Sport Sciences for Health
https://doi.org/10.1007/s11332-023-01159-4

RESEARCH

Predictive performance models in marathon based on half-marathon, age group and pacing behavior

Iker Muñoz-Pérez[1] · Arkaitz Castañeda-Babarro[1] · Aitor Santisteban[1] · Adrián Varela-Sanz[2]

Received: 5 September 2023 / Accepted: 11 December 2023

Abstract

Objective The main aim of this study was to develop an equation for predicting performance in 42.2 km (MRT) using pacing and packing behavior, age group and previous 21.1 km time as possible explanatory variables.

Methods 1571 men and 251 female runners who took part in the Valencia Marathon and Half-Marathon were selected to display the regression models. Stepwise regression analysis showed as explanatory variables for MRT: pacing behavior, age group, and time in 21.1 km.

Results The analysis showed four regression models to estimate accurately MRT based principally on athletes previous performance in half-marathon and pacing behavior for men (R^2= 0.72–0.88; RMSE= 4:03–8:31 [min:s]). For women, it was suggested a multiple linear regression for estimating MRT (R^2 0.95; RSE= 8:06 [min:s]) based on previous performance in half-marathon and pacing behavior. The subsequent concordance analysis showed no significant differences between four of the total regressions with real time in the marathon ($p>0.05$).

Conclusion The present results suggest that even and negative pacing behavior and a better time in 21.1 km, in the previous weeks of the marathon, might accurately predict the MRT. At the same time, nomadic packing behavior was the one that reported the best performance. On the other hand, although the age group variable might partially explain the final performance, it should be included with caution in the final model because of differences in sample distribution, causing an overestimation or underestimation of the final time.

Keywords Endurance · Modeling · Prediction · Running · Testing

Introduction

Recreational running raised as one of the most popular physical activities worldwide, gaining more adepts and increasing the number of running events [1]. In this sense, marathon (i.e., 42.195 km) running is a physically high-demanding challenge that become one of the most popular and beloved running distances in recent years [2, 3]. In fact, a great number of recreational runners systematically train year-by-year seeking their best marathon performance, in some cases trying to imitate world-class marathon runners. Scientific evidence has consistently demonstrated that the optimal manipulation of several training variables (e.g., accumulated training volume, training intensity distribution -TID-, training periodization and peaking strategies), as well as training experience, are key factors that affect running performance [4–6]. In this regard, it was previously suggested that an accurate and realistic prediction of the marathon race final time (MRT) is also crucial for optimizing running performance [7]. Moreover, planning optimal competition strategies (e.g., pacing behavior) increase the possibilities of a runner to successfully complete a marathon, or to reach a personal best MRT [8, 9]. Concerning this, it is important to consider that the age of peak performance in long-distance running races (i.e., from 5 km to marathon) is different between men and women [10]. More specifically, women seem to achieve their best half-marathon and marathon race time 1 year and 3 years earlier in life than men, respectively [10].

✉ Iker Muñoz-Pérez
iker.munoz@deusto.es

1 Department of Physical Activity and Sports, Faculty of Education and Sports, University of Deusto, Bilbao, Spain

2 Performance and Health Group, Department of Physical Education and Sport, Faculty of Sports Sciences and Physical Education, University of A Coruña, A Coruña, Spain

In recent decades, several mathematical models have been proposed to predict MRT in both male and female runners. These models are based on different endurance performance variables, including anthropometric (e.g., body mass, body mass index -BMI-, body fat percentage, calf circumference, trunk-to-leg proportion etc.) and physiological variables (e.g., maximum oxygen consumption -VO2max-, running economy -RE-, physiological thresholds etc.) [11, 12] as well as those related to training (e.g., weekly training distance or volume, training frequency, average weekly running speed or pace, TID etc.) [3, 13–15] and previous experience in the distance [16]. However, some of these variables are commonly assessed by qualified specialists through laboratory tests or by using specific and not easily available equipment for recreational runners and coaches. Although these models are valid and have reported relatively good accuracy [17], predictions based on race final time in shorter distances remain the best option for estimating MRT performance in recreational runners [18, 19]. Thus, participating in a shorter race before the main race, but not too close to cause fatigue during the marathon competition, might be an easy way to test and estimate endurance running performance of recreational runners. In this sense, several cities and organizations worldwide hold competitions over shorter distances (usually half-marathon, 21.1 km), weeks before the marathon race under the claim of “road to”. This is the case of Valencia Marathon, in Spain, one of the fastest, flattest and renamed marathons in the world. Through such preparatory races, athletes might estimate their future performance and pacing on a course profile that is partially similar to that of the main race (i.e., marathon race).

Undoubtedly, accurate prediction of MRT is the main factor in establishing rational and realistic competition pacing, and displaying optimal pacing behavior (PB) during the race is mandatory to follow the previous strategy. Pacing behavior is defined as the way in which effort is distributed along the race [9], and is a crucial skill to develop in order to increase marathon performance in both elite [20] and recreational runners [21]. In fact, Renfree and Casado (2018) suggested that choosing the optimal pacing strategy and PB over the duration of long-distance events will improve running performance. In this regard, there are several factors that affect PB, such as running performance [20], sex [22] and drafting strategies [23]. Concerning running performance, top runners tend to adopt a more even PB (i.e., the pace between the first half-marathon and the second one is almost the same) [22–24], while runners with lower performance level often perform a positive PB (i.e., the first half-marathon is faster than the second one) [24]. Other authors have also suggested that sex might influence runner’s PB, with positive PB being the most common in men, while women show fewer variations in PB during a marathon [25]. Nonetheless, in the last five decades, a negative PB (i.e., the second half-marathon is faster than the first one) has emerged as a good strategy to achieve the best MRT in elite marathon runners [20]. In addition, drafting strategies may also affect PB. Drafting enhances performance in distance running competitions, as demonstrated by previous studies [5]. Accordingly, it has been proposed that running behind an athlete during the second half of a marathon may result in a 5.9% decrease in the metabolic cost of running [26]. In other words, running behind another runner favors RE and therefore, running performance. Similarly, pack formation could help athletes draft one another during a race. Although the influence of pack formation on performance and PB has been investigated in global championships for half-marathons [27] and marathons [23], there are no consensus regarding the best strategy to follow another runner during the race. In this sense, Hanley (2016) suggested that packing strategies might be diverse, mostly when runners display an even PB.

To the best of our knowledge, no study has proposed a regression model based on previous time in half-marathon, age group, PB, and packing behavior during a marathon. Thus, the aims of the present study were: 1) to determine the influence of PB and packing behavior on marathon performance, and 2) to develop an accurate regression model, for men and for women, based on previous half-marathon time, PB, age group, and packing behavior. We hypothesize that the inclusion of these variables might explain MRT better than previous models, especially when runners’ performance level is considered.

Materials and methods

Ethical aspects

Informed consent and ethical approval from university Ethics Committee were not necessary because the data were public and freely available (https://www.valenciaciudaddelrunning.com/maraton/clasificaciones-maraton-2021/). This study was conducted in accordance with the Declaration of Helsinki (1964, amended in 2013) concerning human experimentation.

Study design

An observational approach was followed to conduct this study. Ten race lap splits were examined in order to categorize PB (0–5, 5–10, 10–15, 15–20, 20–21.1, 21.1–25, 25–30, 30–35, 35–40 and 40–42.2 km). Informed consent was not obtained from participants because the data were public and free. This study also met the standards of the Declaration of Helsinki regarding human experimentation.

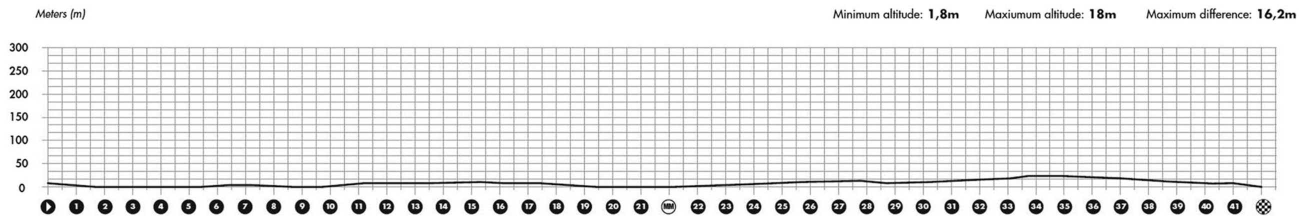

Fig. 1 Valencia Marathon altimetric race profile (recovered from https://www.valenciaciudaddelrunning.com/wp-content/uploads/2022/11/42k-recorrido-2022-con-avituallamiento.pdf)

Participants

Official electronic splits and finishing times were obtained from the official website of each race (Valencia Half-marathon and Marathon: https://www.valenciaciudaddelrunning.com/medio/clasificaciones-medio-maraton-2021/ and https://www.valenciaciudaddelrunning.com/maraton/clasificaciones-maraton-2021/) by means of web scraping technique. Subsequently, through matching technique we identified the names and surnames that took part in both races. Therefore, the inclusion criteria were: 1) to have taken part in the two races in 2021; 2) all official splits times in each race interval were properly recorded and available; 3) the absence of any atypical record (i.e., lack of a lap record or wrong measurement); and 4) to finish the marathon race between 2:06:08 and 5:33:46 [h:min:s] for men and from 2:26:00 to 5:20:03 [h:min:s] for women. This interval time ensures that the different athletes included in the present study run during the most part of the competition. After applying all the inclusion criteria, 1822 runners (1571 men and 251 female) were selected for further analysis.

Pacing behavior criteria

To define PB among runners regardless of their final performance, the average speed of each lap was normalized relative to the average individual race speed for each performance. The average race speed for each runner is represented by a value of 1.00. As a result, a number higher than 1.00 denotes a split pace faster than the average race speed, whereas a value lower than 1.00 shows a lap pace slower than the average race speed.

Pack formation

The split time gap with the nearest runner in each section must be less than one second to determine whether an athlete is running in a pack. This classification criterion, which separates six different forms of packing behavior throughout the race, was based on Hanley's proposal (2016): ever-present pack, in which all athletes run together for at least eight of the nine splits, up to and including the 40th km; halfway pack, when all participants ran together until halfway but were then separated after at least 30th km; nomadic-pack, in which all racers ran with different rivals for at least seven of the nine segments; semi-nomadic pack, where all runners were in packs (not necessarily the same) for at least five of the seven splits until the 30th km, but then ran alone; regrouping-pack, where runners belonged to different packs for more than half of the nine splits and regrouped after having run alone for two splits; and short-lived packs, where athletes ran in a pack for fewer than five of the nine splits.

Race profile

Valencia Marathon altimetric race profile (i.e., course gradient) displays a minimum and maximum altitude over sea level of 1.8 and 18 m, respectively, which makes it one of the fastest and flattest marathons worldwide (Fig. 1).

Marathon final time

Marathon range final time was established for ensuring runners included in our study to run during most time of the competition. Based on previous studies that established the preferred gait transition speed between walking and running at ~ 7.2–7.4 km/h [28–30], we set the average marathon speed threshold at 7.5 km/h for guarantying running the most part of the marathon.

Statistical analysis

The data homogeneity of variance test was performed using Levene's test, and the Kolmogorov–Smirnov, Cramer-von Mises, and Anderson-Darling tests were used to analyze the normal distribution of all continuous variables. Thereafter, unpaired t-tests (between the first and second half-marathon times) were used to identify runners' PB (i.e., negative, positive, or even) [31].

The χ^2 Pearson test was performed to check possible dependencies between variables: PB × packing behavior, age group × packing behavior, and age group × PB. If the previous analysis showed a significant association, Cramér's V was set to establish the effect size (ES); thresholds for effects were: < 0.2 "small", 0.2 ≤ 0.6 "medium", and > 0.6 "large". A subsequent correspondence analysis was carried out to determine the proximity relationship between the variables.

After these preliminary tests, we created two multiple regression models (one for men and another for women) with MRT as the dependent variable. The following independent variables were considered as possible predictors of MRT: age category, half-marathon time 6 weeks before marathon, PB and packing behavior displayed during marathon, the interactions between PB and packing behavior, age group and PB, and age group and packing behavior. The interaction is a combination of variables, making a new one that has a significantly larger effect on the dependent variable than the sum of individual independent variables alone.

The final regression model was selected using stepwise forward and backward method of the "caret" R package [32]. For internal validation, k-fold cross-validation (10 folds and five repetitions) was performed. Internal validation was performed to reduce possible overfitting of the model [33]. There was no sign of multicollinearity (small Variance Inflation Factor for all independent variables).

The R package dplyr was used to identify possible outliers and improve the fitting of the regression model [34]. The outlier data in the multiple regression model were identified and removed when the absolute value of the studentized residual (SRE) was ≥ 2. After this analysis, a final sample of 1416 men and 186 women was taken into consideration for the final regression models.

The Jarque-Bera test was used to check the normal distribution of the residuals in both regression models (men and women). In turn, homoscedasticity of the regression models was checked by Breusch-Pagan test.

To evaluate whether our proposed predictive models differed significantly from the actual time of the runners a concordance test was conducted.

Model performance was assessed using the root mean square error (RMSE) and R^2. The root-mean-square error is the error of the model reported in the outcome units (i.e., min:s). All statistical analyses were two-sided, and the significance level was set at $p < 0.05$. Statistical analysis was conducted using R software 4.2.2 (R Core Team, 2022) and RStudio version 2022.12.0.353 (Rstudio Team, 2022).

Results

The participants' characteristics are presented in Tables 1 and 2 for both men and women, respectively. Descriptive data are structured regarding PB (i.e., even pace, positive pace or negative pace); age group according to the official categorization proposed by the World Athletics (i.e., Under 20, Under 23, Senior, 35, 40, 45, 50, 55, 60, 65 and 70 for both men "M" and women "W"); as well as marathon and half-marathon final time in hours, minutes and seconds including standard deviation. The main behaviors regarding PB and packing were even pace and nomadic in men, while in women were positive and nomadic.

A linear and positive association between Marathon time and Half-marathon time for men and women was observed (r = 0.89; 95% CI= 0.88–0.89; $p < 0.001$ in men; r = 0.83; 95% CI= 0.77–0.87; $p < 0.001$ in women).

The chi-square test showed a significant association between PB and packing behavior (χ^2 (10) = 19.7, p = 0.032; Cramer's V = 0.101) in men but not in women ($p > 0.05$). Similarly, there was a significant dependency between age group and PB (χ^2 (18) = 31.0, p = 0.029; Cramer's V = 0.126) only for men. There was no association between age group and packing behavior of either men or women ($p > 0.05$). Subsequent analyses did not reveal any clear associations between these factors. However, because of the low ES, these data did not allow us to infer the results obtained.

Regression models in men

Owing to the large heterogeneity of the sample, four groups of runners were established according to their performance in the half-marathon. The sample was divided into four quarters (Table 3).

After stepwise regression, four different models for men were constructed considering the time in 21.1 km, PB, Packing, and age group as explanatory variables, but not in all models. There were no significant interactions between the independent variables (i.e., PB × packing behavior; PB × age group; packing behavior × age group). The final regression models for men are showed in Table 3 (R^2= 0.88, 95%CI-0.86-0.9; RSE= 5:08 [min:s]; R^2= 0.72, 95%CI-0.65–0.75; RSE= 4:03 [min:s]; R^2= 0.87, 95%CI-0.85–0.88; RSE= 8:03 [min:s]; R^2= 0.82, 95%CI-0.79–0.85; RSE= 8:31 [min:s], for 1st, 2nd, 3th, and 4th regression model, respectively). After applying concordance test, there were no significant differences between the real MRT and prediction time (p>0.05) (Fig. 2 A 95%CI-617.56–631.92, B 95%CI-476.42–478.45, and D 95%CI-982.6–1022.7). However, the regression model for the 3th performance group (Figure 2C) showed significant differences between the real and predicted times ($p < 0.01$; t = 220, df = 609, 95%CI-931–1150). Although possible outliers were removed, analysis of the residuals from the final regression model showed that they did not follow a normal distribution. Other assumptions (linearity, multicollinearity, and homoscedasticity) were also checked. Thus, the final regression model was constructed using a generalized linear method.

Table 1 Men pacing behavior, age-group, pack formation, marathon, and half marathon performance distribution

PB	Age-group	Pack formation	N	Marathon [h:m:s]	SD ± [h:m:s]	Half marathon [h:m:s]	SD± [m:s]
Even Pace	M-Under 23	Nomadic	2	3:32:24	1:05:01	1:38:13	24:56
	M-Senior	Halfway	10	2:55:49	0:34:59	1:21:43	15:50
		Nomadic	84	3:18:29	0:28:25	1:31:36	13:17
		Regrouping	4	2:40:13	0:05:39	1:13:40	02:39
		Semi-nomadic	3	3:14:29	1:11:14	1:24:52	23:20
		Short-lived	1	2:15:56	–	1:04:21	–
	M35	Halfway	14	3:28:10	0:40:07	1:35:54	15:47
		Nomadic	87	3:23:35	0:32:42	1:34:42	14:37
		Regrouping	2	3:39:19	0:56:56	1:38:30	19:18
		Semi-nomadic	3	3:26:55	1:07:45	1:30:59	20:23
	M40	Halfway	13	3:31:13	0:21:42	1:36:19	09:37
		Nomadic	144	3:26:19	0:28:16	1:34:42	12:04
		Regrouping	6	3:01:20	0:54:00	1:22:42	19:58
		Short-lived	2	4:02:40	2:08:50	1:51:36	58:54
	M45	Everpresent	1	3:24:06	–	1:35:06	–
		Halfway	21	3:40:54	0:27:01	1:41:52	13:33
		Nomadic	132	3:32:42	0:24:08	1:37:59	11:26
		Regrouping	2	4:45:21	0:05:36	1:58:50	09:20
		Semi-nomadic	1	3:06:00	–	1:28:15	–
	M50	Everpresent	2	3:26:21	0:01:43	1:35:07	–
		Halfway	8	3:50:11	0:11:30	1:43:30	05:11
		Nomadic	78	3:43:11	0:29:59	1:41:57	12:59
		Regrouping	2	2:41:41	–	1:15:53	02:19
		Semi-nomadic	2	2:44:59	0:04:32	1:17:59	–
		Short-lived	1	3:34:29	–	1:39:17	–
	M55	Halfway	2	3:47:06	0:22:56	1:44:12	16:26
		Nomadic	37	3:45:20	0:22:55	1:43:47	12:31
		Semi-nomadic	1	4:38:48	–	2:08:43	–
	M60	Halfway	1	4:17:48	–	1:52:42	–
		Nomadic	8	3:40:58	0:11:54	1:41:54	11:40
		Regrouping	2	4:45:03	0:13:07	2:08:44	02:09
		Short-lived	1	4:34:56	–	2:05:36	–
	M65	Nomadic	2	3:23:11	0:39:57	1:30:11	12:05

Table 1 (continued)

PB	Age-group	Pack formation	N	Marathon [h:m:s]	SD ± [h:m:s]	Half marathon [h:m:s]	SD± [m:s]
Negative Pace	M-Under 23	Halfway	1	3:28:05	–	1:35:10	–
		Nomadic	2	3:27:08	–	1:31:00	08:56
	M-Senior	Halfway	2	3:40:27	0:13:55	1:44:41	04:47
		Nomadic	30	3:33:06	0:21:24	1:38:22	09:12
		Short-lived	1	3:18:18	–	1:30:39	–
	M35	Halfway	1	3:18:35	–	1:35:23	–
		Nomadic	35	3:35:09	0:22:07	1:41:33	13:15
		Regrouping	1	4:10:29	–	1:44:45	–
		Semi-nomadic	1	2:44:36	–	1:15:00	–
	M40	Everpresent	1	3:05:50	–	1:29:10	–
		Halfway	6	3:50:18	0:28:22	1:47:59	11:07
		Nomadic	66	3:34:52	0:22:57	1:38:07	10:20
		Regrouping	1	3:41:35	–	1:41:10	–
	M45	Halfway	8	3:38:07	0:13:23	1:40:52	07:27
		Nomadic	59	3:34:25	0:17:59	1:39:17	09:23
		Regrouping	1	4:00:35	–	1:47:25	–
	M50	Halfway	3	3:54:54	0:42:13	1:43:17	15:03
		Nomadic	41	3:32:53	0:20:14	1:38:07	08:17
		Regrouping	2	3:58:03	0:56:15	1:48:35	26:09
	M55	Halfway	3	3:42:21	0:19:32	1:45:49	09:56
		Nomadic	6	3:43:20	0:17:03	1:41:24	04:57
	M60	Halfway	1	4:02:42	–	1:48:24	–
		Nomadic	5	3:59:31	0:17:51	1:47:49	11:48
	M70	Nomadic	1	3:36:00	–	1:40:23	–

Table 1 (continued)

PB	Age-group	Pack formation	N	Marathon [h:m:s]	SD ± [h:m:s]	Half marathon [h:m:s]	SD± [m:s]
Positive Pace	M-Uunder 20	Nomadic	1	4:29:48	–	1:43:30	–
	M-Under 23	Nomadic	3	3:17:35	0:42:52	1:27:39	16:35
	M-Senior	Halfway	5	3:21:20	0:53:47	1:29:46	22:40
		Nomadic	81	3:36:19	0:39:25	1:37:41	17:11
		Regrouping	7	2:35:36	0:15:05	1:12:53	07:03
		Semi-nomadic	4	2:59:19	0:57:53	1:22:09	25:17
		Short-lived	5	2:59:24	1:17:32	1:23:16	34:59
	M35	Everpresent	1	3:15:18	–	1:25:22	–
		Halfway	5	3:56:27	0:54:28	1:43:51	21:04
		Nomadic	72	3:34:51	0:39:11	1:35:02	15:36
		Regrouping	5	3:14:48	0:55:59	1:28:15	22:47
		Semi-nomadic	2	3:36:40	1:17:08	1:41:53	33:49
		Short-lived	3	3:29:24	0:53:58	1:32:40	20:36
	M40	Halfway	13	4:02:45	0:38:17	1:42:46	15:05
		Nomadic	120	3:38:11	0:34:31	1:36:34	14:45
		Regrouping	3	3:21:43	1:08:17	1:29:08	22:11
		Semi-nomadic	5	3:37:03	1:16:46	1:37:40	31:27
		Short-lived	2	5:18:42	0:04:01	2:08:07	16:23
	M45	Everpresent	1	3:21:16	–	1:33:22	–
		Halfway	21	4:08:29	0:35:40	1:45:16	13:29
		Nomadic	107	3:46:29	0:31:05	1:38:59	11:41
		Regrouping	5	3:47:26	0:55:29	1:40:49	22:13
		Semi-nomadic	5	4:01:24	1:08:08	1:45:23	26:34
		Short-lived	1	5:11:24	-	1:33:56	–
	M50	Halfway	4	3:27:54	0:26:26	1:27:25	08:20
		Nomadic	64	4:00:31	0:33:57	1:43:54	12:00
		Regrouping	4	4:40:34	0:11:29	1:57:42	07:16
		Semi-nomadic	3	5:05:33	0:19:56	2:09:55	11:38
		Short-lived	1	2:35:37	–	1:12:39	–
	M55	Halfway	5	4:40:40	0:36:37	1:58:54	20:31
		Nomadic	33	3:57:54	0:19:38	1:44:06	08:27
		Regrouping	2	4:32:46	0:08:53	1:56:34	04:14
	M60	Halfway	3	4:39:17	0:11:26	1:54:37	03:43
		Nomadic	13	4:05:19	0:28:28	1:46:19	09:47
		Semi-nomadic	1	4:30:31	–	2:01:16	–
	M65	Nomadic	4	4:21:20	0:27:34	1:52:03	10:07

PB: pacing behavior; Age group according to World Athletics categorization: M-Under 20: men runners under 20 years, M-Under 23: men runners under 23 years, M-Senior: men runners belonging to Senior category, M35: men runners aged between 35 and 39 years, M40: men runners aged between 40 and 44 years, M45: men runners aged between 45 and 49 years, M50: men runners aged between 50 and 54 years, M55: men runners aged between 55 and 59 years, M60: men runners aged between 60 and 64 years, M65: men runners aged between 65 and 69 years, M70: men runners older than 70 years; h:m:s: hours, minutes and seconds; SD: standard deviation.

Note. Age and pack formation groups without runners were omitted to improve the readability of the table.

Table 2 Women pacing behavior, age-group, pack formation, marathon, and half marathon performance distribution

PB	Age group	Pack formation	N	Marathon [h:m:s]	SD± [h:m:s]	Half marathon [h:m:s]	SD ± [m:s]
Even Pace	W-Senior	Halfway	1	3:36:31	–	1:39:46	–
		Nomadic	13	3:39:44	0:38:51	1:40:31	17:35
	W-35	Halfway	4	3:50:44	0:19:09	1:47:46	16:02
		Nomadic	8	3:43:21	0:34:39	1:41:34	12:47
	W-40	Halfway	1	3:22:31	–	1:33:49	–
		Nomadic	17	4:02:10	0:28:46	1:51:27	12:22
	W-45	Nomadic	13	4:02:32	0:30:37	1:49:06	13:28
		Regrouping	2	4:32:20	0:07:37	2:08:26	04:57
	W-50	Nomadic	5	4:19:56	0:39:49	1:56:35	15:32
	W-55	Nomadic	5	4:01:13	0:40:34	1:50:35	20:10
	W-60	Nomadic	1	4:23:16	–	2:06:00	–
Negative Pace	W-Senior	Nomadic	6	4:13:23	0:24:14	1:57:53	11:48
	W-35	Nomadic	9	3:50:45	0:28:44	1:45:44	13:03
	W-40	Halfway	1	4:29:52	–	2:09:21	–
		Nomadic	11	3:50:20	0:28:16	1:47:43	10:07
		Short-lived	1	4:30:12	–	1:58:43	–
	W-45	Halfway	2	4:48:20	0:22:07	2:16:58	20:33
		Nomadic	13	4:04:41	0:34:05	1:49:33	16:46
		Regrouping	1	4:48:09	–	2:22:37	–
	W-50	Halfway	2	4:08:02	1:07:40	1:55:23	33:41
		Nomadic	2	3:35:03	0:13:16	1:41:02	04:26
	W-55	Nomadic	3	4:09:01	0:17:23	1:50:28	10:21
	W-65	Nomadic	1	3:28:55	–	1:44:43	–
Positive Pace	W-Senior	Halfway	3	4:24:09	0:02:15	2:01:21	02:49
		Nomadic	16	4:04:04	0:33:05	1:45:40	15:19
	W-35	Nomadic	16	4:09:04	0:25:13	1:48:04	11:12
		Regrouping	1	4:50:22	–	2:03:27	–
		Semi-nomadic	2	5:05:40	0:05:30	2:11:44	02:20
		Short-lived	1	2:26:00	–	1:07:48	–
	W-40	Halfway	3	3:51:05	0:51:44	1:40:46	21:23
		Nomadic	14	3:53:06	0:34:12	1:43:24	16:54
		Regrouping	1	4:30:35	–	1:59:14	–
		Short-lived	1	5:04:06	–	2:06:58	–
	W-45	Halfway	1	3:54:50	–	1:44:44	–
		Nomadic	28	4:15:35	0:28:26	1:51:20	13:54
		Regrouping	3	4:40:20	0:21:57	2:00:57	09:38
		Semi-nomadic	3	4:47:17	0:18:10	2:10:14	02:52
	W-50	Halfway	1	5:04:52	–	2:18:35	–
		Nomadic	11	4:22:56	0:26:08	1:52:15	12:08
		Regrouping	3	4:56:00	0:01:49	2:02:31	06:24
		Short-lived	1	5:09:22	–	2:15:07	–
	W-55	Nomadic	5	5:03:38	0:18:46	2:07:41	08:34
		Regrouping	1	5:08:12	–	2:13:27	–
		Semi-nomadic	1	5:15:38	–	1:58:44	–
	W-60	Nomadic	2	4:55:57	0:36:55	1:59:13	05:19
		Regrouping	1	5:12:13	–	2:19:41	–

PB: pacing behavior; Age group according to World Athletics categorization: W-Senior: women runners belonging to Senior category, W35: women runners aged between 35 and 39 years, W40: women runners aged between 40 and 44 years, W45: women runners aged between 45 and 49 years, W50: women runners aged between 50 and 54 years, W55: women runners aged between 55 and 59 years, W60: women runners aged between 60 and 64 years, W65: women runners older than 65 years; h:m:s: hours, minutes and seconds; SD: standard deviation.

Note. Age and pack formation groups without runners were omitted to improve the readability of the table.

Table 3 Regression models for men runners based on performance group in half marathon

Performance Group [h:m:s]	Predictor	Estimate	Std.Error	95% CI Lower	95% CI Upper	t-value	p
	Intercept	– 895.94	227.14	– 1341.12	– 450.76	– 3.94	< 0.001
Group 1 (N=342)	Half-time	2.33	0.05	2.23	2.42	49.54	< 0.001
1:00:06-1:27:41	PB:Negative	178.09	70.19	40.52	315.67	2.54	< 0.01
	PB:Positive	243.96	34.85	175.65	312.26	7	< 0.001
	Intercept	508.37	549.20	– 568.03	1584.78	0.93	0.36
	Half-time	2.08	0.10	1.89	2.27	21.00	< 0.001
	PB:Negative	– 26.53	36.83	– 98.71	45.66	– 0.72	0.47
	PB:Positive	312.46	39.97	234.12	390.79	7.82	< 0.001
Group 2 (N=259)	Under-23	247.56	181.92	– 109.0	604.11	1.36	0.18
1:27:42-1:37:11	M-35	– 180.40	57.59	– 293.27	– 67.52	– 3.13	< 0.01
	M-40	76.41	52.23	– 25.97	178.79	1.46	0.15
	M-45	– 24.26	50.53	– 123.30	74.79	– 0.48	0.63
	M-50	– 126.37	59.27	– 242.53	-10.20	– 2.13	< 0.05
	M-55	138.22	84.94	– 28.26	304.70	1.63	0.11
	M-60	201.45	120.05	– 33.85	436.75	1.68	0.1
	Intercept	533.31	285.68	– 26.62	1093.24	1.87	0.06
	Half-time	2.13	0.04	2.06	2.22	50.38	< 0.001
	PB:Negative	– 224.98	52.83	– 328.52	– 121.43	– 4.26	< 0.001
Group 3 (N=610)	PB:Positive	521.42	44.10	434.98	607.85	11.82	< 0.001
1:37:12-1:46:10	Pack-nomadic	– 266.03	61.65	– 386.81	– 145.25	– 4.32	< 0.001
	Pack-regrouping	514.10	125.87	267.41	760.79	4.08	< 0.001
	Pack-semi-nomadic	102.14	185.35	– 261.14	465.42	0.55	0.58
	Pack-short-lived	– 210.23	223.17	– 647.64	227.18	– 0.94	0.35
	Intercept	3404.05	490.95	2441.81	4366.29	6.93	< 0.001
	Half-time	1.71	0.07	1.57	1.85	24.37	< 0.001
	PB:Negative	– 168.89	87.23	– 339.85	2.07	– 1.94	0.05
Group 4 (N=300)	PB:Positive	712.60	65.65	583.93	841.27	10.86	< 0.001
1:46:11-2:33:15	Pack-nomadic	– 379.92	84.53	– 545.61	– 214.24	– 4.49	< 0.001
	Pack-regrouping	581.64	146.02	295.46	867.83	3.98	< 0.001
	Pack-semi-nomadic	128.62	186.82	– 237.53	494.78	0.69	0.49
	Pack-short-lived	362.64	252.30	– 131.87	857.15	1.44	0.15

PB: pacing behavior; Age group according to World Athletics categorization: M-Under 23: men runners under 23 years, M35: men runners aged between 35 and 39 years, M40: men runners aged between 40 and 44 years, M45: men runners aged between 45 and 49 years, M50: men runners aged between 50 and 54 years, M55: men runners aged between 55 and 59 years, M60: men runners aged between 60 and 64 years; Half-time: final time in half-marathon; h:m:s: hours, minutes and seconds.

Note. PB even and halfway-packing were selected as reference.

Regression model in women

Applying stepwise regression, the final 42.2 km model for women considered the time in 21.1 km and PB as explanatory variables. The resultant model was a multiple linear regression (R^2= 0.95, 95%CI-0.94–0.96; RSE= 8:06 [min:s]) (Table 4). No meaningful interactions were observed between the independent variables (i.e., PB × packing behavior, PB × age group, packing behavior × age group). No significant differences were observed between real MRT and predicted times (p >0.05, Figure 2E). The residuals of the linear regression model were normally distributed and the, linearity, multicollinearity, and homoscedasticity were checked.

For a better comprehension, Table 5 summarizes the five equation models (four for men and one for women).

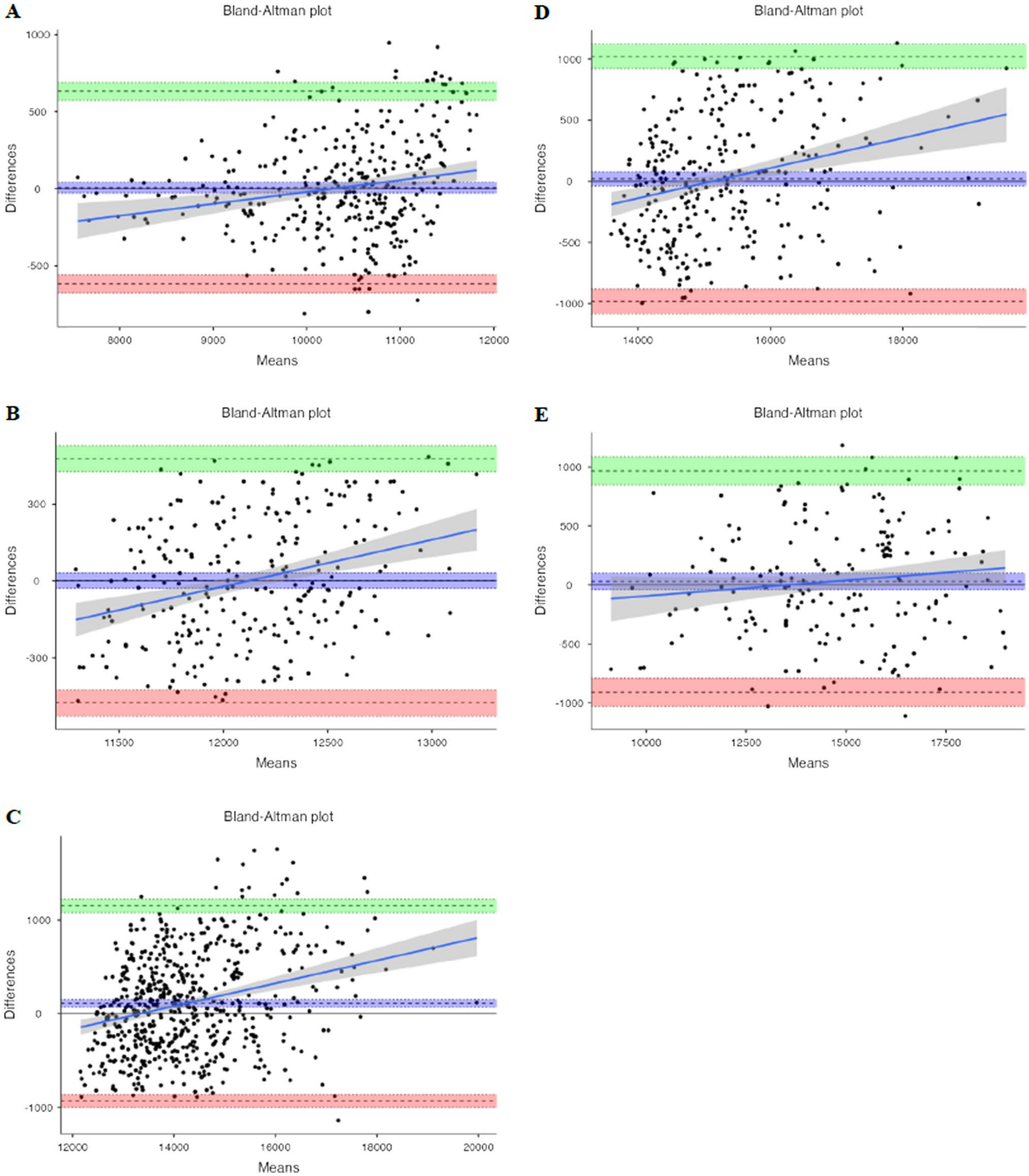

Fig. 2 Differences between the real marathon time and predicted time regarding performance group and sex. **A**: men performance group 1; half-marathon time (in h:min:s) between 1:00:06 and 1:27:41; **B**: men performance group 2; half-marathon time (in h:min:s) between 1:27:42 and 1:37:11; **C**: men performance group 3; half-marathon time (in h:min:s) between 1:37:12 and 1:46:10; **D**: men performance group 4; half-marathon time (in h:min:s) between 1:46:11 and 2:33:15; **E**: real vs predicted marathon time in women

Table 4 Regression model using Marathon as the criterion for women

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]
Intercept	– 397.91	[– 903.21, 107.38]		
Half-time	2.24**	[2.17, 2.32]	.87	[.80, .95]
PB: Negative	– 112.95	[-310.21, 84.31]	.00	[-.00, .00]
PB: Positive	760.24**	[597.19, 923.28]	.02	[.01, .03]

PB: pacing behavior; Half-time: final time in half-marathon; CI: confidence interval.

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.

* indicates p < .05. ** indicates p < .01.

Discussion

The main aim of this study was to propose different regression models for men and women based on the runner's age group, time spent in the previous half marathon, PB, packing behavior, and performance group (only for men models). However, all predictor variables considered differed in relevance regarding performance groups in men and might partially explain the dependent variable (i.e., MRT), thus our hypothesis was in part validated. To the best of our knowledge, this is the first study that consider runners' behavior during competitions for improving the accuracy of the regression models. Moreover, the participants of our study presented a great heterogeneity regarding marathon performance, which makes our predictive model implementation suitable in a wide performance range of runners, from recreational to high-level runners.

Analyzing our models for predicting marathon performance in male runners of different levels, it seems that, apart from the previous time in half-marathon, there is a tendency of PB being more relevant as runner's performance level increase, while pack formation becomes more important in lower-level runners. Further, age group may also play a small role in predicting marathon performance in middle-level male runners (i.e., performance group 2). For instance, the variables explaining the marathon predicted performance in male runners under one and a half hour in half-marathon (i.e., performance group 1) are PB displayed and previous time in half-marathon, with a coefficient of determination (R^2) of 0.88. On the other hand, marathon time prediction relies more on half-marathon time, pack formation (i.e., nomadic and regrouping), as well as positive, even, and negative PB (only performance group 3) pace in middle-to-low-level runners with a previous time in half-marathon between 1 h and 37 min and 2 h and 33 min (i.e., performance groups 3 and 4; R^2 = 0.87 and 0.82, respectively).

The novelty of our model relies in the inclusion of PB and pack formation as explanatory variables. In both men's and women's regression models, only positive PB was significantly correlated with a reduction in runner's performance (from 4:04 to 11:53 [m:s] for men and 11:40 [m:s] for women), while pack formation seems to be more relevant as runner's level decrease. This tendency to decline the performance time during marathons by displaying a positive PB has been well documented in several studies of elite [20] and recreational athletes [21]. Although the present models did not show superior statistical significance of negative PB compared to even PB, it might be observed a small tendency (Tables 3 and 4) of improving performance following a negative PB for almost models. This superiority of negative and even PB compared to positive PB has been consistently reported in previous studies [22–24]. Therefore, our proposal allows athletes to evaluate their performance based on different PBs prior to marathon. Moreover, our models check how could PB affects runners' MRT. Taking together, in almost all the proposed models, for both men and women, negative PB showed a tendency to perform better, although in most cases, this performance was not statistically different. Likewise, in the regression models (Groups 3 and 4) in which packing behavior was selected as an explanatory variable, nomadic packing was shown to be the best strategy for performance enhancement. In this sense, previous studies indicate that this packing behavior may be the most

Table 5 Regression equations summary

Grouped by half marathon time [h:m:s]	Sex		Equation model
1:00:06-1:27:41	M	Marathon time (in s) =	– 895.94+2.33*(time in half marathon in s)+ PB(see the estimate in Table 3)
1:27:42-1:37:11			508.37+2.08*(time in half marathon in s)+ PB(see the estimate in Table 3) + Age group (see the estimate in Table 3)
1:37:12-1:46:10			533.31+2.13*(time in half marathon in s)+PB(see the estimate in Table 3)+Packing behavior (see the estimate in table 3)
1:46:11-2:33:15			3404.05+1.71*(time in half marathon in s)+PB(see the estimate in Table 3)+Packing behavior (see the estimate in Table 3)
-	W		– 397.91+2.24*(time in half marathon in s)+PB(see the estimate in table 4)

appropriate because it allows the athletes to develop a race pace according to their fitness [23, 35]. Further, adopting nomadic packing could prevent runners to follow those runners' groups performing a too fast race pace they could not maintain during the whole competition.

On the other hand, age group was another major factor that slightly influenced the prediction models. However, only one regression model (men performance group 3) took into account age-group as explanatory variable. In this sense, based on the limitation of our data it is difficult to draw some conclusions. Thus, age might not be more determinant than the runner's previous experience in marathon distance [18] or other predictive variables. Our results are in accordance with previous studies, since Nikolaidis and Knechtle (2017) reported only a trivial interaction between marathon performance and age group in the New York City Marathon.

As expected, the main prediction variable in our models was the time in the previous half-marathon for both men and women. In this sense, some popular approaches based on performance in shorter distances, close to the main sports event, might be right for establishing future marathon pacing and assessing a runner's fitness status [18]. Prior studies have underlined the utility of using time in shorter races, such as 10 km, a mile [18, 19], and half-marathon to predict MRT [18], thus the use of these regression models allows the assessment of marathoners' performance. On the other hand, other factors that can influence the prediction of time based on previous distances are weather conditions and race profile (i.e., relative humidity, temperature, course gradient etc.) [36]. In this regard, previous studies have highlighted the negative influence of increasing relative humidity and temperature on runners' performance [3]. Taking this into account, our models have been based on practically the same circuit in the same city, with a flat profile (Valencia, Spain) and similar weather conditions (12 °C vs. 11 °C and relative humidity of 40% vs. 50% for half-marathon vs. marathon, respectively). Based on the relative humidity and temperature data for both distances, it can be assumed that these variables have not influenced the performance of a single race (21.1 or 42.2 km) and therefore, did not affect the predictive model.

Regarding the robustness of the different proposed regression models, it is also important to consider that the coefficients of determination of our models ranged from 0.72 to 0.88 for men, and 0.95 for women, which in all cases are superior than other studies that considered previous race time in shorter distances (10 and 21.1 km) [19]. Scientific literature reported other regression models for estimating MTR mainly based on training variables (i.e., average weekly training distance or volume, training frequency, mean weekly training speed or pace, maximum workout distance per week etc.).[13, 14, 37] In this regard, the equation proposed by Schmid and colleagues (2012) reported R^2 values of 0.50 considering calf circumference and average running speed during training. Similarly, other works reported prediction equation R^2 values of 0.44 based on body fat percentage and running speed during training,[38] while Nikkolaidis and coworkers (2021) reported higher R^2 values (R^2 = 0.61) when VO_{2max}, weekly training distance and BMI are included in the prediction equation. Only two studies [14, 37] reported R^2 values similar to those of our regression models (R^2 = 0.72 for Tanda's model and R^2 = 0.81 for Tanda and Knechtle's model vs R^2 = 0.87-0.91 for ours). However, compared to our models, some of these proposals require monitoring of specific training and physiological variables that the majority of recreational runners do not conduct, making it more difficult to apply these regression models.

The present study has several limitations. The main one is based on the small sample size in the U-23, M-60, M-65, and M-70 age groups in males and in the U-23, W-60, and W-65 age groups in females. Regarding this, we encourage not to use this predictor (Table 3, group 3), which corresponds to each respective age group. This fact could limit the capacity of our models to accurately predict MRT in these specific age groups. A second limitation of our study is the estimation error (RMSE). In men, this prediction error is close to 08:30 [min:s] for lower-level runners, while for runners performing better this error is reduced in 4 min. For female runners, this estimation error rise to 8:06 [min:s]. These results might be interpreted considering the great heterogeneity of our sample when regarding marathon performance, since both high- and low-performance level runners were included in our work. However, the RMSE for our models is in accordance with previous regression models, with lower R2 values for men [14, 39, 40] and women [41]. More recent studies have been able to improve this estimation error based on previous race times [42] and several training variables (i.e., typical mileage, number of tempo and interval training sessions) [43, 44].

In summary, this is the first study to propose different regression models for both men and women based on a previous time in half-marathon, runners' age, PB and pack formation, as well as performance level (only for men models). The high R^2 values reported for all models and the heterogeneity of the sample included make them accurate for easily estimate MRT in runners of different performance levels.

Practical applications

This study has focused on the prediction of the marathon time based on the time achieved in a half marathon in the previous weeks. Bearing in mind that the vast majority of marathoners run a half marathon weeks before their main race, the MRT prediction capacity obtained in this study is very useful and easily applicable because it has no economic cost, has a great reliability in runners with different performance levels and it can be considered a pre-competition estimation tool that help runners to set realistic time and race pace goals for their marathons.

Conclusion

Participating in a half-marathon before the main competition (i.e., marathon) may be one of the easiest ways to accurately estimate MRT. In addition, selecting a competitive strategy that develops a negative or even PB during the marathon race will enhance marathon performance. Consequently, we encourage runners to plan in detail the competitive strategy to optimize his/her race pace during marathon according to their performance level and therefore, improve MRT. On the other hand, other factors such as packing formation and the runner's age group can help to partially explain the final performance and should be taken into account for the prediction of the MTR.

Acknowledgements The authors reported there is no funding associated with the work featured in this article.

Author contributions Conception and design of study: I.M-P; A.V-S. Acquisition of data: I.M-P; A.S; A.C-S. Analysis and interpretation of data: I.M-P; A.V-S; A.C-B. Drafting the manuscript: I.M-P; A.V-S; A. S; A.C-B.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability statement The data that we used is open at: https://www.valenciaciudaddelrunning.com/medio/clasificaciones-mediomaraton-2021/ and https://www.valenciaciudaddelrunning.com/maraton/clasificaciones-maraton-2021/ as we have written in the manuscript. However, if a person is interested, he/she may contact the corresponding author to obtain them.

Conflict of interest No potential conflict of interest was reported by the authors.

References

1. Anderson LM, Martin JF, Barton CJ, Bonanno DR (2022) What is the Effect of Changing Running Step Rate on Injury, Performance and Biomechanics? A Systematic Review and Meta-analysis. Sport Med - Open 8:112. https://doi.org/10.1186/s40798-022-00504-0
2. Knechtle B, Di Gangi S, Rüst CA, Nikolaidis PT (2020) Performance Differences Between the Sexes in the Boston Marathon From 1972 to 2017. J Strength Cond Res 34:566–576. https://doi.org/10.1519/JSC.0000000000002760
3. Reusser M, Sousa CV, Villiger E, Alvero Cruz JR, Hill L, Rosemann T, Nikolaidis PT, Knechtle B (2021) Increased Participation and Decreased Performance in Recreational Master Athletes in "Berlin Marathon" 1974–2019. Front Physiol 12:631237. https://doi.org/10.3389/fphys.2021.631237
4. Mølmen K, Øfsteng SJ, Rønnestad BR (2019) Block periodization of endurance training – a systematic review and meta-analysis. Open Access J Sport Med 10:145–160. https://doi.org/10.2147/OAJSM.S180408
5. Casado A, González-Mohíno F, González-Ravé JM, Foster C (2022) Training Periodization, Methods, Intensity Distribution, and Volume in Highly Trained and Elite Distance Runners: A Systematic Review. Int J Sports Physiol Perform 17:820–833. https://doi.org/10.1123/ijspp.2021-0435
6. Issurin VB (2010) New Horizons for the Methodology and Physiology of Training Periodization. Sport Med 40:189–206. https://doi.org/10.2165/11319770-000000000-00000
7. Liverakos K, McIntosh K, Moulin CJA, O'Connor AR (2018) How accurate are runners' prospective predictions of their race times? PLoS One 13:e0200744. https://doi.org/10.1371/journal.pone.0200744
8. Muñoz-Pérez I, Varela-Sanz A, Lago-Fuentes C, Navarro-Patón R, Mecías-Calvo M (2022) Central and Peripheral Fatigue in Recreational Trail Runners: A Pilot Study. Int J Environ Res Public Health 20:402. https://doi.org/10.3390/ijerph20010402
9. Renfree A, Casado A (2018) Athletic Races Represent Complex Systems, and Pacing Behavior Should Be Viewed as an Emergent Phenomenon. Front Physiol 9:1432. https://doi.org/10.3389/fphys.2018.01432
10. Knechtle B, Nikolaidis PT, Di Gangi S (2018) World Single Age Records in Running From 5 km to Marathon. Front Psychol. https://doi.org/10.3389/fpsyg.2018.02013
11. Esteve-Lanao J, Del Rosso S, Larumbe-Zabala E, Cardona C, Alcocer-Gamboa A, Boullosa DA (2021) Predicting Recreational Runners' Marathon Performance Time During Their Training Preparation. J strength Cond Res 35:3218–3224. https://doi.org/10.1519/JSC.0000000000003199
12. Muñoz Pérez I, Moreno Perez D, Cardona Gonzalez C, Esteve-Lanao J (2012) Prediction of race pace in long distance running from blood lactate concentration around race pace. J Hum Sport Exerc 7:763–769. https://doi.org/10.4100/jhse.2012.74.04
13. Schmid W, Knechtle B, Knechtle P, Barandun U, Rüst CA, Rosemann T, Lepers R (2012) Predictor Variables for Marathon Race Time in Recreational Female Runners. Asian J Sports Med 3:90–98. https://doi.org/10.5812/asjsm.34704
14. Tanda G (2011) Prediction of marathon performance time on the basis of training indices. J Hum Sport Exerc 6:511–520. https://doi.org/10.4100/jhse.2011.63.05
15. Tanda G, Knechtle B (2013) Marathon performance in relation to body fat percentage and training indices in recreational male runners. Open Access J Sport Med 4:141–149. https://doi.org/10.2147/OAJSM.S44945

16. Hynie M, MacKinnon L, MacKinnon C, Jones G (2015) The relationship between 5 km running performance and marathon completion time. J Sport Exerc Psychol 37:144–152
17. Takeshima N, Tanaka K (1995) Prediction of endurance running performance for middle-aged and older runners. Br J Sports Med 29:20–23. https://doi.org/10.1136/bjsm.29.1.20
18. Coquart JBJ, Alberty M, Bosquet L (2009) Validity of a Nomogram to Predict Long Distance Running Performance. J Strength Cond Res 23:2119–2123. https://doi.org/10.1519/JSC.0b013e3181b3dcc3
19. Slovic P (1977) Empirical study of training and performance in the marathon. Res Q 48:769–777
20. Díaz JJ, Fernández-Ozcorta EJ, Santos-Concejero J (2018) The influence of pacing strategy on marathon world records. Eur J Sport Sci 18:781–786. https://doi.org/10.1080/17461391.2018.1450899
21. Kais Ü, Pind R, Pehme A, Kaasik P, Mooses M (2019) Pacing strategy of the finishers of the world marathon majors series. Kinesiology 51:22–27
22. March DS, Vanderburgh PM, Titlebaum PJ, Hoops ML (2011) Age, Sex, and Finish Time as Determinants of Pacing in the Marathon. J Strength Cond Res 25:386–391. https://doi.org/10.1519/JSC.0b013e3181bffd0f
23. Hanley B (2016) Pacing, packing and sex-based differences in Olympic and IAAF World Championship marathons. J Sports Sci 34:1675–1681. https://doi.org/10.1080/02640414.2015.1132841
24. Renfree A, Clair Gibson AS (2013) Influence of Different Performance Levels on Pacing Strategy During the Women's World Championship Marathon Race. Int J Sports Physiol Perform 8:279–285. https://doi.org/10.1123/ijspp.8.3.279
25. Cuk I, Nikolaidis PT, Knechtle B (2020) Sex differences in pacing during half-marathon and marathon race. Res Sport Med 28:111–120. https://doi.org/10.1080/15438627.2019.1593835
26. Hoogkamer W, Kram R, Arellano CJ (2017) How Biomechanical Improvements in Running Economy Could Break the 2-hour Marathon Barrier. Sport Med 47:1739–1750. https://doi.org/10.1007/s40279-017-0708-0
27. Hanley B (2015) Pacing profiles and pack running at the IAAF World Half Marathon Championships. J Sports Sci 33:1189–1195. https://doi.org/10.1080/02640414.2014.988742
28. Rotstein A, Inbar O, Berginsky T, Meckel Y (2005) Preferred Transition Speed between Walking and Running: Effects of Training Status. Med Sci Sport Exerc 37:1864–1870. https://doi.org/10.1249/01.mss.0000177217.12977.2f
29. Sentija D, Markovic G (2009) The Relationship Between Gait Transition Speed and the Aerobic Thresholds for Walking and Running. Int J Sports Med 30:795–801. https://doi.org/10.1055/s-0029-1237711
30. Kung SM, Fink PW, Legg SJ, Ali A, Shultz SP (2018) What factors determine the preferred gait transition speed in humans? A review of the triggering mechanisms. Hum Mov Sci 57:1–12. https://doi.org/10.1016/j.humov.2017.10.023
31. Muñoz-Pérez I, Mecías-Calvo M, Crespo-Álvarez J (2018) Estrategias competitivas en maratón. Una aproximación hacia una categorización objetiva. In: Sánchez Molina JA, Torres Tobío G, González Valeiro M (eds) Libro de resúmenes / X Congreso Internacional de la Asociación Española de Ciencias del Deporte. Facultade de Ciencias do Deporte e a Educación Física, A Coruña, p 156
32. Kuhn M (2008) Building Predictive Models in R Using the caret Package. J Stat Softw. https://doi.org/10.18637/jss.v028.i05
33. Bullock GS, Hughes T, Sergeant JC, Callaghan MJ, Riley RD, Collins GS (2021) Clinical Prediction Models in Sports Medicine: A Guide for Clinicians and Researchers. J Orthop Sport Phys Ther 51:517–525. https://doi.org/10.2519/jospt.2021.10697
34. Wickham H, François R, Henry L, Müller K (2023) dplyr: A Grammar of Data Manipulation. Foundation for Statistical Computing, Vienna, Austria. https://dplyr.tidyverse.org
35. Muñoz-Pérez I, Lago-Fuentes C, Mecías-Calvo M, Casado A (2023) Pacing and packing behaviour in elite and world record performances at Berlin marathon. Eur J Sport Sci 23:1472–1479. https://doi.org/10.1080/17461391.2022.2111278
36. Keogh A, Smyth B, Caulfield B, Lawlor A, Berndsen J, Doherty C (2019) Prediction Equations for Marathon Performance: A Systematic Review. Int J Sports Physiol Perform 14:1159–1169. https://doi.org/10.1123/ijspp.2019-0360
37. Tanda G (2022) A simple relationship for predicting marathon performance from training: Is it generally applicable. J Hum Sport Exerc 17:293–301. https://doi.org/10.14198/jhse.2022.172.05
38. Barandun U, Knechtle B, Knechtle P, Klipstein A, Rüst CA, Rosemann T, Lepers R (2012) Running speed during training and percent body fat predict race time in recreational male marathoners. Open Access J Sport Med. https://doi.org/10.2147/OAJSM.S33284
39. Foster C (1983) VO2 max and training indices as determinants of competitive running performance. J Sports Sci 1:13–22. https://doi.org/10.1080/02640418308729657
40. Hagan RD, Upton SJ, Duncan JJ, Gettman LR (1987) Marathon performance in relation to maximal aerobic power and training indices in female distance runners. Br J Sports Med 21:3–7. https://doi.org/10.1136/bjsm.21.1.3
41. Bale P, Rowell S, Colley E (1985) Anthropometric and training characteristics of female marathon runners as determinants of distance running performance. J Sports Sci 3:115–126. https://doi.org/10.1080/02640418508729741
42. Lerebourg L, Saboul D, Clémençon M, Coquart JB (2023) Prediction of Marathon Performance using Artificial Intelligence. Int J Sports Med 44:352–360. https://doi.org/10.1055/a-1993-2371
43. Vickers AJ, Vertosick EA (2016) An empirical study of race times in recreational endurance runners. BMC Sports Sci Med Rehabil 8:26. https://doi.org/10.1186/s13102-016-0052-y
44. Blythe DAJ, Király FJ (2016) Prediction and Quantification of Individual Athletic Performance of Runners. PLoS One 11:e0157257. https://doi.org/10.1371/journal.pone.0157257