Erasmus
School of
Economics

ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Master Thesis | Strategy Economics

The effects of and solution to the relative age effect on young Dutch soccer players

Rens van der Star | 472637

Supervisor: T.L.P.R. Peeters

Second assessor: M. van de Velden

Word count: 13623| Date final version: 31-10-2023

The views stated in this thesis are those

of the author and not necessarily those

of the supervisor, second assessor,

Erasmus School of Economics or

Erasmus University Rotterdam.

Abstract

This research addresses the importance of understanding and minimizing the Relative Age Effect (RAE) in soccer. The RAE is a phenomenon that can significantly impact team composition and talent development. The research problem revolves around the challenges posed by the RAE in the Dutch Soccer Association (KNVB). The study investigates the influence of player performance and team level on the RAE, as well as the impact of shifting cut-off dates. The study employs simple regressions, a between-subject analysis and several simulations to explore these factors across different shift scenarios.

The results from the simple regression and between-subject analysis show that the RAE is present in the Dutch soccer association. The proceeding simulations demonstrate that shifting cut-off dates can influence the manifestation of the RAE to some extent, albeit with small effect sizes. This research adds to the body of knowledge on the RAE by giving a numerical answer to what the impact could be when the shifts to cut-off dates would be implemented. So far little to no research has provided this and so this research could be a basis for further potential studies and policy makers. The way the research has been performed has also yet to be done in papers.

To conclude, this research provides valuable insights into the RAE in soccer and its implications for team composition. While further research is necessary, the study contributes to the understanding of the RAE and offers a new foundation for developing fair and equitable systems that promote equal opportunities for all players, irrespective of their birth month.

**Disclaimer*

The research that has been conducted contains certain aspects that involve the use of ChatGPT. ChatGPT was mainly used to aid in the construction of R formulas. This was done to enhance the efficiency and accuracy of the R formulas

Table of contents

1. Introduction

Cristiano Ronaldo, renowned as one of the greatest soccer players in history, amazes audiences with his footwork and sharpshooting skills. Born on February 5, Ronaldo's exceptional talent and achievements are known around the world, yet they beckon an intriguing inquiry: Would Ronaldo's journey to stardom have shifted if he had been born just a few months earlier, for example in December? While this may seem far-fetched, the Relative Age Effect (RAE) invites us to ponder over this fascinating theory (Musch & Grondin, 2001).

The RAE is an observable bias influencing various sectors, from academia to sports. This bias stems from the common practice of grouping individuals into cohorts based on their birth year, resulting in age differences of up to 12 months within a single group. As a result, individuals born closer to the cut-off date often outperform their peers born further from the cut-off date due to their relative physical and cognitive maturity, which significantly impacts the talent identification and selection processes (Cobley et al., 2009)

In sports, especially youth sports, the implications of the RAE are significant. A European study on youth soccer found a noticeable overrepresentation of players born closer to the cut-off date and a corresponding underrepresentation of those born further from the cut-off date (Helsen et al., 2005). Similar patterns have been uncovered across various sports, like basketball (Delorme et al., 2010). This birthdate bias challenges the fairness and inclusivity of talent development systems and could potentially overlook late-born talents.

The RAE is not just about the selection process; it also significantly impacts performance estimators, like player rankings, scores, or team selection. For example, in sports such as soccer and basketball, early-born athletes generally have higher performance scores and are more likely to be selected for teams due to their physical superiority, a product of their relative age (Delorme et al., 2010). However, as these athletes age and physical differences diminish, the RAE can skew these performance indicators, potentially overvaluing early-born athletes and undervaluing those born later in the year. This initial advantage tends to accumulate over time and make the advantage bigger. This difference leads to an ascending growing difference between the two groups. This effect is called the Matthew effect and is also applicable to the people who experience a disadvantage (Acar, 2011).

Tackling the RAE is multifaceted due to the great number of contributing factors, including socio-cultural context, the nature of the sport, and the specific policies of sports organizations. Researchers have proposed various solutions, such as recognizing that talent doesn't solely rely on physical attributes, but also on factors like motivation, resilience, and autonomy, which could help mitigate the RAE (Collins & MacNamara, 2012). Another proposed solution to the RAE problem has been adjusting age groupings or implementing a rotating cut-off date system to level the playing field (Romann & Cobley, 2015). The rotating cut-off date system refers to a method of determining age eligibility for youth competitions. Instead of using a fixed calendar date, it employs a dynamic date that shifts each year with a set amount of months. This works as follows: Imagine a soccer league where team selection is based on birth year, and those born in January have an advantage over December-born players because of the extra months of growth and experience. By shifting the cut-off date for team selection from January to July, a player originally born in July would now have the relative age advantage, potentially levelling the playing field for all participants. This is the research topic this paper predominantly focusses on:

“What are the effects of the rotating cut-off date system on the RAE in the Dutch football association”.

This solution to the RAE has been chosen as the most promising solution to the issue by the KNVB (KNVB, 2021). However, because this paper is one of the first papers to investigate the repercussions and influence of the rotating cut-off date system, most results are based on experimental research.

This study delves into the intricate relationship between the RAE, player performance, and team composition within a specific sports dataset from the KNVB. Through this investigation, I aim to contribute to the growing body of literature on the RAE and offer insights with practical implications for the KNVB and the associated talent development programs. Understanding and acknowledging the bias posed by the RAE is a crucial step towards creating more equitable strategies for identifying and nurturing talent. It paves the way for a sports ecosystem that values and nurtures every potential player, irrespective of when they were born in the year.

This paper talks a lot about team level. In this case, being part of a higher team level would indicate a less skilled and capable athlete and a lower team level would indicate a more skilled and capable athlete.

Furthermore, this paper also mentions being born further and closer to the cut-off date. To illustrate, consider two children born in the same year, one in March and the other in December. When the cut-off date is set in January, being born closer to this date would relate to the child born in March. This implies that the child born in March would be relatively older than someone born further from the cut-off date, such as the child born in December.

2. Literature review

In this section, I analyse the findings of various empirical studies on the impact of the RAE. Many researchers have examined this topic by considering multiple types of effects. In this case, I started by looking at the effects on academic performance, mental health, and physical attributes. In the second part of the literature review, I explored the effects on different kinds of sports and dove into football specifically. In the final part of the literature review I explored the different solutions offered by the literature.

2.1 The relative age effect

The calendar cut-off date principle, commonly known as the RAE or birthdate effect, sheds light on a notable phenomenon where individuals born closer to a specific cut-off date often enjoy advantages in various aspects of life. In contrast, individuals born further from a specific cut-off date often experience disadvantages (Green, 2023). According to research by Chen et al. (2021), children born further from the school cut-off date, as observed in Taiwan, face a higher risk of being diagnosed with conditions such as ADHD, anxiety disorder, and depressive disorder. Furthermore, Crawford et al. (2013) suggest that relative age plays a role in fostering risky and rebellious behaviours among children. In a separate study, Ando et al. (2019) highlight that children born further from the cut-off date tend to exhibit lower levels of emotional development. These findings align with the research conducted by Patalay et al. (2015), who specifically focused on children born further from the cut-off date aged 11 and 12 and found that they are particularly susceptible to these effects. Moreover, Patalay et al. (2015) also explored gender differences and revealed that boys, in particular, tend to face greater challenges when they are born further from the cut-off date within this age group. Thus, it can be inferred that there are neurocognitive effects associated with children who are born around the calendar cut-off date.

These observations in the mental health of children highlight the pervasive nature of the calendar cut-off date principle. Its influence extends beyond mere eligibility assessments and diagnostic outcomes, demonstrating its impact on various aspects of human development and behaviour. Recognizing and understanding these effects is crucial for developing equitable policies and interventions that consider the potential disadvantages faced by individuals born further from the cut-off date. By doing so, we can strive to mitigate any unfair advantages or disparities arising from this phenomenon and promote a more inclusive and equitable society.

Research has provided evidence of the effects associated with the calendar cut-off date principle in education. Numerous studies have demonstrated that children born closer to the cut-off date are more likely to excel academically. Bedard and Dhuey (2006) have shown that children born further from the cut-off date, in particular, are significantly impacted and tend to achieve lower grades compared to their older peers. This disparity in grades also extends to higher education, as children born further from the cut-off date are less likely to attend university. This can be attributed to the cumulative effects of early childhood skill deficiencies, which result in a continued lack of skills as they progress through their academic journey. This discrepancy is particularly pronounced in subjects such as mathematics, reading ability, and partially in science performance.

The educational disadvantage experienced by children born further from the cut-off date continues to accumulate throughout their lives. They often lack certain essential soft skills, which ultimately diminishes their prospects of becoming effective leaders (Dhuey & Lipscomb, 2008). Dhuey and Lipscomb (2008) emphasize that these soft skills not only contribute to higher earning potential in adulthood but are also influenced by school structure rather than genetics. Students born closer to the cut-off date also exhibit an increased likelihood of assuming leadership roles during high school, further supporting this claim.

The distribution of students in gifted classes further reinforces the impact of relative age. Maddux et al. (1981) found that classes for gifted children consist predominantly of children born closer to the cut-off date, while classes for slow learners and children with learning disabilities are disproportionately populated by children born further from the cut-off date (Diamond, 1983). Importantly, these patterns hold for both sexes.

These findings underscore the long-lasting consequences of the calendar cut-off date principle in education. They highlight the importance of acknowledging and addressing the inherent advantages and disadvantages associated with relative age. Implementing strategies to mitigate these effects and provide equitable opportunities for all students, regardless of their birthdate, is crucial for fostering a fair and inclusive educational environment. By recognizing and addressing these disparities, we can strive to maximize the potential and success of all students, irrespective of their age relative to the cut-off date.

2.2 The influence of the relative age effect on sports

The calendar cut-off date effect has been the subject of extensive research in various sports, resulting in a multitude of papers exploring this phenomenon. Researchers such as Hurley et al. (2001), Barnsley and Thompson (1988), Kelly et al. (2020), Edgar and O'Donoghue (2005), and Okazaki et al. (2011) have investigated this effect in ice hockey, squash, tennis, and volleyball, respectively. While the majority of studies confirm the presence of a calendar cut-off date effect, many do not delve into potential solutions. However, it is worth noting that some papers, like Kelly et al. (2020), conclude that there is no significant effect of relative age in the sport of squash.

The findings of Okazaki et al. (2011) on female volleyball players shed light on the nuanced nature of the calendar cut-off date effect. They discovered that the psychological aspects of ageing have a greater impact than the physical aspects. This highlights the fact that researchers do not always reach a consensus regarding the magnitude of the effect and the specific areas within a sport that it influences.

The examples discussed so far represent only a fraction of the research conducted on this topic. Some studies have explored the effect across multiple sports, revealing its significant impact, albeit with varying outcomes—some indicating a negative effect and others a positive effect (NCAA, 2019).

Thus far, the calendar cut-off date effect has been primarily attributed to developmental differences between children born closer to and further from the cut-off date, encompassing physical, emotional, and cognitive abilities (Mann & Van Ginneken, 2017). These discrepancies naturally arise as children grow and mature. However, due to the artificial distinctions imposed by the calendar cut-off date, these differences have long-term consequences. Mann and Van Ginneken (2017), for instance, demonstrate that this bias is closely linked to long-term success in sports.

In summary, the research surrounding the calendar cut-off date effect in sports is extensive and spans a wide range of sports disciplines. While some papers confirm the existence of the effect, others contest its significance or even refute its presence. Moreover, the specific aspects of a sport that are affected by the calendar cut-off date vary among studies. Understanding and addressing this phenomenon is vital for promoting fairness and equal opportunities for athletes of all birthdates, thus fostering a more inclusive and equitable sporting environment.

2.3 The influence of the relative age effect on soccer

The RAE in soccer is a phenomenon that becomes particularly evident in highly competitive activities where performance is strongly correlated with age and maturity level. In many European countries, soccer teams are composed of participants born within the same calendar year, although sometimes they might span two consecutive years. This means that a child born closer to the cut-off date could be almost 12 months older than another athlete born towards the end of the same year, despite competing together (Votteler & Höner, 2013).

The implications of RAE are significant. Players born further from the cut-off date, who may not have experienced a growth spurt and are less physically or mentally developed compared to their older peers, often face challenges. As a result, they may be overlooked for academy selection, missing out on opportunities for training under top coaches. This oversight of younger players can lead to diminished enjoyment and increased dropout rates (*Relative Age Effect in Youth Soccer*, 2019). Additionally, Sidelines (*Relative Age Effect in Youth Soccer*, 2019) highlights that older players who do receive preferential treatment may develop a sense of entitlement, potentially hindering their development by practising less and taking things for granted.

Various studies have examined the RAE in elite soccer, revealing different effects depending on the analysed category or qualifying round achieved by the national team. The RAE is observed in the lower youth categories, with a stronger influence on younger age groups, such as the under-17 category (González-Víllora et al., 2015). This claim is supported by Peña-González et al. (2020) who have found that the RAE is stronger for younger players playing at a higher competitive level. This also impacts the number of players who advance to play professionally in the later parts of their lives. Research done by Işın (2021) shows that there is a significant difference between the number of players born in the first quartile of the year who have been overrepresented on the podium of soccer championships in comparison to those born in the last quartile, who have been underrepresented.

Barnsley et al. (1992) focuses on players in national teams, specifically analysing the under-17 and under-20 age groups. Their research confirmed the existence of a significant RAE within these age categories, indicating that players born closer to the cut-off date are more likely to be chosen for national team representation. However, one must consider alternate reasons then the RAE for these observations. One such alternative explanation, as proposed by Barnsley et al. (1992), is the presence of certain regulations, structures, or practices within player development systems that might contribute to the manifestation of the RAE. For example, the cut-off dates used for age group selection or the emphasis on physical attributes in player evaluations could inadvertently favour older, more physically developed players.

The influence of the RAE extends beyond national teams to club-level soccer as well. Salinero et al. (2013) explored potential differences in the RAE based on players' positions on the field. Their study revealed that while the RAE persists across positions, its magnitude varies among them. For instance, the effect was found to be more pronounced in the goalkeeper position compared to other positions on the field. These findings highlight the complex interplay between relative age and player positions, suggesting that certain positions may be more susceptible to the influence of the RAE than others.

To conclude, the RAE is a prominent factor in highly competitive soccer activities, impacting player selection, development, and enjoyment of especially players born further from the cut-off date. While its influence is more pronounced in younger age categories, it still affects professional players most likely due to the early advantages they have received. The existence of RAE within the sport raises important questions about fairness and equal opportunities for all young players, prompting the need for ongoing research and potential adjustments in player development practices.

2.4 Strategies to mitigate the relative age effect

To address the issue of the RAE in soccer, several recommendations have been made in the literature. One common suggestion, proposed by Romann and Fuchslocher (2013), is to educate trainers and implement a rotating cut-off date to minimize the effect. This approach aims to raise awareness among coaches about the potential bias associated with RAE and encourage them to consider individual player development rather than solely relying on birthdates. Webdale et al. (2020) investigated various papers proposing solutions for RAE and highlighted the lack of implementation of these solutions in practice. They emphasized the importance of approaching the implementation of solutions cautiously, considering the potential impact on athletes' career and life outcomes. While it is essential to address the RAE, it is crucial to ensure that the implemented strategies do not inadvertently create new challenges or inequalities.

One effective approach to tackling the RAE is the adoption of “age banding” or “biological maturation”. Age banding involves grouping players based on their biological maturity rather than their birthdate. By considering players' individual stages of growth and development, this method aims to level the playing field and minimize the influence of age-related advantages. Cobley et al. (2009) conducted a study on age banding in academy soccer and found that it significantly reduced the RAE, allowing for a more balanced representation of players across different birth months.

Another potential solution is to implement a "playing up" strategy, where talented players are encouraged to compete in higher age groups. This approach helps to identify and nurture players based on their skill level rather than their birthdate alone. Wattie et al. (2008) explored the effects of playing up in youth ice hockey and found that it reduced the RAE, allowing younger players to gain valuable experience and challenge themselves against older, more developed opponents. Implementing similar strategies in soccer could help address the disadvantages faced by those born further from the cut-off date players and provide them with opportunities to compete and develop their skills against stronger competition.

Furthermore, providing specialized training programs and resources for players who are affected by the RAE can also help mitigate its negative impact. Creating development pathways that focus on individual player growth rather than chronological age can ensure that talented players receive appropriate support and opportunities for improvement. Helsen et al. (1998) conducted a study that demonstrated the effectiveness of a talent development program in reducing the influence of RAE in soccer. This program provided targeted training, mentorship, and guidance to players affected by RAE, resulting in improved performance and increased retention rates.

Implementing changes to the selection and evaluation processes can also help to minimize the influence of RAE. Utilizing multiple sources of information, such as skill assessments, physical evaluations, and performance observations, can provide a more comprehensive picture of a player's potential and minimize the reliance solely on birthdate as a determining factor. This approach encourages a holistic evaluation that takes into account various aspects of player development. Jiménez and Pain (2008) suggested that combining objective measurements with subjective evaluations could contribute to a fairer player selection process.

Votteler and Höner (2014) suggest that creating smaller age groups can help reduce the range of age differences within a group. Instead of having broad age categories like under 17, organizations can implement narrower divisions that take into account smaller age intervals. This way, the differences in physical and mental maturity among players within the same group are reduced, providing a fairer competition environment. Another solution involves the use of a weighted scoring system to account for the potential influence of RAE in competition. Currently, the ability of a player may be tested by for example sprinting contests that favour taller players. By assigning appropriate weights, adjustments to the scores or outcomes based on players' ages, the impact of RAE can be mitigated. This ensures that performances are evaluated and compared fairly. However, it is essential to acknowledge that informing coaches about RAE advantages may also have unintended consequences. It was also pointed out that some coaches are aware of the advantages of selecting older players and may intentionally favour them to gain a competitive edge. This could lead to these coaches receiving better offers for bigger teams due to their higher match-winning rates.

In conclusion, addressing the RAE in soccer requires a multifaceted approach. Strategies such as age banding, playing-up, specialized training programs, increased awareness, changes to the selection process, and coach education can all contribute to mitigating the influence of RAE and promoting a more equitable and inclusive environment for young players. Continued research, collaboration, and implementation of these solutions are crucial in levelling the playing field and unlocking the potential of all players, regardless of their birthdates. By creating fairer opportunities and nurturing talent based on individual merit and potential, soccer organizations can foster a more inclusive and diverse sporting culture.

3. Data

The data that has been used for this research, is from the KNVB. This data set has been chosen specifically by the KNVB to gain more knowledge about the RAE in their teams. The KNVB provided a data set with a total of 1113 different clubs with an average of 4.5 teams in the given age category per club. The data set contains a total of exactly 40000 observations separated over a period of 4 years, from 2019 to 2022. There are 5,000 players observed over this period, resulting in an observation averaging around every 6 months. These observation dates are observed and reported as the "Peildatum" within the data set. This variable is used as a time variable to gain information about the development of the players' performance. The players' performance is based on a formula from the KNVB that mostly uses the same system as the chess elo system. This means, that when people lose, they would lose a certain amount of elo rating and they would gain a certain amount of elo rating when they win. This amount is based on several factors, like the elo of the team they are playing in, the elo rating of the other team, whether or not they are playing in the matches and other factors that are not to be disclosed (Bosma & Vuegen, 2020). In Table 1 it can be observed that the player performance variable has an average value of 1573 with a standard deviation of 241. The range for this metric spans from 621 to 2732.

Table 1

Summary statistics

Variable	Observations	Mean	Std. dev.	Min	Max
Team level	40,000	1.8111	1.1301	1	10
Date of birth	40,000	6.4176	3.4447	1	12
Player performance	40,000	1573.054	241.5282	621	2732
Age of the team	40,000	15.3378	1.3889	12	19

Table 1 also gives more information about the following important variables:

Team level: This numerical variable outlines the level or category of the team the player is part of. This variable is mostly focussed on the classic sports system where team 1 is considered the best team and moving down to team 2 would mean a worse team in general. In the dataset, the average team level is 1.8111 with a standard deviation of 1.130176. The range of this variable extends from a minimum of 1 to a maximum of 10.

Date of birth: This field reflects the month of birth for each player. Represented by an integer value that ranges from 1 (January) to 12 (December). In this case, all observations are within the time period of 13th of June 2006 and 12th of June 2007, or in other words, within 1 calendar year. The average month of birth stands at 6.4176, corresponding to June when rounded to the nearest whole number.

Age of the team: This represents the age of the team someone is playing for. In the data set this is represented as for example playing for the team under 17. The mean team age is 15.34 years. The youngest and oldest teams in the dataset are under 12 and under 19 years old, respectively.

Other variables that are used in the regressions and simulations are:

Personal id: This attribute is a unique identifier assigned to each player within the dataset.

Club id: This attribute is a unique identifier assigned to each club within the dataset. Multiple players can be part of the same club, however, players within this data set do not change clubs over the observation period

Season ID: This is the season in which the observation was made.

To test my hypotheses, I also created extra variables. These were based on the variables mentioned previously or on other ways of creating variables.

Birth date quarter: I divided the birthdays into quarters based on the birth month of the players.

- The first quarter is made up of people born in January, February and March.
- The second quarter is made up of people born in April, May and June.
- The third quarter is made up of people born in July, August and September.
- The fourth quarter is made up of people born in October, November and December.

Team level Category: This variable is created as a means to test whether or not a player is playing in a team that can be considered as playing on a higher or lower level. However because this also depends on the club size, I have considered a team a “selection” team, a team that has the most skilled and capable athletes, under these conditions:

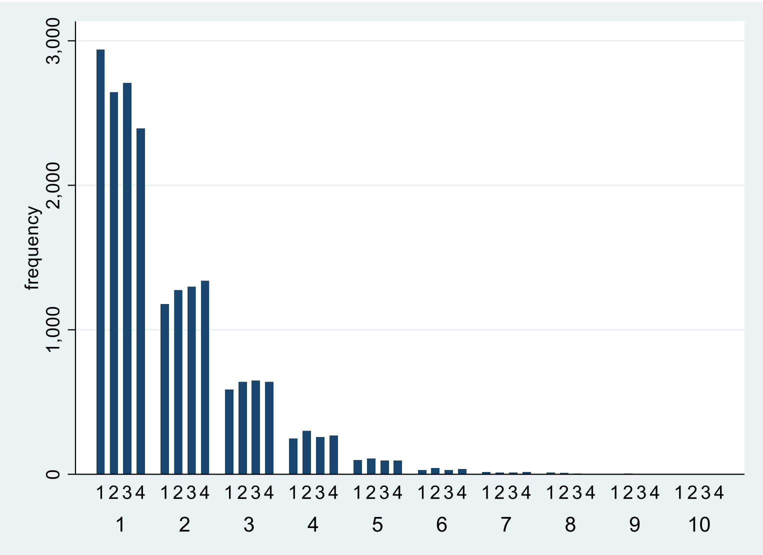
- If the club has 3 teams, only the first team is considered “selection”
- If the club has 4 to 6 teams, the first 2 teams are considered as “selection”
- If the club has more than 6 teams, the first 3 teams are considered as “selection”

Adjusted birth month: To test the second hypothesis, I have to shift the original birth month to test whether or not shifting the cut-off date significantly impacts the magnitude and presence of the RAE. This is done through this new variable that shifts the original birth date by an X amount of months depending on the tested shift of the cut-off date.

Because the research mainly relies on the fact that there are differences between the birth quarters, the first objective is to check if there are visual and statistical differences between the different birth quarters.

Figure 1

Distribution of players per team per quarter

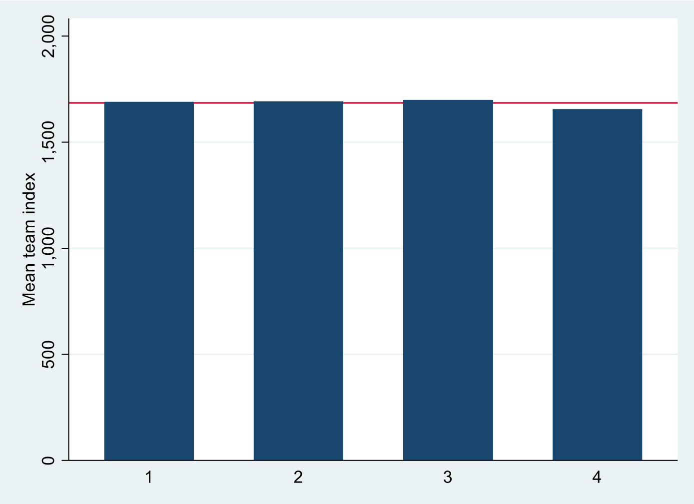


In this distribution, it can be observed that in especially the first two teams there is a large difference between the number of players born in each quarter. In the first team, there seems to be an overrepresentation of players born in the first quarter, while there is an underrepresentation of players born in the fourth quarter. The opposite seems to be the case in the second team. The difference appears to be smaller when going further down the team rankings, however, this can also be because there are far fewer observations for those teams. Another possible reason for the disparity in the number of players from the different quarters can be due to a difference in skill.

Zooming in on the players' performance of the first team, we can make some key observations (Figure 2).

Figure 2

Actual mean of player performance versus the expected player performance for players of the first team



It can be observed that the expected player performance and actual player performance are rather similar for all quarters, as all quarters are within a 10 per cent interval of each other. The players' performance of the fourth quarter is lower than the average and the other 3 quarters are above the average, with the third quarter producing the highest player's performance. This small difference could explain a small piece of why the fourth quarter is underrepresented in the first teams, but the difference seems too small to create such a large difference. Furthermore, according to this graph, most players should come from the third quarter, however, this is also not the case. This could be an indication of the RAE, but does not confirm anything yet. However, for these graphs to represent a good picture of reality, it is necessary for there to be a similar number of people born in each quarter. It is expected that ¼ of all observations, in this case 1250 observations because I have 5000 unique player IDs, fall within 1 quarter. The chi-square test was conducted to evaluate the difference between the observed and expected values.

Table 2

Chi-square test

Observations	=	5000
Chi-Square (χ2)	=	2.9008
p-value	=	.4071741

Note: The results of the chi-square test for an even spread of birthdays

The results show an insignificant result at a 10% significance level ($p > 0.1$), indicating that the observed data did not significantly deviate from the expected data. Therefore, the null hypothesis that there is no difference between the observed and expected values was not rejected. This means that it is expected that all teams should roughly have the same amount of players born in each quarter.

4. Hypotheses

The RAE suggests that individuals closer to the cut-off date have developmental advantages compared to those born further from the cut-off date, which can significantly affect their performance in various areas such as sports, education, and other social settings. The motivation behind this study was to delve into the intricate dynamics of the RAE within the context of the Dutch Soccer Association. Given the extensive body of research highlighting the widespread existence of RAE across various sports, the first hypothesis is aimed at investigating whether this phenomenon was also prevalent in this dataset.
Based on the established theory and previous research findings, my first hypothesis is as follows:

Hypothesis 1:

"The Relative Age Effect significantly influences the outcomes within the Dutch soccer association (KNVB), with individuals born closer to the cut-off date experiencing more positive outcomes compared to individuals born further from the cut-off date."

This hypothesis, while it is based on the existing understanding of RAE, creates a link between the RAE and the KNVB, reaffirming the urgency for equitable talent development strategies.

I also explored the potential of minimising the RAE. Building upon the exploration of the RAE and the research of the KNVB, this study seeks to determine if a strategic manipulation of cut-off dates could potentially reduce the impact of the RAE.
Based on the established theory and previous research findings, the second hypothesis is as follows:

Hypothesis 2:

"Shifting the cut-off dates used to group individuals significantly impacts the magnitude and presence of the Relative Age Effect within the KNVB."

Therefore, I hypothesise that introducing shifts in cut-off dates can lead to discernible alterations in the manifestation of the RAE. By crafting various shift scenarios that span over six months, I aim to investigate not only whether these changes influenced the RAE but also to what extent. This hypothesis also stands as a cornerstone for assessing the feasibility of practical interventions to mitigate the RAE's effects in the literature.

5. Methods

My first objective is to discern whether the RAE is present within my dataset and, if so, quantify its impact. This is done by the analyses of this dataset by making use of regression models that allows for a closer study of how important factors relate to one another. Regression models are mathematical tools that help to understand and predict how one factor changes when another factor changes.

To test the first hypothesis, I applied multiple simple linear regression models and a between-subjects regression model to the data, focusing on several key predictors: Player performance, team level and birth quartile. The birth quarter has been used as a proxy for the individual's relative age within their cohort.

The between regression in a panel data analysis is used to explore the relationship between variables that vary between entities over time. In a panel data setup, you typically have data for multiple entities (clubs and/or individuals) observed over multiple periods. The "between regression" focuses on the variation in the variables between these entities, ignoring the within-entity variation over time. Between variation refers to the variability observed when comparing different groups or categories in a dataset. These groups can represent distinct treatments, conditions, categories, or any other grouping criteria. The variation between groups captures the differences that exist among the groups being compared. The between regression always has an underlying formula of which it makes use. The forms are impacted by the number of variables. The following formula is the formula that applies to my analysis:

$$Yi=\beta0+\beta1Xi+ \beta2Zi +\epsilon i$$

Yi: represents the team level for entity **i** over all time periods in the sample.

Xi: is players' performance for entity **i** over all time periods in the sample.

Zi: is the birth quarter of an entity **i** over all time periods in the sample**.**

β0: is the intercept term, representing the average value of team level when the other variables are 0.

β1: is the coefficient of player performance

B2: is the coefficient of birth quarter

ϵi: is the error term, representing the unobserved factors affecting team level that are not accounted for in the model.

The second hypothesis focusses on simulations that shift the cut-off date in the data and represents the effects of the shift on the RAE. The first task is to transform these dates, categorizing individuals into specific birth quarters of the year. From this transformation, it is possible to perform a series of simulations. These were designed to understand the patterns and variances associated with birth months when they are adjusted or "shifted." This is then followed by the examination of a range of shifts, from six months ahead to six months behind, moving in two-month intervals. For every shift scenario, I execute the simulation multiple times, reshuffling and reassessing the dataset to capture diverse outcomes. This then leads to matching the original data against these varied, shifted scenarios, seeking to identify significant deviations or patterns that emerged from these adjustments. This hypothesis is mostly based on previous findings of the KNVB. The KNVB was confident that a shifting cut-off date could be a viable solution to the RAE. However, there has yet to be any paper to investigate the effects of the shifting cut-off date. This means that this part of the research is not based on previous research and is a new approach to tackle this multifaceted problem.

After the first simulation with a six-month forward and backward simulation that only accounts for one season, I proceed with other simulations that take multiple seasons into account. Subsequent simulations were conducted based on seasons, with each new season shifting the cut-off date by a pre-determined number of months. Two vital variables, player performance and team level, are introduced. As the seasons progressed, birth months were systematically adjusted, allowing for a deeper understanding of how the time of year might influence observed patterns.

Each simulation made use of a statistical model that calculated the impact of these shifts on the average birth month. This model aimed to identify possible connections between birth months, their adjustments, the performance metrics of players and team level across various shift scenarios, spanning from one to six months.

Lastly, this methodology was also applied specifically to "selection" tier teams in the dataset. By focusing on this elite subset, it was possible to discern if the patterns noticed across the broader dataset were consistent within these top-performing groups.

6.1 Regression results

In this section, the results are shown of the statistical models used to test for the RAE and show the simulation results of changing the cut-off date. The first objective is to look at the impact of the birth quarter and player performance on a team level. This is done by a regular Ordinary Least Squares (OLS) model as this allows for the testing of the assumptions. Subsequently, the variables are tested through tests fit for panel data. This is because the data consists of multiple observations of the same person over a selected amount of years. Among the various methodologies available for panel data analysis, the Between-Effects (BE) model offers a great option for a deeper analysis. By focusing on variations between individual units, rather than within them over time, the BE model shows patterns that other models might overlook.

I began by examining a simple linear regression without extra variables. I then progress to models that incorporate additional variables and this leads to the BE model, where I analyse the factors influencing the team level of players.

Table 3

Linear regression with dependent variable team level

					95% CI	
Variable	Coefficient	Standard error	T	*p*	*Lower bound*	*Upper bound*
Birth quarter						
2	.105776	.0158534	6.67	0.000	.0747035	.1368496
3	.0483	.0158408	3.05	0.002	.0172517	.0793483
4	.1163946	.0160551	7.25	0.000	.0849262	.1478631
Constant	1.744323	.0111724	156.13	0.000	1.722424	1.766221
Adj. R^2	0.0016					
Observations	40,000					

Note: A linear regression of team level on birth quarter

Table 3 displays the results of the linear regression model analysing the influence of birth quarter on Team level. The coefficient for players born in the second quarter is 0.1058, which is statistically significant at the 1% significance level ($p < 0.01$). This indicates that players born in the second quarter have a team level that is higher by 0.1058 units in comparison to those born in the first quarter, holding all other factors constant, with higher teams being seen as worse. Players born in the third quarter exhibit a higher team level with a coefficient of 0.0483 and players born in the fourth quarter exhibit a higher Team level effect with a coefficient of 0.1164. This suggests that, compared to players born in the first quarter, those born in the third quarter have a Team level that is higher by 0.0483 units and those born in the fourth quarter have a Team level that is elevated by 0.1164 units, with all other variables held constant. This relationship is statistically significant at the 1% significance level ($p < 0.01$). All birth quarters, when compared to the first, show a significant impact leading to a higher team level, which is considered as worse in sports. This suggests the presence of RAE at a preliminary level.

After analysing the influence of birth quarters on team levels, the focus now shifts to understanding player selection. Does the RAE manifest itself in the likelihood of an athlete's selection? To investigate this, an logistic regression is used, a method suited for binary outcomes like 'selection' or 'no selection' team. This analysis aims to determine if individuals from the first team are more frequently selected for “selection” teams.

Table 4

Logistic regression with selection as dependent variable

					95% CI	
Variable	Odds ratio	Standard error	z	*p*	*Lower bound*	*Upper bound*
Birth quarter						
2	.8371193	.0252548	-5.89	0.000	.7890558	.8881104
3	.8260468	.0248757	-6.35	0.000	.7787023	.8762698
4	.693639	.0209095	-12.13	0.000	.6538442	.7358557
Constant	2.340746	.0505685	39.37	0.000	2.243702	2.441986
Adj. R^2	0.0029					
Observations	40,000					

Note: A logistic regression of player performance on the binary variable of selection team

Table 4 provides the outcomes from a logistic regression model, revealing how birth quarter influences the odds of a player being selected for a team. These odds are presented as ratios, where a value greater than 1 indicates an increased likelihood of selection relative to the reference category (in this instance, the first birth quarter), while a value less than 1 indicates a decreased likelihood.

Players born in the second quarter have an odds ratio of 0.8371193. This suggests that they are approximately 16.3% (or 1 - 0.8371193) less likely to be part of the selection team when compared to players born in the first quarter, assuming all other variables are held constant. This relationship is statistically significant at the 1% level ($p < 0.01$). Players born in the third quarter are around 17.4% less likely to be chosen for the selection team relative to those born in the first quarter, with all other factors being equal. This relationship is statistically significant at the 1% level ($p < 0.01$).

Players born in the fourth quarter are around 30.6% less likely to be chosen for the selection team relative to those born in the first quarter, with all other factors being equal. This relationship is statistically significant at the 1% level ($p < 0.01$).

After examining the basic linear and logistic regression using only the birth quarter as a variable, it is important to evolve the analysis by adding other variables. In the next linear regression, player performance is also added, offering a more comprehensive understanding of the factors at play.

Table 5

Linear regression with dependent variable team level

					95% CI	
Variable	Coefficient	Standard error	T	*p*	*Lower bound*	*Upper bound*
Player performance	-.0022663	.0000205	-110.41	0.000	-.0023065	-.0022261
Birth quarter						
2	.0812828	.0138806	5.86	0.000	.0540765	.1084891
3	.0108619	.0138719	0.78	0.434	-.0163274	.0380512
4	.0032423	.0140927	0.23	0.818	-.0243798	.0308644
Constant	5.352131	.0341077	156.92	0.000	5.28528	5.418983
Adj. R^2	0.2349					
Observations	40,000					

Note: A linear regression of team level on player performance and birth quarter

Table 5 displays the results of the linear regression model. The coefficient for player performance is −0.0023, which is statistically significant at a 1% significance level ($p < 0.01$). This indicates that for every 100-unit increase in Player performance, the expected team level diminishes by 0.23, on average, with all other factors being held constant. The table also shows the influence of being born in a specific quarter with the first quarter serving as the reference period. Players born in the second quarter exhibit a higher team level with a coefficient of 0.0813. This implies that players born in the second quarter have a team level that is higher by 0.0813 units in comparison to those active in the first quarter while keeping other variables constant. This relationship is statistically significant at a 1%

significance level. The third and fourth quarters do not present any statistically significant association with team performance.

To delve deeper, another logistic regression is conducted, now including the player performance variable. This analysis aims to determine if individuals from the first team are more frequently selected for “selection” teams.

Table 6

Logistic regression with selection as dependent variable

Variable	Odds ratio	Standard error	z	*p*	95% CI	
					Lower bound	*Upper bound*
Player performance	1.007113	.0000788	90.57	0.000	1.006959	1.007267
Birth quarter						
2	.845405	.0305501	-4.65	0.000	.7875992	.9074535
3	.8988795	.0327252	-2.93	0.003	.8369741	.9653635
4	.8411029	.0304194	-4.78	0.000	.783546	.9028877
Constant	.0000448	5.37e-06	-83.54	0.000	.0000354	.0000567
Adj. R^2	0.2718					
Observations	40,000					

Note: A logistic regression of player performance and quarter on the binary variable of selection team

Table 6 provides the outcomes from a logistic regression model, revealing how birth quarter influences the odds of a player being selected for a team. These odds are presented as ratios, where a value greater than 1 indicates an increased likelihood of selection.

The coefficient for player performance is an odds ratio of 1.007113, which is statistically significant at the 1% level ($p < 0.01$). This means that for every 1-unit increase in player performance, the odds of being in the selection team increase by 0.7113% (or equivalently, multiply by 1.007113), holding all else constant.

Players born in the second quarter have an odds ratio of 0.845405. This indicates that they are about 15.5% (or 1 - 0.845405) less likely to be in the selection team as compared to players born in the first quarter, keeping everything else the same. This relationship is statistically significant at the 1% level ($p < 0.01$). Players born in the third quarter are approximately 10.1% less likely to be selected for the team relative to those born in the first quarter and players born in the fourth quarter born in the fourth quarter are approximately 15.9% less likely to be selected for the team relative to those born in the first quarter. This relationship is also statistically significant at the 1% level ($p < 0.01$). This logistic regression confirms the results of the first regression in Table 3, showing that people born outside of the first quarter, are less likely to participate in selection teams.

Following the initial linear regression analysis, I recognized the potential for unobserved heterogeneity within the data. This refers to the unmeasured differences between teams that might be constant over time and could influence the outcome. Ignoring such heterogeneity could result in biased and inconsistent estimates. There are 2 ways to identify this heterogeneity, graphically and through a test. The Breusch–Pagan/Cook–Weisberg test is a method to assess heteroskedasticity in the residuals of a regression model. This violates one of the assumptions of the linear regression model and can invalidate the usual t- and F-statistics that are derived from the model.

Table 7

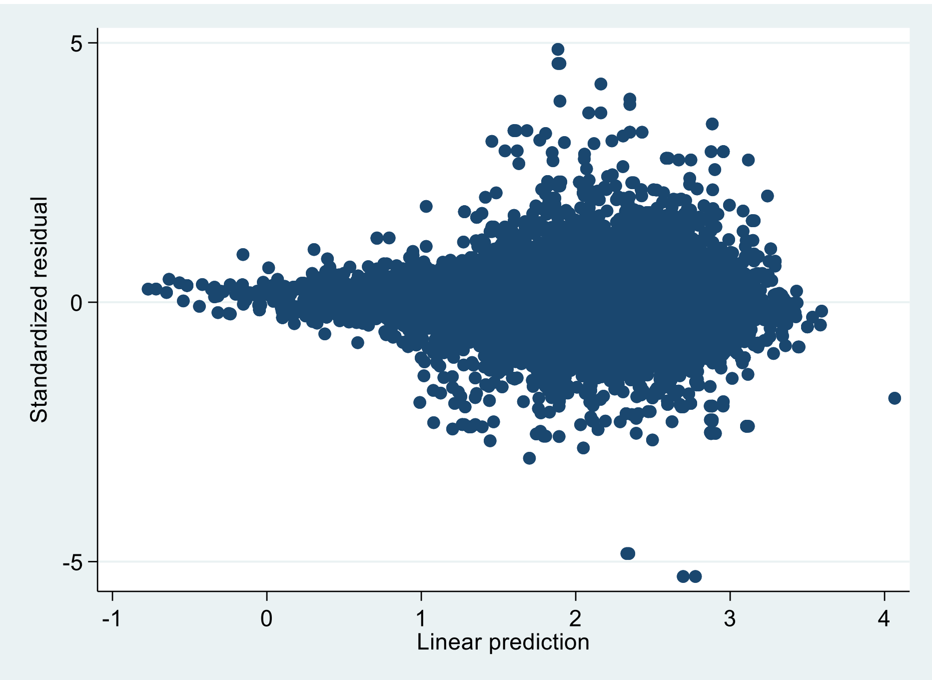
Breusch–Pagan/Cook–Weisberg test for heteroskedasticity

H0: Constant variance		
Chi-Square (χ2)	=	5753.72
Prob > chi2	=	.000

Note: Breusch–Pagan/Cook–Weisberg test for heteroskedasticity. H0: Constant variance

Figure 3

Figure that presents the results of a test on homoscedasticity



The Breusch–Pagan/Cook–Weisberg test returned a significant p-value at the 1% significance level ($p < 0.01$), indicating the presence of heteroskedasticity in the residuals. This implies that the variances of the errors are not constant across levels of the fitted values, and this can distort the standard errors of the regression coefficients, potentially leading to incorrect conclusions about the significance of predictors. Moreover, the shape of the graph is consistent with the presence of heteroskedasticity. The 'fan shape' is a classic visualization of heteroskedasticity: as the predicted values (on the x-axis) increase, the range or spread of residuals (on the y-axis) becomes wider. This fan shape visually confirms what the Breusch–Pagan/Cook–Weisberg test statistically indicated.

This model only uses between-unit variation. It averages data over time and runs a regression on these averages. The between estimator does not solve for unobserved individuals but is only affected by heterogeneity when there is time-invariant heterogeneity that is correlated with the independent variables, the between estimator can be biased. Because the variables are mostly time-variant, except for birth quarter, in the regression, the between estimator should help to mitigate the problem. To account for this, I transitioned to a Between Effects (BE) model using the xtreg command in Stata. This approach allows for the decomposition of variability into within and between variations, and by focusing on the between variation, the BE model provides robust estimates when considering the influence of time-invariant unobserved characteristics of the teams. This transition offers a more nuanced understanding of the relationships and ensures the robustness of the findings.

Table 8

The results of the between regression

					95% CI	
Variable	Coefficient	Standard error	T	*p*	*Lower bound*	*Upper bound*
Player performance	-.0024244	.0000562	-43.17	0.000	-.0025345	-.0023143
Birth quarter						
2	.0792	.0336322	2.35	0.019	.0132661	.1451339
3	.0079631	.0336127	0.24	0.813	-.0579326	.0738588
4	-.0049468	.0341699	-0.14	0.885	-.0719348	.0620412
Constant	5.604173	.0924886	60.59	0.000	5.422855	5.785491
Adj. R^2	0.2468					
Observations	20,000					

Note: This table shows the regression results of a between regression of the quarter in which someone is born and player performance on team level

Table 8 shows the results of the BE model. The coefficient for Player performance is -0.0024, which is statistically significant at a 1% significance level ($p < 0.01$). This suggests that for a 100-unit increase in player performance, the expected Team level decreases by 0.24, on average, holding all other factors constant. The table also shows the effect of being born in a specific quarter. The first quarter is used as the reference quarter. The second quarter showed a positive relationship with team performance, with a coefficient of 0.08. This suggests that teams with players born in the second quarter of the year play at a higher team level by 0.08 units, compared to those born in the first quarter, holding other factors constant. This effect is significant at a 5% significance level. However, the third and fourth quarters did not show any statistically significant relationship with team performance.

6.2 Simulation results

After doing panel data regressions to determine the effect of player performance and birth quarter on team level, I proceeded to simulate the most optimal cut-off to create teams through a simulation. This is done as follows: First, a total of 1000 simulations were conducted and the same player can be used more than once in a simulation and some players will not be used at all in some simulations. In each simulation, the players were randomized within the existing teams, and the average RAE (mean relative age effect) for each team was calculated. This randomization process simulates a situation where the birth quarters are distributed randomly across teams, irrespective of the team's composition or any potential biases. It creates an optimal scenario where the birth quarters are equally represented. This process was repeated a thousand times for each team, providing insight into possible outcomes for every individual team. The overall mean and standard deviation of these team effects across all teams were then computed. This is important as it shows the relative birth quarter mean that should be targeted in the following sections.

Table 9

Partial results of 1000 random simulations

Team number	Mean effect	Standard deviation
1	2.43	0.97
2	2.43	0.96
3	2.43	0.96
4	2.43	0.97
5	2.43	0.97
6	2.43	0.96

Note: This table shows a part of all the simulations that have been run and the corresponding mean birth quarter and the standard deviation per team. These are randomly created from actual observations.

When examining the overall simulation results in Table 9 the mean effect and its standard deviation across all simulations remained fairly consistent. The numbers within the table might not be self-explanatory, but what the simulation results tell us, is that the mean effect is 2.43. This implies that the average birth month of a team should be around early June in the best-case scenario when all teams are created randomly. These results represent the distribution of the mean relative age effect across all teams. The consistency of the mean effects and their standard deviations across simulations suggests that the RAE is somewhat present, even under conditions of randomized birth quarter assignments. This is largely because there are slightly more people born in the first 2 quarters in the data set. While this might not be a significant difference in general, it skews these results from being exactly in the middle, July 1st.

Yet, simulations of the optimal mean birth quarter are only half the story. I then calculated the actual relative age effect for each team based on the original dataset. This is done through a t-test to statistically test if there is a significant difference between the optimal simulations and the observations. If there is a significant difference, there would be an opportunity to try and close the gap between the optimal randomised scenario and the current situation. The same 1000 simulations were used for the simulated mean effect.

Table 10

Paired t-test of the actual mean effect and the simulated mean effect

Statistic	t	df	p-value	95% CI [Lower	95% CI [Upper]	Mean Difference
Value	2.6	7320	0.00926	0.00233	0.000328	0.00133

Note: This table shows the results of the paired t-test to see if there is a difference between the actual mean effect and the simulated mean effect

Table 10 shows the results from the paired t-test. The results suggest that there is a significant difference between the actual mean effect and the simulated mean effect at a 1% significance level ($p < 0.01$). Given this information, I reject the null hypothesis that the true mean difference is equal to zero and is different from the simulated results. The positive mean difference suggests that the simulated average birth month would be higher than that in the actual data, meaning teams in the actual data tend to have players born earlier in the year. This implies that teams predominantly consist of individuals born closer to the cut-off date, resulting in a potential underrepresentation of those born later in the year.

6.3 Simulation results of the cut-off date shift

Because of the previous section that showed the significant difference between the optimal randomized scenario and the current situation, there is an opportunity to minimise this difference. The following step was the first step towards checking whether it is possible to minimize the existing RAE. While the literature suggests multiple solutions, the majority are very hard to theoretically test without a real test group. However, shifting the cut-off date is a theoretical practice recommended by the KNVB that is possible to test through simulations.

The code sets up a series of simulations, which, in simpler terms, are "what if" scenarios. Each simulation explores how changes in birth months might ripple through the dataset and affect overall outcomes. In essence, the code is an in-depth exploration of a simple question: If I shift players' birth months, how might that influence their journey or standing in the sports world? For instance, given a shift value of one, an athlete originally born in January, who is currently the oldest player, would now be recalibrated to be the youngest player. Upon implementing these shifts, the next phase is to group the athletes. They're grouped by the parameters of team level and the season they participate in.

Here, the 'season' used boils down to the maximum number of players engaged in the same age category of a team. This suggests that among possible age groups like under 15, 16, and under 17, the category housing the majority of players is selected. Additionally, one critical aspect is that while the birth months are altered, athletes maintain their original team level. This would imply that a player who plays in team level 1, would not shift to another team level after the cut-off date shift.

With the birth months now adjusted under these restrictions, the mean birth month for each of these groupings is calculated. This offers an insight into the birth month distributions across teams post-shift. This is done through a linear regression that tries to minimise the RAE, in this case, make the teams as balanced as possible. Although it is impossible to change the actual birth months of players, the altered positions grant an opportunity to predict potential team dynamics had birth months been different. For each unique shift scenario, I constructed a corresponding regression model to gauge how these birth month adjustments might influence team metrics. The results give an insight into whether modifying age groupings might level the playing field. This analysis is a new approach to addressing the RAE issue and the simulations conducted in this study are not based on prior research.

The initial analysis underscores the influence of altering players' birth months on RAE within teams for a single season. A single-season shift implies that with a 2-month change, a February-born player would now rank as the youngest, whereas a shift of -2 months would position a November-born player as the youngest. The range for these shifts spans from -6 to +6 months, progressing in 2-month intervals. A 0-month shift serves as a measure of consistency.

Table 11

Simulation effects on the team birth month for different shifts with a range of -6 months and +6 months

Months shifted	p-value	Mean difference
Shifted by -6 months	0.000	-0.115
Shifted by -4 months	0.001	-0.0574
Shifted by -2 months	0.002	-0.0421
Shifted by 0 months	0.185	0.000396
Shifted by 2 months	0.000	-0.0597
Shifted by 4 months	0.000	-0.101
Shifted by 6 months	0.000	-0.114

Note: These are the results of the effect of shifts in the cut-off dates in increments of 2 months. The range of the months is from -6 months to +6 months with 0 months as a control variable

Table 11 shows the results of the simulations that test the effects for different shifts with a range of -6 months and +6 months in a singular year. The result that is analysed first is the -6-month shift, however, the same logic can be applied to the other months. The shift of -6 months shows that the mean birth quarter of teams was significantly lower than the original with a mean difference of -0.115. This effect is statistically significant at a 1% significance level ($p < 0.01$). The effect appears to be significant at a 1% significance for every shift that has been measured, but the mean difference appears to change depending on the number of months shifted from the original cut-off date. The negative shifts by 4 months and 2 months also resulted in a lower mean birth quarter with a mean difference of -0.057 and -0.042. Additionally, when there is a shift that moves the cut-off date forward by 2 months, 4 months and 6 months, the mean birth quarters became significantly lower with a mean difference change of -0.060, -0.101 and -0.114. In general, no matter the shift, it always results in a negative and significant change in this simulation. This negative effect suggests that players would be born earlier in the year and would thus become older as a consequence. This is the opposite of what is trying to be achieved.

As a robustness check, the simulation was also run on a shift of 0 months. When no shift was applied, there was no significant difference between the original and the shifted mean birth quarters at a 10% significance level ($p > 0.1$). This is to be expected as the original should not statistically change, as no change has been applied.

These findings indicate that both forward and backward shifts in the birth months of players substantially affect the RAE in the same way in teams. This suggests that even minor adjustments to the cut-off dates for age grouping can potentially influence the RAE, underscoring the importance of considering birth dates in team formation and talent identification.

However, as mentioned before, the shift itself only creates a small change in the RAE and there are likely to be some variables that have been excluded. Because of this, I ran a simulation that also investigates how the RAE is affected by the skill of a player and the level of team play through a linear regression with the team birth month as a dependent variable. This linear regression uses data from previous seasons to predict the coming four seasons. The same selection criteria and restrictions are applicable as with the previous simulation. This adjustment is performed on multiple shifts. The shifts that are measured, go from a 1-month shift to a 6-month shift. Instead of a single shift, multiple shifts of birth months were tested over several seasons, creating a spectrum of scenarios. This means that a shift of 2 months would lead to a combined 6-month shift after 3 seasons. Each shift led to its own simulation, which was then analysed using a statistical model to discern patterns or insights. In this simulation, the players are shifted to teams they have played in before. The data is then resampled with a replacement for a specified number of simulations, in this case, 100 simulations. Additionally, the average player performance and number of people who are shifted per shift are shown.

Table 12

Simulation effects for a shift of 1 month with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	-2.15e-17	1.23e-17	-1.75	0.08
Team level	2.71e-15	2.57e-15	1.06	0.291
Constant	6.49	2.2e-14	2.94e+14	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1561	402
2	1539	385
3	1534	411
4	1542	417

Note: This table shows the results of the simulation on the RAE for a shift of 1 month. The simulation includes the variables player performance and team level

Table 12 outlines the findings from the regression models. The main observations from the table are as follows:

The intercept is approximately 6.49. This tells us that the mean birth month of players is halfway through the 6th month when all predictor variables are zero. This holds for every shift.

Player performance influences the average team birth month in an interesting way. The results show a minimal negative influence on birth month. Specifically, for every 100-unit augmentation in the player performance score, the team birth month decreases, albeit slightly. It's imperative to keep in mind that this observation remains valid while keeping all other influencing factors unchanged. The significance of this relationship is endorsed at a 10% threshold ($p < 0.1$). Drawing from this, one could infer that players showcasing better performance tend to have their birthdays in the initial months of the year and this creates a team that is relatively older, even older than the original data set, in the case of a 1-month shift. Implementing this shift would produce the opposite outcome of what was intended.

The variable team level measures whether a player playing at a higher team level would be made up of players older or younger. The variable mean team level has a very small positive effect on the average team birth month, indicating that higher-level teams, which are worse, are made up of relatively younger players than in the original data set. However, this relationship is insignificant at a 10% significance level ($p > 0.1$) and therefore the results cannot be interpreted.

When observing the results of all the other shifts, they mostly show the same pattern as the model where only 1 month is shifted. All models show a significant intercept at a 1% level ($p < 0.01$), indicating that there's a significant mean adjusted birth month even when player performance and team level are zero.

Reviewing the other shifts, the player performance predictor shows a significance level of 10% in each model, suggesting that it might be related to the mean adjusted birth month. The effect size, however, is very small as indicated by the estimate close to zero for every observation. What is very interesting, is that the sign changes when looking at the 2-month and 3-month shifts. For those months it would mean that the implementation would be effective as higher-rated players would be younger players.

Conversely, the team-level predictor does not show a significant effect on the mean adjusted birth month for all shifts, as indicated by the relatively high p-values and thus being insignificant at a 10% significance level ($p > 0.1$). It can be noted that the sign changes when looking at the 2-month and 3-month shifts. Table 12 also shows the results on the average player performance. The results suggest that while the number of players shifted varies slightly with the number of shifts, the average player performance does not exhibit a linear trend. This is the case for all shifts, with the total average player performance ranging from 1544 to 1558 with the shift of 5 and 6 months both reaching 1558.

In summary, while certain patterns emerge, the data suggests minimal impact from the team level and a minor influence from player performance on the mean adjusted birth month across varied shift frameworks. However, there's a discernible discrepancy in average player performance among the shifts. Not all teams are homogeneous. Elite teams are disproportionately influenced by the RAE (Peña-González et al., 2020). Hence, it's intriguing to contemplate how divergent results might manifest when analysing a narrower sample where player performance holds greater sway. In light of this, analogous simulations and examinations were reiterated, but focusing solely on selection teams.

Table 13

Simulation effects for a shift of 1 month for players in the "selection" category with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	2.7e-17	9.82e-18	2.75	0.006
Team level	5.89e-15	4.97e-15	1.18	0.236
Constant	6.48	1.89e-14	3.43e+14	0

Number of shifts	Average Player performance	Number of players shifted
1	1670	235
2	1630	249
3	1603	368
4	1614	287

Note: This table shows the results of the simulation on the RAE for a shift of 1 month for players in the "selection" category. The simulation includes the variables player performance and team level

Table 13 displays the results of the output of the multiple linear regression analysis performed on data for only the people who play in “selection” teams. The main observations from the table are as follows:

The intercept of the results is approximately 6.49. This has not changed when compared to the original simulation results. This means that the intercept remains roughly the same, no matter the shift.

In this context, player performance is more closely associated with the average team birth month as the significance level is at 1% in this model ($p < 0.01$). The effect size, however, is still very small as indicated by the estimate close to zero for every observation. An intriguing observation surfaces when you compare these results to the results where every player was considered. The sign of the player performance variable has reversed. This implies that in the context of the 'selection' category, for every 100-unit enhancement in the player performance score and with a shift of a month, players showcasing better performance tend to have their birthdays in the later months of the year and this creates a team that is relatively younger, even younger than the original data set, in the case of a 1-month shift. Implementing this shift would produce the intended outcome. Notably, the second and third-month shifts again present an anomaly with a negative sign for player performance, setting them apart from the trend seen in other months. This is an interesting phenomenon that is tough to explain with the current date set.

The team-level predictor again shows a significant effect on the mean adjusted birth month for all shifts at a 10% significance level ($p < 0.236$). Although I cannot interpret these results, it is interesting to see the same sign change when looking at the 2-month and 3-month shifts.

Table 13 also shows the results on the average player performance. The results again suggest that while the number of players shifted varies slightly with the number of shifts, the average player performance does not exhibit a linear trend. This is the case for all shifts, with the total average player performance ranging from 1629 to 1687. In this case, the difference is larger and the shift of 6 months leads to the highest average player performance. In this case, choosing the higher average player performance would also increase the average birth month of the selection teams, creating younger selection teams.

7. Discussion

This study aimed to investigate if there is an RAE in the KNVB and if it was possible to minimise the RAE within the KNVB. The investigation of the RAE was done through a between-estimation. The minimisation of the RAE was done through the two factors of player performance and team level, and their effect on the mean adjusted birth month among soccer players across different shift scenarios. The results of the simulations indicate a modest but significant impact of player performance on the adjusted birth month, with no substantial effect on a team level. This finding is critical and yields meaningful insights in the context of RAE in sports, a well-documented phenomenon with ramifications in developmental opportunities and performance outcomes.

The regression results show an interesting development. The primary results of looking at the effect of the birth quarter on the team level shows that people born in the first quarter are more often in higher teams and selection teams. When looking at the impact of the birth quarter and player performance on the team level, a significant negative effect of player performance and a significant effect for the second birth quarter can be observed. However, when you observe the odds of being in the selection team with the same variables, they are all significant and being born in the first quarter and having higher player performance tends to increase the odds of being in the selection team.

Following the regression results, I ran multiple simulations to test the effect of shifting the cut-off date and the effect of player performance and team level on the mean birth month. Player performance was found to be significantly related to the mean adjusted birth month, although the size of the effect was rather small. The findings also align with previous literature highlighting the influence of individual characteristics, such as skill level and talent, on RAE (Cobley et al., 2009). This study provides further empirical evidence supporting the contention that the RAE is not merely a product of environmental factors, but is influenced by individual characteristics such as a player's skill level.

The simulations show that the effect of team level is insignificant when looking at the average team birth month. One thing that is very interesting for both the results of player performance and team level (even though this one is insignificant), is the sign change for the shifts of 2-month and 3-month. This pattern might imply an interaction effect between the size of the shift and the player's skill level. It might be that more skilled players adapt to changes in their relative age more effectively. Alternatively, it could suggest that the impact of skill level on the RAE is context-dependent, with different influences under different conditions. This is an area that warrants further investigation in future research. However, because most of the decisions of the KNVB will be mainly based on higher teams, it is also interesting to look at the relationship between player performance and team level for players in high-level teams. Observing these results, they show that the signs are completely the opposite of the cut-off date shift for all teams. The sign change for the shifts of 2-month and 3-month remains and would thus reinforce that there is a significant difference between those 2 shifts and the other shifts. This would be interesting to research for the exact reason. Furthermore, the difference between ”selection” teams and the other teams also suggests that there should be a consideration about shifting cut-off dates affecting different team levels differently.

But what effect does this study have and what would the implications be? The study's results could have significant implications for policymakers and possible implementation. First of all, this is the first time research like this has been performed with numerical results that show that it could be interesting to apply the shifting cut-off date. Additionally, since player performance has a significant effect, emphasis should be on nurturing individual player skills that have nothing to do with age attributes, considering it as a key factor in selection processes. Given the potential interaction between shift size and skill level, further consideration should be given to the age grouping practices. Policymakers and practitioners should remain alert to the possibility of age groupings inadvertently disadvantaging younger players, especially in selection teams.

When it comes to recommending which shift to use to minimize the RAE, the results suggest that there isn't a clear shift that stands out as being substantially better than the others. However, a shift that could be interesting for the selection teams, would be the 1,2,4, 5 and 6-month shifts. This is because the effect of player performance is positive, meaning the team would become younger with increased skill, balancing out the current overrepresentation of people being born in the first quarter. When choosing between the months, the average mean player performance of the shift of 5 and 6 months seems to increase the average player performance the most. This would result in even more younger players being part of the selection teams due to the positive coefficient of player performance. However, the decision on which shift to use may not solely depend on the results of these models. Other factors might need to be taken into consideration. For example, practical implications, resources available, the specific context of the program, or further research might lead you towards preferring one shift over another.

While statistical analysis is an important part of the decision-making process, the real-world implications and practicality should also be considered. This has been the main problem with researching the RAE as it consumes a lot of resources to implement without a guaranteed and direct result. This is also the reason why there are so few practical solutions in the literature (Webdale et al., 2019). The other solutions face the same problems, like age grouping, adjustments in coaching evaluations, or providing equal opportunities for participation and development regardless of birth month. Additionally, we have to remember that the shifts were arbitrary increments for the purpose of this analysis. They might not represent realistic or feasible changes in the cut-off dates in a specific context. More nuanced adjustments might be necessary to effectively address the RAE. To achieve a more definitive answer on which, if you would implement one at all, shift to use, it would be beneficial to further investigate other potentially influential variables not included in these models, examine potential interaction effects between predictors, or carry out more in-depth analyses on the shifts themselves, such as analysing potential non-linear effects or categorizing shifts differently. The goal is not just to statistically balance out the RAE, but to create an environment where individuals are given equal opportunities to succeed, regardless of their birth month. Therefore, changes should always be implemented with an understanding of the potential impacts on the individuals and the system as a whole. Coaches could be encouraged to adopt a player-centred approach, focusing on individual needs, skill development, and growth rather than solely relying on age-related factors, but till now, this has proven to be difficult. Moreover, it could also help to foster a supportive and inclusive environment within youth soccer organizations. This includes promoting open communication, collaboration, and transparency among coaches, parents, and players. Establishing clear policies and guidelines regarding player selection, development, and competition can help create a level playing field and reduce the potential for bias associated with RAE. This however should also be tested and looked into in other papers. Furthermore, ongoing research and evaluation of the implemented solutions are essential to assess their effectiveness and identify any unintended consequences.

Several limitations can be considered when interpreting the results of the study. First of all, the data that was used, was limited to only the Dutch soccer association. This means that the results are only applicable to the Dutch soccer league. Furthermore, all people observed were born within 365 days of each other. This will lead to both an advantage and a limitation. The advantage is that there are most likely no external shocks that impact the start of soccer players within a certain period, like the coronavirus for example. However, because the birthdays are limited to those 365 days, it is again hard to generalise the results. Another limitation is that the data was very top-heavy. Close to 79% of the players were either in the first or the second team of a club, while not even 1 per cent made up the lower teams, from 7 to 10. Because of this, it will be difficult to apply all the results to a very small sample size in the lower teams. However, this is not a very large limitation, as most of the time the RAE is more prevalent and impactful in the higher levels of play, rather than the lower levels of play (Peña-González et al., 2020).

Last but not least, the simulations are made with very little information. First of all, I had to set several restrictions to set up the simulations, like players only being able to shift towards their own team level. Additionally, the variable of team level only says the current level of team play and not the previous team level. Also, the variable of player performance is most likely correlated with the team level. Most of the time the RAE is most impactful at a young age and tends to decrease in importance when you get older (González-Víllora et al., 2015). Afterwards, it is very hard to reverse the damage that has been done. Most higher teams are given better coaches, trainers and other special advantages that tend to stack up and players also start to get used to these advantages, while relatively younger players are disheartened by the lack of attention (Relative Age Effect in Youth Soccer, 2019). Because of this, it can be very hard for the players' performance to account for all these different factors and it would thus be better if this index would be applied to younger children. It would also be more useful to unpack the index as much as possible to get a true representation of the skill of a player, however, this can lead to subjectivity issues and other biases.

8. Conclusion

The research conducted in this thesis focused on exploring the influence of the RAE in varying team scenarios. The analysis indicated a subtle but statistically significant correlation between player performance and mean adjusted birth month across different shift scenarios, while team level appeared to have a limited effect.

The first research question was whether or not there was an RAE within the Dutch soccer association. My findings would support this hypothesis as there is a significant difference between the team level of people born in the first quarter and people born in other quarters, where people born outside the first quarter are on average in higher teams and are also less likely to be in the selection teams. Because of this, I would say that the results are consistent with the hypothesis that the RAE significantly influences the outcomes within the Dutch Soccer Association (KNVB), with individuals born closer to the cut-off date experiencing more negative outcomes compared to individuals born further from the cut-off date.

The second hypothesis asks whether or not there is a possibility to minimise the magnitude of the RAE through shifting cut-off dates. The results would support this hypothesis as shifting the grouping cut-off dates (evidenced by the shift scenarios) did impact the manifestation of the RAE, evidenced by variations in mean team birth month across different shift scenarios. However, the effect sizes were small, implying that while the cut-off shifts did alter the RAE's presence, the impact's magnitude wasn't substantial. This aligns with the notion that the RAE is a multifaceted phenomenon, influenced by a large number of factors beyond merely the grouping cut-off dates. This would also once again bring up the notion to see whether it would be worth the administration to implement the given solutions or whether there are other ways to tackle the problem.

This study contributes to the growing body of research on the RAE, an area of study which has critical implications for talent identification, development, and team selection processes in various fields. The main aspect that differentiates this paper from other papers, is the in-depth analysis into what can be done through the shifting of cut-off dates instead of the speculation of the effects. Additionally, by delving into relatively unexplored factors such as player performance on the shifting of the cut-off date, the research challenges some assumptions about the RAE and underscores the complexity of this phenomenon, highlighting the multitude of factors that can potentially influence it. Although this paper is unable to give complete advice on what options should be taken to solve the RAE, it could serve as a baseline to explore the option of shifting cut-off dates.

The future direction for this research could take several forms. The first recommendation for future research would most likely be to extend the data set. The data set contained a total of 40,000 observations set over 8 time periods, thus containing 5000 unique players. These players were selected as they were born within the same 365 days and remained at the same club during the data observation. However, it would be very interesting to extend the research to more years with more people.

Additionally, the inclusion of other potential predictors and the exploration of interactions among predictors could lead to a more comprehensive understanding of the factors influencing the mean-adjusted birth month. The variable of Player performance has now been used as an indicator for a lot of variables of a player, speed, height, skill and much more. Including all these different variables would give more accurate and clear results to interpret and would maybe lead to new ways to minimalize or optimise the current solutions to the RAE.

Last but not least, it would also be interesting to look at the long-term effects of the current policies, especially at the higher level. It would be interesting to see if people who are born earlier in the year, are more likely to be selected for national teams, are paid more or receive more play time in comparison to their “younger counterparts”. Additionally, you can also look at the differences between positions as they tend to require different skill sets and physiques.

Reflecting on the paper, it is clear that understanding and solving the RAE is far from simple as shown by previous research. There are a lot of different variables that could impact the RAE and in this paper, the subtle but significant influence of Player performance underscores the potential multifaceted nature of the RAE. This research serves as a stepping stone towards a deeper understanding of the RAE, encouraging future exploration and discussion. The study could serve as a potentially impactful piece of information on how to influence team selection and talent development strategies that will allow the sport to foster a more competitive and fair playing ground through the understanding of the RAE's complexities.

References

Acar, S. (2011). Matthew, Pygmalion, and Founder Effects. In *Elsevier eBooks* (pp. 75–81). https://doi.org/10.1016/b978-0-12-375038-9.00141-2

Ando, S., Usami, S., Matsubayashi, T., Ueda, M., Koike, S., Yamasaki, S., Fujikawa, S., Sasaki, T., Hiraiwa-Hasegawa, M., Patton, G. C., Kasai, K., & Nishida, A. (2019). Age relative to school class peers and emotional well-being in 10-year-olds. *PLOS ONE*, *14*(3), e0214359. https://doi.org/10.1371/journal.pone.0214359

Barnsley, R. H., & Thompson, A. S. (1988). Birthdate and success in minor hockey: The key to the NHL. *Canadian Journal of Behavioural Science*, *20*(2), 167–176. https://doi.org/10.1037/h0079927

Barnsley, R. H., Thompson, A. S., & Legault, P. (1992). Family Planning: Football style. The relative age effect in football. *International Review for the Sociology of Sport*, *27*(1), 77–87. https://doi.org/10.1177/101269029202700105

Bedard, K., & Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *Quarterly Journal of Economics*, *121*(4), 1437–1472. https://doi.org/10.1093/qje/121.4.1437

Bosma, M., & Vuegen, J. (2020). *Documentatie eQuality Project*. Scisports.

ChatGPT. (n.d.). ChatGPT. Retrieved June 1, 2023, from https://chat.openai.com/

Chen, M. H., Huang, K., Hsu, J. W., Tsai, S., Su, T. P., Chen, T. J., & Bai, Y. M. (2021). Effect of relative age on childhood mental health: A cohort of 9,548,393 children and adolescents. *Acta Psychiatrica Scandinavica*, *144*(2), 168–177. https://doi.org/10.1111/acps.13327

Cobley, S., Baker, J., Wattie, N., & McKenna, J. (2009). Annual Age-Grouping and Athlete Development. *Sports Medicine*, *39*(3), 235–256. https://doi.org/10.2165/00007256-200939030-00005

Collins, D., & MacNamara, Á. (2012). The Rocky Road to the top. *Sports Medicine*, *42*(11), 907–914. https://doi.org/10.1007/bf03262302

Crawford, C., Dearden, L., & Greaves, E. (2013). *When you are born matters: evidence for England*. https://doi.org/10.1920/re.ifs.2013.0080

Delorme, N., Boiché, J., & Raspaud, M. (2010). Relative age and dropout in French male soccer. *Journal of Sports Sciences*, *28*(7), 717–722. https://doi.org/10.1080/02640411003663276

Dhuey, E., & Lipscomb, S. (2008). What makes a leader? Relative age and high school leadership. *Economics of Education Review*, *27*(2), 173–183. https://doi.org/10.1016/j.econedurev.2006.08.005

Diamond, G. H. (1983). The birthdate effect. *Journal of Learning Disabilities*, *16*(3), 161–164. https://doi.org/10.1177/002221948301600306

Edgar, S., & O'Donoghue, P. (2005). Season of birth distribution of elite tennis players. *Journal of Sports Sciences*, *23*(10), 1013–1020. https://doi.org/10.1080/02640410400021468

González-Víllora, S., Vicedo, J. C. P., & Cordente, D. (2015). Relative age effect in UEFA Championship Soccer players. *Journal of Human Kinetics*, *47*(1), 237–248. https://doi.org/10.1515/hukin-2015-0079

Green, T. (2023). *RELATIVE AGE EFFECT*. Scienceforsport. https://www.scienceforsport.com/relative-age-effect/

Helsen, W., Van Winckel, J., & Williams, A. M. (2005). The relative age effect in youth soccer across Europe. *Journal of Sports Sciences*, *23*(6), 629–636. https://doi.org/10.1080/02640410400021310

Hurley, W. G., Lior, D., & Tracze, S. (2001). A proposal to reduce the age discrimination in Canadian minor hockey. *Canadian Public Policy-analyse De Politiques*, *27*(1), 65. https://doi.org/10.2307/3552374

Işın, A. (2021). THE RELATIVE AGE EFFECT IN SUCCESSFUL NATIONAL FOOTBALL TEAMS. *Kinesiologia Slovenica*, *27*(2), 40–51. https://doi.org/10.52165/kinsi.27.3.40-51

Jiménez, I. P., & Pain, M. T. (2008). Relative age effect in Spanish association football: Its extent and implications for wasted potential. *Journal of Sports Sciences*, *26*(10), 995–1003. https://doi.org/10.1080/02640410801910285

Kelly, A. G., Jackson, D. J., Taylor, J. A., Jeffreys, M., & Turnnidge, J. (2020). "Birthday-Banding" as a strategy to moderate the relative age effect: A case study into the England squash talent pathway. *Frontiers in Sports and Active Living*, *2*. https://doi.org/10.3389/fspor.2020.573890

KNVB. (2021). *Het geboortedatum effect*.

Maddux, C. D., Stacy, D., & Scott, M. (1981). School entry age in a group of gifted children. *Gifted Child Quarterly*, *25*(4), 180–184. https://doi.org/10.1177/001698628102500408

Mann, D. A., & Van Ginneken, P. J. M. A. (2016). Age-ordered shirt numbering reduces the selection bias associated with the relative age effect. *Journal of Sports Sciences*, *35*(8), 784–790. https://doi.org/10.1080/02640414.2016.1189588

Musch, J., & Grondin, S. (2001). Unequal competition as an Impediment to Personal development: A review of the Relative Age Effect in sport. *Developmental Review*, *21*(2), 147–167. https://doi.org/10.1006/drev.2000.0516

NCAA. (2019, October). *The birthday effect in college athletics*. Retrieved May 19, 2023, from https://www.ncaa.org/sports/2013/11/19/the-birthday-effect-in-college-athletics.aspx

Okazaki, F. H. A., Keller, B., Fontana, F., & Gallagher, J. D. (2011). The relative age effect among female Brazilian youth volleyball players. *Research Quarterly for Exercise and Sport*, *82*(1), 135–139. https://doi.org/10.1080/02701367.2011.10599730

Patalay, P., Belsky, J., Fonagy, P., Vostanis, P., Humphrey, N., Deighton, J., & Wolpert, M. (2015). The extent and specificity of relative age effects on mental health and functioning in early adolescence. *Journal of Adolescent Health*, *57*(5), 475–481. https://doi.org/10.1016/j.jadohealth.2015.07.012

Peña-González, I., Javaloyes, A., Sarabia, J., & Moya-Ramón, M. (2020). Relative age-related differences between different competitive levels and field positions in young soccer players. *Research in Sports Medicine*. https://doi.org/10.1080/15438627.2020.1853540

Rampinini, E., Coutts, A. J., Castagna, C., Sassi, R., & Impellizzeri, F. M. (2007). Variation in top level soccer match performance. *International Journal of Sports Medicine*, *28*(12), 1018–1024. https://doi.org/10.1055/s-2007-965158

Relative Age Effect in Youth Soccer. (2019, October 21). The Soccer Sidelines. https://thesoccersidelines.com/relative-age-effect-in-youth-soccer/

Romann, M., & Cobley, S. (2015). Relative age effects in athletic sprinting and corrective adjustments as a solution for their removal. *PLOS ONE*, *10*(4), e0122988. https://doi.org/10.1371/journal.pone.0122988

Romann, M., & Fuchslocher, J. (2013). Relative age effects in Swiss junior soccer and their relationship with playing position. *European Journal of Sport Science*, *13*(4), 356–363. https://doi.org/10.1080/17461391.2011.635699

Salinero, J. J., Pérez, B. O., Burillo, P., & Lesma, M. L. (2013). Relative age effect in European professional football. Analysis by position. *Journal of Human Sport and Exercise*, *8*(4), 966–973. https://doi.org/10.4100/jhse.2013.84.07

Sherar, L. B., Baxter-Jones, A. D., Faulkner, R., & Russell, K. C. (2007). Do physical maturity and birth date predict talent in male youth ice hockey players? *Journal of Sports Sciences*, *25*(8), 879–886. https://doi.org/10.1080/02640410600908001

Votteler, A., & Höner, O. (2013). The relative age effect in the German Football TID Programme: Biases in motor performance diagnostics and effects on single motor abilities and skills in groups of selected players. *European Journal of Sport Science*, *14*(5), 433–442. https://doi.org/10.1080/17461391.2013.837510

Webdale, K., Baker, J., Schorer, J., & Wattie, N. (2019). Solving sport's 'relative age' problem: a systematic review of proposed solutions. *International Review of Sport and Exercise Psychology*, *13*(1), 187–204. https://doi.org/10.1080/1750984x.2019.1675083

Appendix

Table A.1

Simulation effects for a shift of 2 months with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	3.09e-17	1.762e-17	1.75	0.08
Team level	-3.684e-15	3.684e-15	-1.06	0.291
Constant	6.5	3.16e-14	2.05e+14	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1565	786
2	1544	829
3	1551	847
4	1549	850

Note: This table shows the results of the simulation on the RAE for a shift of 2 months. The simulation includes the variables player performance and team level

Table A.2

Simulation effects for a shift of 3 months with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	3.454e-17	1.973e-17	-1.75	0.08
Team level	-4.356e-15	4.124e-15	1.06	0.291
Constant	6.49	3.54e-14	1.83e+14	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1564	1198
2	1550	1262
3	1550	1261
4	1534	1198

Note: This table shows the results of the simulation on the RAE for a shift of 3 months. The simulation includes the variables player performance and team level

Table A.3

Simulation effects for a shift of 4 months with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	-1.667e-17	9.518e-18	-1.75	0.08
Team level	2.102e-15	1.990e-15	1.06	0.291
Constant	6.46	1.71e-14	3.78e+14	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1576	1615
2	1559	1695
3	1542	1615
4	1549	1695

Note: This table shows the results of the simulation on the RAE for a shift of 4 months. The simulation includes the variables player performance and team level

Table A.4

Simulation effects for a shift of 5 months with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	-7.474e-17	4.269e-17	-1.75	0.08
Team level	9.427e-15	8.924e-15	1.06	0.291
Constant	6.5	7.66e-14	8.48e+13	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1581	2016
2	1569	2123
3	1534	1978
4	1549	2023

Note: This table shows the results of the simulation on the RAE for a shift of 5 months. The simulation includes the variables player performance and team level

Table A.5

Simulation effects for a shift of 6 months with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	-2.684e-17	1.533e-17	-1.75	0.08
Team level	3.386e-15	3.205e-15	1.06	0.291
Constant	6.46	2.75e-14	2.35e+14	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1549	2462
2	1586	2538
3	1551	2462
4	1547	2538

Note: This table shows the results of the simulation on the RAE for a shift of 6 months. The simulation includes the variables player performance and team level

Table B.1

Simulation effects for a shift of 2 months for players in the "selection" *category with team birth month as a dependent variable*

Variable	Coefficient	Standard error	T	*p*
Player performance	-9.31e-17	3.38e-17	-2.75	0.006
Team level	-2.03e-14	1.71e-14	-1.18	0.236
Constant	6.5	6.51e-14	9.98e+13	0

Number of shifts	Average Player performance	Number of players shifted
1	1668	478
2	1641	515
3	1623	751
4	1620	600

Note: This table shows the results of the simulation on the RAE for a shift of 2 months for players in the "selection" category. The simulation includes the variables player performance and team level

Table B.2

Simulation effects for a shift of 3 months for players in the "selection" *category with team birth month as a dependent variable*

Variable	Coefficient	Standard error	T	*p*
Player performance	-8.82e-17	3.2e-17	-2.75	0.006
Team level	-1.92e-14	1.62e-14	-1.18	0.236
Constant	6.49	6.17e-14	1.05e+14	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1670	722
2	1647	823
3	1621	1093
4	1603	722

Note: This table shows the results of the simulation on the RAE for a shift of 3 months for players in the "selection" category. The simulation includes the variables player performance and team level

Table B.3

Simulation effects for a shift of 4 months for players in the "selection" category with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	*p*
Player performance	1.37e-18	4.98e-19	2.75	0.006
Team level	2.99e-16	2.52e-16	1.18	0.236
Constant	6.46	9.59e-16	6.73e+15	0

Number of shifts	Average Player performance	Number of players shifted
1	1686	981
2	1652	1121
3	1614	1119
4	1620	1225

Note: This table shows the results of the simulation on the RAE for a shift of 4 months for players in the "selection" category. The simulation includes the variables player performance and team level

Table B.4

Simulation effects for a shift of 5 months for players in the "selection" category with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	p
Player performance	4.36e-17	1.58e-17	2.75	0.006
Team level	9.5e-15	8.02e-15	1.18	0.236
Constant	6.5	3.05e-14	2.13e+14	0.000

Number of shifts	Average Player performance	Number of players shifted
1	1693	1245
2	1657	1447
3	1603	1332
4	1620	1512

Note: This table shows the results of the simulation on the RAE for a shift of 5 months for players in the "selection" category. The simulation includes the variables player performance and team level

Table B.5

Simulation effects for a shift of 6 months for players in the "selection" category with team birth month as a dependent variable

Variable	Coefficient	Standard error	T	p
Player performance	7.19e-17	2.61e-17	2.75	0.006
Team level	1.57e-14	1.32e-14	1.18	0.236
Constant	6.46	5.03e-14	1.28e+14	0

Number of shifts	Average Player performance	Number of players shifted
1	1555	1696
2	1744	1623
3	1714	1696
4	1737	1623

Note: This table shows the results of the simulation on the RAE for a shift of 6 months for players in the "selection" category. The simulation includes the variables player performance and team level