Sports Medicine and Health Science 5 (2023) 101–105

Contents lists available at ScienceDirect

Sports Medicine and Health Science

journal homepage: www.keaipublishing.com/smhs

Original Article

The efficacy of a verification stage for determining $\dot{V}O_{2max}$ and the impact of sampling intervals

Emily J. Kontos, Nicholas D. Luden, Stephanie Kurti, Christopher J. Womack *

Human Performance Laboratory, Department of Kinesiology, James Madison University, Harrisonburg, VA, 22801, USA

ARTICLE INFO

Keywords:

VO_{2max} testing
Verification stage
VO_{2max} criteria specificity
VO_{2max} criteria sensitivity

ABSTRACT

It is unknown whether oxygen uptake (VO_2) sampling intervals influence the efficacy of a verification stage following a graded exercise test (GXT). Fifteen females and 14 males (18–25 years) completed a maximal treadmill GXT. After a 5 min recovery, the verification stage began at the speed and grade corresponding with the penultimate stage from the GXT. Maximal oxygen consumption (VO_{2max}) from the incremental GXT (iVO_{2max}) and VO_{2max} from the verification stage ($verVO_{2max}$) were determined using 10 seconds (s), 30 s, and 60 s from breath × breath averages. There was no main effect for VO_{2max} measure (iVO_{2max}vs. $verVO_{2max}$) 10 s ([47.9 ± 8.31] ml•kg^{-1}•min^{-1} vs [48.85 ± 7.97] ml•kg^{-1}•min^{-1}), 30 s ([46.94 ± 8.62] ml•kg^{-1}•min^{-1} vs [47.28 ± 7.97] ml•kg^{-1}•min^{-1}), and 60 s ([46.17 ± 8.62] ml•kg^{-1}•min^{-1} vs [46.00 ± 8.00] ml•kg^{-1}•min^{-1}]. There was a stage × sampling interval interaction as the difference between ($verVO_{2max}$−iVO_{2max}) was greater for 10-s than 60-s sampling intervals. The $verVO_{2max}$ was > 4% higher than iVO_{2max}in 31%, 31%, and 17% of the tests for the 10-s, 30-s, and 60-s sampling intervals respectively. Sensitivity for the plateau was < 30% for 10-s, 30-s, and 60-s sampling intervals. Specificity ranged from 44% to 60% for all sampling intervals. Sensitivity for heart rate + respiratory exchange ratio was > 90% for all sampling intervals; while specificity was < 25%. Findings from the present study suggest that the efficacy of verification stages for eliciting a higher VO_{2max} may be influenced by the sampling interval utilized.

Introduction

Maximal oxygen consumption (VO_{2max}) is one of the most measured variables in the field of exercise science. Historically, the achievement of VO_{2max} during a graded exercise test (GXT) has been confirmed by the presence of a "plateau" or a failure for oxygen uptake (VO_2) to increase despite the increasing workload.[1] However, a plateau is not always evident and the frequency with which a plateau can be observed varies in prior research,[2–4] necessitating the use of other "secondary" criteria such as reaching a predetermined heart rate (HR) or respiratory exchange ratio (RER) value to increase the likelihood that true VO_{2max} is reached.

Several studies have found that when a subject completes a VO_{2max} test, a subsequent verification stage typically results in a lower, same, or higher number than the determined VO_{2max} value.[4–6] Mier and colleagues[6] investigated the effectiveness of a supramaximal verification stage in college athletes who did not achieve a VO_{2max} plateau during a GXT. VO_{2max} values from the verification stage were not significantly different than the VO_{2max} values from the continuous GXT. Similarly, Foster et al.[4] observed a verification stage resulted in similar VO_{2max} values as those recorded during a GXT for both athletes and non-athletes. Bhammar et al.[3] actually observed a verification stage resulted in higher VO_{2max} values than the GXT. Thus, the verification stage represents an extra opportunity to give such an effort and thus represents a way to increase the likelihood that VO_{2max} is reached; although it should be realized that it is also possible that a verification stage could result in a value below that achieved during the GXT.

Another factor to consider when conducting VO_{2max} tests is sampling intervals. Higher incidences of a plateau during VO_{2max} testing have been observed for 11-s and 15-s sampling intervals when compared to 30-s sampling averages for breath by breath VO_2 measurements.[7] Furthermore, 15-s and 30-s intervals have been shown to result in higher VO_{2max} values than 60-s intervals.[6,8] One of the issues that could affect the efficacy of a verification stage is that subjects could fatigue quickly due to prior activity and not enough time would be provided for VO_2 to reach VO_{2max}. Thus, it is possible the use of shorter sampling intervals would limit this concern, as the subject would only need to reach VO_{2max} for a shorter window of time. As a result, a greater portion of people could potentially exceed the highest VO_2 achieved during the GXT. Therefore, the efficacy of a verification stage may depend upon the duration of the

* Corresponding author. Department of Kinesiology, MSC 2302, James Madison University Harrisonburg, VA, 22801, USA.
E-mail address: womackcx@jmu.edu (C.J. Womack).

https://doi.org/10.1016/j.smhs.2023.04.001
Received 29 July 2022; Received in revised form 8 April 2023; Accepted 23 April 2023
Available online 3 May 2023

Abbreviation

$\dot{V}O_{2max}$	Maximal oxygen uptake
$\dot{V}O_2$	Oxygen uptake
$\dot{V}O_{2peak}$	Peak oxygen uptake-
$i\dot{V}O_{2max}$	Maximal oxygen uptake from the graded exercise test
$ver\dot{V}O_{2max}$	Maximal oxygen uptake from the verification stage
Delta $\dot{V}O_{2max}$	$ver\dot{V}O_{2max}$-$i\dot{V}O_{2max}$ MMC Metabolic measurement cart
HR	Heart rate
RER	Respiratory Exchange Ratio
s	seconds
min	minutes

sampling interval.

Despite the widespread use of VO_{2max} testing, there are limited data assessing the sensitivity and specificity of traditional secondary criteria. Bhammar et al.[3] found poor sensitivity and specificity for traditional criteria used to verify the achievement of VO_{2max}. However, this study was performed with a limited number of test subjects and these subjects were children. Thus, their findings are limited in their generalizability. Furthermore, it is not known if the use of different sampling intervals will impact the effectiveness of traditional primary and secondary criteria for determining VO_{2max} from a continuous GXT. It is important to study the sensitivity and specificity of VO_{2max} secondary criteria as Poole and Jones[9] have suggested that the use of secondary criteria may lead to an increase in both false negatives and false positives. As such, the use of sensitivity and specificity to assess the suitability of primary and secondary criteria for VO_{2max} testing would improve the objective evaluation of these criteria as they measure the degree to which false negatives, as well as false positives, occur.

The purposes of this present study were to: 1) determine the influence of VO_2 sampling intervals on the efficacy of a verification stage; and 2) to determine the influence of VO_2 sampling intervals on sensitivity and specificity of primary and secondary VO_{2max} test criteria.

Methods

Subjects

This study evaluated 29 test subjects (14 men and 15 women) who were free of known cardiovascular, metabolic, or renal diseases. Additionally, test subjects had no known injuries or other health concerns that would preclude them from exercising or limit their ability to perform a maximal GXT. Written informed consent was obtained from all participants prior to participation. The study and consent form were approved by the Institutional Review Board at James Madison University.

Treadmill test

The protocol employed has been shown to result in fatigue in healthy college-aged students in approximately 12 minutes (min).[10] All subjects were monitored for VO_2 and RER with a VMax metabolic measurement cart (MMC) (CareFusion; San Diego, CA) that was calibrated prior to each test. This MMC utilizes a mass flow sensor to detect expired air volume. The oxygen analyzer is an electrochemical fuel cell and the carbon dioxide analyzer is a non-disperse infrared thermopile. MMC data were collected using breath x breath measurements which were then converted to the respected sampling interval averages (10 s, 30 s, 60 s). A Polar heart rate monitor (Lake Success, NY) was utilized to measure heart rate throughout the test. The GXT began at an initial stage at 3.0 mph and 0% grade. The treadmill speed was increased by 0.5 mph each minute until a speed of 6.0 mph was achieved. After this, the incline of the treadmill was increased by 3% every minute until volitional exhaustion, defined as the point at which the participant felt they could no longer continue. Subjects then walked at a comfortable speed for 5 min. After this active rest period, the verification stage was initiated by increasing the speed and grade to values corresponding to the stage preceding the test subject's prior maximal effort. Unpublished data from our laboratory suggest that this 5 min rest duration is at least as effective as 15 min rest for eliciting the highest possible VO_{2max} values. The test then proceeded as described previously until the participant indicated they could no longer continue. VO_{2max} from the GXT (iVO_{2max}) was defined as the highest VO_2 achieved during the GXT for the respective sampling interval (10 s, 30 s, 60 s). VO_{2max} from the verification stage ($verVO_{2max}$) was defined as the highest VO_2 achieved during the verification stage for the respective sampling interval. Sample data from one of the subjects is

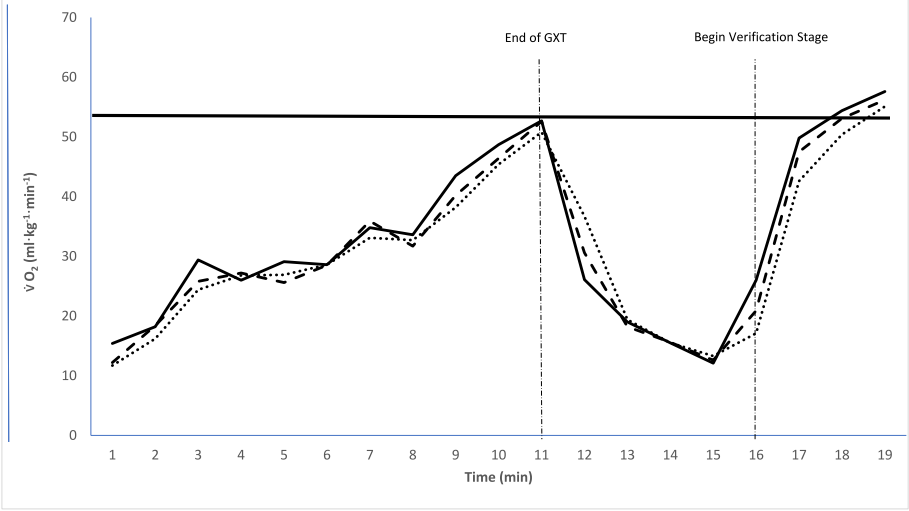

Fig. 1. Oxygen uptake (VO_2) response to the graded exercise test (GXT) and verification stage for one of the subjects. Averaging of VO_2 data is expressed with solid line (10 s), dashed line (30 s) and dotted line (60 s). The solid horizontal line represents VO_{2max} from the GXT. Note: s = seconds.

displayed in Fig. 1.

Determination of Primary and Secondary Criteria for Confirming VO_{2max}

The primary criteria for achievement of VO_{2max} during the GXT was a plateau in VO_2. In order to be considered a plateau, VO_2 had to increase less than the confidence interval of the expected increase in VO_2 for the final stage of the GXT. Expected VO_2, as well as the confidence interval, was determined by plotting VO_2 against treadmill grade from minute seven of the GXT (6.0 mph, 0% grade) to the remaining stages of the test. To eliminate the potential effects of a plateau, the final two stages for each subject were excluded. The slope of this relationship was averaged and the confidence interval was determined by multiplying 1.645 by the standard deviation. Thus, the plateau was defined as an increase in VO_2 less than 1.83 $ml \bullet kg^{-1} \bullet min^{-1}$. The procedures used for this determination have been previously outlined.[11] Secondary criteria were met if at least 90% of age-predicted maximal heart rate ($208 - 0.8 \times age$)[12] and an RER of at least 1.10 were achieved.

Statistical analyses

A two-factor repeated measure of analysis of variance was performed with within-subjects factors of the stage (iVO_{2max}, $verVO_{2max}$) and sampling interval (10 s, 30 s, and 60 s). Post-hoc testings of significant main effects were performed using estimated marginal means with least significant difference (LSD) comparisons. For the interaction effect, estimated marginal means with LSD comparisons were performed on the difference between iVO_{2max} and $verVO_{2max}$ for each sampling interval (delta VO_{2max}). To determine if either duration of the GXT or the verification stage influenced the efficacy of the verification stage, correlations between the respective durations and delta VO_{2max} for all three sampling intervals were established. For all three sampling intervals, the sensitivity, and specificity of the primary and secondary criteria were calculated. $verVO_{2max}$ was considered to be higher if it exceeded iVO_{2max} by more than 4%. This value was used because it is the established coefficient of variation for VO_{2max} for the protocol used in this study for our lab. Sensitivity was calculated by taking the number of True Positives (Criteria achieved and iVO_{2max} is within 4% of $verVO_{2max}$) divided by True Positives plus False Negatives (Criteria not achieved and iVO_{2max} within 4% of $verVO_{2max}$). Specificity was determined by the number of True Negatives (Criteria not achieved and $verVO_{2max}$ more than 4% higher than iVO_{2max}) divided by the number of True Negatives plus False Positives (Criteria achieved but $verVO_{2max}$ greater than 4% higher than iVO_{2max}).

Results

The average age of the participants was (21.3 ± 1.2) years (yr). Height and weight were (men = [179.1 ± 8.5] cm, [80.9 ± 10.4] kg; women = [160.3 ± 5.4] cm, [60.3 ± 5.4] kg) before the maximal treadmill test. The average time to fatigue for the GXT was (10.7 ± 1.9) min (range 5-14 min). The average time to fatigue for the verification stage was (2.0 ± 0.6) min (range 1-3 min). There was no significant correlation between GXT test duration and delta VO_{2max} for 10 s ($R = -0.28$, $p = 0.28$), 30 s ($R = -0.21$, $p = 0.21$), and 60 s ($R = -0.27$, $p = 0.16$) sampling intervals. However, there was a significant correlation between verification stage test duration and delta VO_{2max} for 10 s ($R = 0.52$, $p = 0.004$), 30 s ($R = 0.52$, $p = 0.005$), and 60 s ($R = 0.53$, $p = 0.005$) sampling intervals.

Table 1 shows the effect of the sampling interval on average VO_{2max} values. Table 2 displays the ANOVA table from SPSS. As the sampling interval increased from 10 s to 60 s, average VO_{2max} values significantly ($p < 0.05$) decreased (10 s > 30 s > 60 s). There was no significant main effect for the stage. However, there was a significant stage × sampling interval interaction (*partial* $\eta^2 = 0.25$) as the difference between $verVO_{2max}$ and iVO_{2max} was greater for the 10-s than for the 60-s sampling interval.

Table 3 shows the sensitivity and specificity of the primary and secondary criteria for iVO_{2max}. Sensitivity for the incidence of a plateau for VO_{2max} was ≤ 30% for 10-s, 30-s, and 60-s sampling intervals. Specificity ranged from 44% to 60% for all three sampling intervals. Sensitivity for HR + RER was > 90% for 10-s, 30-s, and 60-s sampling intervals. The highest sensitivity was observed with the 30-s sampling interval. For all three sampling intervals, specificity was < 25%.

Discussion

These data suggest there may be a greater need for verification stages when shorter sampling intervals are implemented. The duration of our verification stage was typically 1–2 min. Because VO_2 is increasing at the onset of a verification stage, it is possible the 60-s sampling interval includes several data points representing lower VO_2 values at the beginning of the onset of exercise and thus fails to deliver an average VO_2 that truly reflects VO_{2max}. However, sampling variability has a greater impact with

Table 1
Average (± *SD*) maximal oxygen uptake from the graded exercise test (iVO_{2max}) and from the verification stage ($verVO_{2max}$) along with the percentage of tests in which $verVO_{2max} > iVO_{2max}$. Note: s = seconds, m = minutes, kg = kilograms.

Sampling Interval*	iVO_{2max} ($ml \bullet kg^{-1} \bullet min^{-1}$)	$verVO_{2max}$ ($ml \bullet kg^{-1} \bullet min^{-1}$)	Delta VO_{2max} ($ml \bullet kg^{-1} \bullet min^{-1}$)	$verVO_{2max} > iVO_{2max}$ (%)
10 s	47.9 ± 8.31	48.85 ± 7.97	0.96 ± 2.88†	31%
30 s	46.94 ± 8.62	47.28 ± 7.97	0.34 ± 2.73	31%
60 s	46.17 ± 8.62	46.00 ± 8.00	−0.17 ± 3.81	17%

*-Main effect for sampling interval (10 s > 30 s > 60 s, $p < 0.05$), †-Significant stage × sample interval interaction (delta VO_{2max} for 10 s > 60 s, $p < 0.05$).

Table 2
ANOVA results from the SPSS analysis. Note main effect for sampling time and the interaction effect between the sampling time and the stages.

Effect	Type III Sum of Squares	*df*	Mean Square	*F*	Significance
Sample Time	152.182	2	76.091	43.67	< 0.001
Error (Sample Time)	97.575	56	1.742		
Verification	6.162	1	6.162	0.463	0.502
Error (Verification)	373.322	28	13.297		
Sample Time × Verification	9.194	2	4.597	4.933	0.011
Error (Sample Time × Verification)	52.192	56	0.932		

df = degree of freedom.

Table 3
Sensitivity and specificity of maximal oxygen uptake (VO_{2max}) primary (plateau) and secondary (heart rate [HR]+ respiratory exchange ratio [RER]) criteria. Note: s = seconds, HR = heart rate, RER = respiratory exchange ratio.

	10 s	30 s	60 s
Plateau			
Sensitivity	30.0	20.0	20.8
Specificity	44.4	55.6	60
HR + RER			
Sensitivity	95	100	91.3
Specificity	25.0	11.1	20.0

shorter sampling intervals, as an aberrant data point would have a larger influence on the shorter time average. It should also be realized that the delta VO_{2max} was small with all sampling intervals; thus, the observed interaction effect may be due to a Type I error. We also observed shorter sampling intervals resulted in higher VO_{2max} values when compared to longer sampling intervals. In support of this, Astorino et al.[8] found that shorter sampling intervals resulted in increased VO_{2max} and increased incidence of a plateau. The latter also appears to be true of our data as plateau incidence was 11, 8, and 7 tests out of the 20 for 10-s, 30-s, and 60-s sampling intervals respectively. Additionally, we also observed significant correlations between verification stage duration and delta VO_{2max}, suggesting that verification stages are more likely to yield higher VO_{2max} values when the participant is able to achieve a longer duration during the verification stage. This makes intuitive sense in that a longer duration verification stage would be associated with a higher maximal workload. Furthermore, a shorter-duration verification stage would likely reflect fatigue occurring before VO_{2max} was able to rise to maximal levels.

Similar to the present study, several studies have found verification stages yield VO_{2max} values were comparable to those achieved during a continuous GXT. Foster et al.[4] observed similar values for VO_{2max} in a verification stage and during GXT's in which a plateau was evident. This was true for both treadmill and cycling tests. Midgley et al.[5] found no statistically significant differences between VO_{2max} values during a running GXT and a verification stage. Furthermore, Rossiter et al.[13] observed no significant differences between the highest VO_2 achieved during an incremental ramp cycling test (VO_{2peak}) and a verification stage performed afterwards. This was true when the verification stage was performed above and below peak workload.[13] Therefore the present study confirms that a verification stage will not result in large changes in VO_{2max} compared to what is obtained during a graded exercise test. However, it should also be realized that verification stages resulted in a higher VO_{2max} in 17%-31% of our tests (depending on the sampling interval). Thus, it may be useful to include a verification stage to account for this fraction of tests in which VO_{2max} is not achieved during the GXT.

Although calculations of sensitivity and specificity are very illuminating concerning criteria traditionally used for achieving VO_{2max}, few studies have reported these parameters. In the present study, we observed poor sensitivity and specificity for the use of a VO_2 plateau (primary criterion) to confirm VO_{2max}. Both sensitivity and specificity appeared to be fairly low for the use of a plateau (Table 2). Because sensitivity is inversely related to the number of false negatives, it is reasonable to expect there would not be a high degree of sensitivity for a plateau as it occurs only 15% of the time in non-athletes and ~50% of the time in athletes.[4] Howley et al.[2] proposed these numbers may be even lower as children, sedentary, and elderly populations have a harder time achieving a plateau. Furthermore, Day et al. observed that a plateau is often not evident when VO_{2max} has been achieved.[14] In short, the low incidence of a plateau observed in the present study (24%–38%) makes it likely that several GXT's will result in a false-negative or a test where VO_{2max} was achieved but a plateau is not evident. However, the finding of low specificity was surprising as it suggests a high number of test subjects that exhibit a plateau do not actually achieve a VO_{2max} during the incremental GXT. The fact that specificity appeared to increase with increasing sample duration may be due to the fact that a plateau is more valid at higher sampling rates, thus resulting in fewer false-positives. Few studies have evaluated the sensitivity and specificity of a plateau. Bhammar and colleagues[3] found similar values for sensitivity and specificity for plateau as the current study in non-obese and obese children (22% and 44%, respectively).

Our data showed a high sensitivity and low specificity for the use of common secondary criteria utilizing HR + RER, which did not appear to substantially change with the sampling interval. This suggests reaching these criteria is a common occurrence and it is rare for someone to fail to achieve these criteria and achieve VO_{2max} based on the inverse relationship between sensitivity and false-negative tests. However, the low specificity, which is inversely related to the number of false-positive tests, suggests these secondary criteria could easily result in a VO_{2max} score below what should be assigned. In support of this, Poole et al.[15] found that during a GXT, VO_2 at an RER of 1.10 averaged over 1 L/min below VO_{2max}. Furthermore, in that same study, five subjects (out of eight) that achieved the criteria of HR ± 10 bpm of age-predicted max achieved that HR at 76% of $\dot{V}VO_{2max}$.[15] Therefore, the secondary criteria as employed in the current study appear to be ineffective for verifying VO_{2max}.

It is possible that the use of a plateau (primary criteria) or the use of HR + RER values (secondary criteria) are good for identifying VO_{2max}, but we failed to identify the right cut-off values for these criteria. However, the specific primary and secondary criteria were chosen because they are commonly used in research. Midgley et al.[16] identified the most common criteria for HR and RER were the same as those used in the present study. Furthermore, the plateau criteria used were specifically tailored to our treadmill protocol and used previously.[12] Thus, our findings show common criteria for confirmation of VO_{2max} result in poor sensitivity and/or specificity. Regardless, the findings from this study should include the caveat that the stated sensitivity and specificity values from secondary criteria are for the values and variables used in the present study. Furthermore, the sensitivity and/or specificity of the plateau may substantially change if plateau incidence is increased due to changes in protocol and/or sample demographics. The characteristics of an optimal verification stage are currently unknown. However, the protocol employed in the current study with respect to both the recovery time after the GXT and the initial intensity of the verification stage were within suggested ranges.[17] Improvements in the verification stage protocol would likely lead to an increased verVO_{2max} and would result in an even greater proportion of tests that required a verification stage and even worse sensitivity/specificity for traditional criteria. Results of the present study confirm that verification stages yield similar VO_{2max} values as the GXT regardless of the sampling interval used, although a larger, albeit physiologically small, difference between verVO_{2max} and iVO_{2max} is evident with shorter sampling intervals. Furthermore, secondary criteria commonly used to verify VO_{2max} may not be ideal for confirming attainment of VO_{2max} during a GXT.

Submission statement

All authors have read and agree with the content of this manuscript. This manuscript will not be submitted elsewhere for review and

publication while it is being reviewed by *Sports Medicine and Health Science*.

Ethical approval statement

Written informed consent was obtained from all participants prior to participation. The study and consent form were approved by the Institutional Review Board at James Madison University.

Authors' contributions

Emily J. Kontos-study design, data collection, manuscript preparation, manuscript review/revision. Nicholas D. Luden-study design, manuscript preparation, manuscript review/revision. Stephanie Kurti-study design, manuscript preparation, manuscript review/revision. Christopher J. Womack-study design, data collection, data analysis, manuscript preparation, manuscript review/revision.

Conflict of interest

There are no conflicts of interest from the authors of this study.

References

1. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J Appl Physiol.* 1955;8(1):73–80. https://doi.org/10.1152/jappl.1955.8.1.73.
2. Howley ET, Bassett jr DR, Welch HG. Criteria for maximal oxygen uptake: review and commentary. *Med Sci Sports Exerc.* 1995;27(9):1292–1301.
3. Bhamar DM, Stickford JL, Bernhardt V, Badd TG. Verification of maximal oxygen uptake in obese and nonobese children. *Med Sci Sports Exerc.* 2017;49(4):702–710. https://doi.org/10.1249/MSS.0000000000001170.
4. Foster C, Kuffel E, Bradley N, et al. VO2max during successive maximal efforts. *Eur J Appl Physiol.* 2007;102(1):67–72. https://doi.org/10.1007/s00421-007-0565-x.
5. Midgley AW, McNaughton LR, Carroll S. Verification phase as a useful tool in the determination of the maximal oxygen uptake of distance runners. *Appl Physiol Nutr Metabol.* 2006;31(5):541–548. https://doi.org/10.1139/h06-023.
6. Mier CM, Alexander RP, Mageean AL. Achievement of VO2max criteria during a continuous graded exercise test and a verification stage performed by college athletes. *J Strength Condit Res.* 2012;26(10):2648–2654. https://doi.org/10.1519/JSC.0b013e31823f8de9.
7. Astorino TA, Robergs RA, Ghiasvand F, Marks D, Burns S. Incidence of VO2 plateau at VO2 max during exercise testing to volitional fatigue. *J Exerc Physiol.* 2000;3(4):1–12.
8. Astorino TA. Alterations in VO2 max and the VO2 plateau with manipulation of sampling interval. *Clin Physiol Funct Imag.* 2009;29(1):60–67. https://doi.org/10.1111/j.1475-097X.2008.00835.x.
9. Poole DC, Jones AM. Measurement of the maximum oxygen uptake VO_{2max} : Vo_{2peak} is no longer acceptable. *J. Appl. Physiol.* 2017;122(4):997–1002. https://doi.org/10.1152/japplphysiol.01063.2016.
10. Paton CM, Nagelkirk PR, Coughlin AM, et al. Changes in von Willebrand factor and fibrinolysis following a post-exercise cool-down. *Eur J Appl Physiol.* 2004;92(3):328–333. https://doi.org/10.1007/s00421-004-1098-1.
11. Dwyer DB. A standard method for the determination of maximal aerobic power from breath-by-breath VO2 data obtained during a continuous ramp test on a bicycle ergometer. *J Exerc Physiol Online.* 2004;7(5):1–9.
12. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. *J Am Coll Cardiol.* 2001;37(1):153–156. https://doi.org/10.1016/S0735-1097(00)01054-8.
13. Rossiter HB, Kowalchuk JM, Whipp BJ. A test to establish maximum O2 uptake despite no plateau in the O2 uptake response to ramp incremental exercise. *J Appl Physiol (1985).* 2006;100(3):764–770. https://doi.org/10.1152/japplphysiol.00932.2005.
14. Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ. The maximally attainable VO_2 during exercise in humans: the peak vs. maximum issue. *J Appl Physiol (1985).* 2003;95(5):1901–1907. https://doi.org/10.1152/japplphysiol.00024.2003.
15. Poole DC, Wilkerson DP, Jones AM. Validity of criteria for establishing maximal O2 uptake during ramp exercise tests. *Eur J Appl Physiol.* 2008;102(4):403–410. https://doi.org/10.1007/s00421-007-0596-3.
16. Midgley AW, McNaughton LR, Polman R, Marchant D. Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. *Sports Med.* 2007;37(12):1019–1028. https://doi.org/10.2165/00007256-200737120-00002.
17. Schaun GZ. The maximal oxygen uptake verification phase: a light at the end of the tunnel? *Sport Med Open.* 2017;3(1):44. https://doi.org/10.1186/s40798-017-0112-1.