Journal of Sports Sciences

ISSN: 0264-0414 (Print) 1466-447X (Online) Journal homepage: https://www.tandfonline.com/loi/rjsp20

Relative age and birthplace effect in Japanese professional sports: a quantitative evaluation using a Bayesian hierarchical Poisson model

Hideaki Ishigami

Journal of Sports Sciences, 2016
Vol. 34, No. 2, 143–154, http://dx.doi.org/10.1080/02640414.2015.1039462

Routledge
Taylor & Francis Group

Relative age and birthplace effect in Japanese professional sports: a quantitative evaluation using a Bayesian hierarchical Poisson model

HIDEAKI ISHIGAMI

Faculty of Sport Science, Nippon Sport Science University, Tokyo, Japan

(*Accepted 7 April 2015*)

Abstract
Relative age effect (RAE) in sports has been well documented. Recent studies investigate the effect of birthplace in addition to the RAE. The first objective of this study was to show the magnitude of the RAE in two major professional sports in Japan, baseball and soccer. Second, we examined the birthplace effect and compared its magnitude with that of the RAE. The effect sizes were estimated using a Bayesian hierarchical Poisson model with the number of players as dependent variable. The RAEs were 9.0% and 7.7% per month for soccer and baseball, respectively. These estimates imply that children born in the first month of a school year have about three times greater chance of becoming a professional player than those born in the last month of the year. Over half of the difference in likelihoods of becoming a professional player between birthplaces was accounted for by weather conditions, with the likelihood decreasing by 1% per snow day. An effect of population size was not detected in the data. By investigating different samples, we demonstrated that using quarterly data leads to underestimation and that the age range of sampled athletes should be set carefully.

Keywords: *relative age effect, birthplace effect, Poisson regression, Japan*

Introduction

Individuals with early birthdates in a given school year or competitive season have an advantage in their performance; this is known as the relative age effect (RAE). A growing body of literature confirms the existence of the effect in sports and other domains, particularly in school or academic performance (Bedard & Dhuey, 2006; Dhuey & Lipscomb, 2008; Kawaguchi, 2011). While prior studies have largely focused on the prevalence of RAEs in a number of different sports in various countries, they have not demonstrated how large the effects are in an intuitively understandable form. One of the objectives of this paper is to report the effect sizes in such a form using data obtained from Japanese athletes. The participants of the study were athletes participating in two major professional sports in Japan, specifically soccer and baseball. These represent two of the most popular sports in Japan, and consequently, they are highly competitive. To date, although there are only a few studies using Japanese data (Hirose, 2009; Nakata & Sakamoto, 2011), most previous studies have examined the effects in North American and European athletes (Cobley, Baker, Wattie, & McKenna, 2009). We expect that this paper will add to the evidence of RAEs at the professional level in Asian countries.

Moreover, some recent studies provide evidence on the birthplace effect in addition to the RAE (Baker & Logan, 2007; Baker, Schorer, Cobley, Schimmer, & Wattie, 2009; Bruner, MacDonald, Pickett, & Côté, 2011; Côté, MacDonald, Baker, & Abernethy, 2006; MacDonald, Cheung, Côté, & Abernethy, 2009; Turnnidge, Hancock, & Côté, 2014) . These studies report that the place where an athlete was born and developed is another important factor in becoming an elite athlete. According to their findings, the population size of the birthplace affects the likelihood of becoming a top athlete. We therefore examined whether prior findings on the effect of population size hold true for Japanese professional sports. In addition, weather conditions are quite different between Japanese regions. For example, northern areas may receive, on average, over 100 snow days per year, whereas in the southern areas there are few or no snow days. Except for winter sports, lower temperatures and large numbers of snow days seem to be unfavourable for sport activities because, for instance, participants can play neither baseball nor soccer in snow. We thus

Correspondence: Hideaki Ishigami, Faculty of Sport Science, Nippon Sport Science University, 7-1-1, Fukasawa, Setagaya-ku, Tokyo, 158-8508, Japan.
E-mail: ishigami@nittai.ac.jp

investigated whether weather conditions affect the likelihood of becoming a professional baseball or soccer player. Hence, our estimation strategy is to decompose the variation in the likelihood between birthplaces into the effects of population size, weather conditions and the residual difference between birthplaces. This study, to the best of our knowledge, is the first attempt to explore the effect of weather conditions in an analysis of relative age and birthplace effects. We used regression analysis to infer the size of these effects, and thus, we report various estimates that one could not obtain by just comparing the number of the players in each category (e.g. born in the first quarter, born in a large city). Finally, since our study focuses on the quantitative aspects of these effects, we describe the method used in detail. A description of the mechanisms (e.g. physical and psychological maturity, experience; see van den Honert, 2012) and remedies (e.g. the quota system proposed by Barnsley & Thompson, 1988; Raschner, Muller, & Hildebrandt, 2012) related to such effects is beyond the scope of the study.

Method

Sample

The participants of this study are professional soccer and baseball players in Japan. At present, there are 12 teams in the Nippon Professional Baseball Organization (NPB). Each team has its own farm team, and the total number of players listed on the rosters during the 2012 competitive season was 811. The Japan Professional Football League (J. League) consists of 40 teams divided into two divisions, representing a total of 1013 players registered in the 2012 season. In both the NPB and J. League, all of the players are male. Birth dates and birthplace of the players were collected via the official NPB and J. League website. As with the preceding analysis (Nakata & Sakamoto, 2011), we restricted the sample to Japanese players. In this study, Japanese was defined as those who were born in Japan regardless of their nationality. The school year in Japan begins April and ends on March of the following calendar year, which corresponds with the competitive season of most professional sports. The selection process for elite athletes in Japan is related with sports activity in school, particularly in baseball. All Japanese players in the NPB had experience of playing baseball in extracurricular clubs. Only one-quarter of the players in the J. League were from youth teams in the J. League (Kanekiyo & Hirata, 2012), while the remainder were members of extracurricular clubs or university soccer teams. Taking this relationship into consideration, the sample was restricted to Japanese players. The sample was further restricted by the player's age. The three panels of Figure 1 illustrate the age distribution among the players whose ages ranged from 19 to 33 years. There is a clear break at 23 years of age, indicating that there are many players who play as university team members and join professional teams following their graduation. In order to be included in the sample, we set the lower age limit for players at 23 years. It is also

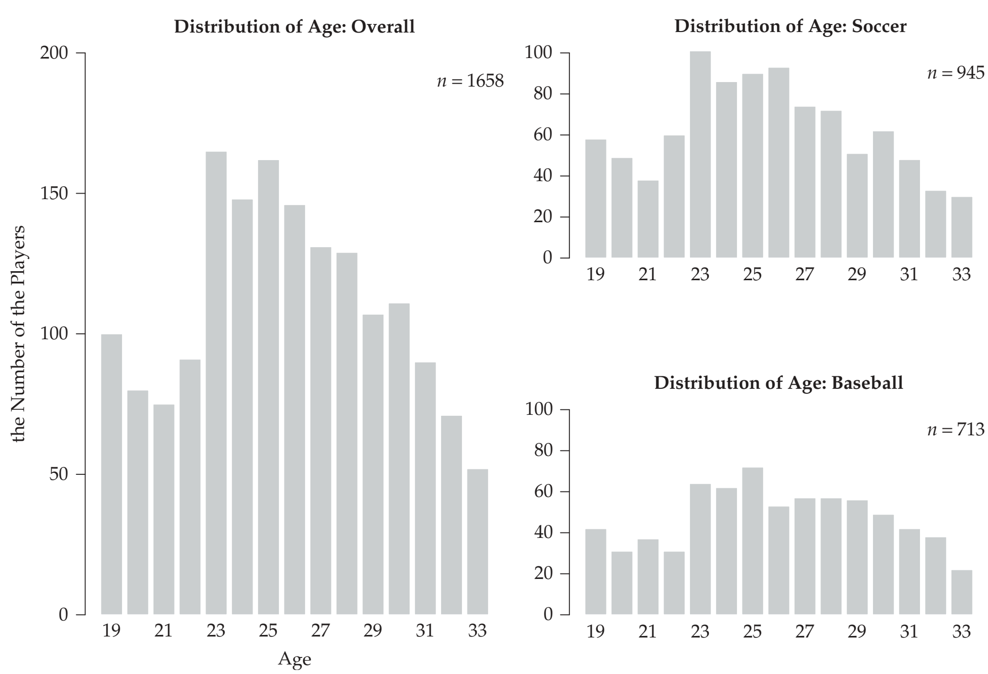

Figure 1. Distributionof age: (a) overall (left panel); (b) soccer (upper-right panel); (c) baseball (lower-right panel).

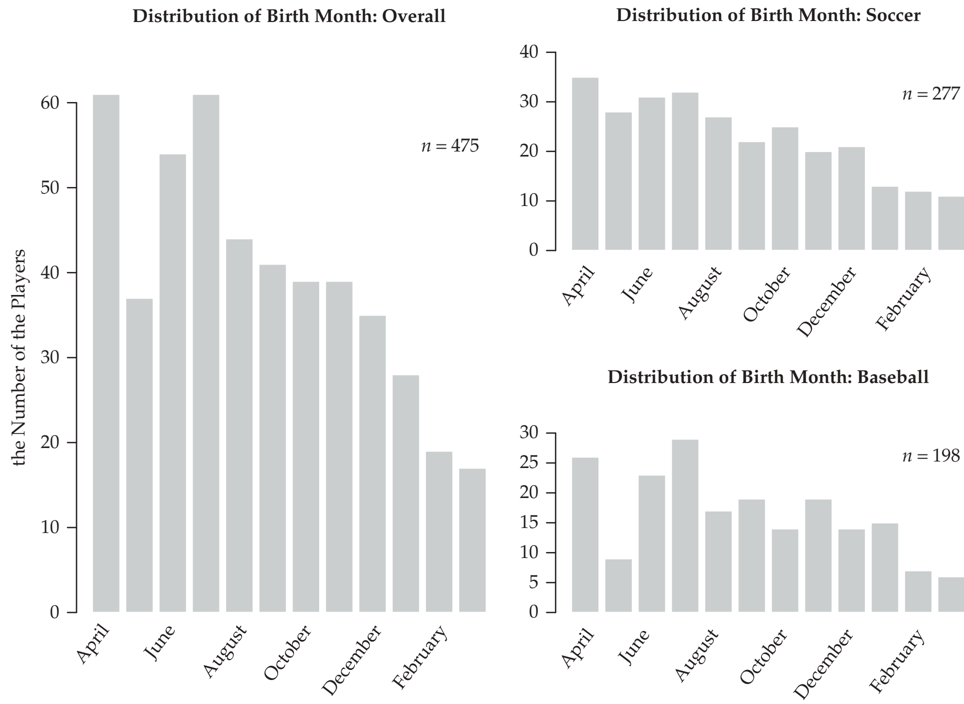

Figure 2. Distribution of birth month: (a) overall (left panel); (b) soccer (upper-right panel); (c) baseball (lower-right panel).

necessary to set the upper limit. Since the proportion of retiring players increases with age, and retired players are not the subject of this study, the number of top players dropping out of the sample increases with age. As shown in Figure 1(a), because the number of players over 25 years begins to decrease, we set the upper age limit at 25 years. The total number of players included in the study was 475: 198 baseball and 277 soccer players. The distribution of birth months is depicted in Figure 2. To assess the robustness of the results, we estimated and reported the effect sizes for another two samples consisting of players in the age ranges of 21–27 years (918 players) and 19–30 years (1445 players).

Statistical framework

Most of the previous studies examining RAEs (e.g. Baker et al., 2014; Côté et al., 2006; MacDonald et al., 2009; Raschner et al., 2012; Turnnidge et al., 2014) have used the χ^2 test to compare the birth date distribution of players to that of the general population, or to detect any significant deviation from a uniform distribution over quarters. Although it is possible to detect differences in the distribution by the test, it is not possible to estimate the magnitude of the effects by χ^2 test. We therefore employed an alternative statistical method to estimate the magnitude of these effects. Becoming a professional sport player can be seen as an "event", and a number of them can be regarded as a "count of events". The standard statistical method applied to the analysis of count data is the Poisson regression model. Utilising this type of regression model makes it possible to estimate simultaneously, not separately, the magnitude of the effects of several factors influencing the likelihood of playing a sport professionally. While we considered three factors in this study (relative age, population size and weather conditions), we could have used regression analysis to examine four or more factors that are expected to influence the likelihood. Another benefit of the regression model is that we can estimate the magnitude of the RAE for a distinct period of time (e.g. per month, per quarter), whereas it is impossible to obtain such an estimate by comparing the numbers of players in each month or quarter. The statistical model is thus very useful to investigate the effect of relative age; nevertheless, only a few studies have thus far used the Poisson regression model. In line with the studies of Hollings, Hume, and Hopkins (2014) and Bruner et al. (2011), our analysis demonstrates the advantages of this method for examining the RAE.

Variables

With the exception of a few previous studies (Augste & Lames, 2011; Vaeyens, Philippaerts, & Malina, 2005; van den Honert, 2012), most previous studies have converted daily birthdate data into quarterly data, leading to a loss of information contained in the original data. We treated birthdate as monthly data to reduce this loss of information. Because school year and competitive season both begin in April (March in the case of J. League) in Japan, relative age was coded as 0 (April) to 11 (March). The J. League season begins in March because of scheduling issues such as international "A" matches arranged by FIFA. Nonetheless, the relative age of soccer players was coded similar to that of baseball players for the following two reasons. First, the teams at which professional soccer players begin their careers organise boys' teams in the same manner as schools. Many of the players in the J. League began their careers at extracurricular clubs organised by school grade. J. League teams also have junior youth (U-15) and junior (U-12) teams. Junior youth teams consist of junior high school students. Team members are not organised strictly according to actual age. For instance, students who reach their 15th birthday during the school year are eligible to join the team. Similarly, youth teams consist of high school students, not necessarily players aged under 18 years. Although these teams are independent of schools, the requirement for membership is the same as that for schools. It is thus reasonable to categorise the relative age of soccer players according to school years. Second, Figure 2(b) clearly shows that a negative trend begins in April, not in March.

A participant's birthplace is defined as the prefecture the player was born in. A prefecture is a first-order administrative district in Japan, and there are 47 prefectures in total. The population size in birthplace is defined as the number of residents in prefecture. Weather condition is defined as the number of snow days per year in the prefecture. In the estimation, these two variables were averaged over the years when the participants were born.

The number of professional players in a month is small if the number of children born in that month is small. For example, since February has only 28 days (or 29 days in a leap year), the number of births in February is smaller; accordingly, the number of professional players born in February should also be smaller. The assumption that the distribution of the number of births is uniform across months of the year is clearly invalid. Nevertheless, most of the studies make this assumption when estimating RAEs (Baker & Logan, 2007; Baker et al., 2009; Côté et al., 2006). Instead, our model includes the total number of male children born in each month as an offset term, which is a variable whose coefficient is fixed at one. The variable in the estimated equation is the total number of male births over the years when the sampled players were born.

Bayesian hierarchical regression model

We estimated the effects using a Bayesian hierarchical Poisson model (Gelman & Hill, 2007; Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2013; Ntzoufras, 2009). The likelihood function of the model is given by

$$y_{ijk} \sim \text{Poisson}(\lambda_{ijk}), i = 1, \ldots, 12; j = 1, \ldots, 47; k = 1, 2;$$

$$\lambda_{ijk} = \theta_{ij} \exp\Big(\alpha_j + \beta_k \text{RA}_i + \gamma \text{Weather}_j + \delta_1 \text{Pop}_j + \delta_2 \text{Pop}_j^2 + \varepsilon_i + \eta I[k = 1]\Big),$$

where the symbol "~" means "is distributed as", y_{ijk} is the number of professional players and subscript i, j and k indicate birth month, prefecture and sport (soccer or baseball), respectively. θ_{ij} is the total number of male children born in month i in prefecture j. The intercept terms, α_j, aim to capture the difference in the likelihoods between the birthplaces that remains after controlling for the two factors of population size and weather conditions. By comparing the 47 intercept terms, we can assess the extent to which birthplace affects the likelihood of playing a professional sport. RA_i is the relative age of those born in month i, while the coefficients on the variable, β_k, measure the RAE. Since the sizes of the RAEs may be different depending on the type of sport, we have formulated the model such that the difference can be evaluated. The variables Weather_j and Pop_j are standardised, i.e. subtract the mean and divide by the standard deviation, in the estimation. We also include a quadratic term of Pop_j because previous studies examining the effect of population size suggest that the effect is not linear (Bruner et al., 2011; Côté et al., 2006; Turnnidge et al., 2014). The parameters εs are intended to control for season of birth effects (Bedard & Dhuey, 2006; Wattie, Cobley, & Baker, 2008). Note that εs are common to all prefectures and both types of sports and that they are constrained to sum to zero. The last term captures the difference in the probability of becoming a professional player in soccer versus baseball, where I is an indicator function that takes the value of one if the condition in brackets is true (in the case of soccer) and zero otherwise. The sign of the coefficient is expected to be positive because the number of professional soccer players is larger than that of baseball. The exponential term in the second equation gives the probability of becoming a professional player. We can then compare the probabilities of becoming a professional player for children under different conditions. Comparing the number of players born in each month yields an imprecise estimate of the RAE because this number is a product of the total number of births in each month and the probability, and thus the estimates are influenced by the total number of births, which is irrelevant to the RAE. Therefore, studies that do not consider the total number of births (e.g. Côté et al., 2006; Raschner et al., 2012) have yielded imprecise estimates. In this regard, our estimation method provides more reliable estimates.

The prior distribution for parameters α_j is given by

$$\alpha_j \sim \text{Normal}\left(\mu_\alpha, \sigma_\alpha^2\right), \mu_\alpha \sim \text{Normal}\left(0, 100^2\right),$$
$$\sigma_\alpha \sim \text{Uniform}(0, 100),$$

and for β_k is given by

$$\beta_k \sim \text{Normal}\left(\mu_\beta, \sigma_\beta^2\right), \mu_\beta \sim \text{Normal}\left(0, 100^2\right),$$
$$\sigma_\beta \sim \text{half - Cauchy}(5).$$

The priors for the variance parameters (${\sigma_\alpha}^2$ and ${\sigma_\beta}^2$) are recommended by Gelman (2006), and other priors used in the analyses are also uninformative or weakly informative. We estimated the three models, which are distinct in the explanatory variables, to compare with each other, thereby we can make an inference regarding the birthplace effect. Comparing the posterior distribution of ${\sigma_\alpha}^2$ in models reveals whether the population size and/or the weather condition have an influence on the likelihood. The posterior estimate of ${\sigma_\alpha}^2$ in a model should be smaller than in other models if the variables included in the model are associated with the variations in αs.

We ran the Markov chain Monte Carlo (MCMC) algorithm for 25,000 iterations for five chains. After the first 5000 samples were discarded as burn-in, samples were saved every 100 iterations to obtain 1000 MCMC samples. Posterior computations were run using JAGS (Plummer, 2012) from R (R Development Core Team, 2013). We also sampled the posterior distribution by using Stan (Stan Development Team, 2013) and confirmed that the results were consistent. The convergence of the MCMC algorithms was then checked by using the Gelman–Rubin statistic (Gelman & Rubin, 1992) and the Geweke's convergence diagnostic (Geweke, 1992).

Measures

One cannot assess the effect sizes directly from the parameter estimates of the Poisson regression model. Furthermore, one cannot compare our estimates with those of previous studies unless we report them in commonly used terms. For instance, studies using a Poisson regression model (Bruner et al., 2011; Hollings et al., 2014) have reported their estimates of RAEs in terms of relative risk. The method employed in this study is flexible in that the effect sizes can be computed and presented in several terms often used in the literature. We therefore report the magnitude of these effects by using three measures: "disadvantage", relative risk and odds ratio. Of these three measures, relative risk and odds ratio have often been used to report the estimate of the effect size. "Disadvantage", which we propose in this study to present the effect size, is designed to evaluate the extent to which children have a relative disadvantage. For example, suppose that the estimated RAE is 10% per month measured by "disadvantage". Then, the probability of becoming a professional player for May-born children is 90% of that for those born in April, other conditions (including the total number of boys born in each month) being equal. By contrast, relative risk, which is well known in epidemiology, measures the degree of relative advantage. In the study of RAEs, "risk" means the probability of becoming a top athlete. Therefore, relative risk indicates the ratio of the probabilities of becoming a top athlete for children under different conditions. The larger relative risk is, the more advantageous children with early birthdates in a given school year have in the competition to become a professional player. The probability of becoming a professional player for April-born children is 1.5 times that for those born in May if the estimated relative risk is 1.5. Unlike these two measures, the odds ratio, which is also often used in epidemiology, does not have an intuitive interpretation. However, odds ratios approximately equal relative risk when the risk of an event is low (Vittinghoff, Glidden, Shiboski, & McCulloch, 2012). Since the ratio of the number of professional players to the total number of births is 0.02% for our sample, we can expect that approximation to work well. To be certain, we note the definitions formally. Let p_i and p_{i+1} denote the probabilities of becoming a professional player for children born in month i and the next month in a given school year, respectively. "Disadvantage", relative risk and odds ratio are then defined as $(p_i - p_{i+1})/p_i$, p_i/p_{i+1}, and $(p_i/(1-p_i))/(p_{i+1}/(1-p_{i+1}))$, respectively. While these are the effect sizes per month, we can estimate the maximum size (i.e. per year) of the RAE by comparing the probability for April-born with that for March-born children. Let p_3 and p_4 denote the probabilities for March-born and April-born children, respectively; the maximum effect size measured by relative risk is then computed as p_4/p_3. Likewise, we can obtain estimates of the maximum size expressed in terms of "disadvantage" and odds ratio. The effect size of weather conditions and population size measured by these terms can be computed similarly.

Results

Table I reports the posterior distribution of the parameters of interest. The convergence of Markov chains to their stationary distributions is confirmed by the statistics given in Table II. Model 2 includes the weather conditions, and model 3 also adjusts for the population size. From the results shown in Table I, we can confirm that the relative age has an effect on the likelihood of becoming a professional sports player. The third row "Sports: Soccer" presents the estimates of the parameter η. The estimates are positive in all models, as expected. The results of models 2 and 3 indicate that the weather conditions also affect the likelihood. The last row labelled "DIC" reports deviance information criterion, which is a measure used to compare models in Bayesian analysis (Spiegelhalter, Best, Carlin, & van der Linde, 2002). As with other information criteria such as the Akaike information criterion, a model with lower DIC is favoured. By comparing DICs, it is evident that models 2 and 3 are more favourable than model 1. The DIC of model 2 is slightly smaller than that of model 3. Moreover, the 95% highest posterior density (HPD) intervals for the parameters of population size and its squared term include zero. We included the population size in model 3 with the expectation that it would account for the difference in the likelihoods between birthplaces. Nevertheless, the posterior estimate of σ_α from model 3 was larger than that from model 2. These results strongly suggest that population size has almost no influence on the likelihood of becoming a professional athlete. Hence, the following analyses are based on the results of model 2. The

Table I. Posterior distribution for the parameters (aged 23–25 years).

	Model 1				Model 2				Model 3			
			95% HPD				95% HPD				95% HPD	
Variable	Mean	s	2.5%	97.5%	Mean	s	2.5%	97.5%	Mean	s	2.5%	97.5%
Relative age: soccer	−0.0934	0.0214	−0.1382	−0.0553	−0.0945	0.0217	−0.1355	−0.0526	−0.0958	0.0223	−0.1385	−0.0522
Relative age: baseball	−0.0800	0.0242	−0.1280	−0.0337	−0.0810	0.0251	−0.1289	−0.0320	−0.0838	0.0245	−0.1302	−0.0373
Sports: soccer	0.3938	0.1515	0.0750	0.6608	0.4000	0.1565	0.0848	0.6999	0.3896	0.1554	0.0726	0.6786
Weather					−0.3307	0.0742	−0.4727	−0.1875	−0.3353	0.0724	−0.4733	−0.1894
Population									0.0125	0.0922	−0.1765	0.1844
Population squared									−0.0101	0.0327	−0.0733	0.0547
σ_α	0.3157	0.0940	0.1532	0.5196	0.1373	0.0851	0.0017	0.2918	0.1524	0.0944	0.0006	0.3217
DIC	1669.05				1654.27				1656.90			

Notes: s, standard deviation; DIC, deviance information criterion.

Table II. Convergence check.

	Model 1				Model 2				Model 3			
Parameter	Gelman	Geweke (1)	Geweke (2)	Geweke (3)	Gelman	Geweke (1)	Geweke (2)	Geweke (3)	Gelman	Geweke (1)	Geweke (2)	Geweke (3)
Birth month: soccer	1.04	1.63	1.16	−0.39	1.01	−1.26	0.55	1.90	1.04	0.28	0.69	−0.43
Birth month: baseball	1.00	0.28	0.86	1.22	1.01	−0.94	1.06	1.40	1.06	0.34	0.89	−0.94
Sports: soccer	1.01	−1.19	1.13	1.44	1.00	1.13	−0.04	0.54	1.01	0.00	1.82	−1.72
Weather					1.01	−0.08	1.92	−0.57	1.02	0.48	0.14	0.16
Population									1.01	0.94	0.79	1.44
Population squared									1.01	−0.71	−1.38	−1.34
σ_α	1.01	−0.06	0.26	1.01	1.01	−1.04	1.50	1.88	1.02	−0.39	−0.84	1.02

Notes: "Gelman" indicates the Gelman–Rubin scale reduction statistic; "Geweke" indicates Geweke's Z-diagnostic. The numbers in parentheses indicate the order of chains. Only the first three of the five chains are reported to save space.

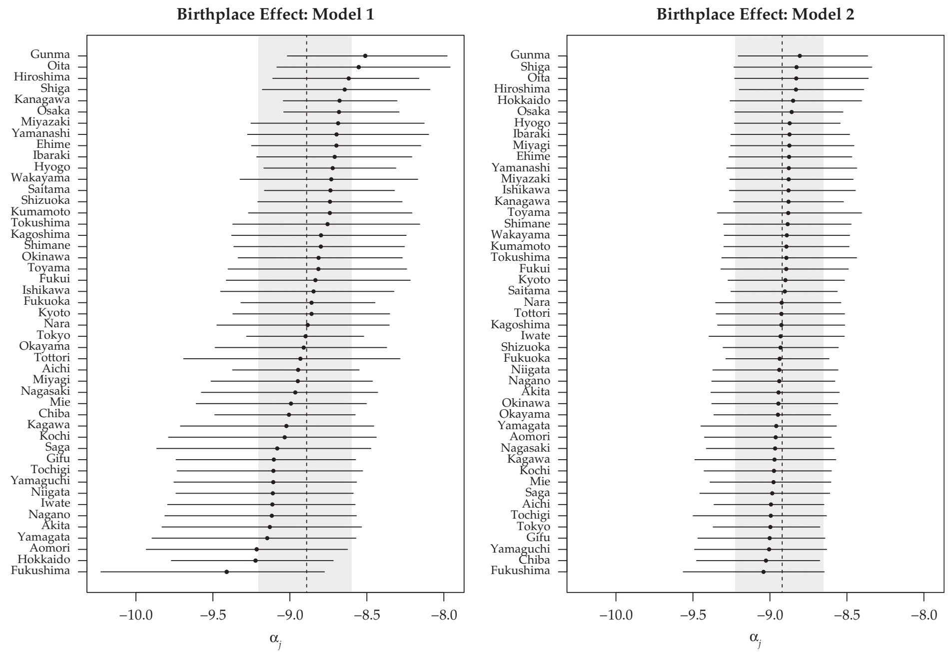

Figure 3. Birthplace effect: (a) model 1 (left panel); (b) model 2 (right panel). The dashed vertical line indicates the mean of hyper-parameter, and the shadowed area indicates 95% HPD intervals.

Table III. The magnitude of the effects (aged 23–25 years).

	Disadvantage (%)				Relative risk				
Effect	Mean	s	2.5%	97.5%	Mean	s	2.5%	97.5%	Odds ratio
RAE: soccer (monthly)	9.0	2.0	5.1	12.7	1.099	0.024	1.054	1.145	1.099
RAE: baseball (monthly)	7.7	2.3	3.2	12.1	1.085	0.027	1.033	1.138	1.085
RAE: soccer (April vs March)	63.6	8.9	43.9	77.5	2.909	0.724	1.717	4.294	2.909
RAE: baseball (April vs March)	57.4	12.0	35.2	79.1	2.531	0.715	1.344	3.979	2.532
Weather (per snow day)	1.1	0.2	0.6	1.5	1.011	0.002	1.006	1.015	1.011
Weather (Hokkaido vs Okinawa)	72.0	8.3	56.4	86.8	3.893	1.195	1.952	6.163	3.893

estimate for σ_α from model 2 was 43% of that from model 1. This indicates that 57% of the difference in the likelihoods between prefectures is accounted for by weather conditions. By comparing the two panels of Figure 3, we can see that the variation in the birthplace parameters is reduced considerably.

Table III presents the estimates of the effects of relative age and weather conditions on the likelihood of becoming a professional player, which are given in terms of the three measures explained above. These estimates were computed directly from the MCMC samples of the model parameters. The RAE evaluated in terms of "disadvantage" was 9.0% per month and the maximum size was 63.6% in soccer, while in baseball these percentages were 7.7% and 57.4%, respectively. The likelihood decreases by 1.1% per snow day. The maximum size of the weather effect is observed between the southernmost prefecture, Okinawa, and the northernmost prefecture, Hokkaido, whereby the estimate was 72.0%. The RAEs evaluated in terms of relative risk were about 1.1 per month for both types of sports. This finding means that the probability of becoming a professional player for April-born children is 1.1 times that for those born in May. Further, April-born children have the advantage by a factor of 2.9 over those born in March in becoming a professional soccer player, whereas the figure for baseball is 2.5. As expected, the odds ratio estimates coincided with those of relative risk up to the second decimal place. The four panels in Figure 4 depict the magnitudes of these effects.

Table IV provides the results of comparing the magnitude of the effects. The first row reports the difference in the RAE measured in terms of the maximum size of the "disadvantage" between the sports. The last column contains the posterior probability of the RAE in soccer being larger than that in baseball. Since the probability is 67.6%, we conclude that there is no statistical

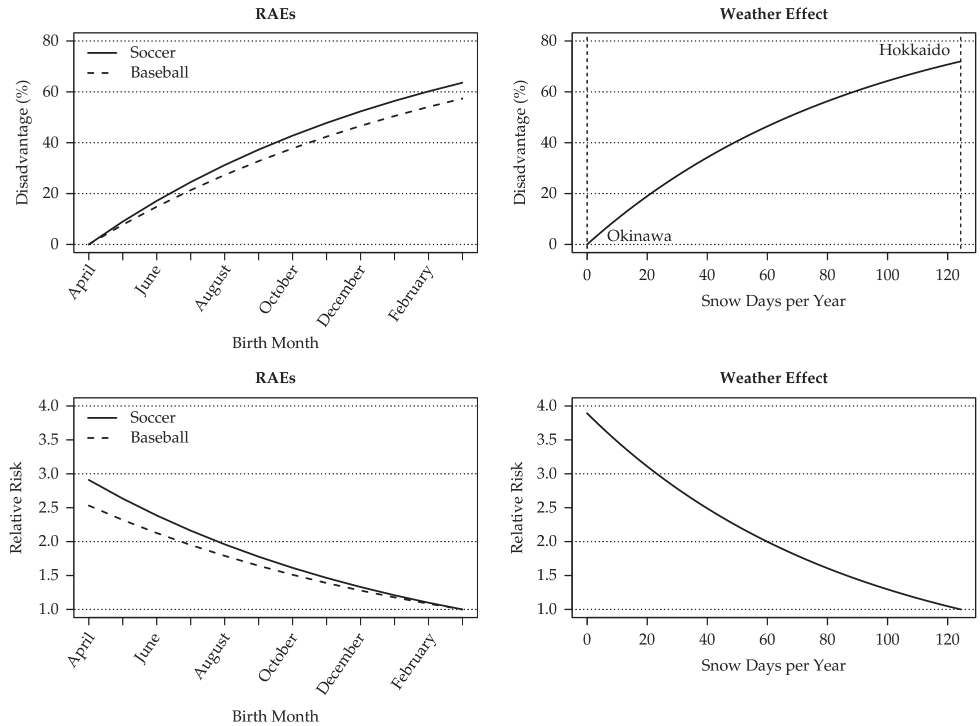

Figure 4. (a) RAEs (upper-left panel); (b) weather effect (upper-right panel); (c) RAEs (lower-left panel); (d) weather effect (lower-right panel).

Table IV. Comparing the effects (aged 23–25 years).

Effect	Mean	s	95% HPD 2.5%	95% HPD 97.5%	Prob($x > y$)
RAE: soccer (March) vs RAE: baseball (March)	6.2	12.3	−14.3	33.1	0.676
Weather (Hokkaido) vs RAE: soccer (March)	8.4	12.4	−16.9	31.9	0.764
Weather (Hokkaido) vs RAE: baseball (March)	14.6	14.8	−13.1	45.8	0.854
Place (Gunma) vs place (Fukushima)	0.2	0.3	−0.2	0.9	0.843

Notes: Estimates for "x vs y" were computed as $x - y$. Prob($x > y$): estimated probability of x being larger than y. The birthplace effects were compared by computing the difference in α_j.

difference in the RAE between these sports. It is not straightforward to compare the sizes of the RAE with weather effects. RAEs were measured in months, whereas the weather effect was measured per snow day. This implies that one cannot directly compare them because they do not have common units of measurement. We address this question by comparing the maximum sizes of the effects: that of the weather condition was 72%, and was larger than those of the RAE. However, the 95% HPD intervals of the difference between the maximum sizes include zero, and the posterior probabilities that the maximum size of weather conditions is larger than that of the RAE in soccer and baseball are 76.4% and 85.4%, respectively (see the second and third rows in Table IV). From these results, we cannot assert that the maximum size of weather conditions is larger than those of the RAE. This finding implies that weather conditions are an equally important factor influencing the likelihood of becoming a professional player in Japan. The last row of the table presents the results of comparing the birthplace effect. The largest estimate (smallest in absolute value) for α is found for Gunma prefecture, while the smallest is found for Fukushima (right panel of Figure 3). The 95% HPD interval for the difference in the birthplace parameters between Gunma and Fukushima includes zero, and the posterior probability of α for Gunma being larger than that for Fukushima is 84.3%. We thus find no conclusive evidence of a difference in the likelihoods between prefectures after adjusting for weather conditions. This result implies that whichever

Table V. The magnitude of the effects.

Age (years)	Effect	Disadvantage (%) Mean	s	2.5%	97.5%	Relative risk Mean	s	2.5%	97.5%	Odds ratio
21–27	RAE: soccer (monthly)	7.5	1.4	4.8	10.1	1.081	0.016	1.051	1.113	1.081
	RAE: baseball (monthly)	5.8	1.6	2.9	9.1	1.062	0.018	1.023	1.094	1.062
	RAE: soccer (April vs March)	57.1	7.1	42.0	69.1	2.394	0.400	1.681	3.188	2.394
	RAE: baseball (April vs March)	47.5	9.8	29.2	66.4	1.973	0.373	1.290	2.681	1.973
	Weather (per snow day)	1.2	0.2	0.8	1.5	1.012	0.002	1.008	1.015	1.012
	Weather (Hokkaido vs Okinawa)	75.8	5.4	64.5	85.4	4.337	0.994	2.501	6.279	4.337
19–30	RAE: soccer (monthly)	8.4	1.1	6.1	10.4	1.092	0.013	1.065	1.117	1.092
	RAE: baseball (monthly)	6.6	1.2	4.3	9.1	1.071	0.014	1.045	1.100	1.071
	RAE: soccer (April vs March)	61.7	5.1	52.0	71.9	2.656	0.357	1.975	3.345	2.656
	RAE: baseball (April vs March)	52.6	6.9	38.5	65.1	2.152	0.305	1.529	2.729	2.152
	Weather (per snow day)	1.0	0.1	0.7	1.2	1.010	0.002	1.007	1.013	1.010
	Weather (Hokkaido vs Okinawa)	70.3	5.7	58.8	80.6	3.491	0.666	2.268	4.742	3.491

factors explain the remaining difference between prefectures, there is little difference other than weather conditions in Japan.

Table V reports the estimates using a considerably different samples consisting of the players aged 21–27 and 19–30 years. We find that the estimates of the RAEs for these two sample sets were slightly smaller than those for the samples consisting of players aged 23–25 years. Consequently, the maximum size of the RAE in baseball was smaller than that of weather conditions (see the third and sixth rows in Table VI). In soccer, the results for effect size compared with weather conditions were mixed (the second and fifth rows in Table VI). As with the results from the sample consisting of players aged 23–25 years, differences in the sizes of RAEs between the sports were not observed in the two samples (the first and fourth rows in Table VI). Furthermore, for comparative purposes, we estimated the model by using quarterly data on players aged 23–35 years, as shown in Tables VII and VIII. The magnitudes of the RAEs in soccer and baseball were 23.2% and 19.6% per quarter, respectively. The maximum RAEs were obtained by comparing the probabilities for the boys born in the first quarter (from April to June) with the probabilities for those

Table VI. Comparing the effects.

Age (years)	Effect	Mean	s	95% HPD 2.5%	95% HPD 97.5%	Prob($x > y$)
21–27	RAE: soccer (March) vs RAE: baseball (March)	9.6	10.5	−8.2	33.3	0.822
	Weather (Hokkaido) vs RAE: soccer (March)	18.7	9.0	1.3	36.1	0.982
	Weather (Hokkaido) vs RAE: baseball (March)	28.2	11.2	6.0	48.4	0.997
19−30	RAE: soccer (March) vs RAE: baseball (March)	9.1	7.7	−5.3	24.1	0.882
	Weather (Hokkaido) vs RAE: soccer (March)	8.6	7.4	−5.9	23.1	0.882
	Weather (Hokkaido) vs RAE: baseball (March)	17.7	8.7	1.2	34.0	0.990

Notes: Estimates for "x vs y" were computed as $x - y$. Prob($x > y$): estimated probability of x being larger than y.

Table VII. The magnitude of the effects: quarterly data (aged 23–25 years).

Effect	Disadvantage (%) Mean	s	2.5%	97.5%	Relative risk Mean	s	2.5%	97.5%	Odds ratio
RAE: soccer (quarterly)	23.2	4.3	14.6	31.2	1.306	0.074	1.171	1.453	1.306
RAE: baseball (quarterly)	19.6	5.1	10.3	29.4	1.249	0.079	1.115	1.416	1.249
RAE: soccer (Q1 vs Q4)	54.2	7.7	39.2	68.6	2.247	0.386	1.604	3.065	2.247
RAE: baseball (Q1 vs Q4)	47.5	10.1	27.8	64.8	1.973	0.373	1.276	2.696	1.973
Weather (per snow day)	1.1	0.2	0.6	1.5	1.011	0.002	1.007	1.015	1.011
Weather (Hokkaido vs Okinawa)	72.4	7.7	59.0	86.9	3.908	1.115	1.978	6.268	3.908

Table VIII. Comparing the effects: quarterly data (aged 23–25 years).

Effect	Mean	s	95% HPD 2.5%	95% HPD 97.5%	Prob($x > y$)
RAE: soccer (Q4) vs RAE: baseball (Q4)	6.7	12.1	−17.7	29.3	0.700
Weather (Hokkaido) vs RAE: soccer (Q4)	18.2	10.9	−3.9	38.6	0.947
Weather (Hokkaido) vs RAE: baseball (Q4)	24.9	12.7	0.3	49.0	0.978
Place (Gunma) vs place (Fukushima)	0.2	0.3	−0.2	0.9	0.799

Notes: Estimates for "x vs y" were computed as $x - y$. Prob($x > y$): estimated probability of x being larger than y. The birthplace effects were compared by computing the difference in α_j.

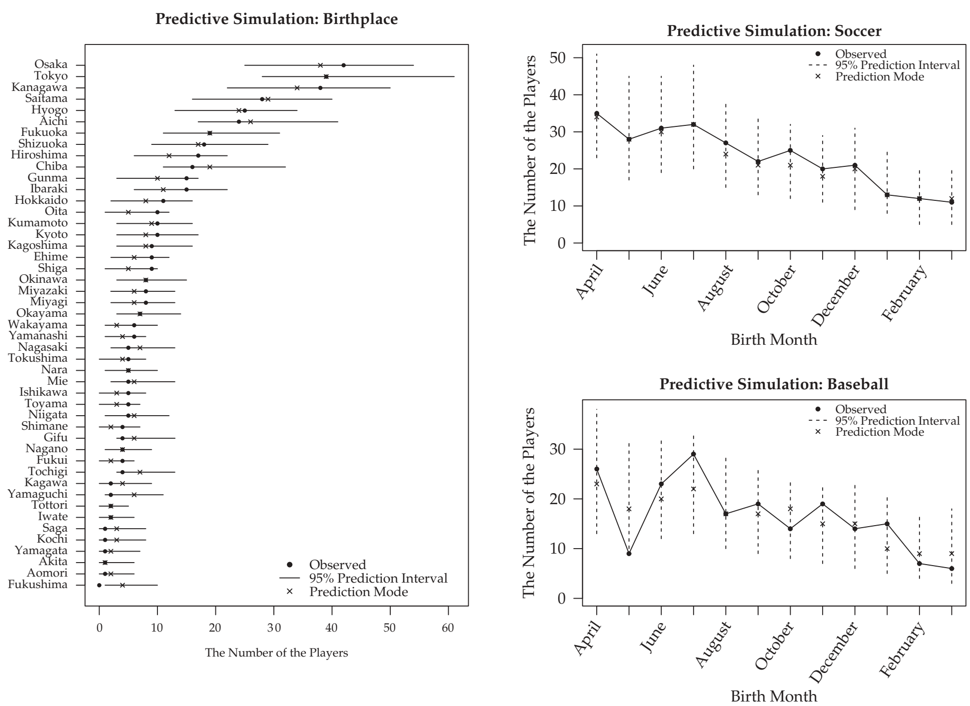

Figure 5. Predictive simulation: (a) birthplace (left panel); (b) soccer (upper-right panel); (c) baseball (lower-right panel).

born in the last quarter (from January to March). The estimates were 54.2% and 47.5% for soccer and baseball, respectively. It is worth noting that these are smaller than the corresponding estimates from the monthly data. The estimates presented in the third row of Table VIII show that relative age in baseball has a smaller effect on the likelihood than weather conditions. Similarly, the magnitude of the RAE in soccer is smaller than that of weather conditions because the probability that the latter is larger than the former is about 95% (the second row of Table VIII). We therefore conclude that the magnitudes of the RAEs are smaller than that of weather conditions when using quarterly data. Finally, the three panels of Figure 5 are posterior predictive plots for the number of players aged 23–25 years by birthplace and birth month. The model used to estimate the effects seems to describe the data well. As already seen in Figure 2, the birth month effect in baseball is considerably large; consequently, any deviation of the predicted value from the observed is large.

Discussion

Cobley et al. (2009) reviewed the literature on RAEs and calculated the magnitude of the effect from the data provided by original studies. Since their study is based on quarterly data, we compare their estimates with ours reported in Table VII. They reported that the maximum size (i.e. first quarter vs fourth quarter) of the RAE in professional soccer ranged from 1.44 to 2.20 (excluding non-significant results) in terms of the odds ratio, whereas our estimate was 2.25. In professional baseball, their estimates ranged from 1.12 to 2.39, and ours was 1.97. Although the estimate for soccer lies slightly outside the range of previous estimates, our estimates are mostly consistent with theirs. This consistency of the estimates across the samples from different countries shows that the strength of the RAE is independent of the sports system (e.g. talent development, selection process) in the country.

Comparing the estimates reported in Table III with those in Table VII shows that the maximum size of the RAE in both types of sports is smaller when using quarterly data. The distance between the compared points (April and March) is 10 months for the monthly data in the estimation of maximum size, whereas the shortest distance, namely the interval from the last month of the first quarter (June) to the first month of the last quarter (January), is only 6 months for the quarterly data. As already found by Cobley et al. (2009), this shortness leads to the underestimation of the effect size. Most previous studies have used quarterly data to examine the RAE; thus, they have underestimated its magnitude. The degree of underestimation in terms of relative risk is over 20% for both types of sports. This figure is not negligible, especially in comparison with the effects of other factors. In this regard, although Côté et al. (2006) found that the birthplace effect was larger than the RAE, their finding seems to be inconclusive because they used quarterly data. Moreover, the results of the studies (e.g. Helsen et al., 2012; MacDonald et al., 2009) that did not detect the RAE may alter if using monthly data. Comparing the estimates given in Table III with those in Table V indicates another source of underestimation for the RAEs for professional players. The latter was estimated for players aged 21–27 and 19–30 years, ranges that were wider than those in the former. The sample consisting of players within a wider age range includes neither players who are members of university teams as of the 2012 competitive season nor those who retired before age 27 or 30. This fact implies that the sample does not cover all top players. Our findings thus demonstrate that such inadequate coverage is likely to degrade the estimation accuracy. The wider the age range for the players to be sampled, the greater is the number of sampled participants. Extending the range, however, paradoxically excludes certain players from the sample. Thus, we should be cautious when setting the range of players' ages to estimate the size of the RAE for professionals.

In studies examining birthplace effects, despite the fact that the number of residents is a continuous variable, many researchers treat it as a categorical variable with several levels (Baker & Logan, 2007; Bruner et al., 2011; Côté et al., 2006; Turnnidge et al., 2014). Whatever the number of categories they use may be, the cut-off values are set arbitrarily and are not based on a statistical analysis; therefore, the results from the procedure may be altered depending on them. Our analyses treat number or residents as a continuous variable, and thereby avoid this problem.

This study has several limitations. First, despite the fact that one's birthplace does not always coincide with the place he/she is developed (Baker et al., 2009; Turnnidge et al., 2014), our analysis used the birthplace as a proxy for the latter. Second, due to data availability, we used the prefecture as the unit of observation for analysis of birthplace effect. A prefecture may be too large to be treated as one's birthplace or the place of development. There is substantial heterogeneity among cities in a prefecture. For instance, population density of Nagoya city in Aichi prefecture is 15 times higher than that of Tahara city. It is suitable to use the city level for examining the birthplace effect. Population size may affect the likelihood of becoming a top athlete if we use city-level data. Thus, our results concerning population size should be interpreted with caution.

Conclusions

We confirmed the presence of RAE in two major professional sports in Japan. The contribution of this study is that it estimated effect sizes that are more precise than those obtained previously and reported them in a readily interpretable form based on a refined statistical method. RAEs were 9.0% and 7.7% per month for soccer and baseball, respectively. These estimates imply that boys born one month earlier in a given school year have a 1.1 times greater chance of becoming a professional soccer or baseball player. Children born in the first month of a school year have a 2.5–3 times greater chance than those born in the last month of the year. We found that birthplace was one of the factors influencing the likelihood of becoming a professional player, but it was not because of the population size. Differences in weather conditions could explain over half of the difference in likelihoods between birthplaces. After adjusting for weather condition, there was almost no difference in likelihoods. Another finding was that the maximum size of the weather effect was almost the same as that of the RAE. Furthermore, we demonstrated that studies using quarterly data have underestimated the maximum size of the RAE. This underestimation causes problems, particularly when comparing the effect sizes of factors influencing the likelihood of becoming a top athlete. Further research is required, particularly examining the birthplace effect through population size.

Acknowledgement

I would like to thank Seinosuke Yano for providing me with valuable information about J. League.

Disclosure statement

No potential conflict of interest was reported by the author.

References

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

Baker, J., Janning, C., Wong, H., Cobley, S., & Schorer, J. (2014). Variations in relative age effects in individual sports: Skiing, figure skating and gymnastics. *European Journal of Sport Science*, *14*, SS183–S190.

Baker, J., & Logan, A. J. (2007). Developmental contexts and sporting success: Birth date and birthplace effects in national hockey league draftees 2000-2005. *British Journal of Sports Medicine*, *41*, 515–517.

Baker, J., Schorer, J., Cobley, S., Schimmer, G., & Wattie, N. (2009). Circumstantial development and athletic excellence: The role of date of birth and birthplace. *European Journal of Sport Science*, *9*, 329–339.

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

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics*, *121*, 1437–1472.

Bruner, M. W., MacDonald, D. J., Pickett, W., & Côté, J. (2011). Examination of birthplace and birthdate in world junior ice hockey players. *Journal of Sports Sciences*, *29*, 1337–1344.

Cobley, S., Baker, J., Wattie, N., & McKenna, J. (2009). Annual age-grouping and athlete development: A meta-analytical review of relative age effects in sport. *Sports Medicine*, *39*, 235–256.

Côté, J., MacDonald, D. J., Baker, J., & Abernethy, B. (2006). When "where" is more important than "when": Birthplace and birthdate effects on the achievement of sporting expertise. *Journal of Sports Sciences*, *24*, 1065–1073.

Dhuey, E., & Lipscomb, S. (2008). What makes a leader? Relative age and high school leadership. *Economics of Education Review*, *27*, 173–183.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 515–534.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Sciences*, *7*, 457–472.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. Berdardo, J. Berger, A. David, & A. Smith (Eds.), *Bayesian statistics 4*. Oxford: Claredon Press.

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

Hirose, N. (2009). Relationships among birth-month distribution, skeletal age and anthropometric characteristics in adolescent elite soccer players. *Journal of Sports Sciences*, *27*, 1159–1166.

Hollings, S. C., Hume, P. A., & Hopkins, W. G. (2014). Relative-age effect on competition outcomes at the World Youth and World Junior Athletics Championships. *European Journal of Sport Science*, *14*, S456–S461.

Kanekiyo, F., & Hirata, T. (2012). J-League Kurabu ni okeru youth shusshin senshu ni kansuru chousa [Research on players from youth academy in J-League clubs]. *Journal of Japan Society of Sports Industry*, *22*, 91–96. in Japanese.

Kawaguchi, D. (2011). Actual age at school entry, educational outcomes, and earnings. *Journal of the Japanese and International Economies*, *25*, 64–80.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.

MacDonald, D. J., Cheung, M., Côté, J., & Abernethy, B. (2009). Place but not date of birth influences the development and emergence of athletic talent in American football. *Journal of Applied Sport Psychology*, *21*, 80–90.

Nakata, H., & Sakamoto, K. (2011). Relative age effect in Japanese male athletes. *Perceptual and Motor Skills*, *113*, 570–574.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Hoboken, NJ: John Wiley and Sons.

Plummer, M. (2012). Just another Gibbs sampler (ver. 3.4.0). Retrieved from http://mcmc-jags.sourceforge.net/

R Development Core Team. (2013). R: A Language and environment for statistical computing (ver. 3.0.1). Retrieved from http://www.R-project.org/

Raschner, C., Muller, L., & Hildebrandt, C. (2012). The role of a relative age effect in the first winter Youth Olympic Games in 2012. *British Journal of Sports Medicine*, *46*, 1038–1043.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B*, *64*, 583–639.

Stan Development Team. (2013). Stan: A C++ library for probability and sampling (ver. 1.3.0). Retrieved from http://mc-stan.org/

Turnnidge, J., Hancock, D. J., & Côté, J. (2014). The influence of birth date and place of development on youth sport participation. *Scandinavian Journal of Medicine & Science in Sports*, *24*, 461–468.

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

van den Honert, R. (2012). Evidence of the relative age effect in football in Australia. *Journal of Sports Sciences*, *30*, 1365–1374.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2012). *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models*. New York, NY: Springer.

Wattie, N., Cobley, S., & Baker, J. (2008). Towards a unified understanding of relative age effects. *Journal of Sports Sciences*, *26*, 1403–1409.