Fading Shooting Stars – The Relative Age Effect, Ability, and Foregone Market Values in German Elite Youth Soccer

Lukas Tohoff* Mario Mechtel†

March 19, 2025

Forthcoming in:

Journal of Sports Economics

Abstract

This paper analyses the Relative Age Effect (RAE) in German elite youth soccer. We hypothesize that, to get selected by elite youth academies, players with relative age disadvantages must be relatively more talented; especially at the margin of getting selected. Using data on 2,383 former elite youth players and their later market values, we show that the RAE was substantial in German elite youth soccer between 2000 and 2020. Moreover, in the sample of former elite youth academy players, those with relative age disadvantages reach significantly higher market values. Our results indicate that relative age disadvantages of elite academy players are positively correlated with their unobserved ability; suggesting that, on average, the RAE results in a loss of talent – and market value. We find that clubs could generate 30.6 to 72.8% higher market values without the RAE.

JEL Codes: J24, Z22, M51, M53, I24, I26, D71
Keywords: Relative age effect in elite youth soccer, selection, market values, unobserved ability, misallocation of talent.

*ROCKWOOL Foundation Berlin, Humboldt-Universität zu Berlin, and Berlin School of Economics, e-mail: lut@rfberlin.com.

†**Corresponding author.** Leuphana University of Lueneburg, Institute of Economics, Universitaetsallee 1, 21335 Lueneburg, Germany, e-mail: mario.mechtel@leuphana.de, phone: +49 4131 6772636.

1 Introduction

In talent selection, it is crucial to distinguish between current and potential performance levels. The person who is the best right now need not be the same as the person who will be the best in five years. Evaluating a heterogeneous talent pool based on current performance only can therefore have negative effects in the long run. In soccer, for instance, many young athletes who are considered elite today will no longer be elite tomorrow. These young athletes will be called fading shooting stars in this paper. Fading shooting stars, in this sense, shine bright today but will never appear on the sky of professional soccer.

It would, of course, require clairvoyant abilities to predict which soccer talent will eventually make it to the professional stage. Performance development in soccer as well as in most other settings is too complex and ambiguous to make exact predictions. Yet, there is also a systematic reason for the abundance of fading shooting stars, which can easily be identified and targeted: the *relative age effect* (RAE). In youth soccer, athletes are grouped by years of birth in most countries. This creates an arbitrary age cut-off. Consequently, adolescents born in January are almost one year older than their December born peers. When playing in the same team, players born in January have a relative age advantage. They are relatively faster, stronger, more mature, and therefore momentarily better athletes on average. Under the 'non-astrology' (Allen and Barnsley 1993) assumption that ability is uncorrelated with birth dates, observing elite youth academies to select more relatively older players would mean that clubs focus too much on current performance rather than the potential performance level of players. The resulting over-representation of relatively older players likely leads to a waste of talent and resources.

A large literature documents the existence of the RAE, drawing on data from several countries, soccer teams, and time periods (e.g., Barnsley et al. 1985, Musch and Hay 1999, Musch and Grondin 2001, Ashworth and Heyndels 2007, Cobley et al. 2008, Mujika et al. 2009, Tribolet et al. 2019, Jackson and Comber 2020, and Pérez-González et al. 2021). Many publications also provide recommendations on how to mitigate the RAE in talent selection processes (Martindale et al. 2012, Mann and Ginnecken 2017, Cumming et al. 2018, Lagestad et al. 2018, and Roberts et al. 2020). Yet, little has changed: Even though the RAE in professional soccer is known since 35 years, it is still very prevalent and, overall, even intensified over time (Sierra-Díaz et al. 2017). Roberts et al. (2020) thus argue that researchers need to consider new approaches to target the RAE in professional soccer to better understand the phenomenon and quantify its consequences.

Our paper investigates the RAE in German elite youth soccer. We use data on 2,383 former youth players of the 17 most successful German Bundesliga Youth Academies (BYA) and their market values in the period between 2002 and 2020. Our paper contributes to the literature in four dimensions. First, despite the attempts to mitigate the problem, we show that the RAE is still prevalent in German BYAs. Second, we introduce a new theoretical model of a player's performance development over time, which facilitates the understanding and analysis of the RAE, and allows us to derive testable hypotheses. Third, using econometric methods and novel data on former German BYA players, we test the implications of the RAE for talent selection and the distribution of unobserved ability in elite youth academies. Fourth, using data on players' market values, we aim to quantify the costs caused by the RAE in professional German soccer, which, in today's highly capitalized soccer, could be a strong argument for changing talent selection practices.

Specifically, we derive from our theoretical model the hypothesis that, among all players selected into BYAs, players with relative age disadvantages have on average higher unobserved ability than their relatively older peers. This builds on the simple observation that relatively younger players must compensate the disadvantages caused by their relative age with more ability to still get selected. We further hypothesize that this effect is particularly pronounced among those players who, at the point of selection, just met the threshold requirements.

Our first finding is that the RAE in BYAs is both substantial and persistent. 71.5% of former Under-19 (U19) BYA players were born in the first half and 44.6% in the first quarter of the year. Moreover, the RAE has even increased slightly in the last two decades. Second, we find that, elite youth players that were born towards the end of the year, in fact, reach significantly higher market values on average. Using an instrumental variable approach, we further show that, at the margin of getting selected, relative age disadvantages are positively correlated with (unobserved) ability. This empirical result supports our theoretical hypotheses and rationalizes also why we find a positive correlation between relative age disadvantages and market values on average in our sample of former elite youth players. Our third finding is that the RAE is very expensive for BYAs: We estimate that Bundesliga clubs could generate 30.6 to 72.8% higher market values through their BYAs when eliminating the RAE in talent selection. This result can be considered as rather conservative as we only model the costs of bad selection related to the RAE. Relative maturity differences during adolescence presumably cause additional costs.

The findings and mechanisms we describe are also relevant for talent identification, development, and recruitment outside of sports. Various studies from different fields show that initial differences in (relative) performance have significant consequences on selection outcomes and achievement, and that eliminating structural biases in recruitment comes with sizeable (economic) gains (e.g., Cullen et al. 2006, Hanushek and Rivkin 2009, Dustmann et al. 2016, Friebel et al. 2019, Hsieh et al. 2019, Murphy and Weinhardt 2020, and Balboni et al. 2022). The world of soccer lends itself, in particular, to the analysis of how initial performance differences affect selection and individual careers in the short and long run because of the excellent data available[1].

The paper is organized as follows. Section 2 discusses the related literature on the existence and consequences of the RAE. Section 3 proposes a model of player's performance development, which allows to illustrate the mechanisms involved in the RAE as well as derive hypotheses for the empirical analysis. Section 4 describes the institutional setting in Germany and our data. Section 5 presents our empirical analysis and results. Finally, section 6 discusses the implications of our analysis and concludes.

2 Related Literature

2.1 The Relative Age Effect in Soccer

The existence of the RAE in sports was for the first time shown by Barnsley et al. (1985), who report skewed birth date distributions in Canadian youth ice hockey. In the 1990s, first soccer-related RAE studies were published. Musch and Hay (1999), for example, find evidence for strong RAEs in professional soccer across several countries including Germany. Decades of research have produced a large body of evidence on the RAE. Yet, the RAE has continued to exist in both youth and professional soccer. Therefore, Roberts et al. (2020) see the need to identify new data capture techniques and more sensitive measures of the RAE to foster a deeper understanding of the effect and its consequences.

While Allen and Barnsley (1993) outline a basic model, the only formalized model of the RAE in sports so far is developed by Pierson et al. (2014), who model the RAE as a reinforcing feedback loop and apply it to Canadian youth hockey. Moreover, Dawid and Muehlheusser (2015) present a dynamic model of repeated talent selection with heterogeneity in ability and relative age, which can also be applied to sports. Besides that, most publications have only relied on descriptive statistics so far.

Cobley et al. (2008) track the RAE in professional German soccer from 1963 to 2007. Using χ^2 tests, they show that the RAE grew consistently and progressively within the period examined. The proportion of players born in the first half of the year is a very popular estimate for the RAE. Referring to the review by Musch and Grondin (2001), early studies on elite youth soccer players in the UK and Sweden found that the proportion of players born in the first half of the year amounts to between 62 and 87%. Based on the same descriptive measure, more recent studies on elite youth academies report the following figures: 85.9% for U9 British Premier League players (Jackson and Comber 2020), 65.4% for Australian U19 elite soccer academies (Tribolet et al. 2019), 75.2% for the AC Bilbao elite youth (Mujika et al. 2009), and 65.6% for international youth championships between 2017 and 2019 (Pérez-González et al. 2021). Overall, age groups examined and measures used differ largely across studies, while the results are unequivocal: The RAE still exists in elite (youth) soccer teams.

Approaches to relate the RAE to players' monetary valuations, have not yet yielded conclusive results. Pérez-González et al. (2020), for instance, analyse players of ten highly successful European soccer clubs and show that players' market values are not significantly correlated with their relative age. Moreover, Fumarco and Rossi (2018) show that professional soccer players born in the last quarter of the year earn significantly lower wages than players born in the first quarter. However, with similar statistical precision, they find that players born in the third quarter of the year earn substantially *more* than those born in the first quarter. Ashworth and Heyndels (2007) use data from professional soccer players in the German Bundesliga for the seasons 1997/1998 and 1998/1999. Based on estimated gross wages, they find that the late-born players in a cohort earn higher wages. Hence, within the sample of professional soccer players – those shooting stars that made it to the professional stage – the relation between players' monetary valuation and relative age is not unambiguously clear.

2.2 Production Function of Elite Youth Academies and the Optimal Selection Policy

The existence and implications of the RAE in German elite youth soccer highly depend on the production function of elite youth academies; in other words, on how elite youth academies employ different kinds of training and selection strategies to optimally exhaust the talent pool. Dawid and Muehlheusser (2015) show that, when initial relative age advantages are strong, clubs can maximize the quality of the talent pool in the long term if they initially resist the temptation to select players based on momentary performance signals[2]. In other words, scarce training resources are misallocated if clubs always select the momentarily best despite strong relative age advantages.

While Dawid and Muehlheusser (2015) assume that "planners" want to maximize the average talent level in a given population at the end of the training process, which we will call the *average shooting star strategy*, it could be possible that soccer clubs have different objectives and thus a different production function. For example, clubs could consider it most effective to focus on the performance development of a small subgroup of 3 to 5 very promising players, which we will denote as the *top shooting star strategy*. To support the few top shooting stars optimally, clubs might surround them at every given stage with the currently best players available which tend to be relatively older and more mature on average. This strategy of largely utilizing the RAE might even be necessary to retain and attract the best. To give a better-informed assessment of the production function of the elite youth academies, we briefly summarize the relevant literature.

In terms of short-term success, it is optimal to fully follow the *average shooting star strategy*. Grossmann and Lames (2013) show that youth clubs can increase their momentary competitiveness by exploiting the RAE. As the RAE tends to be more pronounced in elite youth leagues (Del Campo et al. 2010 and Jackson and Comber 2019) and in clubs which are regarded as successful and have an excellent reputation (Jimenez and Pain 2008), elite youth clubs indeed show a preference for short-term success and momentary competitiveness. Moreover, Jimenez and Pain (2008) argue that the first aim of clubs is to be successful at all stages instead of promoting the greatest talents and taking a long-term perspective. This short-term orientation is further intensified by coaches' incentives who perceive pressure to select players based on short-term goals (Hill and Sotiriadou 2018, and Roach 2022). While these findings do not necessarily contradict the *top shooting star strategy* which also largely relies on the utilization of the RAE, it is apparent that talent development does not just follow a long-term plan but is subject to many short-term constraints.

Furthermore, the *top shooting star strategy* requires that elite youth academies are able to identify top talents already at early stages of selection and that the selection of these top shooting stars is independent of the RAE. Both are rather strong assumptions. The RAE, in fact, is still significant in adult elite leagues (see Sierra-Díaz et al. 2017 and Figure 8 in Appendix C) which indicates the inability of elite youth academies to identify their top players independently of the

RAE. Although we cannot fully dismiss the *top shooting star strategy*, in this paper, we will assume that elite youth academies cannot identify the most promising talents at early stages of selection but, being subject to short-term constraints, primarily aim at maximizing the average talent level.

2.3 Stylized Facts on the RAE and Performance Development

Before developing our model, we present stylized facts from the literature on the RAE and youth players' performance development. A model that is faithful to the evidence must recognize these empirical findings. First, relative age and maturity advantages are generally beneficial in soccer (Malina et al. 2000, Rösch et al. 2000, Malina et al. 2007, Votteler and Höner 2014, Lovell et al. 2015, and Rommers et al. 2018). Second, relative maturity differences can be substantial during adolescence, are greatest around the age of 13 and decline afterwards (Malina et al. 2004 and Walker 2016). Third, the RAE in elite youth soccer follows this maturity pattern, increasing initially and peaking around the age of 13 to 15. Yet, the RAE does not disappear eventually but remains significant even at the professional level (Cobley et al. 2008, Pierson et al. 2014, Sierra-Díaz et al. 2017, and Patel et al. 2019). Fourth, initial age and maturity advantages likely lead to a path dependency due to access to better training and other factors such as players' increased self-confidence, parents' behaviour, and coaches' perceptions (Musch and Grondin 2001, and Pierson et al. 2014). Fifth, as discussed above, the RAE is more pronounced in elite leagues and youth clubs can increase their momentary competitiveness by exploiting the RAE (Jimenez and Pain 2008, Del Campo et al. 2010, Grossmann and Lames 2013, and Jackson and Comber 2019).

From these stylized facts, it is also apparent that the RAE is complemented by a relative maturity effect (RME), i.e., differences in maturation status which are independent from relative age (see Malina et al. 2000). Hence, analysing the impact of only the RAE (and not the RME) on talent allocation will most likely lead to conservative results when it comes to skewed talent selection and misallocation of talent.

3 Theoretical Framework

3.1 Basic Setup

The simple theoretical model introduced in this section aims to illustrate the problems caused by the RAE. The model goes beyond the previous analysis of ability, relative age, selection, training, and monetary valuation in soccer (e.g., Ashworth and Heyndels 2007) and will serve as the basis for deriving our hypotheses. Let P_i denote player i's realized performance level as a function of time t, which is measured in years and refers to his age. P_i aims to represent a player's ability, exercise, and routine as well as soccer specific attitudes (e.g., tactical sense) and physical characteristics (e.g., fitness, height, and speed) – in short, everything that determines how good a player is (see Reilly et al. 2000). As we focus on the earlier stages of a player's career from childhood to the professional age, we rely on a logistic growth function. Here, players' realized performance levels increase with age. This approach allows to incorporate heterogeneity in ability, training, and relative age, but, owing to simplicity, misses to represent the decline in performance at later stages of the career.[3] Finally, $P_i^\star$ captures player i's maximum performance level, which is the performance level that the logistic growth function will eventually converge to. Player i's performance level in period t is $P_i(t) = \dfrac{P_i^\star}{1 + (P_i^\star - 1) \times exp(-t)}$, which implies $\lim_{t \to \infty} P_i(t) = P_i^\star$.

Let m_i denote player i's month of birth which shifts the performance development function to the right according to the player's relative age. The development of a player born in December starts 11/12 of a year later than the development of his peers born in January of the same year. The starting point of the performance development function is, thus, defined by player i's birth month. This yields

$$P_i(t) \quad = \quad \frac{P_i^\star}{1 + (P_i^\star - 1) \times exp\left(-t - \dfrac{m_i}{12}\right)}. \tag{1}$$

Figure 1: Simple Performance Development Model with Two Different Birth Months and Two Different Talent Levels

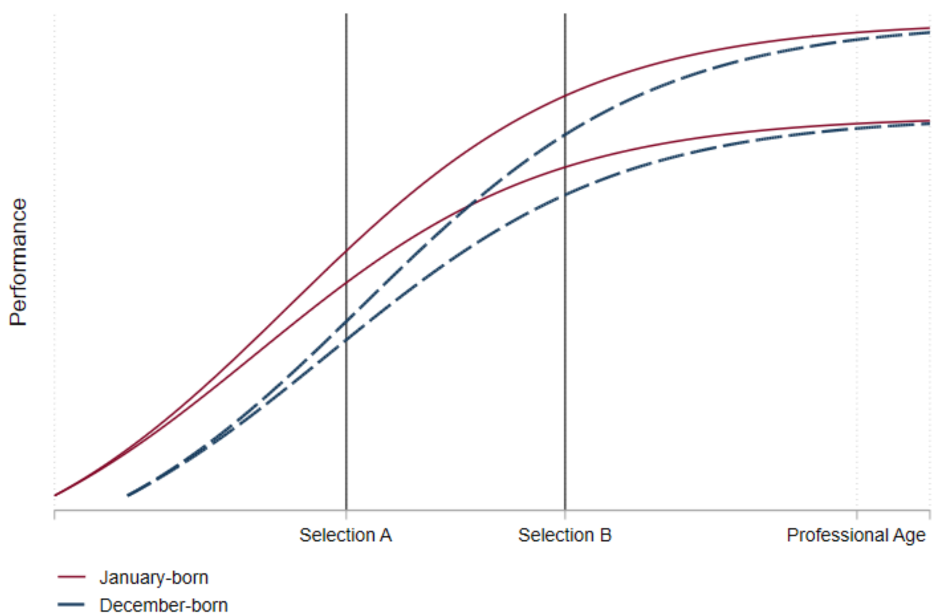


Figure 1 plots the performance development functions of 4 distinct players born in the same year. Two of these players were born in January (red solid lines) and two in December (blue dashed lines). Moreover, per birth month, one player is relatively talented (higher $P_i^\star$) and the other player is relatively untalented (lower $P_i^\star$). The performance development function of players of the same talent level is specified equivalently apart from the fact that they are shifted according to the respective birth month. Figure 1 furthermore indicates two points of selection, A and B, where elite youth academies choose a certain number of adolescents. The plotted performance development functions show, first, that players' realized performance levels increase with age. The slope of the function initially increases and eventually decreases, which perfectly represents the fact that male adolescent height velocity peaks around the age of 13 (Walker, 2016), so that marginal maturity and performance levels around that age are greatest (Malina et al. 2004). Second, Figure 1 illustrates that, among players with the same $P_i^\star$, the performance level of the December-born player is always temporally behind the performance level of his January-born peer. This reflects the stylized fact that relative age advantages are generally beneficial in soccer. Finally, when players approach professional age, relative age differences lose their significance.

For illustration, assume that only two players can get selected by an elite youth academy at selection point A. If the selection is based on current performance levels, both January-born players are chosen. It is obvious that this is not the best choice from a long-run perspective. One might, however, argue that the academy could still pick both players with relative higher ability at selection point B and end up with the two most talented players. The next subsection, however, shows why the RAE might still continue to affect selection decisions.

3.2 The Effect of 'Superior' Elite Academy Training

So far, the model did not incorporate the training effect of soccer elite academies relative to other youth clubs. We assume that elite youth academies indeed offer superior training – the *treatment* – and let the maximum performance level of player i, $P_{id_i}^\star$, depend on treatment $d_i = \{0, 1\}$.[4] We make the established assumption that training and ability are complements in the sense that the former is more effective for individuals with higher potential (see Cunha and Heckman 2007 and Dawid and Muehlheusser 2015). After the point of selection s, player i's maximum performance level depends on whether he receives elite youth academy training ($d_i = 1$) or not ($d_i = 0$): $P_{i1}^\star > P_{i0}^\star$. Player i's realized performance level as a function of time t, thus, reads:

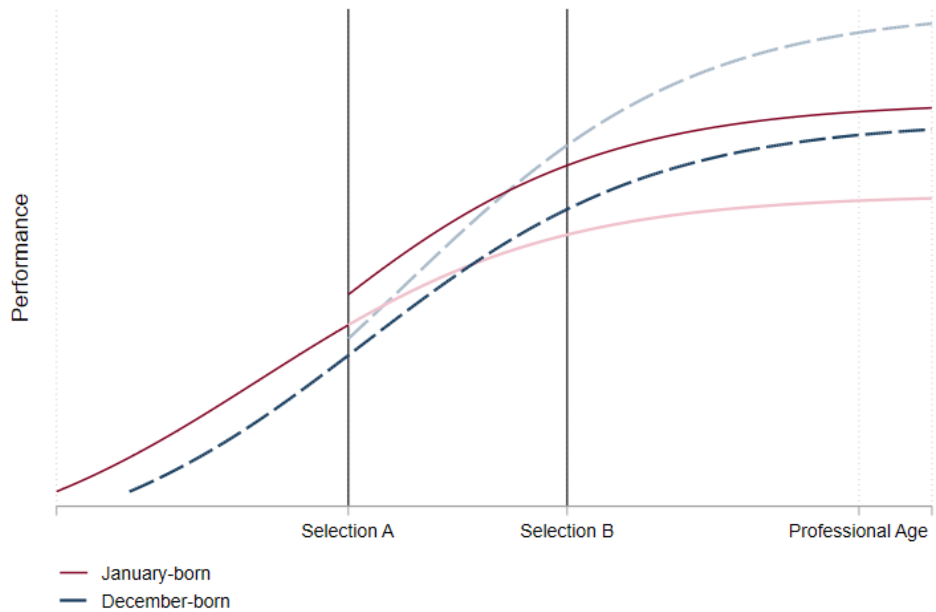
$$\text{for } t \leq s: \qquad P_i(t) = \frac{P_i^\star}{1 + (P_i^\star - 1) \times exp\left(-t - \frac{m_i}{12}\right)}, \tag{2}$$

$$\text{for } t > s: \qquad P_{id_i}(t) = \frac{P_{id_i}^\star}{1 + (P_{id_i}^\star - 1) \times exp\left(-t - \frac{m_i}{12}\right)}. \tag{3}$$

The expansion of the maximum performance level through elite training is illustrated in Figure 2, which shows a relatively untalented January-born (red) and a relatively talented December-born (blue). Out of mathematical ease, the BYA training effect is modelled as an immediate jump to a higher performance development curve after selection.[5]

The lighter red and blue lines represent possible examples of the counterfactuals. Selection of the relatively untalented January-born player lifts his performance level after selection point A, so that even at selection point B it remains higher. This visualizes the main problem caused by the RAE: Although, at selection point A, the relatively talented December-born was currently worse than the relatively untalented January-born, the long-term return of selecting the December-born is much higher. At selection point B, the counterfactual performance of the relatively talented and treated December-born is higher than the performance of the relatively untalented and treated

Figure 2: Performance Development Model: Illustrating Fading and Late Blooming Shooting Stars



January-born. Eventually, the maximum performance level of the relatively talented and treated player exceeds the maximum performance level of the relatively untalented and treated considerably. Meanwhile the performance curve of the relatively untalented January-born presents the case of fading shooting stars vividly: Shining lightly in early selection rounds due to their relative age advantage, they eventually fade before entering the professional soccer stage. In line with existing evidence on the RAE, the model illustrates how the RAE remains even when maturity differences vanish, in particular, in a highly competitive environment. Based on this, we derive our first hypothesis:

Hypothesis 1 – RAE Existence and Path Dependency: *Given (i) selection cut-offs during performance development, (ii) relative age-based performance differences, and (iii) positive effects of elite youth academy training, the relative age effect occurs in a competitive environment and is sustained even when relative age differences fade.*

As described in Section 2.1, the existence of the RAE in professional youth sports is a well-established result, as is its persistence when relative age differences fade. While Hypothesis 1 can therefore almost be considered a stylized fact, it also provides a mechanism behind the RAE which can be related to other fields.

3.3 Marginally Selected Players and Ability

Our theoretical considerations suggest that the ability of those players who were just good enough to get selected into elite academies is not evenly distributed over birth months, although, in the general population, ability is uncorrelated with birth dates. Players who just got selected will be denoted as the *marginally selected*. To define the concept of the marginally selected, we assume that, at selection point A ($t = s$), all players above a certain current performance level $P_\delta(s)$ get selected into youth elite academies ($d_i = 1$), while all players below are rejected:

$$d_i = \begin{cases} 1 & \text{if} \quad P_i(t = s) \quad \geq \quad P_\delta(s) \\ 0 & \text{if} \quad P_i(t = s) \quad < \quad P_\delta(s) \end{cases} \tag{4}$$

The marginally selected is the player for which $P_i(t = s) = P_\delta(s)$. Conditioning the performance level on the birth month m, the performance level of the marginally selected can be denoted as $P_\delta(t|m)$. It becomes evident that the maximum performance level of the January-born marginally selected is lower than the maximum performance level of the December-born marginally selected:

$$\lim_{t\to\infty} P_\delta(t|m = 1) = \quad P^\star_\delta(m = 1) \quad < \quad P^\star_\delta(m = 12) \quad = \lim_{t\to\infty} P_\delta(t|m = 12). \tag{5}$$

This can be further generalized. The marginally selected player of a certain month has a higher maximum performance level than the marginally selected from the previous month apart from the December-January cut-off:

$$P^\star_\delta(m + 1) \quad > \quad P^\star_\delta(m). \tag{6}$$

However, no further statements can be made about the exact relation of the talent of the marginally selected players from different months. A function of the marginally selected depending on month of birth could be convex, concave, or approximately linear, depending on the performance level cut-off at selection $P_\delta(s)$, the point of selection s, and a general scaling parameter determining the course of the function, which we omitted for the sake of simplicity.

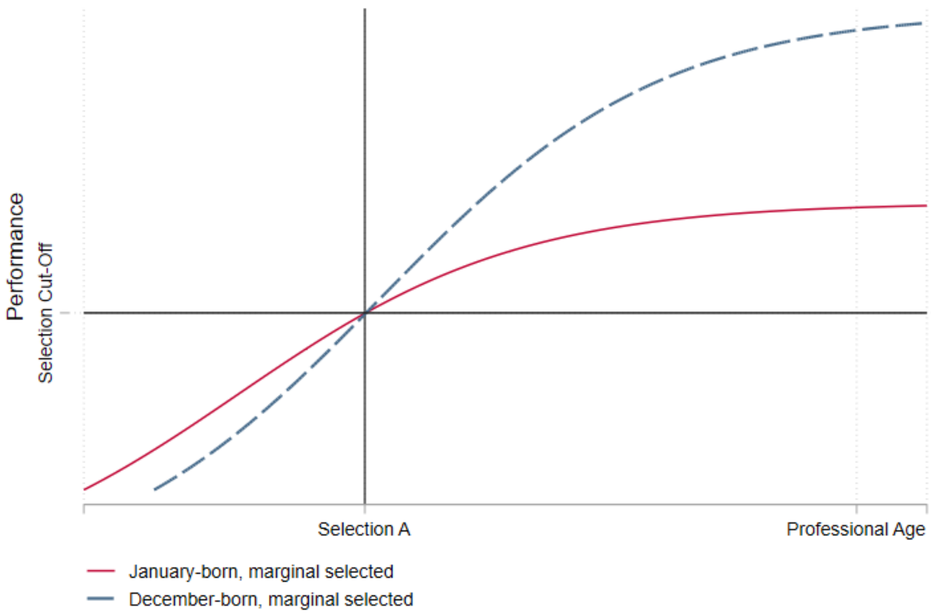
Based on equation 6, we derive the following hypothesis:

Hypothesis 2 – Ability Premium at the Margin: *Among the marginally selected players, relative age disadvantages (later birth months) are positively correlated with players' maximum performance levels:* $Corr(P^\star_{\delta,i}, m_i) > 0$.

As maximum performance levels reflect players soccer-specific ability, we can rephrase this hypothesis as: Among the marginally selected, those with relative age disadvantages have relatively higher unobserved ability. We will test this hypothesis in section 5.2.

Figure 3 illustrates that marginally selected players from different birth months eventually end up having very different maximum performance levels. The player born in December who was just good enough to get selected has a much higher maximum performance level than the January-born marginally selected.

Figure 3: Marginally Selected Players and Maximum Performance Levels



In terms of ability, the upper bound of the players – the *very best* – selected into BYAs per birth month is assumed to be identical for all birth months, as ability is evenly distributed in the general population. However, the lower bound – the *marginally selected* – is skewed with the January-born marginally selected being less talented than the December-born marginally selected. Consequently, the theoretical model suggests that also the average ability of selected January-born players is lower than the average ability of their December-born peers. In general, the average ability – and thus their average maximum performance level $\bar{P}^{\star}(m)$ – of players born in a certain month of the year exceeds the average talent of players born in the previous month apart from the December-January cut-off. Hence, because the marginally selected ability is skewed, also the *average ability* is skewed over birth months.

$$\bar{P}^{\star}(m+1) \quad > \quad \bar{P}^{\star}(m). \tag{7}$$

Being born just before the cut-off, in the end of the year, is thus related to an average ability premium. Based on equation 7, we derive our third hypothesis:

Hypothesis 3 – Ability Premium on Average: *Average performance levels are positively correlated with players' relative age disadvantages in elite youth academies.*

3.4 Maximum Performance Levels and Foregone Market Values

Our main empirical approach is to rely on the highest market values HMV as a proxy for the maximum performance level $P_i^\star$ of players with different treatment status d_i. Concentrating merely on the maximum performance level has the advantage that the individual performance development function can remain unknown. Assuming that development functions are subject to the same development processes and determinants on average, a lot can still be inferred about performance development. Yet, the HMV is not just a function of a player's maximum performance level, but also depends on other characteristics (see Kempa 2022). Let player i's highest market value HMV_i be determined by $P_i^\star$ plus the influence of other factors X_i such as position, youth team or year born and an unobserved error term u_i. Logarithmising the HMV takes the positive skew of market values into account and establishes linearity between HMV and the covariates. When controlling for the covariates X_i, the $logHMV$ is therefore plausibly an applicable proxy for the maximum performance level of players, based on which the RAE in BYAs can be analysed in more detail. Based on that and hypothesis 3, we derive our fourth hypothesis:

Hypothesis 4 – Foregone Market Value: *Among former youth elite academies players, market values are positively correlated with relative age disadvantages; implying that recruiting players more evenly across birth months increases average market values generated by elite youth academies.*

In the following sections, we will empirically test our hypotheses after presenting our data and the setting we analyse.

4 Empirical Setting

4.1 Bundesliga Youth Academies

As a response to the bad performance of the German national team in the World Cup of 1998 and the European Championship of 2000[6], German youth soccer was radically reorganized and modernized. A new licensing regulation, passed in 2001, required every club in the first two divisions (Bundesliga and 2. Liga) to build up Bundesliga youth academies (BYA, German: 'Nachwuchsleistungszentren'). The two primary goals of BYAs are 'internationally outstanding Bundesliga and German national teams' and 'optimal exhaustion of the talent pool' (DFL 2020a)[7]. Linking BYAs' talent selection to the RAE, unobserved ability, and market values, this paper will particularly address the second goal by asking whether BYAs exhaust their talent pool optimally.

BYAs are highly standardized, which will prove to be of great advantage for our analysis.[8] The focus of soccer training is accurately regulated for certain age cohorts. Only from the U15 onwards, BYAs are allowed to conduct 'performance-oriented training', where specializations are stabilized and further developed as direct preparation for a professional soccer career (DFL 2020a).[9] Between the U15 and U19, investments are highest, competition is biggest, and training is most intensive. As players develop most during this performance-oriented training, U15 to U19 squad selection is pivotal.

Today's Bundesliga teams invest millions in their BYAs, while most money is spent on the U15 to U19 teams (Sponsors 2019). Hoffenheim, for instance, has a staff of more than 50 full-time employees responsible for about 150 youth players which play in Hoffenheim's seven BYA teams (Sponsors 2019). In total, 5.400 adolescents played for 279 teams in 54 BYAs[10] in Germany in 2017 (Franzke 2017). To put this figure into context, about 484.000 adolescents between the age of 15 to 19 play soccer in Germany (DFB 2020). Hence, only about 1% of active adolescent players make it to a BYA. From this top one percent, again less than 5% (60–70 players per year) will eventually succeed in getting a professional contract in Europe's top leagues (Franzke 2017, Sponsors 2019). The total investment of the 36 Bundesliga and 2. Liga clubs in BYAs amounted to 177 and 186 million Euro in the seasons 2017/18 and 2018/19, respectively (DFL 2019, DFL 2020b). Overall, more than 1.6 billion Euro have been invested in BYAs since 2001 (DFL 2018). Due to these high investments, it is safe to assume that BYAs indeed offer superior training compared to non-BYA youth clubs, which generally have much fewer resources available.

4.2 Data

In this subsection, we summarize the most important aspects of the data. A more detailed description is provided in Appendix B. We use data on former BYA youth players retrieved from the sports website transfermarkt.de.[11] Next to other information about professional soccer players (name, birth date, strong foot, height, transfer history, etc.), the focus of the website relies on market values. Market values are estimated and discussed by non-expert users for more than 800,000 soccer players worldwide and are regularly updated (Keppel and Claessens 2020). Data from transfermarkt.de was used before in different scientific publications (e.g., Augste and Lames 2011, Grossmann and Lames 2013, Herm et al. 2014, Bryson et al. 2018, and Pérez-González et al. 2020). While the data quality was viewed with criticism first (e.g., Sundermeyer 2009), market values on transfermarkt.de were found to be highly correlated with expert estimates from well-respected sources (Franck and Nüesch 2012). Peeters (2018) finds that transfermarkt.de data on market values performs better than other indicators in predicting a team's strength. Moreover, he does not find evidence for 'wishful thinking bias', which would result in overestimating market values of popular players and teams. Müller et al. (2017) show that the crowd-based estimates from transfermarkt.de are equally accurate as estimates from a multiple regression algorithm and even outperform the algorithm for high-priced players.

When constructing the data set, there was a trade-off between quality and quantity. In other words, the aim was to include as many BYAs as possible without jeopardizing completeness and quality of the data. As a baseline, we examined the aggregated standings of the U19 Bundesliga since 2001. We further supplemented this information with rankings of the most successful BYAs from two different websites (ran.de 2015, fussballfieber.de 2017) and compiled a short list of the 36 most successful BYAs. Yet, going from the top to the bottom of the list, the data became increasingly incomplete. Finally, our data set consists of the U17 and U19 Bundesliga cadres of

Table 1: Descriptive Statistics of Key Variables

	Obs.	Mean	Min	Max	Std. Dev.
HMV in 1,000EUR	2383	1283.702	0	128856.5	7054.63
logHMV	2383	3.83	0	11.8	2.82
BYAyears	2383	2.964	0.1	5.0	1.344
yearBorn	2383	1995.23	1988	2001	3.76
monthBorn	2383	4.69	1	12	3.16
weekBorn	2383	18.17	0.14	52.28	13.8
Born Jan-Jun, dummy	2383	0.715	0	1	0.45
Born Jan-Mar, dummy	2383	0.446	0	1	0.49
Specific Positions, categorical	2281	5.152	1	12	3.38
BuLi Pro, dummy	2383	0.242	0	1	0.43
Right-Footed, dummy	2383	0.469	0	1	0.50
Left-Footed, dummy	2383	0.185	0	1	0.39
Two-Footed, dummy	2383	0.176	0	1	0.38
U19 BYA Team, categorical	2383	9.0	1	17	4.89
U17 BYA Team, categorical	1688	9.3	1	17	4.85
National team, dummy	2383	0.022	0.0	1.0	0.146
Height in cm	1977	182.12	163	202	6.28

Data on the 17 most successful BYA U19 clubs from transfermarkt.de. Players born between 1988 and 2001. Variables on individual player level: 2020 highest market values adjusted for inflation in 1,000EUR (HMV), logarithmised values of HMV (logHMV), years spent in BYA (BYAyears), birth year (yearBorn), birth month (monthBorn), week born in players' respective birth year (weekBorn), dummy variables for being born in the first half (Born Jan-Jun) and first quarter of the year (Born Jan-Mar), specific positions (goalkeeper, center back, right back, left back, central defensive, central midfield, central offensive, right midfield, left midfield, center forward, left wing, or right wing), dummy variable if played in the Bundesliga at least once (BuLi Pro), dummy variables for strong foot (Right-Footed, Left-Footed, Two-Footed), categorical variables for the 17 selected U19 BYA clubs (U19 BYA Team) and the U17 BYA clubs (U17 BYA Team), dummy variable for having played at least once for the German national team (National team, dummy), and height in cm (Height).

the 17 most successful youth teams between 2001 and 2020[12]. Every additional club would have implied incomplete data.

We restrict our data to players with German nationality, as other players might have undergone elite youth academies of different qualities in their home countries before being selected. Additionally, players who were mentioned in BYA cadres but without concordant reference to this in their transfer history were dropped. This was necessary because we need to calculate the number of days that youth players spent in BYAs based on their transfer histories. The final data set contains 3,835 observations. Among them, 2,383 played for a U19 BYA and were born between 1988 and 2001, i.e. could potentially have gotten five full years (U15-U19) of BYA performance-oriented training.

The variable $BYAyears$ captures the time a player spent in one of the 17 BYAs chosen, ranging continuously from zero up to a maximum of five years. We only consider the period of performance-oriented training between the U15 and U19 as competition, investment, and training quality are highest in these years. $BYAyears$ excludes spells during which players were first trained at one of the remaining 37 BYAs and joined one of the 17 selected clubs later.[13] Two main arguments justify this specification: First, close examination of the data reveals that transfers from other BYAs (out of the sample) to the 17 first-tier BYAs (in the sample) are rather rare. Second, not all BYAs provide the same quality of training. More than 70% of total BYA investment is made by the 18 Bundesliga clubs (Sponsors, 2019). Investment in BYAs is, thus, likely to be skewed towards the most successful ones. Hence, $BYAyears$ is an appropriate measure for the years that adolescents received distinguished soccer training, guaranteeing the highest possible level of homogeneity by not treating first- and second-tier BYAs as the same.

In the last two decades, a sharp increase in market values could be observed, so that highest market values are hardly comparable across years. To overcome this issue, we calculate Bundesliga market value inflation rates based on the total market values of all Bundesliga teams' 11 most expensive players in all years between 2005 and 2020. We chose the 11 most expensive players from all 18 Bundesliga clubs in every given year because this yields a 'player basket' of 198 players in each year which remains comparable over time. When merely looking at absolute market values or average market values, the inflation rate might be skewed by the number of players which clubs register in different years. While the number of players per club is also motivated by the 'starting eleven', the 198 players in the player basket is a large enough number that the absolute market values are not influenced too much by individual players.

Absolute market values of all Bundesliga teams' top 11 players and the respective inflation rates are shown in Figure 6 in Appendix C. On first sight, inflation rates of over 30% might appear unrealistic, but Poli et al. (2019) also find inflation rates above 30% for European soccer leagues between 2011 and 2019. Using the calculated inflation rates and the date when a player's highest market value was reached, we convert highest market values to 2020 *inflation-adjusted* highest market values (HMV).[14] In our analyses, we rely on logarithmized values ($logHMV$) to counteract the progressive nature of market values.

All variables are available for all observations except for the players' specific positions and body height which are missing for about 5 and 20% of the observations, respectively. Table 1 reports descriptive statistics of our data set.

5 Empirical Results

5.1 Relative Age Effect in Bundesliga Youth Academies

To test Hypothesis 1 and quantify the RAE in BYAs, we calculate the share of players born in the first half and the first quarter of the year; two well established RAE indicators (see, e.g., Musch and Grondin 2001, Mujika et al. 2009, Tribolet et al. 2019, and Jackson and Comber 2020). Table 2 shows that 71.5% of U19 youth players were born in the first half and 44.6% in the first quarter of the year. Both numbers are well above the equal birthday distributions, 50% and 25%.[15] Table 2 reveals that the RAE is very pronounced across all the 17 BYAs. While certain differences exist, they are not extremely large. The proportion of players born in the first half of the year varies between 77.1% (VfL Wolfsburg) and 65.8% (Schalke 04), while the share of players born in the first quarter of the year ranges between 56.7% (Borussia Dortmund) and 38.0% (Hoffenheim).

Table 6 in Appendix C replicates these findings for the U17 BYA teams, showing an even larger RAE than in U19 BYA teams. The pattern of stronger RAE in U17 teams and a slightly smaller RAE in U19 teams, presumably owing to declining maturity differences, was also found in other studies which we discussed in the literature review (e.g., Malina et al. 2004, Patel 2019, Jackson and Comber 2019).

Relative age differences should have faded to a large extent already in U19 BYA teams, and fully in professional adult leagues. The finding that a significant RAE still exists in U19 BYA teams (see Table 2) as well as in the two top German professional adults leagues (see Figure 8), highlights the persistence of the RAE and is in line with Hypothesis 1. In the context of selection cut-offs during performance development, initial relative age-based performance differences, and positive effects of elite youth academy training, the RAE does not only arise, but it also persists.

Table 2, furthermore, presents average highest market values (HMV). As there is only little variation in the size of the RAE and market values are influenced by various other factors, it is not surprising that the size of the RAE and HMV do not seem to be correlated across clubs. Average HMV, however, need to be treated with caution as values are likely to be affected by a few very expensive players. Yet, it is clear that the existence of the RAE is economically interesting given BYA players' (future) market values.

Figure 4 illustrates the development of the RAE over time, by showing the proportion of players born in the first half of the year between 1985 and 2005. The two main insights from this figure are that, first, the RAE did not decline since the introduction of BYAs and, second, the proportion of players born in the first half of a year is significantly different from 50% (i.e. the equal distribution) at the 95% confidence interval for every birth cohort. At the beginning of the period examined in this paper (birth cohorts 1988 and 1989), the proportion of players born in the first half of the year was around 65%. The RAE indicator increased to around 75% in the following ten years and remained approximately unchanged since then.

Table 2: The Relative Age Effect: Summary Statistics by U19 BYA

	% born Jan-Jun	% born Jan-Mar	Mean HMV in 1,000€	Obs.
Full sample	**71.5**	**44.6**	**1283.702**	**2383**
VfL Wolfsburg U19	77.1	45.0	1054.406	131
Borussia Dortmund U19	75.6	56.7	1960.316	127
FC Bayern München U19	75.4	46.5	3599.560	114
VfB Stuttgart U19	74.4	45.1	2349.844	134
Bayer 04 Leverkusen U19	73.1	50.0	914.710	130
TSV 1860 München U19	71.8	46.6	1067.845	163
Eintracht Frankfurt U19	72.6	42.5	262.893	146
Werder Bremen U19	71.5	47.2	886.345	144
1.FSV Mainz 05 U19	71.5	42.3	1270.778	130
SC Freiburg U19	71.2	42.9	707.105	156
Hamburger SV U19	70.8	44.6	966.160	129
TSG 1899 Hoffenheim U19	70.1	38.0	1008.937	137
1.FC Köln U19	69.5	42.4	1013.030	151
Borussia Mönchengladbach U19	69.2	45.3	1101.972	159
Hannover 96 U19	69.7	41.0	528.013	121
Hertha BSC U19	68.7	43.6	956.346	164
FC Schalke 04 U19	65.8	40.4	2867.211	147

Data on the 17 most successful BYA U19 clubs from transfermarkt.de. Players born between 1988 and 2001. Differences in the number of observations per club can be attributed to missing data and different proportions of foreign youth players, who are not considered here.

Overall, the descriptive statistics show that the RAE has not declined, but was rather amplified since the introduction of BYAs. The primary goal of BYAs, the 'optimal exhaustion of the talent pool' (DFL, 2020a), is thus probably missed.[16] As unobserved ability is plausibly independent of birth dates, the preferred selection of relatively older adolescents suggests that talent is lost: Some late blooming shooting stars are deprived of the chance to shine.

Figure 4: The RAE over Time: Proportion of BYA Players Born in First Half of the Year by Age Cohorts

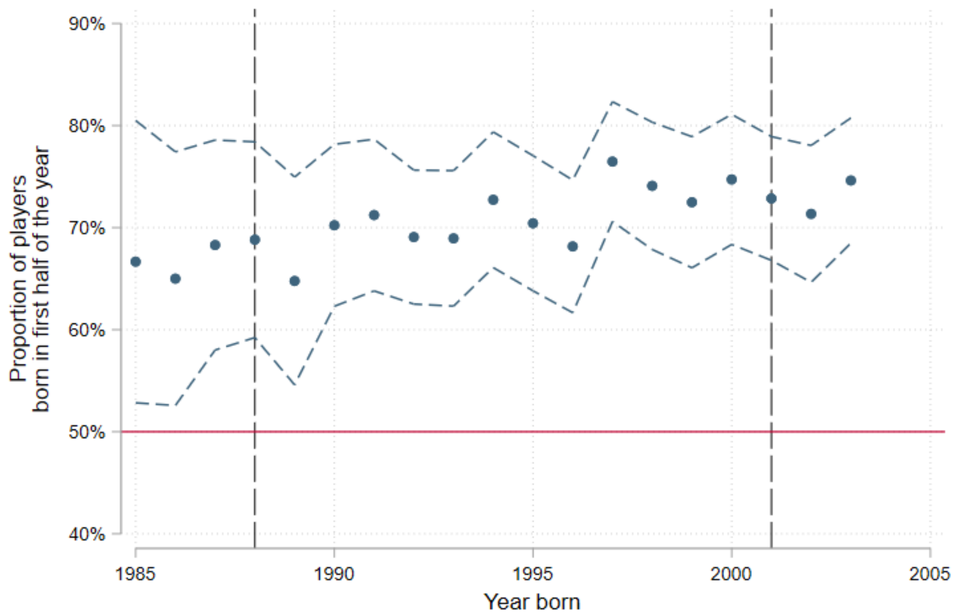


Figure displays values of all former U19 BYA players born between 1988 and 2001. The respective age cohorts are indicated by vertical lines. Confidence intervals at 95% and equal distribution as reference.

5.2 Relative Age Advantages, Training, Market Values, and Ability

To test our Hypotheses 2 and 3 – that performance levels are positively correlated with players' relative age disadvantages, at the margin of getting selected as well as on average – we first run regressions of *logHMV* on players' quarters of birth. In doing so, the *logHMV* serve as a proxy for players' maximum performance levels. In all specifications, we control for year of birth and U19 club fixed effects. Column 1 of Table 3 shows a clear picture: BYA players born later in the year reach significantly higher market values during their careers. Specifically, former U19 youth players that were born in the third (fourth) quarter of the year reach 41.9% (58.7%) higher market values compared to their peers born in the first quarter of the year. Our results are supported by quantile regressions (see Table 8 in Appendix C).

At first sight, this result appears to contradict the findings of Pérez-González et al. (2020) and Fumarco and Rossi (2018) who show that professional soccer players' monetary valuations are not significantly correlated with their relative age. However, this can partly be explained by different

Table 3: OLS Regressions of Market Values, BYA Training, and Birth Quarters

	(1) logHMV	(2) logHMV	(3) BYAyears	(4) logHMV	(5) logHMV
Q2: Apr-Jun	0.0599	0.303	-0.0656		0.0122
	(0.126)	(0.186)	(0.0634)		(0.129)
Q3: Jul-Sep	0.364***	0.315*	-0.191**		0.322**
	(0.140)	(0.178)	(0.0759)		(0.135)
Q4: Oct-Dec	0.462***	0.212	-0.224**		0.453***
	(0.177)	(0.213)	(0.0965)		(0.168)
BYAyears				0.614***	0.532***
				(0.0384)	(0.0396)
Height (cm)					0.0627***
					(0.00937)
Position Control	No	No	No	No	Yes
Strong Foot Control	No	No	No	No	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes
U19 Club FE	Yes	Yes	Yes	Yes	Yes
Observations	2383	577	2383	2383	1873
R squared	0.231	0.276	0.0826	0.306	0.332
Sample	Full	BuLiPro	Full	Full	Full
Cov. with weekBorn	2.626	2.062	-1.383	2.626	2.092
Cov. with BYAyears	0.736	0.106	1.807	0.736	0.650

The full sample includes all former U19 BYA players who were born between 1988 and 2001. The BuLiPro sample in column 2 includes only players that have at least once played in the Bundesliga or 2. Bundesliga during their career. In columns 1, 2, 3, and 5 players born in the first quarter of the year (Q1: Jan-Mar) are omitted and constitute the baseline. Because the logarithm of the market values is the dependent variable (in column 1, 2, 4 and 5), the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, a selection of coefficients and their respective percentage changes are shown in the format $\beta = x\%$: $-0.2 = -18.1\%$, $0.212 = 23.6\%$, $0.322 = 38.0\%$, $0.364 = 43.9\%$, $0.453 = 57.3\%$, $0.462 = 58.7\%$, $0.532 = 70.2\%$, and $0.614 = 84.8\%$. The 'position control' consists of 12 different position specializations. The 'strong foot control' refers to player's strong foot: left, right, or both. The smaller sample in column 5 can be explained by the fact that the 'height'-variable is missing for some players. The last two rows report the unconditional covariance of the dependent variable with the variables 'weekborn' and 'BYAyears', respectively. Heteroskedasticity-robust Huber-White standard errors in parentheses. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

samples. We study former U19 BYA players, while Pérez-González et al. (2020) and Fumarco and Rossi (2018) concentrate only on professional adult players. Our results, however, are more in line with Ashworth and Heyndels (2007) who analyze the (estimated) wages of professional players from the German Bundesliga for two seasons between 1997 and 1999 and find that the late-born players receive a wage premium. In Column 2 of Table 3, we similarly restrict our sample to those former U19 BYA players who have at least once played in one of the top two German professional adult leagues. When doing so, the association between players' quarter of birth and their market values declines and loses significance; while still being positive. Our findings are thus generally in line with the literature. Within the sample of those players who made it to the professional stage, the association between players' monetary valuations and relative age does neither seem to be clearly positive nor clearly negative. In fact, we would not expect significant differences, if players with relative age advantages are a bit less talented, on average, but receive a bit more time of distinguished elite academy training. Yet, in our sample of all former U19 BYA players – including the fading shooting stars that never make it to the professional stage – the unambiguously positive association between relative age disadvantages and market values is striking.

As players' quarter of birth is, by nature, uncorrelated with their talent, skill and thus maximum performance level, the positive correlation between relative age disadvantages and $logHMV$ can be explained by unobserved ability and sample selection: Players born towards the end of the year need to compensate with higher ability for their relative age disadvantage in order to get selected by BYAs.

Still, the simple estimates presented in Column 1 of Table 3 might be biased: Column 3 of Table 3 shows that players with relative age disadvantages are selected relatively later by BYAs and therefore receive, on average, less distinguished elite training. At the same time, years of BYA training are strongly positively associated with market values (see Column 4 of Table 3). Moreover, other factors such as players' positions, their height or even their strong foot, might be correlated with their relative age, unobserved ability and thus market values. In Table 7 in Appendix C, we report additional results on the association between players' market values and player specific characteristics. We obtain three main results: (i) the positions of former BYA players that are associated with the highest market values are right wing, left wing, and central midfield, while center backs, right midfielders, and left midfielder reach the lowest market values, (ii) taller players reach on average higher market values, and (iii) two-footed players reach on average significantly lower market values than right-footed players, which indicates that BYAs overrate the importance of two-footedness when it comes to talent selection.

In Column 5 of Table 3, we therefore add controls for these player characteristics and also control for the number of BYA training years the players received. This does, however, not affect the results: The association between former BYA players' relative age and their market values is still significantly positive. Players born in the last quarter of the year still have market values which are 57.3% larger than the market values of their peers born in the first quarter of the year. This indicates that in the sample of BYA players, relative age and unobserved ability are not evenly distributed. In other words, relatively disadvantaged players that still get selected are relatively more talented. While this finding strongly supports Hypothesis 3 – that average performance levels are positively correlated with relative age disadvantages in elite youth academies – it does not directly verify Hypothesis 2. Based on the regressions in Table 3 alone we cannot conclude that it is indeed the marginally selected players who drive this positive correlation between relative age disadvantage and ability as stated in Hypothesis 2.

To test Hypothesis 2, we therefore need to go one step further. We do this by slightly abusing an instrumental variable (IV) approach: Instead of using an unbiased IV estimate to measure a causal effect, we use a biased IV estimate to learn about the covariance of some part of the error term (unobserved ability) and the instrument (relative age disadvantage) at the margin of getting selected. In the following paragraphs, we describe how we do this in detail.

The following two-stage least squares (2SLS) estimation equations provide the basis for testing Hypothesis 2:

$$BYAyears_i = \pi_0 + \pi_1 weekBorn_i + \delta_y + \gamma_c + \nu_i \quad (8)$$

$$logHMV_i = \beta_0 + \beta^{IV} \widehat{BYAyears}_i + \delta_y + \gamma_c + \epsilon_i, \quad (9)$$

where we first use players' relative age as an instrument for the years of BYA training they received (first stage, equation 8), and then regress players' market values on their instrumented years of BYA training (second stage, equation 9). Here, δ_y denote year of birth fixed effects and γ_c U19 BYA club fixed effects.

The instrument is motivated by the fact that relatively older adolescents have a higher propensity of getting selected early by BYAs. The literature suggests that performance differences between boys of contrasting maturity status are most pronounced between the age of 13 and 16 (see Section 2.3) so that boys with a relative age advantage are more likely to be selected at the U15 stage. As the maturity advantage of relatively older players decreases subsequently, some relatively younger players make it into the team at later stages. In the U19 team, then, relatively older players should have gotten more years of BYA training on average. Moreover, Column 3 of Table 3 and also Columns 1 and 3 of Table 4 show that relative age is significantly correlated with $BYAyears$.

Following the idea of Angrist and Krueger (1992), birthdays are generally a valid instrument as they are plausibly random and uncorrelated with possible confounders. As Musch and Grondin (2001) argue, there are no seasonal circumstances which could explain why youth players are more likely to be selected by elite youth academies apart from relative age. Players born in December of one year and those born in January of the next year are exposed to the same conditions while growing up[17]. It is therefore safe to assume that unobserved ability is distributed equally across birth months in the whole population.

However, unobserved ability is likely not independently distributed in our sample of those players that were selected into BYAs while it is in the general population. As unobserved ability likely influences how early a player gets selected and also how much BYA training he gets, it is also linked to market values. In our sample, the exclusion restriction – the necessary assumption to identify causal effects using IV – therefore likely fails. The IV estimation is biased. This implies that we cannot identify the causal effect of $BYAyears$ on market values. Still, we can use the biased IV estimation to learn something about the covariance of the instrument (the relative age) and the confounder (unobserved ability).

Note that the biased IV estimator, not conditioning on fixed effects for simplicity, can be expressed as follows:[18]

$$\beta^{IV} = \underbrace{\beta^{unbiased}}_{>0} + Cov(\epsilon_i, weekBorn_i) / \underbrace{Cov(BYAyears_i, weekBorn_i)}_{<0}, \tag{10}$$

where $\beta^{unbiased}$ is the unbiased (causal) effect of one additional year of BYA training on players' market values, which we cannot estimate but are safe to assume to be non-negative given the large amount of resources used to train and promote youth players in BYAs compared to soccer training outside of the BYA elite academies. Columns 4 and 5 of Table 3 also suggest that the effect of one additional year of BYA training is significantly positive. Moreover, we know from Column 3 of Table 3 that the covariance of *weekBorn* and *BYAyears* is negative. This is further supported by the first stage of the 2SLS estimation reported in Columns 1 and 3 of Table 4.

The only part of equation 10 which we do not know the sign of is the covariance of the 2SLS regression error ϵ and *weekBorn*. Note from the discussion of our instrument above that the only part of the 2SLS regression error that can be correlated with relative age (*weekBorn*) is players' unobserved ability which determines if they get selected despite having relative age disadvantages. If the sign of the 2SLS estimator β^{IV} is negative, we can therefore conclude that covariance of the 2SLS regression error ϵ and the relative age disadvantage *weekBorn*[19] is positive: $\beta^{IV} < 0 \Rightarrow Cov(\epsilon_i, weekBorn_i) > 0$.

In fact, Columns 2 and 4 of Table 4 show that the 2SLS estimator is significantly negative at the 10% level. Based on equation 10, we thus conclude that the covariance of unobserved ability and relative age disadvantage is positive.

Using another feature of IV estimation, we can go even one last step further. The IV estimator β^{IV} does not estimate an average effect, but a local average treatment effect (LATE). In general, this is the average 'treatment' effect for those that got induced into 'treatment' by the instrument and were not 'treated' otherwise. In our context, this refers to those that only got more (or less) years of BYA training exclusively because of their relative age. It is apparent that this group is almost analogous to our definition of the *marginally selected*, for which small differences determine if they get selected into BYAs or not. As β^{IV} can be considered a (biased) local estimator of the training effect for the *marginally selected*, we can also consider $Cov(\epsilon_i, weekBorn_i)$ as the local covariance between the *marginally selected* players' unobserved ability and their relative age disadvantage. The negative sign of the (biased) LATE in Columns 2 and 4 of Table 4 therefore implies that the covariance of unobserved ability and relative age disadvantages is negative for the *marginally selected*. Hence, we can conclude that there is an ability premium at the margin of getting selected as we have argued theoretically in Section 3 and have stated in Hypothesis 2.

In conclusion, we have utilized a biased IV estimator to establish that the covariance of unobserved ability and relative age disadvantage is positive at the margin of getting selected. This does not only support Hypothesis 2, but also explains why we find an ability premium for those with relative age disadvantages on average (Hypothesis 3).

Table 4: Two-Stage Least Squares: Identifying the Marginally Selected Talent Bias

	First-Stage	Second-Stage	First-Stage	Second-Stage
	(1)	(2)	(3)	(4)
	BYAyears	logHMV	BYAyears	logHMV
$\widehat{BYAyears}$		-1.617*		-1.723*
		(0.863)		(0.881)
weekBorn	-0.00646***		-0.00647***	
	(0.00198)		(0.00197)	
Birth Year FE	Yes	Yes	Yes	Yes
U19 Club FE	No	No	Yes	Yes
Observations	2383	2383	2383	2383
F-Test	10.61		10.76	

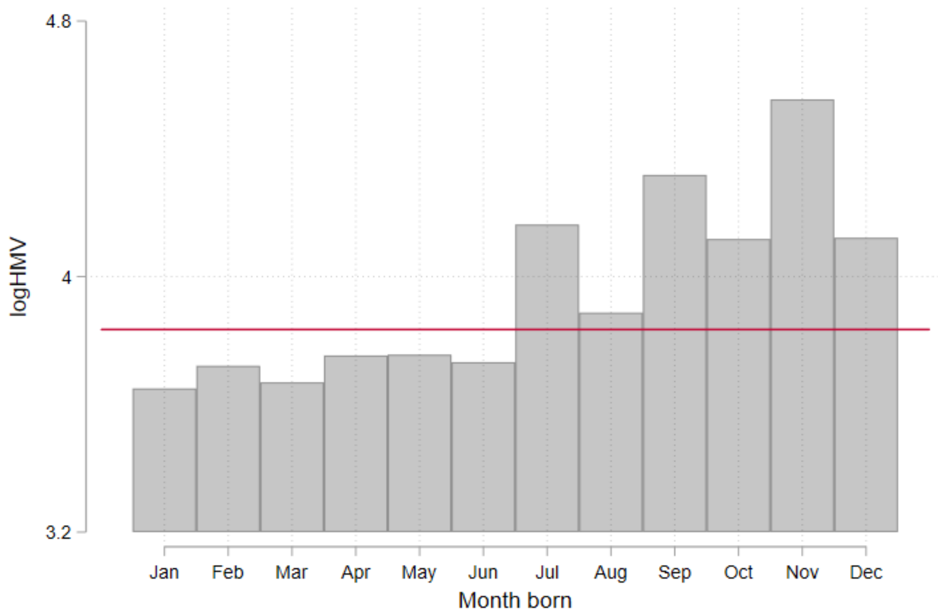
The sample includes all former U19 BYA players who were born between 1988 and 2001. Columns 1 and 3 show the first-stage results. Columns 2 and 4 show the second-stage results. Because the logarithm of the market values is the dependent variable (in columns 2 and 4), the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, coefficients can be interpreted as changes of the following size: -77.4% (column 2) and -82.1% (column 4). The Wu-Hausman F-statistics of the 2SLS regressions are 16.57 and 18.73, respectively, when running regressions analogue to those in columns 2 and 4 without robust standard errors, as required by the Wu-Hausman test. We can therefore reject that $BYAyears$ is exogenous in OLS regressions. Heteroskedasticity-robust Huber-White standard errors in parentheses. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Finally, we want to note that the instrumental variable approach is very promising for further analyses of the mechanisms involved in the RAE. New data on both elite and non-elite youth players would promise two advantages. First, the prevalence of the RAE in BYAs would likely further increase the instrument's strength. Second, the exclusion restriction would hold, as talent in the whole population is independent from birth dates so that a causal effect of elite youth academy training on market values could be estimated. IV regressions can therefore be a promising starting point for further research.

5.3 Foregone Market Values

This section aims to test Hypothesis 4. The key questions are: Do BYAs forego higher market values by not selecting youth players evenly across birth months? And, can we quantify how much additional market value, if any, BYAs could generate when eliminating the RAE in talent selection? To answer these questions, we start with the findings derived in the previous section. We found that, in the sample of former BYA youth players, those with relative age disadvantages have on average higher ability levels; in particular, at the margin of getting selected. This implies, for example, that the December-born player who was *just not selected* by a U15 BYA is likely to have higher (unobserved) ability than the January-born player who *just got selected* by the U15 BYA. From Figure 3 in Section 3, it is apparent that this exemplary selection decision can be improved: By selecting the marginal December-born instead of the marginal January-born, ability in the BYA could be increased; as well as market values in the long term.

Figure 5: Mean Logarithmized Highest Market Values by Month of Birth



Note: The sample includes all former U19 BYA players who were born between 1988 and 2001. The red line refers to the average logHMV of all players born between 1988 and 2001.

Of course, one would need clairvoyant abilities to predict the actual performance potential of youth players. The relationships established in Hypotheses 2 and 3 are obviously not valid for each individual player. Still, the positive association of relative age disadvantages and unobserved ability, which exists on average, indicates that selecting more evenly across months of birth will also have effects on average. To test Hypothesis 4, we therefore compare groups of players, not individual players. The idea is rather simple: We compare a group of players that is representative for the current average level of ability in BYAs to a group of players that is representative for the average level of ability that would emerge if selection was independent of the RAE.

Specifically, using month of birth as a grouping variable, we compare a credible *status quo* reference group (*ref*) to a plausible *state-of-no-RAE* group D. By taking differences in (conditional) mean *logHMV*, we obtain an estimate for the costs of the RAE in BYAs.

$$\frac{HMV^D - HMV^{ref}}{HMV^{ref}} \approx logHMV^D - logHMV^{ref} = \beta_D^{ref} \tag{11}$$

Figure 5 shows the distribution of average *logHMV* by birth month. In line with the regression results in Table 3, *logHMV* increase over birth months. The mean *logHMV* (the red horizontal line) is surpassed in the middle of the year. We, therefore, consider players born in June and July as the natural choice for the reference group for the *status quo* talent level in BYAs. Alternative choices for the *status quo* group are players born between May and June or players born between July and August. As *logHMV* of these players are just below (above) the mean *logHMV*, we expect to obtain upper (lower) bound cost estimates when using these alternative *status quo* groups as a baseline.[20]

We now turn to the question of how high the average talent level in BYAs could be when eliminating the RAE. The natural choice for the *state-of-no-RAE* group are players born between September and December. We split this group into a September-October and a November-December group. In forming these groups, we can account a bit for the fluctuations in *logHMV* across birth months (see Figure 5). We consider the November-December group as a more optimistic *state-of-no-RAE* group, as among these players the average talent level should be the highest, while the September-October *state-of-no-RAE* group is a more conservative choice.

Building on equation (11), we estimate the cost of the RAE in BYAs using OLS regressions. For each *status quo* reference group (*ref*), we estimate differences in means with respect to a set of *state-of-no-RAE* groups $\Gamma = \{d_1, d_2, ...D\}$. Groups are defined by month of birth (*monthBorn*) of player i. Building on the notation introduced above[21], we estimate the following regression model:

$$logHMV_i \quad = \quad \beta_0 + \sum_{d=1}^{D} \beta_d \times \mathbf{1}[monthBorn_i \in d] + X_i\Lambda + \gamma_c + \delta_y + u_i \qquad (12)$$

$$\text{if} \quad monthBorn_i \in \{\, ref \;\cup\; \Gamma \,\}.$$

Our coefficient of main interest, β_d, captures the difference in mean *logHMV* between players born in the *state-of-no-RAE* month of birth group d and players born in the *status quo* reference group (ref). As shown in equation (11), we can interpret β_d as the proportional average market value premium of the *state-of-no-RAE* group relative to the *status quo* reference group. In the regressions, only players who were born in either the *status quo* reference group or one of the *state-of-no-RAE* groups are considered. All players born in other months are not included in the respective regression samples.

Table 5 reports the coefficient estimates for different *status quo* reference groups and *state-of-no-RAE* groups. Columns 3 to 5 show estimates with players born between May and August as the reference group which, as discussed above, constitute the most credible baseline for how high the average talent level currently is in BYAs. Taking players born between September and October and those born between November and December as the *state-of-no-RAE* groups and following our reasoning above, we find that BYAs could generate 30.6 to 72.8% higher HMV when eliminating the RAE.

Taking the natural choice for the *status quo* group, players born in June and July, as reference (Column 4 of Table 5), eliminating the RAE in talent selection is associated with 38.8 to 64.5% larger HMV. The estimator of the September-October *state-of-no-RAE* group, however, is not statistically significant. The estimate of the November-December *state-of-no-RAE* group is meanwhile statistically significant at the 5%-level. Overall, the estimates show that eliminating the RAE could lead to sizeable effects on average market values of former elite youth players.

Table 5: The Foregone Market Values and the Relative Age Effect in Bundesliga Youth Academies

	(1) logHMV	(2) logHMV	(3) logHMV	(4) logHMV	(5) logHMV
Reference Months	Jan-Feb	Mar-Apr	May-Jun	Jun-Jul	Jul-Aug
May-Jun	0.0329	0.00994			
	(0.156)	(0.170)			
Jul-Aug	0.229	0.231	0.193		
	(0.158)	(0.175)	(0.187)		
Sep-Oct	0.470**	0.457**	0.450**	0.337	0.262
	(0.196)	(0.209)	(0.218)	(0.214)	(0.220)
Nov-Dec	0.590***	0.556**	0.540**	0.478*	0.365
	(0.226)	(0.234)	(0.246)	(0.248)	(0.247)
Birth Year FE	Yes	Yes	Yes	Yes	Yes
U19 Club FE	Yes	Yes	Yes	Yes	Yes
Observations	1846	1606	1069	711	680
R^2	0.244	0.215	0.227	0.255	0.269

The sample includes all former U19 BYA players who were born between 1988 and 2001. Column 1 compares logarithmized market values of youth players born in January and February with those born in the other birth months groups shown. In the other columns, the reference birth month combinations are March and April (column 2), May and June (column 3), June and July (column 4), and July and August (column 5). Because the logarithm of the market values is the dependent variable, the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, a selection of coefficients and their respective percentage changes are shown in the format β = x%: 0.193=21.3%, 0.23=25.9%, 0.337=40.1%, 0.45=56.8%, 0.478=61.3%, 0.54=71.6%, and 0.59=80.4%. Heteroskedasticity-robust Huber-White standard errors in parentheses. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Considering the alternative *status quo* groups, May-June (Column 3) and July-August (Column 5), we see that the estimates follow the expected patterns. Mean differences are larger in size and statistically significant for the May-June reference group, while they are smaller and insignificant for the July-August reference group. Moreover, the November-December *state-of-no-RAE* group leads to generally larger mean differences than the September-October group. Hence, the difference in *logHMV* between the July-August and the September-October groups (Column 5, 30.6%) plausibly constitutes a lower bound for the opportunity costs of the RAE in BYAs. Likewise, the difference in means between the May-June and November-October groups (Column 3, 72.8%) can be considered an upper bound. The estimates are robust to controlling for individual player's height and strong foot (see Table 9 in Appendix C).

Overall, these results support Hypothesis 4: BYAs forego higher average market values by maintaining selection characterized by the RAE instead of selecting youth players evenly across birth months. Market values differ considerably over birth months with players born later in the year having generally higher market values. Eliminating the RAE in talent selection, average unobserved ability levels in BYAs could be increased. We find that this higher average ability would translate into 30.6 to 72.8% higher market values of former U19 BYA players. Hence, professional German clubs could generate substantially more value through their BYAs than they are currently doing. To express this in numbers, if the average former BYA player sells for 1.284 million EUR today (see Table 1), he could sell for 1.677 to 2.219 million EUR in absence of the RAE.

One could criticize our calculations arguing that relatively younger players might compensate for their relative disadvantage with greater effort and might even adapt their performance level to the relative physiological advantage of their peers (see Votteler and Höner 2013 and Mann and Ginnecken 2017). In other words, relative age disadvantages could come with positive spillovers from relatively advantaged team mates. This is also discussed as *peer effects* in the literature (see, e.g., Ashworth and Heyndels 2007). A disproportionately higher share of older players within a youth team would then not be an indication of the RAE, but could be a strategic tool to promote a few exceptionally promising players (see Section 2.2). Two important observations speak against this argumentation. First, evidence suggests that relatively younger and relatively less physically developed youth players tend to receive less match playing time than their relatively older and stronger peers (Vaeyens et al. 2005, Deprez et al. 2015, and Sæther 2016). This is also in line with the observation that clubs aim to be successful at all stages and utilize the RAE in pursuit of short-term success (Jimenez and Pain 2008). For relatively younger players, positive spillovers during training might thus be balanced out by the negative effect of less and shorter match experience. Second, the RAE can still be observed at the professional level (see Sierra-Diaz et al. 2017 and Figure 8 in Appendix C) which indicates the inability of elite youth academies to identify their top players independently of the RAE.

Finally, we need to emphasize again that we only estimate the possible gains of eliminating the RAE. We have not considered the opportunity costs of the *relative maturity effect* which co-exists next to the RAE and is also influential in BYA talent selection (see Malina et al. 2000). The overall costs of selecting along the lines of momentary instead of potential performance levels are therefore expected to be even higher.

6 Conclusion

This paper investigates the relative age effect (RAE) in German elite youth soccer. We develop a simple theoretical model which illustrates the underlying mechanism in talent selection and the negative consequences of the RAE. Based on our model, we derive four hypotheses: First, given selection cut-offs during performance development, relative age-based performance differences, and positive effects of elite youth academy training, we hypothesize that the RAE occurs in an competitive environment, and is sustained even when relative age differences fade. Second, we hypothesize that among *marginally selected* players, relative age disadvantages (later birth months) are positively correlated with players' unobserved ability. Third, we derive from the second hypothesis that, among all youth players selected by elite academies, average ability levels are positively correlated with players' relative age disadvantages. Fourth, we postulate that elite youth academies forego the creation of higher market values by not selecting youth players more evenly across months of birth. Our data includes information on the players of the most successful German Bundesliga Youth Academies (BYAs) for the period 2002–2020.

Even though the RAE has been well-documented since decades, our results show that it still exists in German elite Bundesliga youth academies. While there is no reason to assume that ability is not distributed evenly across birth dates, we find that 71.5% (44.6%) of the players in BYAs in the under 19 teams were born in the first six (three) months of a year. In this competitive environment with key date assessments, relatively older players within a cohort accordingly have a higher probability of getting selected. Moreover, the RAE also persists in professional German adult soccer. This supports our first hypothesis.

We show that elite youth players who were born towards the end of the year, on average, reach significantly higher market values during their career. This suggests that, within the sample of all players that were selected by BYAs, elite youth players must compensate for their relative age disadvantage by having higher ability. This is also expressed in market values and supports our third hypothesis.

With respect to our second hypothesis, we rely on an instrumental variable approach to learn about the covariance of some part of the error term (unobserved ability) and the instrument (relative age disadvantage). We find that relative age disadvantages are positively correlated with unobserved ability *at the margin* of getting selected. This supports our second hypothesis and explains the significantly positive relation between market values and relative age disadvantages.

Finally, the results of our analyses reveal that the RAE causes substantial financial losses for the clubs as it reduces players' market values. According to our estimations, future market values of BYA players could be between 30.6 and 72.8% higher if the clubs were able to eliminate the RAE in talent selection. These figures support our fourth hypothesis and show that the RAE does not only cause substantial costs in terms of team performance, but also in the financial dimension.

The mechanisms we have described are also relevant for talent identification, development, and recruitment outside of sports. There are many (structural) reasons that give individuals short-term advantages or disadvantages (e.g., parental background, gender, ethnicity, networks, ordinal ranks, language, mobility, environmental shocks etc.) which might mask their real potential. This can be the case in several contexts: firms' hiring or promotion decisions, the admission to certain schools or study programs, tracking decisions in school, allocation to math or reading groups or other enrichment programs in (primary) school, the award of scholarships, program participation among unemployed or in development aid, et cetera. Failing to account for these short-term factors reproduces and deepens inequalities (e.g., Hanushek and Rivkin 2009, and Murphy and Weinhardt 2020) and poverty (e.g., Balboni et al. 2022), leads to a waste of talent, and makes later compensatory investments more expensive, especially if relative disadvantages occur in early childhood (see Cunha and Heckmann 2007).

We show that distinguishing between adolescents' current and potential performance levels is crucial for the efficient allocation of talent and resources. Beyond that, our paper contributes in two ways to the debate on how to improve the allocation of talent in society. First, we offer a conceptional framework and an exemplary application, highlighting the key mechanisms and implications of talent selection in the nexus between current and potential performance levels. Second, we show that the economic gains can be large if initial differences are eliminated rather than perpetuated.

Notes

[1]Economists have, thus, frequently used sports data to analyse relevant questions from their field (see, e.g., Mechtel et al. 2011, González-Díaz et al. 2012, Bryson and Chevalier 2015, Feess et al. 2015, Kitchens 2015, Berger and Nieken 2016, Cohen-Zada et al. 2018, Muehlheusser et al. 2018, Harb-Wu et al. 2019, Hill and Remer 2020, and Flepp and Franck 2021).

[2]Specifically, Dawid and Muehlheusser (2015) show that, in early selection stages, pro-competitive selection, counter-competitive selection, and no selection can all be the optimal policies, depending on the size of the relative age differences, the timing of selection, and the degree of heterogeneity with respect to ability in the population. However, as relative age advantages are large in youth soccer (Malina et al. 2007, Rommers et al. 2018), the findings of Dawid and Muehlheusser (2015) imply that pro-competitive selection policies are certainly not optimal at all stages of selection in the context of elite youth academies.

[3]As we describe below, the functional form is consistent with the knowledge of male peak height velocity (see Malina et al. 2004, and Walker 2016). We do not elaborate on the exact shape of the performance development function in this section because our analysis will focus on the limit of the function and uses a player's highest market value as a proxy for his maximum performance level. However, the performance development function can be specified for other applications by introducing additional parameters. Appendix A provides a short discussion of adoptions of this theoretical approach.

[4]For now, we consider d_i as a binary but, without loss of generality, we can relax this assumption later and consider d_i as the continuous treatment status of each player, capturing heterogeneous treatment duration.

[5]Note that the model is mainly used for illustrative reasons. The results discussed in this section would also apply for (at least some) versions of the performance development function in which BYA training does not yield an immediate jump, but an increased slope in the first years of BYA training.

[6]In 1998, the German national team lost 0:3 against Croatia in the quarter finals of the World Cup. In 2000, they were already eliminated from the European Championship in the group stage.

[7]After another debacle in the 2018 World Cup, the DFL and DFB started the 'Projekt Zukunft' which aims at improving and modernizing the BYAs in Germany. It is, for instance, planned to decrease short-term competition. Moreover, measures such as bio-banding are discussed which might help to mitigate the RAE in BYAs.

[8]Each club needs to employ several full-time coaches, at least one full-time physiotherapist, and a full-time sports psychologist. Boarding schools and certain fitness and recreational facilities need to be built up. Regular medicine checks are mandatory and squad sizes are capped (DFL 2020a).

[9]The U8 to U14 youth teams are characterized as 'basic training' and 'development training' where having fun with soccer is still paramount and basic soccer skills and specializations are developed.

[10]Although only 36 clubs play in the Bundesliga and 2. Liga, relegated teams continued having licensed BYAs even in lower leagues, so that, in 2021, 56 BYAs existed in Germany.

[11]Being owned by Springer Verlag (Sundermeyer 2009), transfermarkt.de has more than three million unique monthly users in Germany (Statista 2020) and one billion page views per month globally (Keppel and Claessens

2020).

[12]As market values were also significantly affected by the Covid-19 pandemic, we concentrate on the period prior to the pandemic to avoid possible data distortions.

[13]For example, a player who was part of a second-tier BYA for two years and joined one of the 17 first-tier BYAs for the remaining three years ends up with $BYAyears = 3$ in our data set.

[14]For the sake of brevity, we refer to these *inflation-adjusted highest market values* as *highest market values.*

[15]Figure 7 in Appendix C shows the number of children born in Germany for the years 1990 and 2000 by birth month. Birth figures were highest in July, August, and September and smallest in November and December. However, the differences between the months are relatively small. Importantly, the total number of children born cannot account for a distribution of birth dates skewed towards the beginning of the year in the period examined.

[16]Note that Dawid and Muehlheusser (2015) show that the empirical observation of the RAE cannot per se be taken as an indication of non-optimal selection practices, since the RAE is present to some extent even under the optimal selection policy. However, the large RAE in BYAs strongly suggests that BYAs are not optimally using their talent pool.

[17]It should also be noted that, in Germany, age cut-offs for school enrolment are set in the summer months, not between December and January, while legal rights and obligations are determined by absolute age, not by year of birth.

[18]In general, equation 10 can be derived by expressing the unconditional IV estimator as the ratio of the slope coefficients of the reduced form and first stage: $\beta^{IV} = \frac{Cov(Y,Z)/var(Z)}{Cov(D,Z)/var(Z)} = \frac{Cov(\beta_0^* + \beta_1^* D + \epsilon, Z)}{Cov(D,Z)} = \beta^* + \frac{Cov(\epsilon,Z)}{Cov(D,Z)}$ where D denotes the treatment, Y the outcome, Z the instrument, β^* the unbiased estimator, and ϵ the error term (see Angrist and Pischke 2008).

[19]Note that we can consider *weekBorn* as a measure for relative age disadvantage because higher values of *weekBorn* indicate that an individual was born later in the year which implies relative age disadvantages.

[20]Owing to small sample size in individual birth months, we use a combination of two birth months as a baseline to facilitate statistical power.

[21]Let $logHMV_i$ denote the logarithmized highest market value of player i from club c and with year of birth y. γ_c are U19 BYA club fixed effects and δ_y are year of birth fixed effects. X_i represents control variables such as player's position and height, while u_i is the error term.

References

Allen, J., & Barnsley, R. (1993). Streams and tiers: The interaction of ability, maturity, and training in systems with age-dependent recursive selection. *The Journal of Human Resources*, 28(3), 649-659.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Angrist, J. D., & Krueger, A. B. (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87(418), 328-336.

Ashworth, J., & Heyndels, B. (2007). Selection Bias and Peer Effects in Team Sports. *Journal of Sports Economics*, 8(4), 355-377.

Augste, C., & Lames, M. (2011). The relative age effect and success in German elite U-17 soccer teams. *Journal of Sports Sciences*, 29(9), 983-987.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., & Heil, A. (2022). Why Do People Stay Poor? *Quarterly Journal of Economics*, 137(2), 785-844.

Barnsley, R. H., Thompson, A. H., & Barnsley, P. E. (1985). Hockey success and birthdate: The relative age effect. *Canadian Association for Health, Physical Education, and Recreation*, 51(1), 23-28.

Berger, J., & Nieken, P. (2016). Heterogeneous Contestants and the Intensity of Tournaments. *Journal of Sports Economics*, 17(7), 631-660.

Bryson, A., & Chevalier, A. (2015). Is there a taste for racial discrimination amongst employers? *Labour Economics*, 34, 51-63.

Cobley, S. P., Schorer, J., & Baker, J. (2008). Relative age effects in professional German soccer: A historical analysis. *Journal of Sports Sciences*, 26(14), 1531-1538.

Cohen-Zada, D., Krumer, A., & Shapir, O. M. (2018). Testing the effect of serve order in tennis tiebreak. *Journal of Economic Behavior & Organization*, 146, 106-115.

Cullen, J. B., Jacob, B. A., & Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, 74(5), 1191-1230.

Cumming, S. P., Brown, D. J., Mitchell, S., Bunce, J., Hunt, D., Hedges, C. & Malina, R. M. (2018). Premier League academy soccer players' experiences of competing in a tournament bio-banded for biological maturation. *Journal of Sports Sciences*, 36(7), 757-765.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31-47.

Dawid, H., & Muehlheusser, G. (2015). Repeated selection with heterogeneous individuals and relative age effects. *Journal of Economic Behavior & Organization*, 116, 387-406.

Del Campo, D. G. D., Vicedo, J. C. P., Villora, S. G., & Jordan, O. R. C. (2010). The relative age effect in youth soccer players from Spain. *Journal of sports science & medicine*, 9(2), 190.

Deprez, D. N., Fransen, J., Lenoir, M., Philippaerts, R. M., & Vaeyens, R. (2015). A retrospective study on anthropometrical, physical fitness, and motor coordination characteristics that influence dropout, contract status, and first-team playing time in high-level soccer players aged eight to eighteen years. *The Journal of Strength & Conditioning Research*, 29(6), 1692-1704.

Deutsche Fußball Liga. (2018). Die Entwicklung der Leistungszentren. Retrieved from `https://www.dfl.de/de/hintergrund/nachwuchsleistungszentren/wende-am-tiefpunkt-die-nachwuchsleistungszentren/`

Deutsche Fußball Liga. (2019). Wirtschaftsreport 2019. Retrieved from `https://media.dfl.de/sites/2/2019/02/DFL_Wirtschaftsreport_2019_DE_M-1.pdf`

Deutsche Fußball Liga. (2020a). Anhang V: Richtlinien für die Unterrichtung und Unterhaltung von Leistungszentren der Teilnehmer der Lizenzligen. Retrieved from `https://media.dfl.de/sites/2/2019/12/Anhang-V-zur-LO-2019-12-04-Stand.pdf`

Deutsche Fußball Liga. (2020b). Wirtschaftsreport 2020. Retrieved from `https://media.dfl.de/sites/2/2020/02/DE_DFL_Wirtschaftsreport_2020_M.pdf`

Deutscher Fußball Bund. (2020). Mitglieder-Statistik 2020. Retrieved from `https://www.dfb.de/fileadmin/_dfbdam/223584-Mitgliederstatistik.pdf`

Dustmann, C., Puhani P. A., & Schönberg, U. (2016). The Long-Term Effects of Early Track Choice.

Economic Journal, 127(603), 1348-1380.
Feess, E., Gerfin, M., & Muehlheusser, G. (2015). Contracts as rent-seeking devices: Evidence from German soccer. *Economic Inquiry*, 53(1), 714–730.
Flepp, R., & Franck, E. (2021). The Performance Effects Of Wise And Unwise Managerial Dismissals. *Economic Inquiry* 59(1), 186-198.
Franzke, J. (2017). Nachwuchsförderung: 1,23 Milliarden Euro für die NLZs, *Kicker*. Retrieved from `https://www.kicker.de/nachwuchsfoerderung_123-milliarden-euro-fuer-die-nlzs-711418/artikel`
Franck, E., & Nüesch, S. (2012). Talent and/or popularity: what does it take to be a superstar?. *Economic Inquiry*, 50(1), 202-216.
Friebel, G., Heinz, M., Hoffman, M., & Zubanov, N. (2019). What Do Employee Referral Programs Do? Measuring the Direct and Overall Effects of a Management Practice. *NBER Working Paper, (w25920)*.
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.
fupa.net (2020). Retrieved from `https://www.fupa.net/`
fussballfieber.de (2017). Top 7 - Die Besten Jugendakademien Deutschlands: Sie legen besonders viel Wert auf Jugendarbeit. *Fussballfieber*. Retrieved from `https://www.fussballfieber.de/p/2ha9td/top-7-die-besten-jugendakademien-deutschlands`
González-Díaz, J., Gossner, O., & Rogers, B. W. (2012). Performing best when it matters most: Evidence from professional tennis. *Journal of Economic Behavior & Organization*, 84(3), 767-781.
Grossmann, B., & Lames, M. (2013). Relative Age Effect (RAE) in football talents–the role of youth academies in transition to professional status in Germany. *International Journal of Performance Analysis in Sport*, 13(1), 120-134.
Hanushek, E. A., & Rivkin, S. G. (2009). Harming the best: How schools affect the black-white achievement gap. *Journal of policy analysis and management*, 28(3), 366-393.
Harb-Wu, K., & Krumer, A. (2019). Choking under pressure in front of a supportive audience: Evidence from professional biathlon. *Journal of Economic Behavior & Organization*, 166, 246-262.
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.
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.
Hill, N., & Remer, M. (2020). Race and Employment Outcomes: Evidence from NBA Coaches. *Economic Inquiry* 58(3), 1469-1486.
Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The Allocation of Talent and U.S. Economic Growth. *Econometrica* 87(5), 1439-1474.
Jackson, R. C., & Comber, G. (2020). Hill on a mountaintop: A longitudinal and cross-sectional analysis of the relative age effect in competitive youth football. *Journal of Sports Sciences*, 1-7.
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.
Kempa, K. (2022). Task-specific human capital and returns to specialization: evidence from association football. *Oxford Economic Papers*, 74(1), 136-154.
Keppel, P. & Claessens, T. (2020). How the volunteers of data website Transfermarkt became influential players at European top football clubs. *Follow The Money*. Retrieved from `https://www.ftm.nl/artikelen/transfermarkt-volunteers-european-football`
Kitchens, C.T. (2015). Are Winners Promoted Too Often? Evidence From The NFL Draft 1999–2012. *Economic Inquiry* 53(2), 1317-1330.
Lagestad, P., Steen, I., & Dalen, T. (2018). Inevitable relative age effects in different stages of the selection process among male and female youth soccer players. *Sports*, 6(2), 29.
Lovell, R., Towlson, C., Parkin, G., Portas, M., Vaeyens, R., & Cobley, S. (2015). Soccer player characteristics in English lower-league development programmes: The relationships between relative age, maturation, anthropometry and physical fitness. *PloS one*, 10(9), 1-14.
Malina, R. M., Chamorro, M., Serratosa, L., & Morate, F. (2007). TW3 and Fels skeletal ages in elite youth

soccer players. *Annals of Human Biology*, 34(2), 265-272.

Malina, R. M., Eisenmann, J. C., Cumming, S. P., Ribeiro, B., & Aroso, J. (2004). Maturity-associated variation in the growth and functional capacities of youth football (soccer) players 13–15 years. *European Journal of Applied Physiology*, 91(5-6), 555-562.

Malina, R. M., Reyes, M. P., Eisenmann, J. C., Horta, L., Rodrigues, J., & Miller, R. (2000). Height, mass and skeletal maturity of elite Portuguese soccer players aged 11–16 years. *Journal of Sports Sciences*, 18(9), 685-693.

Mann, D. L., & van Ginneken, P. J. (2017). Age-ordered shirt numbering reduces the selection bias associated with the relative age effect. *Journal of Sports Sciences*, 35(8), 784-790.

Martindale, R. J., Collins, D., & Daubney, J. (2005 Talent development: A guide for practice and research within sport.). *Quest*, 57(4), 353-375.

Mechtel, M., Bäker, A., Brändle, T., & Vetter, K. (2011). Red Cards: Not Such Bad News for Penalized Guest Teams. *Journal of Sports Economics*, 12(6), 621–646.

Muehlheusser, G., Schneemann, S., Sliwka, D., & Wallmeier, N. (2018). The Contribution of Managers to Organizational Success: Evidence from German Soccer. *Journal of Sports Economics*, 9(6), 786–819.

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.

Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, 263(2), 611-624.

Murphy, R., & Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, 87(6), 2777-2826.

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.

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.

Patel, R., Nevill, A., Cloak, R., Smith, T., & Wyon, M. (2019). Relative age, maturation, anthropometry and physical performance characteristics of players within an Elite Youth Football Academy. *International Journal of Sports Science & Coaching*, 14(6), 714-725.

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.

Peña, P. A. (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.

Pérez-González, B., Fernandez-Luna, A., Castillo, D., & Burillo, P. (2020). Are european soccer players worth more if they are born early in the year? Relative age effect on player market value. *International Journal of Environmental Research and Public Health*, 17(9), 3301.

Pérez-González, B., León-Quismondo, J., Bonal, J., Burillo, P., & Fernández-Luna, Á. (2021). The New Generation of Professional Soccer Talent Is Born under the Bias of the RAE: Relative Age Effect in International Male Youth Soccer Championships. *Children*, 8(12), 1117.

Pierson, K., Addona, V., & Yates, P. (2014). A behavioural dynamic model of the relative age effect. *Journal of Sports Sciences*, 32(8), 776-784.

Poli, R., Revenel, L. & Besson, R. (2019). Financial analysis of the transfer market in the big-5 European leagues (2010-2019). *CIES Football Observatory Monthly Report*, 47, 1-8.

ran.de (2015). Top 10: Die besten Jugendakademien Deutschlands. *ran*. Retrieved from `https://www.ran.de/fussball/bundesliga/news/top-10-die-besten-jugendakademien-deutschlands-153443`

Reilly, T., Williams, A. M., Nevill, A., & Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences*, 18(9), 695-702.

Roach, M.A. (2022). Career concerns and personnel investment in the Major League Baseball player draft. *Economic Inquiry* 60(1), 413-426.

Roberts, S. J., McRobert, A. P., Rudd, J., Enright, K., & Reeves, M. J. (2020). Research in Another un-Examined (RAE) context. A Chronology of 35 Years of Relative Age Effect Research in Soccer: Is it time to move on?. *Science and Medicine in Football*.

Rommers, N., Mostaert, M., Goossens, L., Vaeyens, R., Witvrouw, E., Lenoir, M., & D'Hondt, E. (2019). Age and maturity related differences in motor coordination among male elite youth soccer players.

Journal of Sports Sciences, 37(2), 196-203.

Rösch, D., Hodgson, R., Peterson, L., Graf-Baumann, T., Junge, A., Chomiak, J., & Dvorak, J. (2000). Assessment and evaluation of football performance. *The American Journal of Sports Medicine*, 28(5), 29-39.

Sæther, S. A. (2016). Presence of the relative age effect and its effect on playing time among under-20 players in the Norwegian premier league Tippeligaen–a four-year follow up. *Montenegrin Journal of sports science and medicine*, 5(1), 11.

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*, 5(1), 5.

Sponsors. (2019). Nachwuchsarbeit in der Bundesliga: Ein Geschäftsmodell? *Sponsors online.* Retrieved from https://www.sponsors.de/news/themen/nachwuchsarbeit-der-bundesliga-ein-geschaeftsmodell

Statista (2020). Anzahl der Unique User von transfermarkt.de von Mai 2018 bis November 2020. *Statista.* Retrieved from https://de.statista.com/statistik/daten/studie/945403/umfrage/besucherzahlen-von-transfermarktde-direkt-als-zeitreihe/

Sundermeyer, O. (2009). Manipulierter Marktwert. *ZEIT online.* Retrieved from https://www.zeit.de/sport/fussball/2009-12/transfermarkt-spieler-marktwert

transfermarkt.de (2020). Retrieved from https://www.transfermarkt.de/

Tribolet, R., Watsford, M. L., Coutts, A. J., Smith, C., & Fransen, J. (2019). From entry to elite: The relative age effect in the Australian football talent pathway. *Journal of Science and Medicine in Sport*, 22(6), 741-745.

Vaeyens, R., Philippaerts, R. M., & Malina, R. M. (2005). The relative age effect in soccer: A match-related perspective. *Journal of sports sciences*, 23(7), 747-756.

Votteler, A., & Höner, O. (2014). 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.

Walker, O. (2016). Peak height velocity: The period of time a child experiences the largest increase in stature. Retrieved from https://www.scienceforsport.com/peak-height-velocity/#top

Appendix

Appendix A: Possible Adaptions of the Theoretical Model

For defining the exact shape of the performance development function, the function needs to be scaled by a development speed parameter θ, and a starting point t_0 of the function needs to be specified. Moreover, the effect of superior training in elite academies can potentially be heterogenous. For instance, it could be assumed that the effect of elite soccer academies is a function of maturity differences α, which are independent from relative age. Based on that, an expanded model for investigating the RAE could be defined as follows:

$$P_i(t - t_0) \quad = \quad \frac{\triangle_i(\alpha_i) \times P^{\star}_{i0}}{1 + (\triangle_i(\alpha_i) \times P^{\star}_{i0} - 1) \times exp\left(-\theta\left(t - t_0 - s + \frac{m_i}{12}\right)\right)} \tag{13}$$

Yet, in order to estimate the performance development curve during adolescence, a proxy for performance other than market values needs to be found, because market values are non-existent or highly regulated in youth soccer. Maybe an index incorporating different performance components can be calculated based on data which elite youth academies gather. In summary, using the logistic function allows to illustrate different mechanisms regarding the RAE in sports and it can even be further adapted if needed.

Appendix B: Detailed Description of the Data

As a baseline, the aggregated standings of the U19 Bundesliga since 2001 were examined. This information was further supplemented with rankings of the most successful BYA from two different websites (ran.de, 2015; fussballfieber.de, 2017) and a short list of the 36 most successful BYAs was compiled. Yet, going from the top to the bottom of the list, at a certain point, complete U19 squad lists by club were no longer available for the entire period between 2002 and 2020. This can be explained by how the database at transfermarkt.de is extended and maintained. The data entry of complete U19 team squads from the past and the linking of the players to their player profiles depends on individual football experts, fans and club employees. Especially for the early years of the BYA, either the complete squad lists are available for the respective clubs or very incomplete ones, which consist entirely of later professional players. In this way, it quickly becomes clear which clubs provide a suitable database for our analysis. Examples for clubs with incomplete data especially in the early 2000s are FC Augsburg, 1.FC Nürnberg, and VfL Bochum.

Another reason for incomplete squad lists and exclusion of clubs from our sample is if a club is relatively new on the professional soccer stage. RB Leipzig, for instance, has a very competitive BYA today but data is missing for years before 2008 when the club still had a different name, no wealthy sponsor and played in the fifth league. Our sample selection is thus highly driven by a tradeoff between data availability and size. Deciding against a larger sample, we only included clubs for which complete squad lists were available for the whole period between 2002 and 2020. Among the 17 clubs selected, all U19 teams played almost the entire time (at tleast 80% of the years) in the youth Bundesliga, the highest league. Moreover, the clubs either belong to the top 20 clubs in the aggregated standings of the U19 Bundesliga since 2001 or are regularly rated among the top 10 BYA that bring about most professional players (ran.de, 2015; fussballfieber.de, 2017).

Therefore, only 13 BYA of today's Bundesliga clubs (FC Bayern Munich, Borussia Dortmund, Schalke 04, VFL Wolfsburg, Bayer 04 Leverkusen, Werder Bremen, Hertha BSC Berlin, 1.FC Köln, VFB Stuttgart, TSG 1899 Hoffenheim, Borussia Mönchengladbach, SC Freiburg, Mainz 05, and Eintracht Frankfurt), two BYA of today's 2.Liga clubs (Hamburger SV and Hannover 96) and one

BYA of today's 3.Liga clubs (1860 München) were selected. Every additional club would have implied increasingly incomplete data.

Then, the players' data from transfermarkt.de was acquired in two steps. Using a *crawler* written in Python, first, all U17 and U19 Bundesliga cadres of the 17 most successful youth teams between 2001 and 2020 were downloaded including player names and player-IDs. Second, using the player-IDs, the crawler downloaded information on every individual player from their respective profiles.

We restrict our data to players with German nationality, as other players might have undergone elite youth academies of different qualities in their home countries before being selected. Missing birthdates were added by hand for about 700 players on basis of the website fupa.net, which allows amateur clubs to populate the database with information on their players. Additionally, players who were mentioned in BYA cadres but without concordant reference to this in their transfer history were dropped. This was necessary because we need to calculate the number of days that youth players spent in BYAs based on players' transfer histories. Finally, the dataset contains 3,835 observations. Among those, 2,383 played for a U19 BYA team and were born between 1988 and 2001, i.e. could potentially have gotten five full years (U15-U19) of BYA performance-oriented training.

We code a variable indicating the days a player spent in one of the 17 BYAs chosen. We only considered the period of performance-oriented training between the U15 and U19, because during this time competition, investment and BYA training quality are highest. Out of interpretative ease, we convert this variable into years spent in BYAs ($BYAyears$), going continuously from zero up to a maximum of five years. It needs to be acknowledged that we use only the 17 most successful BYA to define $BYAyears$. This means that our data could possibly include cases in which players were first trained at one of the remaining 37 BYAs and joined one of the 17 selected clubs later. However, our specification is justified by two main arguments: First, close examination of the data showed that transfers from other BYAs (out of the sample) to the 17 most successful BYAs (in the sample) are rare. Second, not all BYAs provide the same quality of training. More than 70% of total BYA investment is made by the 18 Bundesliga clubs (Sponsors, 2019). Investment in BYAs is, thus, likely to be skewed towards the most successful ones. Hence, $BYAyears$ is an appropriate measure for the years that adolescents received distinguished soccer training.

Appendix C: Additional Figures and Tables

Figure 6: Inflation of Bundesliga Market Values

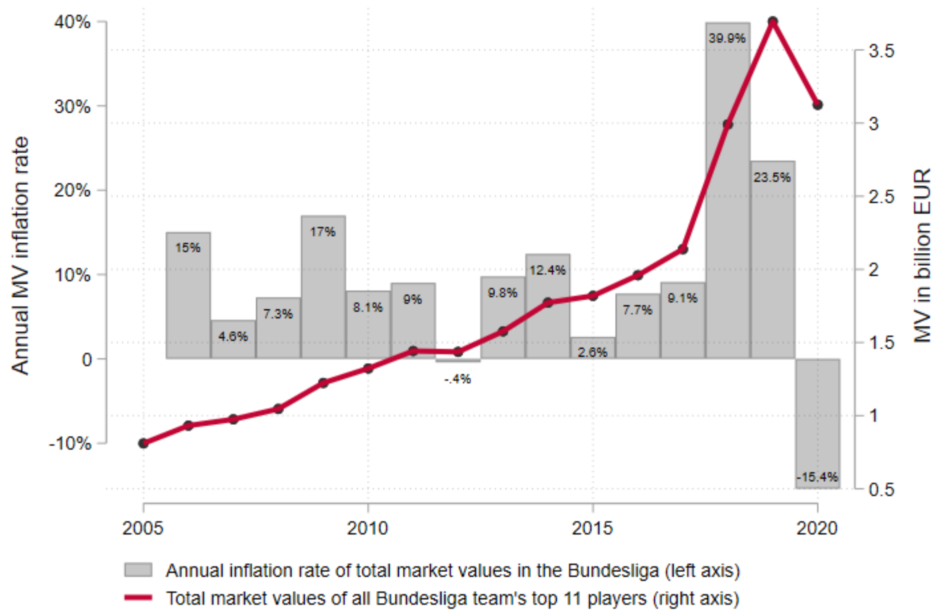


Figure 7: Number of Children Born in Germany in 1990 and 2000 across Birth Months

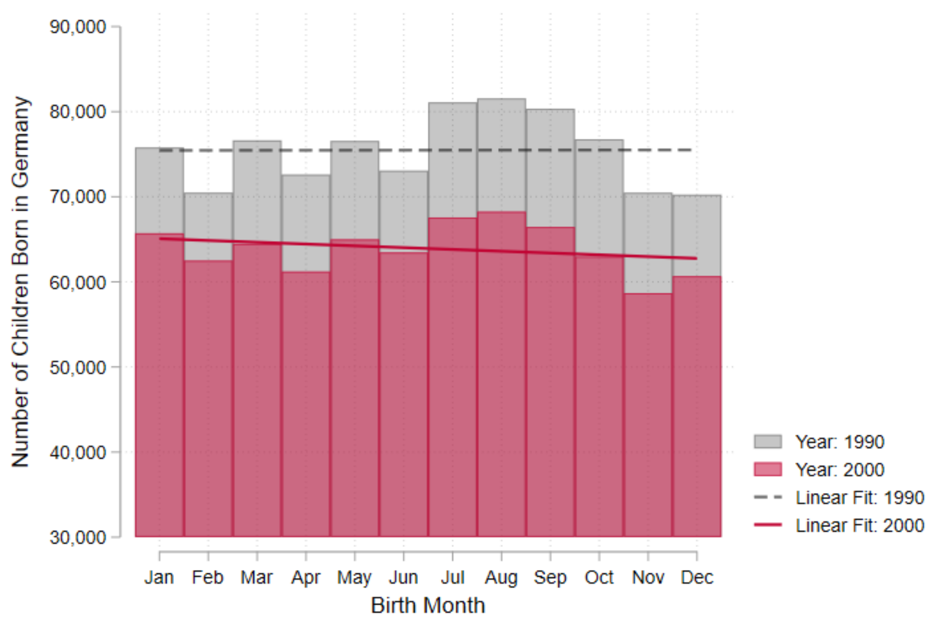


Figure 8: The RAE over Time: Proportion of 1st and 2nd Bundesliga Professional Players Born in First Half of the Year by Season

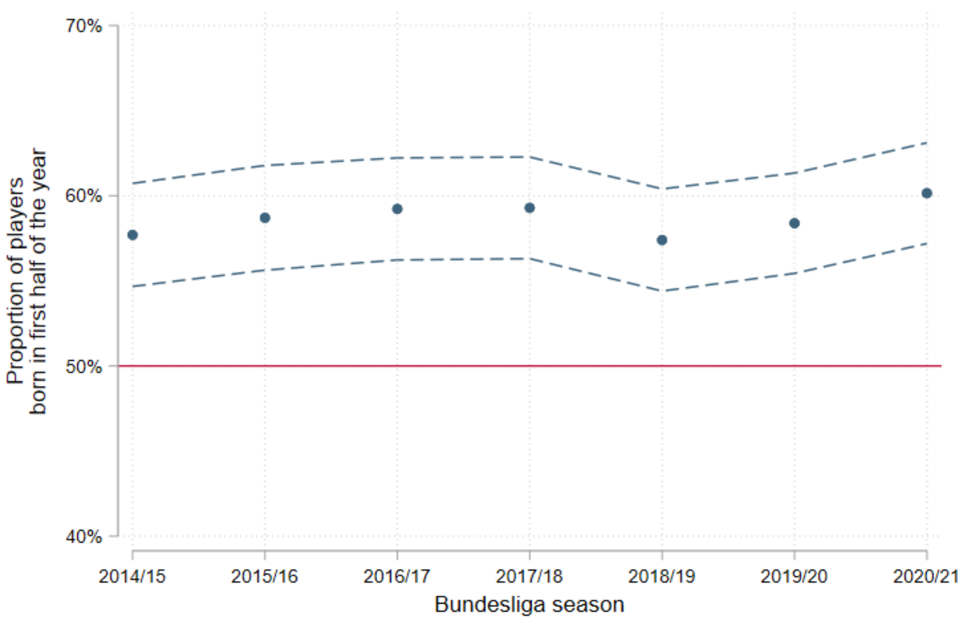


Figure displays values of all 1st and 2nd Bundesliga adult players for the seasons 2014/15 to 2019/2020 at the beginning of the respective season. Confidence intervals at 95% and equal distribution as reference. FIFA data from https://sofifa.com.

Table 6: The Relative Age Effect: Summary Statistics by U17 BYA

	% born Jan-Jun	% born Jan-Mar	Mean HMV in 1,000€	Obs.
Full sample	**74.5**	**47.6**	**1333.413**	**2233**
SC Freiburg U17	79.4	49.4	603.095	170
VfB Stuttgart U17	79.1	49.3	3063.749	134
Bayer 04 Leverkusen U17	78.5	54.8	1728.756	93
FC Schalke 04 U17	76.7	49.3	1701.896	150
Werder Bremen U17	76.1	52.8	630.192	163
Hamburger SV U17	75.4	48.5	1074.453	134
FC Bayern München U17	75.2	45.5	3395.633	101
Eintracht Frankfurt U17	75.4	44.8	874.055	134
Borussia Dortmund U17	74.4	51.3	2515.123	117
TSG 1899 Hoffenheim U17	73.8	40.5	1024.075	126
VfL Wolfsburg U17	73.8	47.6	1018.544	126
Hertha BSC U17	73.1	49.2	1206.956	130
Borussia Mönchengladbach U17	71.7	47.8	834.626	159
1.FSV Mainz 05 U17	71.8	49.1	1241.834	110
1.FC Köln U17	71.5	43.0	667.252	158
Hannover 96 U17	70.1	43.9	579.183	107
TSV 1860 München U17	69.4	41.3	1622.765	121

Data on the 17 most successful BYA U17 clubs from transfermarkt.de. Players born between 1988 and 2001. Differences in the number of observations per club can be attributed to missing data and different proportions of foreign youth players, who are not considered here.

Table 7: Effect of BYA Training, Positions, Height and Strong Foot on Market Values

	(1) logHMV	(2) logHMV	(3) logHMV	(4) logHMV	(5) logHMV
BYAyears	0.583***	0.510***	0.511***	0.517***	0.504***
	(0.0374)	(0.0385)	(0.0385)	(0.0383)	(0.0378)
Center back	0.106	0.359**	0.378**	0.277	0.341**
	(0.167)	(0.169)	(0.171)	(0.171)	(0.167)
Right back	0.00800	0.749***	0.801***	0.633**	0.741***
	(0.199)	(0.241)	(0.243)	(0.242)	(0.239)
Left back	0.168	0.921***	0.972***	0.712**	0.921***
	(0.214)	(0.248)	(0.250)	(0.271)	(0.247)
Central defensive	-0.0491	0.475**	0.534**	0.362*	0.490**
	(0.179)	(0.201)	(0.207)	(0.199)	(0.200)
Central midfield	0.418*	1.166***	1.209***	1.202***	1.125***
	(0.237)	(0.258)	(0.259)	(0.258)	(0.254)
Central offensive	-0.0127	0.917***	0.941***	0.821***	0.836***
	(0.234)	(0.280)	(0.281)	(0.284)	(0.275)
Right midfield	-0.647**	0.539	0.580*	0.416	0.528
	(0.306)	(0.340)	(0.341)	(0.340)	(0.340)
Left midfield	-0.754***	0.0921	0.138	-0.0520	0.0997
	(0.277)	(0.317)	(0.318)	(0.323)	(0.316)
Center forward	0.215	0.545***	0.598***	0.421**	0.519**
	(0.197)	(0.206)	(0.271)	(0.268)	(0.261)
Left wing	0.261	1.168***	1.212***	1.105***	1.053***
	(0.241)	(0.270)	(0.271)	(0.268)	(0.261)
Right wing	1.094***	2.021***	2.053***	1.907***	2.015***
	(0.262)	(0.288)	(0.288)	(0.287)	(0.287)
Height in cm		0.0852***	-0.450	0.0822***	0.0822***
		(0.0109)	(0.372)	(0.0112)	(0.0108)
Height squared			0.00147		
			(0.00102)		
Foot: Left				-0.0206	
				(0.149)	
Foot: Both				-0.586***	
				(0.121)	
Birth Year FE	Yes	Yes	Yes	Yes	Yes
Observations	2281	1914	1914	1826	1904
R^2	0.314	0.333	0.334	0.334	0.330
Data	Sample	Sample	Sample	Sample	HMV < 50m

OLS regression results. The sample includes all former U19 BYA players who were born between 1988 and 2001. Goalkeeper is the omitted reference position. Because the logarithm of the market values is the dependent variable, the coefficients need to be converted as following: $100 \times (exp(-\hat{\beta}) - 1)$. Based on that, coefficients of BYAyears can be interpreted as changes of the follwoing size: 79.1% (column 1), 66.5% (column 2), 66.7% (column 3), 67.7% (column 4), and 65.5% (column 5). The coefficient of height translates into 8.9, 8.6 and 8.6 percent in columns 2, 4, and 5 respectively. Regarding specific positions, a selection of coefficients from column 2 and their respective percentage changes are shown in the format $\beta = x\%$: 2.021 = 654.6%, 1.168 = 221.6%, 0.921 = 151.2%, 0.749 = 111.5%, 0.545 = 72.5%, and 0.359 = 43.2%. Column 4 shows that two-footedness is associated with 44.3% lower HMV. Heteroskedasticity-robust Huber-White standard errors in parentheses.

* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 8: Quantile Regressions of Market Values, BYA Training, and Birth Quarters analogue to Columns 1 and 4 of Table 3

Quantile	q50	q75	q90	q50	q75	q90
	(1)	(2)	(3)	(4)	(5)	(6)
	logHMV	logHMV	logHMV	logHMV	logHMV	logHMV
Q2: Apr-Jun	0.0	0.0	0.214			
	(0.0478)	(0.0662)	(0.178)			
Q3: Jul-Sep	0.0364	0.176	0.567***			
	(0.0506)	(0.111)	(0.210)			
Q4: Oct-Dec	0.0700	0.281	0.456***			
	(0.0599)	(0.179)	(0.172)			
BYAyears				0.266***	0.370***	0.553***
				(0.0211)	(0.0272)	(0.0503)
Birth Year FE	Yes	Yes	Yes	Yes	Yes	Yes
U19 Club FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2383	2383	2383	2383	2383	2383

Consistent with our theoretical model, the quantile regressions are biased due to player selection (see section 3.3). Among players with relative age advantages, more *untalented* players get selected by youth academies. This pushes a very talented player born in the first quarter of the year up the conditional ability distribution. The quantile regressions therefore compare players who have different underlying ability (and market value) distributions. The player at the 75th percentile of the last-quarter-of-the-year-ability-distribution is arguably more talented than the 75th percentile player from the first-quarter-of-the-year-ability-distribution. Hence, it is consistent with our theoretical framework that the positive relation between birth quarter and market values is more pronounced at higher quantiles (see Columns 1, 2 and 3). This argument can be applied analogously to the relationship between years of youth elite academy training and market values (see Columns 4, 5, and 6). The sample includes all former U19 BYA players who were born between 1988 and 2001. In all columns players born in the first quarter of the year (Q1: Jan-Mar) are omitted and constitute the baseline. A quantile regression for the 25th percentile is omitted as the 25th percentile of market values is 0, conditional on all birth quarters. Because the logarithm of the market values is the dependent variable, the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Heteroskedasticity-robust standard errors in parentheses. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 9: Differences in Means with Additional Controls: The Cost of the Relative Age Effect in Bundesliga Youth Academies

	(1) logHMV	(2) logHMV	(3) logHMV	(4) logHMV	(5) logHMV
Reference Months	Jan-Feb	Mar-Apr	May-Jun	Jun-Jul	Jul-Aug
May-Jun	-0.102	-0.122			
	(0.172)	(0.186)			
Jul-Aug	0.0507	0.0664	0.168		
	(0.164)	(0.180)	(0.204)		
Sep-Oct	0.399**	0.395*	0.516**	0.354	0.338
	(0.194)	(0.206)	(0.226)	(0.220)	(0.224)
Nov-Dec	0.350	0.351	0.470*	0.412*	0.316
	(0.230)	(0.237)	(0.259)	(0.255)	(0.250)
Height in cm	0.0569***	0.0495***	0.0351**	0.0254	0.0543***
	(0.0114)	(0.0121)	(0.0153)	(0.0171)	(0.0189)
Position Control	Yes	Yes	Yes	Yes	Yes
Strong Foot Control	Yes	Yes	Yes	Yes	Yes
Birth Year FE	Yes	Yes	Yes	Yes	Yes
U19 Club FE	Yes	Yes	Yes	Yes	Yes
Observations	1453	1278	858	574	558
R^2	0.276	0.244	0.254	0.301	0.310

The sample includes all former U19 BYA players who were born between 1988 and 2001. Birth month comparison groups are given in the header. Because the logarithm of the market values is the dependent variable, the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, a selection of coefficients and their respective percentage changes are shown in the format $\beta = x\%$: 0.25 = 28.4%, 0.30 = 35.0%, 0.338 = 40.2%, 0.4 = 49.2%, 0.412 = 51.0%, 0.470 = 60.0%, 0.5 = 64.9%, and 0.516 = 67.5%.
Heteroskedasticity-robust Huber-White standard errors in parentheses. * Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.