Sports Medicine - Open

ORIGINAL RESEARCH ARTICLE Open Access

Modeling Performance in IRONMAN® 70.3 Age Group Triathletes

Mabliny Thuany[1], David Valero[2], Elias Villiger[3], Pedro Forte[4,5,6], Katja Weiss[2], Marilia Santos Andrade[7], Pantelis T. Nikolaidis[8], Ivan Cuk[9], Thomas Rosemann[3] and Beat Knechtle[2,10*]

Abstract

Background Individual factors related to performance in age group triathletes competing in different race distances have been explored in scientific literature. However, only a few studies have been conducted using machine learning (ML) predictive models to explore the importance of those individual factors. This study intended to build and analyze machine learning regression models that predict the performance of IRONMAN® 70.3 age group triathletes, considering sex, age, country of origin, and event location as predictive factors. A total of 823,464 finishers´ records (625,398 men and 198,066 women) of IRONMAN® 70.3 age group triathletes participating in 197 different events in 183 different locations between 2004 and 2020 were analyzed. The triathletes' sex, age, country of origin, event location and year, and race finish times were thus obtained and considered for the study. Four different ML regression models were built to predict the triathletes' race times from their age, sex, country of origin, and race location. The model with the best performance was then selected and further analyzed using model-agnostic interpretability tools to understand which factors would contribute most to the model predictions.

Results The Random Forest Regressor model obtained the best predictive score. This model's partial dependence plots indicated that men under 30 years, from Switzerland or Denmark, competing in IRONMAN®70.3 Austria/St. Polten, IRONMAN® 70.3 Switzerland, IRONMAN® 70.3 Sunshine Coast, and IRONMAN® 70.3 Busselton presented the best performance.

Conclusions Our results prove that ML models can be used to examine the complex, non-linear interactions between the factors that influence performance and gain insights that can help IRONMAN® 70.3 age group triathletes better plan their races.

Key Points

Performance in amateur IRONMAN® 70.3 is most strongly predicted by age, nationality and event location.
Machine learning models can be used to examine the complex, non-linear interactions between the factors that influence performance in endurance activities.

Keywords Machine learning, Performance, Endurance, Swimming, Cycling, Running

*Correspondence:
Beat Knechtle
beat.knechtle@hispeed.ch

Full list of author information is available at the end of the article

Background

The participation of non-professional athletes in endurance events such as triathlons has significantly increased in the last few years [1]. Considering the full-distance IRONMAN® triathlon—which consists of 3.8 km swimming, 90 km cycling, and 42.2 km running—North American and European countries were over-represented among the best athletes competing in IRONMAN® Hawaii as the IRONMAN® World Championship [2, 3]. The fastest women and men also originated from the USA [2, 4]. Regarding performance, IRONMAN® athletes' performance involves many factors related to individual and environmental characteristics while also being dependent on the distance of the competition [5, 6].

As outdoor sports practitioners, IRONMAN® athletes are exposed to different constraints during training and competition, including equipment, pacing strategy, aerodynamics, and contextual features [7]. Different environmental characteristics (e.g., water temperature, wind speed, altimetry, temperature, etc.) directly influence the athletes' performance [8, 9]. Water characteristics such as temperature have been studied for swimming related to physiological adaptations and performance outcomes [10]. Weather factors (such as wind, temperature, humidity, and altitude) will affect the fluid dynamics (air and water) and, with it, drag [11, 12]. The higher the drag, the higher the resistance acting on the athletes. In addition to the many factors that can improve an athlete's aero- and hydrodynamics, in IRONMAN® swimming split, athletes can also draft behind or on the hip of a fellow swimmer to save a massive amount of energy [13].

For cycling, the race course is an important variable to consider, given that technical features can differ based on natural and physical challenges [14–16]. Wind and rain negatively correlate with drafting and positional resources [7]. For running, temperature, barometric pressure, humidity and sunshine duration were negatively correlated to the endurance runners' performance at different competitive levels [17–19]; also, running technique plays an important role in the runners' performance [14]. These different environmental characteristics seem to influence the performance of endurance athletes differently.

Among the individual factors related to performance, age and sex have been previously investigated. For athletes competing in the IRONMAN® 70.3 (1.9 km swimming, 90 km cycling, 21.1 km running), the performance decline in swimming tended to start in early young athletes, while the decline in running performance started at 26 to 28 years in both sexes. For cycling, the age-related performance decline started at 34 years for men and at 35 years for women [20–22]. These characteristics shed light on sex differences [23, 24], in which women present a lower performance baseline than men [23]. To minimize the decrease in performance, triathletes' tend to select competitions based on environmental conditions (i.e. lower temperature, humidity and higher barometric pressure will result in higher drag). Altitude also seems to play an important role in performance. Every detail counts where marginal gains may be decisive for the finishing time [7, 12].

Despite the vast scientific knowledge regarding the different variables that are related to the performance of amateur (age group or recreational) triathletes, few studies have been developed using machine learning models considering the environmental factors as predictors for performance. Although amateur athletes have less time to dedicate to training since their professional activities are not directly related to the sport [25], training characteristics and competition participation are still important factors in maximizing performance. Runners should consider training plans, internal and external load monitoring, recovery strategies, nutrition, hydration, and sleep [5].

Two different groups of athletes compete in triathlon races: professional and amateur (age group or recreational) triathletes. A study reported that the fastest race times for professional IRONMAN® 70.3 triathletes would be achieved in the Ironman® 70.3 World Championship [26]. Although the World Championship is the most important race for professional triathletes, we do not know the fastest race courses and the fastest athletes for amateur (age group or recreational) IRONMAN® 70.3 triathletes. This study intended to build machine learning regression models to predict the performance of IRONMAN® 70.3 age group triathletes, considering the triathlete's sex, age, country of origin, and event location, and to further analyze the underlying logic of how each of these factors contributes to the predictions. Since amateur (age group or recreational) triathletes also compete in the IRONMAN® 70.3 World Championship alongside professional triathletes, we hypothesized that the IRONMAN® 70.3 World Championship would also be the fastest racecourse for amateur (age group or recreational) triathletes.

Methods

Ethics Approval

The study was approved by the Institutional Review Board of Kanton St. Gallen, Switzerland. The Institutional Review Board of Kanton St. Gallen, Switzerland waived the need for informed consent of the participants as the study involved the analysis of publicly available data Kantonale Ethikkommission St. Gallen [EKSG]. The study was conducted following recognized ethical standards according to the Declaration of Helsinki adopted in 1964 and revised in 2013.

Data Set and Data Preparation

Race data were downloaded from the official IRONMAN® 70.3 website (www.ironman.com) using a Python script. The athletes' sex, age, country of origin, event location and year, and times for swimming, running, cycling, and transitioning were thus obtained. We analyzed age group finishers' race data of all IRONMAN® 70.3 races recorded on the IRONMAN® 70.3 website between 2004 and 2020. Data considered for this analysis were the race finish time, the triathlete's age, sex, and country of origin, and the race location.

Statistical Analysis and Machine Learning Modeling

Descriptive statistics are presented using mean, standard deviation, frequencies, and percentages. Four machine-learning regression models were built and compared using four different algorithms: Random Forest Regressor, XG Boost Regressor, Cat Boost Regressor, and Decision Tree Regressor. The mean absolute error (MAE) and R-squared (R^2) accuracy metrics were calculated for each algorithm, as were the features' relative importance. The MAE represents the mean of the absolute values of the individual prediction errors over all instances in the test set [27], in which higher values mean higher prediction errors. R^2 is a statistical measure representing the proportion of the variance of the dependent variable (predictions) that can be explained by the model's independent variables (predictors). The feature's relative importance represents a score for each feature in a specific model. Higher values correspond to features that present higher importance in predicting the target variable. The model output, or predicted variable, was the finish time in seconds, while sex, age group, country of origin, and event location were used as predictors. All predictors were categorical variables; hence they were encoded into numerical values before they could be fit into the models. Sex was encoded as 0 for women and 1 for men. Age group was encoded using the lower value in each group (18 for age group 18–24 years, 25 for 25–29 years, etc.). At the same time, the event location and country were sorted by the number of records and then encoded based on their position in the resulting list (rank encoding). The full dataset was used to train the models (no separate evaluation sample) since we were more interested in the model's interpretability than in its predictive power. All data processing and analysis were done using Python (https://www.python.org) and a Google Colab (Jupyter) notebook (https://colab.research.google.com/notebook).

Results

A total of 823,464 IRONMAN® 70.3 finishers' records (625,398 men and 198,066 women) from 197 different events across 183 different race locations between 2004 and 2020 were sampled. Figure 1 presents histograms for the finish time in both sexes, with men presented lower mean values compared to women (05 h:46 min:53 s vs. 06 h:16 min:18s, respectively).

Table 1 presents the results of the predictive models, considering the different algorithms. All models rate age as the least important factor in their predictions. Sex,

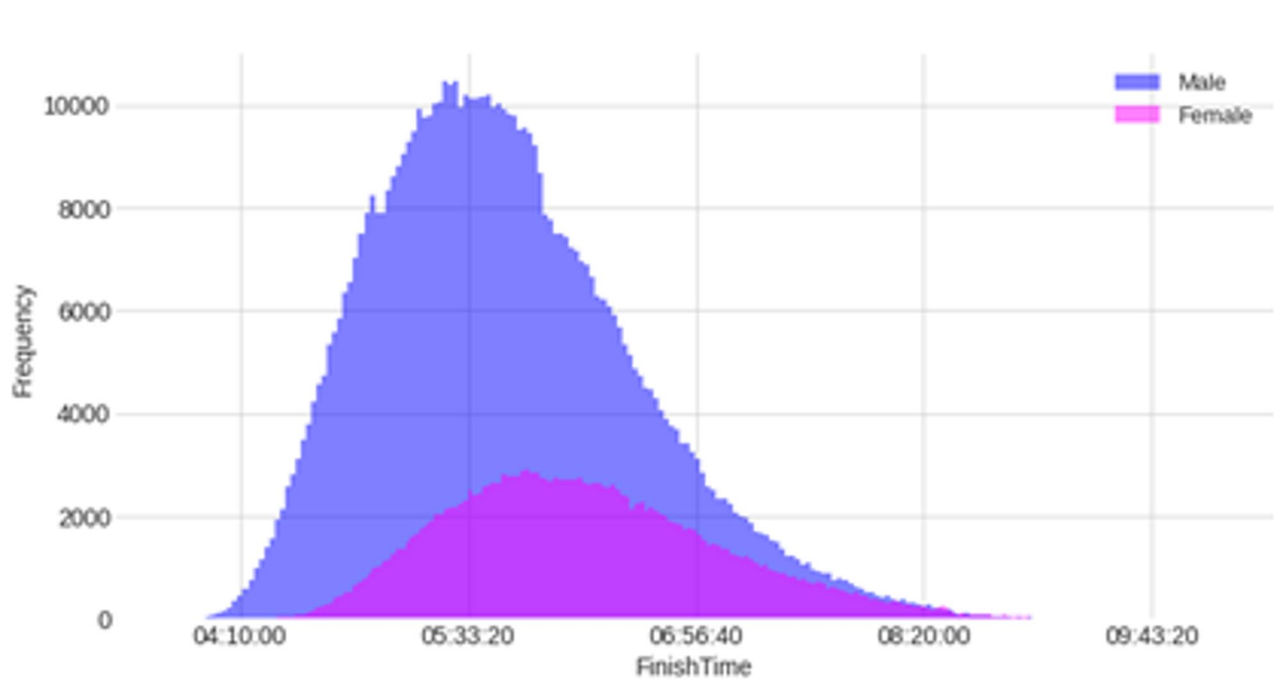

Fig. 1 Histograms of IRONMAN® 70.3 amateur finish times. Times are presented in h: min: s

Table 1 Importance of metrics and features in the four predictive models

	Random forest regressor	XG boost regressor	Cat boost regressor	Decision tree regressor
Feature	Importance			
Event Location	0.33	0.13	0.35	0.31
Country	0.31	0.15	0.40	0.32
Sex	0.22	0.63	0.18	0.25
Age group	0.14	0.09	0.07	0.13
	Algorithm parameters			
Nbr. Estimators	400	2000	2000	1(max_depth = 20)
Training time	5 min 16s	2 min 3s	3 min 19s	1.15s
	Evaluation metrics			
R^2	0.27	0.25	0.24	0.26
MAE	2015.5	2044.9	2076.7	2030.9

Nbr number, *MAE* mean absolute error

country of origin, and event location present different importance according to the algorithm used in each case. Based on the R^2 score, the best model is the Random Forest Regressor, with a value of 0.27. It is acknowledged that this is quite a low score for a predictive model. But still, there is a 27% variability in the models' output that can be explained through the four factors considered.

All models exhibit a low R^2 score (calculated over the training sample), indicating other variables should be considered to improve the predicting power of the models. The best score corresponds to the Random Forest model ($R^2 = 0.27$), while the MAE shows that the difference between the best and worst model predictions is about a minute. The partial dependence plots (PDP) show how the different values of each predictor (on the x-axis) influence the model predictions (on the y-axis, in seconds) on average. The second chart, the actual predictions plot, shows the model predictions' distribution over the same sample it has been trained with. A close inspection of traditional descriptive statistics in the dataset showed very similar trends and patterns. Considering the first factor, sex, the PDP plot shows the model deducts about 1800 s from a men prediction for a women prediction (Fig. 2).

The actual predictions chart shows the distributions through boxplots with the median values super-imposed (above) (Fig. 3).

Considering the age group factor, the model only starts taking this variable into account from the age group 30 (30 to 34 years), increasing linear relationship that accelerates with age (Fig. 4).

The actual predictions plot shows the distribution of the number of records along with the predictions for each age group. The model's best prediction (best median) is achieved by 25–29 years (Fig. 5).

Regarding the country, the PDP plot indicates that the model places very little importance on this factor from the country with ID 50. Indeed, it tells us also the best country (meaning the one making the most negative contribution to the race finish time prediction) is within the top 20 (Fig. 6).

The actual predictions chart shows only the top 20 countries by the number of records. The best median *FinishTime* predictions are given to the countries with ID 12 and 19 (Switzerland and Denmark) with a times of 19525.40 and 19359.24 s, respectively (Fig. 7).

The event location factor includes 183 different locations. With the available data, there was no possibility to assess to what extent one location would be different between events on different dates, so data were

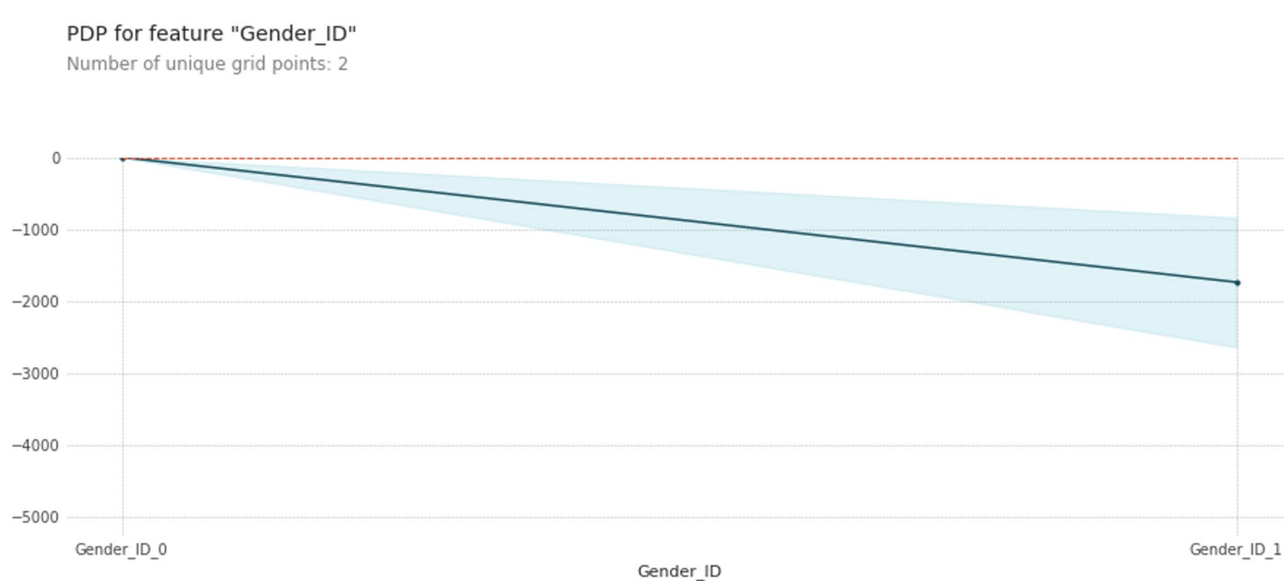

Fig. 2 Partial dependence plot (PDP) for sexes. Sex ID 0 = women, Sex ID 1 = men. Y-axis graduated in seconds

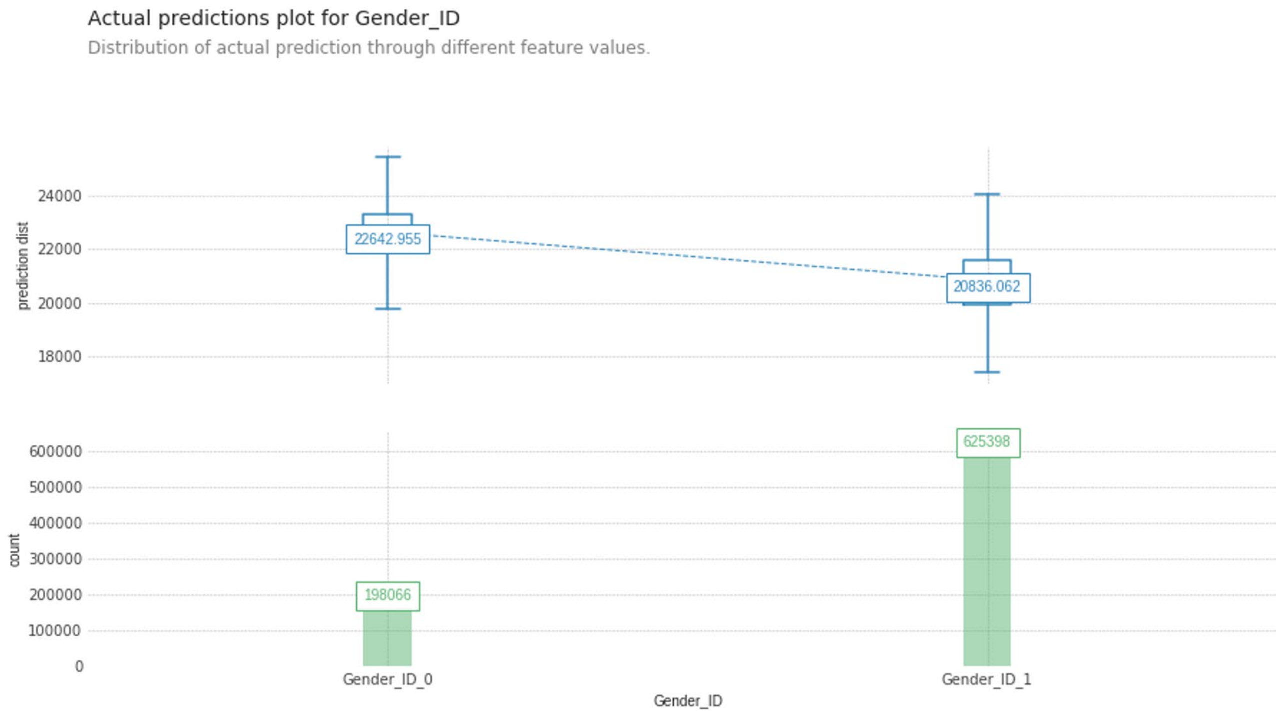

Fig. 3 Prediction distribution plot for sexes

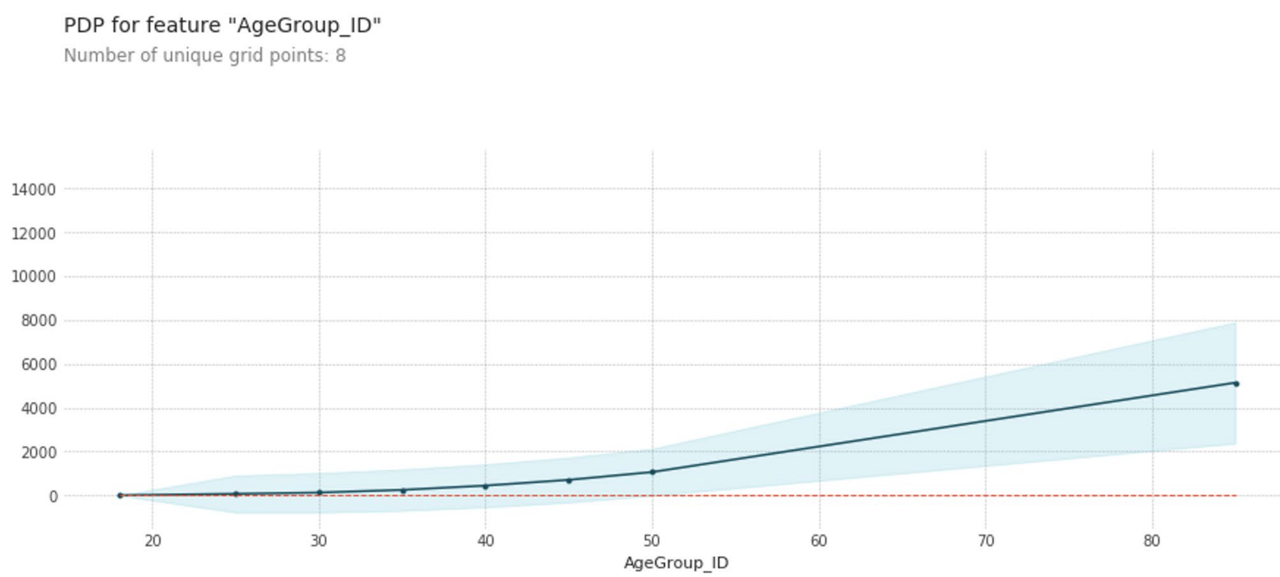

Fig. 4 Partial dependence plot (PDP) for age groups. Y-axis graduated in seconds

aggregated by actual (physical) location irrespective of the event date (Fig. 8).

The PDP plot shows three minimums within the first 20 locations. The actual predictions chart helps us identify the event locations with the fastest predicted times. These have IDs 7, 13, 18, and 19 (attending to the median value of predictions), corresponding to IRONMAN® 70.3 Austria/St. Polten (15056 records), IRONMAN® 70.3 Switzerland (13468), IRONMAN® 70.3 Sunshine Coast (10777) and IRONMAN® 70.3 Busselton (10507) (Fig. 9).

Discussion

The purpose of this study was to build and analyze a machine learning regression model that would predict the performance of amateur IRONMAN® 70.3 triathletes (in the form of the race finish time), considering the role of sex, age, country, and event location as predictive

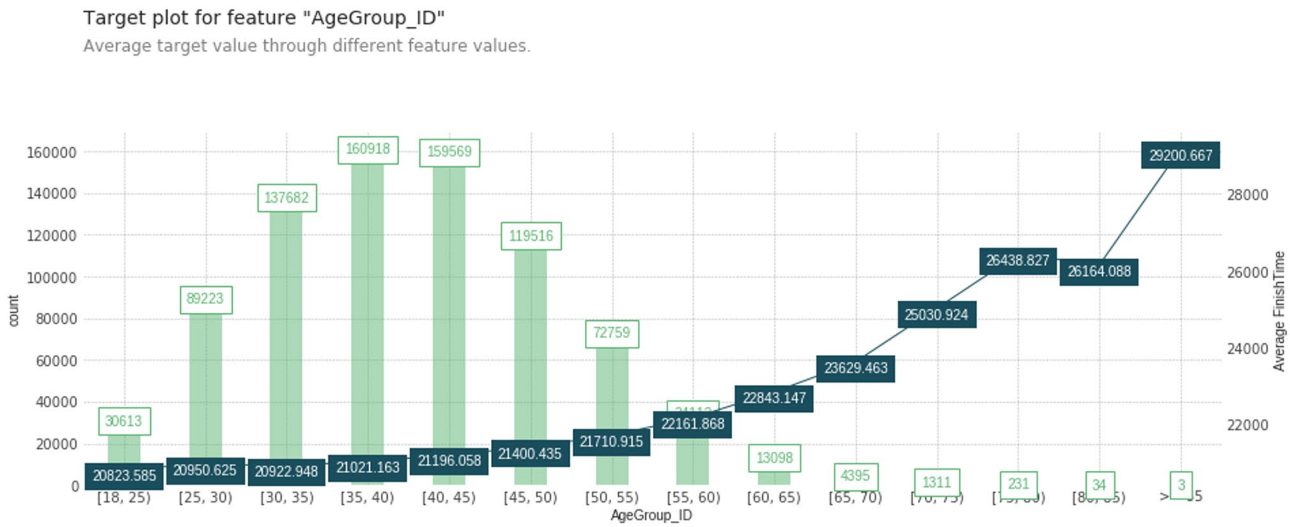

Fig. 5 Prediction (top chart) and count distribution (bottom chart) for age groups

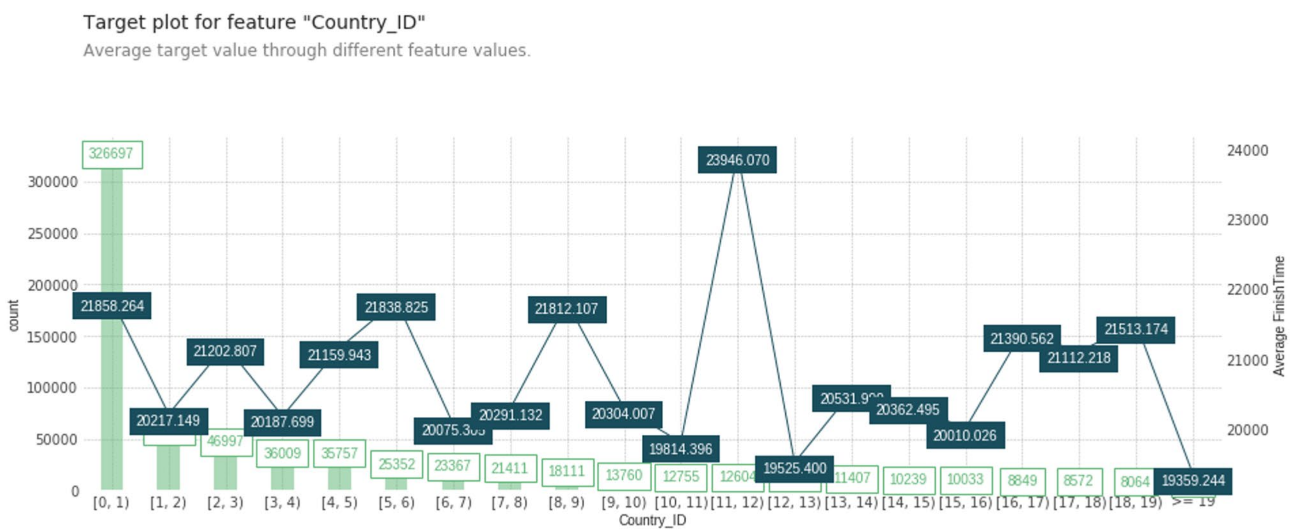

Fig. 6 Partial dependence plot (PDP) for countries. Y-axis graduated in seconds

factors. Based on recent findings for professional IRONMAN® 70.3 triathletes, we hypothesized that the best race times in amateur IRONMAN® 70.3 triathletes would also be achieved in the IRONMAN® 70.3 World Championship.

Firstly, our results showed that the fastest courses were IRONMAN® 70.3 Austria and IRONMAN® 70.3 Switzerland. Interestingly, our hypothesis that amateur IRONMAN® 70.3 triathletes would achieve their fastest race times in the IRONMAN® 70.3 World Championship could not be confirmed. Adapting to the climate, the reduced time of travel, as well as the event hosting effect increases the possibility of attending events and maintaining motivation [28, 29]. In addition, the anxiety status, commonly associated with very big events, such as World Championships, can be seen as a phenomenon that hampers performance [30]. Future studies should investigate the relationship between the availability of events, the motivation to take part in these events, and the performance at country levels. Because amateur practitioners are prone to higher trainability, the performance in amateur IRONMAN® 70.3 may possibly be better explained by training variables than environmental factors. Therefore, age and sex are prone to have a higher contribution than country and event location. However, it is important to highlight that training monitoring variables were impossible to obtain in this research, so future research with training session data is encouraged.

Secondly, Swiss and Danish athletes achieved the best IRONMAN® 70.3 race times (around 5 hr 28 min). Few studies were developed to understand the relationship between country of residence, geographical locations/

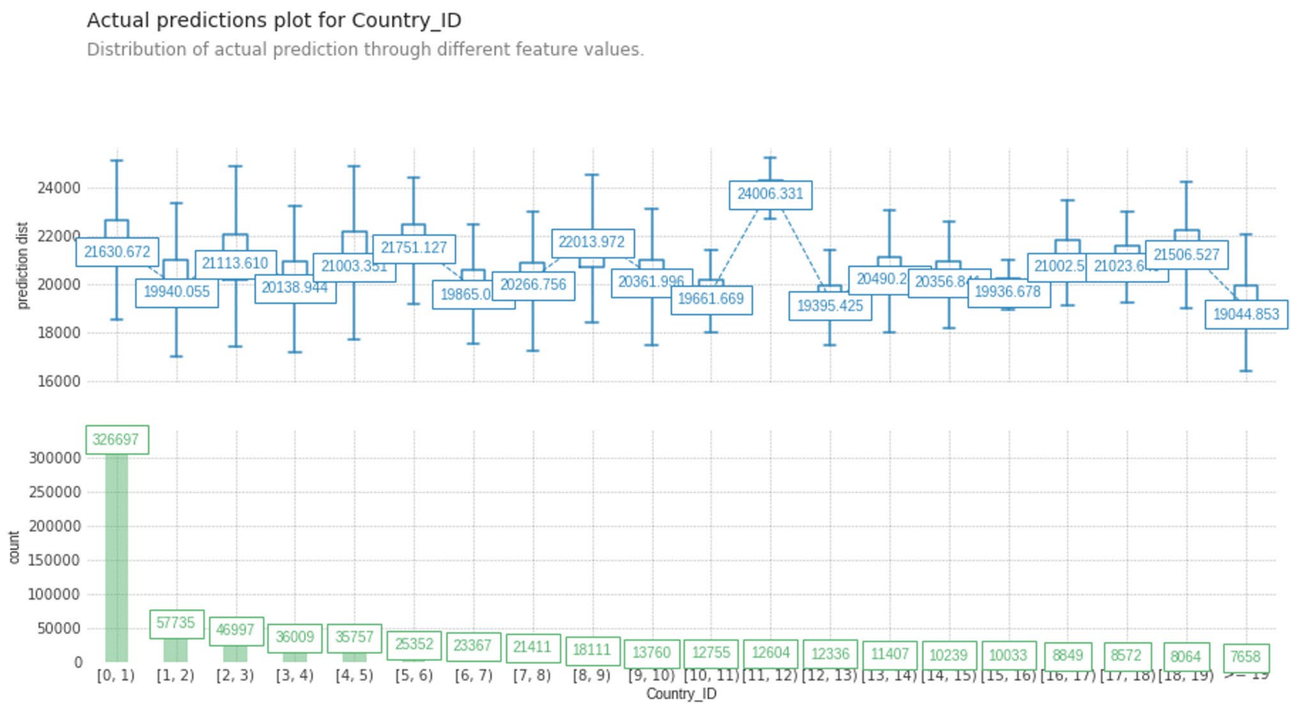

Fig. 7 Prediction (top chart) and count distribution (bottom chart) for countries

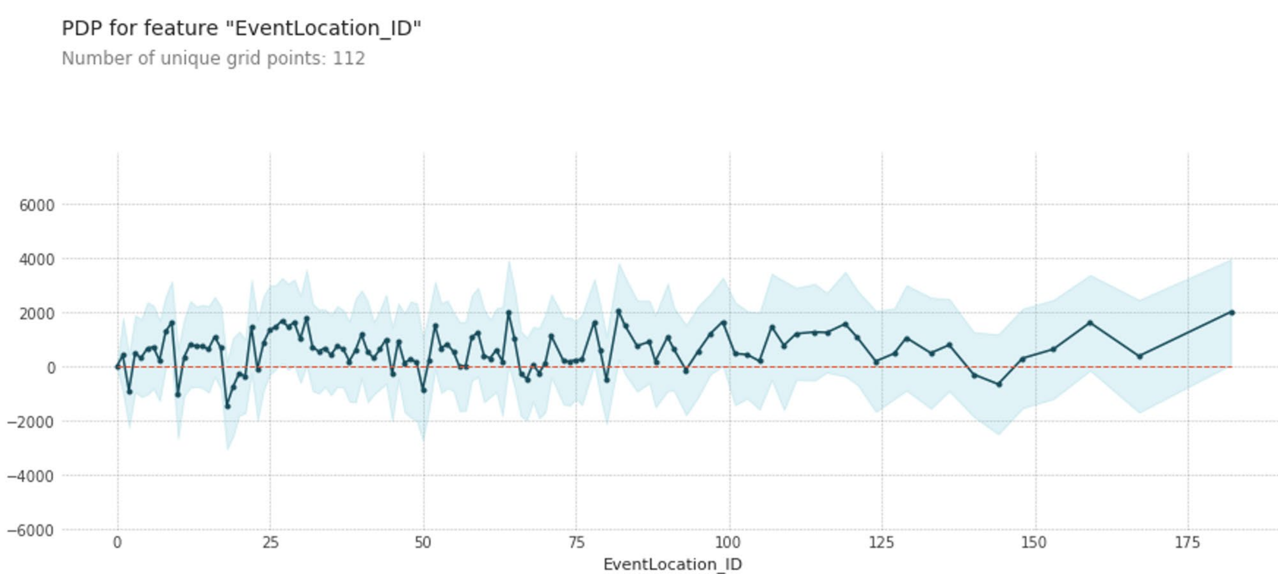

Fig. 8 Partial dependence plot (PDP) for event location. Y-axis graduated in seconds

characteristics, and triathlon performance [31, 32]. Most of the studies developed in this context considered the influence of the country of residence on participation in the event [33]. For example, for 'IRONMAN® Hawaii', a higher portion of the finalists are from the USA, Germany, and Switzerland [2, 34], while participants in 'IRONMAN® Switzerland' are from Europe, especially Switzerland and Germany [2].

Factors that explain the best results achieved by Swiss and Danish athletes are not clarified in the literature. Since sports performance is a complex phenomenon, it is possible that a combination of sports culture, environmental conditions (including social supports, such as terrain and climate), technological development, and the wealth of the country are associated with these results [2, 28]. Since triathlon is a highly specialized sport, economic status is also considered an important predictor of participation [35].

Thirdly, our predictive algorithm showed that the best performance will be achieved by men aged under 30 years

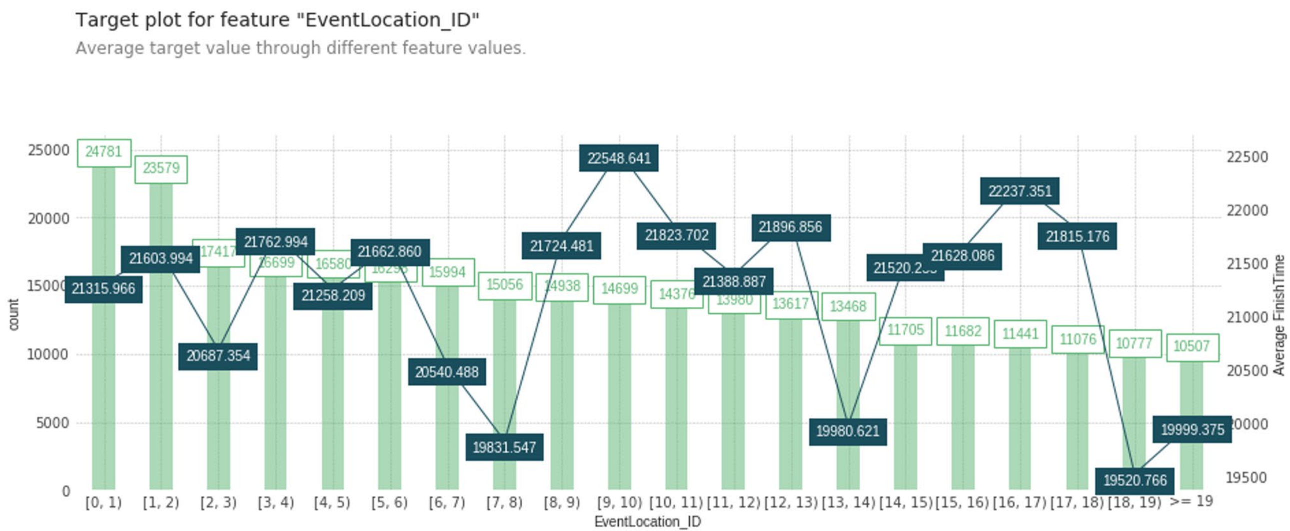

Fig. 9 Prediction (top chart) and count distribution (bottom chart) for event location

(i.e., 25–29 years). In summary, sex performance differences are around 12–18% [36], with differences between disciplines, and an overall superior performance by men [37]. Anthropometric differences (e.g. muscle mass, body fat, skin-fold thicknesses, the circumference of the upper arm) [38], physiological indicators (e.g., VO_2max is about 10% lower in women athletes, ventilatory threshold, and movement economy) [23, 37], and social aspects (e.g., social pressure, financial and emotional support) are all related to the differences in performance between sexes. In addition, nutritional characteristics, training load, planning, and relative energy deficiency need to be investigated to increase the scientific evidence regarding women triathletes and performance characteristics [39].

Our results showed that the best performance would be achieved by athletes under 30 years, according to previous findings that showed a performance decline after 40 years [22]. The age of peak performance was previously investigated and seemed to differ considering the race distance and athletes' characteristics. For the Olympic distance triathlon, the age of peak performance was between 26 and 27 for women and men, respectively, while for IRONMAN® athletes, the age of peak performance was between 34 and 35 years for women and men [40]. Therefore, it is reasonable to assume that older IRONMAN® athletes should pay more attention to environmental factors and the competition location.

The present study did not focus on understanding the age of peak performance. Nonetheless, the results present important practical applications for young athletes who can use their experience in endurance activities to increase the competition race distance over time and reach higher performance levels. However, transferable skills (e.g., from short to long-distance events) were not studied for different triathletes' distance events. A similar idea was investigated using cross-training methods [41, 42], showing positive effects between cycling training and running performance cross-transfer [43]. Future investigations should consider understanding transferable skills among triathlon distances.

The lack of qualitative information regarding the motivational characteristics and natural environment variables are important limitations of the present study. The analysis was not adjusted for possible confounders, such as weather characteristics, training background, racecourse characteristics (Table 2), and psychological and physiological variables.

Previous literature presented the role of both individual and environmental factors related to performance in triathlon [25, 44]. Besides the performance expressed through the finishing times, geographical characteristics can also influence each discipline differently. Despite these limitations, the present study indicates the importance of considering a more holistic approach during the preparation of IRONMAN® 70.3 amateur athletes. Furthermore, the study provides practical applications for professionals working with triathletes about the age of peak performance, nationality's role, and the location where a race takes place. For instance, considering the best performance shown in the age group 25–29 and the subsequent rates of performance decline would aid in setting long-term training goals.

Conclusions

Using a Random Forest Regressor ML model over a sample of over 800,000 IRONMAN® 70.3 race records in 16 years, our results indicate that age group 25–29 is the fastest, that Swiss and Danish are the fastest IRONMAN® 70.3 athletes, and that IRONMAN® 70.3 Austria/St. Pölten, IRONMAN® 70.3 Switzerland, IRONMAN®

Table 2 Specific descriptions of IRONMAN® 70.3 race courses for swimming, cycling, and running with average temperatures, number of laps and changes in altitudes per lap and overall

IRONMAN® 70.3					
Zell am See	August	19 °C	Lake, 1 lap	1 lap, change in altitude of 1270 m per lap, overall change in altitude of 1270 m	2 laps, change in altitude of 26 m per lap, overall change in altitude of 52 m
European Championship Elsinore	June	21 °C	Bay,1 lap	1 lap, change in altitude of 600 m per lap, overall change in altitude of 600 m	3.5 laps, change in altitude of 28.6 m per lap, overall change in altitude of 100 m
Greece Costa Navarino	October	23 °C	Ocean, 1 lap	2 laps, change in altitude of 540 m per lap, overall change in altitude of 1080 m	3 laps, change in altitude of 66.7 m per lap, overall change in altitude of 200 m
Indian Wells La Quinta	December	18 °C	Lake, 1 lap	1 lap, change in altitude of 161 m per lap, overall change in altitude of 161 m	2 laps, change in altitude of 57.7 m per lap, overall change in altitude of 115 m
Middle East Championship Bahrain	December	22 °C	Bay, 1 lap	1 lap, change in altitude of 300 m per lap, overall change in altitude of 300 m	3 laps, change in altitude of 24.3 m per lap, overall change in altitude 73 m
Texas	April	21 °C	Bay, 1 lap	1 lap, change in altitude of 45 m per lap, overall change in altitude of 45 m	3 laps, change in altitude of 10.7 m per lap, overall change in altitude of 32 m
Tallinn	August	17 °C	Lake, 1 lap	1 lap, change in altitude of 300 m per lap, overall change in altitude of 300 m	2 laps, change in altitude of 57.7 m per lap, overall change in altitude of 115 m
Western Sydney	September	23 °C	Lake, 1 lap	2 laps, change in altitude of 165 m per lap, overall change in altitude of 330 m	1 lap, change in altitude of 25 m per lap, overall change in altitude of 25 m
Maceio	August	28 °C	Ocean, 1 lap	1 lap, change in altitude of 200 m per lap, overall change in altitude of 200 m	3 laps, change in altitude of 23.3 m per lap, overall change in altitude of 70 m
Liuzhou	September	28 °C	River, 1 lap	2 laps, change in altitude of 168,5 m per lap, overall change in altitude of 337 m	2 laps, change in altitude of 171 m per lap, overall change in altitude of 342 m
Luxembourg	June	18 °C	River, 1 lap	1 lap, change in altitude of 580 m per lap, overall change in altitude of 580 m	2.5 laps, change in altitude of 40 m per lap, overall change in altitude of 100 m
Marbella	May	21 °C	Ocean, 1 lap	1 lap, change in altitude of 1400 m per lap, overall change in altitude of 1400 m	2 laps, change in altitude of 25 m per lap, overall change in altitude of 50 m
Vichy	August	28 °C	Lake, 1 lap	1 lap, change in altitude of 1000 m per lap, overall change in altitude of 1000 m	2 laps, altitude difference of 50 m per lap, overall change in altitude of 100 m
Gdynia	August	25 °C	Bay, 1 lap	1 lap, change in altitude of 1860 m per lap, overall change in altitude of 1860 m	2 laps, change in altitude of 70 m per lap, overall change in altitude of 340 m
Mallorca	May	25 °C	Ocean, 1 lap	1 lap, change in altitude of 850 m per lap, overall change in altitude of 850 m	3 laps, change in altitude of 20 m per lap, overall change in altitude of 60 m
Panama	March	30 °C	Ocean, 1 lap	3 laps	3 laps
Augusta	September	27 °C	River, 1 lap	1 lap	2 laps
North Carolina	October	18 °C	Ocean, 1 lap	1 lap	1 lap
Dubai	March	24 °C	Ocean, 1 lap	1 lap, change in altitude of 87 m per lap, overall change in altitude of 87 m	1.5 laps, change in altitude of 5 m per lap, overall change in altitude of 7,5 m
Sao Paulo	September	25 °C	Reservoir, 1 lap	2 laps, change in altitude of 20 m per lap, overall change in altitude of 40 m	3 laps, change in altitude of 15 m per lap, overall change in altitude of 45 m
Alagoas	August	28 °C	Ocean, 1 lap	1 lap, change in altitude of 100 m per lap, overall change in altitude of 100 m	3 laps, change in altitude of 20 m per lap, overall change in altitude of 60 m
Emilia Romagna	September	25 °C	Ocean, 1 lap	1 lap, change in altitude of 185 m per lap, overall change in altitude of 185 m	3 laps, change in altitude of 5 m per lap, overall change in altitude of 15 m
Pays D'Aix	May	20 °C	Lake, 1 lap	1 lap, change in altitude of 390 m per lap, overall change in altitude of 390 m	3 laps, change in altitude of 15 m per lap, at an overall change in altitude of 45 m
Sunshine Coast	September	26 °C	Ocean, 1 lap	2 laps, change in altitude of 35 m per lap, overall change in altitude of 70 m	2 laps, change in altitude of 20 m per lap, overall change in altitude of 40 m
Geelong	March	19 °C	Bay, 1 lap	2 laps, change in altitude of 75 m per lap, overall change in altitude of 150 m	2,5 laps, change in altitude of 25 m per lap, overall change in altitude of 67 m
Steelhead	June	20 °C	Lake, 1 lap	1 lap	2 laps

Table 2 (continued)

IRONMAN® 70.3					
Turkey	November	24 °C	Ocean, 1 lap	1 lap, change in altitude of 20 m per lap, overall change in altitude of 40 m	3 laps, change in altitude of 5 m per lap, overall change in altitude of 15 m
California	April	17 °C	Ocean, 1 lap	1 lap, change in altitude of 220 m per lap, overall change in altitude of 220 m	2 laps, change in altitude of 10 m per lap, overall change in altitude of 20 m
Astana	June	25 °C	River, 1 lap	1 lap, change in altitude of 20 m per lap, overall change in altitude of 20 m	2 laps, change in altitude of 5 m per lap, overall change in altitude of 10 m
Florianopolis	April	21 °C	Ocean, 1 lap	1 lap	3 laps

70.3 Sunshine Coast and IRONMAN® 70.3 Busselton are the fastest IRONMAN® 70.3 event locations. Future studies should consider the interaction between individual and environmental factors to better understand the factors that lead to the success of Switzerland and Denmark.

Abbreviations

ML Machine learning
MAE Mean absolute error
PDP Partial dependence plots

Acknowledgements

Not applicable.

Author Contributions

MT drafted the manuscript, DV performed the statistical analysis and prepared methods and results, EV obtained the data, PF, KW, MSA, PTN, IC, TR and BK helped in drafting the final version. All authors read and approved the final manuscript.

Funding

No funding.

Data Availability

Race data were downloaded from the official IRONMAN® website (www.ironman.com). The datasets used and/or analyzed during the study are available from the corresponding author on reasonable request.

Declarations

Ethics Approval and Consent to Participate

This study was approved by the Institutional Review Board of Kanton St. Gallen, Switzerland, with a waiver of the requirement for informed consent of the participants as the study involved the analysis of publicly available data (EKSG 01/06/2010). The study was conducted in accordance with recognized ethical standards according to the Declaration of Helsinki adopted in 1964 and revised in 2013.

Consent for Publication

Not applicable.

Competing Interests

Mabliny Thuany, David Valero, Elias Villiger, Pedro Forte, Katja Weiss, Marilia Santos Andrade, Pantelis T. Nikolaidis, Ivan Cuk, Thomas Rosemann, and Beat Knechtle declare that they have no competing interests. This includes financial interests as well as any professional, personal, or other relationships or beliefs that could be perceived to influence the work reported in this manuscript.

Author details

[1]Department of Physical Education, State University of Para, Pará, Brazil
[2]Ultra Sports Science Foundation, Pierre-Benite, France
[3]Institute of Primary Care, University Hospital Zurich, Zurich, Switzerland
[4]CI-ISCE, Higher Institute of Educational Sciences of the Douro, Penafiel, Portugal
[5]Instituto Politécnico de Bragança, Bragança, Portugal
[6]Research Center in Sports, Health and Human Development, Covilhã, Portugal
[7]Department of Physiology, Federal University of Sao Paulo, Sao Paulo, Brazil
[8]School of Health and Caring Sciences, University of West Attica, Athens, Greece
[9]Faculty of Sport and Physical Education, University of Belgrade, Belgrade, Serbia
[10]Medbase St. Gallen Am Vadianplatz, Vadianstrasse 26, 9001 St. Gallen, Switzerland

Received: 19 March 2024 / Accepted: 9 November 2025
Published online: 25 November 2025

References

1. Shoak MA, Knechtle B, Knechtle P, Rüst CA, Rosemann T, Lepers R. Participation and performance trends in ultracycling. Open Access J Sports Med. 2013;4:41–51. https://doi.org/10.2147/oajsm.s40142. Epub 2014/01/01.
2. Dähler P, Rüst CA, Rosemann T, Lepers R, Knechtle B. Nation related participation and performance trends in 'Ironman hawaii' from 1985 to 2012. BMC Sports Sci Med Rehabil. 2014;6:16. https://doi.org/10.1186/2052-1847-6-16. Epub 20140415.
3. Stiefel M, Knechtle B, Rüst C, Rosemann T. Analysis of performances at the 'Ironman Hawaii triathlon' and its qualifier events with respect to nationality. J Sci Cycl. 2013;2(2):27–34.
4. Knechtle B, Rosemann T, Lepers R, Rüst CA. A comparison of performance of deca iron and triple deca iron ultra-triathletes. SpringerPlus. 2014;3(1):461. https://doi.org/10.1186/2193-1801-3-461.
5. Etxebarria N, Mujika I, Pyne DB. Training and competition readiness in triathlon. Sports. 2019. https://doi.org/10.3390/sports7050101. Epub 20190429.
6. Ofoghi B, Zeleznikow J, Macmahon C, Rehula J, Dwyer DB. Performance analysis and prediction in triathlon. J Sports Sci. 2016;34(7):607–12. https://doi.org/10.1080/02640414.2015.1065341.
7. Phillips KE, Hopkins WG. Determinants of cycling performance: a review of the dimensions and features regulating performance in elite cycling competitions. Sports Med - Open. 2020;6(1):23. https://doi.org/10.1186/s40798-020-00252-z.
8. Baillot M, Hue O. Hydration and thermoregulation during a half-ironman performed in tropical climate. J Sports Sci Med. 2015;14(2):263–8. Epub 20150508.
9. Del Coso J, González C, Abian-Vicen J, Salinero Martín JJ, Soriano L, Areces F, et al. Relationship between physiological parameters and performance during a half-ironman triathlon in the heat. J Sports Sci. 2014;32(18):1680–7. https://doi.org/10.1080/02640414.2014.915425. Epub 20140513.
10. Baldassarre R, Bonifazi M, Zamparo P, Piacentini MF. Characteristics and challenges of Open-Water swimming performance: A review. Int J Sports Physiol Perform. 2017;12(10):1275–84. https://doi.org/10.1123/ijspp.2017-0230.

11. Forte P, Morais JE, Neiva P, Barbosa H, Marinho TM. The drag crisis phenomenon on an elite road Cyclist—A preliminary numerical simulations analysis in the aero position at different speeds. Int J Environ Res Public Health. 2020;17(14):5003.
12. Knechtle B, Forte P, Weiss K, Cuk I, Nikolaidis PT, Sousa CV, et al. Biophysical characterization of the first ultra-cyclist in the world to break the 1,000 Km barrier in 24-h non-stop road cycling: A case report. Front Cardiovasc Med. 2022. https://doi.org/10.3389/fcvm.2022.990382.
13. Janssen M, Wilson BD, Toussaint HM. Effects of drafting on hydrodynamic and metabolic responses in front crawl swimming. Med Sci Sports Exerc. 2009;41(4):837–43. https://doi.org/10.1249/MSS.0b013e31818f2a9b.
14. Forte P, Marinho D, Barbosa TM, Morais JE. Estimation of an elite road cyclist resistive forces and performance wearing standard and aero helmets: an analytical procedure and numerical simulations approach. Eur J Hum Mov. 2021;46:4–15.
15. Forte P, Marinho DA, Barbosa TM, Morouço P, Morais JE. Estimation of an elite road cyclist performance in different positions based on numerical simulations and analytical procedures. Front Bioeng Biotechnol. 2020. https://doi.org/10.3389/fbioe.2020.00538.
16. Forte P, Marinho DA, Morouço P, Pascoal-Faria P, Barbosa TM. Comparison by computer fluid dynamics of the drag force acting upon two helmets for wheelchair racers. AIP Conf Proc. 2017;1863(1):520005. https://doi.org/10.1063/1.4992669
17. Weiss K, Valero D, Villiger E, Thuany M, Scheer V, Cuk I, et al. Temperature and barometric pressure are related to running speed and pacing of the fastest runners in the 'Berlin marathon'. Eur Rev Med Pharmacol Sci. 2022;26(12):4177–287. https://doi.org/10.26355/eurrev_202206_29054.
18. Weiss K, Valero D, Villiger E, Scheer V, Thuany M, Cuk I, et al. The influence of environmental conditions on pacing in age group marathoners competing in the new York City marathon. Front Physiol. 2022;13(842935). https://doi.org/10.3389/fphys.2022.842935.
19. Nikolaidis PT, Di Gangi S, Chtourou H, Rüst CA, Rosemann T, Knechtle B. The role of environmental conditions on marathon running performance in men competing in Boston marathon from 1897 to 2018. Int J Environ Res Public Health. 2019. https://doi.org/10.3390/ijerph16040614. Epub 20190220.
20. Gadelha AB, Sousa CV, Sales MM, Dos Santos Rosa T, Flothmann M, Barbosa LP, et al. Cut-Off values in the prediction of success in olympic distance triathlon. Int J Environ Res Public Health. 2020. https://doi.org/10.3390/ijerph17249491. Epub 20201218.
21. Stones MJ, Hartin A. Aging and Half-Ironman performance. Exp Aging Res. 2017;43(2):178–91. https://doi.org/10.1080/0361073x.2017.1276378.
22. Knechtle B, Rüst CA, Knechtle P, Rosemann T, Lepers R. Age-related changes in ultra-triathlon performances. Extrem Physiol Med. 2012;1(1):5. https://doi.org/10.1186/2046-7648-1-5.
23. Lepers R. Sex difference in triathlon performance. Front Physiol. 2019. https://doi.org/10.3389/fphys.2019.00973.
24. Sousa CV, da Silva Aguiar S, Rosemann T, Nikolaidis PT, Knechtle B. American masters road running Records-The performance gap between female and male age group runners from 5 Km to 6 days running. Int J Environ Res Public Health. 2019. https://doi.org/10.3390/ijerph16132310. Epub 20190629.
25. Sinisgalli R, de Lira CAB, Vancini RL, Puccinelli PJG, Hill L, Knechtle B, et al. Impact of training volume and experience on amateur ironman triathlon performance. Physiol Behav. 2021;232:113344. https://doi.org/10.1016/j.physbeh.2021.113344.
26. Thuany M, Weiss K, Villiger E, Scheer V, Ouerghi N, Gomes TN, et al. A macro to micro analysis to understand performance in 100-mile ultra-marathons worldwide. Sci Rep. 2023;13(1):1415. https://doi.org/10.1038/s41598-023-28398-2.
27. Mean Absolute Error. In. In: Sammut C, Webb GI, editors. Encyclopedia of machine learning. Boston, MA: Springer US; 2010. p. 652.
28. Asgari B, Khorshidi R. The effect of hosting on performance of host countries in summer and winter olympic games. Eur J Sports Exerc Sci. 2013;2(2):1–6.
29. Mölenberg FJM, de Waart F, Burdorf A, van Lenthe FJ. Hosting elite sport events to target recreational sport participation: an interrupted time series analysis. Int J Sport Policy Politics. 2020;12(4):531–43. https://doi.org/10.1080/19406940.2020.1839530.
30. Palazzolo J. Anxiety and performance. Encephale. 2020;46(2):158–61. https://doi.org/10.1016/j.encep.2019.07.008. Epub 20190918.
31. Song M, Zhang Y. Research on the relationship between geographical Factors, sports and culture. Adv Phys Educ. 2018;8:66–70. https://doi.org/10.4236/ape.2018.81008.
32. Knechtle B, Rosemann T, Nikolaidis P. The role of nationality in Ultra-Endurance sports: the paradigm of Cross-Country skiing and Long-Distance running. Int J Environ Res Public Health. 2020. https://doi.org/10.3390/ijerph17072543. Epub 20200408.
33. Sousa CV, Nikolaidis PT, Knechtle B. Ultra-triathlon-Pacing, performance trends, the role of nationality, and sex differences in finishers and non-finishers. Scand J Med Sci Sports. 2020;30(3):556–63. https://doi.org/10.1111/sms.13598. Epub 20191126.
34. Rüst CA, Bragazzi NL, Signori A, Stiefel M, Rosemann T, Knechtle B. Nation related participation and performance trends in 'Norseman xtreme triathlon' from 2006 to 2014. SpringerPlus. 2015;4(1):469. https://doi.org/10.1186/s40064-015-1255-5.
35. Eime RM, Charity MJ, Harvey JT, Payne WR. Participation in sport and physical activity: associations with socio-economic status and geographical remoteness. BMC Public Health. 2015;15:434. https://doi.org/10.1186/s12889-015-1796-0. Epub 20150429.
36. Lepers R, Knechtle B, Stapley P. Trends in triathlon performance: effects of sex and age. Sports Med. 2013;43(9):851–63. https://doi.org/10.1007/s40279-013-0067-4.
37. Puccinelli PJ, de Lira CAB, Vancini RL, Nikolaidis PT, Knechtle B, Rosemann T, et al. The Performance, physiology and morphology of female and male Olympic-Distance triathletes. Healthcare. 2022;10(5):797. https://doi.org/10.3390/healthcare10050797.
38. Barandun U, Knechtle B, Knechtle P, Klipstein A, Rüst CA, Rosemann T, et al. Running speed during training and percent body fat predict race time in recreational male marathoners. J Sports Med. 2012;3:51–8. https://doi.org/10.2147/OAJSM.S33284. Epub 2012/01/01.
39. Mountjoy M, Thomas A, Levesque J. The female triathlete. In: Migliorini S, editor. Triathlon medicine. Cham: Springer International Publishing; 2020. pp. 207–23.
40. Knechtle R, Rüst CA, Rosemann T, Knechtle B. The best triathletes are older in longer race distances – a comparison between Olympic, Half-Ironman and ironman distance triathlon. SpringerPlus. 2014;3(1):538. https://doi.org/10.1186/2193-1801-3-538.
41. Loy SF, Hoffmann JJ, Holland GJ. Benefits and practical use of cross-training in sports. Sports Med. 1995;19(1):1–8. https://doi.org/10.2165/00007256-199519010-00001.
42. Tanaka H. Effects of cross-training. Transfer of training effects on VO2max between cycling, running and swimming. Sports Med. 1994;18(5):330–9. https://doi.org/10.2165/00007256-199418050-00005.
43. Millet GP, Candau RB, Barbier B, Busso T, Rouillon JD, Chatard JC. Modelling the transfers of training effects on performance in elite triathletes. Int J Sports Med. 2002;23(01):55–63.
44. Puccinelli PJ, Lima GHO, Pesquero JB, de Lira CAB, Vancini RL, Nikolaids PT, et al. Previous experience, aerobic capacity and body composition are the best predictors for olympic distance triathlon performance: predictors in amateur triathlon. Physiol Behav. 2020;225:113110. https://doi.org/10.1016/j.physbeh.2020.113110.

Publisher's Note