International Journal of EXERCISE SCIENCE

Original Research

Agreement of Air Bike and Treadmill Protocols to Assess Maximal Oxygen Uptake: An Exploratory Study

TRAVIS LIPSCOMB†[1], MEGAN SAX VAN DER WEYDEN†[1], ALI BOOLANI‡[2], BRYAN ST. ANDREWS‡[3], and JOEL MARTIN‡[1]

[1]Sports Medicine Assessment Research & Testing (SMART) Laboratory, George Mason University, Fairfax, VA, USA; [2]Honors Program, Clarkson University, Potsdam, NY, USA; [3]State College Strength and Conditioning, State College, PA, USA

†Denotes graduate student author, ‡Denotes professional author

ABSTRACT

International Journal of Exercise Science 17(4): 633-647, 2024. Maximal oxygen consumption ($\dot{V}O_{2max}$) is an important measure of aerobic fitness, with applications in evaluating fitness, designing training programs, and assessing overall health. While treadmill assessments are considered the gold standard, airbikes (ABs) are increasingly popular exercise machines. However, limited research exists on AB-based $\dot{V}O2max$ assessments, particularly regarding agreement with treadmill graded exercise tests. To address this gap, a randomized crossover study was conducted, involving 15 healthy adults (9M, 6F, 7 familiar with AB) aged 30.1 ± 8.6 years. Paired *t*-tests, intraclass correlation coefficients (ICC), Bland-Altman and Principal component (PC) analyses were used to assess agreement between protocols. The results demonstrated good to excellent agreement in $\dot{V}O_{2max}$, maximum heart rate (HR), and rating of perceived exertion (ICC range: 0.89-0.92). However, significant differences were observed in several measures, including $\dot{V}O_{2max}$ and maximum HR ($p < 0.01$). Overall a systematic bias 3.31 mL/kg/min (treadmill > AB, 95%CI[1.67,4.94]) was observed, no proportional bias was present; however, regular AB users (systematic bias: 1.27 (95%CI[0.20,2.34]) mL/kg/min) exhibited higher agreement in $\dot{V}O_{2max}$ measures compared to non-regular users (systematic bias: 5.09 (95%CI[3.69,6.49]) mL/kg/min). There were no significant differences in cardiorespiratory coordination, between the AB and the treadmill. These findings suggest that for individuals familiar with the AB, it can be a suitable alternative for assessing $\dot{V}O_{2max}$ compared to the treadmill. Future research with larger samples should focus on developing prediction equations for field AB tests to predict $\dot{V}O_{2max}$. Practitioners should consider using the AB to assess $\dot{V}O_{2max}$ in individuals who prefer it over running.

KEY WORDS: $\dot{V}O_{2max}$, aerobic fitness assessment, exercise physiology, cardiorespiratory, maximum heart rate

INTRODUCTION

A maximal assessment on a treadmill is widely considered the gold standard for assessing aerobic fitness, as determined by $\dot{V}O_{2max}$ (23). However, additional modes of assessment could be safer for maximal exercise testing as there is a risk of injury from falling off a treadmill and more viable for injured populations who cannot withstand impact (10, 28). Different exercise modes, such as swimming, the cycle ergometry, and arm-crank ergometer protocols exercise, tend to yield lower $\dot{V}O_{2max}$ values compared to the treadmill assessment in the general population (29). However, it elite swimmers and cyclists have demonstrated similar $\dot{V}O_{2max}$ values during a swimming, cycle ergometer, respectively, when compared to their treadmill $\dot{V}O_{2max}$ protocol values (29). Ultimately, each mode of $\dot{V}O_{2max}$ testing has its own advantages and disadvantages, and the choice of mode may depend on the individual's preferences and physical capabilities. A non-treadmill running $\dot{V}O_{2max}$ protocol would be valuable as many individuals do not frequently engage in running or may have inefficient running technique (11).

An airbike (AB), also known as fan bike, is a type of stationary exercise bike that uses air resistance. ABs feature a large fan or fan blades at the front of the bike that generates resistance and are designed to spin faster as you pedal harder, creating more resistance and requiring more effort. ABs typically have dual-action handlebars that move back and forth; therefore the participant engages their upper and lower body for a full-body workout (57). In recent years, ABs have become popular in various groups, including CrossFit, the military, professional sports teams, and the general public. However, there is limited research focused on $\dot{V}O_{2max}$ assessments utilizing AB. Most existing studies have examined the effects of AB training on increases in $\dot{V}O_{2max}$ (8) and cardiorespiratory adaptations through sprint interval training on an AB (12). Kokinda and colleagues compared $\dot{V}O_{2max}$ values obtained from an Ergoline Ergoselect 200P cycle ergometer and an AirBike Renegade Pro in 5 power athletes, showing that the AB assessment was shorter in duration and yielded slightly higher median $\dot{V}O_{2max}$ values (~3 ml/kg/min) compared to the cycle ergometer (24). The small number of participants, specificity of the participants' training, and lack of comparison to treadmill-based $\dot{V}O_{2max}$ test are key limitations of the study (24). Establishing a valid and reliable AB $\dot{V}O_{2max}$ testing protocol that adheres to standard guidelines is of value considering the popularity of the AB and the importance of aerobic fitness. Such a protocol would provide an alternative testing option for individuals who prefer not to or cannot run on a treadmill (29). Moreover, developing an AB $\dot{V}O_{2max}$ protocol would enable individuals to use familiar training equipment and tailor their training based on specific heart rate (HR) zones to enhance their ventilatory threshold, $\dot{V}O_{2max}$, and overall aerobic conditioning.

The purpose of the present study was to develop and assess a new protocol for assessing $\dot{V}O_{2max}$ on an AB. Within this overall purpose, the first aim of this study was to investigate the agreement of an individual's $\dot{V}O_{2max}$ during a conventional treadmill assessment and an AB assessment. Given that ABs engage both upper and lower body muscles, it was hypothesized that the AB protocol would yield $\dot{V}O_{2max}$ values comparable to the treadmill protocol (29). Secondly, based on the previously mentioned effects of training experience on $\dot{V}O_{2max}$ protocol validity (29, 30), the effect of participants' familiarity with the AB on the agreement of $\dot{V}O_{2max}$ values with the treadmill assessment was investigated. It was hypothesized that those familiar, regularly using the AB, would display better agreement between the $\dot{V}O_{2max}$ on the AB with a treadmill.

METHODS

Participants

Fifteen participants (6 females, 9 males) completed the experimental procedures. To be eligible for the intervention, participants were required to: (1) be healthy adults 18-45 years of age; (2) not have surgery or injury in the last 3 months; (3) ability to run or bike without pain; (4) no history of cardiovascular, respiratory, or musculoskeletal diseases or use of medications that affect the responsiveness of the cardiovascular system; and (5) engage on average in 30 min of physical activity daily. All participants were informed of the benefits and risks of the study and signed the informed consent. The sample size was arrived at based on a combination of an a priori power analysis (17), prior literature (9, 24, 32) and practical limitations (27) due to resources (i.e., unfunded project, scheduling logistics, and lack of compensation for participants, etc.) available for the study. An a priori power analysis was conducted using G*Power to determine the sufficient sample size. Given the exploratory nature of the study the parameters used were an alpha of 0.05, power of 0.80, a large effect size (d = 0.8) and two tails which yielded a desired sample of 15 participants (17). The study was approved by George Mason University's Institutional Review Board (IRB #: 2028014). This research was carried out fully in accordance to the ethical standards of the International Journal of Exercise Science (31).

Protocol

A randomized, controlled crossover design was employed for this study, where participants performed $\dot{V}O_{2max}$ protocols on an AB and treadmill during two testing sessions (Figure 1). Prior to the testing, participants were instructed to refrain from engaging in intense exercise for 48 hours and to abstain from eating or drinking for 2 hours, although water intake was allowed after the body composition assessment. Upon arrival at the laboratory, participants provided informed consent and completed electronic questionnaires related to their frequency of AB use, mood, personality, and lifestyle behaviors. Resting heart rate and anthropometric measures, including height, mass, and body composition, were recorded in the first data collection session. Subsequently, participants were randomly assigned to either the AB or treadmill $\dot{V}O_{2max}$ protocol and completed the corresponding assessment. During the second data collection session, participants performed the alternative $\dot{V}O_{2max}$ protocol. Standard instructions were provided to all participants, who were also familiarized with the assessments beforehand. During the familiarization, participants were provided demonstrations of the assessments, performed by the researcher, and required to perform at least 2 practice attempts to demonstrate proficiency performing the assessments. The researcher used their professional discretion if any participants needed more practice attempts. To mitigate the short-term effects of fatigue on $\dot{V}O_{2max}$ and minimize the possibility of true changes in $\dot{V}O_{2max}$ between testing sessions, all participants were required to return for the second session within a window of 48 hours to 10 days (21, 35, 38). The second $\dot{V}O_{2max}$ test was performed during the same time of day (± 2 hours) to avoid any large effects of circadian rhythm on performance (14). The testing sessions took place in the same exercise physiology laboratory, where at least 1 researcher held a certification as a National Strength and Conditioning Association Certified Strength and Conditioning Specialist (CSCS).

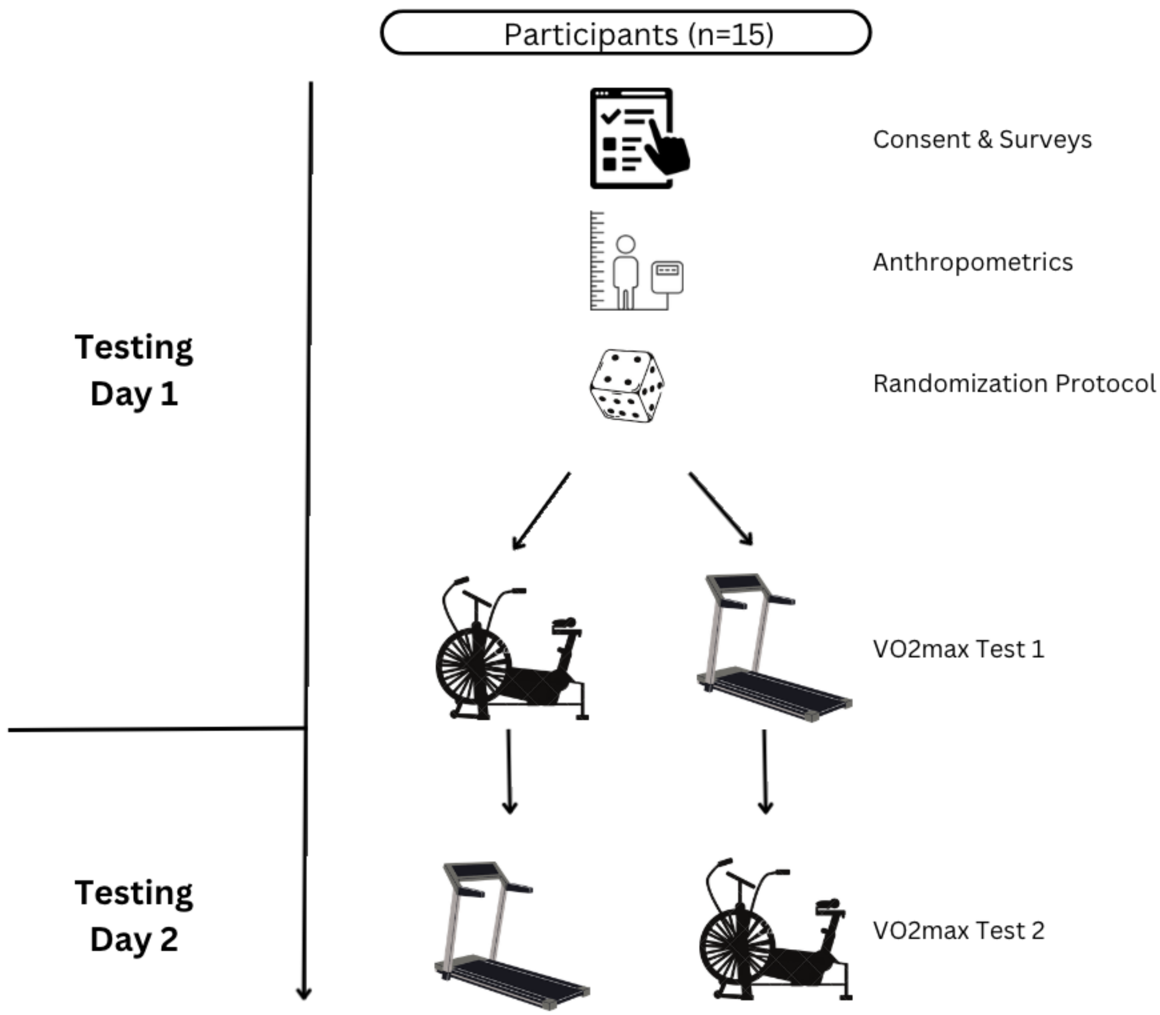

Figure 1. Overview of experimental procedures.
Note: There were 6.5 ± 1.7 (M ± SD) days between day 1 and 2 for participants.

Height and mass were recorded using a stadiometer (Detecto, Webb City, MO, USA) and scale (BOD POD model 2000A; Cosmed USA, Concord, CA, USA). Percent body fat, fat mass, and fat-free mass were measured using air displacement plethysmography (BOD POD model 2000A; BOD POD, COSMED USA, Concord, CA, USA) following standardized procedures by the manufacturer. Prior to testing, the BOD POD was warmed up and calibrated according to the manufacturer's procedures (18). Air displacement plethysmography has been shown to be a reliable and valid method of assessing body composition (15).

A novel AB VO_{2max} testing protocol was developed through pilot testing that incorporated key elements of a graded exercise test, including a warm-up period, a test duration of 8-12 minutes, and a linear increase in intensity. Two separate AB protocols were employed in this study for male and female participants, respectively, to equate the times to exhaustion. Specifically, male participants maintained a power output of 100 watts for the first 60 seconds, followed by a 25-

watt increase each minute until they could no longer maintain the required wattage. Female participants, began with a power output of 80 watts and increased by 20 watts per minute until they could no longer maintain the required power output. In our pilot testing, we opted to increase the intensity of each stage by a power output (males: 25 watts, females: 20 watts) rather than RPM, as AB power output has a non-linear relationship with RPM. By increasing the intensity through power output rather than RPM the current study ensured consistent increases workload whereas small changes in RPM drastically affects the change in power output on a Rogue Echo AB (model used in our study). For treadmill testing, participants underwent the Wellness-Fitness Initiative (WFI) Treadmill ramp protocol, which has been validated in tactical athlete populations (16). The protocol started with 30 seconds at 1.5 mph followed by 3 minutes at 3 mph. Subsequently, the intensity of the protocol increased every minute by alternating between increments in speed (0.5 mph) and incline (2% increase).

During both protocols, physiological variables were continuously monitored and recorded during all stages of the test. Indirect calorimetry was conducted with a calibrated metabolic cart (TrueOne 2400, Parvo Medics, Salt Lake City, UT, USA) and metabolic data were obtained in 15 second intervals. HR was continuously recorded with a wearable chest strap monitor (H10, Polar-Electro, Kempele, Finland). Exertion was self-reported at each stage of the test via the 6 to 20 Borg rating of perceived exertion (RPE) scale (58). Prior to that start of the graded exercise test, participants were shown a visual of the 15-point RPE scale and were verbally instructed that 6 was considered extremely easy and 20 was maximal effort. In the last 30 seconds of each stage, the researcher held up the RPE visual and participants were instructed to point or verbally indicate the number that corresponded with their physical efforts. The termination criteria of VO_2 plateau, respiratory exchange ratio (RER) of 1.10 or higher, and age-predicted HR maximum were monitored for both protocols (1, 4, 29). The criteria were used to determine valid maximal effort by participants and termination of the test occurred when participants either signaled they desired to stop or physically stopped.

Statistical Analysis

A combination of histograms and Shapiro-Wilks tests for normality were used to examine the distribution of the data. Descriptive characteristics were computed for all demographics and measures from the $\dot{V}O_{2max}$ testing. Individual paired *t*-tests were used to determine whether significant differences in measures existed between the modes of $\dot{V}O_{2max}$ testing for all participants (n = 15). Individual paired *t*-tests were also used to determine whether significant differences in measures existed between the modes of $\dot{V}O_{2max}$ for a subgroup of those who reported regular use of the AB (n = 7). For the purposes of the study, regular use of the AB was defined to be using the AB at least 1 time per week for the past 6 months. To further assess agreement between the treadmill and AB, intraclass correlation coefficients (ICC)(25) and Bland-Altman plots were utilized (5). Confidence intervals for ICC estimates, bias and limits of agreements were also calculated to determine the precision of estimated limits of agreement (6). ICC were interpreted as poor (ICC < 0.5), moderate (ICC ≥ 0.5 to ICC < 0.75), good (ICC ≥ 0.75 to ICC < 0.9), and excellent (ICC ≥ 0.9). Proportional bias was assessed by linear regression between the differences in the results obtained with each VO_2 protocol with the $\dot{V}O_{2max}$ from the WFI treadmill protocol.

To assess the agreement between $\dot{V}O_{2max}$ tests on an AB with a treadmill, the network physiology framework (3, 22) may provide greater insight than more traditional measures of agreement, such as intraclass correlation coefficients (25) and Bland-Altman analyses (5). The theory of network physiology of exercise (NPE) provides a comprehensive framework to understand the non-linear dynamic interactions among different physiological systems, such as the synergism between the cardiovascular and respiratory systems during exercise stimuli (3, 22). Traditional analyses, such as correlation and Bland-Altman analyses of assessing agreement of $\dot{V}O_{2max}$ testing are limited in addressing these complex interactions as they are focused on linear relationships between measures. The NPE framework, on the other hand, utilizes non-linear modeling and time series analysis to investigate the coordination and synchronization of physiological systems (34). A method within this NPE framework, called cardiorespiratory coordination (CRC), assesses the synergism between cardiorespiratory variables using principal component analysis (PCA) (2, 19, 32). Papadakis and colleagues recently found the CRC method was able to distinguish novice and intermediate collegiate rowers completing maximal aerobic tests to exhaustion on a rowing ergometer (32). For CRC analysis, the sklearn library in Python was used to conduct PCA analyses on time series data of ventilation (VE), fraction of expired oxygen (FeO_2), fraction of expired carbon dioxide ($FeCO_2$), and HR (32), excluding variables such as oxygen ventilatory equivalent ($VEqO_2$), carbon dioxide ventilatory equivalent ($VEqCO_2$), oxygen (O_2) pulse, RER, and VO_2 due to their linear relationship with the selected variables. The suitability of the PCA implementation was assessed through Bartlett's test for sphericity and the Kaiser Mayer-Olkin (KMO) test. The number of significant PCs was determined using the Kaiser-Gutmann criterion (eigenvalues > = 1.00). Due to the linear nature of the longitudinal data for HR, VE, FeO_2 and $FeCO_2$, missing values for data were accounted for using K-Nearest Neighbor ($k = 5$) method (62). There were 2 participants who had an issue with the heart rate monitor during testing and the signal was not recorded continuously; however, the values at the end of test were obtained. Further, to account for outliers, data were winsorized at the 99% CI (20). The data (HR, VE, FeO_2, $FeCO_2$) were then normalized using the StandardScaler function in sklearn prior to performing the PCA analysis. CRC comparisons between the treadmill and AB protocols were performed on PC1 eigenvalues using the Wilcoxon signed rank test and number of PCs using Chi-Square goodness of fit test. Effect size (Cohen's *d*) was calculated when applicable to indicate standardized mean differences (26, 36). Cohen's d values were interpreted as negligible ($d < 0.2$), small ($d = 0.2$-0.5), medium ($d > 0.5$-0.8) and large ($d > 0.8$) (13). A Benjamini-Hochberg false discovery rate of 10% was used to control for multiple analyses (37). All CRC analyses were conducted in Python 3.11.0 while all other analyses were completed using the R Environment (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $\alpha < 0.05$.

RESULTS

Participant (*n* = 15) demographics and anthropometrics are provided in Table 1. Regular AB users were found to be older (Mean = 9.5 years, Standard Deviation = 6.7, W = 8.5, $p = 0.027$, d = 1.27); however, there were no differences in any of the anthropometric measures between AB regular and non-regular users. All subsequent analyses met the Benjamini-Hochberg false discovery rate of 10%.

Table 1. Participant characteristics.

	All (*n* = 15)			AB Regular Users (*n* = 7)	AB Non-Regular Users (*n* = 8)
	Mean (SD)	Minimum	Maximum	Mean (SD)	Mean (SD)
Age	28.9 (8.3)	21	53	34.8 (10.5)	25.3 (2.7)
Height (cm)	172.5 (7.8)	152.5	186.5	173.5 (7.4)	174.0 (5.2)
Body Mass (kg)	74.6 (15.5)	54.1	110.3	76.1 (12.0)	75.9 (17.5)
BF (%)	19.0 (6.0)	10.2	30.8	17.0 (5.0)	21.0 (7.0)
FFM (kg)	63.5 (11.2)	44.2	86.6	63.3 (12.7)	65.7 (9.0)
FM (kg)	15.9 (9.1)	8.4	40.7	12.8 (3.2)	18.6 (11.4)
Resting HR (bpm)	62.3 (11.8)	47	91	59.2 (16.0)	64.9 (9.2)

Agreement measures between the $\dot{V}O_{2max}$ protocols were evaluated using several statistical analyses, including ICC (Table 2). The ICC results demonstrated good to excellent agreement in $\dot{V}O_{2max}$ (ICC = 0.92 [0.32, 0.98], F(14,14) = 27, $p < 0.001$), maximum HR (ICC = 0.89 [0.49, 0.97], F(14,14) = 14, $p < 0.001$), and RPE (ICC = 0.91 [0.74, 0.97], F(14,14) = 11, $p < 0.001$). However, paired *t*-tests indicated significant differences between the $\dot{V}O_{2max}$ protocols in several measures. Specifically, there were significant differences observed in $\dot{V}O_{2max}$ (t(14) = 4.344, $p < 0.001$), maximum HR (t(14) = 3.137, $p = 0.007$), HR at ventilatory threshold (t(14) = 3.543, $p = 0.003$), and test duration (t(14) = 5.572, $p < 0.001$). Furthermore, Bland-Altman analyses revealed a systematic bias of 3.31 mL/kg/min (treadmill > AB, 95%CI[1.67, 4.94]), with a lower limit of agreement of -2.59 (95%CI[-5.42, 0.24]) and an upper limit of agreement of 9.20 (95%CI[6.38, 12.03]). Notably, no proportional bias was observed between the two protocols (Figure 2).

There was no difference in $\dot{V}O_{2max}$ on the AB between participants who reported regular use of the AB compared to those who did not use the AB regularly (W = 13.5, $p = 0.105$, $d = 0.80$). The only $\dot{V}O_{2max}$ test parameter found to differ between the regular and non-regular AB users was test duration (regular user > non-regular user duration; W = 9, $p = 0.029$, $d = 1.26$). However, subsequent analyses (Table 3) revealed a higher level of agreement in $\dot{V}O_{2max}$ measures among participants who regularly use the AB (Figure 3) compared to those who do not (Figure 4). Specifically, a systematic bias of 1.27 (95%CI[0.20, 2.34]) mL/kg/min was observed among regular AB users, with a lower and upper limit of agreement of -2.60 (95%CI[-4.46, -0.74]) and 5.15 (95%CI[3.29, 7.00]) mL/kg/min, respectively. Whereas, participants who did not regularly use the AB exhibited a systematic bias of 5.09 (95%CI[3.69, 6.49]) mL/kg/min, with a lower limit of agreement of 0.03 (95%CI[-2.40, 2.45]) mL/kg/min and an upper limit of agreement of 10.15 (95%CI[7.72, 12.58]) mL/kg/min. No proportional bias was found for either group based on AB user experience. Notably, no significant differences were found in $\dot{V}O_{2max}$, RER, or maximum

HR obtained from the maximal tests performed on the treadmill and AB for participants who regularly use the AB (Table 3).

Table 2. Comparison of VO_{2max} measures between treadmill and airbike protocols.

	Treadmill M (SD)	Airbike M (SD)	Paired *t*-test		Intraclass Correlation Coefficient	
			P-value	ES	ICC [95% CI]	*p*-value
VO_{2max} (mL/kg/min)	45.1 (7.5)	41.8 (7.8)	< 0.001	0.42	0.92 [0.32,0.98]	< 0.001
RER	1.16 (0.06)	1.13 (0.04)	0.113	0.52	0.40 [-0.54,0.79]	0.15
Max HR (bpm)	185.1 (10.9)	182.3 (9.5)	0.007	0.42	0.89 [0.49,0.97]	< 0.001
Final RPE	19.0 (1.3)	19.1 (1.3)	0.498	0.10	0.91 [0.74, 0.97]	< 0.001
VO_{2max} at Ventilatory Threshold (%)	68.0 (9.0)	59.0 (17.0)	0.065	0.64	0.34 [-0.58, 0.76]	0.19
HR at Ventilatory Threshold (bpm)	151.9 (14.9)	130.9 (28.0)	0.003	1.00	0.43 [-0.28, 0.78]	0.064
Duration (s)	731.7 (100.9)	567.7 (135.3)	< 0.001	1.38	0.44 [-0.26, -0.80]	0.016
CRC Eigenvalue	2.51 (0.27)	2.82 (0.53)	0.132	0.73	-	-

Notes:
1) The *p*-value and effect sizes are results from paired *t*-tests.
2) Eigenvalues were compared with a Wilcoxon signed rank test.
3) The ICC(2,1), intraclass correlation coefficients, are presented in the final column with 95% confident intervals. ICC were not computed for the eigenvalues.
4) The treadmill protocol included a 3 minute (180 second) warm-up period).
5) Abbreviations: M, mean; SD, standard deviation; ES, effect size; CI, confidence interval; CRC, Cardiorespiratory Coordination.

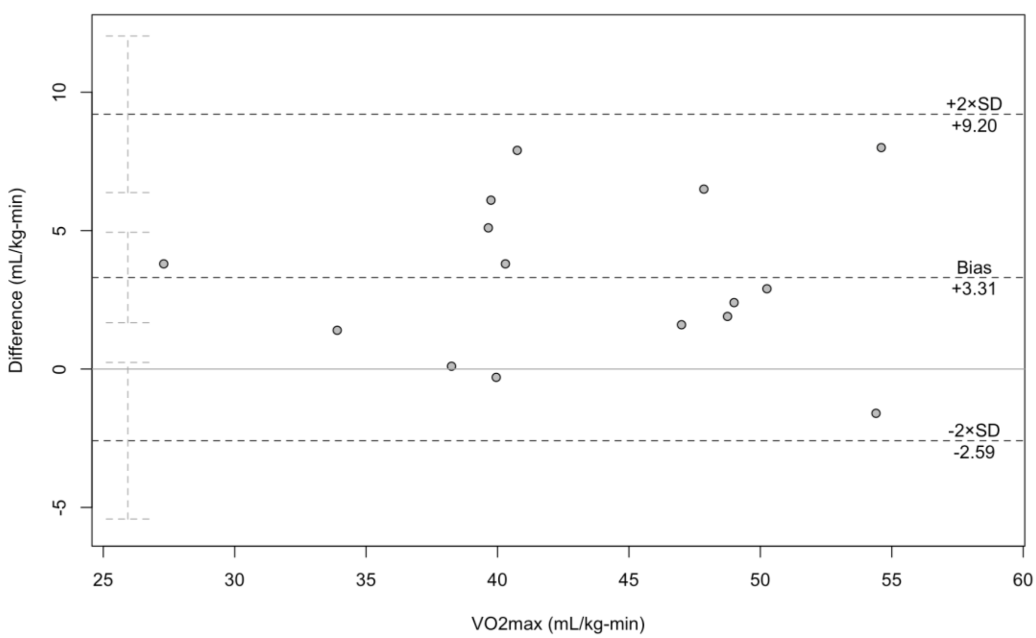

Figure 2. Bland-Altman plot of agreement for treadmill versus airbike VO_{2max} protocols for all participants.

Table 3. Comparison of VO_{2max} measures between treadmill and airbike protocols for regular and non-regular airbike users.

	Regular Airbike Users (*n* = 7)				Non-regular Airbike Users (*n* = 8)			
	Treadmill Median (Q1-Q3)	Airbike Median (Q1-Q3)	*P*-value	ES	Treadmill Median (Q1-Q3)	Airbike Median (Q1-Q3)	*P*-value	ES
VO_{2max} (mL/kg/min)	47.8 (41.0-51.0)	46.2 (39.3-48.3)	0.219	0.20	43.75 (40.30-50.05)	36.95 (35.83-45.4)	**0.008**	0.58
RER	1.15 (1.11-1.16)	1.13 (1.11-1.16)	0.932	0.07	1.19 (1.14-1.22)	1.13 (1.12-1.15)	**0.042**	0.92
Maximum HR (bpm)	186.0 (173.0-192.0)	187. (172.5-192.5)	0.400	0.08	189.0 (186.0-191.2)	180.5 (179.0-185.0)	**0.016**	1.26
Final RPE	20.0 (19.0-20.0)	20.0 (19.0-20.0)	1.000	0.00	19.0 (17.0-20.0)	19.5 (17.8-20.0)	0.484	0.15
VO_{2max} at Ventilatory Threshold (%)	72.0 (71.5-76.5)	64.0 (47.0-74.0)	0.078	0.96	62.5 (59.5-66.3)	51.5 (49.5-57.3)	0.383	0.40
HR at Ventilatory Threshold (bpm)	158.0 (151.5-170.0)	136.0 (129.5-162.5)	0.078	0.86	148.5 (141.2-155.8)	112.0 (101.8-129.2)	**0.039**	1.26
Duration (s)	788.0 (658.5-824.0)	624.0 (594.0-652.0)	0.078	1.00	762.0 (678.8-788.5)	491.0 (441.5-579.0)	**0.008**	1.87
CRC Eigenvalue	2.59 (2.28-2.68)	2.59 (2.53-3.13)	0.208	0.94	2.47 (2.38-2.74)	2.95 (2.31-3.34)	0.313	0.55

Abbreviations: IQR, interquartile range; ES, effect size; HR, heart rate; RPE, rating of perceived effort; VT, ventilatory threshold; CRC, Cardiorespiratory Coordination. Note: Due to small sample sizes the non-parametric Wilcoxon signed rank test (paired samples) was used to assess differences between the treadmill and airbike measures.

There were no significant differences (V = 28, $p = 0.132$) in the eigenvalues of PC1 (Table 2), which accounted for the highest proportion of the data variance, between the AB (Mean = 2.82; SD = 0.53) and the treadmill (Mean = 2.52; SD = 0.27; Supplementary Table 1). However, there were significant differences in the projections of HR (treadmill > AB, V = 91, $p = 0.017$, $d = 1.20$) and fraction of expired oxygen (treadmill < AB, V = 6, $p = 0.004$, $d = 1.27$) in PC1. Notably, a greater number of participants ($n = 15$) exhibited a second principal component for the treadmill (86.7%, $n = 13$) compared to the AB (46.7%, $n = 7$); however, a chi-square goodness of fit indicated no significant difference in the number of PCs ($X^2(1, N = 15) = 3.75$, $p = 0.053$). Further analyses were performed to examine possible disparities in the principal component analysis (PCA) outcomes between and within individuals who regularly use the AB and those who do not. However, none of the PCA metrics, including eigenvalues and PC1 projections, exhibited statistically significant differences between experienced AB users and inexperienced individuals. Thus, the CRC results provided additional insight than the traditional ICC and Bland-Altman analyses. Suggesting that while the overall responses to each protocol were similar there were some subtle differences in the dynamical coordination of the cardiovascular and respiratory systems for some, but not all, participants.

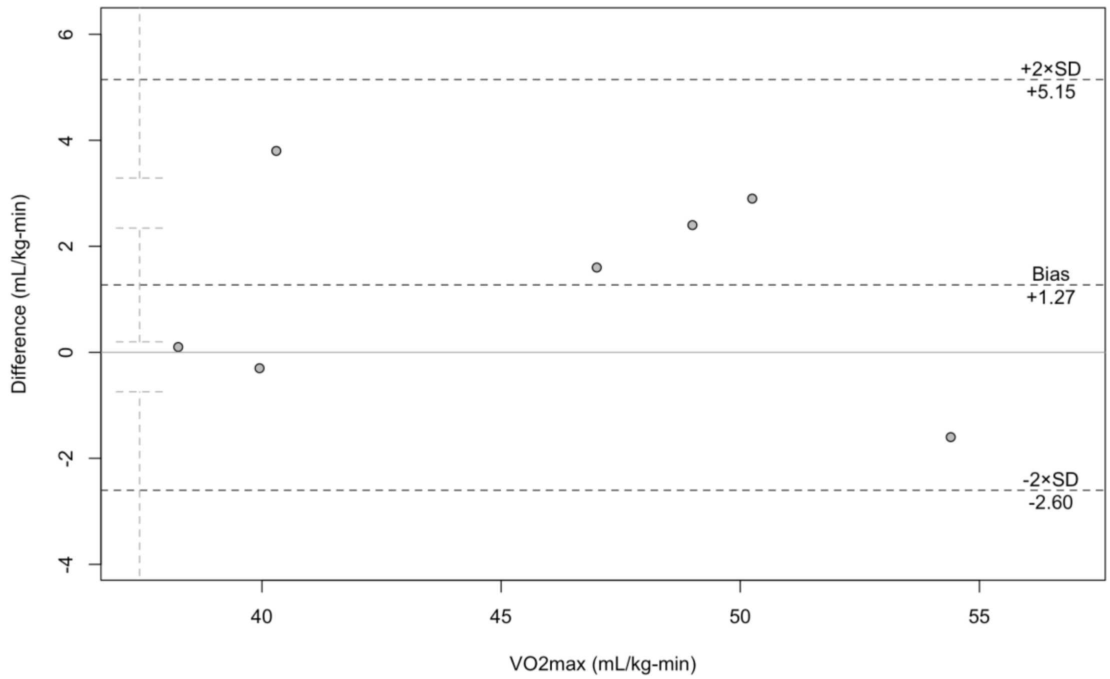

Figure 3. Bland-Altman plot of agreement for treadmill versus airbike VO_{2max} protocols for regular airbike users.

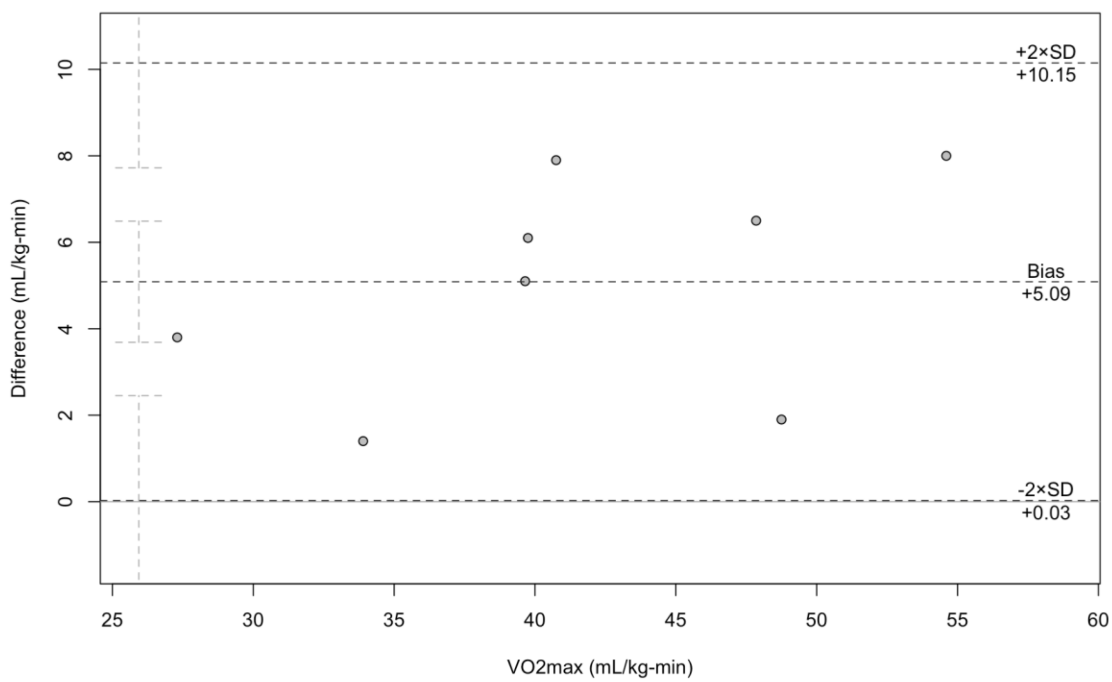

Figure 4. Bland-Altman plot of agreement for treadmill versus airbike VO_{2max} protocols for non regular airbike users.

DISCUSSION

The primary aim of this study was to examine the agreement between $\dot{V}O_{2max}$ values obtained from a traditional treadmill assessment and an AB assessment, while also investigating the potential impact of familiarity with the AB on this agreement. Our hypotheses were partially supported by the findings. Significant differences were observed in $\dot{V}O_{2max}$, maximum HR, HR at ventilatory threshold, and duration between the two assessment methods. However, there were no differences in CRC metrics, indicating a comparable response of the cardiorespiratory system to both protocols. Notably, among individuals who regularly use the AB, a stronger agreement was observed between the treadmill and AB results, supporting previous literature emphasizing the importance of considering an individual's training history when selecting the mode of $\dot{V}O_{2max}$ testing (30). Thus, the findings suggest that while there is insufficient evidence to support the use of the AB for assessing $\dot{V}O_{2max}$ in all individuals, it appears to be a suitable alternative to treadmill testing for those who are familiar with the AB.

As previously mentioned, prior research (24) compared the AB to cycle ergometer protocol and found a median difference of about 3 mL/kg/min in $\dot{V}O_{2max}$ values, with the AB testing producing greater values than the cycle ergometer. In addition to differences in sample size and inclusion criteria, another notable distinction between our study was the duration of the AB protocols (24). This prior study (24) reported an average AB protocol duration of 252 seconds (~4.2 minutes), while our study had an average AB protocol duration of 567.7 seconds (~9.5 minutes). This is likely attributed to the methodology of the previous study (24), which used a protocol starting at a baseline pedaling speed of 55 rpm and increasing by 3 rpm every 20 seconds. Considering the recommended duration for a $\dot{V}O_{2max}$ test of 8 to 12 minutes (29), the protocol used in our study appears to be more appropriate for a broader population. Another study (9) compared the agreement between $\dot{V}O_{2max}$ tests on an AB and a cycle ergometer and found that the AB testing produced a greater $\dot{V}O_{2max}$ than on a cycle ergometer by about 9% (AB $\dot{V}O_{2max}$: 53.1 ± 8.7 mL/kg/min; Cycle ergometer $\dot{V}O_{2max}$: 47.4 ± 9.2 mL/kg/min) in 18 physically active young males men. Notably, Canário-Lemos et al. (9) had participants start at an RPM of 30 (equivalent to 27 watts) and increase by 5 rpm every minute on the AB, with a mean AB testing duration of 598 seconds (~10 minutes), closer to the duration in our study (567.7 seconds). This approach ensured a more graded test and prevented premature termination of the AB testing. Collectively, these studies offer contrasting viewpoints on the validity of the AB protocol for $\dot{V}O_{2max}$ assessment (9,24). The divergent populations and protocol durations indicate the need for additional research to establish a comprehensive understanding of the AB's suitability as a method for measuring $\dot{V}O_{2max}$.

Interestingly, while some of the $\dot{V}O_{2max}$ test metrics showed similarities between the protocols (i.e., RER, final RPE, % $\dot{V}O_{2max}$ at ventilatory threshold, CRC), other metrics exhibited differences (i.e., maximum HR, HR at ventilatory threshold; Table 2). When considering all participants, the treadmill protocol elicited higher exercise intensity in terms of $\dot{V}O_{2max}$, maximum HR, and HR at ventilatory threshold. This discrepancy could be attributed to the greater intensity of the treadmill protocol, resulting in statistically higher values for these physiological measure or the duration of the treadmill test being longer. However, it is important to note that the difference in test duration between the treadmill and AB protocols in this study may be misleading, as the WFI protocol included a 210-second walking warm-up (8). When considering the actual exercise duration, the treadmill duration without the warm-up was 521.7 seconds, which is closer to the average duration of the AB protocol (567.7 seconds), but still significantly shorter. Participants subjectively reported that the initial minutes of the AB protocol were primarily a warm-up based on their RPE. Another plausible explanation for the higher HR-related measures during the treadmill protocol is the documented difference in HR response between running and cycling at similar intensities (30). During maximal exercise testing, the HR response has typically been found to be approximately 5% lower during an incremental cycle test compared to a treadmill test (30, 33). Training specificity is a factor known to influence HR response during maximal exercise testing (30), which leads to what we believe explains the differences in results for AB and non-AB users of this study.

Although there was a significant difference in $\dot{V}O_{2max}$ between the treadmill and AB protocols for all participants, these differences tended to be primarily driven by individuals who were not regular AB users. It is widely recognized that untrained individuals or recreational athletes typically achieve higher $\dot{V}O_{2max}$ values on maximal treadmill tests compared to other modes, such as cycling (29). Conversely, highly trained individuals who specialize in a specific mode of exercise generally attain higher $\dot{V}O_{2max}$ values in their trained mode (30). Bouckaert et al. conducted a study on highly trained runners and cyclists, finding a 14% higher $\dot{V}O_{2max}$ on the treadmill for runners, while cyclists had an 11% higher $\dot{V}O_{2max}$ on the bicycle ergometer (7). Triathletes, who engage in multiple modes of exercise, offer a unique opportunity to investigate $\dot{V}O_{2max}$. Consistent research findings indicate that regular training in multiple modes results in similar $\dot{V}O_{2max}$ values, regardless of the testing mode employed (30). Our findings further emphasize the importance of considering training history and associated adaptations when selecting testing protocols and interpreting $\dot{V}O_{2max}$ results.

The present study has several limitations. First, the sample of 15 healthy adults, including both sexes and a wide age range of 21 to 53, may restrict the generalizability of the findings to specific populations of interest. To enhance the applicability of the results, future research should use larger samples and include participants from specific populations relevant to the study's objectives. Furthermore, developing a prediction equation for estimating $\dot{V}O_{2max}$ from a field test on the AB would be valuable for practical applications. A larger sample size would allow for the development of a robust prediction equation incorporating multiple predictors such as age, sex, body mass, and HR. Lastly, it is important to note that our assessment of regular AB users was based on self-reporting, and the categorization criterion of at least once per week may be considered as a crude measure. Future studies would benefit from more precise quantification of AB usage to accurately categorize participants as regular or non-regular users.

The AB protocol showed promise as a valid assessment of $\dot{V}O_{2max}$ when compared to the treadmill protocol in individuals who regularly use the AB. These results suggest that the AB protocol can be effectively utilized as a measure of $\dot{V}O_{2max}$ in specific populations, such as athletes or individuals who frequently engage in AB exercises as part of their training regimen. Furthermore, the AB protocol showed similar values of the RER, final RPE, and percent $\dot{V}O_{2max}$ at the ventilatory threshold, indicating its ability to provide valuable information on metabolic responses and exertion levels during exercise. However, practitioners should exercise caution when using the AB protocol interchangeably with treadmill-based $\dot{V}O_{2max}$ testing, as significant differences were observed between the two protocols.

REFERENCES

1. Astorino T, Rietschel J, Tam P, Taylor K, Johnson S, Freedman T, et al. Reinvestigation of optimal duration of VO2 max testing. JEPonline 7(6): 1-8, 2004.

2. Balagué N, González J, Javierre C, Hristovski R, Aragonés D, Álamo J, et al. Cardiorespiratory coordination after training and detraining. A principal component analysis approach. Front Physiol 7: 35, 2016.

3. Balagué N, Hristovski R, Almarcha M, Garcia-Retortillo S, Ivanov PC. Network physiology of exercise: Beyond molecular and omics perspectives. Sports Med Open 8(1): 119, 2022.

4. Beltz NM, Gibson AL, Janot JM, Kravitz L, Mermier CM, Dalleck LC. Graded exercise testing protocols for the determination of VO_2max: Historical perspectives, progress, and future considerations. J Sports Med 2016: e3968393, 2016.

5. Bland JM, Altman DG. Applying the right statistics: Analyses of measurement studies. Ultrasound Obstet Gynecol 22(1): 85–93, 2003.

6. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Int J Nurs Stud 47(8): 931–6, 2010.

7. Bouckaert J, Vrijens J, Pannier JL. Effect of specific test procedures on plasma lactate concentration and peak oxygen uptake in endurance athletes. J Sports Med Phys Fitness 30(1): 13–8, 1990.

8. Burke EJ, Franks BD. Changes in VO2 max resulting from bicycle training at different intensities holding total mechanical work constant. Res Q Am Alliance Health Phys Educ Recreat 46(1): 31–7, 1975.

9. Canário-Lemos R, Machado-Reis V, Garrido ND, Rafael-Moreira T, Peixoto R, Nobre-Pinheiro B, et al. Validity and reliability of maximum oxygen uptake on an air bike arm- and leg-ergometer. Motricidade 18(3): 404–9, 2022.

10. Catapano JS, Chapman AJ, Farber SH, Horner LP, Morgan C, Brigeman S, et al. Treadmill associated head injuries on the rise: An 18 year review of U.S. emergency room visits. Brain Inj 32(6): 800–3, 2018.

11. Cavanagh PR, Williams KR. The effect of stride length variation on oxygen uptake during distance running. Med Sci Sports Exerc 14(1): 30–5, 1982.

12. Cervantes M. Sprint interval training on stationary air bike shows benefits to cardiorespiratory adaptations while being time efficient. Med Sci Sports Exerc 53(8S): 6, 2021.

13. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L. Erlbaum Associates; 1988.

14. Davies CT, Sargeant AJ. Circadian variation in physiological responses to exercise on a stationary bicycle ergometer. Br J Ind Med 32(2): 110–4, 1975.

15. Davis JA, Dorado S, Keays KA, Reigel KA, Valencia KS, Pham PH. Reliability and validity of the lung volume measurement made by the BOD POD body composition system. Clin Physiol Funct Imaging 27(1): 42–6, 2007.

16. Dolezal BA, Barr D, Boland DM, Smith DL, Cooper CB. Validation of the firefighter WFI treadmill protocol for predicting VO2 max. Occup Med 65(2): 143–6, 2015.

17. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39(2): 175–91, 2007.

18. Fields DA, Hunter GR, Goran MI. Validation of the BOD POD with hydrostatic weighing: Influence of body clothing. Int J Obes Relat Metab Disord 24(2): 200–5, 2000.

19. Garcia-Retortillo S, Gacto M, O'Leary TJ, Noon M, Hristovski R, Balagué N, et al. Cardiorespiratory coordination reveals training-specific physiological adaptations. Eur J Appl Physiol 119(8): 1701–9, 2019.

20. Ghosh D, Vogt A. Outliers: An evaluation of methodologies. JSM Conference Proceedings 304068, 2012.

21. Helm M, Carrier B, Davis D, Cruz K, Barrios B, Navalta J. Validation of the Garmin Fenix 6S maximal oxygen consumption (VO2max) estimate. Int J Exerc Sci Conf Proc 14(1): 29, 2021.

22. Ivanov PC, Liu KKL, Bartsch RP. Focus on the emerging new fields of network physiology and network medicine. New J Phys 18: 100201, 2016.

23. Kieu NTV, Jung S-J, Shin S-W, Jung H-W, Jung E-S, Won YH, et al. The validity of the YMCA 3-minute step test for estimating maximal oxygen uptake in healthy Korean and Vietnamese adults. J Lifestyle Med 10(1): 21–9, 2020.

24. Kokinda M, Ružbarský P, Kozák T, Fečík M, Kičura D. An innovative approach to functional spiroergometric examination among power athletes. J Phys Educ Sport 22(11): 2833–9, 2022.

25. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15(2): 155–63, 2016.

26. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 4: 863, 2013.

27. Lakens D. Sample size justification. Collabra Psychol 8(1): 33267, 2022.

28. Martinez A, Snyder AJ, Smith GA. Home exercise equipment-related injuries among children in the United States. Clin Pediatr (Phila) 50(6): 553–8, 2011.

29. McArdle WD, Katch FI, Katch VL. Exercise physiology: Nutrition, energy, and human performance. Philadelphia, PA: Lippincott Williams & Wilkins; 2010.

30. Millet GP, Vleck VE, Bentley DJ. Physiological differences between cycling and running: Lessons from triathletes. Sports Med 39(3): 179–206, 2009.

31. Navalta JW, Stone WJ, Lyons TS. Ethical issues relating to scientific discovery in exercise science. Int J Exerc Sci 12(1): 1-8, 2019.

32. Papadakis Z, Etchebaster M, Garcia-Retortillo S. Cardiorespiratory coordination in collegiate rowing: A network approach to cardiorespiratory exercise testing. Int J Environ Res Public Health 19(20): 13250, 2022.

33. Roecker K, Striegel H, Dickhuth H-H. Heart-rate recommendations: Transfer between running and cycling exercise? Int J Sports Med 24(3): 173–8, 2003.

34. Schulz S, Adochiei F-C, Edu I-R, Schroeder R, Costin H, Bär K-J, et al. Cardiovascular and cardiorespiratory coupling analyses: A review. Philos Transact Royal Society A Math Phys Eng Sci 371(1997): 20120191, 2013.

35. Sporer BC, Wenger HA. Effects of aerobic exercise on strength performance following various periods of recovery. J Strength Cond Res 17(4): 638-44, 2003.

36. Sullivan GM, Feinn R. Using effect size—or why the P Value is not enough. J Grad Med Educ 4(3): 279–82, 2012.

37. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg Procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 27(1): 77–83, 2002.

38. Zheng J, Pan T, Jiang Y, Shen Y. Effects of short- and long-term detraining on maximal oxygen uptake in athletes: A systematic review and meta-analysis. BioMed Res Int 2022: 2130993, 2022.