



Oxygen uptake plateau: calculation artifact or physiological reality?

Max Niemeyer¹ · Tobias G. J. Bergmann¹ · Ralph Beneke¹

Received: 8 August 2019 / Accepted: 13 November 2019 / Published online: 20 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Purpose To test whether the oxygen uptake ($\dot{V}O_2$) plateau at $\dot{V}O_{2\max}$ is simply a calculation artifact caused by the variability of $\dot{V}O_2$ or a clearly identifiable physiological event.

Methods Forty-six male participants performed an incremental ramp and a $\dot{V}O_{2\max}$ verification test. Variability of the difference between adjacent sampling intervals (difference) and of the slope of the $\dot{V}O_2$ -workload relationship (slope) in the submaximal intensity domain were calculated. Workload defined sampling intervals used for the calculation of the difference and slope were systematically increased from 20 to 100 W until the expected risk of false plateau diagnoses based on the Gaussian distribution function was lower than 5%. Overall, more than 1500 differences and slopes were analyzed. Subsequently, frequencies of plateau diagnoses in the submaximal and maximal intensity domains were compared.

Results Variability of the difference and slope decreased with increasing sampling interval ($p < 0.001$). At a sampling interval of 50 W, the predefined acceptable risk of false plateau diagnoses ($\leq 5\%$) was achieved. At this sampling interval, the actual frequency (1.4%) of false-positive plateau diagnoses did not differ from the expected frequency in the submaximal intensity domain (1.6%; $p = 0.491$). In contrast, the actual frequency at maximal intensity (35.7%) was significantly higher compared to the submaximal intensity domain ($p < 0.001$) and even higher than the expected frequency of false-positive diagnoses ($p < 0.001$).

Conclusion The $\dot{V}O_2$ plateau at $\dot{V}O_{2\max}$ represents a physiological event and no calculation artifact caused by $\dot{V}O_2$ variability. However, detecting a $\dot{V}O_2$ plateau with sufficient certainty requires large sampling intervals.

Keywords $\dot{V}O_2$ · $\dot{V}O_{2\max}$ · Cutoff · Cardiorespiratory fitness · Endurance

Abbreviations

BLC _{max}	Maximal blood lactate concentration
Difference	Difference between the upper and lower half of a sampling interval
GET	Gas exchange threshold
MRT	Mean response time of ramp test oxygen uptake kinetics
P_{GET}	Workload at gas exchange threshold
P_{max}	Maximum workload
RER _{max}	Maximal respiratory exchange ratio
RPE _{max}	Maximal rating of perceived exertion
SD _{IND}	Standard deviation of the intra-individual variability of the plateau criteria

SD _{GROUP}	Standard deviation of the inter-individual variability of the plateau criteria
Slope	Slope of the $\dot{V}O_2$ -workload relationship as a plateau criteria
$\dot{V}O_2$	Oxygen uptake
$\dot{V}O_{2\max}$	Maximal oxygen uptake

Introduction

The oxygen uptake ($\dot{V}O_2$) plateau is defined as a flattening of the $\dot{V}O_2$ -power- or $\dot{V}O_2$ -velocity relationship at the end of a continuous or between the final stages of a discontinuous maximal exercise test (Howley et al. 1995). The $\dot{V}O_2$ plateau is considered as the only valid indicator that the maximum oxygen uptake ($\dot{V}O_{2\max}$) has been achieved (Poole and Jones 2017). Usually, the $\dot{V}O_2$ plateau is determined by calculating the difference between two adjacent sampling intervals (e.g., the last and the next-to-last 30 s) or the slope of the $\dot{V}O_2$ -workload relationship at the end of an incremental load

Communicated by I. Mark Olfert.

✉ Max Niemeyer
niemeyer@staff.uni-marburg.de

¹ Abt. Medizin, Training Und Gesundheit, Institut für Sportwissenschaft Und Motologie, Philipps-Universität Marburg, Jahnstr. 12, 35037 Marburg, Germany

test (Gordon et al. 2011, 2012; Taylor et al. 1955; Vella et al. 2006; Yoon et al. 2007). If the difference or slope is less than a previously defined cutoff value, it is believed that a $\dot{V}O_2$ plateau occurs (Howley et al. 1995). The most common cutoff values are an increase in $\dot{V}O_2 < 2.1 \text{ ml kg}^{-1} \text{ min}^{-1}$ for running exercise or $< 150 \text{ ml min}^{-1}$ for cycling exercise (Midgley et al. 2007). However, most studies did not consider that these cutoff values were validated using a discontinuous exercise test, which induced an increase in $\dot{V}O_2$ of about $4.2 \text{ ml kg}^{-1} \text{ min}^{-1}$ or 300 ml min^{-1} between subsequent stages in the submaximal intensity domain (Taylor et al. 1955). Instead arbitrary time intervals (also called sampling intervals) were used to check whether the difference between the lower and upper half or slope of the $\dot{V}O_2$ –workload relationship during this time interval is less than the chosen cutoff value (Astorino et al. 2005; Astorino 2009; Edvardsen et al. 2014; Gordon et al. 2011, 2012; Vella et al. 2006; Yoon et al. 2007). A common time interval for the diagnoses of a $\dot{V}O_2$ plateau is the final minute of an incremental ramp test (Astorino et al. 2005; Astorino 2009; Edvardsen et al. 2014; Gordon et al. 2011, 2012). For example, in an incremental ramp test with a workload increment of 30 W min^{-1} the average difference in workload between the upper and lower half (i.e., the difference between the last and next-to-last 30 s) of a 60 s time interval amounts exactly to 15 W. Based on an increase in $\dot{V}O_2$ per increase in workload of about $\sim 10 \text{ ml min}^{-1} \text{ W}^{-1}$ (Boone and Bourgois 2012) it can be calculated that the average difference between adjacent 30 s intervals is only 150 ml min^{-1} in the submaximal intensity domain. If the cutoff of Taylor et al. (1955) is used in this case, a plateau will be diagnosed despite no flattening of the $\dot{V}O_2$ –workload relationship occurs. Based on this example, it becomes clear that a fixed cutoff that is applied to an arbitrary sampling-interval leads to a high risk of false-plateau diagnoses (Beltrami et al. 2014; Marsh 2019).

Therefore, cutoff values which consider the test-specific increase in $\dot{V}O_2$ depending on ramp slope or increments (relative cutoffs) are strongly recommended (Midgley et al. 2007, 2009; Marsh 2019). Nevertheless, even when a relative cutoff value of 50% of the expected $\dot{V}O_2$ increase was applied, several false-positive plateau diagnoses were detected by Beltrami et al. (2014). The authors concluded that the $\dot{V}O_2$ plateau “is a calculation artifact rather than an indicator of true physiological events”. However, like many others, Beltrami et al. (2014) also did not take into account that a relative cutoff value is just a necessary, but not per se a sufficient requirement for a valid $\dot{V}O_2$ -plateau definition. A valid $\dot{V}O_2$ -plateau definition should allow detecting a plateau with low risks of false-positive and false-negative diagnoses (Howley et al. 1995). A false-negative plateau diagnosis is thereby defined as a no plateau diagnosis despite a plateau occurring. A false-positive plateau diagnosis is defined as the diagnosis of a plateau despite no plateau occurring.

Plateaus detected in the submaximal intensity domain are per se false-positive diagnoses (Beltrami et al. 2014).

Breath-by-breath measured $\dot{V}O_2$ shows substantial variability (Lamarra et al. 1987; Myers et al. 1990) which is mainly caused by irregularities in the rate and depth of ventilation (Myers et al. 1990; Robergs et al. 2010). This may cause false $\dot{V}O_2$ -plateau diagnoses (Beltrami et al. 2014; Myers et al. 1990). To discriminate between plateaus that are simply caused by the variability of ventilation (false plateaus) and those which are caused by the limitation of the body to transport or utilize $\dot{V}O_2$ (real plateaus), plateaus at $\dot{V}O_{2\text{max}}$ must be more pronounced than plateaus occurring in the submaximal intensity domain. Interestingly, it has never been checked whether plateaus at $\dot{V}O_{2\text{max}}$ exist which are of larger magnitude than plateaus in the submaximal intensity domain. Therefore, the existence of a real plateau (at least in continuous incremental tests) is still unclear.

Since the $\dot{V}O_2$ plateau is defined as a flattening of the $\dot{V}O_2$ –workload relationship, the magnitude of a plateau is given by the degree of deflection and the size of the workload interval during which the deflection occurs. Therefore, the risk of false-positive plateau diagnoses can be reduced by either using more restrictive cutoffs or larger workload defined sampling intervals. The first approach has been chosen by several studies (Astorino et al. 2005; Gordon et al. 2011; Mitchell et al. 1958; Yoon et al. 2007). However, even if the cutoff will be set at a slope or difference equal to zero, false-positive plateau diagnoses cannot be excluded as long as the workload defined sampling interval is of insufficient magnitude. For example, Myers et al. (1990) found negative $\dot{V}O_2$ –workload slopes in several participants in the submaximal intensity domain during an incremental ramp test using 30 breath sampling intervals. Additionally, the use of restrictive cutoffs leads to higher risks of false-negative plateau diagnoses (Howley et al. 1995). The variability of $\dot{V}O_2$ data reflects a Gaussian distribution curve (Lamarra et al. 1987; Myers et al. 1990). Consequently, the cutoff must be set at 50% of the average increase in $\dot{V}O_2$ in the submaximal intensity domain to enable an equal risk of false-positive and false-negative plateau diagnoses. Therefore, the only option to reduce both false-positive and false-negative diagnoses is to use larger workload defined sampling intervals for the diagnoses of a $\dot{V}O_2$ plateau.

The aims of the present study were

1. To test whether the $\dot{V}O_2$ plateau at $\dot{V}O_{2\text{max}}$ is simply a calculation artifact caused by the variability of $\dot{V}O_2$ or a clearly identifiable physiological event.
2. To establish a valid $\dot{V}O_2$ -plateau definition for continuous cycling tests.

For these purposes, the workload defined sampling interval, which is used for the calculation of the difference

between its upper and lower half, as well as the corresponding slope of the $\dot{V}O_2$ –workload relationship should be gradually increased in a continuous ramp test until an acceptable probability of false plateau diagnoses has been achieved (see Fig. 1). In inference based statistics, a commonly accepted level of erroneous decisions is 5% (p value). Therefore, we set the intended risk of false plateau diagnoses at this value. Assuming a Gaussian distributed variability of both plateau criteria, the cutoff should differ more than two standard deviations from the mean increase or slope and even more than two standard deviations from zero (the assumed increase in $\dot{V}O_2$ when a plateau occurs) to enable a risk of false plateau diagnoses below 5%. Therefore, the workload defined sampling interval will be gradually increased until the average induced difference or slope of $\dot{V}O_2$ in the submaximal intensity domain is higher than four of its corresponding standard deviations (i.e., the coefficient of variation (CV) is less than 25%). Since the variability of $\dot{V}O_2$ differs considerably between participants (Myers et al. 1989), this should be performed based on group and individual approaches.

Furthermore, both common plateau criteria (difference and slope) will be compared.

We tested the following hypotheses:

1. The variability of the plateau criteria and, therefore, the frequencies of plateau diagnoses in the submaximal intensity domain (false-positive diagnoses) decreases with increasing workload defined sampling intervals.
2. The variability at a given workload defined sampling interval in the submaximal intensity domain differs between the two plateau criteria (difference and slope).
3. The frequencies of plateau diagnoses in the maximal intensity domain are higher than the frequencies in the submaximal intensity domain and even higher than the expected frequencies of false-positive plateau diagnoses due to the Gaussian distributed variability of $\dot{V}O_2$.
4. A plateau definition that is based on the individual mean slope or difference as well their corresponding individual standard deviations results in a lower risk of false plateau diagnoses compared to a definition that is

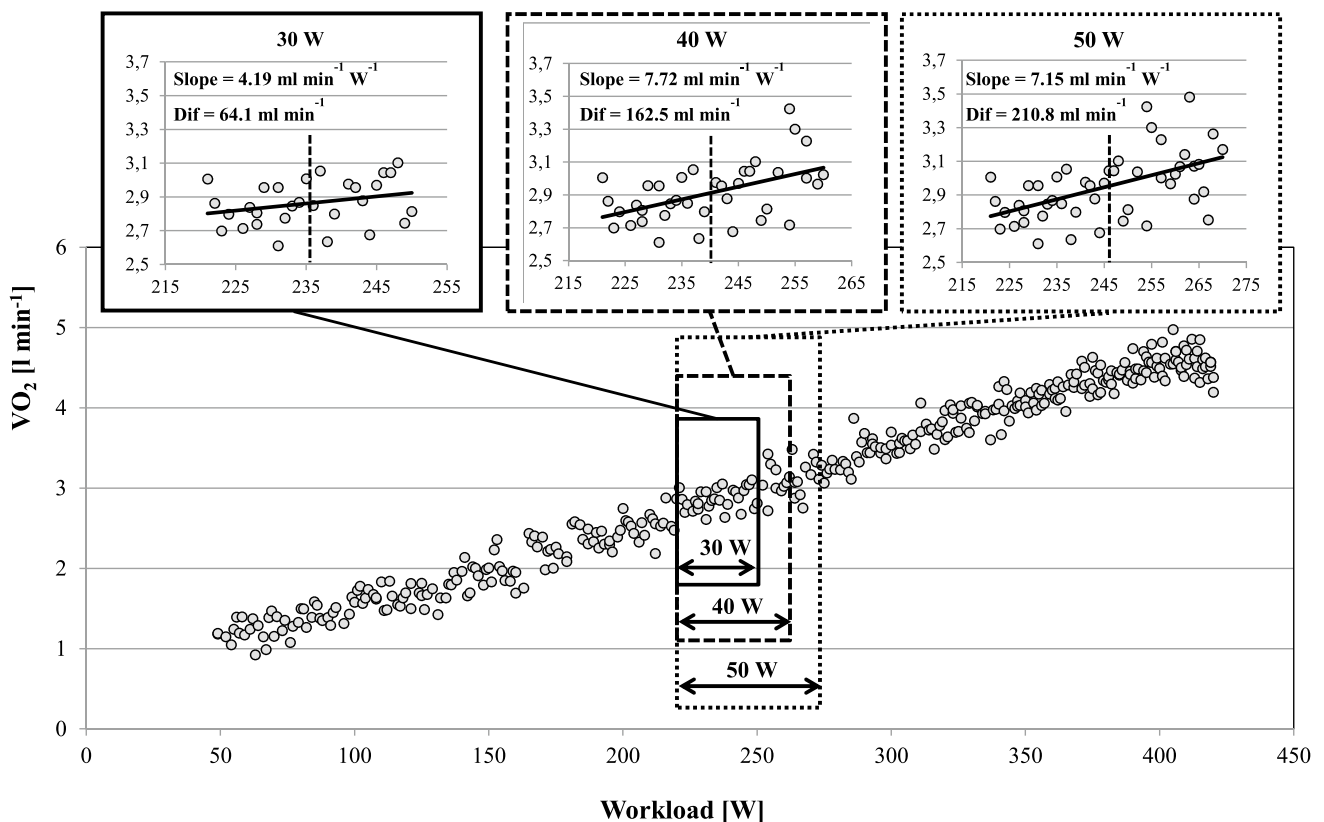


Fig. 1 Determination of the slope and difference in the submaximal intensity domain using workload defined sampling intervals of 30, 40 and 50 W shown for a representative $\dot{V}O_2$ ramp test response. Note that the workload defined sampling intervals were shifted in 10 W steps from 200 to 60 W below maximum workload to get several slopes and differences per participant. The slopes were calculated

from linear regressions that were fitted into the whole workload defined sampling intervals. The differences were calculated by subtracting the mean $\dot{V}O_2$ of the lower from the upper half of a given workload defined sampling interval (the half is represented by the dashed line in small figures)

based on the means and standard deviations of the whole group.

Methods

Participants

Forty-six male, recreationally active participants (age 25.6 ± 3.0 years; height 181.2 ± 6.2 cm; body mass 78.9 ± 8.0 kg) volunteered to participate in the study. After being thoroughly informed about potential risks and discomfort, they gave written consent before each testing day. The participants were instructed to refrain from alcohol, caffeine and exercise 24 h before the tests. The study was approved by the ethics committee of Philipps University Marburg, Department of Education Science (FB-21-AZ1-12-15) and adhered to the Declaration of Helsinki.

Study design

The study comprised two tests per participant on a cycle ergometer (Cyclus-2, RBM elektronik-automation GmbH, Germany) at pedaling rates of 80 ± 2 rpm. On the first testing day, the participants performed an incremental ramp test. After a 3 min warm up at 50 W, the workload was continuously increased at a rate of 30 W min^{-1} until exhaustion. The second test was a square wave $\dot{V}O_{2\text{max}}$ verification test. In accordance with other studies (Day et al. 2004; Sedgeman et al. 2013) and to allow for a minimum time to exhaustion of at least 2 min for each participant, the workload of the verification bout was set at 90% of the maximum workload (P_{max}) of the ramp test. The verification bout was preceded by a warm-up at moderate and severe constant load exercises of 6 min duration each. The workloads of the moderate and severe intensity steps were set at the workload at the gas exchange threshold (P_{GET}) and at 50% of the difference between P_{GET} and P_{max} of the ramp test, respectively. P_{GET} was previously determined from ramp test raw data using the V-slope method (Beaver et al. 1986). Between the warm-up and the verification bout, a 10-min active recovery at 50 W was performed. The warm-up was performed to induce a speeding of $\dot{V}O_2$ kinetics and to prolong time to exhaustion of the verification bout (Bailey et al. 2009). The ramp and verification tests were performed at the same time of day ± 1 h to avoid circadian effects on $\dot{V}O_{2\text{max}}$ (Knaier et al. 2019).

Measurements

Breathing gases were continuously measured during both tests using a breath-by-breath device (MetaMax 3b, Cortex Biophysik GmbH, Germany). The device was calibrated

before each test with a 3 l syringe and a known gas mixture (15% O_2 , 5% CO_2 , and 80% N). Blood samples (20 μl) were collected from hyperaemic earlobes at the end of ramp and verification tests as well as at minute 1, 3, 5, and 7 of the post-exercise phases to determine the maximal blood lactate concentration (BLC_{max}). The blood samples were analyzed with an enzymatic–amperometric device (Biosen C-Line, EKF-diagnostic GmbH, Germany). Rating of perceived exertion (RPE) was assessed at the end of the ramp test with a Borg scale ranging from 6 to 20.

Data analysis

First, all outlining $\dot{V}O_2$ values that differed more than three standard deviations from a previously fitted non-linear function were removed using the 99% prediction bands (Lamarra et al. 1987; Keir et al. 2014). $\dot{V}O_{2\text{max}}$ and maximal respiratory exchange ratio (RER_{max}) were calculated from the 30 s interval with the highest values during the ramp and verification test, respectively. If the $\dot{V}O_{2\text{max}}$ from the verification test was more than 5% higher than the highest 30 s— $\dot{V}O_2$ at ramp test termination, the ramp test was classified as not maximal, which excluded four participants from further analysis.

Second, the mean differences and slopes of the $\dot{V}O_2$ –workload relationship and their corresponding standard deviations were calculated for varying workload defined sampling intervals in the submaximal intensity domain. As shown in Fig. 1, for each workload interval the slope of the $\dot{V}O_2$ –workload relationship was approximated via linear least-squares method regression analysis. Additionally, each workload defined sampling interval was divided into a lower and upper half and the mean of the lower half was subtracted from the mean of the upper half for the calculation of the corresponding difference. As can be seen in Fig. 1, we started with workload defined sampling intervals of 30 W, equivalent to time intervals of 60 s, which is a frequently used time interval for the diagnosis of a $\dot{V}O_2$ plateau (Gordon et al. 2011, 2012). With respect to the group-based coefficient of variation (CV) of the 30 W workload interval, subsequently workload intervals of 40 and 50 W were analyzed by increasing the corresponding time interval to 80 and 100 s, respectively. To get sufficient numbers of slopes and differences, the workload defined sampling intervals were shifted in 10 W steps between 200 and 60 W below P_{max} . An equal number of slopes and differences for every participant independent of their P_{max} could be guaranteed by this approach, and the potential effects of non-linear dynamics of the ramp test $\dot{V}O_2$, caused by the mean response time (MRT) at the beginning of a ramp test or a potential $\dot{V}O_2$ plateau at the end of the ramp test, were eliminated. In practice, this means that if a participant had a P_{max} of 350 W, the first 30 W workload defined sampling interval ranged from 150 up to 180 W, the second from 160 to 190 W and so on, up to 260 to 290 W. At workload defined sampling intervals of

30, 40 and 50 W, this procedure provides 12, 11 and 10 slopes and differences per participant and a total of 504, 462 and 420 slopes and differences, respectively.

Based on the 504, 462 and 420 slopes and differences, we calculated the corresponding means, standard deviations, and CV of the whole group. These were used to determine group-based workload defined sampling intervals and cutoffs, which achieves a risk of false plateau diagnoses below 5%. As written in “Introduction”, this value was chosen because it is a commonly accepted level of erroneous decisions in interference statistics. The group-based cutoffs were set at 50% of the mean slopes and difference of these average values to guarantee an equal risk of false-positive and false-negative plateau diagnoses.

Since the variability of the plateau criteria differs substantially between participants, the CV was calculated additionally for each set of 12, 11 and 10 slopes and differences per participant at workload defined sampling intervals of 30, 40 and 50 W, respectively. In cases where the individual CV of the 30 W workload defined sampling intervals was less than 25%, the slopes and differences were determined at workload defined sampling intervals of 20 W too. If the individual CVs of the 50 W workload defined sampling intervals were higher than 25%, the slopes and differences were determined additionally for 10 W larger workload intervals (= 60 W). This was extended until the individual CV was less than 25% for each participant, meaning that in some participants with a high variability of the plateau criteria workload defined sampling intervals of 60, 70, 80, 90 and 100 W were analyzed. The individual workload defined sampling interval, which enables a risk of false plateau diagnoses < 5%, was defined as the lowest workload defined sampling interval at which the individual CV was less than 25%. The individual cutoff was set at 50% of the mean slope and difference of the individual workload defined sampling interval.

In the next step, the individual and group-based cutoffs were used to detect the frequencies of plateau diagnoses in the submaximal and maximal intensity domain. If the slope or the difference was less than 50% of the mean slope or difference of the corresponding workload defined sampling interval, a plateau was accepted. $\dot{V}O_2$ plateaus occurring in the submaximal intensity domain were classified as false positive. Frequencies of plateau diagnoses in the submaximal and maximal intensity domain were compared with the expected frequencies of false plateau diagnoses. The expected frequencies were calculated based on the Gaussian distribution function (Eq. 1) using Microsoft Excel:

$$F(x) = \frac{1}{SD\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{1}{2}\left(\frac{x-\mu}{SD}\right)^2} dx \quad (1)$$

where μ and SD are the mean and the standard deviation of the plateau criteria and b is equal to the cutoff calculated as

50% of the corresponding mean. The area under the Gaussian distribution function and to the left of the cutoff (dotted vertical line) represents the expected probability of false-positive plateau diagnoses (see Fig. 3). To calculate the expected probability of false-negative plateau diagnoses, it was assumed that the plateau criteria show a similar variability in the maximal compared to the submaximal intensity domain and that the mean slope or difference in case of a $\dot{V}O_2$ plateau is equal to zero. Thus, another Gaussian distribution function with the same standard deviation, but with a mean slope or difference equal to zero, was plotted (see the dashed Gaussian distribution function in Fig. 3). The area under this Gaussian distribution function and to the right of the cutoff represents the probability of false-negative plateau diagnoses (see Fig. 3).

Statistics

Descriptive data were calculated as means and standard deviations, as well as absolute and relative frequencies. The effect of the workload defined sampling interval on the means and standard deviations of both plateau criteria was tested with analyses of variance with repeated measurements and Bonferroni correction. The same approach was used to compare the CVs of the slopes and differences. Normal distributions of plateau criteria at different workload defined sampling intervals were tested via Shapiro–Wilk tests. Furthermore, the distributions were visualized with probability density plots. The frequencies of plateau diagnoses in the submaximal and maximal intensity domain were compared with χ^2 tests. If the expected frequency of plateau diagnoses was less than 5 in more than 20% of the observed cells, we used Fisher’s exact test. For comparison of observed and expected frequencies of false-positive plateau diagnoses we used binomial tests. The level of significance was set at $p = 0.05$.

Results

Four out of the 46 participants attained a more than 5% higher $\dot{V}O_{2\max}$ in the verification test than in the incremental ramp test and were excluded from the subsequent analyses. The remaining 42 participants achieved a P_{GET} , P_{max} , $\dot{V}O_{2\max}$, RER_{max} , BLC_{max} , and RPE_{max} of 148.6 ± 22.0 W, 365.0 ± 39.3 W, 4.01 ± 0.47 l min⁻¹, 1.22 ± 0.06 , 13.55 ± 1.65 mmol l⁻¹ and 18.7 ± 1.5 in the ramp test. $\dot{V}O_{2\max}$, RER_{max} , BLC_{max} , and RPE_{max} of the verification bout were 3.95 ± 0.51 l min⁻¹, 1.13 ± 0.07 , 14.08 ± 2.47 mmol l⁻¹ and 19.2 ± 1.2 , respectively.

Table 1 presents the means and standard deviations of the plateau criteria (slope and difference), which were

calculated in the submaximal intensity domain using workload defined sampling intervals of 30, 40 and 50 W. The mean slope of the $\dot{V}O_2$ -workload relationship did not differ between the workload defined sampling intervals, as indicated by the lack of a main effect [$F(1.12, 45.79) = 0.57$, $p = 0.471$, $\eta^2 = 0.014$]. However, the intra-individual variability of the slope, which is represented by the individual standard deviation (SD_{IND}), decreased when larger workload defined sampling intervals were used to calculate the slope [$F(1.21, 49.60) = 203.1$, $p < 0.001$, $\eta^2 = 0.832$]. The group-based standard deviation (SD_{GROUP}) also decreased with increasing size of workload defined sampling intervals but was about $0.2 \text{ ml min}^{-1} \text{ W}^{-1}$ higher than the SD_{IND} .

The mean difference between the upper and lower half of the workload defined sampling intervals increased with increasing size of workload defined sampling intervals, as indicated by significant main effect [$F(1.64, 67.1) = 1364.8$, $p < 0.001$, $\eta^2 = 0.971$]. In contrast, the intra-individual (SD_{IND}) and the group-based (SD_{GROUP}) variability of the difference decreased with increasing size of workload defined sampling intervals [$F(1.95, 79.9) = 19.74$, $p < 0.001$, $\eta^2 = 0.325$]. The SD_{GROUP} was about $1.6\text{--}6.4 \text{ ml min}^{-1}$ higher than the SD_{IND} .

As shown in Fig. 2, the CVs of the plateau criteria, calculated from the means and standard deviations in the submaximal intensity domain, decreased with increasing size of workload defined sampling intervals [$F(1.35, 55.17) = 172.6$, $p < 0.001$, $\eta^2 = 0.808$]. At 50 W the individual, and group-based CVs of the slope were lower than 25%, which was previously defined as the cutoff for a valid $\dot{V}O_2$ -plateau definition. The individual CV of the difference was also below 25% at a workload defined sampling interval of 50 W. In contrast, the group-based CV of the difference was slightly higher than 25% at the 50 W workload defined sampling interval.

The variability of the slopes was significantly lower at all workload defined sampling intervals than the variability

of the difference [$F(1, 41) = 84.87$, $p < 0.001$, $\eta^2 = 0.674$]. However, the differences between the plateau criteria (slope vs. difference) decreased with increasing size of workload defined sampling intervals, as indicated by an interaction effect between the size of workload defined sampling intervals and the plateau criterion [$F(1.63, 66.69) = 5.14$, $p = 0.013$, $\eta^2 = 0.111$].

When using the slope as a plateau criterion, 3, 11, 17, 6, 3, and 1 out of the 42 participants attained a $CV \leq 25\%$ at workload intervals of 30, 40, 50, 60, 70, and 100 W, respectively. When using the difference as a plateau criterion, 1, 13, 12, 12, 1, 1, and 2 out of the 42 participants attained a $CV \leq 25\%$ at workload intervals of 30, 40, 50, 60, 70, 80, and 90 W, respectively.

As shown in the frequency distribution plots in Fig. 2, the variabilities of the slopes and differences were approximately Gaussian distributed at all workload intervals. This was confirmed by the results of the Shapiro–Wilk test (slope 30 W: $p = 0.329$; slope 40 W: $p = 0.716$; slope 50 W: $p = 0.824$; difference 30 W: $p = 0.433$; difference 50 W: $p = 0.152$), except for the difference at the 40 W workload interval ($p = 0.018$).

Table 2 presents the frequencies of $\dot{V}O_2$ plateau diagnoses of the plateau criteria in the submaximal and maximal intensity domain at different workload defined sampling intervals. There was a reduction of the frequency of plateau diagnoses in the submaximal intensity domain (false-positive diagnoses) when larger workload intervals were analyzed (slope: $\chi^2 = 71.66$; $p < 0.001$; difference: $\chi^2 = 91.99$; $p < 0.001$). The frequencies of plateau diagnoses of both plateau criteria at the end of the ramp test also showed a reduction with increasing magnitude of the workload defined sampling intervals (slope: $\chi^2 = 9.14$; $p = 0.024$; difference: $\chi^2 = 10.40$; $p = 0.015$).

The frequencies of plateau diagnoses at the end of the ramp test (maximal intensity domain) were significantly higher compared to the submaximal intensity domain at all workload defined sampling intervals (slope 30 W: $p < 0.001$;

Table 1 Descriptive data (mean \pm SD) of the $\dot{V}O_2$ -plateau criteria at workload defined sampling intervals of 30, 40 and 50 W

Criterion		30 W	40 W	50 W
Slope ($\text{ml min}^{-1} \text{ W}^{-1}$)	Mean	10.35 ± 1.19	10.30 ± 1.07	10.28 ± 1.06
	SD_{IND}	4.39 ± 1.31	$2.86 \pm 1.00^{\S}$	$2.15 \pm 0.77^{\S,\dagger}$
	SD_{GROUP}	4.54	3.07	2.41
Difference (ml min^{-1})	Mean	156.3 ± 18.4	$204.4 \pm 22.0^{\S}$	$258.5 \pm 27.6^{\S,\dagger}$
	SD_{IND}	74.9 ± 20.3	$67.3 \pm 24.7^{\S}$	$60.5 \pm 22.0^{\S,\dagger}$
	SD_{GROUP}	76.5	71.7	66.9

Slope slope of the $\dot{V}O_2$ -workload relationship, *Difference* difference in $\dot{V}O_2$ between the upper and lower half of the workload defined sampling interval, SD_{IND} individual standard deviations, SD_{GROUP} group-based standard deviations

§ Significantly different from 30 W

† Significantly different from 40 W

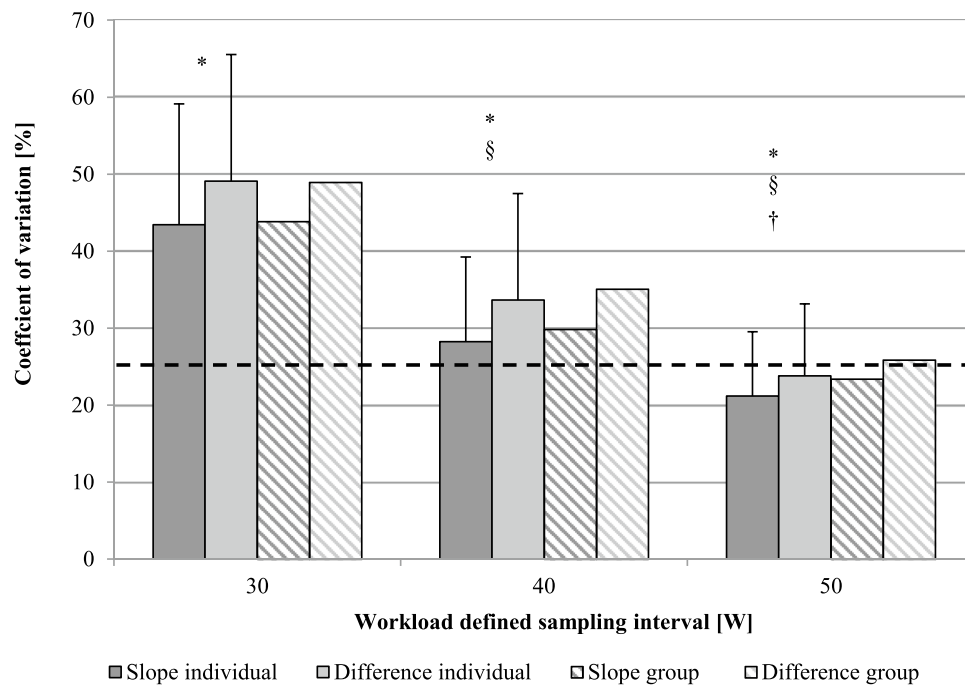


Fig. 2 Coefficient of variation of the individual and group-based variability of the $\dot{V}O_2$ -plateau criteria (slope and difference) at workload defined sampling intervals of 30, 40 and 50 W. Slope individual: mean CV of the slope of the $\dot{V}O_2$ -workload relationship within every single participant, slope group: mean CV of the slope of the $\dot{V}O_2$ -workload relationship of the whole group, Difference individual: mean CV of the difference between the lower and upper half of work-

load defined sampling intervals within every single participant, difference group: mean CV of the difference between the lower and upper half of workload defined sampling intervals of the whole group; § significantly different from 30 W, † significantly different from 40 W, * significant differences between slope and difference; the dashed line marks the predefined cutoff for a valid $\dot{V}O_2$ -plateau definition (coefficient of variation $\leq 25\%$)

Table 2 Frequencies and percentages of $\dot{V}O_2$ -plateau diagnoses in the submaximal and maximal intensity domain using both plateau criteria at different workload defined sampling intervals

Criterion	Workload defined sampling interval	Submaximal					Maximal				
		n total	exp. n	n	exp. %	%	n total	exp. n	N	exp. %	%
Slope	30	504	64.1	53	12.7	10.5	42	5.3	21 ^{*,#}	12.7	50.0 ^{*,#}
	40	462	20.7	19 [§]	4.7	4.1 ^{§,†}	42	2.0	19 ^{*,#}	4.7	45.2 ^{*,#}
	50	420	6.8	6 ^{§,†}	1.6	1.4 ^{§,†}	42	0.7	15 ^{*,#}	1.6	35.7 ^{*,#}
	Individual	423	3.7	1 ^{§,†}	0.9	0.2 ^{§,†}	42	0.4	16 ^{*,#}	0.9	38.1 ^{*,#}
Difference	30	504	77.4	68	15.4	13.5	42	6.5	22 ^{*,#}	15.4	52.4 ^{*,#}
	40	462	35.6	25 ^{§,*}	7.7	5.4 ^{§,*}	42	3.2	17 ^{*,#}	7.7	40.5 ^{*,#}
	50	420	11.2	9 ^{§,†}	2.7	2.1 ^{§,†}	42	1.1	13 ^{*,#,§}	2.7	31.0 ^{*,#,§}
	Individual	410	3.7	0 ^{§†,*}	0.9	0.0 ^{§,†,*}	42	0.4	17 ^{*,#}	0.9	40.5 ^{*,#}

Data are absolute (n) and relative frequencies (%) of $\dot{V}O_2$ plateau diagnoses in the submaximal and maximal intensity domain. Difference: difference between the upper and lower half of the corresponding workload defined sampling interval as a plateau criterion

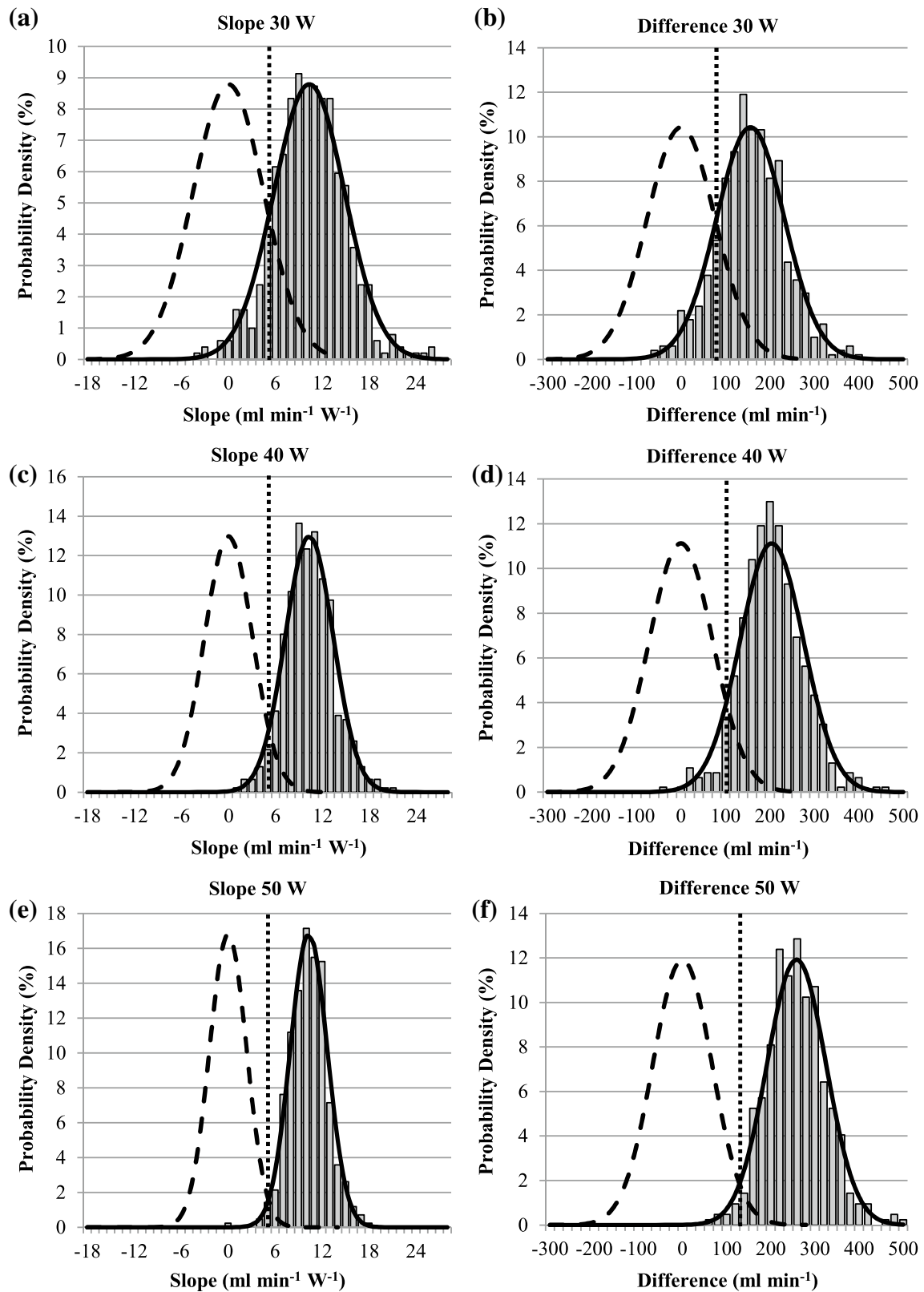
Slope slope of the $\dot{V}O_2$ -workload relationship of the corresponding workload defined sampling interval as a plateau criterion, exp. n expected absolute frequency of false-positive plateau diagnoses, exp. % expected relative frequency of false-positive plateau diagnoses

* Significantly different from the expected frequency of false-positive plateau diagnoses

Significantly different from the submaximal intensity domain

§ Significantly different from the 30 W workload defined sampling interval

† Significantly different from the 40 W workload defined sampling interval



slope 40 W: $p < 0.001$; slope 50 W: $p < 0.001$; slope individual: $p < 0.001$; difference 30 W: $p < 0.001$; difference 40 W:

$p < 0.001$; difference 50 W: $p < 0.001$; difference individual: $p < 0.001$).

Fig. 3 Probability density plots of the plateau criteria at varying workload defined sampling intervals; **a** slope at 30 W, **b** difference at 30 W, **c** slope at 40 W, **d** difference at 40 W, **e** slope at 50 W, **f** difference at 50 W. The solid line represents the Gaussian distribution function of the corresponding plateau criteria and workload defined sampling interval in the submaximal intensity domain. The dashed line represents the assumed Gaussian distribution function of the corresponding plateau criteria and workload defined sampling interval in case of a $\dot{V}O_2$ plateau. The dotted vertical line represents the cutoff, which is set at 50% of the mean slope/difference. Note that the area under the intersection point of the Gaussian distribution functions represents the expected probability of false-positive (left to the dotted vertical line) and false-negative (right to the dotted vertical line) plateau diagnoses

Furthermore, the frequencies of plateau diagnoses at the end of the ramp test were also significantly higher compared to the expected frequency of false-positive plateau diagnoses, which were calculated from the Gaussian distribution function (slope 30 W: $p < 0.001$; slope 40 W: $p < 0.001$; slope 50 W: $p < 0.001$; slope individual: $p < 0.001$; difference 30 W: $p < 0.001$; difference 40 W: $p < 0.001$; difference 50 W: $p < 0.001$; difference individual: $p < 0.001$).

In the submaximal intensity domain, the observed frequencies of plateau diagnoses did not differ from the expected frequencies, except for the difference at 40 W and the difference at the individual workload defined sampling interval (slope 30 W: $p = 0.077$; slope 40 W: $p = 0.322$; slope 50 W: $p = 0.491$; slope individual: $p = 0.106$; difference 30 W: $p = 0.129$; difference 40 W: $p = 0.034$; difference 50 W: $p = 0.302$; difference individual: $p = 0.025$).

Discussion

The main findings of the present study are: (1) that a workload defined sampling interval of at least 50 W is required to reduce the risk of false plateau diagnoses below 5%; (2) that the slope has a lower CV at a given workload defined sampling interval than the difference and is, therefore, more suitable for the diagnosis of a $\dot{V}O_2$ plateau; (3) that the $\dot{V}O_2$ -plateau incidences at the end of ramp tests are higher than the plateau incidences in the submaximal intensity domain and even higher than the expected frequencies of false-positive plateau diagnoses, indicating that the $\dot{V}O_2$ plateau in continuous ramp tests is not a calculation artifact; (4) that a $\dot{V}O_2$ -plateau definition which is based on the individual slope or difference and their corresponding variabilities' results in an almost similar plateau incidence compared to a definition which is based on the mean and standard deviation of a whole group.

Effect of workload defined sampling interval on the variability of the plateau criteria

We found a reduction of the CV of both plateau criteria with increasing magnitude of workload defined sampling intervals, which is consistent with our first hypothesis. At a workload defined sampling interval of 50 W, the CV of the slope is lower than the previously defined threshold of 25%. Since the variability of the plateau criteria is Gaussian distributed (see Fig. 3), the expected frequencies of false-positive and false-negative plateau diagnoses at a CV of $\leq 25\%$ are $\leq 2.5\%$ each, resulting in a combined total of $\leq 5\%$. This indicates that a workload defined sampling interval of at least 50 W is required to achieve a sufficiently low risk of false plateau diagnoses. A workload defined sampling interval of 50 W is much higher than most previously chosen workload defined sampling intervals (Edvardsen et al. 2014; Gordon et al. 2012; Poole et al. 2008; Rivera-Brown et al. 2001; Vella et al. 2006; Yoon et al. 2007). In these studies, the $\dot{V}O_2$ plateau was calculated from the difference between the last and the next-to-last 15 or 30 s periods of an incremental load test. Since the workload increments were usually $\leq 30 \text{ W min}^{-1}$, there is a high probability that the findings of these studies are strongly affected by several false plateau diagnoses. Our finding that rather large workload defined sampling intervals are required for a valid $\dot{V}O_2$ -plateau diagnosis is also contrary to the recommendations of previous studies (Astorino et al. 2005, Astorino 2009; Thomson et al. 2015). These studies simply checked at which sampling-interval durations the highest plateau incidences at the end of incremental tests occurred. Based on these analyses, they recommend very short sampling intervals (15 s or 15 breaths) for the diagnosis of a $\dot{V}O_2$ plateau. However, they did not consider the variability of $\dot{V}O_2$, which is mainly caused by irregularities in the rate and depth of ventilation (Myers et al. 1990; Robergs et al. 2010). Because of the lower number of breaths that were averaged, shorter sampling intervals showed a substantial higher variability than longer ones (Myers et al. 1990). Additionally, they used fixed cutoffs and did not consider that the difference in workload between adjacent sampling intervals in continuous ramp tests decreased when shorter sampling-interval durations are used, as explained in “Introduction”. The extremely high plateau incidences (91–100%) which have been reported when using 15 s or 15 breath sampling-intervals (Astorino et al. 2005, Astorino 2009; Thomson et al. 2015) are, therefore, very likely caused by several false-positive plateau diagnoses. The common use of time intervals instead of workload intervals for the diagnoses of a $\dot{V}O_2$ plateau is the main reason for the above-explained mismatch between the increase in $\dot{V}O_2$ and the applied cutoff in several studies. Therefore, we

decided to report workload defined sampling intervals instead of time intervals to establish a valid $\dot{V}O_2$ -plateau definition. Since workload is linear function of time in incremental ramp tests, the number of breaths that are averaged during a given workload interval changes when ramp tests with higher or lower incremental rates are used. At a first glance, this seems to limit the transferability of the provided plateau definition to other incremental tests. However, the effect of sampling interval duration on the variability of adjacent sampling intervals decreases in an exponential fashion (Myers et al. 1990) and the sampling interval duration at the 50 W workload defined sampling interval is already rather long (50 s for the upper and lower half of the whole workload interval). Therefore, small changes in the incremental rate (± 10 W) will have only minor effects on the validity of our provided definition.

Slope or difference?

Based on the classical plateau definition by Taylor et al. (1955), the difference between two adjacent sampling intervals at the end of an incremental load test has been used to identify a $\dot{V}O_2$ plateau in most studies (Edvardsen et al. 2014; Gordon et al. 2011, 2012; Rivera-Brown et al. 2001; Vella et al. 2006). However, as shown in Fig. 2, the slope had a lower CV at a given workload defined sampling interval than the difference. Based on this observation, we recommend using the slope to detect a $\dot{V}O_2$ plateau at incremental ramp tests.

However, the slope of the $\dot{V}O_2$ -workload relationship cannot be calculated for stepwise or discontinuous incremental load tests. The mean workload difference of the upper and lower half of a given workload interval in a ramp test amounts exactly to one half of the workload interval. As a consequence, the mean difference between the upper and lower half of the 50 W workload interval is 25 W. This indicates that the difference between consecutive stages at stepwise incremental or discontinuous exercise tests should be > 25 W to detect a $\dot{V}O_2$ plateau with sufficient probability. It also needs to be taken into account that the sampling interval durations which were used to calculate the difference between adjacent sampling intervals increased with increasing magnitude of workload defined sampling intervals in the present study. The durations of the upper and lower halves of workload defined sampling intervals were each 30, 40 and 50 s at workload intervals of 30, 40 and 50 W, respectively. As described in “Introduction”, the variability between adjacent samples of breath-by-breath measured $\dot{V}O_2$ decreases with increasing sampling interval duration (Myers et al. 1990). Thus, sampling interval durations of > 50 s combined with workload steps of > 25 W should be used to apply the findings

of the present study to stepwise incremental or discontinuous exercise tests properly.

Calculation artifact or physiological reality?

Based on a high occurrence of false-positive plateau diagnoses, Beltrami et al. (2014) concluded that the $\dot{V}O_2$ plateau is a calculation artifact rather than an indicator of a true physiological event. In the present experiment, we also found numerous false-positive plateaus in the submaximal intensity domain at the lowest workload defined sampling interval (30 W), despite the use of a relative cutoff. However, with increasing magnitude of workload defined sampling intervals, the frequencies of false-positive plateau diagnoses decreased markedly and they were less than 2.5% when using an individual as well as a 50 W workload defined sampling interval. The reduction of plateau diagnoses in the submaximal intensity domain with increasing magnitude of workload defined sampling interval was accompanied by a reduction of plateau diagnoses at the end of the ramp test. This is mainly caused by a reduced frequency of false-positive plateau diagnoses. Based on the Gaussian distributed variability of the plateau criteria, it can be calculated that in the present study 5 or 6 plateau diagnoses in the maximum intensity domain at a workload defined sampling interval of 30 W are most probably false positives. This number can be reduced to 1 when applying 50 W workload defined sampling intervals.

After minimizing the risk of false-positive plateau diagnosis using the aforementioned methods, we still observed plateaus at the end of the ramp test for about 30–40% of the participants. Furthermore, the plateau incidences at the end of the ramp test were significantly higher compared to the submaximal intensity domain and even compared to the expected frequencies of false-positive plateau diagnoses due to the variability of $\dot{V}O_2$ at all workload defined sampling intervals and both criteria. These findings clearly support the idea that the observation of a $\dot{V}O_2$ plateau at $\dot{V}O_{2\max}$ in continuous incremental load tests does represent a physiological event and is not an artifact caused by the variability of $\dot{V}O_2$, as suggested by Beltrami et al. (2014) and others (Myers et al. 1989, 1990).

Individual or group-based plateau definition?

The variability of the plateau criteria differed considerably between participants. As reported in “Results”, some participants achieved a $CV \leq 25\%$ already at a workload defined sampling interval of 30 W, while others achieved this value only at a workload defined sampling interval of ≥ 70 W. Fixed group-based workload defined sampling intervals fail to take this variability into account. This leads to a higher risk of false positives for participants with a $CV > 25\%$ at

the arbitrary group-based workload defined sampling interval while even lower workload defined sampling intervals would have been sufficient for individuals with low variability. As a result, some participants with a $CV \leq 25\%$ at a workload defined sampling interval of 30 and 40 W showed a plateau at the end of the ramp test when using these values for plateau detection, but not when using a workload defined sampling interval of 50 W. This is most likely caused by the fact that a $\dot{V}O_2$ plateau cannot be sustained for long durations because of a rapid $\dot{V}O_2$ deficit accumulation after the attainment of $\dot{V}O_{2max}$, as recently shown (Niemeyer et al. 2019). Consequently, the risk of individual false plateau diagnoses is higher when using group-based methods.

Therefore, a plateau definition which is based on the individual variability of the plateau criteria seems to be more valid than and thus preferable to a group-based workload defined sampling interval and cutoff. However, the $\dot{V}O_2$ -plateau incidence at the end of the ramp test was just slightly, and not significantly, affected by the group or individual approach. Furthermore, the analysis of the variability in the submaximal intensity domain at different workload defined sampling intervals is very time consuming. Therefore, we recommend the individual approach only for studies where the $\dot{V}O_2$ plateau is the main outcome. If the $\dot{V}O_2$ plateau serves simply as a criterion for the diagnoses of $\dot{V}O_{2max}$, a fixed workload defined sampling interval of 50 W for all participants seems to be appropriate for most issues.

Since the slope of the $\dot{V}O_2$ -workload relationship differs between participants (Boone and Bourgois 2012; Niemeyer et al. 2019), some authors recommended using the individual mean slope or difference to calculate an individual cutoff (Midgley et al. 2007, 2009). However, the findings of our study clearly show that simply using a relative cutoff, which accounts for the increase of $\dot{V}O_2$ in the submaximal intensity domain, does not prevent a substantial risk of false plateau diagnoses. To ensure a sufficiently low risk of false plateau diagnoses, the relative cutoff has to cover also a large workload defined sampling interval as well. Additionally, one has to keep in mind that the $\dot{V}O_2$ -plateau incidence is almost unaffected by the use of an individual cutoff in comparison to a group-based relative cutoff (Beltrami et al. 2014).

To reduce the risk of false-positive plateau diagnoses, several studies used cutoffs of considerably less than 50% of the mean difference or slope (Beltrami et al. 2014; Gordon et al. 2011; Mitchell et al. 1958; Yoon et al. 2007). As shown in Fig. 3, this approach leads to an increased risk of false-negative plateau diagnoses. If for example the $< 54 \text{ ml min}^{-1}$ cutoff of Mitchell et al. (1958) is used at the 30 W workload defined sampling interval, the probability of false-negative plateau diagnoses increases from 15.4 to 24.0%. Since the corresponding reduction of the probability of false-positive plateau diagnoses from 15.4 to 9.1% is less pronounced, the overall risk of false plateau diagnoses (false-negative

and -positive) increases from 30.8 to 33.1%. Consequently, we recommend setting the cutoff at 50% of the mean slope or difference. At incremental rates between 20 and 40 W min^{-1} , the slope of the $\dot{V}O_2$ -workload relationship is about $\sim 10 \text{ ml min}^{-1} \text{ W}^{-1}$ (Boone and Bourgois 2012). Therefore, a cutoff of $5 \text{ ml min}^{-1} \text{ W}^{-1}$ (slope) or $125 \text{ ml}^{-1} \text{ min}^{-1}$ (difference) seems to be appropriate for ramp tests with an incremental rate between 20 and 40 W min^{-1} , as long as a workload defined sampling interval of 50 W is used.

Conclusion

The findings of our study strongly indicate that the $\dot{V}O_2$ plateau at $\dot{V}O_{2max}$ reflects a physiological event and not just a calculation artifact, which is caused by the variability of $\dot{V}O_2$. However, rather high workload defined sampling intervals are necessary to discriminate $\dot{V}O_2$ plateaus with sufficient certainty from the inherent noise of breath-by-breath measured $\dot{V}O_2$. It can also be stated that despite considerable differences between participants in terms of the variability of $\dot{V}O_2$, a workload defined sampling interval of $\geq 50 \text{ W}$ and a cutoff which is set at the half of the expected slope or difference seem to be appropriate for most testing purposes. Lastly, the slope of the $\dot{V}O_2$ -workload relationship has a lower variability at given workload defined sampling interval than the difference between adjacent sampling intervals and is, therefore, more suitable as a plateau criterion for incremental ramp tests.

For the diagnoses of a $\dot{V}O_2$ plateau, we recommend the following approach:

- For studies where the $\dot{V}O_2$ plateau is the main outcome, the required workload interval should be calculated based on the individual variability of $\dot{V}O_2$ as described.
- For all other purposes, $\dot{V}O_2$ of last and the next-to-last 25 W of an incremental ramp test should be compared, or even better the slope of the $\dot{V}O_2$ -workload relationship of the final 50 W should be calculated. If the difference or slope is less than the half of the expected increase in $\dot{V}O_2$ (difference $< 125 \text{ ml min}^{-1}$; slope $< 5 \text{ ml min}^{-1} \text{ W}^{-1}$), a real plateau can be diagnosed with a risk of false plateau diagnoses of $\sim 5\%$.
- Since the number of breaths per workload interval and the slope of the $\dot{V}O_2$ -workload relationship ($\Delta\dot{V}O_2/\Delta P$) changes when higher or lower incremental rates are used this plateau definition should be only used for ramp tests with an incremental rate of $20\text{--}40 \text{ W min}^{-1}$.
- To apply the present findings to stepwise incremental tests properly, the workload steps and the sampling interval durations should be at least 25 W and 50 s, respectively.

Data sharing statement

Data are available upon reasonable request.

Acknowledgements We thank all persons who made this study possible by their participation.

Author contributions MN: conceived and designed research, conducted experiments, analyzed data and wrote the manuscript. TGJB and RB: wrote and critically revised the manuscript. All authors read and approved the manuscript.

Funding No funding was received for this study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Ethics committee of Philipps-University Marburg; FB-21-AZ1-12-15) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

- Astorino TA (2009) Alterations in VO_{2max} and the VO plateau with manipulation of sampling interval. *Clin Physiol Funct Imaging* 29(1):60–67
- Astorino TA, Willey J, Kinnahan J, Larsson SM, Welch H, Dalleck LC (2005) Elucidating determinants of the plateau in oxygen consumption at VO_{2max} . *Br J Sports Med* 39(9):655–660
- Bailey SJ, Vanhatalo A, Wilkerson DP, Dimenna FJ, Jones AM (2009) Optimizing the "priming" effect: influence of prior exercise intensity and recovery duration on O_2 uptake kinetics and severe-intensity exercise tolerance. *J Appl Physiol* 107(6):1743–1756
- Beaver WL, Wasserman K, Whipp BJ (1986) A new method for detecting anaerobic threshold by gas exchange. *J Appl Physiol* (1985) 60(6):2020–2027
- Beltrami FG, del Wong P, Noakes TD (2014) High prevalence of false-positive plateau phenomena during VO_{2max} testing in adolescents. *J Sci Med Sport* 17(5):526–530
- Boone J, Bourgois J (2012) The oxygen uptake response to incremental ramp exercise: methodological and physiological issues. *Sports Med* 42(6):511–526
- Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ (2004) The maximally attainable VO_2 during exercise in humans: the peak vs. maximum issue. *J Appl Physiol* 95(5):1901–1907
- Edvardsen E, Hem E, Anderssen SA (2014) End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. *PLoS One* 14 9(1):e85276
- Gordon D, Hopkins S, King C, Keiller D, Barnes RJ (2011) Incidence of the plateau at VO_{2max} is dependent on the anaerobic capacity. *Int J Sports Med* 32(1):1–6
- Gordon D, Schaitel K, Pennefather A, Gernigon M, Keiller D, Barnes R (2012) The incidence of plateau at VO_{2max} is affected by a bout of prior-priming exercise. *Clin Physiol Funct Imaging* 32(1):39–44
- Howley ET, Bassett DR, Welch HG (1995) Criteria for maximal oxygen uptake: review and commentary. *Med Sci Sports Exerc* 27(9):1292–1301
- Keir DA, Murias JM, Paterson DH, Kowalchuk JM (2014) Breath-by-breath pulmonary O_2 uptake kinetics: effect of data processing on confidence in estimating model parameters. *Exp Physiol* 99(11):1511–1522
- Knaier R, Infanger D, Niemeyer M, Cajochen C, Schmidt-Trucksäss A (2019) In athletes, the diurnal variations in maximum oxygen uptake are more than twice as large as the day-to-day variations. *Front Physiol* 18(10):219
- Lamarra N, Whipp BJ, Ward SA, Wasserman K (1987) Effect of inter-breath fluctuations on characterizing exercise gas exchange kinetics. *J Appl Physiol* 62(5):2003–2012
- Marsh CE (2019) Validity of oxygen uptake cut-off criteria in plateau identification during horizontal treadmill running. *J Sports Med Phys Fitness* 59(1):10–16
- Midgley AW, McNaughton LR, Polman R, Marchant D (2007) Criteria for determination of maximal oxygen uptake: a brief critique and recommendations for future research. *Sports Med* 37(12):1019–1028
- Midgley AW, Carroll S, Marchant D, McNaughton LR, Siegler J (2009) Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Appl Physiol Nutr Metab* 34(2):115–123
- Mitchell JH, Sproule BJ, Chapman CB (1958) The physiological meaning of the maximal oxygen intake test. *J Clin Invest* 37(4):538–547
- Myers J, Walsh D, Buchanan N, Froelicher VF (1989) Can maximal cardiopulmonary capacity be recognized by a plateau in oxygen uptake? *Chest* 96(6):1312–1316
- Myers J, Walsh D, Sullivan M, Froelicher V (1990) Effect of sampling on variability and plateau in oxygen uptake. *J Appl Physiol* 68(1):404–410
- Niemeyer M, Leithaeuser R, Beneke R (2019) Oxygen uptake plateau occurrence depends on oxygen kinetics and oxygen deficit accumulation. *Scand J Med Sci Sports* 29(10):1466–1472
- Poole DC, Jones AM (2017) Measurement of the maximum oxygen uptake VO_{2max} : VO_{2peak} is no longer acceptable. *J Appl Physiol* 122(4):997–1002
- Poole DC, Wilkerson DP, Jones AM (2008) Validity of criteria for establishing maximal O_2 uptake during ramp exercise tests. *Eur J Appl Physiol* 102(4):403–410
- Rivera-Brown AM, Alvarez M, Rodríguez-Santana JR, Benetti PJ (2001) Anaerobic power and achievement of VO_2 plateau in prepubertal boys. *Int J Sports Med* 22(2):111–115
- Robergs RA, Dwyer D, Astorino T (2010) Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Med* 1 40(2):95–111
- Sedgeman D, Dalleck L, Clark IE, Jamnick N, Pettitt RW (2013) Analysis of square-wave bouts to verify VO_{2max} . *Int J Sports Med* 34(12):1058–1062
- Taylor HL, Buskirk E, Henschel A (1955) Maximal oxygen intake as an objective measure of cardio-respiratory performance. *J Appl Physiol* 8(1):73–80
- Thomson AC, Ramos JS, Fassett RG, Coombes JS, Dalleck LC (2015) Optimal criteria and sampling interval to detect a VO_2 plateau at VO_{2max} in patients with metabolic syndrome. *Res Sports Med* 23(4):337–350
- Vella CA, Marks D, Robergs RA (2006) Oxygen cost of ventilation during incremental exercise to VO_{2max} . *Respiratory* 11(2):175–181
- Yoon BK, Kravitz L, Robergs R (2007) VO_{2max} , protocol duration, and the VO_2 plateau. *Med Sci Sports Exerc* 39(7):1186–1192

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.