RELATIVE AGE EFFECT AND CONSTITUENT YEAR EFFECT: AN ANALYSIS OF THE INTERNATIONAL TENNIS FEDERATION RANKING

João Paulo Abreu Moreira¹, Mariana Calábria Lopes¹, Larissa Oliveira Faria¹ e Maicon Rodrigues Albuquerque²

¹Universidade Federal de Viçosa, Viçosa-MG, Brasil.
²Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brasil.

ABSTRACT
Relative Age Effect (RAE) and Constituent Year Effect (CYE) seem to influence the process of training young tennis players. The aim of this study was to investigate the presence of these effects in the ranking of junior tennis players of the International Tennis Federation (ITF) and its influence on the score obtained by such tennis players. The results show the presence of the RAE and the CYE (p ≤ 0.05). The regression model identified that the variables “year” and “month” of birth predict the scores of athletes in the ITF ranking with an explained variance of 19%. In conclusion, the ranking composed of various age groups does not seem to be a classification strategy sensitive enough to prevent the RAE and the CYE among junior tennis players.

Keywords: Relative Age Effect. Constituent Year Effect. Tennis.

Introduction

In tennis, as in other sports, the process involving the selection and training of athletes is quite complex. Achieving excellence and becoming a professional tennis player is a difficult task that few can handle. To be successful in such a selective sport, children are chosen to participate in training programs and competition¹ in order to achieve high performance soon in early adulthood. The development of young athletes requires multivariate procedures², since, in addition to concern with physical and socio-affective characteristics, cognitive and motor-sensory aspects need to be worked on as well³. However, other variables may interfere throughout development stages, including genetic characteristics⁴, the environment where development takes place and social aspects involved⁶ in this process.

Among the several factors that seem to influence the selection and training of athletes, the Relative Age Effect (RAE) has been receiving a lot of attention in recent years and has been described in several collective sports modalities, such as football⁸ and handball⁹, in fighting modalities, such as taekwondo¹⁰ and judo¹¹, and also in individual sports, such as
Data collection

All players’ birth dates and scores in the ITF junior ranking of December 30, 2013 were taken from the entity’s official website.
A total of 2,481 male athletes born in the following years: 1995 (n = 588); 1996 (n = 784); 1997 (n = 653); 1998 (n = 334); 1999 (n = 72) and 2000 (n = 10) were verified.

**Procedures**

To assess the RAE, just as in other studies that were composed of an international sample\textsuperscript{10,27,28}, all tennis players were categorized according to month of birth and denominated: Q1 - January-March; Q2 - April-June; Q3 - July-September; Q4 - October-December.

Finally, to assess the CYE, the athletes were divided by year of birth.

**Statistical analysis**

For data analysis, the Chi-square test ($\chi^2$) was used to verify the RAE (eg. Barnsley, Thompson and Barnsley\textsuperscript{16}, Delorme and Raspaud\textsuperscript{29}, Edgar and O'Donoghue\textsuperscript{13}) and the CYE (eg. Medic, Starkes and Young\textsuperscript{24}).

The present study adopted the expected frequency of birth proportional to the number of days that make up each quartile, in the expectation of further reducing the margin of error in the distribution of births\textsuperscript{11}. Thus, the frequencies expected for each quartile are: Q1 - 24.7% (90 days); Q2 - 24.9% (91 days); Q3 and Q4 - 25.2% (92 days each). For RAE analysis, a uniform distribution of births was adopted for the years that make up the ranking, with the expected frequency of 406.83 athletes per year (16.67%), defined by the ratio between the total number of the sample and the number of years that make up the ITF ranking\textsuperscript{24}.

The odds ratio (OD) analysis, with a 95% confidence interval, was performed to calculate the Effect Size between the frequency of tennis players born in Q1, Q2 and Q3 in relation to Q4. The same procedure was used to analyze the frequency of births between the years that make up the ranking. The reference values for the effect size (small, medium or large) are: 1.22; 1.86; 3.00, respectively\textsuperscript{11,27,28,30}.

A Stepwise regression model was used to determine the specific contribution of the “year of birth” and “month of birth” independent variables to the explanation of the “score in the ranking” dependent variable. The level of significance adopted in all procedures was $p \leq 0.05$. All the statistical treatment was performed using the IBM SPSS Statistics program for Mac, version 20.0.

**Results**

Table 1 shows the total distribution of the birth frequencies of the 2,471 subjects researched and of each of the subgroups formed by years of birth. In relation to the total sample, the $\chi^2$ test identified significant difference between birth quartiles [$\chi^2 (3) = 102.33; p<0.001$]. Significant differences were also found between quartiles in the groups formed by the years of birth: 1995 [$\chi^2(3)=8.61; p=0.024$], 1996 [$\chi^2(3)=38.86; p<0.001$], 1997 [$\chi^2(3)=29.88; p<0.001$], 1998 [$\chi^2(3)=25.55; p<0.001$] and 1999 [$\chi^2(3)=14.44; p<0.001$]. For those born in 2000 it was not possible to perform the $\chi^2$ test, because the sample size was small and the frequency observed in quartile 4 was 0.
Table 1. Frequencies, Chi-square values and p value for the total sample and subgroups by year of birth

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>Total</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Total</th>
<th>(\chi^2)</th>
<th>(p^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>2,441</td>
<td>733</td>
<td>719</td>
<td>540</td>
<td>449</td>
<td>102.33</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>858</td>
<td>210</td>
<td>182</td>
<td>114</td>
<td>122</td>
<td>29.12</td>
<td>(0.017)</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>927</td>
<td>210</td>
<td>184</td>
<td>114</td>
<td>129</td>
<td>29.12</td>
<td>(0.017)</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>1,010</td>
<td>223</td>
<td>191</td>
<td>123</td>
<td>133</td>
<td>36.05</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>1,041</td>
<td>230</td>
<td>200</td>
<td>131</td>
<td>140</td>
<td>39.93</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>2000***</td>
<td>400</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Note: \(p \leq 0.05\); ** \(\Delta\): difference between observed and expected values; ***For the subgroup of athletes born in 2000, 3 cells (100.0%) have expected frequencies lower than 5.

Source: The authors.

Table 2 shows results referring to the odds ratio, used to assess effect sizes between quartiles. A greater probability of athletes born in the initial quartiles was found in relation to the final quartile (Q4), and in most comparisons the effect size found was medium, with the exception of Q3 x Q4 comparisons for the total sample and for those born in 1995, which presented a small effect size, and of Q1 x Q4 and Q2 x Q4 comparisons for tennis players born in 1999, who presented a large effect size.

Table 2. Odds Ratio of Q1, Q2 and Q3 in relation to Q4 for the total sample and subgroups by year of birth

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>Total</th>
<th>Q1 x Q4</th>
<th>Q2 x Q4</th>
<th>Q3 x Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>2,441</td>
<td>1.67** (1.42-1.96)</td>
<td>1.62** (1.37-1.90)</td>
<td>1.20* (1.02-1.42)</td>
</tr>
<tr>
<td>1996</td>
<td>858</td>
<td>1.26** (0.91-1.74)</td>
<td>1.31** (0.95-1.81)</td>
<td>0.99* (0.71-1.39)</td>
</tr>
<tr>
<td>1997</td>
<td>927</td>
<td>1.60** (1.20-2.14)</td>
<td>1.80** (1.35-2.40)</td>
<td>1.23** (0.91-1.65)</td>
</tr>
<tr>
<td>1998</td>
<td>1,010</td>
<td>1.79** (1.31-2.45)</td>
<td>1.52** (1.11-2.10)</td>
<td>1.24** (0.90-1.72)</td>
</tr>
<tr>
<td>1999</td>
<td>1,041</td>
<td>2.18** (1.39-2.41)</td>
<td>1.87** (1.19-2.94)</td>
<td>1.42** (0.89-2.27)</td>
</tr>
<tr>
<td>2000***</td>
<td>400</td>
<td>4.14*** (1.45-11.87)</td>
<td>3.00*** (1.02-8.80)</td>
<td>2.14** (0.71-6.50)</td>
</tr>
</tbody>
</table>

Note: Effect Size (es) *es \(\leq 1.22\) (small), **es \(\leq 1.86\) (medium) and ***es \(\leq 3.00\) (large). ****For the subgroup of athletes born in 2000, 100.0% have expected frequencies lower than 5.

Source: The authors.

Table 3 displays results for the CYE. The \(\chi^2\) test showed significant difference \([\chi^2(5)=1254.98; p<0.001]\) in the distribution of the year of birth of the athletes present in the ranking.

Table 3. Frequencies, Chi-square values and p value for the distribution of years of birth in the ITF junior ranking

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>Total</th>
<th>(\chi^2)</th>
<th>(p^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>588</td>
<td>2.441</td>
<td>1254.98</td>
</tr>
<tr>
<td>1996</td>
<td>784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>653</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>334</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Expected frequency (ef) = 406.83

Source: The authors.

Analyzing Table 4 with the Odds Ratio of frequency of birth in the years that make up the ITF ranking, results show that the effect sizes were small in the comparisons made...

**Table 4.** Odds Ratio between years of birth, based on observed frequencies (of) and expected frequencies (ef)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>0.75*</td>
<td>(0.63-0.89)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>0.90*</td>
<td>(0.75-1.07)</td>
<td>1.20*</td>
<td>(1.01-1.43)</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>1.76**</td>
<td>(1.45-2.13)</td>
<td>2.35**</td>
<td>(1.95-2.83)</td>
<td>1.96**</td>
</tr>
<tr>
<td>1998</td>
<td>8.16***</td>
<td>(6.17-1.81)</td>
<td>10.89***</td>
<td>(8.25-14.37)</td>
<td>9.07***</td>
</tr>
<tr>
<td>1999</td>
<td>58.80***</td>
<td>(31.01-111.51)</td>
<td>78.40***</td>
<td>(41.40-148.50)</td>
<td>65.30***</td>
</tr>
<tr>
<td>2000</td>
<td>7.20***</td>
<td>(4.62-12.37)</td>
<td>6.64***</td>
<td>(3.48-6.20)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Effect Size (es) *es ≤ 1.22 (small), **es ≤ 1.86 (medium) and ***es ≤ 3.00 (large). Expected frequency (ef) = 406.83

Source: The authors.

In relation to the insertion of the “year of birth” and “month of birth” variables as independent variables into the regression model, in which the score in the ranking was the dependent variable, it was verified that the stepwise regression analysis indicated a significant model (F= 24.180, p<0.001, R² = 0.019), showing that 19% of the score obtained by the tennis player is directly influenced by the year and month of his birth. Both variables, “year of birth” (β= -13.931, t = -6.433, p <0.001) and “month of birth” (β= -6.877, t = -3.111, p = 0.002), were significant, and in this case, were inserted into the model.

**Discussion**

The objective of the present study was to investigate the Relative Age Effect (RAE) and the Constituent Year Effect (CYE) on the ITF junior tennis players ranking. In short, the results found showed the presence of RAE (Table 1), revealing a trend of greater representativeness of players born in the first months of the year in relation to those born in the final ones, in addition to the CYE (Table 3), with a statistically greater representation of tennis players born in the years 1995, 1996 and 1997 than in 1998, 1999 and 2000. In addition, it was possible to identify that the year of birth and the month of birth predict the score in the ranking with an explained variance of 19% and that, as expected, the year of birth contributes more to the score in the ranking than the athletes’ month of birth does.

The results confirm the findings of Dudink, Baxter-Jones, Edgar and O’Donoghue, who also investigated the effects of birth dates among tennis players. Although the earliest studies in tennis have demonstrated the RAE more than two decades ago, little seems to have been done to soften it. The RAE seems to continue to influence the training process of young tennis players, along with the CYE produced by the proposed classification system. This fact brings advantages to more mature players and creates obstacles for biologically younger players, and the scores obtained by the athletes and their
positions in the ranking seem to translate not only technical differences between players, but mainly their maturational differences.

Specifically about the RAE, several factors may have influenced its presence among the ranked players. According to the model proposed by Wattie et al., the RAE may be related to numerous variables, but all linked to three interrelated spheres, being: 1) individual characteristics; 2) environmental aspects; and 3) specific characteristics of the modality.

Concerning individual characteristics, the RAE appears to be an aspect to be constantly investigated. According to Cobeley et al., an explanation to the smaller representativeness of athletes born far from the cut-off date would be associated with their late maturation compared to those relatively older, which would result in physical, cognitive and psychological disadvantages, as well as a relatively smaller life experience for the youngest, since the selection stages and throughout the training process. Although differences in growth, biological maturity and cognition decline over the years, in an environment where competition is high, any initial advantage in these aspects can trigger a series of events that may place more mature younger athletes in situations of advantage compared to less mature ones. As an example, older and more developed athletes have greater opportunities to participate in sports with a high level of competitiveness, in which case they can improve their psychological skills (eg. motivation, self-efficacy, etc.), techniques (enhanced motor control due to richer and more frequent practice opportunities), as well as tactical aspects. In this sense, any initial advantage that a certain “older” athlete has over “younger” ones can increase this disadvantage throughout the selection process and, consequently, make the process unequal. For Baxter-Jones, motivation can influence sports training process, interfering with an individual’s desire to continue practicing the modality in question. Thus, face the barriers that stand during training, such as competition for space with other more mature athletes, potential tennis players, relatively younger, may consider the competition level unfair, losing motivation towards the sport and, as a consequence, quitting it.

Studies about the influence of date of birth on sports training generally use the same term, RAE, for differences found in categories with more than one age group. Despite this, Wattie, Cobeley and Baker suggested the use of the term Constituent Year Effect (CYE) as more appropriate for the type of effect observed in groups formed by different age groups. From this perspective, a factor taken as environmental could be related to the biased distribution of births among junior players and their position in the ranking system used by the ITF. Schorer, Wattie and Baker emphasize a diversity of policies used in sports to divide subjects into categories, based on age and cut-off dates. Among junior players, the model adopted by the ITF is the ranking composed of six different age groups. In the specific case of the present study, the ranked tennis players were born in 1995, 1996, 1997, 1998, 1999 and 2000. The results obtained in the analysis of the CYE point to a greater representativeness of tennis players born in the first years that make up the ranking (1995, 1996 and 1997) than of players born in the final years (1998, 1999 and 2000). A possible explanation to this difference could be the poor motivational orientation of younger athletes to compete against older tennis players, resulting in a process of self-selection and abandonment of the sport, based on psychosocial factors, such as low perception of success in relation to their peers. It is important to add that, according to Wattie, Cobeley and Baker, the two phenomena (RAE and CYE) can be observed within the same context/sample, and the results of the present study confirm this trend. Another environmental factor that may be contributing to birth date effects decisively interfering with the composition of the ranking is the popularity of tennis and its high competitive level. The main hypothesis used to explain the effect of the period of birth on the sport seems to be related to the competitiveness of certain modalities. For
instance, if in a big tennis club there are 15 vacancies and 15 athletes are interested in filling them, the level of competitiveness will be low. However, if this same club offers 15 vacancies and 15,000 athletes want them, the level of competitiveness will be high\textsuperscript{18}. Thus, the higher the competitiveness, the higher the influence of the birth period on the selection process of high-performance athletes\textsuperscript{18}, that is, competitiveness may make the training process of athletes increasingly susceptible to the influence of the subject’s month of birth. In this sense, Goldschmied\textsuperscript{18} believes that athletes that play competitive sports born in the final months of the year and who manage to reach a high level of performance in adulthood (eg. being among the 10 best tennis players in the world ranking) should be considered resilient due to the great disadvantage they had in relation to the other ones (born in the beginning of the year) throughout the training process.

On specific characteristics of the modality, it can be inferred that the presence of the RAE and the CYE relate to the high physical demands of tennis, which are important for good performance on court (eg. strength, power and speed). Having such abilities at more advanced stages due to maturation can contribute to older athletes being in advantage over younger ones during childhood and adolescence\textsuperscript{22,25}. That is, these relatively older athletes tend to be taller and have a higher body mass\textsuperscript{22}, which might lead them to perform better and, consequently, have greater chances of success in the sport. Due to this momentary advantage, young athletes end up being chosen and subjected to training, experiencing situations that will contribute to the enhancement of their performance\textsuperscript{9}, achieving high scores and the first places in the ranking of players.

The results of the regression analysis indicated that both the year and the month of birth influence the points accumulated by the player throughout the season. Among the many variables that could influence an aspiring tennis player’s score, approximately one-fifth is related to these two factors (explained variance of 19%). In this specific case, although both influence the obtained score, the CYE has greater impact than the RAE. This reality suggests that the organization of the ranking, in the way it is done, ends up benefiting more mature players, serving as a purely quantitative parameter of the performance of each of them.

It is important to note that the present study presents some limitations, such as the selected sample, which was obtained from different countries. Therefore, it is not possible to carry out a more in-depth analysis of the particularities in the organization of the sport (selection process, number of athletes, categories, form of level/group ascension, among others) in each country. There is also the fact that births are not uniformly distributed throughout the year and are affected by environmental zones and cultural factors\textsuperscript{39}. Thus, other expected distributions can not be used\textsuperscript{40}.

**Conclusions**

Although previous studies on tennis have demonstrated the RAE\textsuperscript{12,13}, the present study seems to confirm that this effect continues to influence the training process of young tennis players, along with the CYE. The presence of the RAE and the CYE in the ITF junior athletes ranking can provide evidence of flaws during the selection and training processes of young tennis players, which could contribute decisively to the loss of potential talents of the modality for withdrawal or abandonment\textsuperscript{40}. The ranking made up by several age groups has not been proving a classification strategy sensitive enough to prevent the RAE and the CYE among the junior players, especially in relation to the (dis)advantages during the selection and development of the athletes\textsuperscript{18}, and which will interfere considerably with their careers in the modality.

Because the main objective of RAE and CYE studies is to promote greater equality during the training process of athletes in order to reduce or eliminate these effects, several
researchers have presented proposals in this sense. For instance, Musch and Grondin\textsuperscript{18} suggested the cut-off date rotation system or the adoption of a classification system based on biological age, similar to weight categories. However, there is already evidence that this system is flawed\textsuperscript{27}. Del Campo et al.\textsuperscript{31} proposed the grouping of categories by quartiles; however, in a system driven by competition, this solution has not been accepted by the professionals involved. Finally, there is still a lot to investigate on this theme, but everything indicates that the solution to the problem will revolve around a proposal that will eliminate formal competition during the training process of young athletes\textsuperscript{42} and/or drop from 12 to 6 months the age groups of competition\textsuperscript{18}.

References