Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

Luca Francesco Masieri

A Reappraisal of Ventilatory Thresholds in Cycling

What is the Evidence?

Master's thesis in Physical Activity and Health - Movement Science
Supervisor: Gertjan Ettema
Co-supervisor: Julia Kathrin Baumgart
May 2023

NTNU
Norwegian University of
Science and Technology

Luca Francesco Masieri

A Reappraisal of Ventilatory Thresholds in Cycling

What is the Evidence?

Master's thesis in Physical Activity and Health - Movement Science
Supervisor: Gertjan Ettema
Co-supervisor: Julia Kathrin Baumgart
May 2023

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

NTNU
Norwegian University of
Science and Technology

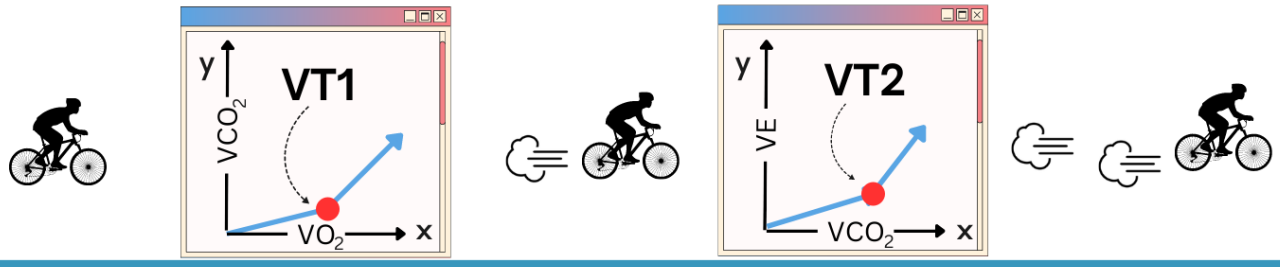
REAPPRAISAL OF VENTILATORY THRESHOLDS IN CYCLING
WHAT IS THE EVIDENCE?

BACKGROUND:

Breakpoints based on ventilatory data are commonly used to separate intensity zones for training. Two distinct ventilatory thresholds (VT1, VT2) are identified during graded exercise testing (GXT).

...BUT: Do breakpoints really exist?

LOW INTENSITY | MODERATE INTENISTY | HIGH INTENSITY



AIM:

1) Compare breakpoint vs NO breakpoint model fit at VT1 and VT2.
2) Reliability of % VO_2 peak at VT1 and VT2 during GXT (Test / Re-Test).
3) Impact of protocol on the breakpoint outcome (Ramp / Graded).

METHOD:

PRIMARY DATA (OWN)

x 19

SECONDARY DATA (PALLARÉS ET AL., 2016)

x 14

RESULTS:

1) Curve fitting:

-VT1: better fit of NO breakpoint for Ramp(Own) and Test(Pallarés).
better fit of breakpoint for Graded (Own).
-VT2: no significant differences in the fit model.

2) Reliability - Test / Re-Test:

-Poor at VT1 and moderate at VT2.
-Large variability Bland Altman plot.

3) Impact of protocol:

-Different protocols (Ramp / Graded) showed significant differences at VT2.

CONCLUSION:

-No evidence of the existence of clear breakpoints in the ventilatory data.
-Caution should be taken in organizing training zones featured by distinct VTs.
-Test protocol may have an impact on VT2 determination.

Master of Science in Physical Activity and Health
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science (INB)

Luca Francesco Masieri

NTNU | Kunnskap for en bedre verden

ABSTRACT

Background:
Intensity zones are commonly used to individualize training programs. These zones are often based on ventilatory thresholds (VTs) retrieved from graded exercise testing (GXT) protocols. The VT1 separates low and moderate intensity, and the VT2 moderate from high intensity. The methods used to detect the VT1 and VT2, a priori, assume the presence of a distinct "breakpoint", where the physiological response changes markedly – however, it is not clear whether these breakpoints actually exist.

Objectives:
The objectives were: (1) compare the fit of a continuous (no-breakpoint) function with a discontinuous (breakpoint) function, used to identify VTs, (2) evaluate the test-retest reliability of thresholds identified during repeated GXT, (3) appraise the relative impact that two different GXT protocol have on the determination of VTs.

Methods:
For primary data, n=19 participants, performed two different GXT (i.e., $RAMP_{Own}$, $GRADED_{Own}$). For secondary data, retrieved from a published study, (Pallarés et al., 2016), n=14 participants performed two identical GXT (RAMP) protocols, (i.e., $TEST_{Pallarés}$, $RETEST_{Pallarés}$). A function with two regression lines were fitted to the oxygen uptake ($\dot{V}O_2$)-carbon dioxide production ($\dot{V}CO_2$) data, to identify VT1, and to the $\dot{V}CO_2$-minute ventilation ($\dot{V}E$) to identify VT2. The same data were also fitted by an exponential function. Wilcoxon signed rank tests were used to compare the R^2 outcome parameters of these functions. ICC analysis for test-retest reliability was performed. In addition, the Bland–Altman analysis assessed the differences between $TEST_{Pallarés}$ -$RETEST_{Pallarés}$ and $RAMP_{Own}$-$GRADED_{Own}$ at VT1 and VT2.

Results:
At VT1, only for $GRADED_{Own}$ a significantly better fit for the discontinuous function (breakpoint) ($p < 0.01$) was found. However, these differences are relatively small. At VT2 no significant difference in fit between the continuous function (no breakpoint) and the discontinuous function (breakpoint). A comparison of the test-retest reliability ($TEST_{Pallarés}$, $RETEST_{Pallarés}$) revealed no significant difference ($p > 0.05$) of % $\dot{V}O_{2peak}$ at VT1 and VT2, although ICC was poor (0.43) and moderate (0.71) at VT1 and VT2 respectively. Data derived from different protocols (i.e., $RAMP_{Own}$, $GRADED_{Own}$) shown statistically significant difference ($p = 0.007$) at VT2.

Conclusions:
A continuous function describes at least equally well the relationship among respiratory variable retrieved during a GXT on cycle ergometer. Thus, our study does not provide evidence for the existence of clear breakpoints in the ventilatory data. Test-retest reliability determination of assumed VT1 and VT2 was low to moderate in the RAMP protocol. The variability at the VT2 suggests that different outcomes can be found when alternative protocols are employed.

Keywords: ventilatory threshold, gas exchange, breakpoint, test-retest reliability

Table of Contents

List of Tables

List of Figures

FREQUENTLY USED ABBREVIATIONS

AN	aerobic threshold
ANT	anaerobic threshold
GET	gas exchange threshold
GRADED	multiple abrupt stepwise increase of constant workload
GXT	graded exercise testing
HR	heart rate
ICC	intraclass correlation coefficient
La-	lactate
RAMP	continuous fashion increase of workload
RCP	respiratory compensation point
RPE	rate of perceived exertion
RPM	revolution per minute
$\dot{V}CO_2$	volume (absolute) of carbon dioxide production
$\dot{V}E$	minute ventilation (absolute)
VT	ventilatory threshold
VT1	ventilatory threshold 1
VT2	ventilatory threshold 2
$\dot{V}O_2$	volume (absolute) of oxygen consumption
W	watt

Introduction

Numerous descriptive studies use intensity zones in an attempt to create a common scale for training intensity (Seiler, 2010). Intensity zones are often related to planning, changing, and evaluating personalized (training) protocol; beside predict, to some degree, endurance performance (e.g., Lucía et al., 2000; Pallarés et al., 2016; Yeh et al., 1983). These intensities are based on the increase in blood lactate (La-) and the subsequent alteration that takes place into respiratory parameters at a certain workload (e.g., Myers & Ashley, 1997; Poole et al., 2021). Different techniques for the identification of these zones have been developed, including the use of power output (Jones & Vanhatalo, 2017), blood La- (Faude et al., 2009; Sales et al., 2019), and/or respiratory data (Beaver et al., 1985; Wasserman et al., 1973). All these methods, a priori, assume the presence of a distinct "breakpoint", where the physiological response shows a "continuous, kinked response" and "a clear point of abrupt increase in slope" (hereafter referred to as "*discontinuous"* response). These physiological processes are claimed to result in clear-cut changes in the oxygen consumption ($\dot{V}O_2$)- carbon dioxide production ($\dot{V}CO_2$) relationship and $\dot{V}CO_2$- minute ventilation ($\dot{V}E$) relationship when exercise intensity is increased. These clear-cut changes identify the first (VT1) and second (VT2) ventilatory threshold, respectively. Various techniques have been utilized to assess these stages, including ($\dot{V}E$) or gas exchange. This thesis will focus on the widely used methods, namely the V-slope for the determination of the gas exchange threshold (GET) at VT1 and the respiratory compensation point (RCP) at VT2 (Beaver et al., 1986; Reinhard et al., 1979).

More precisely, for low exercise intensity zone below the VT1, $\dot{V}O_2$ reaches, within few minutes, a steady-state (Faude et al., 2009). In contrast, the moderate exercise intensity zone, above the VT1 but below the VT2, shows an additional, "slow component of the $\dot{V}O_2$ kinetics". The latter, most likely featured by the consequent recruitment of type II muscle fibers and increased cost of $\dot{V}E$, postpones the achievement of a metabolic steady-state of about 20 min (Pettitt et al., 2013). Conversely, the high exercise intensity zone, above VT2, it is characterized by the impossibility to maintain metabolic equilibrium at a sustained work rate (Galán-Rioja et al., 2020).

Although presented extensively in the literature available, the existence of these “distinct thresholds”, remains controversial (e.g., Morton, 1989; Myers et al., 1994). Theories illustrating breakpoints, have been challenged, in favor of a more smooth curvilinearly changing ventilatory with increasing intensity (Dennis et al., 1992; Hughson et al., 1987). Notwithstanding, critiques arose about these studies (including a low number of participants, exceedingly exhausting protocol and rate of data collection) leaving the controversy unresolved (Myers & Ashley, 1997). In view of the existing dispute, further assessment is needed to shed light on this issue. Hence, the first aim of this study was to evaluate experimental data on the support or rejection of the existence of distinct breakpoints in the gas exchange profile. This was done by analyzing the statistical outcome of the curve fit of the ventilatory data according two different mathematical function (defined hereafter as “*function*”). More precisely, a continuous exponential function was compared with a discontinuous function consisting of two-line segments. A better fit of a continuous function in comparison with the discontinuous function could dispute the presence of clear breakpoints (Baumgart et al., 2018).

Additionally, assuming the presence of these thresholds, a factor that limits the practicality is, together with (poor) test-retest reliability, the difficulty associated with its assertation (Ekkekakis et al., 2008). An important feature of the determination of the ventilatory thresholds (VTs) is its reproducibility, influenced by elements such as within-participant biological variations and methodological inconsistency (Barron et al., 2014). If the presence of breakpoints is questioned, i.e., the ventilatory data determination of such breakpoints are likely close to smooth and continuous, this mathematical determination may be extra sensitive for noise or biological variation and the reliability question is of high importance.

The assessment of dynamic association among exercise workload and cardiorespiratory fitness includes linear increase in intensity over time. Protocols commonly used include “RAMP”, characterized by the smooth continuous increase of workload or “GRADED”, where the increase is abrupt and stepwise (Beltz et al., 2016).

Accordingly, the second aim of this thesis was to evaluate the reliability of the graded exercise testing (GXT) protocol (in our case RAMP). Additionally, for the third aim, i.e., to appraise the impact of test protocol on the VTs outcome, two different GXT protocols (i.e., RAMP versus GRADED) were used.

Materials and Methods

Overall design

The present study included two different samples of participants and relative datasets in order to examine the existence of VT1 and VT2. A primary data set was collected for this purpose on 19 participants in two incremental cycling protocols ($RAMP_{Own}$ and $GRADED_{Own}$), and a secondary data set ($TEST_{Pallarés}$ and $RETEST_{Pallarés}$) was retrieved from a published study including 14 subjects performing twice the same RAMP cycling test (Pallarés et al., 2016).

Primary data collection

Participants

Fifteen men and four women participated in this study (i.e., primary collection) for the "Own" data collection. Standard procedures were used to measure height and body mass at first meet in the laboratory, and the anthropometrics of the participants are depicted in Table 1. Mean $\dot{V}O_2$ max was 53.4 ± 10.5 $ml \cdot kg^{-1} \cdot min^{-1}$ (based on the highest value for each participant, independent of which test, RAMP or GRADED). All participants, ranging from sedentary to well-trained, not specifically in cycling, were in good health and free of injuries during the testing period although the exact training status and overall amount of physical activity was not known. This data collection and storage protocol were approved by the Norwegian Centre for Research Data (Org-ID 10020356) and performed in agreement with the Declaration of Helsinki. All participants received and signed a consent form before joining in the trial. Additionally, withdrawal from the study, without presenting a justification at any time, was allowed. Data were collected between September 2022 and November 2022.

Table 1.Anthropometric of the 19 participants. (Primary "Own" data collection).

Participant	Sex (M/F)	Age (Years)	Height(cm)	Body Mass (Kg)	$\dot{V}O_2$ Max ($ml \cdot kg^{-1} \cdot min^{-1}$)
1	M	41.4	185	84.7	41.3
2	M	36.2	182	78.8	37.6
3	M	25.1	188	73.5	62.7
4	M	29.1	195	87.3	52.4
5	F	25.3	181	61.8	59.8
6	M	23.7	183	76	47.9
7	M	25.7	178	65.3	46.1
8	F	36.3	165	66.2	46.9
9	M	23.2	179	72.5	66.7
10	M	24.3	187	80	72.7
11	M	26.5	182	84.4	54.4
12	M	20.1	184	73	49.2
13	F	23.4	172	64.7	42.4
14	F	22.0	162	58.9	41.3
15	M	25.8	178	73.6	65.2
16	M	22.9	170	63.1	64.0
17	M	22.6	169	74.3	58.5
18	M	28.7	183	69.2	63.1
19	M	59.5	174	87.3	53.4
Mean ± SD	15M/4F	28.5 ± 9.3	178.8 ± 8.4	73.4 ± 8.8	53.4±10.5

Test set-up

The testing, for the primary data (Own) collection, consisted of two different GXT to volitional exhaustion, on cycle ergometer. Test were performed in two different days under similar environmental conditions, scheduled at a similar time of the day (± 3 hour), to control the circadian rhythms effect (Teo et al., 2011). In order to minimize the effect of different levels of fitness among participants, at least 24 hours of full recovery between each test was given. Both GXT protocols, $RAMP_{Own}$ and $GRADED_{Own}$, were performed, in this order for each participant, at the Centre for Elite Sports Research (SenTIF) in Granåsen (Trondheim, Norway) in the laboratory NeXtMove. Tests were conducted on the LODE Excalibur sport cycle ergometer (Lode BV, Groningen, Netherlands).

Each participant was instructed to set the cycle ergometer to their preferred settings. These settings were recorded and used for both tests (i.e., $RAMP_{Own}$ and $GRADED_{Own}$). Each test included a short warm-up at the preferred intensity. Test $RAMP_{Own}$ began at an intensity of 10 Watt (W) followed by a continuous increase of 1W every 3 seconds (20 W·min). In order to perform at least 4 incremental stages, participants were assigned to the starting power in $GRADED_{Own}$ based on the results of the $RAMP_{Own}$ (Plato et al., 2008).

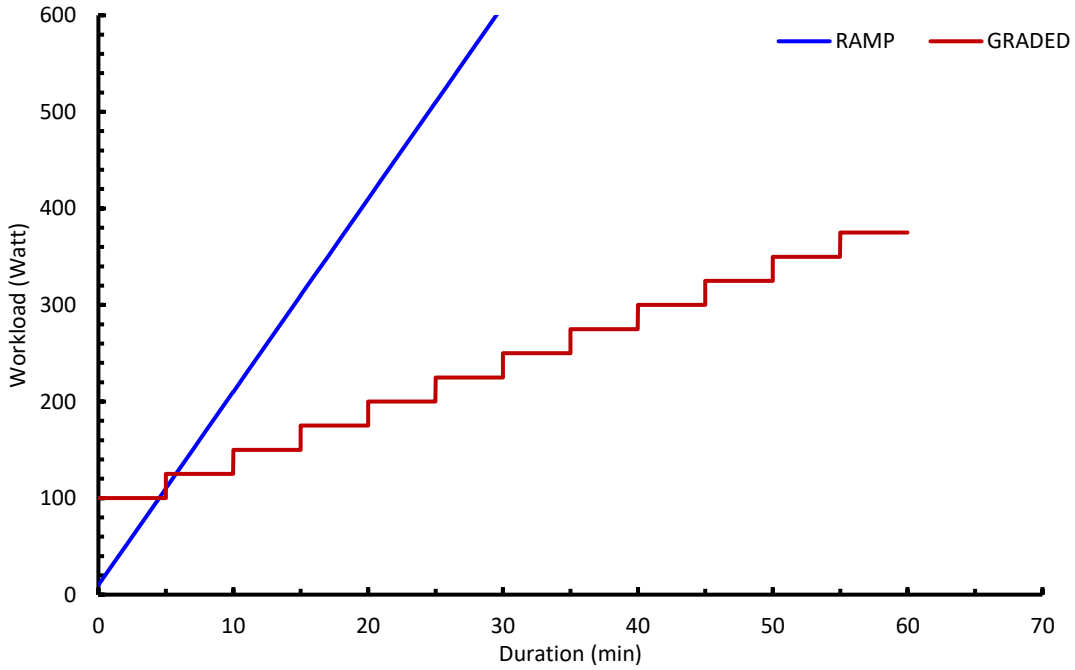


Figure 1.Two test protocols of the primary data collection: $RAMP_{Own}$ (Blue line) and $GRADED_{Own}$ (Red line). Note that 15 participants started $GRADED_{Own}$ at 100 W and 4 participants at 75 W.

Therefore, the order of tests was set and not done randomly. Test $GRADED_{Own}$ consisted of four to eleven 5-min abrupt stepwise stages at increasing effort of 25 W for each stage; with the initial power at either 75 W or 100 W (Fig.1). Participants were instructed to keep the pedal rate in revolution per minute (RPM) as constant as possible throughout testing. Verbal encouragement was given throughout both tests to promote maximal physical exertion.

For the assessment that physiologically maximal effort was reached ($\dot{V}O_2$ max), we followed the criteria set out by Gellish et al. (2007). Four out of five criteria were to be met: 1) Respiratory exchange ratio (RER) of ≥1.10; 2) a plateau in $\dot{V}O_2$ against incremental resistance; 3) observed maximal limit reached, with a fall of 10 (RPM); 4) attainment of highest heart rate (HR) within 10 beats of the age-predicted maximal HR [208–(0.7 × age)](Tanaka et al., 2001) ; and 5) rate of perceived exertion (RPE) equal or higher than 19. At the end of each test the ergometer was unloaded, and the participant continued to cycle for recovery.

Test equipment

Respiratory data were measured with an open-circuit indirect calorimetry apparatus (Vyntus CPX, Vyaire medical GmBH Hoechberg, Germany), and respiratory parameters (i.e., $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}E$) were measured breath-by-breath and averaged over 10 seconds by the in-built software. $\dot{V}O_2$ and $\dot{V}CO_2$ gas were calibrated using known gases (15% $\dot{V}O_2$, 5% $\dot{V}CO_2$, Riessner-Gase GmbH & Co., Lichtenfels, Germany) and air flow was calibrated using the built-in automatic calibration procedure of the indirect calorimetry apparatus. To ensure accuracy of indirect calorimetry, gas and flow calibration was conducted prior to every exercise test and a “dry” 2.4 m TwinTube sample line was utilized each time.

A 20µl blood sample was collected from the participant’s fingertip. Samples were taken at rest and at the end of both tests ($RAMP_{Own}$ and $GRADED_{Own}$), and only in $GRADED_{Own}$ within the last 30 seconds of each of the submaximal stages (i.e., after each 25 W increments stage). If the participant could not finalize the whole stage, samples were collected directly after the participant gave up. Blood La- was analyzed using, the Biosen C-Line Sport lactate measurement system (EKF Industrial Electronics, Magdeburg, Germany) after a regular calibration.

Similarly, the RPE (Borg scale 6–20) was taken at the beginning and at the end of $RAMP_{Own}$ and $GRADED_{Own}$. Additionally, in $GRADED_{Own}$ during the last 30 seconds of each submaximal (25 W incremented) stage. HR was measured every second with a Polar heart rate monitor (Garmin Forerunner 920XT, Garmin International Inc., Kansas, United States).

Secondary data collection

These data were obtained as part of a previous study on determination of ventilatory threshold in fourteen well trained cyclists, age 26.7 ± 8.2 year, body mass 70.3 ± 4.9 kg, height 173.7 ± 4.2 cm, body fat 12.5 ± 3.0%, mean $\dot{V}O_2$ max was 62.1 ± 4.6 $ml \cdot kg^{-1} \cdot min^{-1}$, endurance training experience 10.9 ± 4.9 year (Pallarés et al., 2016). Tests were performed on a cycle ergometer (Ergoselect 200, Ergoline, Germany). This study was done during the period from January 2014 to July 2015. After a standardized warm-up of 10 min at 50 W, all participants performed two identical ramp protocols with increments of 25 W·min until exhaustion, ($TEST_{Pallarés}$-$RETEST_{Pallarés}$). Each test was separated by two-five days.

Data Processing

The breadth-by-breadth ventilatory data were resampled (by the Vyntus software) and retrieved at 0.1 Hz. Following these procedures, it was possible to obtain printouts every 10 s of $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}E$, and their relationship $\dot{V}O_2/\dot{V}CO_2$ and $\dot{V}CO_2/\dot{V}E$.

Data preparation involved three steps on the three outcome parameters (i.e., $\dot{V}O_2$, $\dot{V}CO_2$ and $\dot{V}E$). First, non-realistic physiological (e.g., equipment error or alike) values were removed. Then, lower and upper boundaries for the ventilatory threshold calculations were set. The lower boundary was set at the end of the warm-up (i.e., minute 1 for $RAMP_{Own}$ and $GRADED_{Own}$) at the beginning of the test. The upper boundary was set at the $\dot{V}O_2$ peak. Outcome parameters (e.g. $\dot{V}O_2$, % of $\dot{V}O_2$ peak) were interpolated at the thresholds identified with each of the V-slope GET methods and RCP used to determine the VT1 and VT2.

For aim 1, we compared the fit of discontinuous versus continuous function. Two regression lines (Eq 1) and an exponential curve (Eq 2) were fitted to the $\dot{V}O_2/\dot{V}CO_2$ (VT1) and $\dot{V}CO_2/\dot{V}E$ (VT2) gas exchange data (by linear least squares fitting).

The function slmengine of the Shape Language Modeling (SLM) toolkit (D'Errico, 2023) was employed to fit two regression lines to the data, including the function *fmincon* with the interior-point algorithm to detect the best fitting model. A custom-made function was edit to fit the exponential curve to the data, and the function *fminsearch* using the Nelder-Mead approach (Nelder & Mead, 1965)was utilized to ascertain the best fit including a maximum of 40,000 evaluations and 1,500 iterations (Baumgart et al., 2021).

$$y = \begin{cases} a_1 + b_1\, x, x < k \\ a_2 + b_2\, x, x \geq k \end{cases} \quad (1)$$

$$y = a + b * \exp(c + d * x) + e * x \quad (2)$$

y and x are the variables of interest, a, b, c, d, e, and k are fitting constants, of which k identifies the x value at the breakpoint, b_1 and b_2 the slopes of the corresponding line segments.

For aim 1 an adjusted R^2 value was obtained for the fit of each model (i.e., discontinuous function, Eq 1, versus continuous function, Eq 2) to the ventilatory variable pairs of interest.

Statistical analyses

Standard statistical analysis was utilized for the calculation of means and standard deviation (SD) of the descriptive ventilatory data, and the median and 95% confidence interval (CI) of the adjusted R^2 values. The difference between the adjusted R^2 values, and thus fit of the discontinuous function (breakpoint) compared with the continuous function (no-breakpoint), were analyzed by Wilcoxon signed rank test to determine whether different methods yielded significantly different estimates. An alpha level of 0.05 was used to indicate statistical significance. Our rationale was that a better fit of the continuous function as compared with the two lines segment (discontinuous function), would challenge the presence of a distinct breakpoint.

For aim 2 and 3, assuming that breakpoints exist, The Bland–Altman analysis (Giavarina, 2015) assessed the differences in of $\dot{V}O_2$ between $TEST_{Pallarés}$-$RETEST_{Pallarés}$ and $RAMP_{Own}$-$GRADED_{Own}$ at VT1 and VT2. In addition, for aim 2, the test-retest reliability was also assessed by the intraclass correlation coefficient (ICC) estimates with 95% CI (absolute-agreement, 2-way mixed-effects model). Size of the ICC was evaluated as follows; below 0.5 indicate poor reliability, between 0.5 and 0.75 moderate reliability, between 0.75 and 0.9 good reliability, and any value above 0.9 indicates excellent reliability (Koo & Li, 2016). Data were analyzed and figures generated using SPSS software version19.0 (SPSS, Chicago, IL)/ Matlab R2020a (The MathWorks, Inc., Natick, MA, United States).

Results

Comparison of breakpoint versus no breakpoint

Fig.2 shows an example of the individual respiratory value at VT1 ($\dot{V}O_2/\dot{V}CO_2$), for the participants n.14 and n.19, retrieved during $GRADED_{Own}$. The relative continuous (green line) and discontinuous (red line) model fitting are reported. Participant n.14 showed the biggest difference for R^2adjusdted among the two methods (R^2_{diff}=0.0066), while participant n.19 displayed the highest value for R^2adjusted fit in discontinuous function (breakpoint), within the whole group (R^2=0.993). (All individual data provided in Supplementary Excel File).

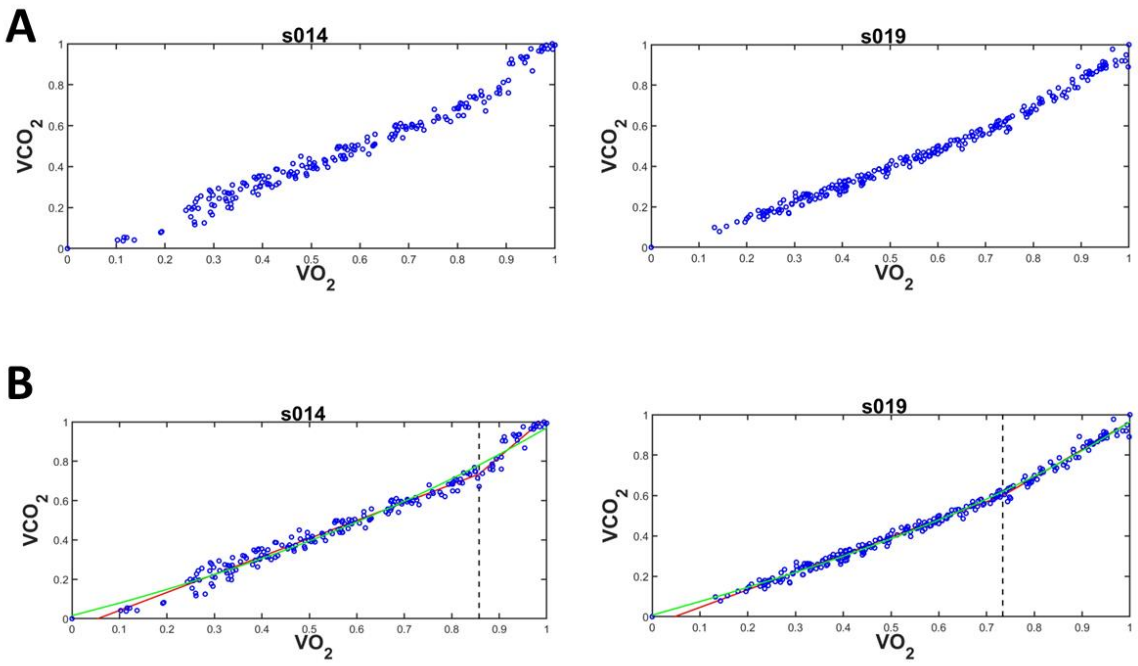


Figure 2. Illustration of ventilatory gas exchange, without (A) and with (B), relative model fitting for participant n.14 and n.19 during $GRADED_{Own}$ at VT1, with $\dot{V}O_2$ uptake vs $\dot{V}CO_2$ production. In each panel with model fitting (B), red line indicates discontinuous function (V-slope) and green line continuous exponential function.

At the VT1, the adjusted R^2 medians ±95% CI interval ranges show an overall better fit of the continuous function (no breakpoint) compared to the discontinuous function (two lines segments - breakpoint) for $TEST_{Pallarés}$ ($p = 0.041$) and for $RAMP_{Own}$ ($p = 0.011$). Conversely, at the VT1 for $GRADED_{Own}$, there was a better fit ($p < 0.01$) of the discontinuous function (breakpoint) compared to the continuous function (no breakpoint) as shown in Fig. 3. At the VT2, there were no significant differences in fit between continuous function as compared with the discontinuous function, all comparisons, ($p > 0.05$), as shown in Fig.4.

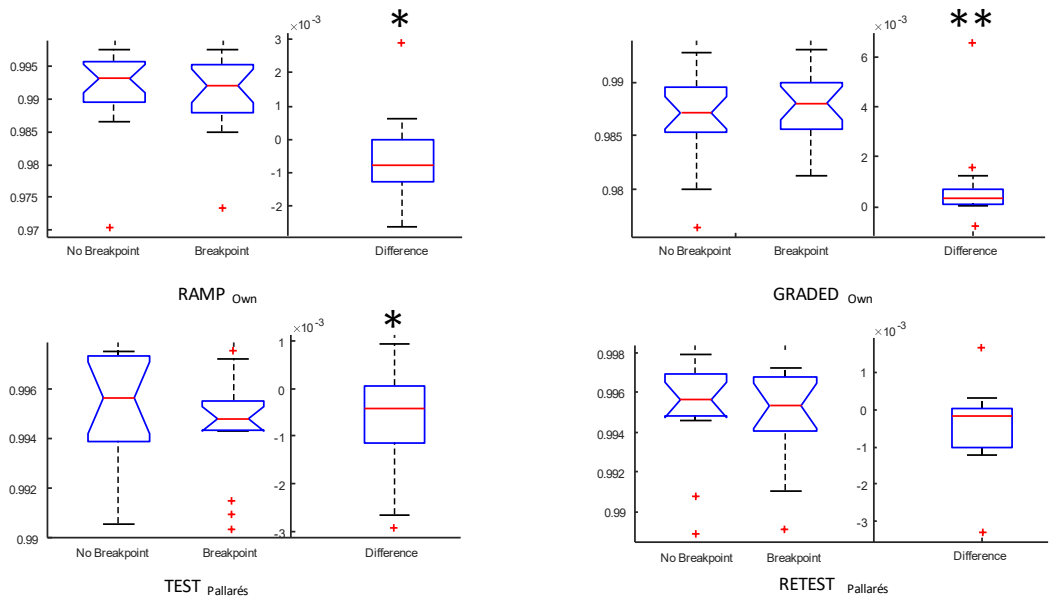


Figure 3. Boxplots of the VT1 fitting outcome.
Notched boxplots are presented to compare adjusted R^2values between the breakpoint and the no-breakpoint model for VT1 ($\dot{V}O_2$ vs $\dot{V}CO_2$). The notches are 95% CIs that are constructed around the median (red line).
If p-value is less than 0.05, it is flagged with one star (*). If p-value is less than 0.01, it is flagged with two stars (**).

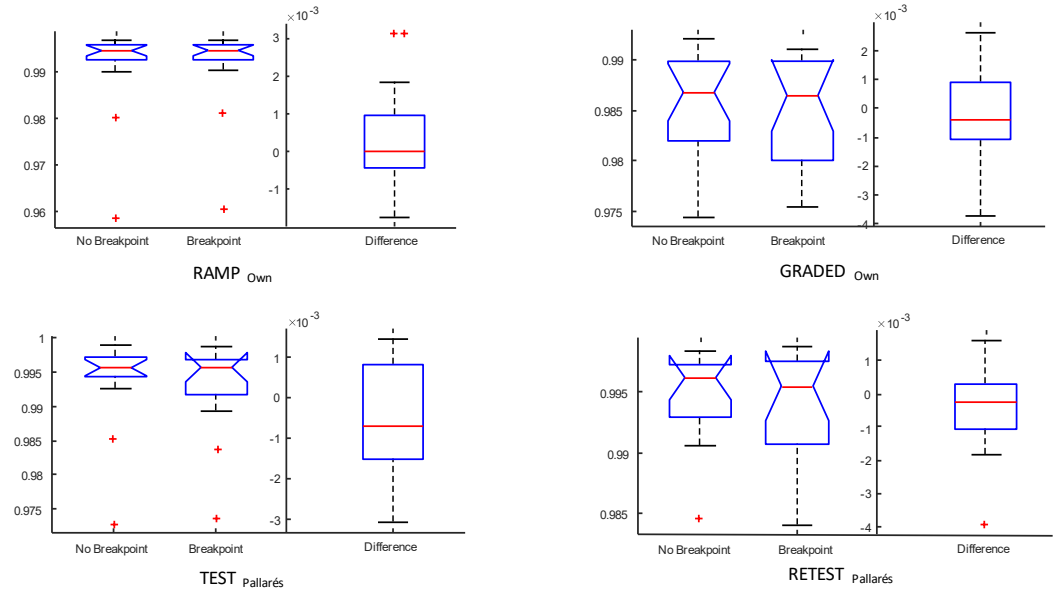


Figure 4. Boxplots of the VT2 fitting outcome.
Notched boxplots are presented to compare adjusted R^2values between the breakpoint and the no-breakpoint model for VT2 ($\dot{V}CO_2$ vs $\dot{V}E$). The notches are 95% CIs that are constructed around the median (red line).
If p-value is less than 0.05, it is flagged with one star (*). If p-value is less than 0.01, it is flagged with two stars (**).

Reliability of determination of VTs

In the comparison of two tests using the identical GXT protocol (i.e., RAMP 25 W·min), both VT1 and VT2 retrieved with two lines segment method (i.e., discontinuous), did not show statistically significant difference in the $TEST_{Pallarés}$- $RETEST_{Pallarés}$ respectively (all comparisons, p>0.05). More precisely as shown in Table 2 the $\dot{V}O_2$ at the determination of each threshold was: for VT1: 69% and 73% of $\dot{V}O_{2peak}$ for $TEST_{Pallarés}$-$RETEST_{Pallarés}$ respectively, and for VT2: 86% and 87% of $\dot{V}O_{2peak}$ for $TEST_{Pallarés}$-$RETEST_{Pallarés}$ respectively. However, ICC was poor at VT1 (0.43) and moderate at VT2 (0.71) as shown in Table 3. Additionally, for one participant in $TEST_{Pallarés}$ VT2 occurred before VT1.

Table 2. Mean ± standard deviation (SD) for oxygen uptake at ventilatory threshold 1 (VT1), ventilatory threshold 2 (VT2) and peak oxygen uptake ($\dot{V}O_{2peak}$) in L min^{-1}, for both $TEST_{Pallarés}$ and $RETEST_{Pallarés}$.

	VT1			VT2			$\dot{V}O_{2peak}$		
	$TEST_{Pallarés}$	$RETEST_{Pallarés}$	P value	$TEST_{Pallarés}$	$RETEST_{Pallarés}$	P value	$TEST_{Pallarés}$	$RETEST_{Pallarés}$	P value
Mean	2.74±0.35	2.87±0.31	0.23	3.31±0.40	3.36±0.32	0.56	3.94±0.28	3.95±0.33	0.84

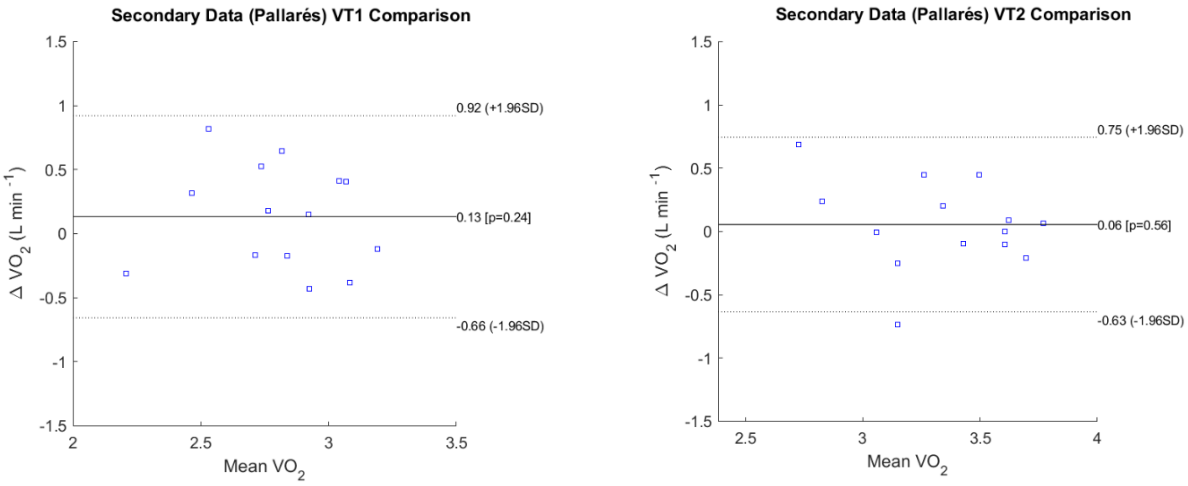


Figure 5. Bland-Altman plots for the VT1 and VT2 differences (L min^{-1}) between TEST and RETEST data of the secondary data (Pallarés).

Table 3.Test-retest reliability measurements of $TEST_{Pallarés}$-$RETEST_{Pallarés}$.

	Overall mean ($TEST_{Pallarés}$-$RETEST_{Pallarés}$)	MD (95% limits of agreement)	SDD	ICC (95% CI)
VT1 $\dot{V}O_2$ (L min^{-1})	2.80	0.13(-0.65, 0.92)	0.4	0.43(-0.64, 0.81)
VT2 $\dot{V}O_2$ (L min^{-1})	3.33	0.05(-0.63, 0.74)	0.35	0.71(0.09,0.90)

Units are defined within the supplementary data.; ICC, intraclass correlation coefficient; MD, mean difference; SDD, standard deviation of the difference; $\dot{V}O_2$, oxygen uptake (L min^{-1}).

Comparison of $\dot{V}O_2$ at breakpoint between protocols

In the comparison of two different GXT protocols (i.e., RAMP, GRADED); at VT1 no statistically significant level was found ($p > 0.05$), and the $\dot{V}O_2$ at the determination of each threshold was 65% and 72% of $\dot{V}O_{2peak}$ for $RAMP_{Own}$-$GRADED_{Own}$ respectively. Conversely at VT2, a statistically significant difference ($p = 0.007$) was found for the $\dot{V}O_2$, at 81% and 77% of $\dot{V}O_{2peak}$ for $RAMP_{Own}$-$GRADED_{Own}$ respectively, as shown in Table 4. Please note that unlike for aim 2 no ICC is provided here because different protocols are compared. Additionally, for one participant in $GRADED_{Own}$ VT2 occurred before VT1.

Table 4.Mean ± standard deviation (SD) for oxygen uptake at ventilatory threshold 1 (VT1), ventilatory threshold 2 (VT2) and peak oxygen uptake ($\dot{V}O_{2peak}$) in L min^{-1} for both $RAMP_{Own}$ and $GRADED_{Own}$.

	VT1			VT2			$\dot{V}O_{2peak}$		
	$RAMP_{Own}$	$GRADED_{Own}$	P value	$RAMP_{Own}$	$GRADED_{Own}$	P value	$RAMP_{Own}$	$GRADED_{Own}$	P value
Mean	2.51±0.67	2.66±0.55	0.14	3.12±0.69	2.93±0.59	0.007	3.85±0.93	3.69±0.69	0.052

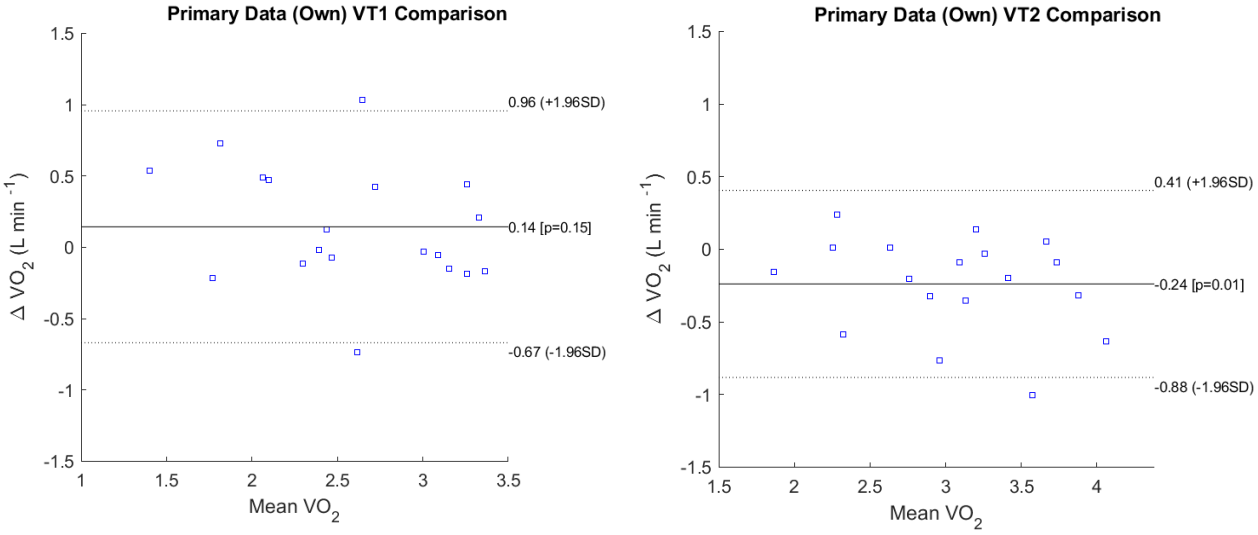


Figure 6. Bland-Altman plots for the VT1 and VT2 differences (L min^{-1}) between RAMP and GRADED data of the primary data (Own).

Discussion

Is there evidence for the existence of ventilatory thresholds?

The first aim of this study was to compare the statistical outcome of the curve fit of discontinuous (breakpoint) versus continuous function (no-breakpoint), relative to respiratory data retrieved during a GXT test, irrespective of the test protocol utilized (i.e., RAMP or GRADED). For all the determined VT1 and VT2, in all tests, there were overlaps in the adjusted R^2 medians ± 95% CI between the statistical outcome of the curve fit of discontinuous versus continuous function. Moreover, apart from VT2 in $RAMP_{Own}$, the average of the fits of discontinuous function (breakpoint) was never better than the continuous ones (no-breakpoint).

In light of the data collected and analyzed, only at VT1 for $GRADED_{Own}$, the discontinuous function (breakpoint) appeared to fit better the gas exchange profile on overall group level ($p < 0.01$). This evidence supports previous findings (Beaver et al., 1986; Caiozzo et al., 1982; Davis et al., 1976; Galán-Rioja et al., 2020; Kim et al., 2021; Lucía et al., 2000; Pallarés et al., 2016; Reinhard et al., 1979; Wasserman et al., 1973).

In contrast, for all the other approaches investigated, the discontinuous function did not better describe the association of the data. Hence, the evidence for the presence of a defined turn-over VT1 and VT2 breakpoints seems slim. Accordingly, similar findings were found in wheelchair athletes (Baumgart et al., 2021) and in male endurance-trained athletes (Dennis et al., 1992) , where: *"…respiratory changes in response to a gradual increase in exercise intensity are curvilinear..."*. Overall, these results suggest that discontinuous function and relative breakpoint, although widely employed, might not be the most appropriate method to describe the relationship of these dynamic human physiological responses during GXT. Furthermore, as shown in this study by the statistical analysis, the minimal differences of fit between these two functions could be extremely subtle. In line with this, it is noteworthy to underline the large presence of 95% CI overlaps among the functions employed at VT1 for $GRADED_{Own}$ (Fig.3) even where statistical difference was present ($p < 0.01$). A small but systematic improvement is present in this case, which might be caused by small but

systematic error in the data. A larger fit improvement would be in fact be expected if there is reality in the existence of a clear breakpoint, i.e., two straight x-y data segments with distinct slope differences (Beaver et al., 1986).

Reliability and protocol effects on VT1 and VT2, using V-slope method and RCP

The second aim was to evaluate the reliability for determination of VTs, derived by discontinuous function (breakpoint), in two repeated GXT RAMP protocols, (i.e., $TEST_{Pallarés}$, $RETEST_{Pallarés}$). In the current study we reported VT1 at 69% and 73% of $\dot{V}O_{2peak}$ and VT2 at 86% and 87% of $\dot{V}O_{2peak}$ for $TEST_{Pallarés}$-$RETEST_{Pallarés}$ respectively. These intensities are overall higher than those available among the literature. More precisely, Beaver et al. (1986) reported that VT1 occurred at 55% and VT2 at 75% of peak $\dot{V}O_2$, Kim et al. (2021) ascertained VT1 and VT2 at 58% and 79% of peak $\dot{V}O_2$ respectively, while a recent meta-analysis (Galán-Rioja et al., 2020), registered that VT1 occurs between 50% to 60% of peak $\dot{V}O_2$. These findings might underline a large range of variability, with potential impact on the effective meaningfulness for identification and implementation of intensity zones.

Furthermore, a comparison of the test-retest reliability ($TEST_{Pallarés,}$ $RETEST_{Pallarés}$) revealed no significant difference ($p > 0.05$) between TEST and RETEST outcome for both VTs. Notwithstanding, ICC displayed poor (0.43) and moderate (0.71) at VT1 and VT2 respectively. Additionally, the employment of Bland–Altman plots, consent a visual qualitative assessment of the reproducibility of a variable. In particular the difference of the two paired measurements (e.g., $TEST_{Pallarés,}$ $RETEST_{Pallarés}$) is plotted against the average of these two measurements for each pair of individuals (Barron et al., 2014). This methodological representation describes only the range of agreements, without defining if those limits are appropriate or not (Giavarina, 2015). Results retrieved from this study shown an overall uniform normality of differences around the mean for both primary (Fig.6) and secondary (Fig.5) data.

Furthermore, the third aim of this study was to quantify difference in $\dot{V}O_2$ at VT1 and VT2 resulting from alternative protocol (i.e., $RAMP_{Own}$, $GRADED_{Own}$). It is interesting to note that the present trial has revealed how diverse protocols ($RAMP_{Own}$, $GRADED_{Own}$) can yield statistically significant ($p = 0.007$) different results in determination of VT2, detected by two regression lines (showing discrepancies as high as 0.22 L min^{-1} or 4 % $\dot{V}O_2peak$).

Inconsistent evidence has been published concerning the possibility that the VTs may be affected by the rate of increase in work rate. Hughson et al. (1987) reported that slow RAMP GXT yielded lower threshold values than those found with fast increment. Conversely, Wasserman et al. (1973) did not find difference in outcome implementing testing with 1 and 4 min increment durations (the increment size was 25 W·min for both tests). The (RAMP) protocols utilized in this thesis (i.e., $TEST_{Pallarés}$, $RETEST_{Pallarés}$ 25 W·min, and $RAMP_{Own}$ 20 W·min) were similar to those described by other studies such as Caiozzo et al. (1982); Yeh et al. (1983) with 20 W·min increment, Davis et al. (1976) with 32,7 W·min and Reinhard et al. (1979) with 16,4 W·min.

It might be noted that the comparison of two identical RAMP protocols yields better agreement for determination of VT1 and VT2. Intuitively a RAMP test, by avoiding sudden increments in work rate between stages, may reduce dramatic alteration in gas exchange parameters and affect the ability to determine the VTs (Shimizu et al., 1991). At the present time, no method can legitimately be acknowledged as ‘‘gold standard’’ for determination of VTs. However, these variations might raise concerns for both research and clinical practice(Ekkekakis et al., 2008). In fact, the absolute and relative $\dot{V}O_2$ at which VTs occurs is also extremely useful in evaluating the normalcy or otherwise of an individual’s response to the stress of exercise. More precisely, VTs have construct validity, and is sensitive and specific to interventions such as, cardiac resynchronization therapy or manipulations of systemic O_2 delivery. Among the others, VTs are employed in assessing physiological function in chronic disease states and categorizing ‘fitness’ to undergo related surgery (Poole et al., 2021).

General controversy about VT: No discussion in “isolation”

The use of a “threshold behavior form”, as determination of the aerobic (AT) and anaerobic (ANT) threshold, has generated an important debate within the field of exercise physiology, throughout the recent years. This is partly because the curve fitting is phenomenological. In other words, most often the outcome (the phenomenon) of a physiological process is fitted, rather than the relative system and/or processes are modeled, and its outcome verified (Morton, 1989). Furthermore, a system containing discontinuous processes, may result in a continuous or discontinuous outcome, while a system containing only continuous processes never results in a discontinuous outcome (Morton, 1989). When modeling the outcome by

curve fitting and not modeling the, (although meaningful), biological processes underlying it, some of the issues are purely academic or unresolvable. The current thesis can therefore not elucidate if the system contains discontinuous processes, and it is limited to investigating the evaluation of the process outcome. The VT is "…*theoretical concept, and that definition is a conceptual definition…*" (Svedahl & MacIntosh, 2003). What Wasserman and associates originally intended, was that these variables showed a continuous, kinked response "…*a clear point of abrupt increase in slope*…", and it was this interpretation which was, and still is disputed (Myers & Ashley, 1997).

Methodological consideration

Choosing to employ a cycle ergometer over treadmill offers an opportunity to utilize progressive ramp protocol with increased reproducibility and retrieve more computable workload data (W). Nevertheless, lower recruitment of skeletal muscle activity and consequent reduced peak $\dot{V}O_2$ are attained using protocols on bike compared to treadmill. Therefore, there is a relative decreased metabolic acidosis at submaximal effort. In addition, also the modification of the stage duration during an incremental exercise test may influence the submaximal and maximal physiological variables (Beltz et al., 2016). Furthermore, single, regression-based computer algorithms, such as those incorporated in commercial metabolic analysis software programs, should be viewed as useful aides but not capable of providing ''automatic'' or "definitive" solutions for the identification of VTs. The subtle variance found also in this study, especially between identical protocol, suggests that reliance on any single computerized method might be imprudent (Ekkekakis et al., 2008). Moreover, in the current study, we selected two very commonly used analytical methods to identify the VT1 and the VT2 (i.e., V-slope and RCP). There are other techniques to ascertain these points. (Gaskill et al., 2001). Nevertheless, all of these methods a priori assume the presence of a breakpoint, which, according to the current findings, remains debatable (Baumgart et al., 2021). We, like others, (e.g., Ekkekakis et al., 2008; Gaskill et al., 2001) argue for the development of a universally agreed standard protocol and method for determining the exact nature of ventilatory changes during incremental exercise intensity, and points or areas of distinct alteration.

Limitation

With the variety of techniques utilized in assessing the VTs, caution should be taken in interpretation of the results as the different protocol may elicit a variety of responses during incremental exercise. Furthermore, it is essential to account for the individual's unique response to such exercise and the possible systematic error in the data (Loat & Rhodes, 1993). The strengths of the current study include two distinct (fairly) heterogeneous groups of participants, combined with a relatively varied methodological approach. The aims employing three protocols were: 1) to verify if there was a tendency for specific pairs of methods (e.g. RAMP vs RAMP) to generate consistent results and 2) whether there was a predisposition for particular methods to constantly produce lower or higher outcomes in comparison with others methodology (e.g. RAMP vs GRADED) (Ekkekakis et al., 2008). Nevertheless, our study also has several limitations including the following. First, the generalizability of the conclusions might be restricted by the features of the participants, overall young and healthy individual. It may occur that the results could vary for other population groups (Pallarés et al., 2016). Second, the results may be limited to cycle ergometer GXT and the detailed incremental protocols used in this trial. Respiratory data reported throughout different exercise procedure may also generate diverse "signal to noise ratio" and consequent outcomes.

Practical Applications

The result of our study indicates that, attempts to compare ventilatory and gas exchange variables responses in various experimental condition should be viewed with some prudence. The sensitivity of these results highlight variability in $\dot{V}O_2$ relative to VTs determination, with modification of GXT design. Consequently, it may change the validity of using these results for predicting performance and prescribing or monitoring training (Bentley 2007).

Similarly, caution should be taken in organizing training zones featured by distinct ventilatory turn-over points as often stated in the literature. While common and recommended in coaching literature, these intensity areas suggest a degree of physiological specificity that does not really seem to be present, as previously suggested by Baumgart et al. (2021). Sports

scientists and coaches should consider these factors when conducting incremental exercise testing for the purposes of training prescription.

Conclusion

In opposition to the breakpoint hypothesis, our findings suggest that continuous function describes at least equally well the relationship among gas exchange variable retrieved during a GXT on cycle ergometer. Thus, our study does not provide evidence for the existence of clear ventilatory breakpoints. These findings reinforce previous and unresolved debates about: if breakpoint exists, and if so, that they may not be detectable due to insufficient sensitivity of model employed.

Moreover, for test-retest reliability, determination of VT1 and VT2, based on the assumption that clear breakpoint exist, was low to moderate in two (identical) RAMP protocol. In line with this, VTs ascertained with two different protocols (i.e., RAMP vs GRADED) displayed a large difference at VT2 value. Thus, caution should be taken in interpretation of these results, since these inconsistencies may (partly) caused by the uncertainty in the assumption, i.e., that clear breakpoint exists.

References

Barron, A., Dhutia, N., Mayet, J., Hughes, A. D., Francis, D. P., & Wensel, R. (2014). Test-retest repeatability of cardiopulmonary exercise test variables in patients with cardiac or respiratory disease. *Eur J Prev Cardiol*, *21*(4), 445-453.

Baumgart, J., Moes, M., Skovereng, K., Ettema, G., & Sandbakk, O. (2018). Examination of gas exchange and blood lactate thresholds in Paralympic athletes during upper-body poling. *PLOS ONE*, *13*, e0205588.

Baumgart, J. K., Ettema, G., Griggs, K. E., Goosey-Tolfrey, V. L., & Leicht, C. A. (2021). A Reappraisal of Ventilatory Thresholds in Wheelchair Athletes With a Spinal Cord Injury: Do They Really Exist? *Front Physiol*, *12*, 719341.

Beaver, W. L., Wasserman, K., & Whipp, B. J. (1985). Improved detection of lactate threshold during exercise using a log-log transformation. *J Appl Physiol (1985)*, *59*(6), 1936-1940.

Beaver, W. L., Wasserman, K., & Whipp, B. J. (1986). A new method for detecting anaerobic threshold by gas exchange. *J Appl Physiol (1985)*, *60*(6), 2020-2027.

Beltz, N. M., Gibson, A. L., Janot, J. M., Kravitz, L., Mermier, C. M., & Dalleck, L. C. (2016). Graded Exercise Testing Protocols for the Determination of VO(2)max: Historical Perspectives, Progress, and Future Considerations. *J Sports Med (Hindawi Publ Corp)*, *2016*, 3968393.

Caiozzo, V. J., Davis, J. A., Ellis, J. F., Azus, J. L., Vandagriff, R., Prietto, C. A., & McMaster, W. C. (1982). A comparison of gas exchange indices used to detect the anaerobic threshold. *J Appl Physiol Respir Environ Exerc Physiol*, *53*(5), 1184-1189.

D'Errico, J. (2023). SLM - Shape Language Modeling (https://www.mathworks.com/matlabcentral/fileexchange/24443-slm-shape-language-modeling), MATLAB Central File. In.

Davis, J. A., Vodak, P., Wilmore, J. H., Vodak, J., & Kurtz, P. (1976). Anaerobic threshold and maximal aerobic power for three modes of exercise. *J Appl Physiol*, *41*(4), 544-550.

Dennis, S. C., Noakes, T. D., & Bosch, A. N. (1992). Ventilation and blood lactate increase exponentially during incremental exercise. *J Sports Sci*, *10*(5), 437-449.

Ekkekakis, P., Lind, E., Hall, E. E., & Petruzzello, S. J. (2008). Do regression-based computer algorithms for determining the ventilatory threshold agree? *J Sports Sci*, *26*(9), 967-976.

Faude, O., Kindermann, W., & Meyer, T. (2009). Lactate threshold concepts: how valid are they? *Sports Med*, *39*(6), 469-490.

Galán-Rioja, M., González-Mohíno, F., Poole, D. C., & González-Ravé, J. M. (2020). Relative Proximity of Critical Power and Metabolic/Ventilatory Thresholds: Systematic Review and Meta-Analysis. *Sports Med*, *50*(10), 1771-1783.

Gaskill, S. E., Ruby, B. C., Walker, A. J., Sanchez, O. A., Serfass, R. C., & Leon, A. S. (2001). Validity and reliability of combining three methods to determine ventilatory threshold. *Med Sci Sports Exerc*, *33*(11), 1841-1848.

Gellish, R. L., Goslin, B. R., Olson, R. E., McDonald, A., Russi, G. D., & Moudgil, V. K. (2007). Longitudinal modeling of the relationship between age and maximal heart rate. *Med Sci Sports Exerc*, *39*(5), 822-829.

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochem Med (Zagreb)*, *25*(2), 141-151.

Hughson, R. L., Weisiger, K. H., & Swanson, G. D. (1987). Blood lactate concentration increases as a continuous function in progressive exercise. *J Appl Physiol (1985)*, *62*(5), 1975-1981.

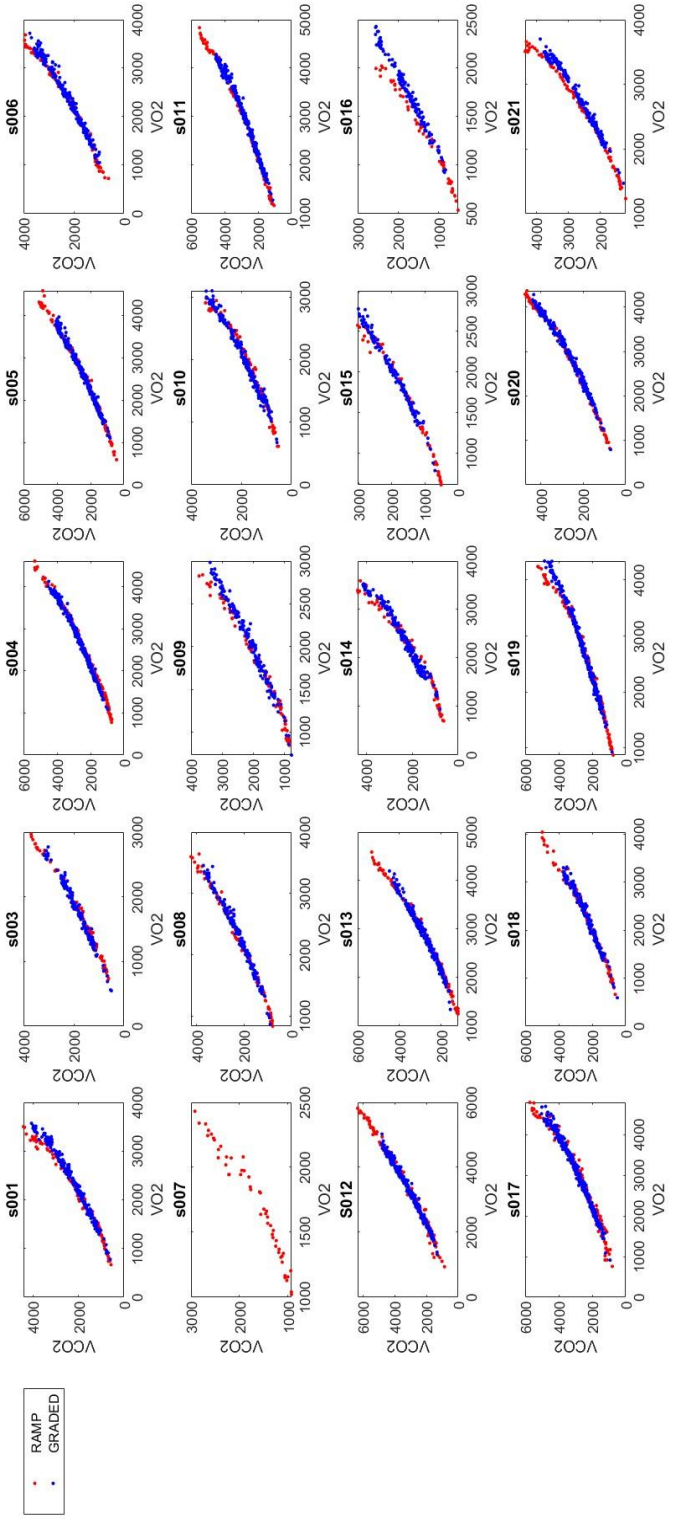
Jones, A. M., & Vanhatalo, A. (2017). The 'Critical Power' Concept: Applications to Sports Performance with a Focus on Intermittent High-Intensity Exercise. *Sports Med*, *47*(Suppl 1), 65-78.

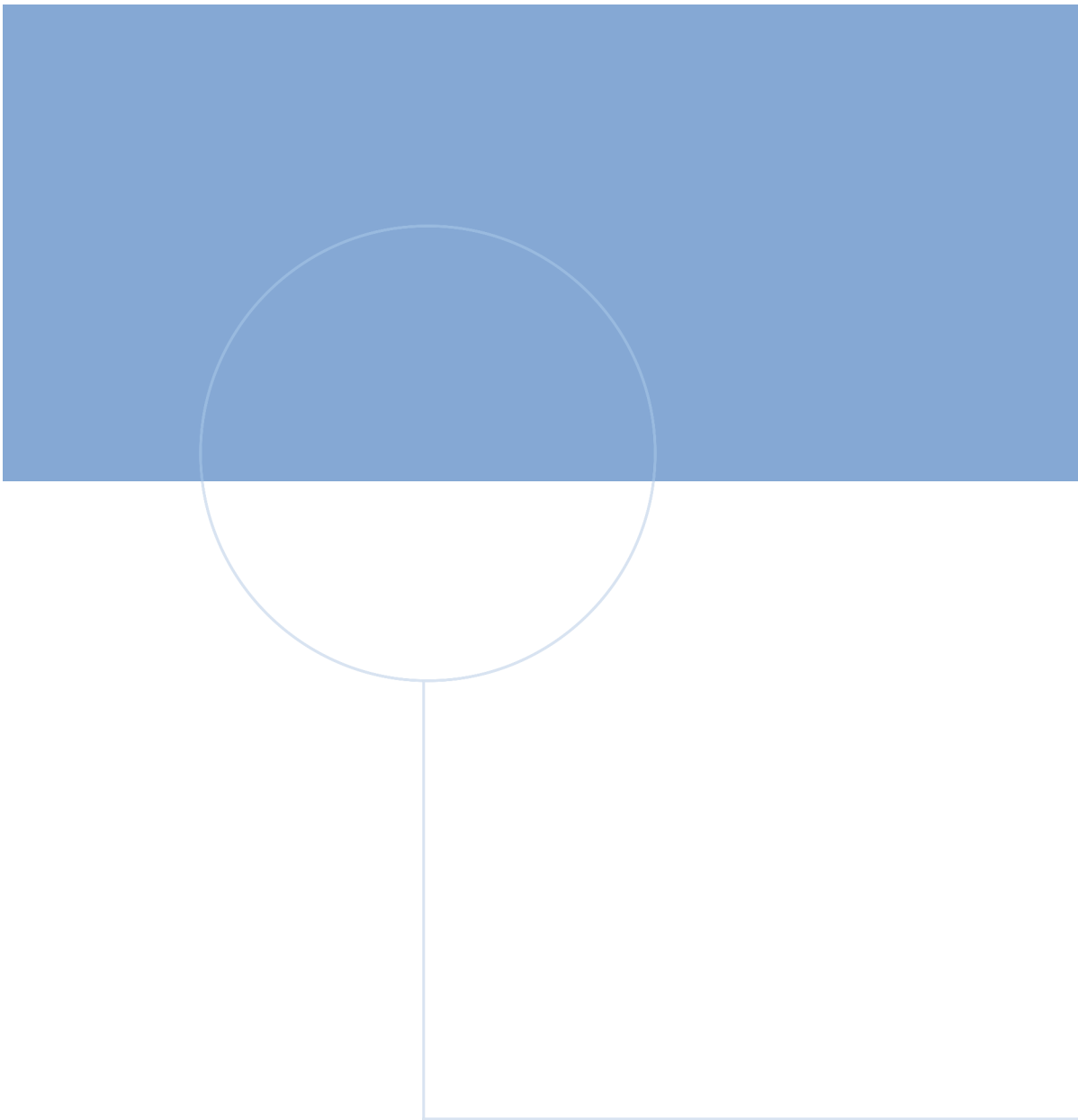
Kim, K. J., Rivas, E., Prejean, B., Frisco, D., Young, M., & Downs, M. (2021). Novel Computerized Method for Automated Determination of Ventilatory Threshold and Respiratory Compensation Point. *Front Physiol*, *12*, 782167.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*, *15*(2), 155-163.
Loat, C. E. R., & Rhodes, E. C. (1993). Relationship Between the Lactate and Ventilatory Thresholds During Prolonged Exercise. *Sports Medicine*, *15*(2), 104-115.
Lucía, A., Hoyos, J., Pérez, M., & Chicharro, J. L. (2000). Heart rate and performance parameters in elite cyclists: a longitudinal study. *Med Sci Sports Exerc*, *32*(10), 1777-1782.
Morton, R. H. (1989). Detection of a lactate threshold during incremental exercise? *J Appl Physiol (1985)*, *67*(2), 885-888.
Myers, J., & Ashley, E. (1997). Dangerous curves. A perspective on exercise, lactate, and the anaerobic threshold. *Chest*, *111*(3), 787-795.
Myers, J., Walsh, D., Buchanan, N., McAuley, P., Bowes, E., & Froelicher, V. (1994). Increase in blood lactate during ramp exercise: comparison of continuous and threshold models. *Med Sci Sports Exerc*, *26*(11), 1413-1419.
Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308-313.
Pallarés, J. G., Morán-Navarro, R., Ortega, J. F., Fernández-Elías, V. E., & Mora-Rodriguez, R. (2016). Validity and Reliability of Ventilatory and Blood Lactate Thresholds in Well-Trained Cyclists. *PLOS ONE*, *11*(9), e0163389.
Pettitt, R. W., Clark, I. E., Ebner, S. M., Sedgeman, D. T., & Murray, S. R. (2013). Gas exchange threshold and VO2max testing for athletes: an update. *J Strength Cond Res*, *27*(2), 549-555.
Plato, P. A., McNulty, M., Crunk, S. M., & Tug Ergun, A. (2008). Predicting lactate threshold using ventilatory threshold. *Int J Sports Med*, *29*(9), 732-737.
Poole, D. C., Rossiter, H. B., Brooks, G. A., & Gladden, L. B. (2021). The anaerobic threshold: 50+ years of controversy. *J Physiol*, *599*(3), 737-767.
Reinhard, U., Müller, P. H., & Schmülling, R. M. (1979). Determination of anaerobic threshold by the ventilation equivalent in normal individuals. *Respiration*, *38*(1), 36-42.
Sales, M. M., Sousa, C. V., da Silva Aguiar, S., Knechtle, B., Nikolaidis, P. T., Alves, P. M., & Simões, H. G. (2019). An integrative perspective of the anaerobic threshold. *Physiol Behav*, *205*, 29-32.
Seiler, S. (2010). What is best practice for training intensity and duration distribution in endurance athletes? *Int J Sports Physiol Perform*, *5*(3), 276-291.
Shimizu, M., Myers, J., Buchanan, N., Walsh, D., Kraemer, M., McAuley, P., & Froelicher, V. F. (1991). The ventilatory threshold: method, protocol, and evaluator agreement. *Am Heart J*, *122*(2), 509-516.
Svedahl, K., & MacIntosh, B. R. (2003). Anaerobic threshold: the concept and methods of measurement. *Can J Appl Physiol*, *28*(2), 299-323.
Tanaka, H., Monahan, K. D., & Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *J Am Coll Cardiol*, *37*(1), 153-156.
Teo, W., Newton, M. J., & McGuigan, M. R. (2011). Circadian rhythms in exercise performance: implications for hormonal and muscular adaptation. *J Sports Sci Med*, *10*(4), 600-606.
Wasserman, K., Whipp, B. J., Koyl, S. N., & Beaver, W. L. (1973). Anaerobic threshold and respiratory gas exchange during exercise. *J Appl Physiol*, *35*(2), 236-243.
Yeh, M. P., Gardner, R. M., Adams, T. D., Yanowitz, F. G., & Crapo, R. O. (1983). "Anaerobic threshold": problems of determination and validation. *Journal of Applied Physiology*, *55*(4), 1178-1186.

Appendix 1

Comparison data collection $\dot{V}O_2$ vs $\dot{V}CO_2$ (ml min^{-1}) during $RAMP_{Own}$ (red) vs $GRADED_{Own}$(blue) for primary (Own) data. Participant n.7 performed only $RAMP_{Own}$.





NTNU

Norwegian University of
Science and Technology