



Article

Modeling Physiological Predictors of Running Velocity for Endurance Athletes

Szczepan Wiecha ^{1,*} , Przemysław Seweryn Kasiak ² , Igor Cieśliński ¹, Marcin Maciejczyk ³ ,
Artur Mamcarz ^{4,†} and Daniel Śliż ^{4,*}

- ¹ Department of Physical Education and Health in Biala Podlaska, Faculty in Biala Podlaska, Józef Piłsudski University of Physical Education in Warsaw, 21-500 Biala Podlaska, Poland
- ² Student's Scientific Circle of Lifestyle Medicine, 3rd Department of Internal Medicine and Cardiology, Medical University of Warsaw, 04-749 Warsaw, Poland
- ³ Department of Physiology and Biochemistry, Faculty of Physical Education and Sport, University of Physical Education in Krakow, 31-571 Kraków, Poland
- ⁴ 3rd Department of Internal Medicine and Cardiology, Medical University of Warsaw, 04-749 Warsaw, Poland
- * Correspondence: szczepan.wiecha@awf.edu.pl (S.W.); daniel.sliz@wum.edu.pl (D.Ś.)
- † Artur Mamcarz has been added as a mentor author.

Abstract: Background: Properly performed training is a matter of importance for endurance athletes (EA). It allows for achieving better results and safer participation. Recently, the development of machine learning methods has been observed in sports diagnostics. Velocity at anaerobic threshold (V_{AT}), respiratory compensation point (V_{RCP}), and maximal velocity (V_{max}) are the variables closely corresponding to endurance performance. The primary aims of this study were to find the strongest predictors of V_{AT} , V_{RCP} , V_{max} , to derive and internally validate prediction models for males (1) and females (2) under TRIPOD guidelines, and to assess their machine learning accuracy. **Materials and Methods:** A total of 4001 EA ($n_{males} = 3300$, $n_{females} = 671$; age = 35.56 ± 8.12 years; BMI = 23.66 ± 2.58 kg·m⁻²; $VO_{2max} = 53.20 \pm 7.17$ mL·min⁻¹·kg⁻¹) underwent treadmill cardiopulmonary exercise testing (CPET) and bioimpedance body composition analysis. XGBoost was used to select running performance predictors. Multivariable linear regression was applied to build prediction models. Ten-fold cross-validation was incorporated for accuracy evaluation during internal validation. **Results:** Oxygen uptake, blood lactate, pulmonary ventilation, and somatic parameters (BMI, age, and body fat percentage) showed the highest impact on velocity. For V_{AT} $R^2 = 0.57$ (1) and 0.62 (2), derivation RMSE = 0.909 (1); 0.828 (2), validation RMSE = 0.913 (1); 0.838 (2), derivation MAE = 0.708 (1); 0.657 (2), and validation MAE = 0.710 (1); 0.665 (2). For V_{RCP} $R^2 = 0.62$ (1) and 0.67 (2), derivation RMSE = 1.066 (1) and 0.964 (2), validation RMSE = 1.070 (1) and 0.978 (2), derivation MAE = 0.832 (1) and 0.752 (2), validation MAE = 0.060 (1) and 0.763 (2). For V_{max} $R^2 = 0.57$ (1) and 0.65 (2), derivation RMSE = 1.202 (1) and 1.095 (2), validation RMSE = 1.205 (1) and 1.111 (2), derivation MAE = 0.943 (1) and 0.861 (2), and validation MAE = 0.944 (1) and 0.881 (2). **Conclusions:** The use of machine-learning methods allows for the precise determination of predictors of both submaximal and maximal running performance. Prediction models based on selected variables are characterized by high precision and high repeatability. The results can be used to personalize training and adjust the optimal therapeutic protocol in clinical settings, with a target population of EA.

Keywords: velocity; respiratory compensation point; anaerobic threshold; endurance training; running; prediction models; machine learning



Citation: Wiecha, S.; Kasiak, P.S.; Cieśliński, I.; Maciejczyk, M.; Mamcarz, A.; Śliż, D. Modeling Physiological Predictors of Running Velocity for Endurance Athletes. *J. Clin. Med.* **2022**, *11*, 6688. <https://doi.org/10.3390/jcm11226688>

Academic Editor: Birna Bjarnason-Wehrens

Received: 26 September 2022

Accepted: 9 November 2022

Published: 11 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The benefits of regular physical exercise are widely debated and include reducing the risk of obesity [1] or cardiovascular diseases [2]. On the other hand, improperly performed training with excessive intensity may negatively affect the organism's homeostasis and increase the risk of injury [3].

The concept of anaerobic threshold (AT) is widely discussed in exercise physiology [4]. As envisioned by Karlman Wasserman, the AT linked the increase in blood lactate concentration ($[La^-]_b$), during a strenuous incremental cardiopulmonary exercise test (CPET), with an excess arterial CO_2 accumulation and its further pulmonary output [5]. Above the AT, $[La^-]_b$ increase leads to temporary acidosis. The endurance capacity of the whole system is usually sufficiently high to cope with the incoming state [6]. During steady-state exercise with intensity above the AT, an equilibrium in $[La^-]_b$ appearance and its elimination is observed [6]. Thus, AT is a useful practical indicator to provide personalized training recommendations (with the aim of adjusting exercise intensity to set goals) and load monitoring [7].

The respiratory compensation point (RCP) is the intensity at which arterial CO_2 begins to decrease during demanding activity due to breathing capacity [8]. Above the RCP, the intensity of acidic ion accumulation exceeds their systemic or respiratory elimination abilities and indicates reduced endurance capacity. This leads to an over-decrease in the serum pH during graded exercise. This threshold indicates how long a high-intensity effort can be sustained [9].

The velocity at the anaerobic threshold (V_{AT}), at the respiratory compensation point (V_{RCP}), and at its maximum (V_{max}) play an essential position in the endurance performance assessment, both for professional and recreational endurance athletes (EA), as well as for the general population under clinical conditions [10,11].

These variables are the shift points of aerobic exercise to anaerobic metabolism and can be used as one of the parameters to evaluate the maximum endurance capacity [7]. Moreover, they closely positively correlate with exercise abilities [4]. They could be incorporated into the prescription for the advancement of training plans [7] or competition strategies [12] for special and narrow populations (e.g., EA), and in sports diagnostics whenever controlled running intensity is required (i.e., in clinical CPET) [7]. Furthermore, currently, these variables most closely correspond to the EA critical power sustainability [13,14].

Apart from V_{max} , maximal aerobic speed (MAS), which is directly related to VO_{2max} , is another important aspect of overall performance evaluation. However, as the aim of this research is to predict V_{max} , we recommend that further studies should be performed to analyze the MAS.

Numerous parameters, such as heart rate (HR), oxygen uptake (VO_2), or anthropometric data (i.e., height, age, and gender), are widely discussed in the development of multivariable prediction models that provide an increasingly more suitable alternatives to direct CPET measurements [15].

Several studies have attempted to develop and validate various non-invasive prediction equations for different sports performance measurements (i.e., for HR, VO_2 , and others) [11,15,16]. However, they were mostly conducted on general populations or on small athletic samples, and thus, they can only be extrapolated to a low degree [17]. In addition, their methodology is widely variable, and only a few of them fulfilled recommended TRIPOD guidelines [18]. Thus, the actual number of V_{AT} , V_{RCP} , and V_{max} predictive models is limited, despite being significant measures of endurance capacity [19,20]. Moreover, although the variables influencing the running performance are well researched, the authors have not yet assessed how accurately they can be used to estimate running velocity by further including them in prediction models.

The aims of this study were: (1) to find the somatic and CPET variables that are the most responsible for running velocity, (2) to develop a prediction method for V_{AT} , V_{RCP} , and V_{max} , in accordance with TRIPOD recommendations [18], (3) to internally validate the obtained formulae, (4) to assess the accuracy of the current machine-learning abilities to predict running velocity based on the primarily determined variables, and (5) to evaluate practical applications of such an approach in sports or clinical conditions based on actual knowledge regarding exercise physiology.

2. Materials and Methods

The TRIPOD guidelines [18] have been applied to this research (see Supplementary Material S1, TRIPOD Checklist for Prediction Model Development and Validation). This was a retrospective data analysis from the registry of CPET performed in the years 2013–2021 at a tertiary care sports diagnostic clinic (SportsLab Clinic, Warsaw, Poland, www.sportslab.pl).

2.1. Ethical Approval

The study protocol was approved by the Institutional Review Board of the Bioethical Committee at the Medical University of Warsaw (AKBE/32/2021) and met the necessary regulations of the Declaration of Helsinki. Mandatory written consents to undergo incremental CPET were obtained from each EA before participating in the study. Written informed consent for participation was not required for this study, in accordance with the national legislation and the institutional requirements.

2.2. Study Design

Participants were endurance runners who underwent CPET on the treadmill (TE). The tests were performed according to individual requests of the EA as an element of the training program prescription. Preliminary inclusion criteria were (1) age ≥ 18 years, and (2) $\leq \pm 3$ SD from mean for all of the tested variables included in Table 1 (only the lowest or the weakest extreme outliers were excluded). Exclusion criteria were (1) CPET was not performed on the TE, (2) any acute or chronic medical condition (including the musculoskeletal system, or addictions), (3) ongoing pharmacological treatment, and (4) smoking. To ensure that each subject reached maximum effort during the CPET, we applied the additional selection protocol consisting of the fulfillment ≥ 6 of the following criteria: (1) plateau in VO_2 (growth $< 100 \text{ mL}\cdot\text{min}^{-1}$ in VO_2 with exercise workload increase), (2) respiratory exchange ratio (RER) ≥ 1.10 , (3) respiratory frequency (fR) $\geq 45 \text{ breaths}\cdot\text{min}^{-1}$, (4) reported exertion during CPET ≥ 18 , according to the Borg scale, (5) $[\text{La}^-]_{\text{b}} \geq 8 \text{ mmol}\cdot\text{L}^{-1}$, (6) increase in speed $\geq 10\%$ of its RCP value post-exceeding the RCP, and (7) peak HR $\geq 15 \text{ beats}\cdot\text{min}^{-1}$ under predicted maximal HR [21]. For the entire selection procedure, along with the exclusion data, see Figure 1.

Table 1. Participants' basic anthropometric characteristics.

Variable (Unit)	Male [<i>n</i> = 3330; 83.23%]	Female [<i>n</i> = 671; 16.77%]	<i>p</i> -Value
Age (years)	35.90 (8.15)	33.86 (7.74)	<0.0001
Height (cm)	179.58 (6.22)	167.19 (6.88)	<0.0001
BM (kg)	77.72 (9.47)	60.60 (8.73)	<0.0001
BMI ($\text{kg}\cdot\text{m}^{-2}$)	24.07 (2.44)	21.64 (2.38)	<0.0001
BF (%)	15.49 (4.53)	22.04 (5.46)	<0.0001
FM (kg)	12.29 (4.71)	13.47 (4.65)	<0.0001
FFM (kg)	65.42 (6.47)	47.08 (6.36)	<0.0001

Abbreviations: BM, body mass; BMI, body mass index; BF, body fat; FM, fat mass; FFM, fat-free mass. The continuous value is presented as mean (SD), while the categorical value is shown as numbers (%), when appropriate. Comparisons between subgroups (*p*-value) were obtained by Student's *t*-test for independent variables.

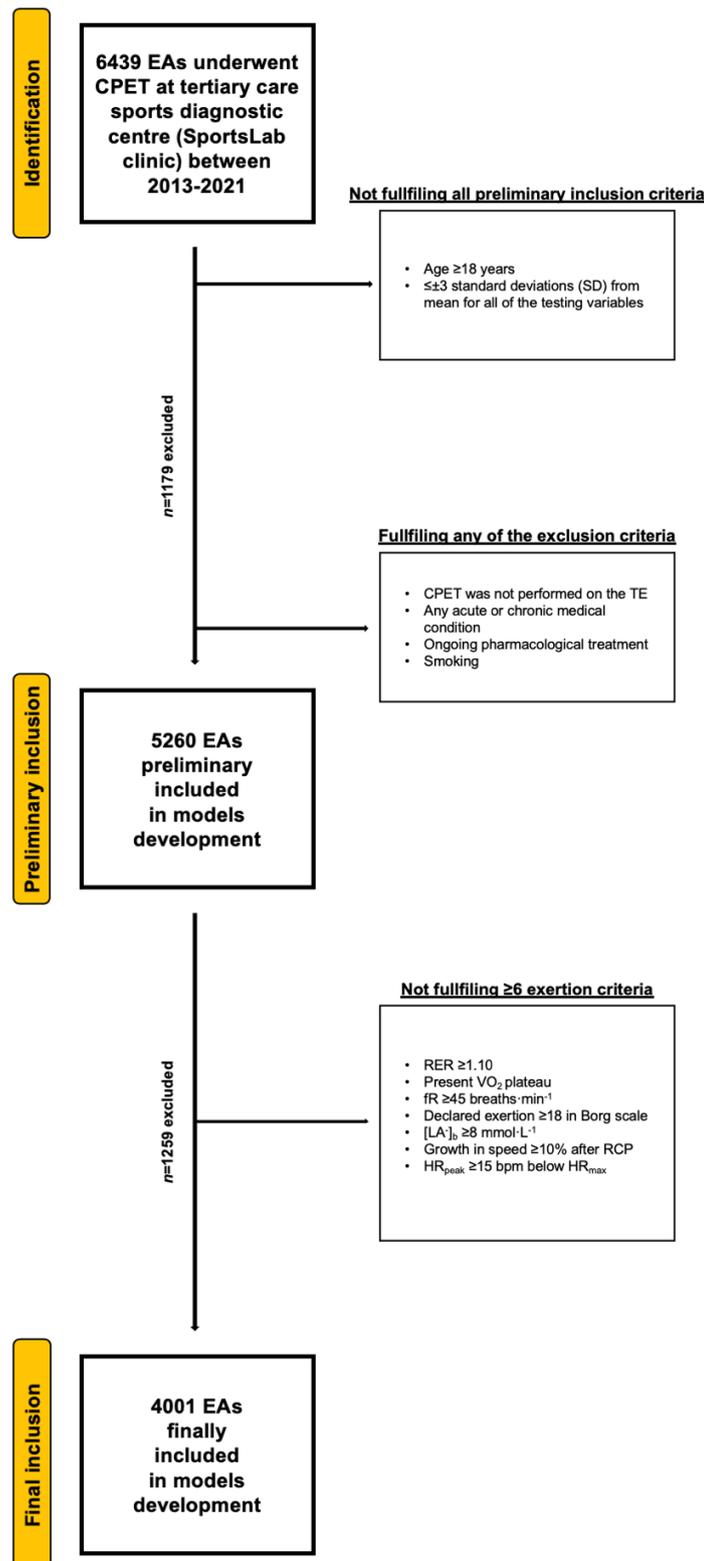


Figure 1. Flowchart of the preliminary inclusion and exclusion process. Abbreviations: EA, endurance athlete; CPET, cardiopulmonary exercise testing; SD, standard deviation; TE, treadmill; RER, respiratory exchange ratio; VO_2 , oxygen uptake ($\text{mL} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$); $[\text{La}^-]_b$, lactate concentration ($\text{mmol} \cdot \text{L}^{-1}$); fR , breathing frequency ($\text{breaths} \cdot \text{min}^{-1}$); RCP, respiratory compensation point; HR_{peak} , peak heart rate ($\text{beats} \cdot \text{min}^{-1}$); HR_{max} , maximal heart rate (bpm). At both stages of the selection, some participants met several (>1) exclusion criteria.

2.3. Somatic, $[La^-]_b$ Measurements, and CPET Protocol

First, body mass (BM) stratified by body fat (BF) and fat-free mass (FFM) measurements were obtained via the bioimpedance method (BIA) using a body composition (BC) analyzer (Tanita, MC 718, Tokyo, Japan) with the multifrequency of 5 kHz/50 kHz/250 kHz. Conditions during BC and CPET were the same: 40 m² indoor, air-conditioned area, 40–60% humidity, temperature 20–22 °C, altitude 100 m ASL, and the subjects had their skin cleaned before testing. In our standardized laboratory practice, each EA had received recovery and dietary instructions via email a few days prior to testing to enable them to prepare appropriately for the CPET and BC tests. Our recommendations included: eating a high carbohydrate meal 2–3 h before the CPET and staying hydrated with sports drinks, and female EAs were advised to be well beyond their menstrual phase [22]. They also received information stating that the CPET would be performed on a mechanical TE and that they should be familiar with the characteristics of this type of effort, as well as the running technique involved.

Running tests were performed on a mechanical TE (h/p/Cosmos quasar, Nussdorf-Traunstein, Germany). CPET indices were measured using the breath-by-breath method during 15 s intervals [23], utilizing a Hans Rudolph V2 Mask (Hans Rudolph, Inc., Shawnee, KS, USA), a gas exchange analyzing device Cosmed Quark CPET (Rome, Italy), and specialized software (Quark PFT Suite powered by Omnia 10.0E). The gas analyzer device was calibrated prior to the testing protocol (16% O₂; 5% CO₂; ventilation accuracy $\pm 2\%$ or 100 mL·min⁻¹). The analyzer measurement mode takes into account the manufacturer's standard settings, i.e., 3-step smoothing and removing erroneous breaths from the analysis. HR was measured through the ANT+ and torso belt as a part of the Cosmed Quark set (accuracy similar to ECG; ± 1 beats·min⁻¹). $[La^-]_b$ was examined using a Super GL2 analyzer (Müller Gerätebau GmbH, Freital, Germany) employing an enzymatic-amperometric electrochemical technique. The lactate analyzer was also calibrated before each round of analysis for each participant.

CPET began with a 5 min preparatory protocol (walking or slow running at a declared "conversation" pace). The primary speed was 7–12 km·h⁻¹ at a 1% inclination (the differences in the starting pace resulted from the training level of the participants and were selected on the basis of an interview on their previous sports results). The pace was increased by 1 km·h⁻¹ every 2 min. VO₂ or HR plateau (no increase in VO₂ or HR with an increase in CPET intensity) or volitional inability to maintain intensity was the moment when the test was terminated [23,24]. Subjects were encouraged verbally to make a maximum effort. HR was considered the highest value at CPET (not averaged). Maximal VO₂ was recorded as an average from stable VO₂ in 10 s intervals directly before the termination of the CPET [23,24]. AT and RCP were assessed via non-direct methods based on the ventilatory concept. AT was achieved if the following measures were fulfilled: (1) VE/VO₂ curve started to grow with the constant VE/VCO₂ curve and (2) end-tidal partial pressure of O₂ started to grow with the constant end-tidal partial pressure of CO₂ [25]. RCP was achieved if the following measures were fulfilled: (1) a reduction in partial pressure of end-tidal CO₂ (PetCO₂) after attaining a maximal intensity; (2) a fast nonlinear growth in VE (second deflection); (3) the VE/VCO₂ ratio achieved the lowest value and started to grow; and (4) a nonlinear growth in VCO₂ versus VO₂ (linearity divergence) was achieved [25]. $[La^-]_b$ was assessed by obtaining a 20 μ L blood sample from a fingertip: before the test, after any speed increase, and 3 min after termination. A sample for $[La^-]_b$ analysis was taken during running without interruption or pace decrease. Each time, the sample was from the same initial puncture. The first few drops were drained onto a swab and the proper blood sample was drawn. In further analysis, the corresponding values of $[La^-]_b$ for AT, RCP, and maximal VO₂ were determined.

2.4. Data Analysis and Predictors Selection

Data were saved into an Excel file (Microsoft Corporation, Redmond, WA, USA) and Python script. Further, they were calculated according to frequency (percentage) and mean (\pm standard deviation; SD, or 95% confidence intervals; CI) for continuous variables and the median for categorical variables. Intergroup differences (each was a continuous variable) were calculated using the Student t-test for independent variables. If there were lacking data (only for $[La^-]_b$; in 1190 cases for males and 266 cases for females in total), imputation with the random forest method (RF) was applied to fill in the gaps [26,27].

The XGBoost machine learning approach was used to select variables with the highest prediction value [28]. In order to select the variables, the population was divided into 3 groups: 60% for derivation (building group), 20% for testing (testing group), and the remaining 20% for validation (validation group). After selection, 11 variables were included in the further analysis: VO_{2max} , VO_{2RCP} , VO_{2AT} , $[La^-]_{bRCP}$, $[La^-]_{bAT}$, VE_{max} , VE_{RCP} , age, BM, BMI, and BF. Next, selected parameters were input into multiple linear regression (MLR) modeling. As a result, only significant predictors (with $p < 0.05$) were included in the final models. The derived equations are characterized by the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). A 10-fold cross-validation technique [29] and the Bland–Altman plots analysis [30] were used to establish the model's precision and accuracy during internal validation. To clarify, in the 10-fold cross-validation, the population is divided into 10 random parts. The candidate model is built on $[10 - 1 = 9]$ training sets; then, the derived model is evaluated on the test set consisting of the remaining one part. By respectively conducting building procedures on training sets and validation on the test set 10 times, we chose the final formula with the lowest possible inaccuracy validation score (defined in this paper as the lowest RMSE and MAE) [29]. Other implemented tests to reach the complete fulfillment of MLR modeling requirements include Ramsey's RESET test (for the correctness of specificity in MLR equations), the Chow test (for stability assessment between different coefficients), and the Durbin–Watson test (for autocorrelation of residuals). Each model was examined under the above-mentioned requirements and any irregularities have not been noted.

Our comprehensive machine learning approach enables the evaluation of each formula according to preliminary variable precision (at the stage of selection), accuracy (during model building), and recall (in internal validation).

The Ggplot 2 package (version-6.0-90; Available from: <https://cran.r-project.org/web/packages/caret/index.html>, accessed on 21 June 2022) in RStudio (R Core Team, Vienna, Austria; version 3.6.4), GraphPad Prism (GraphPad Software; San Diego, CA, USA; version 9.0.0 for Mac OS), and STATA software (StataCorp, College Station, TX, USA; version 15.1) were used for statistical analysis. A two-sided p -value < 0.05 was considered as the significance borderline.

3. Results

3.1. Somatic Measurements and CPET Results

The participants' anthropometric data are presented in Table 1. The full population consisted of 4001 people, of which 3330 (83.23%) were male and 671 (16.77%) were female. All data showed a normal distribution. The mean age was 35.90 ± 8.15 years for males and 33.86 ± 7.74 years for females and the overall age ranged from 18 up to 74 years. BMI was $24.07 \pm 2.44 \text{ kg}\cdot\text{m}^{-2}$ for men, while women had $21.64 \pm 2.38 \text{ kg}\cdot\text{m}^{-2}$. BF percentage was relatively low, estimate at 15.49 ± 4.53 in males and 22.04 ± 5.46 in females. Significant differences between genders has been observed for height ($p < 0.0001$), BM ($p < 0.0001$), BMI ($p < 0.0001$), BF ($p < 0.0001$), and FFM ($p < 0.0001$).

CPET results are presented in Table 2. Among other measured variables, V_{AT} was $10.97 \pm 1.40 \text{ km}\cdot\text{h}^{-1}$ and $9.64 \pm 1.36 \text{ km}\cdot\text{h}^{-1}$ for males and females, respectively. V_{RCP} was $14.02 \pm 1.74 \text{ km}\cdot\text{h}^{-1}$ and $12.29 \pm 1.68 \text{ km}\cdot\text{h}^{-1}$ for males and females, respectively. The V_{max} obtained during CPET was $16.07 \pm 1.93 \text{ km}\cdot\text{h}^{-1}$ and $14.12 \pm 1.85 \text{ km}\cdot\text{h}^{-1}$ for males and females, respectively. The starting protocol velocity was $8.61 \pm 1.28 \text{ km}\cdot\text{h}^{-1}$ for males

and $7.60 \pm 1.08 \text{ km}\cdot\text{h}^{-1}$ for females. When comparing both genders, significant differences (all $p < 0.0001$) were found for all the measured variables except $[\text{La}^-]_{\text{b}}$ at AT ($p = 0.99$), maximal respiratory exchange ratio ($p = 0.77$), and maximal HR ($p = 0.15$).

Table 2. CPET characteristics.

Variable (Unit)	Males [<i>n</i> = 3330]			Females [<i>n</i> = 671]			<i>p</i> -Value
	Mean	CI	SD	Mean	CI	SD	
VO _{2AT} (mL·min ⁻¹ ·kg ⁻¹)	38.42	38.25–38.59	4.96	35.69	35.32–36.05	4.83	<0.0001
VO _{2AT} (mL·min ⁻¹)	2955.15	2942.04–2968.26	385.79	2137.77	2113.25–2162.29	323.48	<0.0001
RER _{AT}	0.87	0.87–0.87	0.04	0.86	0.85–0.86	0.04	<0.0001
HR _{AT} (beats·min ⁻¹)	151.32	150.96–151.68	10.70	156.45	155.66–157.24	10.39	<0.0001
VE _{AT} (L·min ⁻¹)	78.26	77.84–78.68	12.02	58.38	57.66–59.09	9.25	<0.0001
fR _{AT} (breaths·min ⁻¹)	34.88	34.61–35.14	7.85	34.89	34.31–35.47	7.66	<0.0001
[La ⁻] _{bAT} (mmol·L ⁻¹);	1.95	1.92–1.98	0.67	1.86	1.80–1.93	0.66	0.99
VO _{2RCP} (mL·min ⁻¹ ·kg ⁻¹)	47.59	47.37–47.81	6.10	43.05	42.56–43.55	6.14	<0.0001
VO _{2RCP} (mL·min ⁻¹)	3642.72	3626.90–3658.54	465.70	2576.01	2545.15–2606.87	407.12	<0.0001
RER _{RCP}	1.00	1.00–1.00	0.04	0.99	0.99–1.00	0.04	<0.0001
HR _{RCP} (beats·min ⁻¹)	173.43	173.12–173.75	9.33	176.04	175.34–176.73	9.12	<0.0001
VE _{RCP} (L·min ⁻¹)	113.82	113.25–114.39	16.43	81.15	80.20–82.11	12.34	<0.0001
fR _{RCP} (breaths·min ⁻¹)	44.19	43.91–44.48	8.52	43.09	42.49–43.68	7.87	<0.0001
[La ⁻] _{bRCP} (mmol·L ⁻¹);	4.53	4.49–4.58	1.07	4.19	4.09–4.29	1.02	<0.0001
VO _{2max} (mL·min ⁻¹ ·kg ⁻¹)	54.10	53.87–54.34	6.93	48.73	48.23–49.24	6.67	<0.0001
VO _{2max} (mL·min ⁻¹)	4176.37	4157.64–4195.09	551.09	2949.02	2911.51–2986.54	494.89	<0.0001
RER _{max}	1.12	1.12–1.12	0.04	1.12	1.12–1.12	0.04	0.76
HR _{max} (beats·min ⁻¹)	184.81	184.49–185.13	9.54	185.39	184.69–186.09	9.24	0.15
VE _{max} (L·min ⁻¹)	148.86	148.15–149.57	20.46	103.83	102.60–105.05	15.86	<0.0001
fR _{max} (breaths·min ⁻¹)	57.59	57.28–57.90	9.20	55.46	54.83–56.09	8.30	<0.0001
[La ⁻] _{bmax} (mmol·L ⁻¹);	9.91	9.82–10.00	2.02	9.08	8.88–9.28	1.93	<0.0001
V _{AT} (km·h ⁻¹)	10.97	10.92–11.02	1.40	9.64	9.53–9.74	1.36	<0.0001
V _{RCP} (km·h ⁻¹)	14.02	13.96–14.08	1.74	12.29	12.16–12.41	1.68	<0.0001
V _{max} (km·h ⁻¹)	16.07	16.01–16.14	1.93	14.12	13.98–14.26	1.85	<0.0001
V _S (km·h ⁻¹)	8.61	8.56–8.66	1.28	7.60	7.51–7.69	1.08	<0.0001

Abbreviations: CI, 95% confidence interval; SD, standard deviation; VO_{2AT}, oxygen uptake at anaerobic threshold; RER_{AT}, respiratory exchange ratio at anaerobic threshold; HR_{AT}, heart rate at anaerobic threshold; VE_{AT}, pulmonary ventilation at anaerobic threshold; fR_{AT}, respiratory frequency at anaerobic threshold; [La⁻]_{bAT}, lactate concentration at anaerobic threshold; VO_{2RCP}, oxygen uptake at respiratory compensation point; RER_{RCP}, respiratory exchange ratio at respiratory compensation point; HR_{RCP}, heart rate at respiratory compensation point; VE_{RCP}, pulmonary ventilation at respiratory compensation point; fR_{RCP}, respiratory frequency at respiratory compensation point; [La⁻]_{bmax}, lactate concentration at respiratory compensation point; VO_{2max}, maximal oxygen uptake; RER_{max}, maximal respiratory exchange ratio; HR_{max}, maximal heart rate; VE_{max}, maximal pulmonary ventilation; fR_{max}, maximal respiratory frequency; [La⁻]_{bmax}, maximal lactate concentration; V_{AT}, velocity at anaerobic threshold; V_{RCP}, velocity at respiratory compensation point; V_{max}, maximal velocity; V_S, protocol starting velocity. Comparisons between subgroups (*p*-value) were obtained by Student *t*-test for independent variables.

3.2. Prediction Models for V_{AT}, V_{RCP}, and V_{max}

Complete MLR prediction models for males and females are presented in Table 3 (left columns), while Figure 2 shows their performance in the derivation cohort (illustrated as an analysis of observed vs. predicted values). The importance of all CPET variables, based on XGBoost selection, included in the modeling is presented in Figure 3. The following variables showed the strongest impact in building the models: VO₂, [La⁻], VE age, and BMI. Model performance is presented as R², along with RMSE and MAE. Briefly, R² for male equations ranged from 0.57 for V_{AT} and V_{max} to 0.62 for V_{RCP}. For female formulae, R² ranged from 0.62 for V_{AT} to 0.67 for V_{RCP}. The obtained RMSE was the lowest for the female V_{AT} equation (=0.828) and the highest for the male V_{max} (=1.202), while the observed MAE was the lowest for the female V_{AT} equation (=0.657) and the highest for male V_{max} (=0.944).

Table 3. Running velocity prediction equations stratified by gender.

Model Category	Multiple Linear Regression Equation	R ²	Derivation Group Performance		Validation Group Performance	
			RMSE	MAE	RMSE	MAE
V _{AT} Males	$8.00 - 0.01 \cdot \text{Age} - 0.09 \cdot \text{BMI} + 0.04 \cdot \text{VO}_{2\text{max}} + 0.09 \cdot \text{VO}_{2\text{AT}} - 0.65 \cdot [\text{La}^-]_{\text{bAT}} + 0.01 \cdot \text{VE}_{\text{RCP}}$	0.57	0.909	0.708	0.913	0.710
V _{AT} Females	$7.55 - 0.02 \cdot \text{Age} - 0.10 \cdot \text{BMI} + 0.15 \cdot \text{VO}_{2\text{AT}} - 0.70 \cdot [\text{La}^-]_{\text{bAT}} + 0.01 \cdot \text{VE}_{\text{RCP}}$	0.62	0.828	0.657	0.838	0.665
V _{RCP} Males	$10.88 - 0.02 \cdot \text{Age} - 0.11 \cdot \text{BMI} + 0.04 \cdot \text{VO}_{2\text{max}} - 0.99 \cdot [\text{La}^-]_{\text{bAT}} + 0.10 \cdot \text{VO}_{2\text{RCP}} + 0.01 \cdot \text{VE}_{\text{RCP}} + 0.10 \cdot [\text{La}^-]_{\text{bRCP}}$	0.62	1.066	0.832	1.070	0.835
V _{RCP} Females	$9.24 - 0.02 \cdot \text{Age} - 0.11 \cdot \text{BMI} - 1.05 \cdot [\text{La}^-]_{\text{bAT}} + 0.15 \cdot \text{VO}_{2\text{RCP}} + 0.01 \cdot \text{VE}_{\text{RCP}} + 0.19 \cdot [\text{La}^-]_{\text{bRCP}}$	0.67	0.964	0.752	0.978	0.763
V _{max} Males	$12.41 - 0.03 \cdot \text{Age} + 0.01 \cdot \text{BM} - 0.12 \cdot \text{BMI} + 0.10 \cdot \text{VO}_{2\text{max}} - 0.82 \cdot [\text{La}^-]_{\text{bAT}} + 0.07 \cdot \text{VO}_{2\text{RCP}}$	0.57	1.202	0.943	1.205	0.944
V _{max} Females	$9.37 - 0.03 \cdot \text{Age} + 0.06 \cdot \text{VO}_{2\text{max}} - 0.79 \cdot [\text{La}^-]_{\text{bAT}} + 0.09 \cdot \text{VO}_{2\text{RCP}} + 0.01 \cdot \text{VE}_{\text{max}} - 0.04 \cdot \text{BF}$	0.65	1.095	0.861	1.111	0.881

Abbreviations: RMSE, root mean square error; MAE, mean absolute error; R², adjusted R²; V_{AT}, velocity at anaerobic threshold; Age, age in years; BMI, body mass index (kg·m⁻²); VO_{2max}, relative maximum oxygen uptake (mL·min⁻¹·kg⁻¹); VO_{2AT}, relative oxygen uptake at anaerobic threshold (mL·min⁻¹·kg⁻¹); [La⁻]_{bAT}, blood lactate concentration at anaerobic threshold (mmol·L⁻¹); VE_{RCP}, pulmonary ventilation at RCP (L·min⁻¹); V_{RCP}, velocity at respiratory compensation point; VO_{2RCP}, relative oxygen uptake at respiratory compensation point (mL·min⁻¹·kg⁻¹); [La⁻]_{bRCP}, blood lactate concentration at respiratory compensation point; BM, body mass; V_{max}, maximal velocity; VE_{max}, maximal pulmonary ventilation (L·min⁻¹). RMSE and MAE are explained in km·h⁻¹. Model performance at the stage of derivation has been shown in the left columns. Briefly, our equations showed high accuracy and explained approximately 60–70% of the differences between participants. The results of internal validation via the 10-fold cross technique are presented in the right columns, and they showed a precise transferability, despite a limited sample size for internal validation. We are presenting one R² because of the 10-fold cross-validation for the same group of participants as the derived validation.

3.3. Internal Validation

The evaluation of each model is also presented in Table 3 (right columns). In summary, the performance of our prediction equations was similar to that observed in the derivation cohort. A slightly higher RMSE and MAE were noted. Overall, RMSE values are located between 0.838–1.205 km·h⁻¹ and MAE between 0.665–0.944 km·h⁻¹. The best working model (defined as having the highest replicability and the lowest risk of inaccuracies in the test set) was for V_{AT} for females (RMSE = 0.838, MAE = 0.665), and the worst was for males V_{max} (RMSE = 1.205, MAE = 0.944). The most and least accurate models were the same in regards to the derivation and validation. Figure 4 illustrates the Bland–Altman plots, with a comparison of observed vs. predicted velocity using newly derived prediction models at the stage of validation.

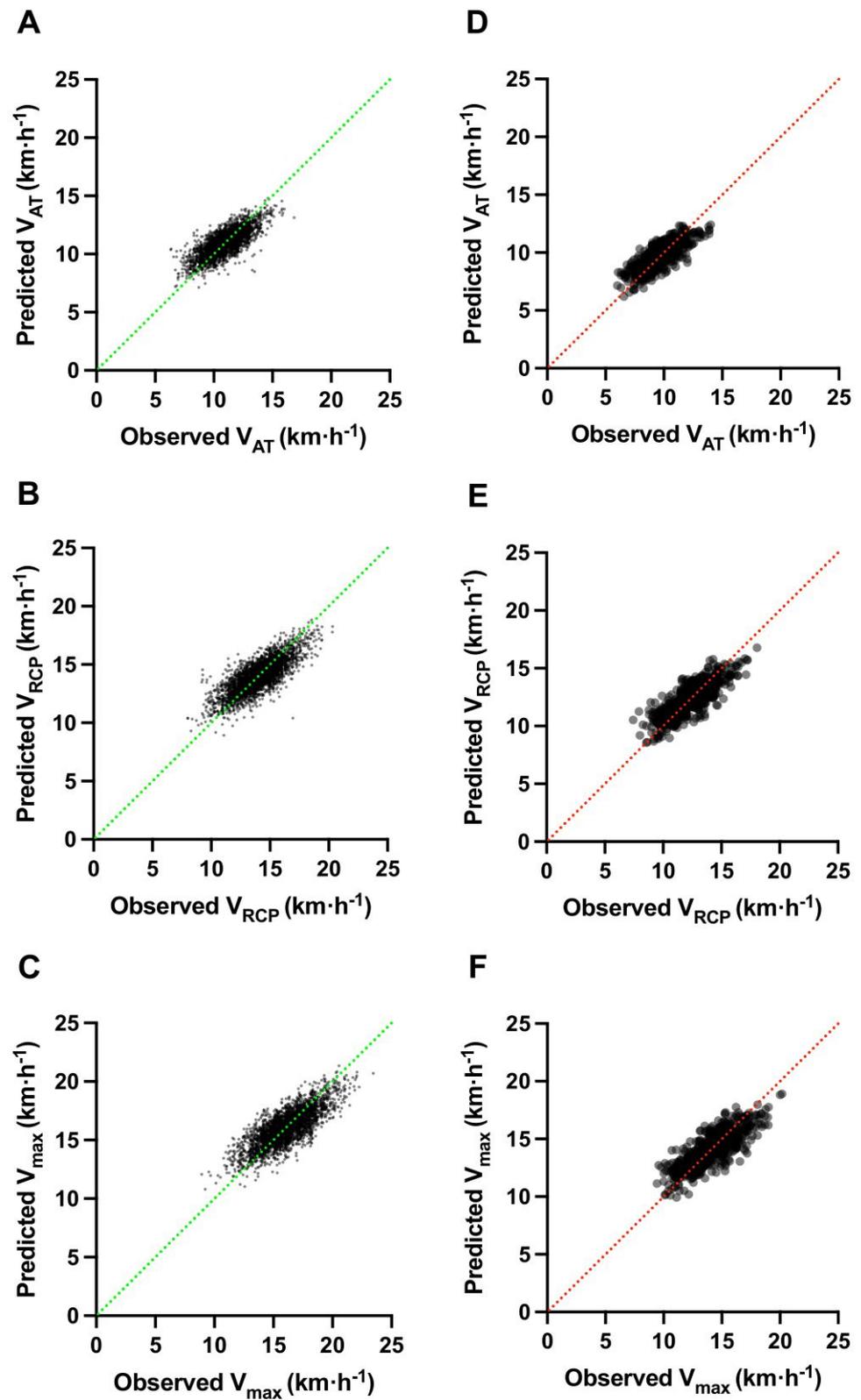


Figure 2. Performance of novel prediction equations for treadmill velocity. Abbreviations: V_{AT}, velocity at anaerobic threshold; V_{RCP}, velocity at respiratory compensation point; V_{max}, maximal velocity. Colored dotted lines illustrate a 1:1 correspondence between measured and predicted velocities, respectively green for males (left row; (A–C) panels) and red for females (right row; (D–F) panels).

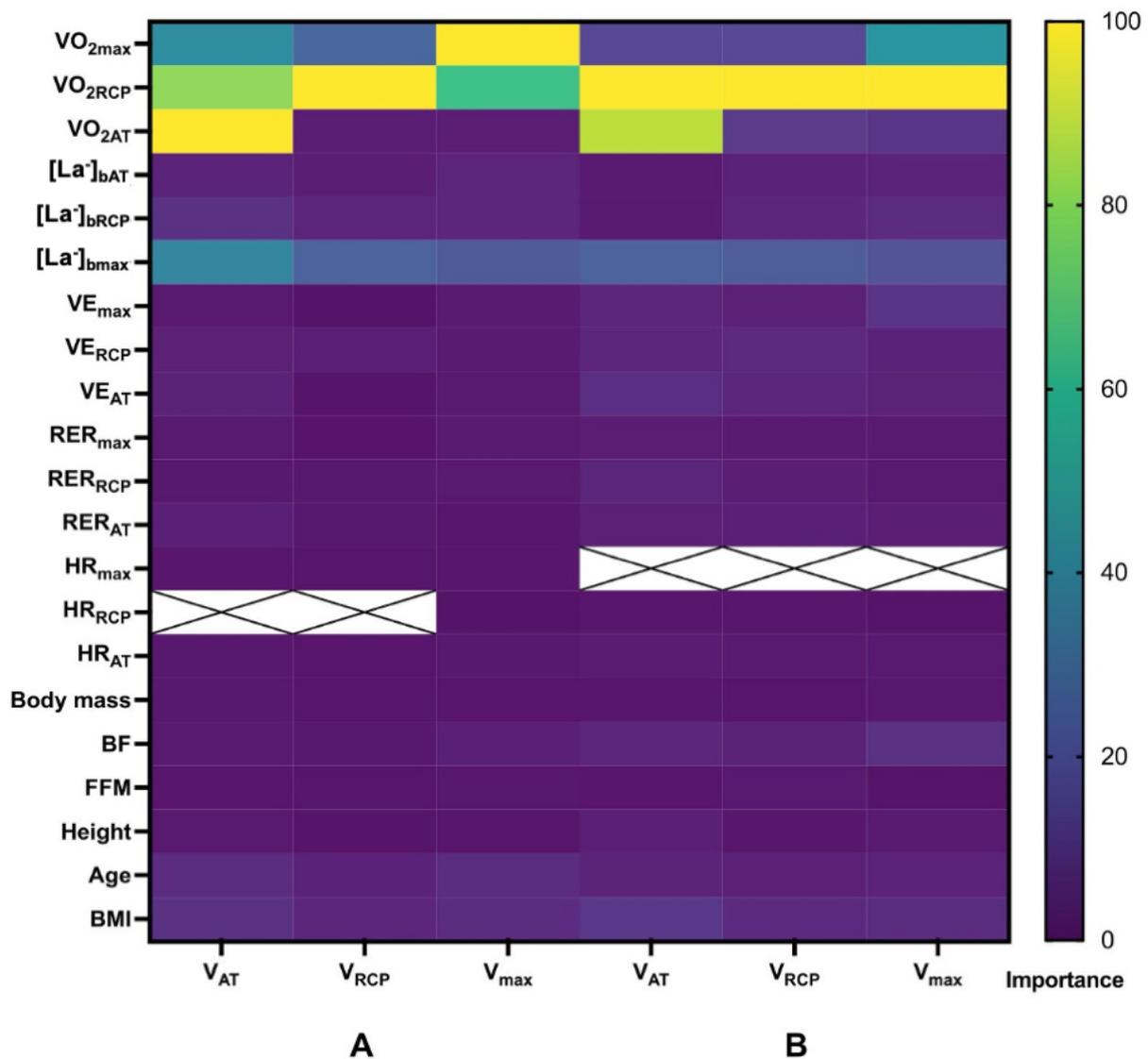


Figure 3. Heat map showing the importance variables regarding predicted velocity based on XGBoost selection. Abbreviation: VO_{2max} , maximal oxygen uptake; VO_{2RCP} , relative oxygen uptake at respiratory compensation point; VO_{2AT} , relative oxygen uptake at anaerobic threshold; $[La^-]_{bRCP}$, blood lactate concentration at respiratory compensation point; $[La^-]_{bAT}$, blood lactate concentration at anaerobic threshold; $[La^-]_{bmax}$, maximal blood lactate concentration; VE_{max} , maximal pulmonary ventilation; VE_{RCP} , pulmonary ventilation at respiratory compensation point; VE_{AT} , pulmonary ventilation at anaerobic threshold; RER_{max} , maximal respiratory exchange ratio; RER_{RCP} , respiratory exchange ratio at respiratory compensation point; RER_{AT} , respiratory exchange ratio at anaerobic threshold; HR_{max} , maximal heart rate; HR_{RCP} , heart rate at respiratory compensation point; HR_{AT} , heart rate at anaerobic threshold; BF, body fat; FFM, fat-free mass; BMI, body mass index; V_{AT} , velocity at anaerobic threshold; V_{RCP} , velocity at respiratory compensation point; V_{max} , maximal velocity. Panel (A) presents data for males, while panel (B) shows the data for females. The cross means that the variable has not fulfilled preliminary selection-stage requirements (only in HR). The maps present a variable’s importance regarding the predicted velocity during the model-building stage. In the final prediction models, only the variables with significant impact ($p < 0.05$) were included.

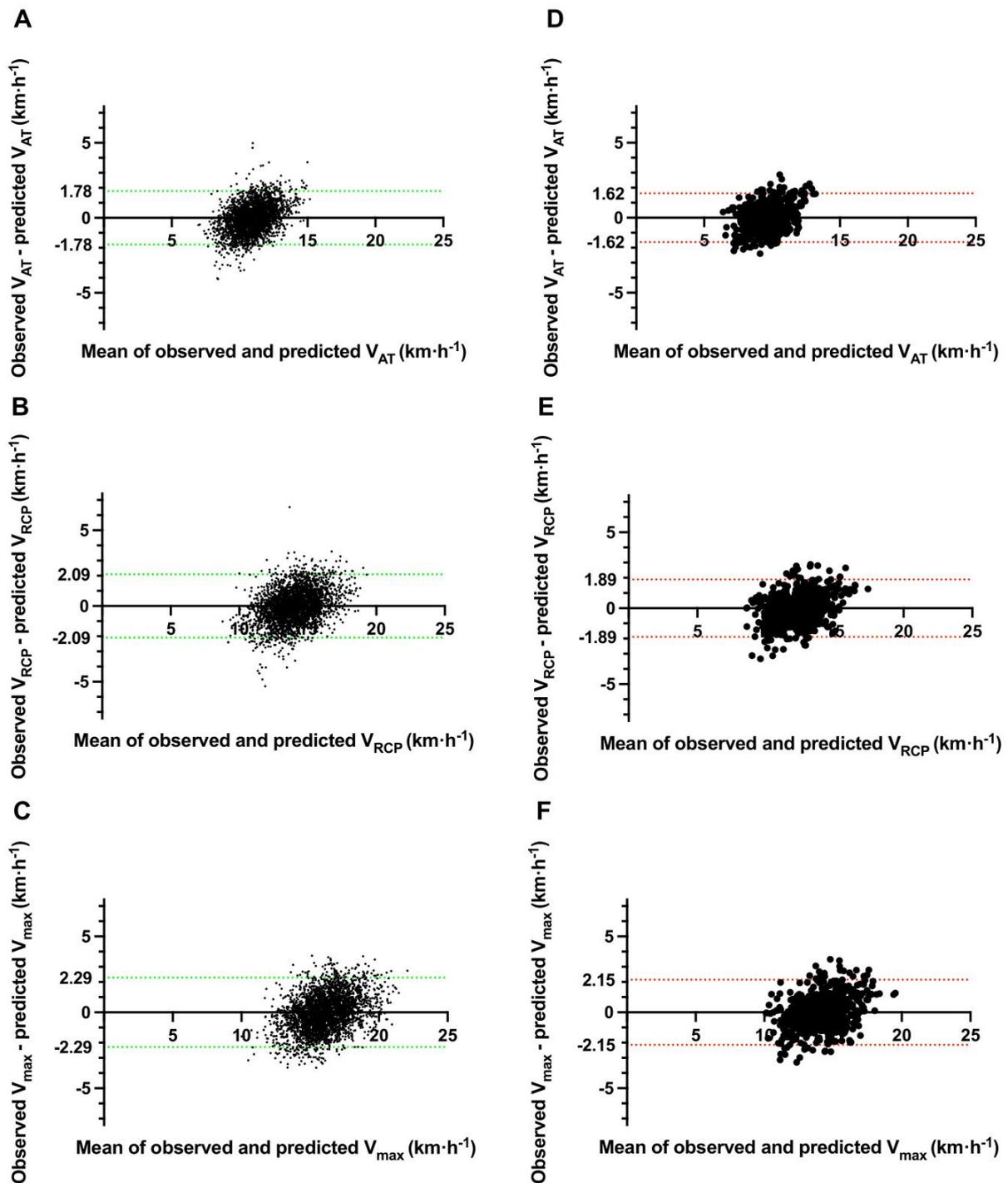


Figure 4. Bland–Altman Plots comparing observed with predicted velocity during internal validation. Abbreviations: V_{AT} , velocity at anaerobic threshold; V_{RCP} , velocity at respiratory compensation point; V_{max} , maximal velocity. Colored dotted lines present a 95% confidence interval of agreement, green for males (left row; (A–C) panels) and red for females (right row; (D–F) panels), respectively.

4. Discussion

In the current study, we applied advanced machine-learning properties in a comprehensive evaluation of running physiology. The obtained equations include several physiological-only measures (both anthropometric and directly measured during CPET) to provide a feasible utility for the prediction of V_{AT} , V_{RCP} , and V_{max} with substantial accuracy. The availability of this type of machine-learning tool in exercise diagnostics

enables better training recommendations for EA and facilitates rehabilitation prescriptions for patients suffering from cardiovascular or respiratory diseases [7,31]. The novelty and main advantage are that there are no comparable studies that first select the variables with the strongest predictive abilities, and then directly evaluate their accuracy in the derived prediction models. An additional attribute is a relatively large group of healthy adult EA ($n = 4001$) who have undergone the CPET under an identical protocol, by which the maximum precision and similarity of the collected data were obtained. This enabled us to better examine whether parameters such as age [32], BC and BF [33] or $\text{VO}_{2\text{max}}$ [16] exerted a possible significant impact on the predictive performance of the model (as they were previously classified as relevant variables in the literature. Moreover, the inclusion criteria enable us to avoid the disturbing influence of factors such as smoking [34] or medications [35].

4.1. Model Performance and Physiological Properties

Performance measurements show precise prediction abilities which were fairly replicable between the training and test sets (see Figures 2 and 3). The obtained R^2 explained approximately 60% to 70% of the differences, while errors were moderate-to-low, under $1 \text{ km}\cdot\text{h}^{-1}$ for most cases. With additional internal validation, they were both still located in the upper sensitivity range. Thus, the model accuracy was only minorly reduced. In previous publications, such as that by Petek et al. for $\text{VO}_{2\text{peak}}$ [36], similar results were observed. However, usually, previous researchers have not carried out an initial selection of the most suitable variables, and so far, studies have been based on previously established parameters, only changing their proportions. Our study showed that VO_2 at RCP and maximal VO_2 were the most important parameters responsible for the prediction of middle- to long-distance running velocity (a lower impact of VO_2 at AT was noted). This confirms previous findings by Thompson et al. and Lanferdini et al. [16,37] that the VO_2 can be described as the universal and comparable performance measure, and that it is strongly related to running speed. Moreover, according to the physiological relationship between exercise performance and $[\text{La}^-]_b$ at AT, at RCP, and maximal VO_2 , they also significantly influence the predicted velocities (but in the varied order compared with VO_2 , with more impact from sub-maximal levels at AT or RCP than maximal $[\text{La}^-]_b$ values). This is confirmed in studies by Tanaka and Matsuura [12] and Schabert et al. [19], as growing $[\text{La}^-]_b$ and training intensity were positively correlated in both. Thus, of greater importance seems to be the ability to rapidly utilize and prevent excess growth in $[\text{La}^-]_b$ by EA than working at maximal value for a prolonged time. Our study confirmed the previous findings by Farrell et al. [38] on this point. Another important variable was pulmonary ventilation. The majority of the influence was created by VE_{RCP} , and only for V_{max} , was there a significant impact of VE_{max} . The higher it was, the better running velocity was observed. Thus, it can be concluded that the higher oxygen (O_2) supply and better carbon dioxide (CO_2) utilization yielded an improvement in running performance. This is a well-documented concept that was stated by Sjodin and Svedenhag in the 1980s [32]. Our insights on both VO_2 and VE also confirmed that performance at RCP is strongly correlated with other running and general exercise indices [15]. When it comes to somatic parameters, they also showed a relevant effect on velocity. Higher BMI [19] and increasing age [39] were associated with lower endurance performance. On the other hand, BC, described as a percentage of BF and FFM, showed some effect on the predicted velocity, despite their impact on males being not enough to be included in the modeling for this gender. It is worth mentioning that the influence of BF was more noticeable in females, perhaps because they naturally have a higher level of BF [40]. HR was one of the variables with the lowest impact on velocity (see Figure 3). Moreover, we emphasize that HR, which shows high inter-individual variability and is difficult to precisely estimate for EA [21], was not included in any of our equations. To summarize, the degree of the relationships between the variables is interesting. It is very promising to assess how precisely we can estimate V_{AT} , V_{RCP} , and V_{max} based on the above-mentioned parameters.

4.2. Clinical Considerations

Our results also have important clinical applications for patients from the general and athletic populations. The development of sports cardiology has resulted in a higher number of EA patients, including former cardiac patients or those suspected of having exertional cardiac abnormalities. TE CPET is often performed to some level of submaximal intensity or until refused. However, those who are less experienced may quit earlier, before reaching their optimal diagnostic intensity level, because they are not mentally adapted to perform such demanding activities [31,41]. The calculation ($\text{MET} \times \text{running velocity}$) is used by medical professionals to provide personalized recommendations for cardiac rehabilitations [31]. Selection of the most important variables and additional comparison of those directly achieved with the predicted velocity verify whether an optimal level of intensity was achieved.

4.3. Practical Applications

The characteristics of selected variables and prediction models could be used in the preparation of exercise recommendations for both professional and recreational EA as patients in clinical settings [7]. The highest accuracy of the observed repeatable values would be for EA, mainly for running activities (i.e., during long-distance running or football), due to the characteristics similar to those in the derivation cohort [36]. Thanks to the use of V_{AT} , V_{RCP} , and V_{max} prediction models, there would be no need to run the full CPET protocol and measure all parameters, but only the most significant and contributing ones [19]. This is a matter of importance, as CPET is often impossible to perform according to the full protocol due to the limited availability of specialized clinics and equipment or the high cost of the procedure [42]. This model can also be used to verify/assess whether the athlete obtains sufficient running speed on the basis of the directly measured parameters. Of course, it currently would not be the gold standard or method of choice. Thus, results should be generalized carefully. However, they could be used as a valuable supplement to direct measurements in the present. We encourage other researchers to test our velocity prediction models and evaluate the proportion of the obtained variables using different populations to assess to what extent the results can be extrapolated and transferred.

4.4. Limitations

A possible limitation is that participants underwent CPET in different phases of the day (circadian rhythm), month (menstrual cycle for female athletes), or season [43,44]. Moreover, we did not evaluate the training volume of the EA. The participants received dietary and preparation tips, but we cannot be sure that they were rigorously implemented; thus, some BIA results for BC should be analyzed with caution. Some data in $[\text{La}^-]_b$ were missed (not all participants decided on the $[\text{La}^-]_b$ test because it was an optional variable in the clinic's CPET portfolio) and RF imputation was applied. RF is recognized as the best method for filling data gaps, and our imputation did not cause a significant negative effect on the $[\text{La}^-]_b$ data precision. The models still showed high prediction abilities at the building and validation (i.e., out-of-bag error) stages. A comparison of both datasets (first set only with directly measured $[\text{La}^-]_b$ and second only with imputed $[\text{La}^-]_b$) did not show significant differences between them ($p = 0.4$) [26]. Volunteers individually declared the intensity level on the Borg scale, and the evaluation could differ between participants. The above limitations result from the specifics of the study, which is population-based, and not a controlled trial. In order to minimize their importance, the above-described internal validation was applied, which revealed the high data precision and replicability of the derived equations.

4.5. Future Directions

We advise that future prediction models used to estimate running velocities should be applied in cohorts with comparable characteristics to those from which they were primarily created (similar to other prediction models used in sports diagnostics) [36]. It is especially

important in narrow and specified populations, including well-trained EA or cardiology patients [36]. We underline that there is a significant necessity for more accurately adjusted contributing factors and the development of new, advanced machine-learning prediction algorithms using unified TRIPOD recommendations [18]. This will enable the subsequent choice of the appropriate protocol to use in medical diagnostic and training prescriptions, depending on the participant's disease type or fitness level [7]. We recommend assessing our methods in an external environment, such as the 3000 m distance run, to cover all evaluation sites [45,46]. It is worth mentioning that, as stated by Figueiredo et al., the critical speed showed a better predictive value for the 5 km running results regarding a steady run than the peak velocity. Although our research focuses mostly on CPET performed in the clinical settings on the mechanical treadmill, we recommend further studies which will investigate the effect of critical speed compared to peak velocity [47].

5. Conclusions

In summary, (1) we found the strongest predictors of running velocity, (2) we derived novel prediction models for running velocities in accordance with TRIPOD guidelines, and (3) we established their fair validation.

Currently, with the use of a machine-learning approach, we can accurately estimate V_{AT} , V_{RCP} , and V_{max} based only on somatic and exertion variables (the precision and repeatability in the study subgroups were comparable to the test-retest error). VO_2 , $[La^-]_b$, VE, and somatic characteristics were the greatest contributing factors. We anticipate that our findings will improve the personalization of training and rehabilitation programs. Models should be primarily applied in disciplines where running is the main form of activity, due to the similar characteristics to those regarding the specificity of the derivation cohort.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11226688/s1>, Supplementary Material S1: TRIPOD Checklist: Prediction Model Development and Validation.

Author Contributions: Conceptualization, S.W.; methodology, S.W. and M.M.; writing—original draft preparation, P.S.K. and S.W.; software and statistics, I.C. and S.W., writing—review and editing, P.S.K., S.W. and M.M.; supervision, D.Ś. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received no funding from an external source.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review of the Bioethical Committee at the Medical University of Warsaw (AKBE/32/2021).

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Saint-Maurice, P.F.; Graubard, B.I.; Troiano, R.P.; Berrigan, D.; Galuska, D.A.; Fulton, J.E.; Matthews, C.E. Estimated Number of Deaths Prevented Through Increased Physical Activity Among US Adults. *JAMA Intern. Med.* **2022**, *182*, 349–352. [[CrossRef](#)]
2. Guazzi, M.; Arena, R.; Halle, M.; Piepoli, M.F.; Myers, J.; Lavie, C.J. 2016 Focused Update: Clinical Recommendations for Cardiopulmonary Exercise Testing Data Assessment in Specific Patient Populations. *Circulation* **2016**, *133*, E694–E711. [[CrossRef](#)] [[PubMed](#)]
3. Bull, F.C.; Al-Ansari, S.S.; Biddle, S.; Borodulin, K.; Buman, M.P.; Cardon, G.; Carty, C.; Chaput, J.P.; Chastin, S.; Chou, R.G.; et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br. J. Sports Med.* **2020**, *54*, 1451–1462. [[CrossRef](#)] [[PubMed](#)]
4. Poole, D.C.; Rossiter, H.B.; Brooks, G.A.; Gladden, L.B. The anaerobic threshold: 50+years of controversy. *J. Physiol.* **2021**, *599*, 737–767. [[CrossRef](#)] [[PubMed](#)]

5. Wasserman, K.; McIlroy, M.B. Detecting threshold of anaerobic metabolism in cardiac patients during exercise. *Am. J. Cardiol.* **1964**, *14*, 844–852. [[CrossRef](#)]
6. Binder, R.K.; Wonisch, M.; Corra, U.; Cohen-Solal, A.; Vanhees, L.; Saner, H.; Schmid, J.P. Methodological approach to the first and second lactate threshold in incremental cardiopulmonary exercise testing. *Eur. J. Cardiovasc. Prev. Rehabil.* **2008**, *15*, 726–734. [[CrossRef](#)]
7. Mann, T.; Lamberts, R.P.; Lambert, M.I. Methods of Prescribing Relative Exercise Intensity: Physiological and Practical Considerations. *Sports Med.* **2013**, *43*, 613–625. [[CrossRef](#)]
8. van den Aardweg, J.G.; De Groot, N. Tight coupling between inspiration and expiration after the respiratory compensation point. *Eur. Respir. J.* **2015**, *46*, PA1553. [[CrossRef](#)]
9. Takano, N. Respiratory compensation point during incremental exercise as related to hypoxic ventilatory chemosensitivity and lactate increase in man. *Jpn. J. Physiol.* **2000**, *50*, 449–455. [[CrossRef](#)]
10. Hill, D.W.; Rowell, A.L. Running velocity at VO_{2max} . *Med. Sci. Sports Exerc.* **1996**, *28*, 114–119. [[CrossRef](#)]
11. Wiecha, S.; Price, S.; Cieslinski, I.; Kasiak, P.S.; Tota, L.; Ambrozy, T.; Sliz, D. Transferability of Cardiopulmonary Parameters between Treadmill and Cycle Ergometer Testing in Male Triathletes-Prediction Formulae. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1830. [[CrossRef](#)] [[PubMed](#)]
12. Tanaka, K.; Matsuura, Y. Marathon performance, anaerobic threshold, and onset of blood lactate accumulation. *J. Appl. Physiol.* **1984**, *57*, 640–643. [[CrossRef](#)] [[PubMed](#)]
13. Galan-Rioja, M.A.; Gonzalez-Mohino, F.; Poole, D.C.; Gonzalez-Rave, J.M. Relative Proximity of Critical Power and Metabolic/Ventilatory Thresholds: Systematic Review and Meta-Analysis. *Sports Med.* **2020**, *50*, 1771–1783. [[CrossRef](#)]
14. Esteve-Lanao, J.; Del Rosso, S.; Larumbe-Zabala, E.; Cardona, C.; Alcocer-Gamboa, A.; Boullosa, D.A. Predicting Recreational Runners' Marathon Performance Time During Their Training Preparation. *J. Strength Cond. Res.* **2021**, *35*, 3218–3224. [[CrossRef](#)] [[PubMed](#)]
15. Rory, P.B.; Thabtah, F. A machine learning framework for sport result prediction. *Appl. Comput. Inform.* **2019**, *15*, 27–33. [[CrossRef](#)]
16. Lanferdini, F.J.; Silva, E.S.; Machado, E.; Fischer, G.; Peyre-Tartaruga, L.A. Physiological Predictors of Maximal Incremental Running Performance. *Front. Physiol.* **2020**, *11*, 979. [[CrossRef](#)]
17. Malek, M.H.; Berger, D.E.; Housh, T.J.; Coburn, J.W.; Beck, T.W. Validity of VO_{2max} equations for aerobically trained males and females. *Med. Sci. Sports Exerc.* **2004**, *36*, 1427–1432. [[CrossRef](#)]
18. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M.; Grp, T. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Eur. Urol.* **2015**, *67*, 1142–1151. [[CrossRef](#)]
19. Schabert, E.J.; Killian, S.C.; Gibson, A.S.; Hawley, J.A.; Noakes, T.D. Prediction of triathlon race time from laboratory testing in national triathletes. *Med. Sci. Sports Exerc.* **2000**, *32*, 844–849. [[CrossRef](#)]
20. Ghosh, A.K. Anaerobic threshold: Its concept and role in endurance sport. *Malays. J. Med. Sci. MJMS* **2004**, *11*, 24–36.
21. Lach, J.; Wiecha, S.; Sliz, D.; Price, S.; Zaborski, M.; Cieslinski, I.; Postula, M.; Knechtel, B.; Mamcarz, A. HR Max Prediction Based on Age, Body Composition, Fitness Level, Testing Modality and Sex in Physically Active Population. *Front. Physiol.* **2021**, *12*, 695950. [[CrossRef](#)] [[PubMed](#)]
22. Meignie, A.; Duclos, M.; Carling, C.; Orhant, E.; Provost, P.; Toussaint, J.F.; Antero, J. The Effects of Menstrual Cycle Phase on Elite Athlete Performance: A Critical and Systematic Review. *Front. Physiol.* **2021**, *12*, 654585. [[CrossRef](#)] [[PubMed](#)]
23. Ross, R.; Blair, S.N.; Arena, R.; Church, T.S.; Despres, J.P.; Franklin, B.A.; Haskell, W.L.; Kaminsky, L.A.; Levine, B.D.; Lavie, C.J.; et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign A Scientific Statement From the American Heart Association. *Circulation* **2016**, *134*, E653–E699. [[CrossRef](#)]
24. Guazzi, M.; Adams, V.; Conraads, V.; Halle, M.; Mezzani, A.; Vanhees, L.; Arena, R.; Fletcher, G.F.; Forman, D.E.; Kitzman, D.W.; et al. Clinical Recommendations for Cardiopulmonary Exercise Testing Data Assessment in Specific Patient Populations. *Circulation* **2012**, *126*, 2261–2274. [[CrossRef](#)]
25. Beaver, W.L.; Wasserman, K.; Whipp, B.J. A new method for detecting anaerobic threshold by gas-exchange. *J. Appl. Physiol.* **1986**, *60*, 2020–2027. [[CrossRef](#)] [[PubMed](#)]
26. Stekhoven, D.J.; Bühlmann, P. MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2012**, *28*, 112–118. [[CrossRef](#)]
27. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C plus plus and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [[CrossRef](#)]
28. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
29. Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **2006**, *7*, 91. [[CrossRef](#)]
30. Altman, D.G.; Bland, J.M. Measurement in medicine—the analysis of method comparison studies. *J. R. Stat. Soc. Ser. D Stat.* **1983**, *32*, 307–317. [[CrossRef](#)]
31. Jette, M.; Sidney, K.; Blumchen, G. Metabolic equivalents (mets) in exercise testing, exercise prescription, and evaluation of functional-capacity. *Clin. Cardiol.* **1990**, *13*, 555–565. [[CrossRef](#)]
32. Sjodin, B.; Svedenhag, J. Applied physiology of marathon running. *Sports Med.* **1985**, *2*, 83–99. [[CrossRef](#)] [[PubMed](#)]
33. Krachler, B.; Savonen, K.; Lakka, T. Obesity is an important source of bias in the assessment of cardiorespiratory fitness. *Am. Heart J.* **2015**, *170*, E7–E8. [[CrossRef](#)] [[PubMed](#)]

34. Cooper, K.H.; Gey, G.O.; Bottenberg, R.A. Effects of cigarette smoking on endurance performance. *J. Am. Med. Assoc.* **1968**, *203*, 189. [[CrossRef](#)]
35. Adami, P.E.; Koutlianos, N.; Baggish, A.; Bermon, S.; Cavarretta, E.; Deligiannis, A.; Furlanello, F.; Kouidi, E.; Marques-Vidal, P.; Niebauer, J.; et al. Cardiovascular effects of doping substances, commonly prescribed medications and ergogenic aids in relation to sports: A position statement of the sport cardiology and exercise nucleus of the European Association of Preventive Cardiology. *Eur. J. Prev. Cardiol.* **2022**, *29*, 559–575. [[CrossRef](#)]
36. Petek, B.J.; Tso, J.V.; Churchill, T.W.; Guseh, J.S.; Loomer, G.; DiCarli, M.; Lewis, G.D.; Weiner, R.B.; Kim, J.H.; Wasfy, M.M.; et al. Normative cardiopulmonary exercise data for endurance athletes: The Cardiopulmonary Health and Endurance Exercise Registry (CHEER). *Eur. J. Prev. Cardiol.* **2021**, *29*, 536–544. [[CrossRef](#)] [[PubMed](#)]
37. Thompson, M.A. Physiological and Biomechanical Mechanisms of Distance Specific Human Running Performance. *Integr. Comp. Biol.* **2017**, *57*, 293–300. [[CrossRef](#)] [[PubMed](#)]
38. Farrell, P.A.; Wilmore, J.H.; Coyle, E.F.; Billing, J.E.; Costill, D.L. Plasma lactate accumulation and distance running performance. *Med. Sci. Sports Exerc.* **1979**, *11*, 338–344. [[CrossRef](#)]
39. Stensvold, D.; Sandbakk, S.B.; Viken, H.; Zisko, N.; Reitlo, L.S.; Nauman, J.; Gaustad, S.E.; Hassel, E.; Moufack, M.; Bronstad, E.; et al. Cardiorespiratory Reference Data in Older Adults: The Generation 100 Study. *Med. Sci. Sports Exerc.* **2017**, *49*, 2206–2215. [[CrossRef](#)]
40. Vogel, J.A.; Friedl, K.E. Body-fat assessment in women—special considerations. *Sports Med.* **1992**, *13*, 245–269. [[CrossRef](#)]
41. McCormick, A.; Meijen, C.; Marcora, S. Psychological Determinants of Whole-Body Endurance Performance. *Sports Med.* **2015**, *45*, 997–1015. [[CrossRef](#)]
42. Ekblom-Bak, E.; Bjorkman, F.; Hellenius, M.L.; Ekblom, B. A new submaximal cycle ergometer test for prediction of VO_{2max} . *Scand. J. Med. Sci. Sports* **2014**, *24*, 319–326. [[CrossRef](#)] [[PubMed](#)]
43. Petek, B.J.; Groezinger, E.Y.; Pedlar, C.R.; Baggish, A.L. Cardiac effects of detraining in athletes: A narrative review. *Ann. Phys. Rehabil. Med.* **2022**, *65*, 101581. [[CrossRef](#)] [[PubMed](#)]
44. Price, S.; Wiecha, S.; Cieśliński, I.; Śliż, D.; Kasiak, P.S.; Lach, J.; Gruba, G.; Kowalski, T.; Mamcarz, A. Differences between Treadmill and Cycle Ergometer Cardiopulmonary Exercise Testing Results in Triathletes and Their Association with Body Composition and Body Mass Index. *Int. J. Environ. Res. Public Health* **2022**, *19*, 3557. [[CrossRef](#)] [[PubMed](#)]
45. Du, Z.; Lu, W.; Lang, D. Comparison between 2000 m and 3000 m time trials to estimate the maximal aerobic speed for collegiate runners. *Front. Physiol.* **2022**, *13*, 1005259. [[CrossRef](#)] [[PubMed](#)]
46. Casado, A.; Tuimil, J.L.; Iglesias, X.; Fernandez-Del-Olmo, M.; Jimenez-Reyes, P.; Martin-Acero, R.; Rodriguez, F.A. Maximum aerobic speed, maximum oxygen consumption, and running spatiotemporal parameters during an incremental test among middle- and long-distance runners and endurance non-running athletes. *PeerJ* **2022**, *10*, e14035. [[CrossRef](#)]
47. Figueiredo, D.H.; Figueiredo, D.H.; Manoel, F.A.; Machado, F.A. Peak Running Velocity or Critical Speed Under Field Conditions: Which Best Predicts 5-km Running Performance in Recreational Runners? *Front. Physiol.* **2021**, *12*, 680790. [[CrossRef](#)]