

SCHOOLING AS A POSSIBLE SUCCESS FACTOR? A NOVEL INVESTIGATION OF DETERMINING FACTORS OF SUCCESS IN FOUR SUMMER OLYMPIC GAMES

SOOS I.^{1,*}, KISS T.², WHYTE I.³, HAMAR P. ¹,
BOROS-BALINT I.⁴, SZABO A.⁵

ABSTRACT. The exploration of factors underpinning Olympic success (number of medals won) extended to the Human Development Index (HDI). Analysis of the Rio Olympic Games supported the influence of geographic and social variables. *Schooling* was one of the most important predictors and together with *population* explained 64%; when adding geographic variables (*North-West* and *temperature*) 67% of Olympic medals won were explained. This relationship was validated in the last four Olympic Games (from 2004 to 2016). This is the first study ever to demonstrate that specific social and geographical factors determine more than two thirds of the variance in Olympic success.

Keywords: *Olympic medals, HDI, schooling, population size.*

Introduction

Investigating Olympic success is an interesting and important field of research, given the importance attached to it by both a nations' citizens and governments alike. While there are some conflicting views, Olympic success is widely recognised in many countries as having positive impacts on three discrete yet often overlapping areas: image, economy, and physical activity and health.

Moosa and Smith (2004) identified that Olympic success is a source of national pride or disappointment for citizens and governments, reinforces public good and national identity, and improves a country's international image, promoting its reputation.

¹ University of Physical Education, Budapest, Hungary.

² University of Pecs, Hungary.

³ University of Sunderland, UK.

⁴ University of Babeş-Bolyai, Cluj-Napoca, Romania.

⁵ ELTE Eötvös Loránd University, Budapest, Hungary.

*Corresponding author: soos.istvan@tf.hu.

Economic gains have also been attributed to Olympic success with more public spending during and after the Olympic Games (O'Brien and Onorato 2018). In economic terms, hosting an Olympic Games helps the host city to develop its infrastructure and attract tourists (Feizabadi *et al*, 2012).

Increasingly, local and national governments have linked Olympic success to the quality of people's lives with contemporary agencies focussing on physical activity and well-being. The relationship is two-fold. Olympic success boosts participation in sport, recreation and/or leisure-time physical activity. To prime these opportunities, Olympic success also encourages governments to invest more in sports and leisure facilities and the development of health-related physical activity provision.

While there have been many claims made about the benefits that may accrue from being a successful sporting nation at the Olympics, it is worth noting that there seems also to be mutual relationships between economic, political and social characteristics which can likely impact on Olympic success, either directly or indirectly (Humphreys, Maresova and Ruseski 2012). Amongst those factors, eliminating barriers such as lack of time, poor sports facilities and little financial support for sports participation seem paramount.

At the inaugural modern Olympic Games in Athens in 1896, there were only 245 participating athletes, representing 14 nations. However, over a century later there were forty times more athletes (approximately 11,250) who were representing 207 nations (International Olympic Committee [IOC] 2018). With greater participation and gender equality (BBC Sport 2018), there are more medals now for countries to win, but the numbers of countries and participants taking part have also increased thus forcing a concomitant expansion of competition for medals. However, this increase in the number of participating countries does not always lead to greater spread of medals won. Tcha and Persin (2003) noted in their work that many countries win no medals, while others with higher GDP do better across more sports.

One of the key issues thrown up by this emphasis on Olympic success and relating this success to medals won, has been considered by De Bosscher *et al* (2008). They contended that sporting success can be considered in both absolute or relative terms. In proposing analyses that veer towards relative success, they argued that absolute definition of success in which medals won created a research bias that is difficult to overthrow. For example, total medals won as a measure neglects the quality of those medals: are ten gold medals (outright winners), given the same rankings as five silver and 5 bronze medal winners? They looked at awarding points for different types of medals and found few differences in the ratings of top 10 countries in 'ranking' tables (De Bosscher *et al*, (2008). In fact, Kuper and Sterken (2001) felt that the number of participants from each country provided a better judgment of what made 'success'.

These results led De Bosscher *et al* (2008) to develop their enquiries to consider items that might be used to re-define success, from absolute to relative success. One way that they felt that success could be considered is by using two of the most obvious variables: size of population and wealth of a country. Yet, both measures show different results. For example, when looking at results from Athens in 2004, when medals or positions were divided by population, Bahamas became the most successful nation. (De Bosscher *et al* (2008). Similarly, when a country's wealth was considered the medal tally of China placed them first when GDP per head of population was involved (De Bosscher, 2008). Thus, De Bosscher *et al* (2003; 2008) work concluded that several determinants at time were needed e.g. geographical area; degree of urbanisation; religion; political system).

Until 1989, monarchies and single-party or communist systems employed different approaches to participation, training, and incentives for success on the world stage. This skewed historical data with respect to medals won due to the role of the ideological or political will of countries' leaders (Green and Oakley, 2001; Green, 2005; Tan and Green, 2008; Andreff *et al*, 2008). Political support leads to institutional support in the form of finance and rewards, which in turn leads to improved services for athletes and coaches. Very often, they also hold numerical advantages in population terms a determining characteristic that was supported by Lui and Suen (2008) and Soos *et al* (2017).

Overall, a tendency has emerged, that many wealthy nations (e.g. Germany, United Kingdom and France) as well as some other countries (e.g. United States and Russia) with large populations and strong athletic traditions, state-of-the-art training facilities, and significant sports science and coaching support have become very successful over a long period and several Olympic cycles (e.g. Andreff, 2001; Lui and Suen, 2008; Soos *et al*, 2017).

Recently, links point to education being of importance in promoting success (Lawrence, 2017; Noland and Stahler, 2017). The latter authors stated: "Rather than per capita income, education is much more positive determinant of medal winners" (Noland and Stahler, 2017, p4).

The complexity of differentiating predictors of success can be highlighted in this area with Noland and Stahler's (2017) work identifying tight correlations between educational attainment and income measures, indicating that more affluent countries invest more heavily in education.

There have been several hypotheses that suggest what might be the most influential factors to affect Olympic medal success. These have included demographic, geographic, political, cultural, and economic hypotheses. Hong (2006) promoted the demographic hypothesis, by describing populous countries, like China or USA, that have a large talent pool based on population size. She suggested

that the size of the population provides greater opportunity to select and train more elite athletes who will be capable of winning medals at the recent Olympic Games (see also Andreff (2001))

The geographic hypothesis (Hoffmann, Ging and Ramasamy 2004) described the benefits of warm weather training conditions in some countries with optimal or more ideal climate and weather conditions being prevalent for sports. This hypothesis has been supported by Soos *et al* (2017), albeit, a more recent study by Vagenas and Palaiothodorou (2019) analysed six Olympic Games and found no climatic trends.

Some governments (like those of countries which were formerly operating with a communist structure) pay a lot of attention to success in sports for political benefits thus shaping what has been termed the political hypothesis (Johnson and Ali 2008).

The cultural hypothesis focuses on participation in mega events (e.g. sport events) that can be identified as a cultural element of a country and thus, there is a will and motivation to make success happen at many levels (Frey, Iraldo, and Mellis 2007).

Finally, focusing on economics and, in particular *GDP*, is a reasonable approach to studying this topic as it can highlight the impact of a nation's wealth on general sporting participation as well as on elite level performance (Moosa and Smith 2004).

Publication of the *HDI* (Human Development Index; United Nations Development Programme 2017) encouraged authors to consider what is now termed a social hypothesis. Halsey (2009) has already used the Gini coefficient (an index of inequality of income or wealth distribution) and the *HDI* (composite) index as independent variables in a regression analysis to investigate the influential factors of the number of medals won. The Gini coefficient did not prove to be useful predictor; the *HDI* also had a low explanation power (R^2 is about 0.13). Vagenas and Vlachokyriakou (2012) found *health expenditure* as a good explanatory variable while Jayantha and Ubayachandra (2015) identified *schooling* as a possible factor affecting Olympic performance. The authors of this current study investigated not only the *HDI* as a composite indicator, but the parts of the *HDI* as well (*Schooling* and *Life Expectancy*) to examine the social hypothesis in more details.

Existing hypotheses first try to explore the main, measured indicators of a given area. In the case of the nations' wealth, the *GDP* is used, albeit it is obvious that this is only a proxy (see the case of Norway and Kenya) and there are other important factors. Other variables, like unemployment, served as a refinement of the economic side. Such a refinement tool can be e.g. the distribution of wealth: the previously mentioned research of Hasley (2009) used Gini coefficient,

which can be applied for the distribution. Therefore, it might be worthwhile to use it as a weight for the GDP. There are numerous ways for the further exploration, however, findings will always be a proxy, because there are always room for further examinations. In the social-cultural field, a GDP-like main indicator is the HDI, which also did not provide the expected effect. This present study refines the social-cultural field one step further with the inclusion of the two main composites: education and healthy life.

Investigation Aims

The current study served three objectives:

First, we wanted to investigate the traditional determinants of Olympic success, in terms of *medals won* in at least two of the last four Olympic Games (more than one medal indicates that it did not “accidentally” happen and could reflect a systematic approach). Specifically, *GDP per capita* (the economic hypothesis) and the *population* (size of a nation; demographic hypothesis) as the strongest predictors were considered.

Second, the *geographic* and *temperature* data (geographic hypothesis) were included to examine further influential factors considered important in the literature.

Third, the effects of the *HDI* variables (social hypothesis) on Olympic medals were examined respecting the hypothesis that both social and economic factors influenced Olympic success to a large extent.

Methods

This investigation was carried out in three studies: Study 1, Study 2 and Study 3, each serving one of the aforementioned objectives. The reason for three studies is that existing theories required validation using the database being used by this current research (see below for greater detail of the studies).

Database

Olympic medal winning countries from the last four Olympic Games were included in the data analyses. Seventy-five countries won medals at the Athens Olympic Games 2004, eighty-six at the Beijing Olympic Games 2008, eighty-five at the London Olympic Games 2012 and eighty-seven at the Rio Olympic Games 2016. Of these countries, eighty-four were selected as a purposive

sample for analysis as they had won at least one medal in a minimum of two of those Olympic Games and were thus considered ‘successful’ countries in Olympic sport for this investigation. The natural logarithm of *medals won* (per capita), *population*, *GDP per capita* and *temperature* data were used throughout this investigation, because these variables in their original form are highly differentiated from a normal distribution. Analyses in this investigation will always use the logarithm version of these variables without mentioning it separately. Categorical values of *population*, *temperature* and *GDP per capita* (median based split) were also used in order to reduce the strong non-normality of the original variables. *Northern and Southern hemisphere* countries as demographic variables were also treated as a categorical variable. The following variables were employed:

- *LnPop*: Logarithm of *population* (in ten million), 2017, World Bank online sources (The World Bank 2017). *PopCat*: A categorical variable of population was also used with the median based cut value.

- *LnGDPpc*: Logarithm of *GDP per capita* in thousand dollar (current/Olympic year), 2017, World Bank online sources (The World Bank 2017). *GDPcat*: Categorical variable of *GDP per capita* was also used with the median based cut value.

- *LnTemp*: Logarithm of *temperature*, yearly average temperature in Celsius (2017), Weatherbase database (WeatherBase n.d.). *TempCat*: Categorical variable of *temperature* was also used with the median based cut value, low: (1), high: (2).

- *NorthSouth*: *Northern and Southern hemisphere nations*, categorical variable, where the North countries are the Eurasian and the North American (as a continent) countries (1) and the South countries are the rest of the world (2).

- *Total medals won*: Gold, Silver, Bronze medals won in total; obtained from the BBC Sport website (BBC Sport 2018).

- *LnTotMedPc*: Logarithm of the *total medals won per capita*.

The Human Development Index (United Nations Development Programme 2017 and Segura and Birson, 2013, for Puerto Rico’s data) is a composite index, composed of:

- *LifeExpect*: The *life expectancy* at birth (in years);
- *Schooling*: The *mean years of schooling* (in years) and
- The *gross national income per capita* (in dollars).

Total Medals per capita was used as dependent variable in all studies, either for one Games or for all Games in the case of repeated measures methods. GDP was also used as a relative measure, as a per capita GDP, therefore, this study examines the *relative* Olympic success of nations.

Study 1. First, *GDP per capita* and national *population* size were used. The hypothesis was that traditionally agreed variables have a significant effect on the dependent variable (economic and demographic hypotheses).

Study 2. Geographic variables suggested by experts (Hoffman et al. 2004) were included into the analysis. The hypothesis was that in addition to the variables examined in Study 1; these variables will also affect Olympic success (geographic hypothesis).

Study 3. Further analysis was run for the eighty-four countries. This time, the number of variables were extended to include the two additional variables (GNP excluded) reported in the *HDI* (social variables). The hypothesis was that these other social variables will affect Olympic success, however, their influence will not be as strong as the impact of those examined in Study 2.

Description of Statistical Processes and Analyses

First, multiple regression analysis was used on the Rio Games' database to examine the relationship between Olympic success and all the variables, according to each study's hypothesis, described above. The regression analysis in Study 2 and Study 3 uses the backward elimination method.

Independent samples t-tests refined and further explained the results of the regression analysis. Validity of regression analyses were also checked. Normality of dependent variables and residual variables were tested by the Shapiro-Wilk test. Multicollinearity was tested by the VIF values.

Heteroskedasticity was tested by the Breusch-Pagan test. Autocorrelation (relationship between a variable's current value and its previous values) in the specific arrangement of the sample was tested by the Durbin-Watson statistic. This approach is also appropriate for proving that there is no trend in the sample that could result in a statistically spurious correlation.

Following regression analysis, the examination was extended to all the Olympic Games to check whether the explored relationships of the regression analyses are generally valid for all Games or only specific to Rio Games. The selected, most significant variables (Results section will specify them) were tested on the four Games with help of repeated-measures analysis of covariance (ANCOVA; as a General Linear Model in SPSS). The reason for the use of ANCOVA is that the Games (participating countries) are not independent from each other (see the sphericity in the explanation below), however, traditional validation of the results with the comparison of the regression slopes will also be performed.

Statistical analyses were performed using IBM SPSS v. 25 software package, apart from the Breusch-Pagan test for testing homoskedasticity. That was conducted using statistical software R (R Development Core Team, 2008).

Results

Total medals per capita (LnTotMedPc) was used in all studies below as a dependent variable, therefore, the normality test of this variable is generally valid to all studies. According to the Shapiro-Wilk test, the variable has a normal distribution at 5% (Shapiro-Wilk (77) =0.969, p=0.055).

Study 1

First, the effects of the traditionally approved variables: *population* and *GDP per capita* as demographic and economic hypothesis were tested on the database, and afterwards, the difference between the four Olympic Games was analysed using repeated-measures ANCOVA.

Regression Analysis

First, regression analysis on the Rio Games' database was performed (Table 1).

Source	<i>B</i>	<i>SE B</i>	<i>B</i>	<i>t</i>	<i>P</i>	<i>VIF</i>
(Constant)	-.269	.381	.000	-.705	.483	.000
<i>LnPop</i>	-.594	.072	-.645	-8.219	.000	1.059
<i>LnGDPpc</i>	.357	.104	.270	3.437	.001	1.059

a. Dependent Variable: LnTotMedPc

Table 1. The results of Multiple Regression Analysis for the *total medal per capita* with *population* and *GDP per capita*.

Note that values in the table are in logarithm: in the case of a 1 unit increase of the *population (LnPop)* there is a -.594 change in total medals per capita (*LnTotMedPc*). This means – in original values – that 2.718*10 millions of an increase in population results in a 44.8% decrease ($e(2.718)$ unit change in x causes $\exp(B)-1$ percentage change in Y , based on Benoit (2011) in the *medals per capita* (which is .37 medals; the average medals per capita per nation in Rio is .827). The reason is that in smaller countries it is easier to have a higher value for one capita. In the case of *GDP per capita*: 2.718*1000-dollar *GDP per capita* increase results in .355 more medals per capita.

Population and *GDP per capita* explained a significant proportion of variance in the *total medal per capita*. Adjusted $R^2 = .559$, $F(2, 74) = 49.074$, $p < .001$.

Validity

There is no multicollinearity in the model (VIF values are much below 4.0; see the last column in Table 1). The Durbin-Watson statistics' value is 2.01 ($D_u=1.514$) which means that this concrete realisation of the sample does not have autocorrelation. The model is homoskedastic (based on the Breush-Pagan test: $\chi^2(1)=1.$, $p=0.317$). The residual variable is not normally distributed by the Shapiro-Wilk test ($W(77)=0.955$, $p=0.008$). These variables were then tested in the four Olympic Games with repeated-measures ANCOVA.

Repeated measures ANCOVA and the Equality of Regression Slopes

The Mauchly's test indicates that there are dependencies between the Games ($W=0.682$, $\chi^2(5) = 22.09$, $p=0.001$). The *LSD* test indicated that there is no one specific Games which differs from the other and the cause of the dependency: the test revealed no statistically significant differences between the individual Games.

As the Mauchly's test indicated significant differences, we therefore used the Greenhouse-Geisser test (Field 2013, p. 1605), which is a corrected form of testing within-subjects' effects (the differences), caused by the different Games.

In the regression analysis it was shown that 56% is already explained by *population* and *GDP per capita*, therefore, only the remaining differences could be explained by the Olympic Games. *Population* and *GDP per capita* are included in the analysis as between subject variables (covariates). Table 2 depicts the result of the ANCOVA.

The first row of within-subjects effects shows how the Olympic Games could explain the remaining differences after the effect of the covariates. In other words: did the *total medals per capita per nation* depend on what Olympic Games was being considered? E.G. was any difference Games location, or time specific? The significance value ($0.828 > 0.05$) shows that it did not matter where the Olympics took place.

Effect	#	Tests of Within-Subjects Effects					
		<i>SS</i> ^a	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	partial η^2
<i>TotMedPc</i>		.157	2.494	.063	.247	.828	.004
<i>TotMedPc * LnPop</i>		.369	2.494	.148	.582	.597	.010
<i>TotMedPc * LnGDPpc</i>		.073	2.494	.029	.115	.927	.002
Error		37.405	147.159	.254			

Tests of Between-Subjects Effects						
Corrected Model	408.992 ^b	11	37.181	36.831	.000	.584
Intercept	.887	1	.887	.879	.349	.003
Olympics	.513	3	.171	.169	.917	.002
LnPop	263.477	1	263.477	260.995	.000	.475
Olympics*LnPop	.392	3	.131	.129	.943	.001
LnGDPpc	57.231	1	57.231	56.692	.000	.164
Olympics*LnGDPpc	.487	3	.162	.161	.923	.002
Error	291.748	289	1.010			
Total	1060.343	301				
Corrected Total	700.740	300				

a. Calculated using SPSS default (Type III)

b. R Squared = ,584 (Adjusted R Squared = ,568)

Table 2. Tests of Within-Subjects and Between-Subject Effects of repeated-measures ANCOVA for the four Olympic Games, testing the effect of the Games. Covariates are the *population* and *GDP per capita* variables; Greenhouse-Geisser statistics are used.

The next two rows' results proved that none of the two covariates have interactions with the different Olympic Games. The effect size (partial η^2) is negligible in all cases.

Equality of Regression Slopes

The lower part of Table 2 (between-subjects effect) shows the results of the (non-repeated measures) ANCOVA. The results are the same as previously: the different Games do not make any difference (see the *p*-value of the *Olympics* row: *p*=.917). Regression slopes, considering the *population* and the *GDP* do not differ significantly, as the two *p*-values in the rows of *Olympics*LnPop* and *Olympics*LnGDPpc* (*p*=.943 and *p*= .923 respectively) indicate. The estimated Adjusted *R*² is practically the same as in the case of the Regression analysis (=,568).

The tests of the normality of the residuals in the Repeated Measures ANCOVA show non-normality in the case of the Rio Games (*W*(62)=.949, *p*=.012), but the test for the other Games are non-significant.

Thus, these results provided evidence that the traditional factors such as *population* and *GDP per capita* have significant effects on Olympic success (*won medals per capita per nation*); a finding that applies across all four Olympic

Games, as the Repeated Measures ANCOVA and the equality of the regression slopes indicate. Both the demographic and economic hypotheses were partly supported because the normality of the residual variable in the regression analysis and the Rio Games' residual variable from the repeated-measures ANCOVA were rejected.

Study 2

In the Introduction section of this paper, it was suggested that there could be support for both a *geographical location*, as well as a *temperature* effect. Therefore, the next study examined the effect of the variables, conveying those effects through backward regression analysis, looking only for variables that really have significant explanation power. Four variables were used in this study: the *population* and *GDP per capita* data (as earlier), together with the *NorthSouth* and the *temperature* variables. Additionally, three categorical variables (for the *GDP per capita*, the *temperature*, and the *population*) were also included (the *NorthSouth* geographical variable is already a categorical variable).

Regression Analysis

Table 3 displays the details of the results of the regression analysis.

Source	<i>B</i>	<i>SE B</i>	β	<i>t</i>	<i>p</i>	<i>VIF</i>
(Constant)	1.267	.503		2.517	.014	
<i>LnPop</i>	-.578	.065	-.627	-8.829	.000	1.063
<i>TempCat</i>	-.924	.221	-.299	-4.187	.000	1.076
<i>LnGDPpc</i>	.259	.097	.196	2.680	.009	1.124

a. Dependent Variable: *LnTotalMedPc*

Table 3: Results of Regression Analysis for Study 2 considering variables of *population*, *GDP per capita* data, together with the *North South* and *temperature* variables and three categorical variables (for the *GDP per capita*, the *temperature*, and the *population*)*

* Note that as a result of the backward method the *population* category also had a strong influence on the dependent variable, but not as strong as the continuous *population* variable, therefore this variable was manually removed from the independent variable list.

The effect of the *population* and the *GDP per capita* decreased. The *temperature* categorical variables' *B* parameter value (-.924) expresses the difference between the low and high temperature countries; the *number of medals*

per capita per nation is lower by 60% - 0.499 medals per cap (1 unit change in x causes $\exp(B)-1$ percentage change in Y , based on Benoit, 2011). Independent samples t-test results also support the fact that low *temperature* countries win more medals either in total or in per capita.

Independent variables explained a significant proportion of variance in the *total medal per capita*. Adjusted $R^2 = .639$, $F(3, 73) = 39.654$, $p < .001$.

Validity

There is no multicollinearity in the model (VIF values are still below 4.0), despite the fact that both the *population* and the *population* category are elements of the final table. The Durbin-Watson statistics' value is 2.002 ($d_u=1.577$) which means that this concrete realisation of the sample does not have autocorrelation. The model is homoskedastic (based on the Breush-Pagan test: $\chi^2(1)=0.658$, $p=0.417$). The residual variable has a normal distribution at 5% significance level by the Shapiro-Wilk test ($W(77)=0.98$, $p=.278$). These variables are also tested in the case of the four Olympic Games with repeated-measures ANCOVA.

Repeated-measures ANCOVA and the Equality of Regression Slopes

The Mauchly's test also suggested that there are significant dependencies between the Games ($W=0.683$, $\chi^2(5) = 21.618$, $p=0.001$), therefore in this case, the Greenhouse-Geisser test is used as well. Table 4 displays the results of the ANCOVA model.

Tests of Within-Subjects Effects						
Effect	<i>SS</i> ^a	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	partial η^2
<i>TotMedPc</i>	.738	2.497	.296	1.166	.321	.020
<i>TotMedPc * LnPop</i>	.300	2.497	.120	.474	.666	.008
<i>TotMedPc * LnGDPpc</i>	.277	2.497	.111	.438	.690	.007
<i>TotMedPc * TempCat</i>	.689	2.497	.276	1.089	.349	.018
Error	36.716	144.828	.254			
Tests of Between-Subjects Effects						
Corrected Model	469.162 ^b	15	31.277	38.493	.000	.670
Intercept	27.648	1	27.648	34.026	.000	.107
Olympics	.172	3	.057	.071	.976	.001
<i>LnPop</i>	242.028	1	242.028	297.860	.000	.511

Olympics*LnPop	.145	3	.048	.060	.981	.001
LnGDPpc	23.043	1	23.043	28.359	.000	.091
Olympics*LnGDPpc	.425	3	.142	.174	.914	.002
TempCat	59.816	1	59.816	73.615	.000	.205
Olympics*TempCat	.191	3	.064	.078	.972	.001
Error	231.578	285	.813			
Total	1060.343	301				
Corrected Total	700.740	300				

a. Calculated using SPSS default (Type III)

b. R Squared = ,670 (Adjusted R Squared = ,652)

Table 4. Tests of Within-Subjects Effects of repeated-measures ANCOVA for the four Olympic Games, testing the effect of the Games. Covariates are the *population*, *GDP per capita* and the *temperature* category variables. Greenhouse-Geisser statistics are used.

Covariates (population, GDP, temperature) partially explained the causes of Olympic success and only the remaining part is left for the explanation by the different Games. The first row of the within-subjects' effects (*TotMedPc*) demonstrates this effect of how the Olympic Games could explain the remaining differences? The significance value ($0.321 > 0.05$) shows that no significant differences were observed between the Olympic Games after the influence of the covariates (see also the small effect sizes in each factor).

Equality of Regression Slopes

The lower part of Table 4 (between-subjects effect) shows the results of the ANCOVA. The results also reinforced previously demonstrated results: different Games do not make any difference ($p=.976$). Regression slopes, considering the population, GDP and the temperature category do not differ significantly, as the p -values in the rows of *Olimpics*LnPop* and *Olimpics*LnGDPpc* and *Olimpics*TempCat* ($p=.981$, $p=.914$ and $p=.972$ respectively) indicate. The estimated Adjusted R^2 is also remarkably similar to the Regression analysis ($=.652$).

The tests of the normality of the Repeated Measures ANCOVA model's residual variables are all non-significant, in other words, they are not significantly different from a normal distribution.

This section provided evidence that the *temperature* category (geographic hypothesis) also has significant explanatory power on the Rio Olympic successes (medals won), and - considering all the four Games - all the three involved variables (*GDP*, *population* and *temperature*) show significant effects. There is no difference between the four Games in this respect.

Study 3

Since influential factors that had previously been suggested, both in the literature and already proved by this investigation, were all such factors which can hardly be changed by a country (*population*, *GDP per capita*), the analyses were extended to consider social factors, including the *HDI* variables. The basic research still focussed exclusively on the Rio Olympic Games, employing also all *HDI* (Human Development Index) variables.

Regression Analysis

When the backward process stopped, there remained only those variables which had a significant relationship with the dependent variable. In this variable set, *schooling* from the *HDI* also proved to be significant (Table 5).

Source	<i>B</i>	<i>SE B</i>	β	<i>t</i>	<i>p</i>	<i>VIF</i>
(Constant)	-1.983	1.146		-1.731	.088	
<i>TemCat</i>	-.886	.310	-.283	-2.855	.006	2.233
<i>LnPop</i>	-.483	.071	-.521	-6.816	.000	1.328
<i>NorthSouth</i>	.772	.317	.226	2.434	.017	1.950
<i>Schooling</i>	.242	.066	.351	3.659	.000	2.086

a. Dependent Variable: LnTotalMedPc

Table 5. The results of Multiple Regression Analysis (Backward method) for the *total medal per capita*, starting with all variables.

The effect of the *population* and the *temperature* category is similar as in Table 3. The *NorthSouth* categorical variable's *B* value is seemingly controversial, but it coincides with the result of the *population*. The *B* parameter of *schooling* (.242) means that one more year spent in education increases the *medals per capita per nation* by 27,4% that is .226 medals.

The final model did not contain the *GDP per capita* in any form, but one of the social factors, *schooling* remained in the final model. Demographic, geographic and social variables explained a high proportion (67.4%) of variance in the total medal per capita per nation. Adjusted $R^2 = .674$, $F(4, 70) = 39.295$, $p = .000$.

Validity

There is no multicollinearity in the model (VIF values are much below 4.0) and there is no autocorrelation: the Durbin-Watson statistics' value is 2.056 ($d_u=1.768$). The model is homoskedastic (based on the Breush-Pagan test ($\chi^2(1)=0.616$, $p=0.432$). The residual variable had a normal distribution at 5% significance level by the Shapiro-Wilk test ($W(76)=0.976$, $p=0.175$).

Variations

If the *NorthSouth* variable is removed, the adjusted $R^2=.652$ ($F(3, 71)=40.396$, $p < .001$) is somewhat lower. *Population* and *schooling* together explains 64% of the variation of the *medals per capita* changes (adjusted $R^2=.64$ ($F(2, 72)=66.77$, $p < .001$). None of these models have multicollinearity, autocorrelation and heteroskedasticity and their residual variables are normally distributed.

The variables, the results of the backward methods (*population*, *schooling*, *temperature* category and *NorthSouth* variable) were also tested in the case of the four Olympic Games with repeated-measures ANCOVA. Results are depicted in Table 6.

Repeated-measures ANCOVA and the Equality of Regression Slopes

The Mauchly's test also proved that there are significant dependencies between the Games ($W=0.691$, $\chi^2(5)=20.570$, $p=0.001$). As a result, during the explanation of the ANCOVA model, the Greenhouse-Geisser tests were used again.

In the first row of the within-subjects effect (*TotMedPc*), the significance value ($0.639 > 0.05$) shows that there were no significant differences between the Olympic Games after the influence of the covariates: none of the variables (interactions of the covariates with the *total medals won per capita per nation*) have significant interaction with the Olympic Games. The table of between-subjects effects (Table 6) shows that all variables remained significant in the

case of more dependent variables (Olympic Games). Effects sizes are higher (in parallel with the *F*-statistic values). In other words, *population* and *schooling* have the highest values (.518 and .361 respectively).

Tests of Within-Subjects Effects						
Effect	SS ^a	df	MS	F	p	partial η^2
<i>TotMedPc</i>	.492	2.506	.196	.764	.494	.013
<i>TotMedPc * TempCat</i>	.971	2.506	.388	1.510	.220	.026
<i>TotMedPc * LnPop</i>	.545	2.506	.217	.847	.453	.015
<i>TotMedPc * NorthSouth</i>	.638	2.506	.255	.991	.388	.017
<i>TotMedPc * Schooling</i>	.291	2.506	.116	.453	.681	.008
Error(<i>TotMedPc</i>)	36.028	140.323	.257			
Tests of Between-Subjects Effects						
Corrected Model	496.589 ^a	19	26.136	35.314	.000	.709
Intercept	6.197	1	6.197	8.373	.004	.030
Olympics	.202	3	.067	.091	.965	.001
LnPop	134.695	1	134.695	181.992	.000	.398
Olympics*LnPop	.047	3	.016	.021	.996	.000
NorthSouth	10.566	1	10.566	14.276	.000	.049
Olympics*NorthSouth	2.134	3	.711	.961	.411	.010
TempCat	22.703	1	22.703	30.675	.000	.100
Olympics*TempCat	1.063	3	.354	.479	.697	.005
Schooling	42.050	1	42.050	56.816	.000	.171
Olympics*Schooling	.100	3	.033	.045	.987	.000
Error	203.532	275	0.74			
Total	1050.37	295				
	4					
Corrected Total	700.121	294				

a. Calculated using SPSS default (Type III)

b.R Squared = ,709 (Adjusted R Squared = ,689)

Table 6. Repeated-measures ANCOVA for the four Olympic Games, testing the within-subjects and between-subjects effect of the Games. Covariates are the *population*, *schooling* and the categorical variables of *NorthSouth* and *temperature*; Greenhouse-Geisser statistics are used.

The normality tests for the model's residual variables did not show significant alteration from normal distribution.

Equality of Regression Slopes

The lower part of Table shows the tests for the equality of regression slopes. The results also coincide with those of previously demonstrated results: different Games do not make any difference ($p=.965$). *Population*, *NorthSouth*, *Temperature* category and *Schooling* do not differ significantly, as the p-values in the rows of *Olympics*LnPop*, *Olympics*NorthSouth*, *Olympics*TempCat* and *Olympics*Schooling* ($p=.996$, $p=.411$, $p=.697$ and $p=.987$ respectively) indicate. The estimated Adjusted R^2 is again similar to the Regression analysis ($=.689$). Further analysis of the between-subject effect in the last column shows that *population* has the highest effect (.398), followed by the *Schooling* (.171), *temperature category* (.100) and the *NorthSouth* variable (.045).

This section provided evidence that social factors are important, and *schooling* can be viewed as the second most important predictor of success, after the *size of the population*. Finally, the demographic and social hypotheses proved to be the strongest predictors. These results are valid for all the four Games examined.

Generalisation of the Schooling effect

The effect of mean years spent in school is valid in this data environment: in this study only those countries who won a minimum of two medals in the last four Olympics are present. However, this influence might be higher or lower if we consider e.g. those nations who won a minimum of one medal or three medals or more. The inclusion of those nations that did not win medals is not sound methodically as it alters the focus of the Games in which medals won is viewed as a key indicator of success: the main tables of success reflect medals won by countries. It might be assumed that nations that won medals in a minimum of two Games have a similar focus on the Olympic Games, therefore, the comparison of those countries provided acceptable results.⁶ Note: it is not proved in this paper that *schooling* is the panacea for Olympic success. However, the results of the study indicate that between the successful Olympic nations

⁶ The use of other statistics on this dataset is rather difficult: such as the Tobit regression as it also needs the normality of the dependent variable, which is not fulfilled with the inclusion of zero medals. Also, the Poisson regression is of no use in this study as it is created for countable dependent variable, but the dependent variable of this study is the *medal per capita*.

those are more successful, where education is more developed. The connection of *schooling* with wealth is not obvious. Statistically, the *GDP per capita* can explain the changing in *Schooling* variable by 37.6% (R^2). Furthermore, the GDP proved to be a weaker explanatory factor than *Schooling* and there are successful nations with Olympic traditions, which are not wealthy.

Discussion

According to the first aim, the Study 1 results appear to strengthen the previous as well as other reports that have claimed that the *population* size of the participating nations and *GDP per capita* are related to Olympic success (Bernard and Busse 2004; De Bosscher *et al.* 2003; Hoffmann, *et al.* 2004; Morton 2002). The results of Study 1 generally affirm a positive relationship between these two variables and Olympic success. However, our findings regarding the connection between *GDP per capita*, *population* and *total medals won* during the last four Olympic Games indicated that *population* surfaced as the more robust indicator for medals won as opposed to *GDP per capita*.

In this study, no difference was proven between the effects at any of the last four Olympic Games. Also, 56% of Olympic success is explained by the *population* size and the *GDP per capita* variables. However, the rest of the causes of success remain unknown. Nevertheless, in the past, other authors (Boudreau *et al.* 2014) disputed this finding, as they believed *GDP per capita* may not be a significant factor in determining success, as in places with lower standards of living (e.g. Russia or China) the decision makers may invest substantial funding for training, support and competition preparation in advance of an Olympic Games. Additionally, such countries may also commit to sending more athletes to the Olympic Games.

In this case, the more athletes that a country has competing, the greater its chance is to gain more medals (Feizabadi *et al.* 2013). Therefore, there is a significant measurable advantage to nations with larger general population sizes succeeding at the Olympic Games (Johnson and Ali 2000), if they send large(r) teams. Contrary to this finding, Johnson and Ali (2008) also stated that in many cases at mega events (e.g. at the Winter Olympic Games) small nations can outperform their larger competitors. Thus, sports specific, cultural, or environmental factors may drive success in specialist areas of sport as is found in the Winter Olympics. This conjecture needs further investigation.

Regarding the second aim, Study 2 added two variables, the geographic location (*North versus South*) and *yearly average temperature*. In this analysis four variables were included, as *temperature* of the participant countries, *population* size, *GDP per capita*, and finally a geographical location variable that was related to

whether ‘winning countries’ were in the *North or South*. Countries of the North and South were categorised being either part of Eurasia or North America; countries of the South were part of South America, Africa and Australia as well as the South Pacific region. This leads to the suggestion that future studies must consider the continental dimension from different perspectives such as using political, cultural and economic delineators as a differentiation tool as opposed to the equatorial or hemispheric line (Reuveny and Thompson, 2018). According to these analyses, there were no significant differences found between the *North and South* countries and the success of countries at the last four Olympic Games, however, the *temperature* category, *GDP per capita* and the *population* size had a significant effect on Olympic success and still did not differ by Olympic Games. The explanation power of this model increased though to 64%, thus strengthening the new model.

With respect to the third aim, Study 3’s analysis involved social factors from *HDI* (see the details of these datasets in the *Database* section). The analysis confirmed that the most influential factors of Olympic success were *population* size and *schooling* along with *NorthSouth* and *temperature*. This is the first developed model of specific demographic, social and geographical factors, which can explain more than two thirds (67.4%) of the variance in Olympic success (medals per capita won). Unsurprisingly perhaps, a warmer climate lends itself to sporting achievements in Summer Olympic Games and this may go some way to explaining this finding. Further studies must consider the Winter Olympic Games to identify trends in that event.

Policy Implications

Given that *schooling* is more developed in the wealthy *NorthSouth*, with more people taking part in structured education, opportunities are greater to participate in structured sport (e.g. the American school and collegiate system). *Schooling*, together with *population* explain 64% (adjusted $R^2=0.640$) of Olympic medals won. Additionally, *schooling* is equally important in all four Olympic Games, therefore categorised as one of the most important roles in Olympic success after the *population* size. The *Schooling* variable in the HDI database measures the number of years that students attend school. *Schooling* is also only a proxy, as discussed in the Introduction, however, it clearly indicates the social-cultural features of the countries. This finding therefore opens such a wide area of research that at this point of the research it would be premature to give more accurate policy recommendations.

Future Research and Limitations

Future Research

For recommendations of future work, a study of Olympic success needs to bring in the Winter Olympics to be truly valid as smaller nations often achieve high results e.g Norway, Switzerland, Austria. Those nations seem to be wealthy, have small populations and well-established health and school systems, AND seem environmentally suited to specialist sports. These latter factors should be considered to compare and perhaps align Summer Olympic success characteristics with those of the Winter Olympics. In fact, this study could also be expanded to include environmental variables alongside those of geography, demographics, schooling and sport culture. Increased schooling implies better resourcing, focused opportunities, and better nutritional status of the population due to improved socio-economic status; the investigation of these factors should also be the topic of further research.

One final topic that could be considered is whether other ‘Mega-Events’, such as the FIFA World Cup, would exhibit similar findings? It may be that different sports and sporting occasions attract culturally, politically, socially, or economically different populations and therefore might have alternative factors that impact on success.

Considering the current literature, these findings are important since they show that the widely hypothesised relationship between winning Olympic medals and *GDPpc* in reality is relatively modest. The outcome of this current study fully matched the results of another recent analysis, which revealed that the *GDPpc* of participating countries was a weak predictor of Olympic medals won in Athens 2004, Beijing 2008 (Boudreau, Kepner and Rondone 2014). Their study identified that the physical size of participating countries and national health care expenditures were stronger indices of winning medals.

Conclusion

The findings of this study are important since they show that the widely hypothesised relationship between winning Olympic medals and *GDPpc* is relatively modest. The outcome of this current study fully matched the results of another analysis, which revealed that the *GDPpc* of participating countries was a weak predictor of Olympic medals won in Athens 2004 and Beijing 2008 (Boudreau, Kepner and Rondone 2014). The conclusions of this work, based on the last four Olympics and supported by previous analyses (Boudreau et al. 2014), is that the role of the *GDP per capita* to predict Olympic medals wins

appears to be inflated in literature. The size of the *population* of participating nations accounted for more than twice as much of the variance in the winning of Olympic medals than the *GDP per capita* in Study 1. This points to the fact that success at the Olympic Games is not as much dependent on a nation's share of international wealth as it was presumed earlier in previous literature.

Schooling is the other most important factor predicting Olympic success. Furthermore, it is probably the *schooling* (social factor) that is the easiest to influence by a country and offers the shortest time for returns on investment as opposed to either *population* size, or *GDP per capita*. Also, *North versus South* geographic location and *temperature* cannot be changed.

In the interim, the scholastic message of this work is that, according to this analysis of the last four Olympics, *wealth* may be a feeble factor in determining which country wins medals at the Olympic Games. These findings deserve replication with data from future Olympics (i.e. Tokyo 2021).

REFERENCES

- Andreff, W. (2001). The correlation between economic underdevelopment and sport. *European Sport Management Quarterly*, 1(4), 251-279. <https://doi.org/10.1080/16184740108721902>
- Andreff, M., Andreff, W., & Poupaux, S. (2008). The economic determinants of Olympic performance: forecasting the medals to be won at the Beijing Games. *Revue d'économie politique*, 118, 135-169. doi: 10.3917/redp.182.0135.
- BBC Sport. (2018, 29 July). *Sport, Olympics, medals*. <http://www.bbc.co.uk/sport/olympics/2016/medals/countries>.
- BBC Sport. (2018, 12 September). *100 Women: do the Olympics have a gender gap?* <https://www.bbc.co.uk/news/world-41272613>.
- Bernard, A., & Busse, M. R. (2004). Who wins the Olympic Games: economic development and medal totals. *Review of Economics and Statistics*, 86(1), 413-417.
- Boudreau, J., Kepner, J. & Rondone, J. (2014, 4 August). *The miracle on thin ice: how a nation's GDP affects its Olympic performance*. Research Paper, Georgia Institute of Technology, School of Economics. <https://smartech.gatech.edu/handle/1853/51651>
- De Bosscher, V., De Knop, P. & Heyndels, B. (2003). Comparing relative sporting success among countries: create equal opportunities in sport. *Journal for Comparative Physical Education and Sport*, 3(3), 109-120.
- De Bosscher, V., Heyndels, B., De Knop, P., van Bottenburg, M., & Shibli, S. (2008). The paradox of measuring success of nations in elite sport. *Belgeo* <http://journals.openedition.org/belgeo/10303>; DOI: 10.4000/belgeo.10303
- Hong, F. (2006). *Sport, nationalism and orientation: The Asian Games*. Routledge.

- Feizabadi, M. S., Khabiri, M., & Hamidi, M. (2013). The relationship between the success of countries at the Guangzhou 2010 Summer Asian Games and demo-economic factors. *Procedia - Social and Behavioral Sciences*, 82, 369-374.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*, (4th Ed) SAGE Publications.
- Frey, M., Iraldo, F., & Melis, M. (2007). *The impact of wide-scale sport events on local development: an assessment of the XXth Torino Olympics through the sustainability report*. RSA, University of STD, 56-81.
- Green, M. (2005). Integrating macro- and meso-level approaches: a comparative analysis of elite sport development in Australia, Canada and the United Kingdom. *European Sport Management Quarterly*, 5(2), 143-166.
- Green, M., & Oakley, B. (2001). Elite sport development systems and playing to win: uniformity and diversity in international approaches. *Leisure Studies*, 20(4), 247-267.
- Tan, T-C, & Green, M. (2008). Analysing China's drive for Olympic success in 2008. *The International Journal of the History of Sport*, 25(3), 314-338.
- Halsey, L.G. (2009). The true success of nations at recent Olympic Games: comparing actual versus expected medal success. *Sport in Society*, 12(10), 1353-1368, DOI: 10.1080/17430430903204892
- Hoffmann, R., Ging, L. C., & Ramasamy, B. (2004). Olympic success and ASEAN countries: economic analysis and policy implications. *Journal of Sport Economics*, 262-276.
- Humphreys, B., Maresova, K., & Ruseski, J. (2012) *Institutional factors, sport policy, and individual sport participation: an international comparison*. Working Papers, University of Alberta, Department of Economics. https://ideas.repec.org/p/ris/albaec/2012_001.html
- International Olympic Committee. (2016). *Rio 2016: highlights of the Games*. <https://www.olympic.org/rio-2016>. [accessed 18 November 2018]
- Jayantha, K., & Ubayachandra, E.G. (2015). Going for the gold medals: factors affecting Olympic performance. *International Journal of Scientific and Research Publications*, 5(6), 1-6.
- Kuper, G. H., & Sterken, E. (2001). Olympic participation and performance since 1896. (University Gronigen). Available at SSRN: <https://ssrn.com/abstract=274295> <http://dx.doi.org/10.2139/ssrn.274295> [Accessed 5 August 2020].
- Johnson, D. K.N., & Ali, A. (2000). *Coming to play or coming to win: participation and success at the Olympic Games*. Working Paper, Department of Economics, Wellesley College, Massachusetts, 1-19. <https://pdfs.semanticscholar.org/3a5e/710f9890e5717349143ed895a3c22db91cf6.pdf> [accessed 21 November 2018]
- Lawrence, D. W. (2017). Sociodemographic profile of an Olympic team. *Public Health*, 148, 149-158.
- Lui, H.-K., & Suen, W. (2008). Men, money and medals: an econometric analysis of the Olympic Games. *Pacific Economic Review*, 13(1), 1-16. <https://doi.org/10.1111/j.1468-0106.2007.00386.x>
- Moosa, I., & Smith, L. (2004). Economic development indicators as determinants of medal winning at the Sydney Olympics: an extreme bounds analysis. *Australian Economic Papers* 9, 288-301.

- Morton, R. H. (2002). Who won the Sydney 2000 Olympics? An allometric approach. *The Statistician*, 5, 147-155.
- Noland, M., & Stahler, K. (2017). An old boys' club no more: pluralism in participation and performance at the Olympic Games. *Journal of Sport Economics*, 18(5), 506-536.
- O'Brien, K., & Onorato, A. (2018, 18 November). *Everything you need to know about marketing during the World Cup, 2018*. <https://www.dnnews.com/customer-experience/article/13034591/everything-you-need-to-know-about-marketing-during-the-world-cup>.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Reuveny, R., & Thompson, W. (2018). *North and south in the world political economy*. Wiley-Blackwell.
- Segura, R., & Birson, K. (2013). *The Human Development Index: How do Puerto Ricans measure up?* https://centopr.hunter.cuny.edu/sites/default/files/data_briefs/Centro_RB2013-03_Human_Dev_Index.pdf.
- Tcha, M., & Pershin, V. (2003). Reconsidering performance at the Summer Olympics and revealed comparative advantage, *Journal of Sport Economics*, 4(3), 216-339
- The World Bank (2017). *GDP per capita*. Retrieved from: <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD>
- The World Bank (2017). *World population data*. Retrieved from: <http://data.worldbank.org/indicator/SP.POP.TOTL>
- United Nations Development Programme (2017). *Human development research: Human Development Index*. <http://hdr.undp.org/en/content/human-development-index-hdi>
- Vagenas, G., & Vlachokyriakou, E. (2012). Olympic medals and demo-economic factors: novel predictors, the ex-host effect, the exact role of team size, and the "Population-GDP" model revisited. *Sport Management Review*, 15, 211-217.
- Vagenas, G., & Palaiothodorou, D. (2019). Climatic origin is unrelated to national Olympic success and specialization: an analysis of six successive games (1996-2016) using 12 dissimilar sports categories. *Sport in Society*, 22(12), 1961-1974.
- WeatherBase (2018, 29 July). <http://www.weatherbase.com>.

