European Sport Management Quarterly

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/resm20

Relative age effect on the market value of elite European football players: a balanced sample approach

András Gyimesi & Dániel Kehl

To cite this article: András Gyimesi & Dániel Kehl (2021): Relative age effect on the market value of elite European football players: a balanced sample approach, European Sport Management Quarterly, DOI: 10.1080/16184742.2021.1894206

To link to this article: https://doi.org/10.1080/16184742.2021.1894206

Published online: 04 Mar 2021.

EUROPEAN SPORT MANAGEMENT QUARTERLY
https://doi.org/10.1080/16184742.2021.1894206

Routledge
Taylor & Francis Group

Check for updates

Relative age effect on the market value of elite European football players: a balanced sample approach

András Gyimesi and Dániel Kehl

Department of Economics and Econometrics of the Faculty of Business and Economics, University of Pécs, Pécs, Hungary

ABSTRACT

Research question: Athletes born late in the selection year are disadvantaged compared to the relatively older in youth age groups. This leads to an overrepresentation of the early-born in professional teams, which is called the relative age effect (RAE). Previous studies examine the RAE on wage or market value with contradictory results. This paper aims to estimate the partial RAE on the market value of top-level European football players with a new methodology.
Research methods: We analysed a data set of players from the biggest five European football leagues over ten seasons (2008-2017). We argue that the conventional estimation of RAE fails to yield a partial effect, due to a sample selection bias. For various regression models, we use a stratified sampling method to balance the birth date distribution and eliminate the bias.
Results and Findings: Unlike prior studies, we found an extremely strong straight RAE in elite European football, especially for younger players. Using our methodology, we interpret RAE as a partial effect, which indicates that an earlier birth date within the calendar year results in a higher market value.
Implications: This paper contributes to the literature by proposing a new methodology to measure the partial RAE on labour market outcomes. Our results imply discrimination based on birth date in several countries, which needs to be reduced by better regulation of youth competitions. Our findings have implications for coaches and managers on how to account for relative age when training, transferring, and selecting players.

ARTICLE HISTORY
Received 29 May 2020
Accepted 19 February 2021

KEYWORDS
Relative age effect; sample selection bias; European football; market value; soccer

Grouping people by age is a common phenomenon: students born within a year are assigned to classes based on a cut-off date to ensure equal opportunities. However, there are inevitable age differences within the same class. This is particularly relevant in the case of adolescents. Many studies found that relatively younger students underperform relative to their classmates. This has a long-lasting effect on their lives in many aspects (Fumarco et al., 2020; Peña, 2017). In addition to lower self-esteem and general well-being, relative age is associated with many other outcomes including learning disabilities, fewer friends, weaker social networks (Fumarco & Baert, 2019), and

CONTACT András Gyimesi gyimesi.andras12@gmail.com Department of Economics and Econometrics of the Faculty of Business and Economics, University of Pécs, 80 Rákóczi street, Pécs, Hungary, 7622

worse long-term health outcomes (Bahrs & Schumann, 2020). Bai et al. (2019) find labour market effects in adulthood namely relatively older fund managers outperform younger ones and are more self-confident. These phenomena are called relative age effects (RAEs) and are found to be consistent in the literature.

Most youth sports competitions set dates determining the age groups for participants, using the participant's birth year to arrange them in groups. The relevant cut-off date between groups is January 1st. In 1997 the international governing body of professional football (FIFA) decided to set the cut-off date to January 1st for all international competitions (Helsen et al., 2005). Since then, most major European domestic football associations have also adopted this date for domestic youth leagues and other competitions. In many European leagues, up until the late 1990s, the cut-off date was August 1st instead.

Young athletes who are born relatively late in their respective age groups are disadvantaged in both physical and cognitive abilities compared to their relatively older competitors (Burgess & Naughton, 2010; Musch & Grondin, 2001). Therefore, the selection of athletes participating in youth sports clubs will be biased towards the month or quarter of birth, even if coaches are aware of the RAE (Hill & Sotiriadou, 2016). This leads to an overrepresentation of athletes born in certain months of the year. RAE in youth teams is widely documented in the sports science literature in various individual and team sports and countries.

Although age groups and age-based physical differences disappear in adulthood, biased selection during their early career might reduce the athlete's chances of succeeding later in their career as is the case in school starting age. The relatively younger might have less competition experience, lower self-esteem, motivation, and a lower chance to receive high-quality training (Musch & Grondin, 2001). However, there are some studies in the sports science literature, which argue that a reverse RAE exists during the adult career. As Ramos-Filho and Ferreira (2021) point out these effects include salaries, the likelihood of being drafted, career length, and performance, moreover they examine the effects on market value and sports performance. These results seem to contradict those seen in the case of school classes and other studies that find RAE (from this point on we'll call this the straight RAE as opposed to reverse RAE).

The straightforward question arises. In what aspects are the two scenarios different? Why do we see reverse RAE in the case of athletes and not in the case of students? Our paper adds to the literature by analysing the relative age effect on market value using a new method. We show that using OLS on the collected data set might result in a reverse RAE whereas a more careful selection of the sample would suggest a straight RAE. The empirical data set is larger than most previously used ones. It contains elite football players from the top 5 leagues in Europe in 10 seasons. We believe, this study is an important step in resolving controversy about straight and reverse RAE and thus helps future empirical work on this field. This is crucial from the sport management perspective, as coaches and managers need straightforward answers to how their player selection affects the future performance and market value of the athletes to help better management decisions. Everything that managers have learned from previous RAE estimates should be examined with caution, because, if the selection bias is taken into account, the RAE in professional football is straight and much larger than previously thought.

Literature review

The RAE of youth football players is well-established in the literature (Sierra-Díaz et al., 2017). Players born in the first quarter of the year are consistently overrepresented in youth football competitions (Barnsley et al., 1992; Gil et al., 2020; Helsen et al., 2005; Li et al., 2020; Williams, 2010). In youth football, the effect is found to be stronger with younger age groups (Mujika et al., 2009), indicating that straight RAE diminishes as the player getting older. However, several studies found that the overrepresentation of the relatively older persists into adulthood at the professional level in football (Doyle & Bottomley, 2018; Fumarco & Rossi, 2018; Helsen et al., 2012; Musch & Hay, 1999). These results suggest that the relatively early-born players have a lower drop-out rate and a higher chance to be signed by professional football clubs.

Most papers agree that there is a clear straight RAE on the number of professional athletes but the effects on other aspects are far more controversial. Fumarco and Rossi (2018) find no RAE on the performance of Italian football players, while several studies even find a reverse RAE on player performance in the NHL (Bryson et al., 2017; Deaner et al., 2013; Fumarco et al., 2017) and Brazilian football (Ramos-Filho & Ferreira, 2021). Papers finding reverse RAE usually explain this phenomenon with peer effects (Ashworth & Heyndels, 2007) or the tendency for later-born talents to be revealed later (Bryson et al., 2017).

We focus this paper on the RAE on the market value of the players rather than player performance. In professional football, labour is not free to move from one employer to another. Instead, they follow a transfer system, where players are tied to clubs by contracts. It is common practice that a club pays a fee for another club to terminate a player's previous contract and to sign a new one with the other team. Therefore a precise valuation of players on the market is crucial for every football club in order to make profitable transfers. In this sense, the market value of a player reflects the human capital they possess and clubs invest in this human capital. Numerous studies attempt to explain the factors that drive the market value of football players (Bryson et al., 2013; Franck & Nüesch, 2012; Herberger & Wedlich, 2017; Majewski, 2016), but they do not consider relative age as a potential determinant.

There are only a few previous studies that examine the relationship between relative age and human capital related measures, and there is no clear empirical consensus on the connection. Ashworth and Heyndels (2007) study the connection between the month of birth and wages of professional German football players in the 1997/1998 and 1998/1999 seasons. In their sample period, the relevant cut-off date for German football is August 1st, so the calendar months are numbered starting with August as number one. These studies estimate the relationship between the month of birth and the log of wage via OLS regression in a human capital earnings function framework, controlling for age, playing position, and the number of games played as a proxy for experience. The authors find that the relatively early-born players earn systematically lower wages, which corresponds to a reverse RAE. They explain this result mainly with peer effects, suggesting that relatively younger players train with older, more experienced peers, which increases their human capital, resulting in reverse RAE.

Fumarco and Rossi (2018) use Italian data for a more recent and longer period to study the same relationship. They measure relative age by dummy variables indicating the quarter of birth. They also include IVG (index of general evaluation) as a performance indicator in their OLS regression model. They find the opposite direction effect as Ashworth and Heyndels (2007), suggesting a straight RAE on human capital. Those who were born in the last quarter of the year (relatively younger) earn significantly less, both with and without controlling for individual performance.

Two studies use wages from the National Hockey League and find a reverse RAE on wages. Fumarco et al. (2017) uses quantile regressions and finds that those players born in the last quarter of the year have a significant wage advantage and the effect is stronger at the higher quantiles of the wage distribution. Bryson et al. (2017) conclude that relatively older players (born between January and April) get lower wages and this wage gap widens over the course of the player's career.

There are only two studies that explicitly study the RAE on market value. Doyle and Bottomley (2018) use the sample of 1,000 highest valued European football players to test the relationship between birth week and market value. They use OLS regression and find that relative age by birth weeks does not influence the market value of football players in the European top 1,000 players. Ramos-Filho and Ferreira (2021) analyse data of 601 Brazilian players using OLS regression on market value and performance measured by the total number of games in the career. They did not find a significant relationship between market value and quarter of birth, but the number of appearances and relative age (measured as Q4 or not) showed significant reverse RAE.

These previous results suggest that if relatively younger players 'survive' the selection process of professional clubs, the youth career disadvantage disappears and they even have some advantage over the older. The studies on the connection between relative age and human capital based measures yielded ambiguous effects, suggesting that if a potential athlete is born a month later in the year *ceteris paribus*, there is no clear effect on their market value as a professional player. However, we argue that these previous estimations cannot be considered a partial effect and underestimate the advantage of the relatively early born that results from the biased selection in their youth career. All of the above-mentioned studies use samples of professional players, who are already signed by prominent sports clubs. Therefore, their estimated RAE could be interpreted as effects conditional on being a good enough player to enter the sample. There is no debate on the straight RAE in terms of representativeness and the authors also report that those players born early in the calendar year are heavily overrepresented in their samples compared to the average population, which leads to a sample selection bias. This paper addresses a methodological limitation of previous RAE estimations. An alternative method for sample selection is introduced so that our estimation could be interpreted as a partial effect.

Methodology

Data

The empirical data set of our analysis is based on the players of the top 5 football leagues in Europe, namely France (Ligue 1), Germany (Bundesliga), Italy (Serie A), Spain (La Liga), and England (Premier League) in the seasons 2008/2009–2017/2018. Data is collected from the website www.transfermarkt.com This website provides independent estimates of players' market value that is regularly updated by up to 190,000 professional and non-professional individuals with the approval of transfermarkt.com experts (Bryson et al., 2013). This source of data is used in several previous scientific studies (Bryson et al., 2013; Franck & Nüesch, 2012; Herberger & Wedlich, 2017; Ramos-Filho & Ferreira, 2021; Scelles et al., 2016), utilised in real transfer negotiations, and proven to be strong predictors of actually completed transfer fees (Herm et al., 2014). These market value estimates are shown to be a good proxy of team performance and useful for predicting international results (Peeters, 2018).

The data set contains information about the players' team, birth date, market value, and primary playing position for each player-season observation. All players with a non-zero market value in the primary squads of the given leagues are retrieved. Zero market value players are dropped because the logarithmic operation is not meaningful for 0 values. There are a few observations with missing birth dates or positions in the raw data set. These, along with duplicated cases are omitted from the cleaned data set. The main variables used in the analysis are listed in Table 1.

The exact youth age groups in different countries might differ, but the relevant admission date for each youth category in UEFA competitions is January 1st from 1997, hence for the whole time scope of the study. Older players in the sample might have faced different cut-off dates during their youth careers, as the cut-off date in many countries for youth competitions was August 1st in earlier seasons (Butler & Butler, 2015; Helsen et al., 2000; Ostapczuk & Musch, 2013). To account for this issue, players born before 1980 are excluded from the data set. There are 25,816 player-season observations in the final data set.

Figure 1 shows that the later seasons are more represented in the data set, as well as the Italian Serie A in contrast to the French and Spanish first divisions. This difference is due to Italian teams working with larger squads compared to the other leagues. The distribution by playing position and age reflects the squad composition of the teams, with extreme age groups and goalkeepers underrepresented.

Table 1 Description of the variables used in the analysis.

Variable name or category	Variable description
LOGVALUE	The natural logarithm of the player's estimated market value in the given season, measured in EUR.
DOB	The player's day of birth, numbered from 1 (first day of the year) to 366 (last day of a leap year).
QOB	The player's quarter of birth, Q1 to Q4.
AGE	The player's age on January 1st during the given season. Using the January 1st age makes it theoretically uncorrelated with the birthday (The correlation coefficient between DOB and AGE is only 0.017) and the starting date of the season. It is measured on a continuous scale, so it can take non-integer values. The day-based age is used for calculation. For example, Lionel Messi was born on 1987.06.24. He was 11149 days old on January 1st 2018, therefore his AGE value for the 2017/2018 season is 11149/365 = 30.545.
SEASON	The starting year of the season.
LEAGUE	The country where the player's league is located.
POSITION	The player's primary playing position separated into 4 main categories listed below.
Goalkeeper	Goalkeeper
Defender	Centre-Back, Defender, Left-Back, Right-Back
Midfielder	Attacking Midfield, Central Midfield, Defensive Midfield, Left Midfield, Midfielder, Right Midfield
Forward	Centre-Forward, Forward, Left Winger, Right Winger, Second Striker

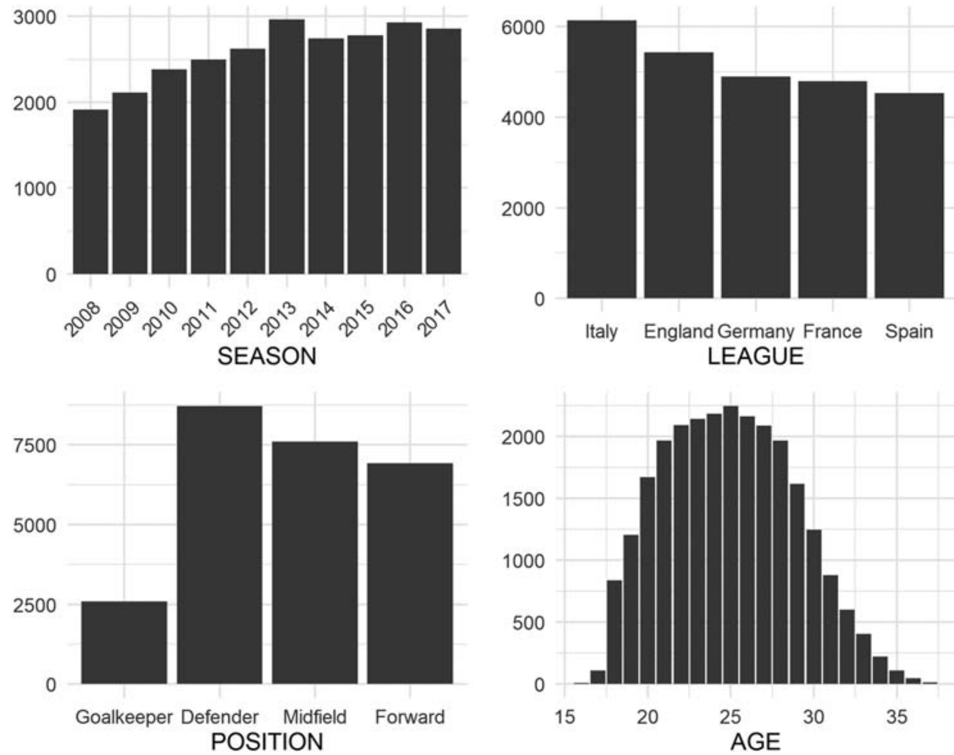

Figure 1. Distribution of players by SEASON, LEAGUE, POSITION, and AGE.

Earlier studies consistently report an overrepresentation of athletes born shortly after the cut-off date. Figure 2 shows that this is also the case in our data set. 11.5% of the players in the data set are born in January, while it continuously drops to only 6% for December born players. The frequencies can be compared to an expected distribution. In cross-country RAE research, the standard procedure is to apply the uniform distribution as expected quarterly birth date frequencies (Barnsley et al., 1992; Doyle & Bottomley, 2019; Li et al., 2020; Williams, 2010). In reality, birth rates differ between quarters but are very heterogeneous between countries and over time, so the exact birth date distribution of the relevant population cannot be defined, but assumed to be adequately uniform. This applies to our data set of the top European leagues containing players coming from all over the world, with a considerable proportion of non-European players. The chi-square test for comparing expected and observed quarterly birth date frequencies yields $p < 0.001$, which shows a highly significant difference from the uniform distribution.

Model specification

A model equation is considered, which is based on a human capital earnings function (Lucifora & Simmons, 2003) similar to the ones that Ashworth and Heyndels (2007) and Fumarco and Rossi (2018) report in their studies. Our main dependent variable is the log market value instead of log wage, but both of them are human capital related measures so the same variables are expected to influence them. Each *i* observation denotes a player in

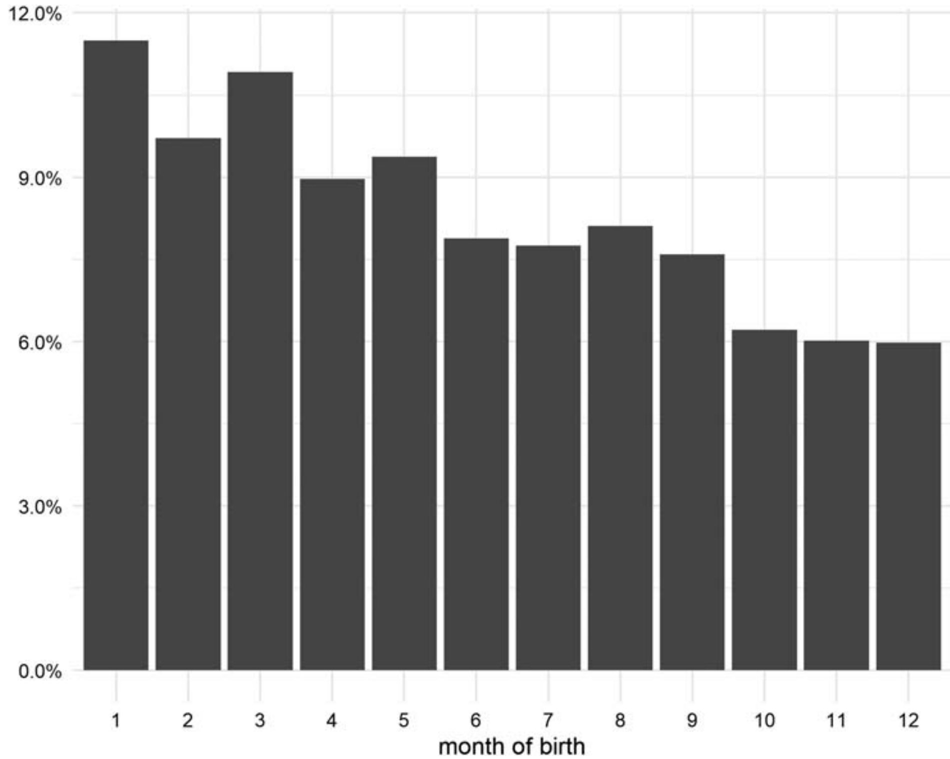

Figure 2. Distribution of players by month of birth.

a single season. The experience of the players is missing from the equation because of the lack of proper data for measuring that variable. The model equation is as follows.

$$\begin{aligned} LOGVALUE_i = & \alpha_0 + \alpha_1 RELAGE_i + \alpha_2 AGE_i + \alpha_3 AGE_i^2 \\ & + \alpha_4 POSITION_i + \alpha_5 LEAGUE_i + \alpha_6 SEASON_i + \varepsilon_i \end{aligned} \quad (1)$$

The variable RELAGE indicates the player's relative age. It either takes the form of DOB or QOB in different specifications. QOB is treated as a categorical variable, while DOB is treated as a continuous scale variable. The latter has the advantage of yielding one coefficient that indicates the overall RAE of one day, however, it requires the assumption that the effect corresponding to a 1 day later birth date is constant throughout the year.

Unlike wages, market value is estimated by parties external to the team. It is theoretically independent of the player's team, so team fixed effects are not included in the main model. However, the player's value still can be affected by the league which he plays for, so league fixed effects are included, along with season fixed effects to account for inflation and capture the increasing trend in market values over time. With the logged dependent variable, the model coefficients can be interpreted as percentage changes on market value, so using nominal market values comes with no interpretation problems for different years. In each model specification, the reference categories for the dummy variables are Defenders for POSITION, Spain for LEAGUE, and 2008 for SEASON.

The AGE variable denotes the absolute age of the player. Measuring age on a continuous scale leads to a more adequate estimation of the RAE, as compared to estimations that use an age variable measured in whole years. In general, it is reasonable to assume a continuous increase of the true market value until the peak of the career, then a continuous decrease. If we measure the players' absolute age discretely, we omit important information on the continuous market value change between birthdays. Then, the relative age variable might unwantedly capture some of this omitted absolute age effect. We measure the absolute age variable as precisely as possible to account for the players' continuous market value development.

The model coefficients of Equation 1 are first estimated using the whole cleaned sample via OLS regression, where the dependent variable is LOGVALUE. Then, the potential shortcomings of this frequently used methodology are discussed, and a balanced sample method is suggested to address those issues.

Handling sample selection bias

First, we estimate the model following a similar methodology as the models in most of the previous studies discussed in the literature review section. While the estimated RAE is valid within the sample, we argue that this methodology suffers from a sample selection bias, so the estimated RAE cannot be interpreted as a partial effect. We work with a sample of professional players who are already signed by prominent sports clubs, but in reality, there are much more less talented players in smaller clubs. Because of the well-documented youth RAE, the less talented players who are relatively younger miss out on a first league signing with a higher probability than the relatively older. This raises the average quality of sampled Q4-born players compared to the Q1-born, as most of the less talented Q4-born players miss out, while some of the less talented Q1-born players make it to first league teams.

This phenomenon is evident in our data set, as similarly to previous studies, Q1-born players are much higher represented than Q4-born players. Therefore, the probability of inclusion in the sample depends on the main explanatory variable: relative age, causing a sample selection bias (Heckman, 1979). Ignoring the sample selection bias leads to an underestimation of straight (and overestimation of reverse) RAE on market value when interpreting the regression coefficients as unconditional partial effects.

If it was really advantageous to be relatively younger, we should not see a pattern of world-class players' market values, which is shown in Figure 3. The 5 highest-valued Q1-born players in the 2017/2018 season were much higher valued on average than the 5 highest-valued Q4-born players (the pattern is similar in other seasons). It is already an indication that we should see a straight RAE instead of a reverse RAE on market value.

To remove the sample selection bias, Heckman (1979) proposed a method that incorporates an additional step of modelling the probability of being included in the sample (heckit model). In this case, however, this probability cannot be modelled, as no data is available for all football players that have not been signed by professional football clubs, so a heckit model is not applicable. However, the sample can be balanced with respect to the quarter of birth to better represent the whole population of (potential) football players.

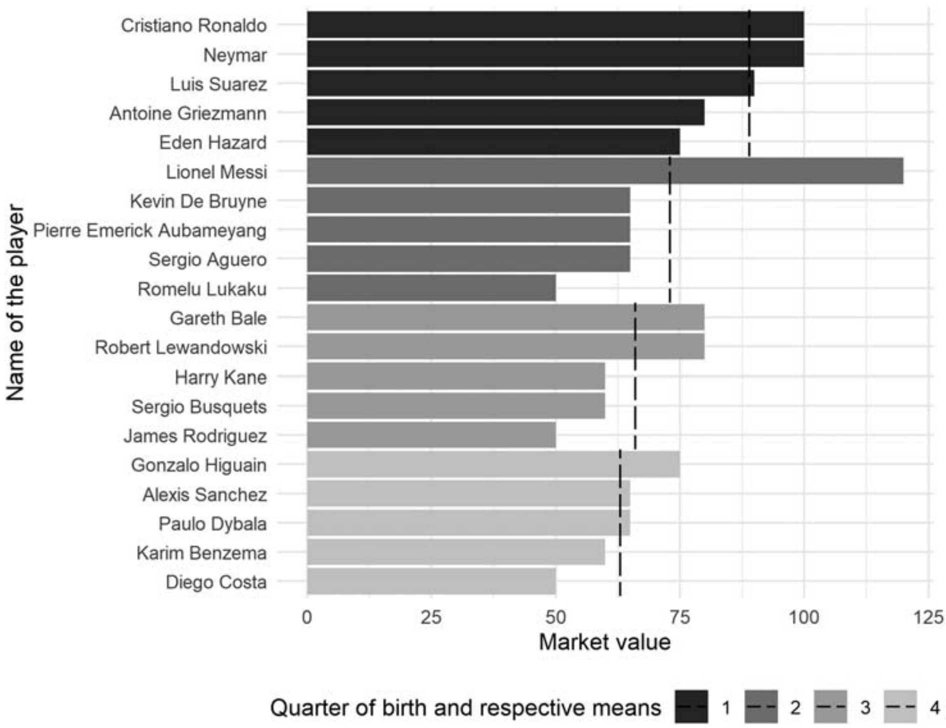

Figure 3. Top 5 highest valued players born in each quarter and mean market value in million EUR (Season 2017/2018).

Our approach is to select an equal number of players from each quarter of birth, creating a new, balanced sample. It comes with the widely used assumption, that the reference population is equally distributed among each quarter of birth, assuming that cross-country birth rate differences are sufficiently close to balancing out (Barnsley et al., 1992; Doyle & Bottomley, 2019; Williams, 2010).

We create a sample of the five leagues for each season in a quasi-experimental setting where the RAE in terms of representativeness does not exist because the selection to prominent teams is not affected by relative age. For that, we artificially drop out players, so that the same number of players remain in the sample from each quarter of birth. We use a stratified sampling method creating sub-groups with respect to QOB, SEASON, POSITION, LEAGUE, and age group (up to 26 or above 26 years old). Each stratum contains N_{sq} players, where s represents each SEASON, POSITION, LEAGUE, and age group combination, and q takes the value of 1–4 depending on the QOB. These N_{sq} stratum sizes differ between the quarters of birth, as Q1-born players are overrepresented in the original sample. For example, there are 31 Forwards in the 2017/2018 Spanish League, who are up to 26 years old and born in Q1, so $N_{s1} = 31$. The stratum-size is 30 for similar Q2-born, 29 for Q3-born, and only 14 for Q4-born players. A random sample of players is drawn without replacement from each stratum, with the sample size equal to the size of the smallest similar stratum among different QOB groups, so $n_{sq} = \min_q N_{sq}$. Considering the previous example, 14 players are drawn from each of these 4 strata, and some Q1, Q2, and Q3-born players are dropped. Also, we assign sampling weights to players according to their market value, so that lower-valued players are more likely to be dropped from the sample. The purpose of this method is to simulate a setting where the relatively older have no better chance to be signed by prominent sports teams.

Thanks to the balanced and weighted sampling mostly lower-valued relatively older player observations were dropped from the sample as if they had been deprived of the advantage of being able to enter prominent sports clubs more easily. The new sample is representative of the full data set regarding SEASON, POSITION, LEAGUE, and age group, but balanced with respect to QOB to avoid the selection bias. We estimate Equation 1 on this balanced sample, with an OLS regression method, and compare the results with the full sample case.

Results

Overall RAE estimation in elite European football

The parameter estimations are shown in Table 2 for both the full sample OLS and the balanced sample models. The first 2 columns of Table 2 show the estimated coefficients

Table 2 OLS parameter estimations.

	Dependent variable:			
	LOGVALUE			
	full sample		balanced sample	
	(1)	(2)	(3)	(4)
DOB	0.0003***		−0.0013***	
	(0.0001)		(0.0001)	
QOB(2)		0.014		−0.182***
		(0.020)		(0.025)
QOB(3)		0.071***		−0.208***
		(0.021)		(0.025)
QOB(4)		0.044*		−0.440***
		(0.022)		(0.025)
AGE	1.690***	1.690***	1.425***	1.420***
	(0.023)	(0.023)	(0.030)	(0.030)
AGE^2	−0.030***	−0.030***	−0.025***	−0.025***
	(0.0004)	(0.0004)	(0.001)	(0.001)
POSITION(Forward)	0.336***	0.336***	0.348***	0.348***
	(0.020)	(0.020)	(0.022)	(0.022)
POSITION(Goalkeeper)	−0.581***	−0.582***	−0.449***	−0.449***
	(0.027)	(0.027)	(0.036)	(0.036)
POSITION(Midfielder)	0.237***	0.236***	0.274***	0.274***
	(0.019)	(0.019)	(0.023)	(0.023)
LEAGUE(France)	−0.407***	−0.407***	−0.491***	−0.493***
	(0.025)	(0.025)	(0.030)	(0.030)
LEAGUE(England)	0.330***	0.332***	0.288***	0.286***
	(0.025)	(0.025)	(0.029)	(0.028)
LEAGUE(Italy)	−0.279***	−0.280***	−0.223***	−0.225***
	(0.024)	(0.024)	(0.029)	(0.029)
LEAGUE(Germany)	−0.213***	−0.213***	−0.228***	−0.232***
	(0.025)	(0.025)	(0.030)	(0.030)
Intercept	−8.994***	−8.981***	−4.516***	−4.479***
	(0.295)	(0.295)	(0.393)	(0.392)
Season dummies	yes	yes	yes	yes
Observations	25,816	25,816	16,367	16,367
Adjusted R^2	0.342	0.342	0.262	0.265

Note: Standard errors in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

of standard full sample OLS regressions. The values of the control variables are roughly the same in all specifications. The dummy variables indicating playing positions show that all else unchanged, Forwards are valued about 33% more than Defenders, while being a Goalkeeper means a value less than half of the Defenders' on average. There are significant differences between the 5 leagues in the market value of similar players. The English Premier League is often considered as the overall highest quality league, which is reflected in the market value premium. The season fixed effects show a steady increase in market values over the years. We find an inverse U-shaped effect of age: the market value increases up to roughly the age of 28 and starts declining afterwards.

The coefficient of DOB and QOB are the parameters of main interest. Generally, if the coefficient value is negative, it indicates a straight RAE, while positive values indicate a reverse RAE on market value. The first model yields a small but significantly positive DOB parameter, which means a reverse RAE of 0.03% with each day of birth. Using the QOB as the relative age variable in Model 2 shows a non-linear effect, indicating Q3-born players having a significant 7.1% market value advantage over the Q1-born. However, this estimated effect does not mean that if an average Q1-born player would have been born in Q3 all else unchanged, he would be valued 7.1% higher. It only means that the average value of Q1-born players is about 7.1% lower than similar Q3-born players in the sample of the top 5 European football leagues.

The last two models in Table 2 are estimated on about two-thirds of the original 25,816 observations, but with uniform birth quarter distribution. In contrary to the full sample models, the balanced sample models yield an exceptionally strong straight RAE on market value. According to the last model, Q4-born players suffer a huge disadvantage, as they are valued 44% less than the Q1-born on average. We have no reason to believe that the sampled Q4-born players are any less talented than the Q1-born players, so this difference must have come from the RAE. The third column reveals that a birth one day later results in a highly significant 0.13% value decrease on average. This can be compared to the traditional full sample OLS estimation of the same model, which yielded a 0.03% reverse RAE each day. The balanced sample estimation leads to a very different conclusion that a strong straight RAE on market value exists. This effect can be interpreted as a partial effect, as it is based on a comparison of similar sets of players with different birth quarters.

It is important to note, that the exact values of the coefficients in Table 2 vary with each run of the model, due to the randomness in the stratified sampling method. To analyse the robustness of the results, the models were run 1000 times. The deviance of the DOB parameter point estimate was always within 0.0001 (0.01 percentage points) from the reported 0.0013 value. The deviance of the QOB(4) parameter point estimate was always within 0.03 (3 percentage points) from the reported 0.44 value. These small deviations do not change the main conclusions of the balanced sample models.

Additional analyses

The DOB coefficient of Model 3 shows a strong straight RAE, as discussed in the previous section, and it is reported again in Table 3. This is an average effect at the elite level of football. The RAE at the professional level can originate from various sources. The relatively older might be signed by better teams, get more game time or experience, or simply get their relative age incorporated in their market value. Models 5 and 6 help to decompose the overall RAE. Team fixed effects are incorporated in Model 5 to filter out the element of RAE that is coming from the phenomenon that the relatively older are playing for stronger teams. Model 6 additionally controls for minutes played during the season.

There are other variables that moderate the effect. As Bryson et al. (2017) highlight, the advantage of the relatively younger might disappear over time. To test this for European elite football, an interaction term between MOB and AGE is incorporated in Model 7. An interaction between MOB and SEASON is introduced in Model 8 to test if RAE has changed over time between 2008 and 2017. Table 3 displays the results of these models.

Controlling for the player's team, in Model 5 reduces the effect to 0.08% with one later day of birth, from the original 0.13%. This suggests that relatively older players are signed by stronger teams, which raises their visibility and hence market value. Controlling for this accounts for about a third of the total RAE, but further team level analysis is required to evaluate the exact effect. RAE is further reduced to 0.07% in Model 6 when game minutes are controlled for. This means that the relatively older also get more game time and some part of the original RAE comes from this advantage.

The MOB and AGE interaction in Model 7 proves to be significantly positive, which underlines the assumption that the straight RAE diminishes throughout the career. However, the decrease is small in magnitude and the effect does not disappear even for older players. At the age of 18 a player is valued about 8% lower than a similar player born a month (30 days) earlier, while at the age of 30, this difference reduces to about 1.5%. This means that early in the career the relatively younger players are significantly undervalued, but the role of physical differences and youth career selection bias decreases over the course of the career.

On the other hand, the MOB-SEASON interaction in Model 8 does not yield a significant coefficient, which means that the level of RAE is unchanged over the sample period.

Table 3 Regressions on balanced sample with additional covariates.

	Dependent variable:				
	LOGVALUE				
	(3)	(5)	(6)	(7)	(8)
DOB	−0.0013***	−0.0008***	−0.0007***	−0.006***	−0.0011***
	(0.0001)	(0.0001)	(0.0001)	(0.001)	(0.0002)
DOB*AGE				0.0002***	
				(0.00002)	
DOB*SEASON					−0.00003
					(0.00003)
AGE	1.425***	1.576***	1.242***	1.367***	1.425***
	(0.030)	(0.025)	(0.022)	(0.031)	(0.030)
AGE^2	−0.025***	−0.028***	−0.022***	−0.025***	−0.025***
	(0.001)	(0.0005)	(0.0004)	(0.001)	(0.001)
MINUTES			0.0004***		
			(0.00001)		
Intercept	yes	yes	yes	yes	yes
TEAM dummies	no	yes	yes	no	no
POSITION dummies	yes	yes	yes	yes	yes
LEAGUE dummies	yes	yes	yes	yes	yes
SEASON dummies	yes	yes	yes	yes	yes
Observations	16,367	16,367	16,367	16,367	16,367
Adjusted R^2	0.262	0.516	0.638	0.265	0.262

Note: Standard errors in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

In parallel with the findings of Helsen et al. (2012), the RAE bias showed no improvement in the past years.

In contrary to the other 4 analysed countries, the cut-off date for English youth tournaments is still September 1st, aligned to the school year instead of the calendar year (Doyle & Bottomley, 2019). The balanced sample regressions are carried out excluding the English Premier League, and the results are presented in Table 4.

Excluding the English Premier League (in the first 2 columns) yields a slightly stronger, but similar straight RAE to Model 3 and 4 in Table 2. This indicates that our main conclusions are not invalidated by the inclusion of English players. Estimating the model solely for the English Premier League (in the last 2 column of Table 4) still yields a very strong straight RAE. Only the Q4-born players face a smaller disadvantage than in the other 4 top leagues, which might be explained by the different cut-off dates. However, the large share of foreign players and the January 1st cut-off date of international competitions have a higher impact on the RAE in the English Premier League.

Conclusions and discussion

Using a new methodology, we found an exceptionally strong RAE in professional football that addresses the issue of biased selection to sports clubs. We use market value data of elite football players to illustrate our findings but we believe that similar issues arise when analysing RAE on performance measures or wages. The results have important implications for researchers, coaches, sports managers as well as the regulation of sport.

The findings contribute to future research on RAE in various ways. There is a consensus on the RAE in terms of representativeness i.e. the number of relatively early-born athletes is usually found to be higher in youth competitions (Barnsley et al., 1992; Gil et al., 2020; Helsen et al., 2005; Li et al., 2020; Williams, 2010), and also in professional football

Table 4. Balanced sample regressions separating the English Premier League.

	Dependent variable:			
	LOGVALUE			
	Rest of leagues		England	
	(9)	(10)	(11)	(12)
DOB	−0.0014***		−0.0010***	
	(0.0001)		(0.0002)	
QOB(2)		−0.166***		−0.216***
		(0.029)		(0.050)
QOB(3)		−0.165***		−0.306***
		(0.029)		(0.050)
QOB(4)		−0.479***		−0.315***
		(0.029)		(0.050)
AGE	1.396***	1.390***	1.476***	1.476***
	(0.034)	(0.034)	(0.061)	(0.061)
AGE^2	−0.025***	−0.025***	−0.026***	−0.026***
	(0.001)	(0.001)	(0.001)	(0.001)
Intercept	yes	yes	yes	yes
POSITION dummies	yes	yes	yes	yes
LEAGUE dummies	yes	yes	yes	yes
SEASON dummies	yes	yes	yes	yes
Observations	12,460	12,460	3,907	3,907
Adjusted R^2	0.232	0.237	0.218	0.220

Note: Standard errors in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

(Doyle & Bottomley, 2018; Fumarco & Rossi, 2018; Helsen et al., 2012; Musch & Hay, 1999). However, this overrepresentation introduces a bias in the estimated RAE coefficients, the true effect of relative age can be better estimated using a sub-sample with balanced birth date distribution. Commonly used data sets are not random samples of all players but only observe those who were talented enough to get to the top level (e.g. Ashworth & Heyndels, 2007; Bryson et al., 2017; Doyle & Bottomley, 2018; Fumarco & Rossi, 2018; Fumarco et al., 2017; Ramos-Filho & Ferreira, 2021). If an estimate is obtained using the full sample, the effect has to be interpreted as: given players have already been signed by the sampled clubs. This however does not take the difference in the chances into consideration i.e. the less talented youth players likely to miss out if they are born at the end of the selection year, while they likely get in the sample if they are relatively older. The effect estimates from a balanced sample approach are closer to true partial effects, comparing similar players born in different months or quarters. The applied methodology can heavily influence the estimated RAE or even its direction.

The extremely strong competition in elite football is responsible for the RAE in terms of representativeness which will cause some of the estimates to support reverse RAE on performance, wage, or market value (e.g. Ashworth & Heyndels, 2007; Ramos-Filho & Ferreira, 2021). There is no such phenomenon (high drop-out rate) in school, this is the reason we do not see reverse RAE in the literature dealing with student performance (Bahrs & Schumann, 2020; Fumarco & Baert, 2019; Fumarco et al., 2020; Peña, 2017). In general, we believe that the higher the drop-out rate of certain age groups, the larger the observable bias, and the more important is to account for this using balanced samples. Proving this belief, however, needs more research and we suggest researchers comparing both methods in their future studies.

Using the proposed new methodology, we found an extremely strong straight RAE on the market value of European footballers. This contradicts the results of several previous studies of RAE on market value (Doyle & Bottomley, 2018; Ramos-Filho & Ferreira, 2021) or wage (Ashworth & Heyndels, 2007; Bryson et al., 2017; Fumarco et al., 2017), that either found a reverse or no effect. The results support the findings of Fumarco and Rossi (2018), who also find a straight RAE on the wages of Italian footballers, but our estimation is much larger in magnitude. This estimate is conducted measuring the absolute age variable on a continuous scale, which is an important contribution. We suggest researchers reporting the way of calculating their absolute age variable as this will possibly influence their results. Finally, it is shown that RAE on market value is stronger at a younger age but the effect decreases with the player getting older, so relative age is an important determinant of the career market value trajectory.

Coaches and football managers can draw valuable conclusions based on these results. Preferring relatively younger players on the transfer market might be beneficial, as they are undervalued compared to similar relatively older players. Later in the career, the value gap narrows, so the transfer is expected to be financially rewarding. Targeting young players that are cheap and overlooked because of their Q4 birth date might be a lucrative transfer strategy. The relatively older are found to be valued higher partly because they play for better teams and get more game minutes. We suggest both youth coaches and professional club managers make their squad selection with respect to RAE. The relatively younger might show worse performance (compared to those born in the same year) only because of their physical disadvantage and lack of confidence. Looking past their inferior performance, and giving them more game time might be beneficial for their confidence and market value development.

Finally, implications for football policy-makers are evident. If the relatively older players are valued higher, they possess more human capital and likely also get higher wages (although the exact partial RAE on wages requires further research). This indicates strong labour market discrimination based on the birth date of athletes, which is against the ideology of equal labour rights. We found no improvement in the RAE bias over the years, which is an indication that coaches and sports decision-makers are not fully aware of the problem. Regulators should consider the recommendations of partial RAE estimates on individual outcomes (i.e. performance, wage, or market value) to foster urgent action and take measures to eliminate the extreme bias. There are already various proposed solutions to the problem in the literature. Sierra-Díaz et al. (2017) provide an overview of the previous recommendations on reducing RAE in football. The calendar-year grouping and the uniform cut-off date in youth competitions are largely responsible for the RAE, so different-length age groups and rotating cut-off dates from year to year are suggested in several studies (e.g. Barnsley et al., 1992; Helsen et al., 2012; Hill & Sotiriadou, 2016; Musch & Grondin, 2001). Other recommendations include the education of coaches and parents, grouping based on anthropometric characteristics (Sierra-Díaz et al., 2017), and the introduction of a birth quarter quota (Fumarco & Rossi, 2018).

Acknowledgment

This work was supported by the ÚNKP-20-3-II New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund; Thematic Excellence Program 2020 - Institutional Excellence Sub-programme of the Ministry for Innovation and Technology in Hungary under grant number: 2020-4.1.1.-TKP2020; Comprehensive Development for Implementing Smart Specialization Strategies at the University of Pécs under grant number: HRDOP-3.6.1-16-2016-00004. The authors would like to thank organizers and participants of the 4th International Conference on Sport Economics & Sport Management (SESM) in Berlin; Jeffrey Anderson, colleagues at the University of Pécs, and anonymous referees for providing helpful comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

András Gyimesi http://orcid.org/0000-0002-6333-3318
Dániel Kehl http://orcid.org/0000-0002-5048-8062

References

Ashworth, J., & Heyndels, B. (2007). Selection bias and peer effects in team sports: The effect of age grouping on earnings of German soccer players. *Journal of Sports Economics*, *8*(4), 355–377. https://doi.org/10.1177/1527002506287695

Bahrs, M., & Schumann, M. (2020). Unlucky to be young? The long-term effects of school starting age on smoking behavior and health. *Journal of Population Economics*, *33*(2), 555–600. https://doi.org/10.1007/s00148-019-00745-6

Bai, J., Ma, L., Mullally, K. A., & Solomon, D. H. (2019). What a difference a (birth) month makes: The relative age effect and fund manager performance. *Journal of Financial Economics*, *132*(1), 200–221. https://doi.org/10.1016/j.jfineco.2018.10.003

Barnsley, R. H., Thompson, A. H., & 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

Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent footedness and player remuneration in European soccer. *Journal of Sports Economics*, *14*(6), 606–628. https://doi.org/10.1177/1527002511435118

Bryson, A., Gomez, R., & Zhang, T. (2017). All-Star or benchwarmer? Relative Age, cohort size and career success in the NHL. In B. Frick (Ed.), *Breaking the Ice. Sports Economics, management and policy* (vol. 16, pp. 57–91). Springer.

Burgess, D. J., & Naughton, G. A. (2010). Talent development in adolescent team sports: A review. *International Journal of Sports Physiology and Performance*, *5*(1), 103–116. https://doi.org/10.1123/ijspp.5.1.103

Butler, D., & Butler, R. (2015). The relative age effect and under-21 Irish association football: A natural experiment and policy recommendations. *The Economic and Social Review*, *46*(4), 511–519. https://doi.org/10.2139/ssrn.2467013

Deaner, R. O., Lowen, A., & Cobley, S. (2013). Born at the wrong time: Selection bias in the NHL draft. *PLoS ONE*, *8*(2), article no e57753. https://doi.org/10.1371/journal.pone.0057753

Doyle, J. R., & Bottomley, P. A. (2018). Relative age effect in elite soccer: More early-born players, but no better valued, and no paragon clubs or countries. *PLoS ONE*, *13*(2), article no e0192209. https://doi.org/10.1371/journal.pone.0192209

Doyle, J. R., & Bottomley, P. A. (2019). The relative age effect in European elite soccer: A practical guide to poisson regression modelling. *PloS ONE*, *14*(4), article no e0213988. https://doi.org/10.1371/journal.pone.0213988

Franck, E., & Nüesch, S. (2012). Talent and/or popularity: What does it take to be a superstar? *Economic Inquiry*, *50*(1), 202–216. https://doi.org/10.1111/j.1465-7295.2010.00360.x

Fumarco, L., & Baert, S. (2019). Relative age effect on European adolescents' social network. *Journal of Economic Behavior and Organization*, *168*, 318–337. https://doi.org/10.1016/j.jebo.2019.10.014

Fumarco, L., Baert, S., & Sarracino, F. (2020). Younger, dissatisfied, and unhealthy – relative age in adolescence. *Economics & Human Biology*, *37*, article no 100858. https://doi.org/10.1016/j.ehb.2020.100858

Fumarco, L., Gibbs, B., Jarvis, J., & Rossi, G. (2017). The relative age effect reversal among the National Hockey League elite. *PLoS ONE*, *12*(8), article no e0182827. https://doi.org/10.1371/journal.pone.0182827

Fumarco, L., & Rossi, G. (2018). The relative age effect on labour market outcomes: Evidence from Italian football. *European Sport Management Quarterly*, *18*(4), 501–516. https://doi.org/10.1080/16184742.2018.1424225

Gil, S. M., Bidaurrazaga-Letona, I., Martin-Garetxana, I., Lekue, J. A., & Larruskain, J. (2020). Does birth date influence career attainment in professional soccer? *Science and Medicine in Football*, *4*(2), 119–126. https://doi.org/10.1080/24733938.2019.1696471

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. https://doi.org/10.2307/1912352

Helsen, W. F., Baker, J., Michiels, S., Schorer, J., Van Winckel, J., & Williams, A. M. (2012). The relative age effect in European professional soccer: Did ten years of research make any difference? *Journal of Sports Sciences*, *30*(15), 1665–1671. https://doi.org/10.1080/02640414.2012.721929

Helsen, W. F., Starkes, J. L., & Van Winckel, J. (2000). Effect of a change in selection year on success in male soccer players. *American Journal of Human Biology: The Official Journal of*

the Human Biology Association, *12*(6), 729–735. https://doi.org/10.1002/1520-6300(200011/12)12:6<729::AID-AJHB2>3.0.CO;2-7
Helsen, W. F., 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
Herberger, T. A., & Wedlich, F. (2017). Does selection bias matter in football players' valuation? A crowdsourced valuation approach on players' athletic characteristics. *Journal of Global Sport Management*, *2*(3), 196–214. https://doi.org/10.1080/24704067.2017.1350593
Herm, S., Callsen-Bracker, H. M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, *17*(4), 484–492. https://doi.org/10.1016/j.smr.2013.12.006
Hill, B., & Sotiriadou, P. (2016). Coach decision-making and the relative age effect on talent selection in football. *European Sport Management Quarterly*, *16*(3), 292–315. https://doi.org/10.1080/16184742.2015.1131730
Li, Z., Mao, L., Steingröver, C., Wattie, N., Baker, J., Schorer, J., & Helsen, W. F. (2020). Relative age effects in elite Chinese soccer players: Implications of the 'one-child' policy. *PLoS ONE*, *15*(2), article no e0228611. https://doi.org/10.1371/journal.pone.0228611
Lucifora, C., & Simmons, R. (2003). Superstar effects in sport: Evidence from Italian soccer. *Journal of Sports Economics*, *4*(1), 35–55. https://doi.org/10.1177/1527002502239657
Majewski, S. (2016). Identification of factors determining market value of the most valuable football players. *Journal of Management and Business Administration. Central Europe*, *24*(3), 91–104. https://doi.org/10.7206/jmba.ce.2450-7814.177
Mujika, I., Vaeyens, R., Matthys, S. P., Santisteban, J., Goiriena, J., & Philippaerts, R. (2009). The relative age effect in a professional football club setting. *Journal of Sports Sciences*, *27*(11), 1153–1158. https://doi.org/10.1080/02640410903220328
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
Musch, J., & Hay, R. (1999). The relative age effect in soccer: Cross-cultural evidence for a systematic discrimination against children born late in the competition year. *Sociology of Sport Journal*, *16*(1), 54–64. https://doi.org/10.1123/ssj.16.1.54
Ostapczuk, M., & Musch, J. (2013). The influence of relative age on the composition of professional soccer squads. *European Journal of Sport Science*, *13*(3), 249–255. https://doi.org/10.1080/17461391.2011.606841
Peeters, T. (2018). Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, *34*(1), 17–29. https://doi.org/10.1016/j.ijforecast.2017.08.002
Peña, A. P. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, *56*, 152–176. https://doi.org/10.1016/j.econedurev.2016.12.001
Ramos-Filho, L., & Ferreira, M. P. (2021). The reverse relative age effect in professional soccer: An analysis of the Brazilian National League of 2015. *European Sport Management Quarterly*, *21*(1), 78–93. https://doi.org/10.1080/16184742.2020.1725089
Scelles, N., Helleu, B., Durand, C., & Bonnal, L. (2016). Professional sports firm values bringing new determinants to the foreground? A study of European soccer, 2005–2013. *Journal of Sports Economics*, *17*(7), 688–715. https://doi.org/10.1177/1527002514538976
Sierra-Díaz, M. J., González-Víllora, S., Pastor-Vicedo, J. C., & Serra-Olivares, J. (2017). Soccer and relative Age effect: A walk among elite players and young players. *Sports (Basel)*, *5*(1), 5. https://doi.org/10.3390/sports5010005
Williams, J. H. (2010). Relative age effect in youth soccer: Analysis of the FIFA U17 World Cup competition. *Scandinavian Journal of Medicine & Science in Sports*, *20*(3), 502–508. https://doi.org/10.1111/j.1600-0838.2009.00961.x