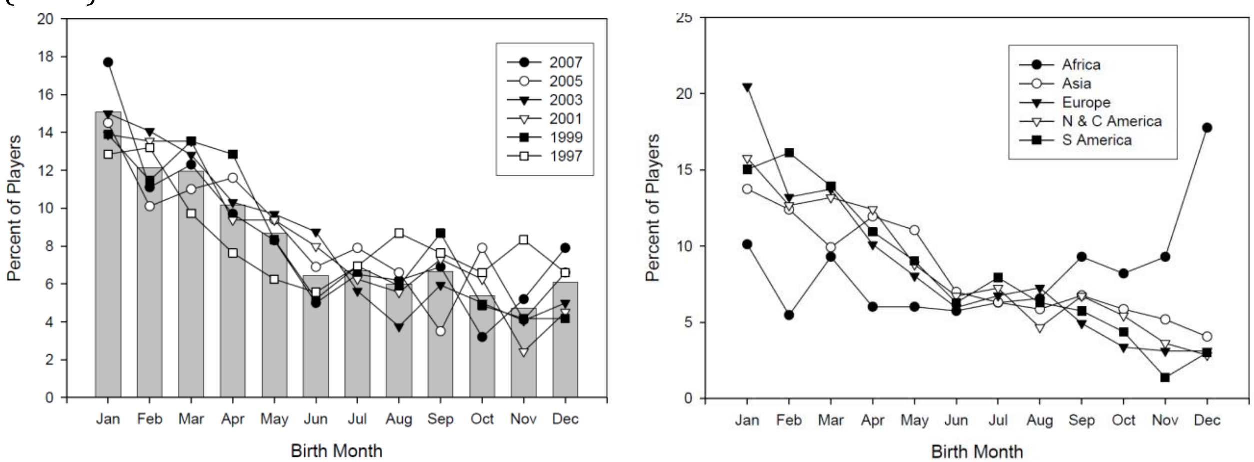


Breznick et al. (2016) researched the relative age effect in mind game of elite chess. The age categories for chess competition were below 10, 12,15,16,18 and 20 years of age. The cutoff date was January 1. Data for this study was obtained from the official international chess federation (FIDE) website. The birthdates of the participants were categorized by birth quarter in a calendar year. To test for significant differences, Chi square tests were used to compare the frequency of the appearance of the birth dates of top chess players across the birth quarters to the expected values. Significant differences in gender and comparing relative age of winners between age categories were statistically tested with the students' *t* test. This study resulted in significant differences in every age category, girls (under the age of 20), boys (under the age of 20), females and males. In conclusion, relative age effect can be displayed in an individual sport like young competitive chess. Another study done by Helsen et al. (2014) found similar results studying the relative age effect in youth chess competition and suggested that these results may be because of 'cognitive maturity and differences in skill acquisition among youth.'

To test different aspects of relative age effect, Norikazu Hirose (2009) conducted a study looking at the relationships among birth month distribution, skeletal age and anthropometric characteristics in adolescent elite soccer players. There were 322 male adolescent elite soccer players between 1997 and 2000 that participated in this study. The average age of these players were 12.2 ±1.5 years and the range was 9.1-15.0 years. The athletes were divided into six age categories (U10-U15) according to chronological age. Measurements that were taken were birthday, height, body mass and skeletal age. Biological maturation was calculated using skeletal age minus chronological age. This study resulted in a significant shift in the birth month category where 37.9-58.8% of athletes were from quarter 1 of the year (January through March) and 3.2-13.5% were represented from quarter 4 (October through December). This study reported that there were no significant differences in the age groups according to skeletal age and maturation difference. The comparison between birth month and height (cm) displayed significant results in the U11, U12, U13, and U14 teams. These results, especially U14, showed that the quarter one players were significantly taller than the quarter four players were. While measuring body mass (kg), the results concluded that the U11 and U14 categories weighed significantly more in the quarter one group than the quarter four group. This study concluded that their results showed a clear bias in selecting the older players born in quarter one over the younger players born in quarter four and that some bias may be due to differences in individual skeletal age and body size.

Williams (2010) focused on the RAE in elite youth soccer. He examined data from the U17 FIFA World Cup. He examined 1985 players and found that there was a large RAE in this competition (Figure 6). This relationship held for the most and least successful teams. There were no differences in the both month distributions between the top four teams in the competition (most successful) and the bottom four (least successful). Interestingly, the RAE held for all geographical regions of the world. The only exception being the African nations. In these countries, there was a "reverse" RAE with the birth month distribution being skewed towards the latter part of the year (Figure 6).

Figure 6. The RAE in the U17 FIFA World Cup. Shown are Birth Month Distributions for the entire study (left) and Geographical Regions of the World (right) Data from Williams (2010).



Based on a review of the research literature, it is clear that the RAE exists in multiple sports and across a wide range of youth competitions. However, it is not clear to what extent it exists in senior, elite level competition. The available research is mixed. For example, Thompson et al. (1991) and Barnesly and Thompson (1988) suggests that it does exist in Major League Baseball and the NHL. However, Ford and Williams (2011) found no such bias when looking at multiple professional sports. Further, it is not clear if the RAE varies between genders, team and individual sports, and the most and least successful athletes. Thus, the research question addressed in this investigation is, does the RAE exist in senior level competition? Specifically, does it exist in the Olympic sport athletes? An answer to this question could provide insight into the development of youth sports participants. Since advanced training and completion are allowed for "elite" teams and players, selection to the developmental programs should be based on the quality and potential of the athletes. If selections are biased due to the relative age of the players, then some talented, well deserving players born in the latter part of the year may be overlooked. On the other hand, less deserving athletes may be selected based on physical maturity resulting from being relatively older that the other participants.

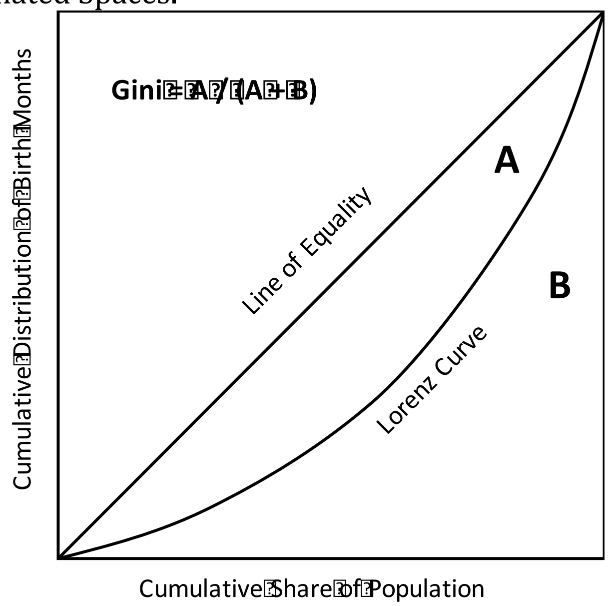
Analyzing the Relative Age Effect

Analyzing and quantifying the RAE has proven to be challenging. Investigators have used several statistical techniques. Williams (2010) used the traditional method of calculating X^2 of the observed birth month distribution against an expected ($\frac{1}{12}$ or 8.33% per month). However, this statistic gives the research information about the extent to which the actual and expected distributions differ. It does not provide information about direction or identify a systematic pattern of difference. In addition, Delorme and Champely (2015) suggest that the risk of a type I error using X^2 may increase with sample size. Other approaches would include linear or non-linear regression or curve fitting (Romann & Cobley, 2015). This may provide information about direction and extent of the RAE, but there is no consensus on the pattern of the RAE. In some cases, it appears linear and in others, it resembles an exponential decay.

The Gini coefficient is a statistical measure of the degree of variation or inequality in the distribution of values (as cited in Gini, 1921). It is often used in economics to analyze income inequality. It is preferable to the chi-square analysis because it compares the actual values to a consistent or equal distribution across group (or months). It is calculated using the Lorenz curve (Lorenz, 1905), which plots the cumulative proportion of each group versus the proportion total population (Figure 7). The Gini coefficient is the ratio of the area that lies between the line of equality and the Lorenz curve and the total area under the line of equality. A ratio 0 represents perfect equality whereas 1 represents total inequality. The Gini coefficient (G) can be approximated using the following equation where Y equals the fraction of players born each month (0 to 1) and X is fraction of the birth year (the sum, X_k - X_{k-1} remains constant as $\frac{1}{12}$ or 0.0833) and *k* represents the month (1 to 12).

$$G = 1 - \sum_{k=1}^{12} (X_k - X_{k-1})(Y_k + Y_{k-1})$$

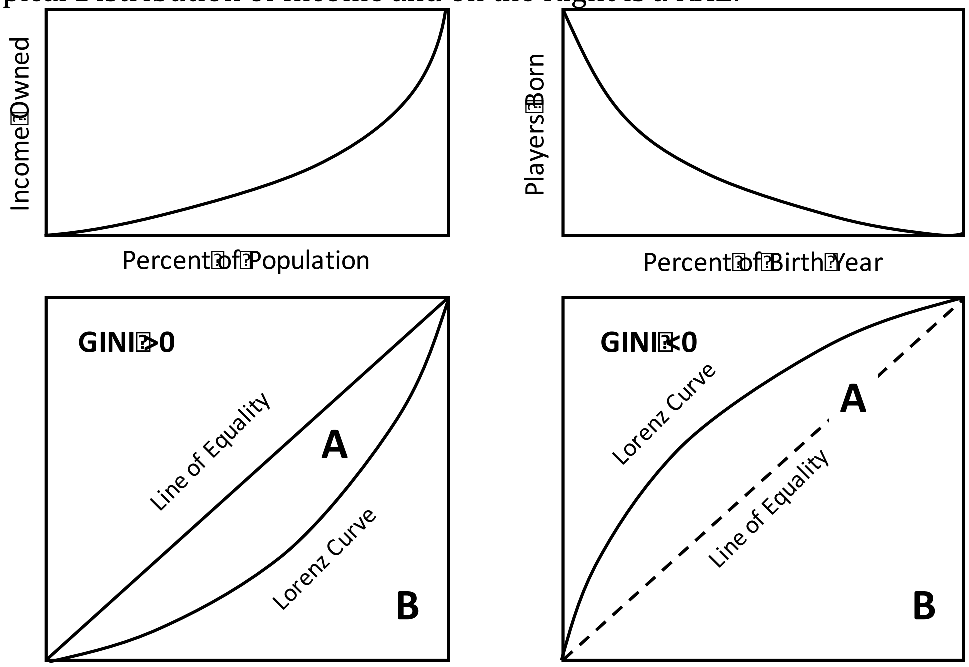
Figure 7. Graphical Representation of the Gini Coefficient (or index). A and B represent the Areas of the Designated Spaces.



The Gini index can by supplemented by X^2 comparison of the expected birth month distribution (8.3% per month) and the actual distribution. Also, the slope of the regression of birth month and fraction of players born each month may be computed to provide better directionality.

The Gini coefficient was originally used to examine income inequality. It is assumed to range from 0 to 1. In general, income is distribution is skewed towards the upper fraction of the population. This results in a Lorenz curve that is situated below the line of equality and a positive Gini coefficient. In the case of the RAE, player birth months are skewed towards the early part of the birth year. This places the Lorenz curve above the line of equality and results in a negative Gini coefficient (Figure 8).

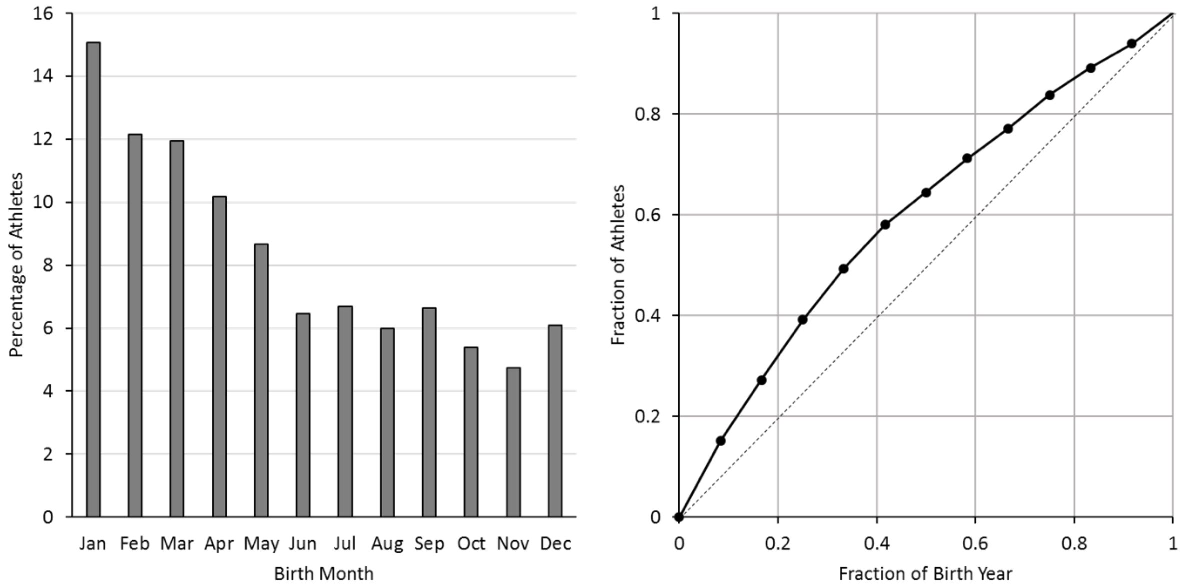
Figure 8. Lorenz Curves and Gini Coefficients for Two Different Skewed Distributions. The Left is a Typical Distribution of Income and on the Right is a RAE.



For a RAE, where the distribution of birth months is skewed, the Gini coefficient might prove more informative than other methods of analysis. For example, using the data of Williams (2010), a negative Gini coefficient is shown (Figure 9). The subjects in this study represented players of the men's U17 World Cup soccer competition. For these data, the Gini index equals -0.198.

Thus, the Gini coefficient may be a more appropriate method to identify the RAE. By combining it with other methods of analyzing distribution disparity, quantification of the RAE may be more accurate and informative.

Figure 9. Birth Month Distributions (left) and Lorenz Curves (right) for U17 World Cup Players. Data obtained from Williams (2010).



Summary

Based on the review of literature, it is clear that the RAE exists in a number of youth sports and other competitions. It is possible that the RAE is greater in males than in females. However, less is known regarding the RAE at elite, senior level competitions. This includes the Olympics. Second, there is debate as to the method of determining a RAE. Past studies have used the goodness of fit approach and X^2. However, other approaches such as the Gini coefficient may be more appropriate.

A better understanding of the existence of a RAE at Olympic level competition and establishment of a consistent method of analysis would play key roles in determining the role of the RAE in talent identification and player development. These data could also for the basis for correcting existing methods of talent identification leading to a reduced RAE.

CHAPTER 3

METHODS

Sample

The Olympic athlete's name, birthdate, place of birth, sport and gender of the 2012 London and 2016 Rio de Janeiro Olympics were obtained from an open-access, internet source (www.kaggle.com and www.theguardian.com). Given that these databases are open-access and freely available to the general public, this research is exempt from Institutional Review Board approval as stated in 45-CFR-46.

In all, there are over 20,000 Olympians representing nearly 30 sports identified in the two data sets. The datasets were examined for obvious errors in the data such as clearly erroneous birth date as well as omissions of key data such as body mass, medals earned, etc. In these cases, the athlete was excluded from the analysis. Sports that are not represented in both Olympic competitions were excluded.

The variables included in the database were the athlete's birth date (date, month and year), gender, sport and medals awarded. Also included were height and body mass. Specific sports were classified as individual or team competition. Individual competition is defined as a sport where the athlete competes without the aid of teammate. For example, gymnastics awards a team medal. However, the gymnasts compete as individuals. Whereas basketball requires teammates be present during competition. It is important to point out that some sports classified as individual have team competitions. For example, relay races in athletics and aquatics are four-person events. Also, rowing is an individual sport but some events such as the "eights" have teams within a single boat. Sports were also categorized whether or not they have weight classifications for competition. These sports included wrestling, boxing, judo, taekwondo and weightlifting. Some specific sports were combined such as swimming and diving that were included in “aquatics”. Also track cycling, road cycling and mountain biking were combined into a “cycling” category.

Data Analysis

To understand the effect of RAE, all athletes were coded into categories including birth date, specific sport, sport type (team or individual), weight class, gender (male or female), and medals earned (0 or ≥1).

Given the extremely large size of the final data set, birth month distribution was used to identify the RAE (as opposed to birth quarter). As most international competitions use calendar year as the competitive year, January was considered the first month of the competitive year (1) and December the last (12). Given that birth month of the world population vary by country and region, an equal distribution across months was considered as the expected value ($\frac{1}{12}$ or 8.3% per month).

To identify a RAE, the distribution of birth months must differ from expected and there must be a systematic pattern in that distribution. For this study, Lorenz curves were constructed and the Gini coefficient was used to identify unexpected distributions and their direction using the equation below. The Gini coefficient was further verified by the X^2 statistic. Also, linear regression was used to identify significance directionality. A RAE was identified as a large, negative Gini coefficient and a significant negative slope of the birth date distribution.

$$G = 1 - \sum_{k=1}^{12} (X_k - X_{k-1})(Y_k + Y_{k-1})$$

A RAE was determined as a Gini coefficient less than 0 and a statistically significant negative slope values. Unfortunately, there are no statistical tests available to evaluate the magnitude of the Gini coefficient. Thus, interpretation of this metric must be qualitative. However, it is expected that noticeable negative Gini coefficient values are associated with significant X^2 value.

Statistical Analysis

Differences between males and females, medal winners and non-winners, team and individual sports, as well as weight class and non-weight class sports were determined using t-tests. For this, the Gini coefficient was calculated for the sport. Then the sports within each category was averaged. The category averages were then compared using statistical procedures. Significance was established at the 0.05 level of confidence.

CHAPTER 4

RESULTS

Entire Cohort

There were a total of 21,890 athletes included in the final dataset. A comparison of the two Olympiads is shown in Table 2. There were eleven percent more athletes in the Rio Olympics that in the London games. There were no significant differences between any of the descriptive variables. For example, the average birthdate of the two groups were similar. Shooting and golf had athletes who were the oldest with the mean age of thirty years. The youngest sport was gymnastics with a mean age of twenty-two years.

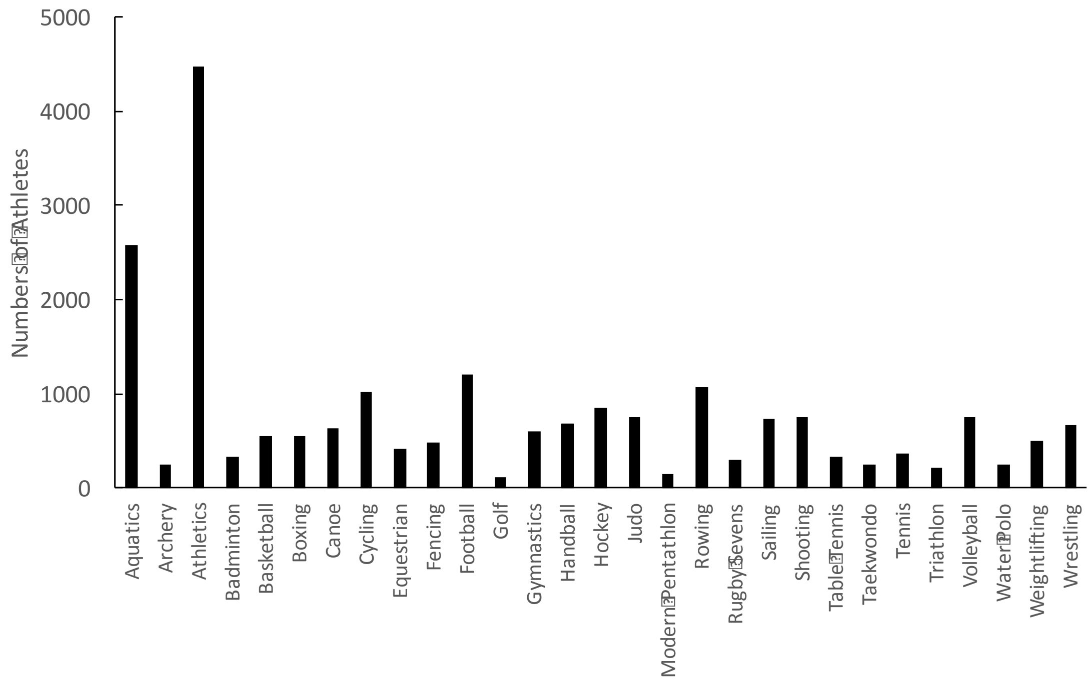
Table 2. Descriptive Data of the Athletes competing in the 2012 and 2016 Olympics.

Variable	London 2012	Rio 2016	Total
n	10353	11537	21890
Age (y)	26.1 ± 5.4	27.2 ± 5.4	27.3 ± 4.9
Body Mass (kg)	72.9 ± 0.2	72.1 ± 0.2	72.6 ± 13.4
Height (cm)	176.9 ± 11.3	176.6 ± 11.3	176.6 ± 9.6
Birthdate	June 24	June 10	June 16

The numbers of males and females competing in each sport classification are shown in Figure 10. Athletics, which is track and field events combined, and aquatics, which is swimming sports combined, had the greatest numbers of participants. Combined, they accounted for thirty two percent of the total athletes. The fewest participants were found in golf with a total number of 120 competitors and modern pentathlon with 141 total competitors.

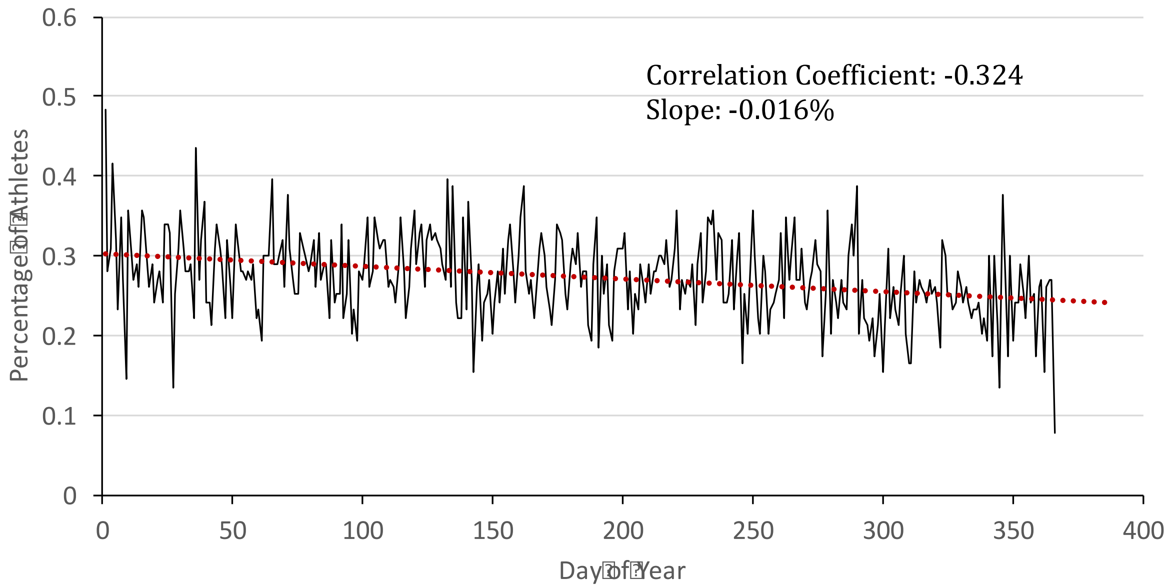
Due to low numbers of athletes in some specific sports, some sports were combined. For example, swimming and diving were collapsed and referred to as "aquatics." In addition, road, trail and track cycling were combined into a "cycling" category. This is consistent with how other sports are typically classified. For example, track (i.e. runners) and field athletes (i.e. throwers) are typically referred to by the group descriptor "athletics."

Figure 10: Number of Athletes Competing in the 2012 and 2016 Summer Olympic games. Shown are both Male and Female Athletes.



The distribution if birthdates for the entire cohort of 2012 and 2016 Olympians is shown in Figure 11. In this figure, January 1 is designated as 1, the first day of the year. All other dates are computed from this value. In the case of leap years, December 31 is designated as 366. For all other years, it is designated at 365. The correlation coefficient for this relationship was -0.324 ($p<.05$) and the slope was -0.016% per day.

Figure 11. The Distribution of Birthdates of 2012 and 2016 Summer Olympic Athletes



The distribution of birth months and Lorenz curve for the entire cohort is shown in Figure 12. As can be seen, there was a tendency for the fraction of athletes born each month to decline as the year progressed. This is reflected in a Lorenz curve that is slightly above the line of equality. For these athletes the Gini coefficient was -0.0324. Both the slope (-.0042) and χ^2 value (113.24) were significant ($p<.05$). Given that the conditions for a RAE are met, these results suggest that there is a RAE, albeit small, for Summer Olympic competition.

The Gini coefficient varied considerably between sports. These values are shown in Figure 13. The largest negative Gini coefficient was seen in the triathlon (see Figure 14). Positive values were seen for boxing, golf, gymnastics and taekwondo. Gini coefficients, r^2 values, slopes and X^2 values for each sport are shown in Table 3. Based on these results, the sports showing noticeable RAE are archery, football, judo, modern pentathlon, and triathlon.

Figure 12. Distribution of Birth Months and Lorenz Curve for the Entire Cohort of 2012 and 2016 Olympians.

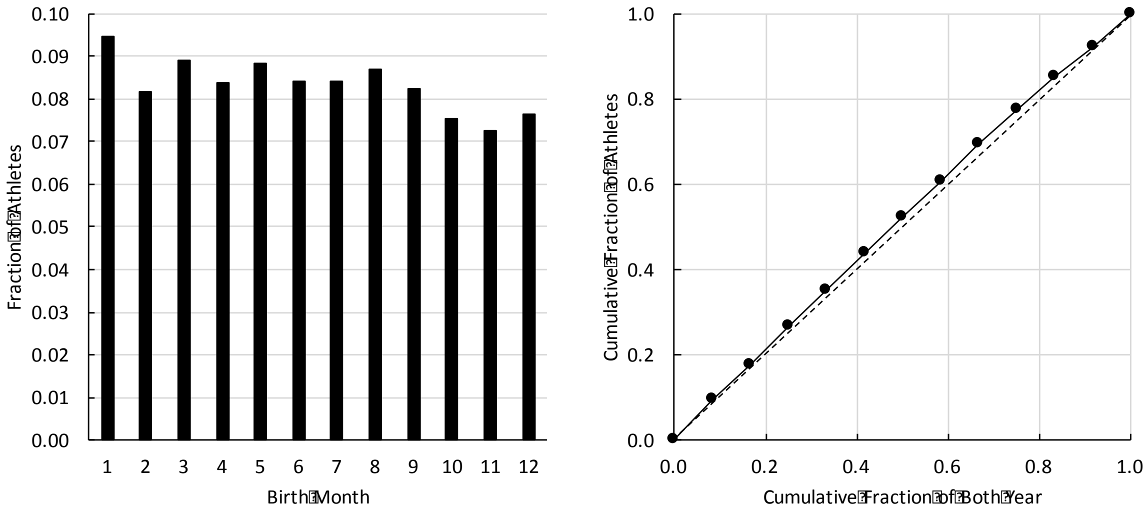


Figure 13. Gini Coefficients for each Sport in the 2012 and 2016 Summer Olympics.

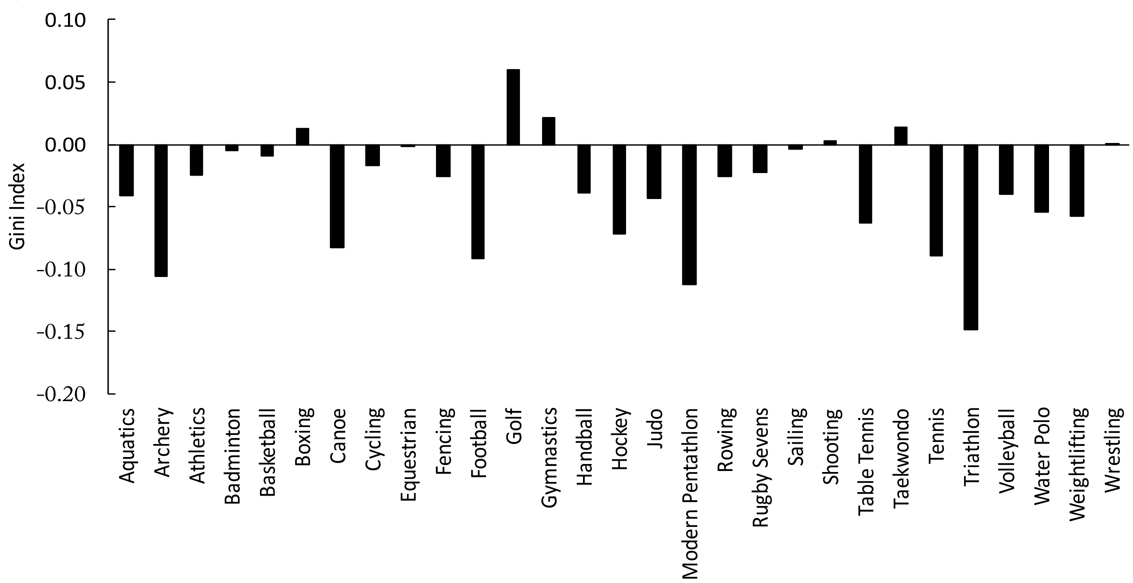
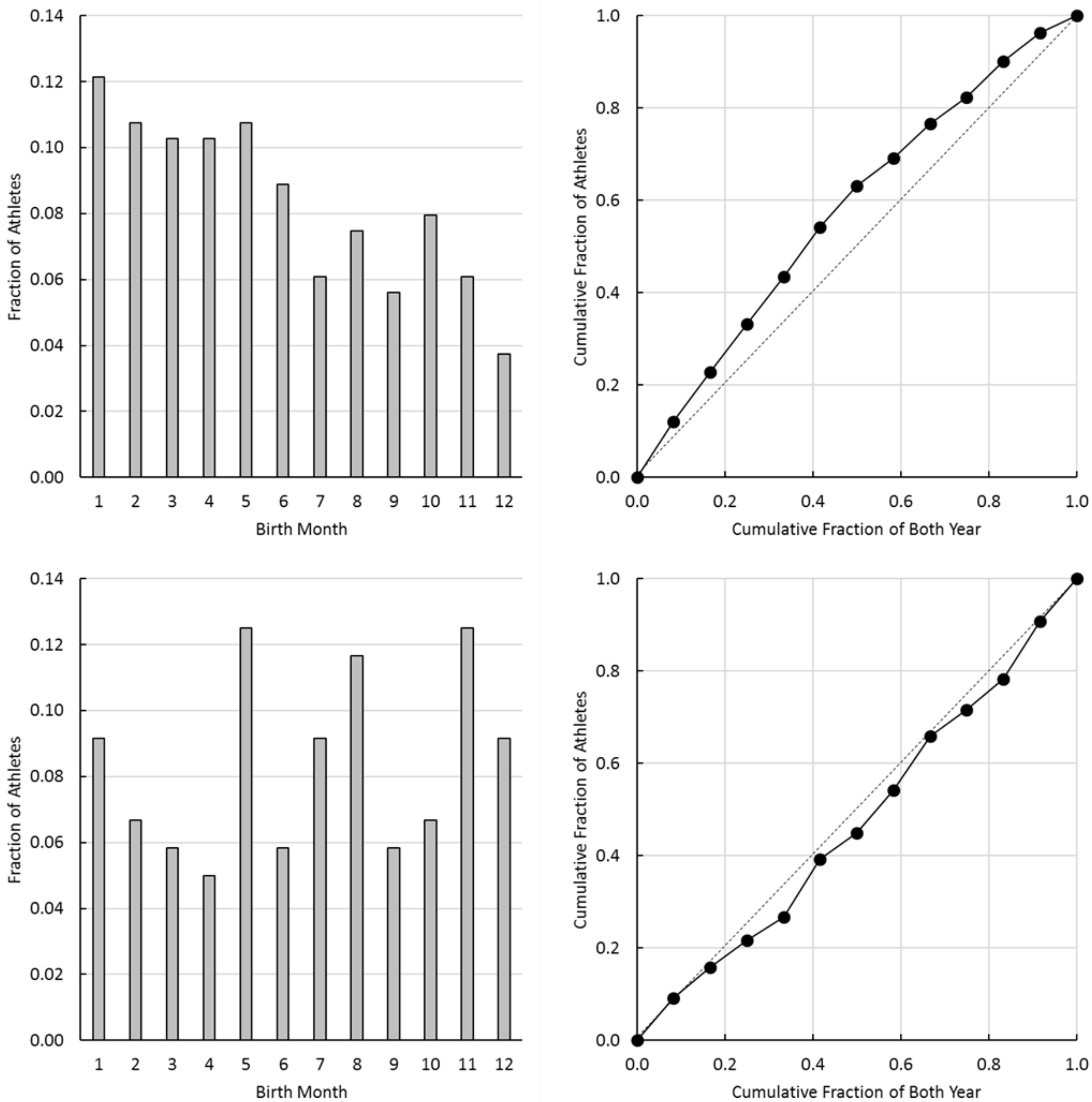


Table 3. Gini Coefficients, r^2, slope and X^2 values for each sport, along with the average for all sports.

SPORT	GINI	r^2	SLOPE	X^2
Aquatics	-0.0428	0.3773	-0.0018	37.92[a]
Archery	-0.1014	0.4690	-0.0043 [a]	16.49
Athletics	-0.0228	0.3226	-0.0010	21.81[a]
Badminton	-0.0148	0.0150	-0.0006	14.90
Basketball	-0.0094	0.0097	-0.0004	15.40
Boxing	0.0075	0.0117	0.0003	8.02
Canoe	-0.0673	0.5495	-0.0028	15.99
Cycling	-0.0232	0.1487	-0.0010	11.07
Equestrian	0.0196	0.0590	0.0008	8.10
Fencing	-0.0113	0.0086	-0.0005	21.53[a]
Football	-0.0988	0.7757[a]	-0.0041[a]	45.88[a]
Golf	0.0528	0.0841	0.0022	12.00
Gymnastics	0.0268	0.1621	0.0011	7.98
Handball	-0.0244	0.1470	-0.0010	8.37
Hockey	-0.0480	0.2541	-0.0020	23.19[a]
Judo	-0.0491	0.5940[a]	-0.0021[a]	9.33
Modern Pentathlon	-0.1200	0.5706[a]	-0.0050[a]	10.74
Rowing	-0.0273	0.2166	-0.0011	11.14
Rugby Sevens	-0.0378	0.1071	-0.0016	12.08
Sailing	-0.0069	0.0068	-0.0003	15.60
Shooting	-0.0011	0.0003	0.0000	8.64
Table Tennis	-0.0696	0.2287	-0.0029	21.67[a]
Taekwondo	0.0079	0.0041	0.0003	11.51
Tennis	-0.0898	0.2594	-0.0038	34.06[a]
Triathlon	-0.1558	0.8284[a]	-0.0065[a]	18.93
Volleyball	-0.0348	0.1510	-0.0015	18.07
Water Polo	-0.0545	0.1568	-0.0023	14.28
Weightlifting	-0.0378	0.1317	-0.0016	16.39
Wrestling	0.0009	0.0001	0.0000	13.13
AVERAGE	**-0.0356[b]**	**0.2293[a]**	**-0.0015[a]**	**16.70[a]**
SEM	**0.0086**	**0.0444**	**0.0004**	**1.69**

[a]Statistically significant from zero ($p<.05$).

Figure 14. Birth month distributions and Lorenz curves for triathlon (top) and golf (bottom) athletes in the 2012 and 2016 Summer Olympics. These sports represent the largest positive and negative Gini coefficients.



Females vs Males

Descriptive data for the males and females included in the study are shown in Table 4. There were 23 percent more male athletes in the sample than female. Not surprisingly, body mass and height were greater in the male athletes versus the females. However, there were no differences between mean ages. Further, the average birthdates for the males and females were nearly identical. The numbers of males and females included in each sport are shown in Figure 15.

Table 4. Descriptive Data for Male and Female athletes in 2012 and 2016 Summer Olympics.

Variable	**Female**	**Male**
n	9,816	12,074
Age (y)	26.1 ± 0.1	27.1 ± 0.1
Body Mass (kg)	62.9 ± 0.1	80.3 ± 0.1[b]
Height (cm)	170.1 ± 0.1	182.2 ± 0.1[b]
Birthdate	June 17	June 16

[b]p<.05 between genders

Figure 15. Number of Male and Female Athletes in the 2012 and 2016 Olympics representing the 29 sports in the entire database.

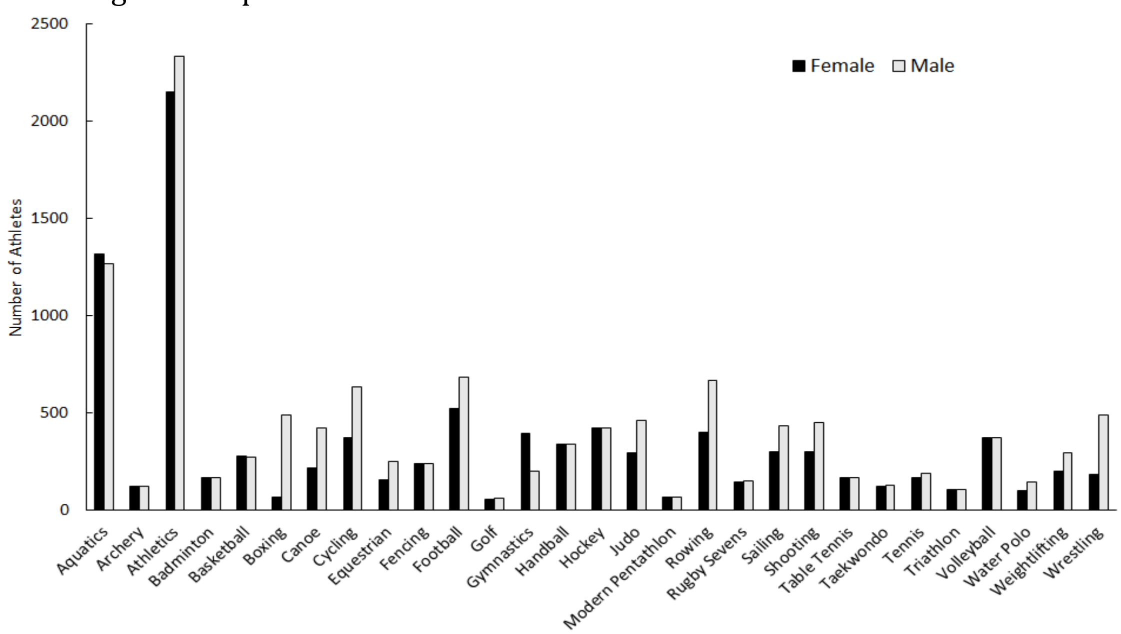


Figure 16 shows the birth month distributions and Lorenz curves for the female and male athletes. As can be seen, there is a tendency for the number of athletes born each month to decline from January to December. In addition, the Lorenz curve is positioned above the line of equality.

Figure 16. Birth Month Distributions and Lorenz Curves for Female (top) and Male (bottom) 2012 and 2016 Summer Olympians.

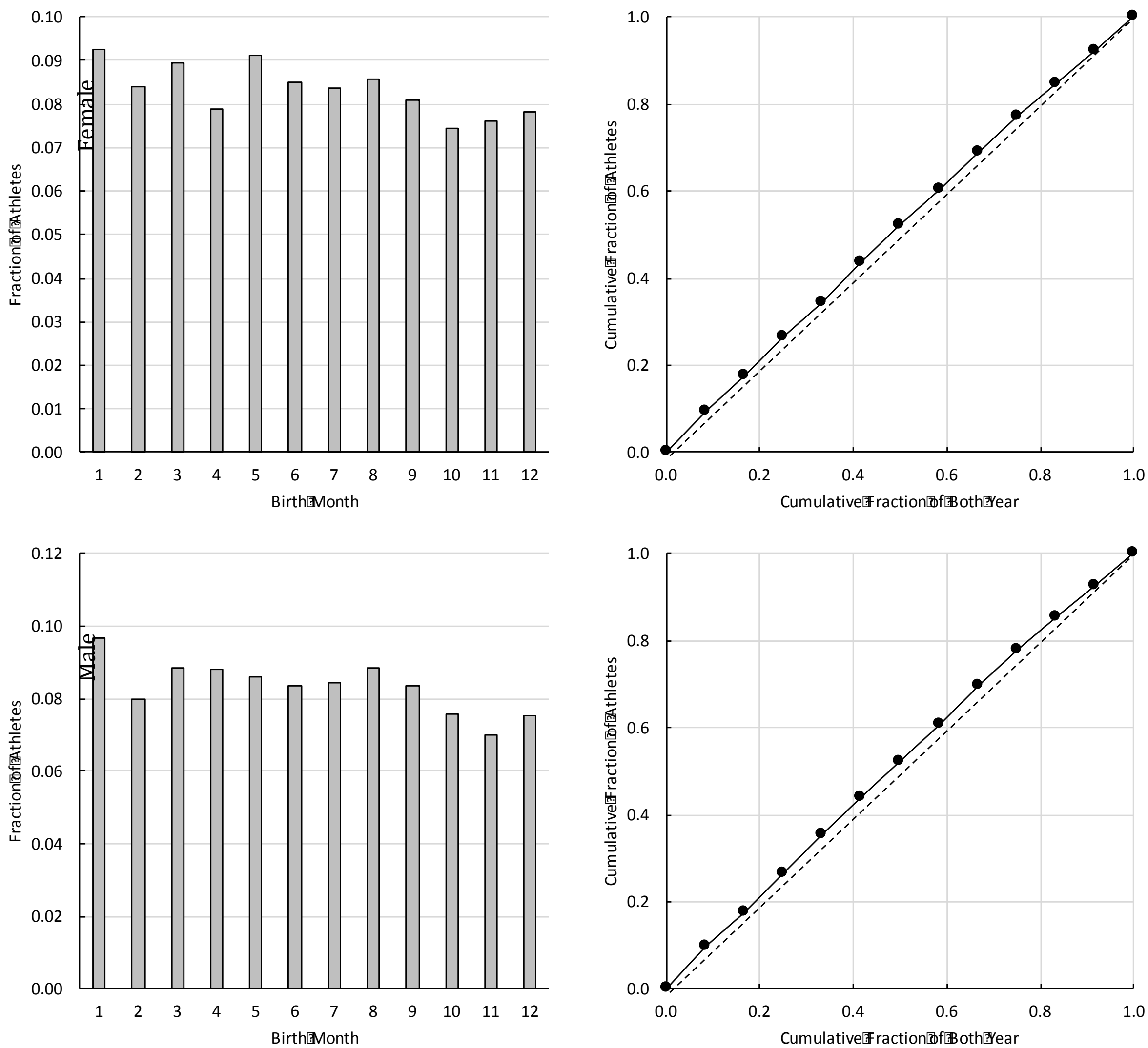


Table 5 shows the GINI coefficient, r^2, slope and X^2 for the males and females. The table shows values for the entire cohort as well as the mean values averaged across sports. For both genders, the Gini coefficients slopes were negative and the p-value, r^2, slope and X^2 values were significant. This suggests that there was a relative age effect for both females and males. However, there were no significant differences between female and male for any of the mean values.

Table 5: RAE Statistics comparing Male and Female 2012 and 2016 Summer Olympians.

Variable	Group	Cohort	Average	SEM
Gini	Female	-0.0279	-0.0321	0.0110
	Male	-0.0360	-0.0345	0.0112
r^2	Female	0.5342[a]	0.1603	0.0308
	Male	0.5650[a]	0.1712	0.0313
Slope	Female	-0.0012[a]	-0.0013	0.0005
	Male	-0.0015[a]	-0.0014	0.0005
X^2	Female	43.31[a]	14.78	1.19
	Male	83.61[a]	16.40	1.57

[a]$p<.05$ versus 0

Team vs Individual Sports

There were 23 sports identified as team and 6 as individual. For the team sports, it was assumed that selection was based on some combination of quantitative and qualitative criteria. Whereas it was assumed that individual sports selections were based on qualitative qualification criteria. This assumption was made because of how qualification is determined in the Olympics. For example, in athletics, times are used for qualifying, but in basketball, popular names in that sport may be selected to represent the country in the

Olympics. Descriptive data for the team and individual sport participants are shown in Table 6.

Table 6. Descriptive Data for Team and Individual Sport Athletes Competing in the 2012 and 2016 Olympics.

Variable	**Team**	**Individual**
n	4,291	17,599
Age (y)	26.0 ± 0.1	27.0 ± 0.1
Body Mass (kg)	76.4 ± 0.1	71.4 ± 0.1[b]
Height (cm)	181.3 ± 0.1	175.6 ± 0.1[b]
Birthdate	June 16	June 16

[b]$p<.05$ between groups.

Birth month distributions and Lorenz curves for team and individual sports are shown in Figure 17. For these two groups, there was also a tendency for birth dates to decline across the year, and for the Lorenz curve to rise above the line of equality to indicate a RAE. However, these tendencies were small, particularly for the individual sports which resulted in no significant differences between individual and team competition in the 2012 and 2016 summer Olympics.

Figure 17. Birth Month Distributions and Lorenz Curves for Individual (top) and Team (bottom) Sports.

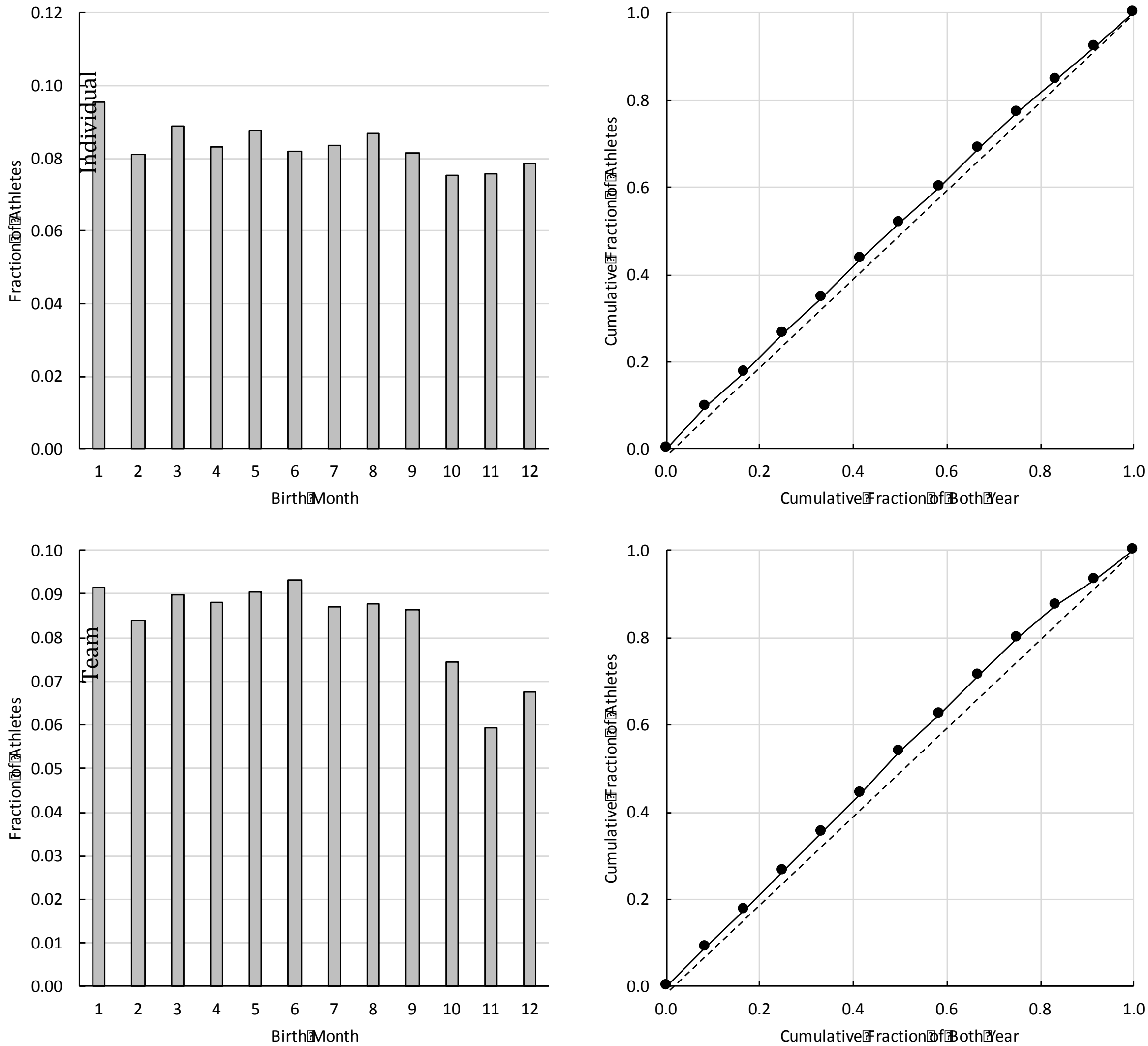


Table 7 shows that cohort and average Gini coefficient, r^2, slope and X^2 values for these groups of athletes. The Gini coefficient and slope values were negative. Also, the r^2, slope and X^2 values were statistically significant. However, there were no significant differences between team and individual sport athletes for any of the average variables.

Table 7: RAE Statistics Comparing Team and Individual Sport in the 2012 and 2016 Summer Olympians.

Variable	Group	Cohort	Average	SEM
Gini	Team	-0.0563	-0.0429	0.0128
	Individual	-0.0284	-0.0333	0.0103
r^2	Team	0.6087[a]	0.2533	0.0932
	Individual	0.5567[a]	0.2194	0.0508
Slope	Team	-0.0024[a]	-0.0018	0.0005
	Individual	-0.0012[a]	-0.0014	0.0004
X^2	Team	61.95[a]	19.42	4.02
	Individual	68.34[a]	14.84	1.59

[a]$p<.05$ versus 0.

Weight Class vs Non-Weight Class Sports

Weight class sports are sports that require weight to be a factor in competition. The competitors weights are utilized to match opponents accordingly. There were five sports that had separate weight class divisions (e.g. wrestling) and 24 identified as non-weight class. Descriptive data for the team and individual sport participants were shown in Table 8. There were no significant differences in age, height and weight between the two groups.

Table 8: Descriptive Data for Weight Class and Non-Weight Class sport Athletes in the 2012 and 2016 Summer Olympics.

Variable	Weight Class	Non-Weight Class
n	2,750	19,140
Age (y)	26.8 ± 0.1	27.2 ± 0.1
Body Mass (kg)	76.6 ± 0.1	72.0 ± 0.1
Height (cm)	172.6 ± 0.1	177.4 ± 0.1
Birthdate	June 15	June 16

Figure 18 shows the birth month distributions and Lorenz curves for the weight class and non-weight class sports. For the weight class sports the distribution of birth months was fairly consistent across months. As such, the Lorenz curve is positioned very close to the line of equality. For the non-weight class sports, there is a tendency for the distribution and Lorenz curve to be consistent with a RAE.

Figure 18. Birth Month Distributions and Lorenz Curves for Weight Class (top) and Non-Weight Class (bottom) sports in the 2012 and 2016 Summer Olympics.

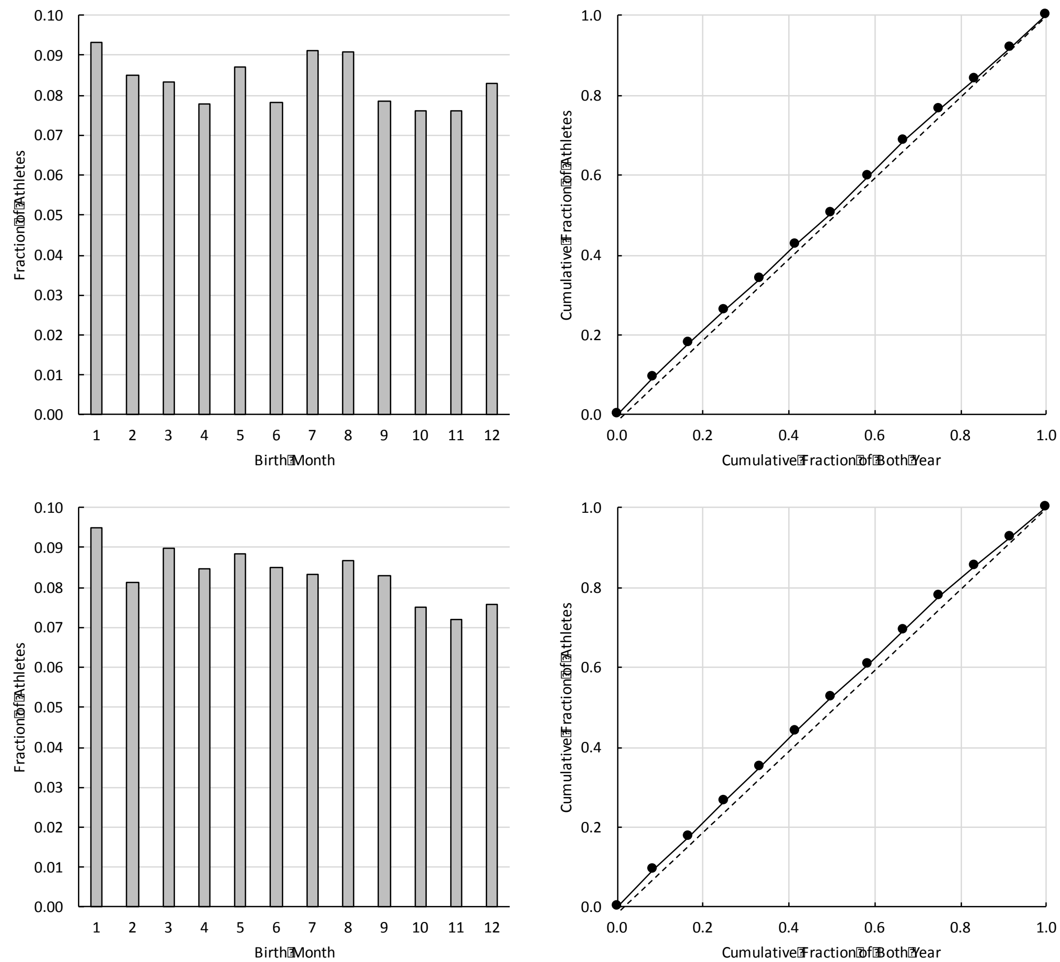


Table 9 shows the cohort and average Gini coefficient, r^2, slope and X^2 values for the two groups of athletes. For the non-weight class sports, the Gini coefficient and slope values were negative. In addition, the r^2, slope and X^2 values are statistically significant from zero resulting in a RAE. For the weight class sports, the Gini coefficient and slopes were negative. However, the r^2, slope and X^2 were not significant from zero. There were no significant differences between weight class and non-weight class sports for any of the average variables.

Table 9: RAE Statistics Comparing Weight Class and Non-Weight Class sports for the 2012 and 2016 Summer Olympics.

Variable	Group	Cohort	Average	SEM
Gini	Non-Weight Class	-0.0344[a]	-0.0401	0.0099
	Weight Class	-0.0180	-0.0141	0.0122
r^2	Non-Weight Class	0.6292[a]	0.2462	0.0485
	Weight Class	0.1946	0.1483	0.1141
Slope	Non-Weight Class	-0.0014[a]	-0.0017	0.0004
	Weight Class	-0.0008	-0.0006	0.0005
X^2	Non-Weight Class	108.99[a]	17.74	1.97
	Weight Class	13.83	11.68	1.47

[a]$p<.05$

Medalists vs Non-Medalists

As expected, there were considerably fewer medalists than not-medalists. Descriptive data for the medalists and non-medalists are shown in Table 10. There were no differences in age, height and weight between the two groups.

Table 10. Descriptive Data for Medalists and Non-Medalist competing in the 2012 and 2016 Summer Olympics.

Variable	Medal	Non-Medal
n	2,347	19,543
Age (y)	27.0 ± 0.1	27.0 ± 0.1
Body Mass (kg)	74.6 ± 0.1	72.2 ± 0.1
Height (cm)	178.2 ± 0.1	176.6 ± 0.1
Birthdate	June 16	June 16

Birth month distributions and Lorenz curves for the medalists and non-medalists are shown in Figure 19. Athletes earning medals show a birth month distribution that was fairly consistent across months. In this case, the Lorenz curve is positioned very close to the line of equality. For the non-medalists, there is a tendency for the distribution and Lorenz curve to be consistent with a RAE.

Table 11 shows that cohort and average Gini coefficient, r^2, slope and X^2 values for the medalists and non-medalists. For the medalists, the Gini coefficient and slope values were negative. However, only the r^2 and X^2 values are statistically significant. For the non-medalists, the Gini coefficient and slopes were negative. In addition, the r^2, slope and X^2 were significant from zero. There were no significant differences between medalists and non-medalists for any of the average variables except for the r^2 values were significantly different between groups.

Figure 19. Birth month distributions and Lorenz curves for medalists (top) and non-medalists (bottom) for the 2012 and 2016 Summer Olympics.

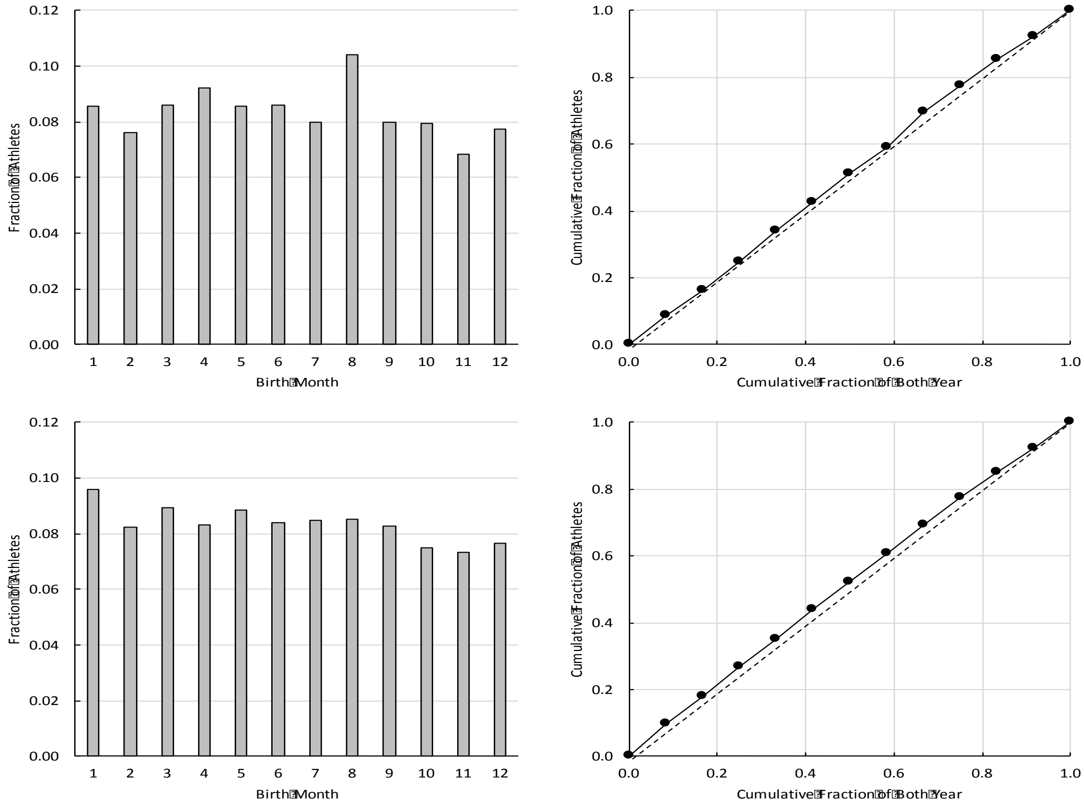


Table 11: RAE Statistics comparing Medalists and Non-Medalists of the 2012 and 2016 Summer Olympians.

Variable	Group	Cohort	Average	SEM
Gini	Non-Medal	-0.0340[a]	-0.0362	0.0089
	Medal	-0.0186	-0.0443	0.0216
r^2	Non-Medal	0.6481[a]	0.2371	0.0462
	Medal	0.0983[a]	0.1166[b]	1.0623
Slope	Non-Medal	-0.0014[a]	-0.0015	0.0004
	Medal	-0.0008	-0.0019	0.0009
X^2	Non-Medal	105.51[a]	16.08	1.66
	Medal	24.88[a]	13.12	1.06

[a]p<.05 versus 0, [b]p<.05 between groups

CHAPTER 5

DISCUSSION

This study examined birth month distributions of more than 20,000 athletes competing in the 2012 and 2016 Summer Olympics. With this data set, the results are consistent with a RAE in the highest sporting event in the world. This conclusion is based on three pieces of evidence. First, there was negative Gini coefficient for the entire cohort. In addition, the mean Gini coefficient for all sports was significantly different from zero. Second, the slope of the distribution of birth months across the year was negative significantly different from zero. Third, there was a significant X^2 value for a comparison of the birth month distribution to an expected consistent distribution. Thus, birth months of athletes competing in the Summer Olympics over represent the beginning of the calendar year versus the later part. In addition, the distribution of birth dates declines from January 1 to December 31 with an average birthdate of June 16. However, the RAE in this cohort is rather small and qualitatively less than previously reported for other sports and ages. The Gini coefficient has not been routinely used to asses the RAE. However, analysis of the data of Williams (2010) yields a Gini coefficient of -0.198 (personal communication). This is considerably larger that that found for the entire Olympic athlete cohort, -0.032. Interestingly, some specific sports display larger RAE values whereas several display a small negative RAE. Lastly, there does not seem to be an effect of gender, type of sport (individual vs team and weight class vs non-weight class) or quality of athlete on the RAE. Given these results, the key findings conclusions from this study are:

1. The RAE exists in Olympic competition
2. The RAE is relatively small compared to other sport age groups
3. The RAE varies from one sport to another
4. The RAE is not dependent on gender or type of sport or success of the athlete

It is important to point out that the two databases used in this study were very similar. Both contained similar numbers of athletes and there were no significant differences in age, body mass and height between the two groups. Also, there are several limitations and delimitations that may affect the outcome of the study. First, it is assumed that the data are complete and without error. Examination of the raw data resulted in several athletes being omitted from the study due to missing data and errors in data. Second, the two databases came from different sources. For the 2012 Olympics, data were obtained from The Guardian website, a London news outlet. For the 2016 data, they were obtained from Kaggle, an online source of various datasets. Third, only the 2012 and 2016 Summer Olympics were analyzed as these were the only two competitions available. Fourth, some sports were combined into a general category. For example, "athletics" comprises of both track and field events. Also "cycling" comprises track and road cycling as well as mountain biking. This was done to maximize the available athletes for specific types of sports. However, Bjerke et al. (2016) suggest that within a sport, the RAE can differ. There study showed RAE differences between speed and technical skiing events.

Despite these limitations, the two databases resulted in a large cohort of athletes used in this study. In fact, the number of athletes used (>21,000) appears to be the largest cohort examined for the existence of a RAE. For example, on one of the largest studies of its kind, Kirkendall (2014) examined the RAE in youth soccer using 5943 players. In addition, this is the first study to simultaneously examine the influence of specific sport, type of sport and gender on the RAE. Lastly, this study utilized multiple approaches to quantify the RAE. In particular, the Gini coefficient and Lorenz curve has only been used to study the RAE in one publication (Jimenez & Pain, 2008). Thus, this study adds to the understanding of the RAE by presenting results using a large cohort of athletes, a comprehensive looks at a variety of sports and a unique approach to quantifying the RAE.

Male vs Female

Several studies show that there are differences in the RAE between youth males and females. Table 1 highlights the RAE in several studies and shows that while both male and females exhibit the RAE, more studies support the effect occurring in males. Vincent and Glamser (2006) found that in US Olympic Development Program youth players a minimal RAE was found for females but a rather large one was found for males. Arrieta et al. (2016) found similar results for U16, U18 and U20 European Basketball Championship male and female teams. Thus, a comparison of the RAE based on gender was warranted. In the comparison of males and females competing in the 2012 London and 2016 Rio de Janerio summer Olympics, there were no notable differences in terms of a RAE between male and female competitors. The number of athletes was similar and the average birthdates were June 17 for males and June 16 for females. Knowing that the middle day of a normal calendar year is July 2, there may be an inclination, but not a trend of RAE. Further, the Gini coefficients were similar and there was no statistical difference between the r^2, slope and X^2 values between genders. Thus, at the Olympic level where there is a small RAE, the magnitude of the RAE is similar for males and females.

Team vs Individual

When selecting athletes to participate in elite competition such as the Olympics, different selection criteria may be employed for team and individual sports. For team sports, selection often involves qualitative criteria. This may be related to evaluation of talent, body type, contribution to the team dynamic, etc. For individual sports, selection is often more qualitative. For example, in US track and field, the 1st, 2nd and 3rd place finisher at the Olympic trials "makes the team" regardless of his or her past performance history or other qualitative measures. The results comparing team sports and individual sports concluded that there was not a clear difference in RAE between sport types. In fact, the largest positive and negative RAEs were found in golf and triathlon, respectively. These are both categorized as individual sports. The individual sports had more athletes primarily due to athletics and aquatics. The team sports were basketball, football, handball, hockey, rugby sevens, volleyball and water polo. The individual sports studied were aquatics, archery, athletics, badminton, boxing, canoe, cycling, equestrian, fencing, golf, gymnastics, judo, modern pentathlon, rowing, sailing, shooting, table tennis, taekwondo, tennis, triathlon, weightlifting and wrestling. The average birthdate for both was June 16.

It should be noted that the are some limitations of this analysis and categorization of sports. First, some sports like rowing are considered individual. However, events such as the rowing "eights" comprise a team selection process. Also, relay events in athletics and aquatics utilize some qualitative selection versus pure qualification. Second, some individual sports violate the quantitative selection process. For example, US gymnastics combines a quantitative component using scores from the Olympic trials competition.

However, final selection of the team also utilizes qualitative assessments. Such an approach likely occurs in other countries and other sports.

Based on the above, it appears that there are no noteworthy differences in RAE between team and individual sport competition.

Medal vs Non-Medal

In ice hockey, studies show that the RAE shrinks as the level of competition increases (Sherar et al., 2007; Barnesly et al., 1988; Daniel & Janssen, 1987). Thus, as age or the level of competition increases, the occurrence of RAE decreases. However, this could be an effect of maturation. On the other hand, Kirkendall (2014) found that the RAE in youth soccer was not influenced by the quality of the team. That is, teams with more wins did not display RAEs different from teams with fewer wins. Similarly, Williams (2010) found no systematic differences in the birth month distributions between U17 World Cup soccer teams who were considered more successful and those less successful teams in the tournament. In this study, medalists were compared to non-medalists. The assumption is that the medalists are more "elite" that their counterparts. There were more non-medalists than medalist, which was to be expected. This comparison sis not identify any RAE differences. In fact, average birthdate for both groups was June 16. Thus, the quality or success of the athlete does not seem to influence the RAE

Weight Class vs Non-Weight Class

Delorme et al. (2014) found no RAE in amateur male and female boxers. Thy suggested that this was due to the use of weight classes. Weight classes would eliminate the "size advantage" enjoyed by a relatively older athlete. On the other hand, Albuquerque et al. (2015) found a RAE in Olympic judo, a weight class sport. However, they used a "semester" approach that monthly or quarterly examination of birthdate distribution. The present results comparing weight class and non-weight class sports concluded that there was no differences between groups of sports. The weight class sports had fewer athletes as expected given the small number of these sports. The weight class sports were boxing, judo, taekwondo, weightlifting and wrestling. The non-weight class sports studied was aquatics, archery, athletics, badminton, basketball, canoe, cycling, equestrian, fencing, football, golf, gymnastics, handball, hockey, modern pentathlon, rowing, rugby sevens, sailing, shooting, table tennis, tennis, triathlon, volleyball and water polo. The average birthdates for these groups of sports differed by a single day (June 15 and 16). Thus it appears that weight classification do not influence the RAE in Olympic sports.

Olympic Competition Comparisons

There are relatively few studies that examine Olympic athletes. O'Neill et al. (2016) examined Olympic athletes from Australia and Spain. They found no RAE for the entire cohort or for team and individual sports. This differs from the present study in terms of the entire cohort but agrees with the team vs individual sport findings. Werneck et al. (2016) found a significant X^2 values for Olympic basketball but the distribution was not indicative of a RAE. The largest number of players were born in the second quarter to the year. The present study also found a significant X^2 values for basketball but a small Gini. In agreement with the present study, Albuquerque et al. (2012) found no RAE in Olympic taekwondo athletes. In a second study, the results of Albuquerque et al. (2015) agree with the present study in that a RAE is present in Olympic judo athletes. While there are relatively few studies on the RAE of Olympic athletes, previous studies generally agree with the present study.

Significance for Athletic Development

The present study suggests that there was a RAE in Olympic competition at the 2012 London and 2016 Rio de Janerio games. However, the effect is much smaller than those found at the youth sport level. With this conclusion, one must wonder why this effect is prominent at the youth sport level then “disappears” a the senior, elite level. The present results do not offer an explanation. It is possible that as players mature, there are more drop outs in the relatively older athletes. Perhaps this results from their selection to advanced training programs, requiring more physical, psychological and time commitment. This could result in injury or burn out. It is also possible that the relatively younger athletes feel the need to prove themselves after being passed over at younger ages. This desire increases their drive and training intensity, ultimately making them superior competitors when they reach adulthood. Similarly, relatively older athletes may be superior because of greater physical stature. This results in them not focusing on developing the technical aspects of their sport. Ultimately this leads to a weaker player at the professional level when stature differences are minimized.

Conclusions

The overall results of this study suggest that the RAE exists at Olympic level competition. However, it is less prominent than youth sports programs and differs from one sport to another. Gender, team type and athlete success do not seem to effect the RAE.

REFERENCES

Albuquerque, M., Lage, G., Costa, V., Ferreira, R., Penna, E., Moraes, L., Diniz, L. Relative age effect in Olympic taekwondo athletes. *Perceptual and Motor Skills.* 2012; 114(2): 461-468

Albuquerque, M., Lage, G., Costa, V., Diniz, L. The relative age effect in combat sports: An analysis of Olympic judo athletes, 1964-2012. *Sage Journals.* 2015; 121 (1): 300-308

Arrieta H, Unda, J., Gil, S., Irazusta, J. Relative age effect and performance in the U16, U18 and U20 European basketball championships *Sports Sciences.* 2016; 34(16): 1530-1534.

Barnsley R, Thompson, A. Birthdate and success in minor hockey: The key to the NHL. *Canadian journal of behavioral science.* 1988; 20(2): 167-176

Barnsley, R., Thompson, A., Barnsley, P. Family planning: football style. The RAE in football. *International Review for the Sociology of Sport.* 1992; 27: 77-88

Baxter-Jones, A., Helms, P., Maffull, N., Baines-Preece, J., Preece, M. Growth and development of male gymnasts, swimmers, soccer and tennis players: A longitudinal study. *Annals of Human Biology.* 1995; 22, 381-394

Baumler, G. The relative age effect in soccer and its interaction with chronological age. Talk at the Soccer Sport Science Symposium in Overreaching, September, 1996.

Bjerke, O., Loras, H., Pedersen, A. Variations of relative age effect within and across groups in elite alpine skiing. *Mental and Physical Health.* 2016; 5: 1-6

Barnsley R., Thompson, A. Family planning: football style. The relative age effect in football. *International Sociology of Sport Association.* 1992; 27(1): 77-87.

Boucher, J., Mutimer, B. The relative age phenomenon in sport: a replication and extension with ice-hockey players. *Research Quarterly for Exercise and Sport.* 1994; 65(4): 377-381.

Breznik K. Relative age effect in mind games: the evidence from elite chess. *Perceptual and Motor Skills.* 2016; 122(2): 583-594.

Brewer, J., Balsom, P., Davis, J. Seasonal birth distribution amongst European soccer players. *Sports Exercise and Injury.* 1995; 1: 154-157

Campos F, Stanganelli, L., Rabelo, F., Campos, L., Pellerinotti, I. The relative age effect in male volleyball championships. *Sports Sciences.* 2016; 6(3): 116-120.

Cobley, S., Schorer, J., Baker, J. Relative age effects in professional German soccer: A historical analysis. *Sports Sciences.* 2010; 26 (14): 1531-1538

Delorme, N., Boiche, J., Raspaud, M. Relative age effect in female sport: a diachronic examination of soccer players. *Scandinavian Journal of Medicine and Science in Sports.* 2009; 20 (3): 509-515

Delorme N. Do weight categories prevent athletes from relative age effect? *Sports Sciences.* 2014; 32(1): 16-21.

Delorme, N. and Champely, S. Relative age effect and chi-squared statistics. *Sociology of Sport.* 2015; 50(6): 740-746

Delorme, N., Radel, R., Raspaud, M. Relative age effect and soccer refereeing: A ‘strategic adaptation’ of relatively younger children? *European Journal of Sport Sciences.* 2011; 13(4): 400-406

Diaz Del Campo, D., Vicedo, J., Villora, S., Jordan, O. The relative age effect in youth soccer players from Spain. *Sports Science and Medicine.* 2010; 9(2): 190-198

Diaz, M., Villora, S., Vicedo, J., Olivares, J. Soccer and relative age effect: a walk among elite players and young players. *Sports.* 2017; 5 (5): 1-20

Dudink, A. Birth date and sporting success. *Nature.* 1994; 369, 592

Edgar S., O'Donoghue, P. Season of birth distribution of elite tennis players. *Sports Sciences.* 2005; 23(10): 1013-1020.

Ford, PR., Williams, MA No relative age effect in the birth dates of award-winning athletes in male professional team sports. *Research Quarterly for Exercise and Sport*, 2013; 82: 570-573.

Gini, C. Measurement of inequality of incomes. *The Economic Journal*. 1921;31: 124–126.

Grondin S, Deshaies P, Nault LP. Trimestres de naissance et participation au hockey et au volleyball. *La Revue Québécoise de l'Activité Physique,* 1984; 2: 97-103.

Grondin , S., Koren, S. The relative age effect in professional baseball: A look at the history of Major League Baseball and at current status in Japan. *Avante*, 2000; 6: 64-74

Helsen W, Starkes, J., Van Winckel, J. The influence of relative age on success and dropout in male soccer players. *Human Biology.* 1998; 10: 791-798.

Helsen W., Hodges, N., Van Winckel, J., Starks, J. The roles of talent, physical precocity and practice in the development of soccer expertise. *Sports Sciences.* 2000; 18: 727-736

Helsen W, Van Winckel, J., Williams, M. The relative age effect in youth soccer across Europe. *Sports Sciences.* 2005; 23(6): 629-636.

Helsen W, Baker, J., Schoerer, J., Steingroever, C., Wattie, N. The relative age effect in youth chess competition. *Exercise, Movement, and Sport.* 2014; 46(1).

Hirose N. Relationships among birth-month distribution, skeletal age and anthropometric characteristics in adolescent elite soccer players *Sports Sciences.* 2009; 1-8.

Jimenez, I., Pain, M. Relative age effect in Spanish association football: its extent and implications for wasted potential. *Sports Science.* 2008; 26 (10): 995-1033

Kirkendall, D. The relative age effect has no influence on match outcome in youth soccer. *Sport and Health Science.* 2014; 3 (4): 273-278

Lorenz,M. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association.,* 1905; 9: 209–219

Malina R, Cumming, S., Kontos, A., Eisenmann, J., Ribeiro, B., Aroso, J. . Maturity- associated variation in sport-specific skills of youth soccer playwers aged 13-15 years. *Sports Sciences.* 2005; 23(5): 512-522.

Malina R, Reyes, M., Eisenmann, J., Horta, L.,Rodrigues, J. Miller, R. Height, mass and skeletal maturity of elite Portuguese soccer players aged 11-16 years *Sports Sciences.* 2000; 18: 685-693.

Musch, J., Hay, R. The relative age effect in soccer: Cross-cultural evidence for a systematic discrimination against children born late in the competition year. *Sociology of Sport Journal.* 1999; 16: 54-64

Musch, J and Grondin, S. 2001. Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental Review*, 21: 147–167.

Olympic Games 2012 Database. 2012 https://www.theguardian.com/sport/series/london-2012-olympics-data Access date, January 20, 2017.

O'Neill, K., Cotton, W., Cuadros, J., O'Conner, D. An investigation of the relative age effect amongst Olympic athletes. *Talent Development and Excellence.* 2016; 8(1): 27-39

Romann, M. and Cobley, S. Relative age effect in athletic sprinting and corrective adjustments as a solution for their removal. *PLoS ONE.* 2015; 10(4): 1-12

Sherar, L., Baxter-Jones, A., Faulkner, R., Russell, K. Do physical maturity and birth date predict talent in male youth ice hockey players? *Sports Sciences.* 2007; 25(8): 879-886.

Stanaway, K. and Hines, T. Lack of season of birth effect among American athletes. *Perceptual and Motor Skills.* 1995; 81: 952-954

Thompson, AH, Barnsley, RH and Stebelsky, G. "Born to play ball", the relative age effect and professional baseball. *Sociology of Sport Journal*, 1991; 8: 146-151.

Thompson, A., Barnsley, R., Stebelsky, G. Baseball performance and the relative age effect: Does Little League neutralize birthdate selection bias? *Nine*. 1992; 1: 19-30

Vaeyens, R., Phillppaerts, R., Malina, R. The relative age effect in soccer; A match-related perspective. *Sports Sciences.* 2004; 23(7): 747-756

Verhulst, J. Seasonal birth distribution of West European soccer players: A possible explanation. *Medical Hypothesis.* 1992; 38: 346-348

Vincent, J. and Glamser, F. Gender differences in the relative age effect among US Olympic development program youth soccer players. *Sports Sciences.* 2005; 24 (4): 405-413

Werneck, F., Coelho, E., Oliveira, H., Junior, D., Almas, S., Lima, J., Matta, M., Figueiredo, A. Relative age effect in Olympic basketball athletes. *Science and Sports.* 2016; 31: 158-161.

Williams J. Relative age effect in youth soccer: analysis of the FIFA U17 World Cup competition. *Scandinavian Journal of Medicine & Science in Sports.* 2010; 20(3): 502-508.

World Rugby Limited. Relative Age Effect. 2010; http://playerwelfare.worldrugby.org/, Access date February 09, 2017.

2016 Rio de Janeiro. 2017:https://www.kaggle.com/rio2016/olympic-games , Access date January 20, 2017.