

External validation of VO_{2max} prediction models based on recreational and elite endurance athletes

Szczepan Wiecha (✉ szczepan.wiecha@awf.edu.pl)

Department of Physical Education and Health, Faculty in Biala Podlaska, Jozef Pilsudski University of Physical Education in Warsaw, 21-500 Biala Podlaska, Poland

Przemysław Seweryn Kasiak

Students' Scientific Group of Lifestyle Medicine, 3rd Department of Internal Medicine and Cardiology, Medical University of Warsaw, 02-091 Warsaw

Igor Cieśliński

Department of Physical Education and Health, Faculty in Biala Podlaska, Jozef Pilsudski University of Physical Education in Warsaw, 21-500 Biala Podlaska, Poland

Tim Takken

Department of Medical Physiology, Child Development & Exercise Center, Wilhelmina Children's Hospital, UMC Utrecht, Utrecht

Tomasz Pałka

Department of Physiology and Biochemistry, Faculty of Physical Education and Sport, University of Physical Education in Krakow, 31-571 Krakow

Beat Knechtle

Institute of Primary Care, University of Zurich, Zurich

Pantelis T Nikolaidis

School of Health and Caring Sciences, University of West Attica, Athens 12243

Łukasz A Małek

Department of Epidemiology, Cardiovascular Disease Prevention and Health Promotion, National Institute of Cardiology, Warsaw

Marek Postuła

Department of Experimental and Clinical Pharmacology, Centre for Preclinical Research and Technology, Medical University of Warsaw, Warsaw

Artur Mamcarz

3rd Department of Internal Medicine and Cardiology, Medical University of Warsaw, 02-091 Warsaw

Daniel Śliż

3rd Department of Internal Medicine and Cardiology, Medical University of Warsaw, 02-091 Warsaw

Article

Keywords: endurance athlete, cardiopulmonary exercise test, cardiorespiratory fitness, maximum oxygen consumption, VO_{2max} , prediction model, validation

Posted Date: April 19th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1441930/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

In recent years, numerous prognostic models have been developed to predict VO_{2max} . Nevertheless, their accuracy in endurance athletes (EA) stays mostly unvalidated. This study aimed to compare predicted VO_{2max} (pVO_{2max}) with directly measured VO_{2max} .

5,260 healthy adult EA underwent a maximal exertion cardiopulmonary exercise test (CPET) (84.76% male; age 34.6 ± 9.5 yrs.; VO_{2max} 52.97 ± 7.39 mL·min⁻¹·kg⁻¹, BMI 23.59 ± 2.73 kg·m⁻²). 13 models have been selected to establish pVO_{2max} . Participants were classified into four endurance subgroups (high-, recreational-, low-, trained, and “transition”) and four age subgroups (18-30, 31-45, 46-60, and ≥ 61 yrs.). Validation was performed according to TRIPOD guidelines. pVO_{2max} was low-to-moderately associated with direct CPET measurements ($p > 0.05$). Models with the highest accuracy were for males on a cycle ergometer (CE) (Kokkinos $R^2 = 0.64$), females on CE (Kokkinos $R^2 = 0.65$), males on a treadmill (TE) (Wasserman $R^2 = 0.26$), females on TE (Wasserman $R^2 = 0.30$). However, selected models underestimated pVO_{2max} for younger and higher trained EA and overestimated for older and lower trained EA.

All equations demonstrated merely moderate accuracy and should only be used as a supplemental method for physicians to estimate CRF in EA. It is necessary to derive new models on EA populations to include routinely in clinical practice and sports diagnostic.

Introduction

The concept of maximal oxygen uptake (VO_{2max}) was first suggested by Hill et al. in the 1920s.[1] VO_{2max} is the highest attained oxygen uptake during an incremental exercise test with large muscle groups (e.g., treadmill or cycling). VO_{2max} is an important parameter to objectively assess cardiorespiratory fitness (CRF) both in healthy people and those suffering from cardiovascular diseases (CVD)[2,3]. The American Heart Association (AHA) recognized that CRF, described mainly as a VO_{2max} , should be used as an essential factor in the comprehensive diagnostic process[3]. Moreover, a lower level of CRF is strongly related to a higher risk of CVDs, death from numerous cancer types, and all-cause mortality[4]. This represents a switch from risk factors widely discussed in recent decades, such as smoking, hypertension, or hyperlipidemia[2,3,5].

VO_{2max} in sports & performance diagnostics

VO_{2max} is an important variable in endurance sports, such as running, cycling, swimming, triathlon, or team sports[6]. VO_{2max} strongly correlates with athlete's aerobic performance, could be applied to prescribe training properly, and is useful to assess adaptation to exercise[7–9]. Furthermore, the VO_{2max} could help in the prediction of a race time[10,11]. Elite athletes achieve varied VO_{2max} values, dependent on their discipline and training experience[12,13]. Males typically have higher VO_{2max} than females[14], and VO_{2max} values decrease with age[15]. Body weight and height are related as well as the testing mode. Higher VO_{2max} values are observed on a treadmill compared to the cycle ergometer[16]. Due to its numerous practical implications and variability, it is important to precisely assess VO_{2max} in different athletic disciplines and populations[17].

VO_{2max} in clinical practice

Measuring VO_{2max} is also especially important under clinical conditions during the examination of the cardiovascular system[3,18]. It could be regarded as the integrated function of (amongst others) lungs, heart, blood vessels, and muscles [9,19]. Recommendations for VO_{2max} -testing include the presence of ambiguous pathologic exertional symptoms, cardiovascular risk estimation, and monitoring response to applied treatment[18]. Moreover, understanding the exercise limitation is crucial information for healthcare professionals to monitor cardiac status and could be used to prescribe treatment properly for those suffering from CVDs[2,20]. Therefore, VO_{2max} is a practically relevant parameter for a new, growing population of patients in cardiologic ambulatory care—endurance athletes (EA)[18]. Both highly trained endurance athletes (HTEA), recreational endurance athletes (REA) and low-trained endurance athletes (LTEA) with suspected CVD and those undergoing cardiopulmonary exercise testing (CPET) for periodic training evaluation are potential candidates for VO_{2max} assessment[18].

CPET protocol and applicability of prediction formulae

The gold standard to measure VO_{2max} is performing a CPET[21]. VO_{2max} is reached when the subject meets the physiological limit and maintains it for some time (usually 15-s, 30-s, or 60-s)[22]. Due to practical reasons, such as high costs of the procedure or a lack of testing devices as well as health contraindications, this form of measuring is often not possible to apply in a sports setting[21].

Parameters such as age, sex, and heart rate (HR) could be used to predict VO_{2max} through various models[21,23]. The reliability of this potentially non-sophisticated and valuable method is complicated and doubtful because of low accuracy, especially in women, extremely small or tall subjects, and in individuals with high BMI values[24,25]. In the 2013 statement, AHA pointed out that there is a need for a universal and transferable prediction standard[26].

Prediction formulae undoubtedly have numerous advantages, however, those currently used were created on different populations and with the incorporation of heterogeneous testing modes[27]. Indeed, proper external validation should be a mandatory stage before the new model will be widely used[28,29]. This study aimed to externally evaluate prediction formulae on EA tested under the same conditions from one tertiary care sports diagnostic center. The secondary aim was to assess the impact of age and CRF on the risk of error and bias in tested models. We hypothesise that their validity may not be sufficient to make them an equivalent method for directly measured VO_{2max} .

Material And Methods

We applied TRIPOD guidelines for the development and validation of prediction models[27,30] (for detailed protocol see Supplementary information. TRIPOD Checklist for Prediction Model Validation). Results from CPETs collected between 2013-2021 were retrospectively analyzed. Maximal-effort examinations consisted of the treadmill (TE) or the cycle ergometry (CE) tests, paired with body composition (BC) analysis took place in the medical clinic (www.sportslab.pl, Warsaw, Poland). Tests were performed on an individual request as a part of regular endurance assessment or training monitoring.

Cardiopulmonary exercise testing protocol

Cardiopulmonary exercise tests (CPET) were preceded by body mass (BM) and fat mass (FM) analysis with 5 kHz/50 kHz/250 kHz electrical bioimpedance method on the body composition (BC) monitor (Tanita, MC 718, Japan. Conditions during BC and CPET were: 40 m² indoor, air-conditioned area, 40–60% humidity, temperature 20–22°C, altitude 100 m MSL. Endurance athletes (EA) were instructed via e-mail on how to prepare: avoid any demanding exercises 24 hours before CPET, consume a high carbohydrate meal and hydrate with isotonic beverages 2-3 hours earlier, and exclude any stimulants or caffeine on the day of the procedure.

Cycle ergometry (CE) examination was performed on a cycle ergometry Cyclus-2 (RBM elektronik-automation GmbH, Leipzig, Germany) and treadmill (TE) examination was conducted on a mechanical treadmill (h/p/Cosmos quasar, Germany). CPET scores were measured breath by breath during 10-s intervals using a Hans Rudolph V2 Mask (Hans Rudolph, Inc, Shawnee, KS, USA), a gas exchange analyzer Cosmed Quark CPET (Rome, Italy), and dedicated manufacturer's software (from PFT Suite to Omnia 10.0E.) HR was measured via ANT and a torso strap as a part of the Cosmed Quark set (product accuracy comparable to ECG; ± 1 bpm.). The CPET device was calibrated with reference gas (16% O₂; 5% CO₂) and turbine flow for each person separately, according to manufacturer recommendations. Equipment software was regularly actualized between 2013-2021. Three gas analyzing devices were utilized and each one has been changed after 36-48 months. Every part of CPET equipment was periodically verified by manufacturer employees to keep their mechanical certificates valid. Blood lactate (LA) was assessed with the usage of Super GL2 analyzer (Müller Gerätebau GmbH, Freital, Germany). The instrument was also individually prepared before each round of analysis and calibrated with reference solution before each sample set.

Exercises begin with a 5-min. warm-up (walking or pedaling with minimal resistance). Participants' endurance capacities were used to assess starting load. The initial power for CE was 60-150W and was increased in 2 min. intervals by 20-30W. The initial speed for TE was 7-12 km·h⁻¹ (described by a person as a "conversation pace") at 1% inclination. The pace was raised by 1 km·h⁻¹ every 2 min. Observer verbally encouraged athletes to keep effort as long as possible due to assess their endurance most exactly. Achievement of oxygen uptake (VO₂) or heart rate (HR) plateau, or volitional inability to maintain intensity were reasons for test termination. LA was measured by taking a 20 µL blood sample from a fingertip: directly prior to exercises, after any resistance or pace modification, and 3 min. after termination. Samples were obtained without an interruption in CE and TE tests. Before a proper sample was obtained, the first drops were gathered in a swab. HR (not averaged) was recorded at the highest point during intervals and used in further analysis [32]. Maximal oxygen uptake (VO_{2max}) was defined as an averaged maximum oxygen uptake during the 15-s period at the end of the CPET.

Derivation cohort

The rigorous inclusion/exclusion process was applied to narrow the validation group to only those EAs who achieved maximum exertion during CPET and were free of any possible VO_{2max} alleviating factors (see Figure 1. Flowchart of the inclusion-exclusion and further groups classification process).

6,439 EAs underwent CPET. Participants were eligible for preliminary inclusion if they had: (1) experience in regular running or cycling training ≥ 3 months, (2) age ≥ 18 years, (3) ± 3 standard deviations (SD) from mean for all of the testing variables (extreme outliers were excluded), (4) lack of any acute or chronic medical condition (also musculoskeletal injuries, or addictions), (5) not taking any medications, (6) not being an active smoker.

Maximum exertion in CPET was defined as fulfilment ≥ 6 criteria: (1) respiratory exchange ratio (RER) ≥ 1.10 , (2) present VO₂ plateau (growth < 100 mL·min⁻¹ in VO₂ with more increased running or cycling intensity), (3) respiratory frequency (fR) ≥ 45 breaths·min⁻¹, (4) declared exertion during CPET ≥ 18 in the Borg scale[31], (5) lactate concentration (LA) ≥ 8 mmol·L⁻¹, (6) growth in speed/power $\geq 10\%$ of RCP after exceeding the respiratory compensation point (RCP), (7) peak heart rate (HR_{peak}) ≥ 15 bpm below predicted maximal heart rate (HR_{max})[32].

Finally, 5,260 EA met all inclusion criteria. The population was divided between males and females into four age groups: 18-30; 31-45; 46-60, ≥ 61 years, and 4 endurance groups: HTEA, REA, LTEA, and "transition". Endurance classification was conducted based on the speed (km·h⁻¹) or power (W·kg⁻¹) at the RCP calculated independently for each sex. Speed/power at RCP was a variable-of-choice because it is currently described as a parameter most closely corresponding to the critical endurance capacity[33,34]. Moreover, the selection of a variable different from VO_{2max} to the classification of participants in terms of their endurance capacity, enabled to make group assignments independent of the factor directly validated in the study. Participants with $> +1.5$ SD were classified as HTEA (n=309), < -0.5 SD/ > -0.5 SD as REA (n=2,033), < -1.5 SD as LTEA (n=339). To precisely distinguish endurance subgroups, those placed between $\geq +0.5$ SD/ $\leq +1.5$ SD and between ≤ -0.5 SD/ ≥ -1.5 SD were classified as "transition" (n=2,579). Models' validation was conducted on each of the age and endurance cohorts independently (except the „transition" group) both for TE_{VO2max} and CE_{VO2max}.

Selected prediction models

Candidate models were found from previous systematic reviews for CPET testing (up to February 2019)[35,36] and additional literature search in PubMed, MEDLINE, EMBASE, Scopus, and Web of Science databases (for a period between March 2019- December 2021 and meta-analyses) for keywords: Cardiopulmonary exercise testing, Cardiorespiratory fitness, Exercise testing, VO_{2max}, VO_{2peak}.

Exclusion criteria were: (1) not reporting VO_{2max} parameters, (2) usage of other ergometers than CE or TE during CPET, (3) consideration of parameters not possible to verify in our sample (declared physical activity level, time to exhaustion), (4) generating unviable results multiple times (< 0 or > 100 mL·min⁻¹·kg⁻¹

¹ VO_{2peak}), (5) being derived from pediatric (the oldest participant <18 years old) or geriatric population (the youngest participant ≥61 yrs.), (6) being derived before 01.01.2000, (7) not reporting R² from internal/external validation, (8) being derived from group <1000 participants (for the general population) or <200 participants (specifically for EA population), (9) methodological quality <7 points according to ATS/ACCP guidelines[37], (10) usage of other testing technique than breath by breath, (11) CPET protocol not carried out following the recommended clinical ATS/ACCP guidelines[37].

Moreover, the Wasserman et al.[19] model was validated in the study due to its well-established reputation. Equations from 2 meta-analyses[38,39] were also considered because of their wide range of applications for EA. 13 equations from 8 different publications were included in the analysis. Their detailed characteristics are presented in the supplementary material (Supplementary table 1. Prediction equations included in the validation).

Statistical analysis

Baseline statistics were exported into the Excel file (Microsoft Corporation, Washington, USA) and are presented as mean (±SD and 95% CI) or frequency (percentage) for categorical variables, and median for continuous variables. Differences between subgroups (all continuous variables) were analyzed using the ANOVA test-of-variance and post-hoc HSD Tukey test. There was not any missing data in the whole population. Thus, an entire cohort has been validated.

External validation was conducted by following the recommendations for the validation and interpretation of diagnostic prediction models(30). In summary, we assessed equations accuracy by comparisons between the originally established formulas and data obtained directly from CPETs and BC examinations (e.g., VO_{2max}, BMI). Linear model regressing measured VO_{2max} on pVO_{2max} was generated for each equation. Performance, considered as the proximity of the observed and expected CRF, was evaluated with the usage of the R², root mean square error (RMSE). Additionally, calibration slope (the slope of a linear regression model that includes the model's linear predictor as the only covariate parameter estimate where 1 being ideal; C1), and calibration-in-the-large (mean observed compared to mean predicted value where 0 being ideal; C2) were calculated.

Ggplot 2 package in RStudio (R Core Team, Vienna, Austria; version 3.6.4), originally written Python script (Python Software Foundation, Delaware, USA; version 3.10.1), and STATA software (StataCorp, College Station, Texas, USA; version 15.1) were used in statistical analysis. The significance borderline was at a two-sided *p*-value <0.05.

Results

From a total of 6,439 endurance athletes (EA) who underwent CPET at a tertiary care sports diagnostic center in Poland, 5,260 EA met the inclusion criteria. Participants' basic anthropometric characteristics are shown in Table 1. The average age of was 35.04 ± 9.58 yrs. in the male population (n = 4,459; 84.76%) and 32.25 ± 8.99 yrs. in the female population (n = 801; 15.24%).

Table 1
Participants' basic anthropometric characteristics

Variable	Male [n = 4459; 84.76%]	Female [n = 801; 15.24%]
Baseline characteristic		
Age (years)	35.04 (9.58)*	32.25 (8.99)
Height (cm)	179.42 (6.60)*	167.19 (6.88)
Weight (kg)	77.23 (10.32)*	60.60 (8.73)
BMI (kg·m ⁻²)	23.95 (2.63)*	21.64 (2.38)
BF (%)	15.68 (4.55)*	22.04 (5.46)
FFM (kg)	64.87 (7.17)*	47.08 (6.36)
Endurance groups characteristic		
HTEA (n = 316; 6.08%)	257 (4.89)	59 (1.12)
REA (n = 2009; 38.19%)	1711 (32.52)	298 (5.67)
LTEA (n = 345; 6.56%)	290 (5.51)	55 (1.05)
„transition“ (n = 2590; 49.24%)	2201 (41.84)	389 (7.40)
Age groups characteristic		
Age 18–30 (n = 1380; 26.24%)	1099 (20.89)	281 (20.89)
Age 31–45 (n = 3310; 62.92%)	2842 (54.03)	458 (8.71)
Age 46–60 (n = 538; 10.23%)	487 (9.26)	51 (0.97)
Age > 61 (n = 32; 0.61%)	31 (0.59)	1 (0.02)

Abbreviations: CE, cycle ergometry; TE, treadmill; BMI, body mass index; BF, body fat; FFM, fat-free mass; HTEA, high-trained endurance athletes; REA, recreational endurance athletes; LTEA, low-trained endurance athletes. Continuous value is presented as mean (SD), while categorical was showed as numbers (%) when appropriate. Comparisons between subgroups (p value) were obtained by one-way ANOVA, Student's t-test, and post-hoc HSD Tukey test. Significant differences ($p < 0.05$) were marked with [†].

CPET data are shown in the supplementary material (Supplementary table 2. CPET characteristics). For male EA, observed VO_{2max} was significantly higher for TE ($n = 3,330$) than for CE ($n = 1,129$) (54.10 ± 6.93 vs 51.92 ± 8.05 mL·min⁻¹·kg⁻¹; $p < 0.05$). In female athletes, VO_{2max} was similar for TE ($n = 671$) and for CE ($n = 130$) (48.79 ± 6.67 vs 49.05 ± 6.64 mL·min⁻¹·kg⁻¹; ($p > 0.05$). HTEA had significantly higher ($p < 0.05$) levels of VO_{2max} , the speed at RCP (S_{RCP}), and the power at RCP (P_{RCP}). Observed VO_{2max} was significantly lower ($p < 0.05$) in the LTEA subgroup.

Briefly, VO_{2max} differed significantly between the selected equations. The performance of prediction models is presented in Table 2 and Table 3 along with R^2 , root mean square error (RMSE), calibration-in-the-large (C1), and calibration slope (C2). Figures 2, 3, 4 and 5 shows the regression analysis of observed vs predicted VO_{2max} stratified by age for the whole population, HTEA, REA, and LTEA, respectively. Subgroups that did not meet the TRIPOD guidelines[30] to consider their validation results as reliable (i.e., $n \geq 100$) were additionally marked in tables and graphs.

Table 2
Performance of selected models stratified by endurance level and

Prior Equations in CE															
Validated subgroup	Males												Females		
	HTEA‡ [n = 69]				REA [n = 429]				LTEA‡ [n = 81]				HTEA‡ [n = 10]		
	R ²	RMSE	C1	C2	R ²	RMSE	C1	C2	R ²	RMSE	C1	C2	R ²	RMSE	C1
Wilson et al.† (mL·min ⁻¹ ·kg ⁻¹)	0.22*	6.62	16.59	0.76*	0.24*	4.74	11.20*	0.68*	0.21*	4.37	8.03	0.53*	n/a	n/a	n/a
Fitzgerald et al.† (mL·min ⁻¹ ·kg ⁻¹)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.03	3.38	40
Wasserman et al. (mL·min ⁻¹)	0.27*	446	2064.6*	0.88*	0.48*	363	1390.49*	0.93*	0.60*	351	1164.51*	0.73*	0.40*	291	29
Kokkinos et al. (1 for males/ 2 for females)§ (mL·min ⁻¹ ·kg ⁻¹)	0.35*	6.06	10.69	0.86*	0.31*	4.53	10.17*	0.90*	0.35*	4.91	36.54*	0.11	0.09	3.27	43
Kokkinos et al. (3)§ (mL·min ⁻¹ ·kg ⁻¹)	0.35*	6.06	10.53	0.91*	0.31*	4.53	9.99*	0.94*	0.00	4.91	36.52*	0.12	0.09	3.27	43
Mylius et al. (mL·min ⁻¹)	0.16	480.53	2147.43*	0.75*	0.35	405.87	925.17*	1.03*	0.22	489.44	498.10	0.99*	0.09	251.25	83
Petek et al. (L·min ⁻¹)	0.15*	0.48	1.91*	0.61*	0.27*	0.43	1.10*	0.73*	0.21*	0.49	0.53	0.75*	0.27	0.32	1.0
Prior Equations in TE															
Validated subgroup	HTEA [n = 188]				REA [n = 1282]				LTEA [n = 209]				HTEA‡ [n = 49]		
	R ²	RMSE	C1	C2	R ²	RMSE	C1	C2	R ²	RMSE	C1	C2	R ²	RMSE	C1
Wilson et al.† (mL·min ⁻¹ ·kg ⁻¹)	0.15	5.56	23.59*	0.66*	0.14*	5.01	21.75*	0.53*	0.09*	5.27	20.23*	0.42*	n/a	n/a	n/a
Fitzgerald et al.† (mL·min ⁻¹ ·kg ⁻¹)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0.10	5.52	36
Wasserman et al. (mL·min ⁻¹)	0.44*	404	1461.98*	1.14*	0.45*	379	1583.39*	0.89*	0.56*	411	1382.82*	0.78*	0.64*	450	– 10
Myers et al. (mL·min ⁻¹ ·kg ⁻¹)	0.20*	5.40	31.67*	0.70*	0.23*	4.75	26.41*	0.62*	0.21*	4.93	24.32*	0.51*	0.08*	5.58	44
Nevill et al. (1) §§ (mL·min ⁻¹ ·kg ⁻¹)	0.18*	5.47	42.55*	0.45*	0.20*	4.84	35.06*	0.42*	0.18*	5.03	31.00*	0.35*	0.10*	5.52	43
Nevill et al. (2) §§ (mL·min ⁻¹ ·kg ⁻¹)	0.18	5.47	32.33*	0.68*	0.21*	4.81	26.77*	0.61*	0.21*	4.93	24.84*	0.50*	0.08*	5.54	40

Prior Equations in CE															
Petek et al. (L·min ⁻¹)	0.18*	0.49*	0.32	1.02*	0.23*	0.45*	0.26	0.94	0.23*	0.23*	0.94*	-0.07	0.53*	0.51	-1

Abbreviations: CE, cycle ergometry; HTEA, high-trained endurance athletes; REA, recreational endurance athletes; LTEA, low-trained endurance athletes; R², adjusted R²; RMSE, root mean square error; C1, calibration-in-the-large; C2, calibration slope; n/a, not applicable; TE, treadmill. Comparisons between subgroups (*p* value) were obtained by one-way ANOVA, Student's t-test, and post-hoc HSD Tukey test. Significant values (*p* < 0.05) were marked as [*]. All values are presented in originally derived units for each model (original unit was added in bracket). †Fitzgerald et al. and Wilson et al. are meta-analyses exclusively for one sex. §Kokkinos et al. presents 3 equations for cycle ergometry: (1) only for males, (2) only for females, (3) for both males and females. §§Nevill et al. presents 2 equations for treadmill: (1) allometric model, (2) additive model. Subgroups that did not meet the TRIPOD guidelines to consider their validation results as reliable (i.e. n ≥ 100) were marked with [#].

Performance calculations for the whole population and each subgroup, with comparison (mean and SD) between observed and predicted VO_{2max} are presented in the supplementary material (Supplementary table 3a-d). For TE, the lowest non-significant differences (mean and CI) were for Petek's equation both in males (mean = -0.11; CI, -0.42, 0.20) and females (mean = -0.52; CI, -1.20, 0.16). For CE, the lowest non-significant differences (mean and CI) were for Petek's equation in the male population (mean = -0.08; CI, -0.68, 0.52). Similarly, for the female population, the lowest but significant differences were also for Petek's equation (mean = 2.65; CI, 1.15, 4.15). For TE, other models significantly overestimated VO_{2max} (Wilson's for males and Fitzgerald's for females) or underestimated (Wasserman's, Myers's, Nevill's both allometric (1) and additive (2) formulae for males and Wasserman's, Myers's, Nevill's (1) and (2) for females). For CE, significant overestimation was observed in Wilson's and Fitzgerald's models respectively for males and females, and underestimation in Wasserman's, Mylius's, and Kokkinos's formulae for both males, females, and the whole population.

For HTEA a group size ≥ 100 was only for male runners (n = 188). For TE, significant differences between the observed and predicted VO_{2max} were both in males and females for the HTEA subgroup. The lowest obtained differences on TE were for the Wilson's for males (mean = 2.52; CI, 1.50, 3.54) and for the Fitzgerald's for females (mean = 4.1; CI, 1.92, 6.28). For CE, there were no significant differences for Wilson's model (mean = 1.82; CI, -0.32, 3.96) and Fitzgerald's model formula (mean = 2.57; CI, -1.38, 6.52), respectively for male and female EA.

The equation based on the general population explained little to moderate of the measured variability in the VO_{2max} in the general derivation cohort. For TE, R² ranged from 0.28 for Nevill's (1) equation up to 0.54 for Petek's equation. In CE, R² ranged from 0.38 for Petek's equation up to 0.64 in Kokkinos's equations. In the HTEA cohort (only one with n ≥ 100), for males TE, R² ranged from 0.15 in Wilson's equation up to 0.44 in Wasserman's equation. Although, they were poorly calibrated (for Wasserman's C1 = 1461.98 mL·min⁻¹).

Discussion

The aim of the current study was to assess the accuracy of common VO_{2max} prediction equations in a large sample of healthy EA tested under standardized conditions. We hypothesized that their accuracy may not be adequate to make them a comparable approach for CPET.

Our main finding is that previously published predictive equations derived from general or athletic cohorts, along with those widely used equations showed moderate to poor performance in EA. Furthermore, a steeper decline in predicted VO_{2max} for older participants was noted.

Current limitations in model's transferability

Until now, most frequently underestimation of results of what in younger EA and overestimation in older ones have been observed[18,23]. Malek et al. found that 16 of 18 commonly used prediction equations were inaccurate when used in an athlete population[23]. Moreover, there was a lack of equations to predict VO_{2max} developed in large samples of trained participants, especially elite athletes[18]. In one recent study, Petek et al. validated previous and developed new VO_{2max} equations for EA, although the sample size was relatively small[18]. Their main finding was that the previously established models, both on general cohorts and EA, perform poorly when used for EA undergoing CPET for clinical reasons.

Valid VO_{2max} prediction equations are important as they can lead to false-negative or false-positive results and inadequate recommendations regarding a safe level of physical activity or the level of advancement of the training plan[8,18]. Furthermore, the normality of the VO_{2max} values is often a very important step to determine the cause of the exercise limitation(40).

Practical application of the most accurately derived predictive equations is a better distinction of physiological vs impaired endurance. Moreover, it undoubtedly improves the clinical usage of VO_{2max} assessment for EA examined with the suspected or confirmed CVD or to precisely prepare individualized training plans.

One of the reasons for obtaining very heterogeneous predicted results is the discrepancy in the methodology^{18,24}. Potential complications were mainly related to CPET- usage of cardiology-specific protocols for TE (e.g., Bruce protocol[42]) which are not commonly used in sports-performance diagnostics[43]. Individual running economy, general fatigue, or nonspecific stress during testing rise the probability of bias[17]. The error could be even up to 40% of the actual value[23]. Our study population is larger and contains EA from the individual- or team-sports disciplines. The testing protocol was strictly standardized, and measurements included advanced parameters influencing performance- LA and BM(43).

Specificity of particular subgroups

Outcome of the present study was that the examined prediction equations of VO₂max had low-to-moderate prediction value in the locomotion (running versus cycling), age, and performance subgroups of participants. An explanation of this low-to-moderate prediction value might be due to the selection of specific predictors (sex, age, and weight) that were not measures of CRF. Among the selected predictors, the only mechanical workload was a measure of CRF. CRF in EA consisted not only of a health-related but also a sport-related physical fitness parameter; thus, it would be of great practical importance that predicted VO₂max could reflect changes in sports performance. Furthermore, performance subgroups of participants might differ for body composition (i.e. lower body fat percentage in HTEA than in LTEA), which in turn might consist a bias in the assessment of CRF[45]. Ceaser and Hunter point out that endurance capacity may also depend on participant ethnicity, so this factor should be considered when deriving new models(45,46).

Directions of future research

We recommend that the formulas used to estimate VO₂max should be applied to groups with a similar profile to the one from which they were originally derived, especially in narrow populations like LTEA, REA, or HTEA[37]. At the same time, we emphasize that there is a significant need to create new, more advanced models under unified guidelines and with the incorporation of PROBAST-AI(47) and TRIPOD checklist[30]. It will facilitate the further selection of the appropriate equation to apply in EA depending on their level of CRF. In addition, the need of selecting other predictors, such as oxygen uptake at submaximal exercise intensity, ethnicity, or a daily number of steps, should be considered in future studies.

Conclusions

To conclude, we have accomplished an independent external validation of prognostic models for the prediction of the CRF level, defined as a VO₂max. Each included prognostic model showed only moderate discriminatory ability, but acceptable performance at derivation population. An updated and unified prognostic formula for clinical and experimental use in EA populations is necessary. Despite no formula being completely exact, the best performance was noted for males on the CE in Kokkinos model ($R^2 = 0.64$) and males on the TE in the Wasserman model ($R^2 = 0.26$), whereas for females on the CE in Kokkinos ($R^2 = 0.65$) and female on the TE in Wasserman ($R^2 = 0.30$) equations. Those models seem to better predict VO₂max in our EA population and may provide utility as a method-of-choice in assessment tool during sports diagnostics or clinical practice. The overall lowest model accuracy has been observed for HTEA and EA 18–30 year. A potential limitation of the study was the ethnic homogeneity of our group, as the subjects were mainly Caucasian.

Declarations

Ethical approval

All parts of the study were approved by the Bioethical Committee-IRB of the Medical University of Warsaw (AKBE/32/2021) and were conducted in line with the Declaration of Helsinki. Moreover, each EA has to provide their written consent in a separate document.

Author contributions

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

Founding

This research received no external funding.

Competing interests

The authors declare no competing interests.

Data Availability Statement

Data are available from the corresponding author upon reasonable request. Conflicts of Interest: The authors declare there are no conflicts of interest

References

1. Hill, V., CNH, L., H, L. & B, C. Muscular exercise, lactic acid and the supply and utilisation of oxygen.— Parts VII–VIII. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **97**, 155–176 (1924).
2. Guazzi, M. *et al.* 2016 focused update: Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Circulation* **133**, e694-711 (2016).
3. Ross, R. *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* **134**, (2016).
4. Harber, M. P. *et al.* *Impact of cardiorespiratory fitness on all-cause and disease-specific mortality: Advances since 2009*. vol. 60 11–20 (W.B. Saunders, 2017).
5. Christina, S. *et al.* A reference equation for maximal aerobic power for treadmill and cycle ergometer exercise testing: Analysis from the FRIEND registry. *European Journal of Preventive Cardiology* **25**, 742–750 (2018).
6. Rowell, L. B., Taylor, H. L. & Wang, Y. Limitations to prediction of maximal oxygen intake. *Journal of Applied Physiology* **19**, 919–927 (1964).

7. Bentley, D. J., Newell, J. & Bishop, D. *Incremental exercise test design and analysis*. vol. 37 575–586 (Sports Med, 2007).
8. F, C. E. Integration of the physiological factors determining endurance performance ability. *Exercise and sport sciences reviews* **23**, 25–63 (1995).
9. BASSETT, D. R. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine & Science in Sports & Exercise* **32**, 70 (2000).
10. Hawley, J. A. & Noakes, T. D. Peak power output predicts maximal oxygen uptake and performance time in trained cyclists. *European Journal of Applied Physiology and Occupational Physiology* **65**, 79–83 (1992).
11. D, N. T., H, M. K. & Schall, R. Peak treadmill running velocity during the VO₂ max test predicts running performance. *Journal of sports sciences* **8**, 35–45 (1990).
12. Millet, G. P., Vleck, V. E. & Bentley, D. J. Physiological differences between cycling and running. *Sports Medicine* **39**, 179–206 (2009).
13. Basset, F. A. & Boulay, M. R. Specificity of treadmill and cycle ergometer tests in triathletes, runners and cyclists. *European Journal of Applied Physiology* **81**, 214–221 (2000).
14. Al-Mallah, Mouaz H *et al.* Sex differences in cardiorespiratory fitness and all-cause mortality. in *Mayo Clinic Proceedings* vol. 91 755–762 (Elsevier Ltd, 2016).
15. Stensvold, D. *et al.* Cardiorespiratory reference data in older adults: The generation 100 study. *Medicine and science in sports and exercise* **49**, 2206–2215 (2017).
16. Myers, J. *et al.* Recommendations for clinical exercise laboratories. *Circulation* **119**, 3144–3161 (2009).
17. D, W. P. New ideas on limitations to VO₂max. *Exercise and sport sciences reviews* **28**, 10–4 (2000).
18. Petek, B. J. *et al.* Normative cardiopulmonary exercise data for endurance athletes: the C ardiopulmonary H ealth and E ndurance E xercise R egistry (CHEER). *European Journal of Preventive Cardiology* (2021) doi:10.1093/eurjpc/zwab150.
19. Wasserman, K. *et al.* Principles of exercise testing and interpretation: Including pathophysiology and clinical applications. in (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2012).
20. Wisloff, U. & Lavie, C. J. Taking physical activity, exercise, and fitness to a higher level. *Progress in Cardiovascular Diseases* **60**, 1–2 (2017).
21. Wiecha, S. *et al.* Transferability of cardiopulmonary parameters between treadmill and cycle ergometer testing in male Triathletes—Prediction formulae. *International Journal of Environmental Research and Public Health* **19**, 1830 (2022).
22. Lavie, C. J. *et al.* *Exercise and the cardiovascular system: clinical science and cardiovascular outcomes*. vol. 117 207–19 (Circulation Research. Lippincott Williams and Wilkins, 2015).
23. Malek, M. H., Berger, D. E., Housh, T. J., Coburn, J. W. & Beck, T. W. Validity of VO₂max equations for aerobically trained males and females. *Medicine and science in sports and exercise* **36**, 1427–32 (2004).
24. Lorenzo, S. & Babb, T. G. Quantification of cardiorespiratory fitness in healthy nonobese and obese men and women. *Chest* **141**, 1031–1039 (2012).
25. Peterson, M. J., Pieper, C. F. & Morey, M. C. Accuracy of VO₂(max) prediction equations in older adults. *Medicine and science in sports and exercise* **35**, 145–9 (2003).
26. Kaminsky, L. A. *et al.* The importance of cardiorespiratory fitness in the united states: The need for a national registry. *Circulation* **127**, 652–662 (2013).
27. Collins, G. S., Omar, O., Shanyinde, M. & Yu, L.-M. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *Journal of Clinical Epidemiology* **66**, 268–277 (2013).
28. Debray, T. P. A. *et al.* A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of Clinical Epidemiology* **68**, 279–289 (2015).
29. Moons, *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* **98**, 691–698 (2012).
30. Kaminsky, L. A., Arena, R. & Myers, J. Reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing: Data from the fitness registry and the importance of exercise national database. *Mayo Clinic proceedings* **90**, 1515–23 (2015).
31. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine* **13**, 1 (2015).
32. Lach, J. *et al.* HR max prediction based on age, body composition, fitness level, testing modality and sex in physically active population. *Frontiers in Physiology* **12**, (2021).
33. Gómez-Molina, J. *et al.* Predictive variables of half-marathon performance for male runners. *Journal of sports science & medicine* **16**, 187–194 (2017).
34. Paap, D. & Takken, T. Reference values for cardiopulmonary exercise testing in healthy adults: a systematic review. *Expert review of cardiovascular therapy* **12**, 1439–53 (2014).
35. Esteve-Lanao, J. *et al.* Predicting recreational runners' marathon performance time during their training preparation. *Journal of Strength and Conditioning Research* **35**, 3218–3224 (2021).
36. Takken, T. *et al.* Reference values for cardiopulmonary exercise testing in healthy subjects – an updated systematic review. *Expert Review of Cardiovascular Therapy* **17**, 413–426 (2019).
37. A.T.S./A.C.C.P. ATS/ACCP statement on cardiopulmonary exercise testing. *American Journal of Respiratory and Critical Care Medicine* **167**, 211–277 (2003).
38. Fitzgerald, M. D., Tanaka, H., Tran, Z. V. & Seals, D. R. Age-related declines in maximal aerobic capacity in regularly exercising vs. sedentary women: a meta-analysis. *Journal of Applied Physiology* **83**, 160–165 (1997).

39. Wilson, T. M. & Tanaka, H. Meta-analysis of the age-associated decline in maximal aerobic capacity in men: relation to training status. *American Journal of Physiology-Heart and Circulatory Physiology* **278**, H829–H834 (2000).
40. Radtke, T. *et al.* ERS statement on standardisation of cardiopulmonary exercise testing in chronic lung diseases. *European respiratory review: an official journal of the European Respiratory Society* **28**, (2019).
41. Myers, J. *et al.* Comparison of the ramp versus standard exercise protocols. *Journal of the American College of Cardiology* **17**, 1334–1342 (1991).
42. Muscat, K. M. *et al.* Physiological and perceptual responses to incremental exercise testing in healthy men: effect of exercise test modality. *Applied Physiology, Nutrition, and Metabolism* **40**, 1199–1209 (2015).
43. J, S. E., C, K. S., Gibson, C., A, H. J. & D, N. T. Prediction of triathlon race time from laboratory testing in national triathletes. *Medicine and science in sports and exercise* **32**, 844–9 (2000).
44. Krachler, B., Savonen, K. & Lakka, T. Obesity is an important source of bias in the assessment of cardiorespiratory fitness. *American Heart Journal* **170**, e7–e8 (2015).
45. Ceaser, T. & Hunter, G. Black and white race differences in aerobic capacity, muscle fiber type, and their influence on metabolic processes. *Sports Medicine* **45**, 615–623 (2015).
46. CEASER, T. G., FITZHUGH, E. C., THOMPSON, D. L. & BASSETT, D. R. Association of physical activity, fitness, and race. *Medicine & Science in Sports & Exercise* **45**, 286–293 (2013).
47. Moons, K. G. M. *et al.* *PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration*. vol. 170 W1 (Annals of Internal Medicine, 2019).

Tables

Table 3 is available in the Supplemental Files section.

Figures

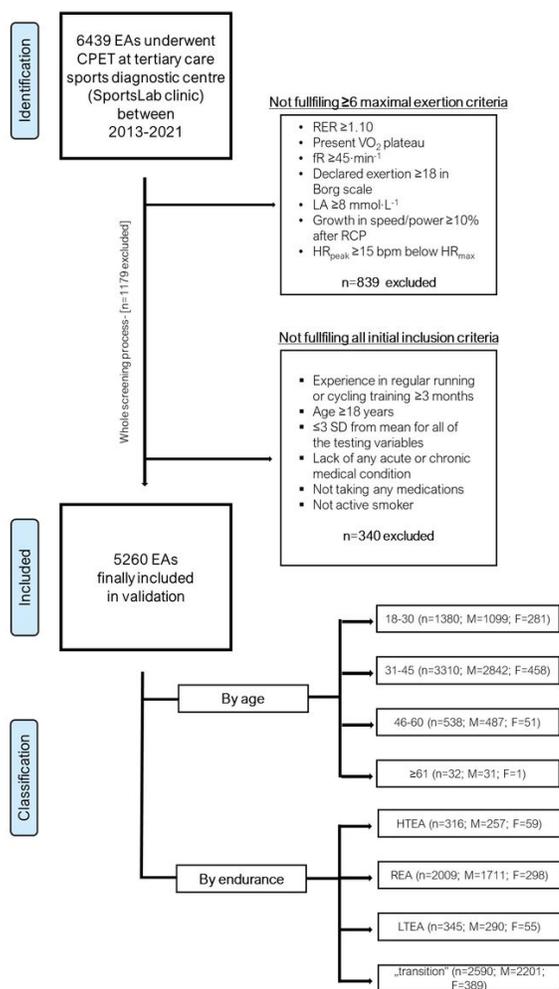


Figure 1

Flowchart of the inclusion-exclusion and further groups classification process. Age classification is presented in years. Endurance classification has been performed based on speed/power at respiratory compensation point (RCP) which is currently described as a variable most closely corresponding to the critical power. Moreover, the selection of a variable different from VO_{2max} to the classification of participants in terms of their endurance capacity, enable to make group assignments independent of the factor directly validated in the study. Abbreviations: EA, endurance athlete; CPET, cardiopulmonary exercise testing; SD, standard deviation; RER, respiratory exchange ratio; VO_2 , oxygen uptake ($mL \cdot min^{-1} \cdot kg^{-1}$); LA, lactate concentration ($mmol \cdot L^{-1}$); fR, breathing frequency ($breaths \cdot min^{-1}$); RCP, respiratory compensation point; HR_{peak} , peak heart rate during CPET (bpm); HR_{max} , maximal heart rate during CPET (bpm); F, female; M, male; HTEA, high-trained endurance athletes; REA, recreational endurance athletes, LTEA, low-trained endurance athletes.

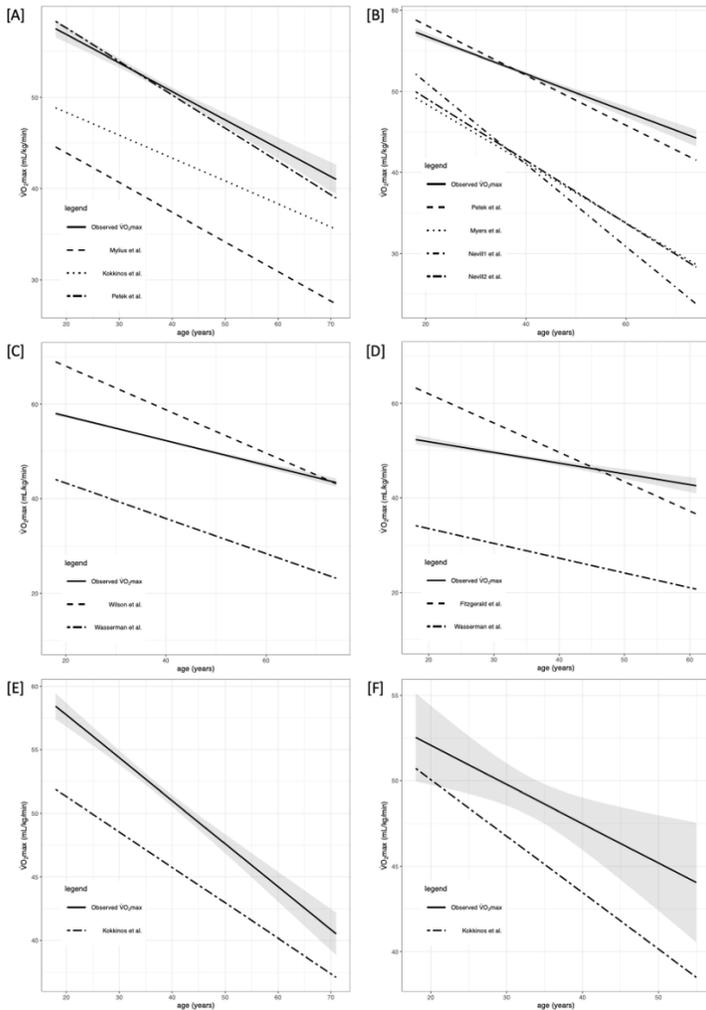


Figure 2
 Regression analysis (RA) of predicted and observed VO_{2max} for whole population (both males and females with included „transition” group) stratified by age. 95% confidence intervals are presented with grey color. Subgroups that did not meet the TRIPOD guidelines to consider their validation results as reliable (i.e. $n \geq 100$) were marked with [‡]. Abbreviations: VO_{2max} , maximal oxygen uptake. [A] RA for males and females for CE models; [B] RA for males and females for TE models; [C] RA for males for TE or CE models (Wilson et al. and Kokkinos et al.[1]); [D] RA for females for TE or CE models (Fitzgerald et al. and Kokkinos et al.[2]); [E] RA for males for TE or CE model (Kokkinos et al.[3]); [F] RA for females for TE or CE model (Kokkinos et al.[3]).

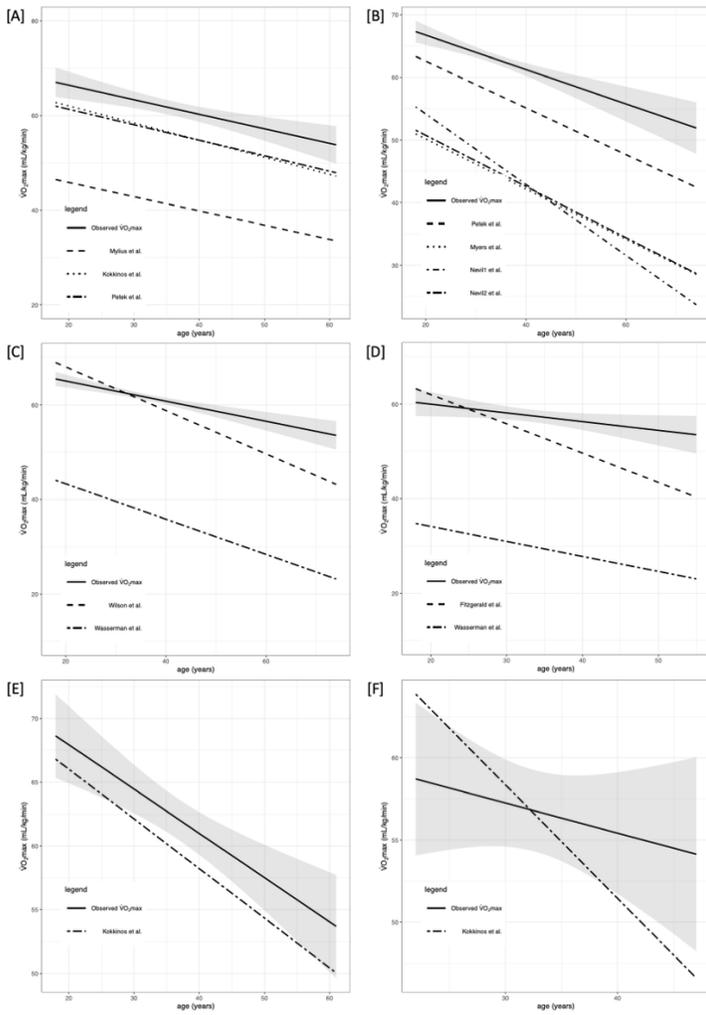


Figure 3
 Regression analysis (RA) of predicted and observed VO_{2max} for HTEA stratified by age. 95% confidence intervals are presented with grey color. Subgroups that did not meet the TRIPOD guidelines to consider their validation results as reliable (i.e. $n \geq 100$) were marked with [‡]. Abbreviations: VO_{2max} , maximal oxygen uptake. [A] RA for males‡ and females‡ for CE models; [B] RA for males and females‡ for TE models; [C] RA for males for TE or CE‡ models (Wilson et al. and Kokkinos et al.[1]); [D] RA for females for TE‡ or CE‡ models (Fitzgerald et al. and Kokkinos et al.[2]); [E] RA for males for TE or CE‡ model (Kokkinos et al.[3]); [F] RA for females for TE‡ or CE‡ model (Kokkinos et al.[3]).

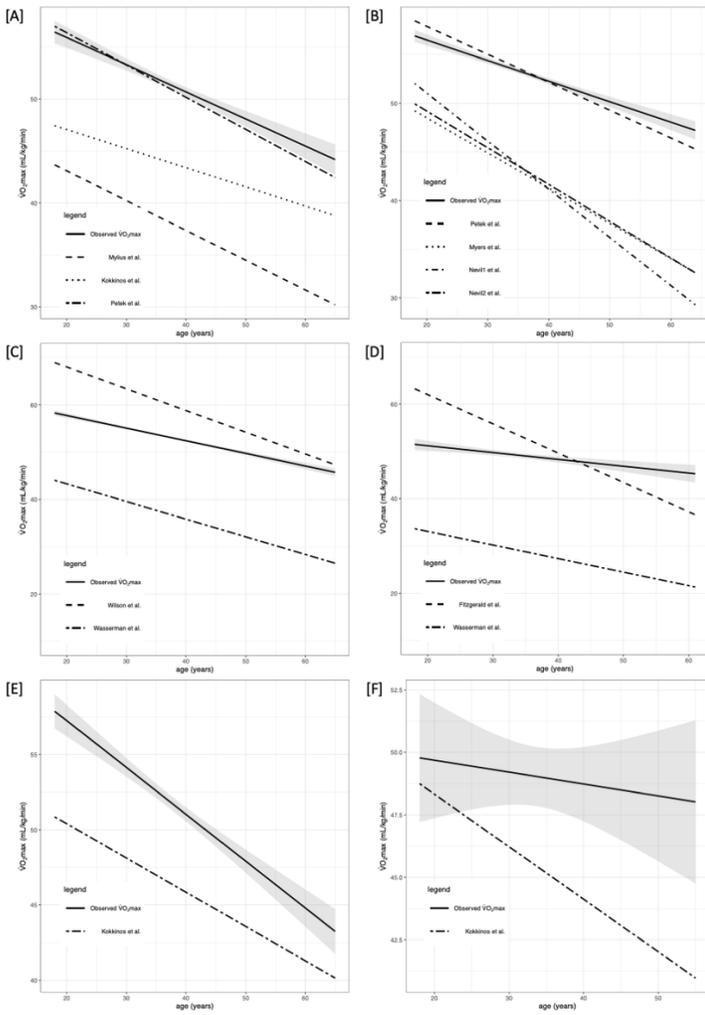


Figure 4
 Regression analysis (RA) of predicted and observed VO_{2max} for REA stratified by age. 95% confidence intervals are presented with grey color. Subgroups that did not meet the TRIPOD guidelines to consider their validation results as reliable (i.e. $n \geq 100$) were marked with ‡. Abbreviations: VO_{2max} , maximal oxygen uptake. [A] RA for males and females‡ for CE models; [B] RA for males and females‡ for TE models; [C] RA for males for TE or CE models (Wilson et al. and Kokkinos et al.[1]); [D] RA for females for TE‡ or CE‡ models (Fitzgerald et al. and Kokkinos et al.[1]); [E] RA for males for TE or CE model (Kokkinos et al.[3]); [F] RA for females for TE‡ or CE‡ model (Kokkinos et al.[3]).

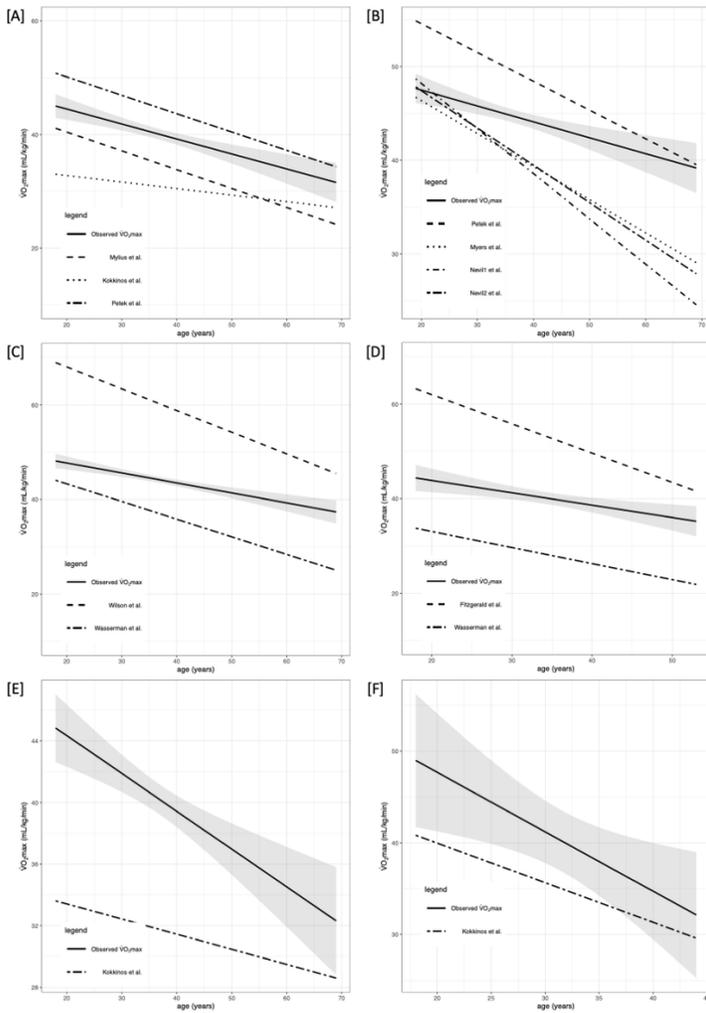


Figure 5

Regression analysis (RA) of predicted and observed VO_{2max} for LTEA stratified by age. 95% confidence intervals are presented with grey color. Subgroups that did not meet the TRIPOD guidelines to consider their validation results as reliable (i.e. $n \geq 100$) were marked with [‡]. Abbreviations: VO_{2max} , maximal oxygen uptake. [A] RA for males‡ and females‡ for CE models; [B] RA for males and females‡ for TE models; [C] RA for males for TE or CE‡ models (Wilson et al. and Kokkinos et al.[1]); [D] RA for females for TE‡ or CE‡ models (Fitzgerald et al. and Kokkinos et al.[2]); [E] RA for males for TE or CE‡ model (Kokkinos et al.[3]); [F] RA for females for TE‡ or CE‡ model (Kokkinos et al.[3]).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table3.docx](#)
- [Supplementaryinformation.pdf](#)