Which Cutoffs for Secondary $\dot{V}O_{2max}$ Criteria Are Robust to Diurnal Variations?

RAPHAEL KNAIER[1], MAX NIEMEYER[2], JONATHAN WAGNER[1], DENIS INFANGER[1], TIMO HINRICHS[1], CHRISTOPHER KLENK[1], SABRINA FRUTIG[1], CHRISTIAN CAJOCHEN[3,4], and ARNO SCHMIDT-TRUCKSÄSS[1]

[1]Department of Sport, Exercise and Health, Faculty of Medicine, University of Basel, Basel, SWITZERLAND; [2]Department Medicine, Training and Health, Philipps-University Marburg, Marburg, GERMANY; [3]Centre for Chronobiology, Psychiatric Hospital of the University of Basel, Basel, SWITZERLAND; and [4]Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, Basel, SWITZERLAND

ABSTRACT

KNAIER, R., M. NIEMEYER, J. WAGNER, D. INFANGER, T. HINRICHS, C. KLENK, S. FRUTIG, C. CAJOCHEN, and A. SCHMIDT-TRUCKSÄSS. Which Cutoffs for Secondary $\dot{V}O_{2max}$ Criteria Are Robust to Diurnal Variations? *Med. Sci. Sports Exerc.*, Vol. 51, No. 5, pp. 1006–1013, 2019. **Purpose**: The aim was to determine the minimum maximum oxygen uptake ($\dot{V}O_{2max}$) criteria cut-offs in highly trained athletes (i.e., maximum RER [RER_{max}], maximum HR [HR_{max}], maximum RPE [RPE_{max}], and maximum blood lactate concentration [BL_{max}]) necessary to determine maximum oxygen uptake ($\dot{V}O_{2max}$) during cardiopulmonary exercise tests (CPET), by balancing type I and type II errors. A further aim was to investigate if the defined cutoffs would be robust to diurnal and to day-to-day variations. **Methods**: Data from two CPET studies involving young athletes were analyzed. In the first study, 70 male participants performed one CPET until exhaustion to define cutoffs. In the second study, eight males and five females performed one CPET on seven consecutive days at six different times of day (i.e., diurnal variation). The time of the CPET was identical on the sixth and seventh days (i.e., day-to-day variation). To ensure comparability both studies were carried out under the same conditions. **Results**: Participants' mean $\dot{V}O_{2max}$ was 63.0 ± 5.3 $mL\cdot kg^{-1}\cdot min^{-1}$. $RER_{max} \geq 1.10$ was reached by 100%, $HR_{max} \geq 95\%$ of age-predicted HR_{max} by 99%, $RPE_{max} \geq 19$ by 100%, and $BL_{max} \geq 8$ $mmol\cdot L^{-1}$ by 100% of participants, respectively. Regarding the intraday variations, latter cutoffs were not reached in two cases for RER_{max} and in one case for HR_{max} and BL_{max}. Intraclass correlations for the day-to-day variability were r = 0.823 for RER_{max}, r = 0.828 for HR_{max}, and r = 0.380 for BL_{max}, respectively. **Conclusions**: The proposed high cut-off values for secondary criteria provide some assurance that $\dot{V}O_{2max}$ may have been achieved in athletes without increasing type II errors. However, type I errors may still occur indicating that further methods such as $\dot{V}O_2$-plateau or $\dot{V}O_2$-validation may be required. **Key Words:** EXERCISE TEST, OXYGEN CONSUMPTION, REPRODUCIBILITY OF RESULTS

The maximum volume of oxygen uptake per minute ($\dot{V}O_{2max}$, $L\cdot min^{-1}$) is considered the gold standard to measure cardiorespiratory fitness. Therefore, $\dot{V}O_{2max}$ is used in intervention studies as a primary outcome to assess changes in physical fitness, in competitive sports to evaluate the effectiveness of training programs or in clinical settings for risk estimation of all-cause mortality (1). To maximize signal-to-noise ratio, it is crucial to measure $\dot{V}O_{2max}$ with sufficient rigor, especially when baseline and follow-up values are being compared. If the "true" $\dot{V}O_{2max}$ were not determined in both, baseline and follow-up tests, the intervention would account for changes in $\dot{V}O_{2max}$, which may actually result from an inaccurate measurement. Participants' maximum possible effort is necessary to ensure that peak oxygen uptake ($\dot{V}O_{2peak}$) and true $\dot{V}O_{2max}$ are achieved (2). The primary criterion to distinguish $\dot{V}O_{2max}$ from $\dot{V}O_{2peak}$ is the appearance of a $\dot{V}O_2$ plateau. However, a $\dot{V}O_2$ plateau may not be found in all participants, even if performing the exercise test at maximal effort (3–5). Furthermore, the criteria to define a $\dot{V}O_2$ plateau are controversial (6). Verification phases (i.e., additional supramaximal exercise tests performed immediately after a short regeneration phase subsequent to the cardiopulmonary exercise test [CEPT]) have been discussed as an additional method to determine true $\dot{V}O_{2max}$ (2,7). However, verification tests are rather time-consuming, and their day-to-day reliability has not been investigated. Additionally, the procedure to perform these verification tests is not standardized (8) and many previous studies that promote the use of verification tests failed to report the required data to support the use of verification phases. In detail, it has been criticized that $\dot{V}O_2$ from the initial phase of exercise testing and $\dot{V}O_2$ from the verification phase were only compared on a group level (9) and it was not reported in which proportion of participants $\dot{V}O_{2max}$ could actually be verified (10–14). In those participants that show no $\dot{V}O_2$ plateau and without a verified $\dot{V}O_{2max}$, secondary criteria are used to distinguish

---

Address for correspondence: Raphael Knaier, Ph.D., Departement für Sport, Bewegung und Gesundheit, Universität Basel, Birsstrasse 320B, 4052 Basel, Switzerland; E-mail: raphael.knaier@unibas.ch.

Submitted for publication August 2018.
Accepted for publication December 2018.

0195-9131/19/5105-1006/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000001869

between $\dot{V}O_{2peak}$ and $\dot{V}O_{2max}$ (2,6,15). However, secondary criteria to define $\dot{V}O_{2max}$ are rarely reported, despite their importance. Only 76% of studies published in *Medicine and Science in Sports and Exercise* from October 1993 to May 1994 (16) and 44% of studies published from October 2005 to May 2006 (6) reported criteria for $\dot{V}O_{2max}$. The most commonly secondary $\dot{V}O_{2max}$ criteria are maximum RER (RER_{max}), maximum HR (HR_{max}), maximum RPE (RPE_{max}), and maximum concentration of blood lactate (BL_{max}) (6). Values used to confirm $\dot{V}O_{2max}$ attainment include ≥1.00 (17) and ≥1.10 (18,19) to ≥1.15 (5,20) for RER_{max}, 85% (21) to 100% (22) of age-predicted HR_{max} (based on the formula 220 – age in years), ≥17 (18) to ≥19 (6) for RPE_{max}, and from ≥8 (16) to ≥10 $mmol \cdot L^{-1}$ (6) for BL_{max}. Interestingly, and as noted by Midgley et al. (6), participants reached these criteria in nearly all studies, leaving room for speculation as to whether exhaustion criteria were defined postanalysis. However, defining secondary $\dot{V}O_{2max}$ criteria is a trade-off between mistakenly assuming that subjects have reached $\dot{V}O_{2max}$ when they have not (i.e., low criteria, type I error) and declaring subjects to have not reached $\dot{V}O_{2max}$ even though they have (i.e., high criteria, type II error). Therefore, this study analyzed data from two studies with endurance-trained athletes, pursuing the following aims:

Aim I was to define secondary $\dot{V}O_{2max}$ criteria based on the results of the first study. Therefore, differences in $\dot{V}O_2$ were assessed at time points during the CEPT when different criteria were reached based on various cutoffs used in various studies. Based on the results, criteria were defined to reduce type I errors without increasing the risk of type II errors. Aim II was to test if the criteria defined in Aim I would be robust to diurnal variations and day-to-day variations.

METHODS

Study design. In this work data from two different studies were analyzed. The first study was conducted between April 2014 and April 2015 (ClinicalTrials.gov Identifier: NCT02203539) and the second study between December 2016 and April 2018 in the laboratories of the Department of Sport, Exercise and Health at the University of Basel, Switzerland under consistent conditions (air humidity, 40%–55%; room temperature, 20°C–22°C) with the same equipment. Both studies were approved by the local ethics committee "Ethikkommission Nordwest- und Zentralschweiz" (EKNZ 2014-056 and EKNZ 2016-01572) and written informed consent was obtained from all participants before the start of studies. Study 1 primarily investigated the influence of different light exposures on cycling performance. All CEPT data analyzed in this work were performed without previous light exposure under normal room lighting conditions. The purpose of study 2 was to investigate the time of the day when athletes achieve their peak performance and the association of this time of the day with participants' chronotype and training habits.

Participants. Inclusion criteria for both studies were physical and mental health, ages 18 to 35 yr, no shift work in the last 3 months and no travel across time zones in the 4 wk before the study, and high cardiorespiratory fitness. Only participants with a $\dot{V}O_{2max} \geq$ 95th percentile of ACSM references values (i.e., ≥56 $mL \cdot kg^{-1} \cdot min^{-1}$ for males and ≥50 $mL \cdot kg^{-1} \cdot min^{-1}$ for females) (23) were included in this analysis. For study 1, a further criterion was male sex. The main sports performed by participants in the first study were cycling (n = 27), triathlon (n = 8), other endurance sports (e.g., running, rowing, kayak) (n = 11), team sports (e.g., football, hockey, volleyball) (n = 14) and other sports (e.g., tennis, squash (n = 10). The main sports performed by participants in the second study were cycling (n = 8) and other endurance sports (e.g., running, rowing) (n = 5). Participants were advised to restrain from alcohol, caffeine, and vigorous exercise through the entire studies and advised to avoid malnutrition and dehydration.

Participant characteristics testing. In both studies, a physician physically examined participants, a 12-channel resting electrocardiograph was acquired, medical history was assessed, and body height was measured.

Cardiopulmonary exercise testing. In both studies, cardiopulmonary exercise testing until exhaustion was performed using a bicycle ergometer (Sport Excalibur; Lode Medical Technology, Groningen, The Netherlands). Participants were free to choose pedaling cadence as long as it remained over 60 rpm. Participants were allowed to cycle with their own pedals and shoes. Before each test body mass and body fat content were measured with four segment bioelectrical impedance analyses (Inbody 720; Biospace, Seoul, South Korea). In both studies workload linearly increased with 25 $W \cdot min^{-1}$ (20 $W \cdot min^{-1}$ for females in study 2) until exhaustion. In both studies, through the entire CPET gas exchange was measured breath by breath (MetaMax 3B; Cortex Biophysik GmbH, Leipzig, Germany). For analysis data were averaged in 10-s intervals. $\dot{V}O_{2max}$ was determined as the highest 30 consecutive seconds of $\dot{V}O_2$. Maximum workload (P_{max}) was defined as the highest value during exercise with a minimum of 60 rpm. RER_{max} was determined as the highest value during exercise. HR was monitored with a 12-channel electrocardiography (Custo med GmbH, Ottobrunn, Germany) to comply with ethical regulations. For data analyses, HR was additionally measured with a Polar T-34 HR belt (Polar Electro Europe AG, Zug, Switzerland). According to the 6-20 Borg scale (24), RPE was assessed at rest, after warm-up, and every 3 min until exhaustion. Capillary blood lactate concentration (analyzed by SuperGL Ambulance; Hitado Diagnostic Systems, Moehnesee, Germany) was measured at rest, immediately after exhaustion, 1, 3, and 5 min after exhaustion. The highest measured value was labeled as BL_{max}.

In study 1, the only CPET performed was scheduled around 4:00 PM (median time, 4:17 PM). In study 2, six CPET were performed at: 7:00 AM, 10:00 AM, 1:00 PM, 4:00 PM, 7:00 PM, and 9:00 PM in randomized order to investigate the diurnal variation in the criteria defined in study 1. Additionally, a seventh test was performed at the same time the sixth test was performed to investigate the day-to-day variability in the criteria. In study 2, a verification test was performed immediately after the CPET. After a regeneration period of 10 min in duration, workload was increased stepwise to 50% of $P_{\max}$ for 2 min and then to 70% of $P_{\max}$ for 1 min. Subsequently, workload was increased to 105% of $P_{\max}$ until exhaustion (25). $\dot{V}O_{2\max}$ verification was accepted if the verification-$\dot{V}O_2$ was ±3% of the $\dot{V}O_{2\max}$ from the initial phase of the exercise test (25). This protocol has been used in a previous study where it led to a successful verification in all participants (25). However, this protocol showed an insufficient Gwet's agreement coefficient for the diurnal variation indicating a rather low reliability (data not shown). Therefore, the verification data was not considered in this manuscript. A $\dot{V}O_2$ plateau was defined as an increase of $\dot{V}O_2 < 125\ \text{mL}\cdot\text{min}^{-1}$ between the last and the second to last minute of the CEPT. This definition was based on the assumption that $\dot{V}O_2$ increases approximately 10 $\text{mL}\cdot\text{min}^{-1}$ in the submaximal intensity domain per increase of each watt (26). Therefore, the cutoff value amounts to half of the expected $\dot{V}O_2$ increase between neighboring 1-min sampling intervals (~250 $\text{mL}\cdot\text{min}^{-1}$), as recommended by Marsh (27).

Further, we compared the secondary $\dot{V}O_{2\max}$ criteria, that is, the time points at which $\dot{V}O_2$ was analyzed, for RER (≥1.05, ≥1.10, ≥1.15) and HR (≥90%, ≥95%, and ≥100% of age-predicted HR). The formula for age-predicted HR 220 minus age (28) was adapted to 210 minus age in years to consider the lower muscle mass involved in a cycle ergometer test which results in a lower maximum HR in comparison to treadmill tests (29,30). Further, the values reached by the participants for $\text{RPE}_{\max}$ and $\text{BL}_{\max}$ were compared with the values stated in various guidelines.

Statistical analysis. In a first step, descriptive statistics were used to compare the number of participants reaching certain exhaustion criteria and the respective $\dot{V}O_2$ reached at that point in time. Based on these results, cutoff values were defined for secondary exhaustion criteria that were least likely to produce type II errors. The cutoff was defined as 2.5% of participants not reaching a criterion, based on the 95% confidence interval. Furthermore, we compared the four secondary exhaustion criteria ($\text{RER}_{\max}$, $\text{HR}_{\max}$, $\text{RPE}_{\max}$, and $\text{BL}_{\max}$) between participants showing a $\dot{V}O_2$ plateau and those showing no $\dot{V}O_2$ plateau to ensure that the recommendations for the cutoff values were not biased by type I errors. For the second aim, we ascertained the number of tests in study 2 where certain exhaustion criteria were not met. To analyze the day-to-day variation, intraclass correlations were calculated for performance parameters and the four secondary exhaustion criteria values between the sixth and the seventh test from study 2. Intraclass correlation coefficient values were interpreted as follows: <0.5: poor, 0.5–0.75: moderate, 0.75–0.90: good, >0.90: excellent (31). Intraclass correlation does not account for the fact that the coefficient is dependent on the standard deviation, which is irrelevant for the intraindividual reliability. Therefore, we further calculated the standard error of measurement (SEM) (32). Normality was assessed using normal quantile-quantile plots of the residuals and variance homogeneity was assessed using Tukey-Anscombe plots. Descriptive data are presented as mean and standard deviation or median and interquartile range. For our analyses, we used IBM SPSS Statistics for Windows, version 24 (IBM Corp., Armonk, NY) and R version 3.3.1 for graphics (R Foundation for Statistical Computing, Vienna, Austria). The data analyzed in this study were used from a previous study on which the sample size was calculated for. Therefore, we did not perform a sample size calculation for this study (33).

RESULTS

Participant flow and characteristics. Seventy males from study 1 and eight males and five females from study 2 were analyzed. Participant characteristics from medical examination and CEPT are presented in Table 1. There were no relevant differences between male participants from study 1 and study 2.

Definition of cutoff values for secondary exhaustion criteria. Table 2 shows the number and percentage of participants from study 1 reaching the different exhaustion criteria defined in the literature. The table shows the $\dot{V}O_2$ uptake at the time point, when the different exhaustion

TABLE 1. Participant characteristics.

Characteristics	Study 1 (*n* = 70)	Study 2 (Males) (*n* = 8)	Study 2 (Females) (*n* = 5)
Age (yr)	24.8 ± 4.4 (18–35)	27.4 ± 4.9 (22–35)	27.6 ± 6.5 (21–35)
Height (cm)	180 ± 7 (164–197)	179 ± 6 (171–185)	168 ± 7 (158–176)
Body mass (kg)	73 ± 7 (60–88)	73 ± 9 (57–83)	64 ± 6 (56–70)
Body fat content (%)	11 ± 4 (3–22)	10 ± 2 (6–13)	14 ± 4 (10–18)
HR at rest (bpm)	61 ± 11 (36–84)	60 ± 5 (55–67)	54 ± 8 (42–64)
Performance[a]			
$P_{\max}$ (W)	408 ± 40 (300–500)	390 ± 40 (342–459)	307 ± 35 (261–354)
$\dot{V}O_{2\max}$ ($\text{L}\cdot\text{min}^{-1}$)	4.62 ± 0.46 (3.52–5.79)	4.58 ± 0.47 (3.93–5.50)	3.53 ± 0.31 (3.20–3.94)
$\dot{V}O_{2\max}$ ($\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$)	63.0 ± 5.3 (56.0–80.0)	62.8 ± 5.3 (57.3–69.6)	55.2 ± 1.4 (53.6–56.8)
Exhaustion criteria[a]			
$\text{RER}_{\max}$	1.18 ± 0.04 (1.1–1.25)	1.22 ± 0.06 (1.15–1.31)	1.23 ± 0.04 (1.17–1.27)
$\text{HR}_{\max}$ (bpm)	192 ± 9 (173–214)	189 ± 7 (181–191)	187 ± 8 (181–200)
$\text{RPE}_{\max}$	19.9 ± 0.3 (19–20)	20 ± 0 (20–20)	20 ± 0 (20–20)
$\text{BL}_{\max}$ ($\text{mmol}\cdot\text{L}^{-1}$)	14.5 ± 2.4 (8.2–21.0)	14.1 ± 2.2 (10.6–16.4)[b]	11.1 ± 2.2 (8.2–14.0)

Mean ± SD (minimum; maximum).

[a] Values reached in the test with the highest $\dot{V}O_{2\max}$ (study 2).

[b] Available in seven participants.

TABLE 2. Values for $\dot{V}O_2$ at the time point when exhaustion is reached based on different cutoffs.

Criteria	*n* (%) Participants Reaching Criteria	Mean ± SD (95% CI) $\dot{V}O_2$ ($mL \cdot kg^{-1} \cdot min^{-1}$)	Mean ± (95% CI) % of $\dot{V}O_{2max}$
RER_{max}			
≥ 1.15	56 (80)	61.4 ± 5.0 (59.9–62.8)	98 ± 2 (98–99)
≥ 1.10	70 (100)	58.7 ± 5.2 (57.2–60.2)	94 ± 3 (93–95)
≥ 1.05	70 (100)	55.6 ± 5.4 (54.0–57.2)	89 ± 4 (88–90)
HR_{max}			
≥ 100% of APHR	60 (86)	58.0 ± 6.7 (56.1–60.0)	93 ± 7 (91–95)
≥ 95% of APHR	69 (99)	52.0 ± 6.4 (50.1–53.9)	83 ± 8 (81–86)
≥ 90% of APHR	70 (100)	46.7 ± 6.5 (44.8–48.6)	75 ± 8 (72–77)
RPE_{max}			
= 20	62 (89)		
≥ 19	70 (100)		
BL_{max}			
≥ 10 $mmol \cdot L^{-1}$	69 (99)		
≥ 8 $mmol \cdot L^{-1}$	70 (100)		

APHR (210 – age).

criteria were met and the relative proportion of each value from the measured $\dot{V}O_{2max}$. RER_{max} ≥1.1, HR_{max} ≥90% of age-predicted maximum HR (APHR), RPE_{max} ≥19 and BL_{max} ≥8 $mmol \cdot L^{-1}$ were reached by all participants suggesting that these values are unlikely to cause type II errors. Thus, exhaustion criteria in the literature with lower values then these may produce type I errors. The 95% of APHR and BL_{max} ≥10 $mmol \cdot L^{-1}$ values were not reached in only one participant. Therefore, these values fall within the defined 2.5% error range.

A $\dot{V}O_2$ plateau appeared in 40 of the 70 (57%) participants in study 1. There were no relevant or significant differences in any secondary exhaustion criteria between participants showing a $\dot{V}O_2$ plateau (RER_{max} 1.19 ± 0.04; HR_{max} 192 ± 8 bpm; RPE_{max} 19.9 ± 0.3; BL_{max} 14.8 ± 2.1 $mmol \cdot L^{-1}$) and participants showing no $\dot{V}O_2$ plateau (RER_{max} 1.18 ± 0.04; HR_{max} 193 ± 10 bpm; RPE_{max} = 19.9 ± 0.3; BL_{max} 14.4 ± 2.7 $mmol \cdot L^{-1}$). However, participants showing a $\dot{V}O_2$ plateau had a significantly higher $\dot{V}O_{2max}$ (64.2 ± 4.9 $mL \cdot kg^{-1} \cdot min^{-1}$) and power output (417 ± 40 W) than the group without $\dot{V}O_2$ plateau (61.6 ± 5.4 $mL \cdot kg^{-1} \cdot min^{-1}$ and 396 ± 35 W). The mean differences were 2.6 $mL \cdot kg^{-1} \cdot min^{-1}$ (95% CI, 0.1–5.1; P = 0.042) and 22 W (95% CI, 3–40: P = 0.022). The coefficient of variation across all trials for $\dot{V}O_{2max}$ ($L \cdot min^{-1}$) was 3.4%. Figure 1 shows the $\dot{V}O_2$–work relationship of three participants and the $\dot{V}O_2$ values at the points in time the respective criteria were achieved.

Secondary exhaustion criteria: Diurnal variation. RER_{max} ≥1.15 was reached in all tests by nine participants. At one or more occasions (i.e., times of the day) four participants did not reach a value of RER_{max} ≥1.15. All participants reached the value of RER_{max} ≥1.10 at every time of the day except of one. BL_{max} ≥10 $mmol \cdot L^{-1}$ was not reached in all tests by four of five women and by two men in all tests.

Secondary exhaustion criteria: Day-to-day variability. There was a significant intraclass correlation for maximum power output (W) and $\dot{V}O_{2max}$ ($L \cdot min^{-1}$) between the two CPET at the same time of the day (Table 3). These correlations indicate excellent reliability (31) and therefore low day-to-day variability in performance. RER_{max} and HR_{max} also showed significant correlations, indicating low day-to-day variability and robustness. Furthermore, the typical percentage error was lower than 3% for the respective value (Table 3). Twelve of 13 participants rated RPE at 20 on both tests and the remaining participant rated RPE at 19 and 20, respectively. Due to all values from the last test day being 20, no variation was present. Hence, an interclass correlation could not be calculated. BL_{max} showed no significant correlation between the two CPET at the same time of the day. There was no significant difference in lactate concentrations between the first and the last test, which indicates sufficient regeneration time between the tests. Furthermore, the percentage typical error was quite high in comparison to RER_{max} and HR_{max}. The coefficients of variation were 1.6% for P_{max}, 2.6% for $\dot{V}O_{2max}$, 1.8% for RER_{max}, 1.4% for HR_{max} and 14.9% for BL_{max}, respectively.

DISCUSSION

To determine $\dot{V}O_{2max}$ in athletes rather high cutoff values for secondary $\dot{V}O_{2max}$ criteria can be used without increasing the risk of type II errors. However, type I errors may still occur indicating that further methods, such as $\dot{V}O_2$ plateau or $\dot{V}O_2$ validation, may be required. Participant A in Figure 1 showed a $\dot{V}O_2$ plateau and therefore reached $\dot{V}O_{2max}$. However, if $\dot{V}O_{2max}$ criteria were chosen too highly (i.e., RER ≥ 1.15, HR_{max} ≥100% of APHR) this participant would be misclassified as not reaching $\dot{V}O_{2max}$ although he in fact did (i.e., type II error). Participant B showed a $\dot{V}O_2$ plateau and also reached the higher $\dot{V}O_{2max}$ criteria. If this participant had stopped at a submaximal $\dot{V}O_2$ he would still have reached all criteria indicating possible type I errors for these criteria. Participant C, on the other hand, showed no $\dot{V}O_2$ plateau although he by far reached all criteria. For participant C it would still be unclear if the measured $\dot{V}O_2$peak is also the $\dot{V}O_{2max}$. A review by Midgley et al. (6) shows that many published studies have used relatively low exhaustion criteria (6). This strongly increases the chance to assume that subjects have reached $\dot{V}O_{2max}$ when they actually have not. From a study with eight subjects, Poole et al. (15) recommended not to use any secondary exhaustion criteria due to the high risk of type I and type II errors. Based on the data from the present work RER ≥ 1.10, HR_{max} ≥95% of APHR and RPE ≥ 19 are highly likely not to produce type II

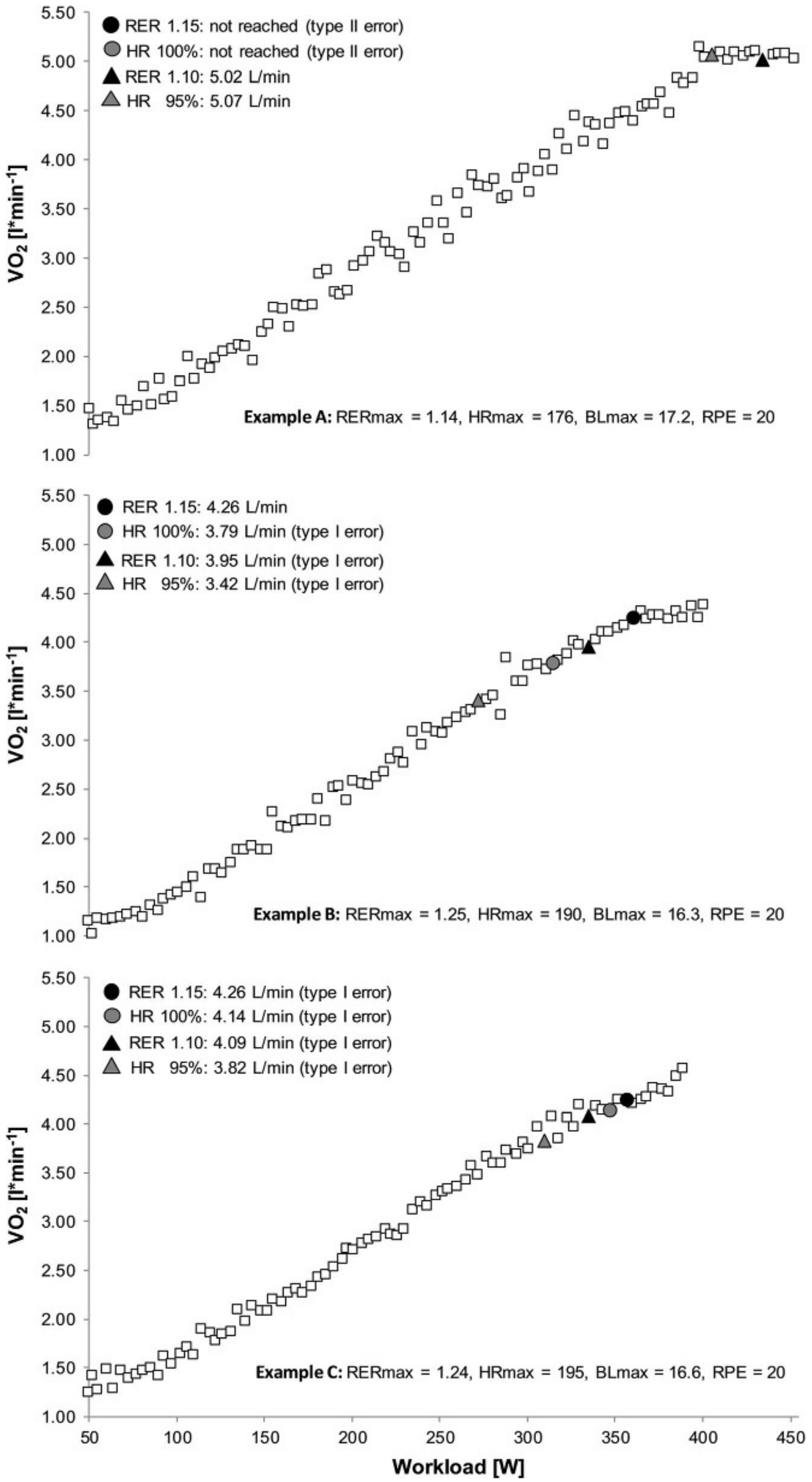

FIGURE 1—$\dot{V}O_2$–work relationship profiles from three participants and the $\dot{V}O_2$ at the time points certain secondary $\dot{V}O_{2max}$ criteria were reached. HR 100%, 100% of APHR; HR 95% of APHR.

errors. Therefore, participants are not misclassified for not reaching $\dot{V}O_{2max}$, they in fact did. If these values are used as the minimum for cutoff values in athletes, it may still occur that subjects are declared as not exhausted even though they are but with a smaller impact on $\dot{V}O_{2max}$ than with lower cutoff values. Misclassification of exhaustion with lower

TABLE 3. Interclass correlation and standard error of measurement of performance and $\dot{V}O_{2max}$ criteria regarding day-to-day variation (day 1, test 6; day 2, test 7 of study 2)—mean ± SD.

	Day 1	Day 2	ICC (95% CI)	SEM (TPE)
Performance				
P_{max} (W)	355 ± 57	356 ± 58	0.990 (0.968 to 0.997)	6 (1.7)
$\dot{V}O_{2max}$ (L·min^{-1})	4.01 ± 0.71	3.99 ± 0.76	0.979 (0.933 to 0.994)	0.10 (2.5)
Exhaustion criteria				
RER_{max}	1.20 ± 0.06	1.20 ± 0.05	0.823 (0.518 to 0.943)	0.02 (1.7)
HR_{max} (bpm)	187 ± 6	187 ± 7	0.828 (0.529 to 0.944)	3 (1.6)
BL_{max} (mmol·L^{-1})	12.2 ± 2.2	12.0 ± 2.4	0.380 (−0.191 to 0.759)	1.7 (13.9)

ICC, intraclass correlation; TPE; typical percentage error.

secondary $\dot{V}O_{2max}$ criteria may lead to the misinterpretation of study results or training progress. In our study population, for example, 75% of participants (i.e., the upper quartile) achieved only 93% or less of their actual $\dot{V}O_{2max}$ when they reached RER ≥1.05 (Table 2). These participants would, therefore, be falsely classified as exhausted. If attainment of $\dot{V}O_{2max}$ is achieved in a subsequent test, a performance increase of 7% is "measured" only due to higher exhaustion levels. Although, there is no ideal cutoff value to prevent all type I and type II errors this does not justify abandoning all secondary exhaustion criteria, because so far there is no acceptable alternative to be used in large-scale studies. Verification-phases are highly dependent of the participants' motivation and $\dot{V}O_2$ plateaus do not appear in all participants (34–36). More importantly, large-scale studies generating reference values for $\dot{V}O_{2max}$ need to use equivalent secondary exhaustion criteria to make the results comparable.

RER ≥ 1.15 and HR_{max} ≥100% of APHR were not reached by several participants during several CEPT in study 1. From the 40 participants showing a $\dot{V}O_2$ plateau in study 1, nine did not reach ≥1.15 and six did not reach HR_{max} ≥100% of APHR. This supports the conclusion drawn, from the results of study 1, that these two values are likely to increase the chance of classifying an exhausted participant as not-exhausted. In contrast, RER_{max} ≥1.10 and HR_{max} ≥95% were reached by all participants showing a $\dot{V}O_2$ plateau in study 1. In study 2, these cutoffs were not reached in only 1 of 78 cases (13 participants × 6 tests each) supporting the recommendation of choosing this cutoff. Furthermore, RER_{max} and HR_{max} showed low day-to-day variability as indicated by the moderate reliability (31) based on the lower bound of the 95% confidence interval of the intraclass correlation (Table 3). The standard error of measurement and typical percentage error are low for both RER_{max} and HR_{max} at 1.7% and 1.6%, respectively. As mentioned intraclass correlations for RPE could not be calculated due to the ceiling effect of this parameter. This ceiling effects lead to the fact that variations can only appear in one direction and therefore increase test–retest reliability. Therefore, it may be discussed if RPE might better serve as an external criterion.

Based on a sample size of 861 participants Edvardsen et al. (37) recommend to use a cutoff value for BL_{max} of 9 mmol·L^{-1} for treadmill exercise testing. Based on our data from study 1, a cutoff value for BL_{max} seems to be applicable; however, data from study 2 investigating the reliability showed a poor and nonsignificant intraclass correlation of 0.380 (−0.191; 0.759) representing poor reliability (31) (Table 3). Additionally, BL_{max} showed a very high typical percentage error of 13.9%. BL_{max} was also highly variable between participants (Table 1). In study 1, lactate concentrations ranged from 8.2 to 21.0 mmol·L^{-1}, making it hard to define generalizable cutoffs for this parameter. Furthermore, blood lactate concentrations showed high variance during the day. This is likely due to varying nutrition status. This variation in nutrition status may lead to lower reproducibility of lactate concentration. However, it is remarkable in this context that RER_{max} showed much better reliability, since RER_{max} values >1.0 are mainly determined by the respiratory compensation of metabolic acidosis, which is affected by the lactate production. Finally, females are less likely to reach the cutoff values for lactate concentration. This may not be due to less exhaustion but rather from fewer fast-twitch fibers (38) and lower activity of total lactate dehydrogenase (38) producing less lactate (39).

Strengths and limitations. The strength of this study is that all CEPT were performed under standardized conditions with the same equipment. All measurements in the first study were performed at the same internal time for each participant and supervised by the same investigator (R.K.). This work is, to the authors' knowledge, the first investigating the robustness of secondary exhaustion criteria regarding diurnal and day-to-day variability.

A limitation of this study is the low sample size for female athletes and the low sample size in study 2 in general, which makes the results not generalizable for female athletes. Furthermore, cutoff values for RER_{max} may differ if much higher (≥50 W·min^{-1}) or lower (<13 W·min^{-1}) increases in workload are chosen than the 25 W·min^{-1} used herein, (40) and that cutoff values for HR_{max} may differ if exercise is performed on a treadmill instead of a bicycle ergometer (29,41). In addition, the formula used for the APHR (210 – age [years]) has not been used in previous studies and therefore reduces comparability with other studies. However, the rationale for choosing this formula was to account for the previously reported lower maximum HR in cycle ergometer tests compared with treadmill tests (29,30). Further, the day-to-day variability was assessed at different times of the day but it is unclear if the biological variability is dependent on the time of the day.

CONCLUSIONS

In trained athletes, high secondary exhaustion criteria cutoffs need to be chosen to reduce type I errors. Based on our analyses we recommend the following cutoffs: (1) RER_{max}

≥ 1.10; (2) $HR_{max} \geq 95\%$ APHR (defined as 210 bpm, age in years); (3) $RPE_{max} \geq 19$. Lower cutoff values are likely to produce type I errors. The defined cutoff values have shown to be robust to diurnal and day-to-day variations. The signal-to-noise ratio in intervention studies and in the evaluation of athlete training programs can only be increased if high secondary exhaustion criteria cutoff values are used. Many of the currently used secondary exhaustion criteria are too low and therefore produce type I errors. However, type I errors may still occur with our defined cutoffs indicating that further methods, such as $\dot{V}O_2$ plateau or $\dot{V}O_2$ validation, may be required.

The authors thank all athletes who made this study possible by their participation.

This study was not funded. None of the authors involved in the present study have any conflict of interest, financial, personal, or otherwise, which would influence this research, and the results do not constitute endorsement by the American College of Sports Medicine. The authors declare that the results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Author contributions: R. K., M. N., and J. W. participated in the writing original draft of the article. D. I., T. H., C. K., S. F., C. C., A. S. T. participated in the review and editing of the article. R. K., M. N., C. C., and A. S. T. participated in the concept and design. R. K., T. H., C. K., and S. F. participated in the data acquisition. R. K., M. N., J. W., and D. I. participated in the data analysis and interpretation. D. I. participated in the statistical expertise.

REFERENCES

1. Blair SN, Kohl HW 3rd, Barlow CE, Paffenbarger RS Jr, Gibbons LW, Macera CA. Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *JAMA*. 1995;273(14):1093–8.
2. Poole DC, Jones AM. Measurement of the maximum oxygen uptake $\dot{V}$o2max: $\dot{V}$o2peak is no longer acceptable. *J Appl Physiol (1985)*. 2017;122(4):997–1002.
3. Barker AR, Williams CA, Jones AM, Armstrong N. Establishing maximal oxygen uptake in young people during a ramp cycle test to exhaustion. *Br J Sports Med*. 2011;45(6):498–503.
4. Lucía A, Rabadán M, Hoyos J, et al. Frequency of the VO2max plateau phenomenon in world-class cyclists. *Int J Sports Med*. 2006;27(12):984–92.
5. Wood RE, Hills AP, Hunter GR, King NA, Byrne NM. Vo2max in overweight and obese adults: do they meet the threshold criteria? *Med Sci Sports Exerc*. 2010;42(3):470–7.
6. 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):1019–28.
7. Scharhag-Rosenberger F, Carlsohn A, Cassel M, Mayer F, Scharhag J. How to test maximal oxygen uptake: a study on timing and testing procedure of a supramaximal verification test. *Appl Physiol Nutr Metab*. 2011;36(1):153–60.
8. Schaun GZ. The maximal oxygen uptake verification phase: a light at the end of the tunnel? *Sports Med - Open*. 2017;3(1):44.
9. Green S, Askew C. $\dot{V}$o2peak is an acceptable estimate of cardiorespiratory fitness but not $\dot{V}$o2max. *J Appl Physiol (1985)*. 2018;125(1):229–32.
10. Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ. The maximally attainable $\dot{V}$o2 during exercise in humans: the peak vs. maximum issue. *J Appl Physiol (1985)*. 2003;95(5):1901–7.
11. 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):764–70.
12. Foster C, Kuffel E, Bradley N, et al. VO2max during successive maximal efforts. *Eur J Appl Physiol*. 2007;102(1):67–72.
13. Astorino TA, White AC, Dalleck LC. Supramaximal testing to confirm attainment of VO2max in sedentary men and women. *Int J Sports Med*. 2009;30(4):279–84.
14. Straub AM, Midgley AW, Zavorsky GS, Hillman AR. Ramp-incremented and RPE-clamped test protocols elicit similar VO2max values in trained cyclists. *Eur J Appl Physiol*. 2014;114(8):1581–90.
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):403–10.
16. Howley ET, Bassett DR Jr, Welch HG. Criteria for maximal oxygen uptake: review and commentary. *Med Sci Sports Exerc*. 1995;27(9):1292–301.
17. Myers J, Kaminsky LA, Lima R, Christle JW, Ashley E, Arena R. A reference equation for normal standards for VO 2 max: analysis from the fitness registry and the importance of exercise National Database (FRIEND registry). *Prog Cardiovasc Dis*. 2017;60(1):21–9.
18. Edvardsen E, Hansen BH, Holme IM, Dyrstad SM, Anderssen SA. Reference values for cardiorespiratory response and fitness on the treadmill in a 20- to 85-year-old population. *Chest*. 2013;144(1):241–8.
19. Nelson MD, Petersen SR, Dlin RA. Effects of age and counseling on the cardiorespiratory response to graded exercise. *Med Sci Sports Exerc*. 2010;42(2):255–64.
20. Duncan GE, Howley ET, Johnson BN. Applicability of VO2max criteria: discontinuous versus continuous protocols. *Med Sci Sports Exerc*. 1997;29(2):273–8.
21. Davis JA, Storer TW, Caiozzo VJ, Pham PH. Lower reference limit for maximal oxygen uptake in men and women. *Clin Physiol Funct Imaging*. 2002;22(5):332–8.
22. Jackson AS, Sui X, Hébert JR, Church TS, Blair SN. Role of lifestyle and aging on the longitudinal change in cardiorespiratory fitness. *Arch Intern Med*. 2009;169(19):1781–7.
23. American College of Sports Medicine. *ACSM's Guidelines for Exercise Testing and Prescription*. 8th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2010. 84 p.
24. Borg G. Ratings of perceived exertion and heart rates during short-term cycle exercise and their use in a new cycling strength test. *Int J Sports Med*. 1982;3(3):153–8.
25. Nolan PB, Beaven ML, Dalleck L. Comparison of intensities and rest periods for VO2max verification testing procedures. *Int J Sports Med*. 2014;35(12):1024–9.
26. Jones AM, Carter H. Oxygen uptake-work rate relationship during two consecutive ramp exercise tests. *Int J Sports Med*. 2004;25(6):415–20.
27. Marsh CE. Validity of oxygen uptake cut-off criteria in plateau identification during horizontal treadmill running. *J Sports Med Phys Fitness*. 2019;59(1):10–6.
28. Cleary MA, Hetzler RK, Wages JJ, Lentz MA, Stickley CD, Kimura IF. Comparisons of age-predicted maximum heart rate equations in college-aged subjects. *J Strength Cond Res*. 2011;25(9):2591.
29. Gordon D, Mehter M, Gernigon M, Caddy O, Keiller D, Barnes R. The effects of exercise modality on the incidence of plateau at VO2max. *Clin Physiol Funct Imaging*. 2012;32(5):394–9.
30. Roecker K, Striegel H, Dickhuth H-H. Heart-rate recommendations: transfer between running and cycling exercise? *Int J Sports Med*. 2003;24(3):173–8.
31. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–63.

32. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30(1):1–15.
33. Knaier R, Schäfer J, Rossmeissl A, et al. Prime time light exposures do not seem to improve maximal physical performance in male elite athletes, but enhance end-spurt performance. *Front Physiol*. 2017;8:264.
34. Astorino TA, DeRevere J. Efficacy of constant load verification testing to confirm VO2 max attainment. *Clin Physiol Funct Imaging*. 2018;38(4):703–9.
35. Bhammar DM, Stickford JL, Bernhardt V, Babb TG. Verification of maximal oxygen uptake in obese and nonobese children. *Med Sci Sports Exerc*. 2017;49(4):702–10.
36. Sawyer BJ, Tucker WJ, Bhammar DM, Gaesser GA. Using a verification test for determination of $\dot{V}O2max$ in sedentary adults with obesity. *J Strength Cond Res*. 2015;29(12):3432–8.
37. Edvardsen E, Hem E, Anderssen SA. End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. *PLoS One*. 2014;9(1):e85276.
38. Esbjörnsson M, Sylvén C, Holm I, Jansson E. Fast twitch fibres may predict anaerobic performance in both females and males. *Int J Sports Med*. 1993;14(5):257–63.
39. Sale DG. Influence of exercise and training on motor unit activation. *Exerc Sport Sci Rev*. 1987;15:95–151.
40. Adami A, Sivieri A, Moia C, Perini R, Ferretti G. Effects of step duration in incremental ramp protocols on peak power and maximal oxygen consumption. *Eur J Appl Physiol*. 2013;113(10):2647–53.
41. McArdle WD, Magel JR. Physical work capacity and maximum oxygen uptake in treadmill and bicycle exercise. *Med Sci Sports*. 1970;2(3):118–23.