



Strategies to Ensure Accurate Calculation of Parameters of the VO_2 Response Profile During Heavy Intensity Cycle Ergometer Exercise

David Wilfred Hill^{1,*}

¹Department of Kinesiology, Health Promotion, and Recreation, University of North Texas, Denton, United States

*Corresponding author: Department of Kinesiology, Health Promotion, and Recreation, University of North Texas, 1155 Union Circle #310769, Denton, United States. Tel: +1-4697636769, Email: david.hill@unt.edu

Received 2019 September 16; Accepted 2019 September 28.

Abstract

Background: The parameters of the VO_2 response profile are obtained by fitting breath-by-breath VO_2 data from an exercise test to an appropriate mathematical model. Several strategies have been recommended to ensure, or at least improve, the accuracy of the values.

Objectives: The purpose of this study was to evaluate two strategies to enhance the accuracy of parameter estimates that describe the two-component VO_2 response during heavy intensity exercise. The first was to use data from a number of tests rather than just one. The second was to 'smooth' the data, using three-breath, five-breath, or seven-breath rolling averages of the breath-by-breath VO_2 data prior to fitting the data to the two-component model.

Methods: Twenty participants (eight women and twelve men) performed six 6-min heavy-intensity (midway between the ventilatory threshold and VO_{2max}) cycle ergometer tests. Breath-by-breath data and smoothed data from each test were fit to a two-component model. The parameter estimates from the first test, and the average of the values from the first two, first three, first four, first five, and all six tests were compared against the criterion value, which was the average of all six values obtained using five-breath averages.

Results: Modeling five-breath averages of data from the first test generated values for the parameters that were closely related to the criterion values. Modeling data from two or three tests improved the accuracy slightly, but improvements were small, and negligible when more than three tests were included.

Conclusions: Depending upon the accuracy required, that is depending upon how close each and every participant's value must be to his or her 'true' value, smoothed data from one or two tests is sufficient to calculate the values that describe the two-parameter VO_2 response profile in heavy intensity cycling exercise.

Keywords: Kinetics, Heavy Intensity, Modeling, Cycling, Slow Component

1. Background

The pulmonary VO_2 response profile in exercise reflects the underlying metabolic activity in the muscles (1, 2). In moderate intensity exercise, work rates below the lactate threshold, the metabolic response is mirrored by the mono-exponential increase in VO_2 leading to rapid attainment of a steady state (1). For exercise in the heavy domain, which comprises work rates above the lactate threshold and up to critical power or critical speed, the asymptote of the relationship between time to exhaustion and work rate or speed (3-5), the rate of lactate production is said to be balanced by the rate of removal, so the blood lactate concentration will increase in the first few minutes and then decrease gradually or stay steady (6); this metabolic profile is reflected in the two-phase VO_2 response, which fea-

tures (i) a primary phase or fast response followed by (ii) a slow phase or slow component, which emerges after ~2 min of exercise and leads to a steady-state VO_2 (7). In severe exercise, there is two-phase VO_2 response and, if exercise is continued long enough, the slow component will bring the VO_2 to VO_{2max} (7).

Characteristics of the VO_2 response profile -or parameters of the kinetics of the VO_2 response- are determined using a three-phase process: data collection, data processing, and data fitting (8). Since the advent of automated gas analysis systems that provide VO_2 data on a breath-by-breath (B×B) basis, these are the systems of choice for collecting data. The second phase, data processing, involves treating the data to ensure that parameters are estimated with the greatest precision and accuracy during the final

phase. This data processing aims to improve the signal (the underlying responses) to noise (breath to breath variability) ratio (1, 9, 10). The final phase is data fitting, the mathematical process of fitting the breath-by-breath VO_2 data to an appropriate mathematical model using iterative nonlinear regression procedures on any number of readily available statistical or graphing packages to identify what the values of the parameters are the best describe how closely the actual data fit the model. Using these packages, parameter estimates are generated, each with an associated SEE, which describes the precision of the estimates.

The focus of the present study is on the second phase, data processing. We assume that data are collected carefully on a B×B basis, using calibrated equipment, under reasonable environmental conditions, from participants who are properly prepared and motivated. We assume that an appropriate statistical analysis package is available and that an acceptable model has been selected.

Two approaches have been taken in order to improve the precision, and ensure the accuracy, of parameter estimates. First, the breath-by-breath data have been ‘smoothed’ -for example, by using interpolation to generate second-by-second values or by generating three-breath (3-B), five-breath (5-B), or seven-breath (7-B) (etc.) rolling averages- prior to performing the iterative progression (9). Second, data have been collected from several identical exercise tests -the parameter estimates generated by mathematical modeling of the data from each test can be combined (averaged) or the data from the tests can be combined prior to mathematical modeling (8, 9). While it may be inherently obvious that smoothing data or replicating the exercise tests would improve the precision of parameter estimates, relatively little research has sought to determine the optimal treatment of exercise data to ensure the accuracy of the parameter values (9).

2. Objectives

The objective of this study was to evaluate the impact of these two strategies to improve the accuracy of the descriptors of the VO_2 response profile during heavy intensity cycle ergometer exercise. Twenty participants performed six identical exercise tests. B×B data from each test, as well as rolling 3-B, 5-B, and 7-B averages, were fitted to a two-component (primary + slow) model. Parameter estimates from the combinations of number of tests used (one to six) and the methods of smoothing (none, 3-B, 5-B, and 7-B rolling averages) were compared against a criterion value. The purpose of this study was to identify the optimal smoothing method and the minimum number of

tests necessary to ensure accurate estimation of the parameters of the VO_2 response in heavy intensity exercise.

3. Methods

3.1. Participants

The study procedures were approved by the Institutional Review Board for the Protection of Human Subjects in Research at the university prior to any recruitment of participants. The study was conducted in accordance with the latest Declaration of Helsinki (11). Eight women (mean \pm SD: age 22 ± 1 y, height 167 ± 9 cm, weight 66 ± 11 kg, $\text{VO}_{2\text{max}}$ 39 ± 6 mL.kg⁻¹.min⁻¹) and twelve men (23 ± 2 y, 182 ± 8 cm, 79 ± 12 kg, 43 ± 5 mL.kg⁻¹.min⁻¹) volunteered to participate and provided informed consent. These 20 participants were involved in recreational sport or fitness activities, but not organized sport activities. They were all familiar with exercise testing procedures and with breathing through a mouthpiece. They verified that they did not change their exercise routines, diet, or sleep habits over the course of the study.

3.2. Overview

Participants performed an incremental test for determination of their $\text{VO}_{2\text{max}}$ and the VO_2 at the ventilatory threshold. Then they performed a series of six 6-min tests at a work rate individually selected so that the oxygen demand would be midway between $\text{VO}_{2\text{max}}$ and the VO_2 at the ventilatory threshold. The testing sessions were separated by at least 24 hours and were scheduled at the same time of day for each participant to avoid the confounding effects of time of day that we have reported on responses associated with the ventilatory threshold (12) and in severe intensity VO_2 kinetics (13); the work rates used in the present study lay between the ventilatory threshold and the lower boundary of the severe intensity domain. Tests were performed under similar conditions in a temperature-controlled laboratory (20°C to 22°C ; $\sim 50\%$ relative humidity), with no distractions. Data collection was completed in a three-week period. Participants were instructed to sleep at least six hours the night before each test; not to exercise and not to ingest carbonated beverages, caffeine, or alcohol for 12 hours before each test; and not to eat a heavy meal in the three hours before each test. Actual dietary intake was at each participant’s discretion and was not recorded. They were tested only if they verified that they had adhered to all these instructions.

3.3. Incremental Tests to Determine VO_{2max}

The incremental tests were performed on an Monark Ergonomic 828E (Varberg, Sweden) cycle ergometer, with pedaling cadence of ~ 80 revolutions per min (rev/min). A digital readout of the cadence was visible during the tests. The tests began with three minutes of baseline data collection during seated rest. The initial work rates were 40 W for women and 80 W for men. Work rate was abruptly increased 20 W each minute.

Throughout each test, expired gases were analyzed using a MedGraphics (St. Paul, Minnesota, USA) Express metabolic cart. The cart was calibrated before each test according to the manufacturer's instructions. Breath-by-breath VO_2 data were reduced to serial 15-s averages. Tests were terminated when the participant allowed the cadence to drop below 70 rev.min⁻¹ for five seconds, despite strong verbal encouragement. VO_{2max} was determined as the highest average of adjacent 15-s averages. The ventilatory threshold was identified as described by Wasserman and colleagues (14).

3.4. Constant Power Heavy Intensity Tests

Tests were performed using the same Monark ergometer as for the incremental tests, and during each test, expired gases were analyzed using the same MedGraphics metabolic cart. The tests began with three minutes of baseline data collection during seated rest. After the rest, the participant began pedaling and rapidly brought the pedaling cadence up to 80 (rev.min⁻¹) as the resistance was abruptly increased to provide the work rate that had been individually pre-determined by the primary investigator.

3.5. VO_2 Kinetics in the Constant Power Tests

For each individual, for each test, data from the first 20 s of exercise were removed (8) and the remaining B×B, 3-B, 5-B, or 7-B data points were fit to the following model (1) using iterative regression procedures in KaleidaGraph 4.5 software (Reading, PA USA) (Equation 1):

$$VO_2(t) = VO_{2baseline} + A_{primary} \times \left(1 - e^{-\left(\frac{t - TD_{primary}}{\tau_{primary}}\right)}\right) + A_{slow} \left(1 - e^{-\left(\frac{t - TD_{slow}}{\tau_{slow}}\right)}\right) \quad (1)$$

$VO_{2baseline}$ is the steady state VO_2 at the end of the three minutes of seated rest prior to exercise, $A_{primary}$ and A_{slow} are the projected increases in VO_2 due to the primary and slow component responses, $TD_{primary}$ and TD_{slow} are the

time delays preceding the two responses, and $\tau_{primary}$ and τ_{slow} are the time constants of the two responses.

The mean response time ($MRT_{primary}$) represents the time from the start of exercise until the VO_2 has increased 63% of the $A_{primary}$. It is calculated as the sum of $TD_{primary}$ and $\tau_{primary}$ and tends to be more stable and reliable than either of the parameters which it comprises. $MRT_{primary}$ was used as a supplementary variable to describe the primary phase of the VO_2 response.

The actual increase in VO_2 due to the slow component, A'_{slow} , was calculated as (Equation 2):

$$A'_{slow} = A_{slow} \times \left(1 - e^{-\left(\frac{t_{exhaustion} - TD_{slow}}{\tau_{slow}}\right)}\right) \quad (2)$$

A'_{slow} and TD_{slow} were used to describe the characteristics of the slow component in all tests.

3.6. Statistical Analyses

Descriptive characteristics of participants were calculated separately for women and men. For all other analyses, data were collapsed across the sexes. Sample size was 20.

First, to identify which smoothing method would be selected to provide the criterion measure for each parameter, the SEE of each parameter that was directly generated using the iterative regression procedure in KaleidaGraph ($TD_{primary}$, $\tau_{primary}$, $A_{primary}$, and TD_{slow}) were compared using a two-way (type of smoothing [B×B, 3-B, 5-B, 7-B] × test number [first, second, third, fourth, fifth, sixth]) repeated-measures analysis of variance (ANOVA) in SPSS V.22 (SPSS, Armonk, NY, USA). Two other descriptors of the VO_2 response profile, $MRT_{primary}$ and A'_{slow} , were not included because they are calculated values, and not directly generated by KaleidaGraph. Data were tested for sphericity using Mauchly's test of sphericity and, if assumptions were violated, results were interpreted using a Greenhouse-Geisser correction. Significance was set at $P < 0.05$. The post hoc comparisons of SEE were performed using paired-means t tests with a fixed level of significance ($P < 0.05$) rather than correcting the P-level because of multiple comparisons; given that these comparisons were a tool to identify the criterion measure and any difference was considered meaningful. Data are presented as mean \pm SD.

Second, the optimal smoothing method and the minimum number of tests necessary to ensure accurate estimation of the parameters of the VO_2 response in heavy intensity exercise, six values for each parameter ($TD_{primary}$, $\tau_{primary}$, $MRT_{primary}$, $A_{primary}$, TD_{slow} , and A'_{slow} , and for the SEE associated with the four parameters that were directly

generated by KaleidaGraph) obtained using each smoothing method were calculated. The first of the six values was simply the value from the first test, the second was the average of the values from the first and second tests, the third was the average of the values from the first three tests, etc. These values were compared using a two-way (type of smoothing [$B \times B$, 3-B, 5-B, 7-B] \times number of tests used to calculate the value [1, 2, 3, 4, 5, 6]) repeated-measures ANOVA. In addition, correlations between each value and the criterion measure were calculated.

Finally, a Bland-Altman plot (15) was created for each comparison. Arguably, use of a Bland-Altman analysis is preferred when the task is to identify methods that produce the same answer (in this case, to identify methods that produce an answer that is the same as the criterion), as opposed to identifying values or methods that are different. As proposed by Krouwer (16), criterion values were on the x-axis and differences between values and the criterion were on the y-axis. Aside from the fact that 23 Bland-Altman plots were needed to assess the agreement between the various means and the criterion measure for each variable, one challenge of using Bland-Altman plots was that levels of agreement (defined by the 95% confidence interval around the mean difference for each comparison) were different for each comparison, even for comparisons for the same variable. This meant, for example, that for any given parameter, some types of smoothing faced stricter limits of agreement than others. In addition, Bland and Altman (15) noted that the levels of agreement that they propose (95% confidence interval around the mean difference) may be unacceptably large in some situations, such as clinical testing, and this was the case in the present study: the 95% confidence interval in many cases was simply too broad, and allowed deviations from the criterion value that would be inappropriate in research or practical applications. In order to address the issue of inequity caused by using the 95% confidence intervals that were unique to each comparison, as well as to address the issue of appropriateness, we used stricter levels of agreement. We constructed the levels of agreement to be $0.0 \pm 1.0 \times \text{SEE}$ of the criterion measure. Thus, we used the same limits of agreement for all comparisons involving a given variable; the levels of agreement were similar to the 85% confidence interval. We also constructed levels of agreement that were $0.0 \pm 1.5 \times \text{SEE}$ of the criterion measure. We estimate that these ranges were similar to the 90% confidence interval. In each case, these plots then were intolerant of bias; they defined a range of 'accepted' individual values that were close enough to the criterion value to meet the requirements of research or practical applications.

4. Results

Results of the two-way ANOVA that was used to identify which smoothing method would be selected to provide the criterion measure for each parameter revealed a significant effect of type of smoothing ($P < 0.05$) for three of the SEE associated with the parameter estimates that were directly generated using KaleidaGraph ($\text{tau}_{\text{primary}}$, A_{primary} , and TD_{slow} , but not $\text{TD}_{\text{primary}}$). The mean values associated with the main effects are presented in the farthest right column of Tables 1-4 and the results of the post hoc t tests are provided below, with differences significant at the 0.05 level:

$$\text{SEE } \text{TD}_{\text{primary}}, B \times B = 7-B = 3-B = 5-B$$

$$\text{SEE } \text{tau}_{\text{primary}}, B \times B > 3-B = 7-B > 5-B$$

$$\text{SEE } A_{\text{primary}}, B \times B > 3-B = 7-B = 5-B$$

$$\text{SEE } \text{TD}_{\text{slow}}, B \times B > 3-B > 5-B = 7-B$$

Based on the mathematically smaller SEE associated with parameters generated using 5-B smoothing, this method was chosen to identify the criterion values for all parameters. We note that the coefficient of variation among the values from the six tests tended to be smallest for 5-B averages, as well (these results not provided, but can be inferred from data in Tables 1-4). We assumed that the average value from all six tests would be most representative of the 'true' or criterion value.

Mean values for the parameter that were obtained using the 24 combinations of kind-of-smoothing ($B \times B$, 3-B, 5-B, 7-B) and number-of-tests used to calculate the values (1 to 6) are presented in Tables 1-5. Results of the two-way ANOVAs that were used to investigate the effects of smoothing and number of tests revealed no significant main or interaction effects. Thus, we cannot argue that there were any differences among the reported values, regardless of the type of smoothing or the number of tests used to calculate the values. Similarly, the results of the correlational analyses, which are also presented in the tables, showed that values from almost all combinations of type of smoothing and number of tests used were strongly correlated with the criterion 5-B six test values. Because of space limitations, results for the $\text{MRT}_{\text{primary}}$, which was calculated as the sum of $\text{TD}_{\text{primary}}$ and $\text{tau}_{\text{primary}}$, are not provided in tabular form. Variability in $\text{MRT}_{\text{primary}}$ is much less than in either parameter individually; thus, the accuracy of $\text{MRT}_{\text{primary}}$ generated with data from one test was acceptable, regardless of how the data were smoothed.

Two hundred and seventy-six Bland-Altman plots were constructed (6 variables \times 2 levels of agreement \times 23 comparisons). Because of space restrictions, the plots are not reproduced here. The important results from these plots

Table 1. Estimates of TD_{primary} (with Units of s) Generated Using the Results from the First Test, the First Two Tests, the First Three Tests, the First Four Tests, the First Five Tests, and All Six Tests^{a, b}

Type of Smoothing	Number of Values Averaged (Number of Tests Included in Calculation)					
	1	2	3	4	5	6
B × B						
Mean	6 ± 9	7 ± 7	7 ± 8	8 ± 7 ^C	9 ± 8 ^B	10 ± 8 ^B
SEE	8 ± 9	6 ± 8	7 ± 7	8 ± 7	7 ± 7	8 ± 7
Corr (r)	0.45	0.66	0.72	0.80	0.83	0.88
3-B						
Mean	12 ± 13	13 ± 11 ^C	11 ± 9 ^B	12 ± 8 ^B	11 ± 8 ^B	11 ± 9 ^B
SEE	6 ± 4	6 ± 3	5 ± 3	5 ± 3	5 ± 3	5 ± 3
Corr (r)	0.54	0.70	0.80	0.83	0.89	0.91
5-B						
Mean	12 ± 7 ^C	11 ± 6 ^C	11 ± 6 ^B	12 ± 6 ^B	11 ± 6 ^B	11 ± 6 ^A
SEE	6 ± 4	5 ± 3	5 ± 4	4 ± 3	4 ± 4	4 ± 3
Corr (r)	0.60	0.87	0.92	0.90	0.95	Criterion
7-B						
Mean	13 ± 9	12 ± 9 ^C	11 ± 8 ^B	11 ± 7 ^B	12 ± 6 ^B	11 ± 7 ^B
SEE	7 ± 5	7 ± 5	6 ± 5	7 ± 3	6 ± 5	6 ± 4
Corr (r)	0.54	0.81	0.88	0.90	0.88	0.92

^aExercise responses from each test were analyzed individually using KaleidaGraph, and the parameter estimates that were generated, and their SEE, were then averaged.

^b(A) identifies the criterion measure (average of six 5-B values). (B) identifies means for which all individual differences (individual's parameter estimate minus their criterion) fell within the limits of agreement that were calculated as $\pm 1.0 \times \text{SEE}$ associated with the mean criterion measure (4 ± 3 s); in each case, these limits of agreement were approximately the same as limits that would be defined by the 85% confidence interval. The range of acceptable differences was -4 s to +4 s, which represents the criterion value $\pm \sim 32\%$. (C) identifies values for which all differences fell within the limits of agreement that were calculated as $\pm 1.5 \times \text{SEE}$; these limits of agreement were approximately the same as limits that would be defined by the 90% confidence interval. The acceptable range of differences was -5 s to +5 s, which represents the criterion value $\pm \sim 48\%$.

Table 2. Estimates of tau_{primary} (with Units of s) Generated Using the Results from the First Test, the First Two Tests, the First Three Tests, the First Four Tests, the First Five Tests, and All Six Tests^{a, b}

Type of Smoothing	Number of Values Averaged (Number of Tests Included in Calculation)					
	1	2	3	4	5	6
B × B						
Mean	45 ± 15	42 ± 13	43 ± 12	42 ± 12 ^C	43 ± 11 ^C	43 ± 11 ^C
SEE	13 ± 9	13 ± 7	12 ± 7	11 ± 7	11 ± 7	11 ± 7
Corr (r)	0.56	0.77	0.80	0.88	0.90	0.92
3-B						
Mean	44 ± 13	44 ± 11	43 ± 9 ^C	43 ± 8 ^C	43 ± 8 ^B	43 ± 8 ^B
SEE	6 ± 4	6 ± 3	5 ± 3	5 ± 3	5 ± 3	5 ± 3
Corr (r)	0.72	0.85	0.89	0.93	0.94	0.93
5-B						
Mean	43 ± 7 ^C	42 ± 6 ^B	42 ± 6 ^B	42 ± 6 ^B	42 ± 6 ^B	42 ± 6 ^A
SEE	5 ± 4	5 ± 3	4 ± 4	4 ± 3	3 ± 4	3 ± 3
Corr (r)	0.91	0.96	0.97	0.98	0.99	Criterion
7-B						
Mean	42 ± 9	43 ± 9 ^C	42 ± 8 ^C	42 ± 7 ^B	42 ± 6 ^B	42 ± 7 ^B
SEE	7 ± 5	6 ± 4	5 ± 4	5 ± 3	4 ± 4	5 ± 4
Corr (r)	0.79	0.90	0.93	0.94	0.94	0.94

^aExercise responses from each test were analyzed individually using KaleidaGraph, and the parameter estimates that were generated, and their SEE, were then averaged

^b(A) identifies the criterion measure (average of six 5-B values). (B) identifies means for which all individual differences (individual's parameter estimate minus their criterion) fell within the limits of agreement that were calculated as $\pm 1.0 \times \text{SEE}$ associated with the mean criterion measure (3 ± 3 s); in each case, these limits of agreement were approximately the same as limits that would be defined by the 85% confidence interval. The range of acceptable differences was -3 s to +3 s, which represents the criterion value $\pm \sim 8\%$. (C) identifies values for which all differences fell within the limits of agreement that were calculated as $\pm 1.5 \times \text{SEE}$; these limits of agreement were approximately the same as limits that would be defined by the 90% confidence interval. The acceptable range of differences was -5 s to +5 s, which represents the criterion value $\pm \sim 12\%$.

Table 3. Estimates of A_{primary} (with Units of mL/min) Generated Using the Results from the First Test, the First Two Tests, the First Three Tests, the First Four Tests, the First Five Tests, and All Six Tests^{a, b}

Type of Smoothing	Number of Values Averaged (Number of Tests Included in Calculation)					
	1	2	3	4	5	6
B × B						
Mean	1313 ± 503	1240 ± 460	1260 ± 435 ^C	1252 ± 442 ^B	1269 ± 435 ^B	1265 ± 433 ^B
SEE	167 ± 80	150 ± 70	142 ± 57	145 ± 45	140 ± 40	143 ± 37
Corr (r)	0.76	0.86	0.91	0.94	0.97	0.98
3-B						
Mean	1293 ± 444	1308 ± 455 ^C	1294 ± 443 ^C	1281 ± 432 ^B	1265 ± 427 ^B	1273 ± 429 ^B
SEE	93 ± 48	84 ± 39	82 ± 30	79 ± 3	78 ± 3	79 ± 3
Corr (r)	0.82	0.93	0.95	0.95	0.98	0.98
5-B						
Mean	1243 ± 467 ^C	1263 ± 446 ^B	1260 ± 438 ^B	1258 ± 430 ^B	1261 ± 426 ^B	1262 ± 425 ^A
SEE	68 ± 21	64 ± 20	66 ± 18	67 ± 18	66 ± 17	68 ± 21
Corr (r)	0.89	0.94	0.98	0.99	0.99	Criterion
7-B						
Mean	1269 ± 451	1229 ± 486 ^C	1244 ± 468 ^B	1251 ± 444 ^B	1255 ± 437 ^B	1257 ± 432 ^B
SEE	75 ± 5	69 ± 5	68 ± 5	67 ± 3	67 ± 5	66 ± 4
Corr (r)	0.88	0.86	0.94	0.96	0.99	0.99

^aExercise responses from each test were analyzed individually using KaleidaGraph, and the parameter estimates that were generated, and their SEE, were then averaged.

^b(A) identifies the criterion measure (average of six 5-B values). (B) identifies means for which all individual differences (individual's parameter estimate minus their criterion) fell within the limits of agreement that were calculated as $\pm 1.0 \times \text{SEE}$ associated with the mean criterion measure (68 ± 21 mL/min); in each case, these limits of agreement were approximately the same as limits that would be defined by the 85% confidence interval. The range of acceptable differences was -68 mL/min to +68 mL/min, which represents the criterion value $\pm \sim 5\%$. (C) identifies values for which all differences fell within the limits of agreement that were calculated as $\pm 1.5 \times \text{SEE}$; these limits of agreement were approximately the same as limits that would be defined by the 90% confidence interval. The acceptable range of differences was -102 mL/min to +102 mL/min, which represents the criterion value $\pm \sim 8\%$. Of note, 102 mL/min is approximately 1.4 mL/kg/min when expressed relative to body weight. Clearly, the two-component model identifies the A_{primary} with very high precision.

Table 4. Estimates of TD_{slow} (with Units of s) Generated Using the Results from the First Test, the First Two Tests, the First Three Tests, the First Four Tests, the First Five Tests, and All Six Tests^{a, b}

Type of Smoothing	Number of Values Averaged (Number of Tests Included in Calculation)					
	1	2	3	4	5	6
B × B						
Mean	119 ± 15	127 ± 13	125 ± 12 ^C	127 ± 12 ^C	126 ± 11 ^C	126 ± 11 ^C
SEE	22 ± 9	21 ± 7	19 ± 7	19 ± 7	18 ± 7	19 ± 7
Corr (r)	0.56	0.77	0.85	0.88	0.90	0.92
3-B						
Mean	118 ± 13	118 ± 11 ^C	121 ± 9 ^B	123 ± 8 ^C	124 ± 8 ^B	124 ± 8 ^B
SEE	11 ± 4	11 ± 3	10 ± 3	10 ± 3	10 ± 3	10 ± 3
Corr (r)	0.72	0.86	0.89	0.93	0.94	0.93
5-B						
Mean	121 ± 17 ^C	117 ± 16 ^B	118 ± 16 ^B	120 ± 14 ^B	120 ± 15 ^B	120 ± 14 ^A
SEE	10 ± 4	6 ± 3	7 ± 4	6 ± 3	6 ± 4	6 ± 3
Corr (r)	0.91	0.96	0.97	0.98	0.99	Criterion
7-B						
Mean	127 ± 22	121 ± 19 ^B	118 ± 18 ^C	118 ± 17 ^B	119 ± 14 ^B	118 ± 16 ^B
SEE	8 ± 5	9 ± 4	8 ± 4	8 ± 3	7 ± 4	8 ± 4
Corr (r)	0.79	0.88	0.93	0.96	0.96	0.95

^aExercise responses from each test were analyzed individually using KaleidaGraph, and the parameter estimates that were generated, and their SEE, were then averaged.

^b(A) identifies the criterion measure (average of six 5-B values). (B) identifies means for which all individual differences (individual's parameter estimate minus their criterion) fell within the limits of agreement that were calculated as $\pm 1.0 \times \text{SEE}$ associated with the mean criterion measure (6 ± 3 s); in each case, these limits of agreement were approximately the same as limits that would be defined by the 85% confidence interval. The range of acceptable differences was -6 s to +6 s, which represents the criterion value $\pm \sim 5\%$. (C) identifies values for which all differences fell within the limits of agreement that were calculated as $\pm 1.5 \times \text{SEE}$; these limits of agreement were approximately the same as limits that would be defined by the 90% confidence interval. The acceptable range of differences was -8 s to +8 s, which represents the criterion value $\pm \sim 7\%$.

Table 5. Estimates of A_{slow}^A (with Units of mL/min) Generated Using the Results from the First Test, the First Two Tests, the First Three Tests, the First Four Tests, the First Five Tests, and All Six Tests^{a, b}

Type of Smoothing	Number of Values Averaged (Number of Tests Included in Calculation)					
	1	2	3	4	5	6
B × B						
Mean	448 ± 111	456 ± 105 ^B	450 ± 100 ^C	452 ± 89 ^C	453 ± 88 ^B	455 ± 89 ^B
Corr (r)	0.70	0.83	0.91	0.95	0.97	0.96
3-B						
Mean	454 ± 99	453 ± 91 ^C	459 ± 88 ^C	459 ± 87 ^B	458 ± 83 ^B	459 ± 81 ^B
Corr (r)	0.82	0.93	0.95	0.95	0.96	0.98
5-B						
Mean	460 ± 89 ^C	452 ± 93	454 ± 87 ^B	457 ± 83 ^B	456 ± 86 ^B	457 ± 82 ^A
Corr (r)	0.86	0.94	0.98	0.99	0.99	Criterion
7-B						
Mean	466 ± 97 ^C	450 ± 93 ^C	454 ± 87 ^B	454 ± 83 ^B	454 ± 86 ^B	454 ± 432 ^B
Corr (r)	0.89	0.88	0.96	0.96	0.98	0.98

^aExercise responses from each test were analyzed individually using KaleidaGraph, and the parameter estimates that were generated, and their SEE, were then averaged.

^b(A) identifies the criterion measure (average of six 5-B values). There were no SEE associated with the MRT parameter, because it was calculated from the values of τ_{slow} and A_{slow} ; we chose to use the same values for this amplitude as were calculated for the $A_{primary}$ parameter. (B) identifies means for which all individual differences (individual's parameter estimate minus their criterion) fell within the limits of agreement that were calculated as ± 68 mL/min, which represents the criterion value $\pm \sim 15\%$. (C) identifies values for which all differences fell within the limits of agreement that were calculated as ± 102 mL/min, which represents the criterion value $\pm \sim 22\%$. Of note, 68 mL/kg and 102 mL/min are less than 1.0 mL/kg/min and 1.5 mL/kg/min, respectively.

are summarized in the tables and can be explained as follows: mean values from type-of-smoothing × number-of-tests combinations for which all individual values fell within very strict limits of agreement ($\pm 1.0 \times$ SEE) are identified by superscript 'B' and mean values for which all individual values fell within 50% broader limits of agreement are identified by superscript 'C'. So, for example, for the 5-B $\tau_{primary}$ parameter value obtained using only data from the first test, all individual values fell within 5 s ($1.5 \times$ SEE) of their associated criterion value; when parameter values from the first two tests were averaged, all fell within 3 s ($1.0 \times$ SEE). This can be interpreted that a single test provides enough 5-B data to closely identify (within 5 s) the value of the $\tau_{primary}$ parameter and that, with data from two tests, individual values will be within 3 s of the 'true' value. Requiring more tests produces rapidly diminishing returns; we cannot even argue that accuracy and precision of these values improves when data from more than two or three tests were included in their calculation.

5. Discussion

The important finding in the present study is that curve fitting of 5-B data from only one exercise test can generate accurate values for parameters of the two-component VO_2 response profile in heavy intensity exercise. 3-B and 7-B smoothing methods were also very good to improve the signal-to-noise ratio and reduce the number of tests that must be performed. The use of two or three

tests may be indicated if acceptable tolerance of deviation from the 'true' value of each parameter is very small, for example, less than 3 s for $\tau_{primary}$, < 1 mL/kg/min for $A_{primary}$, < 6 s for $TD_{primary}$, and < 1 mL/kg/min for $A_{primary}$. Slightly better accuracy may be obtained if more tests are performed, but any improvements may not justify the extra demands on personnel, participants, and other resources.

Subsequent to the work of Lamarra and colleagues (10), it has often been assumed that multiple trials are required to minimize the signal-to-noise ratio, and often this is accomplished by combining data from the trials before smoothing the data, that is, before fitting them to a mathematical model (8). Benson and colleagues (8) applied a variety of smoothing interventions to one to ten sets of simulated moderate intensity. Like Francescato and colleagues (17), who also used simulated data, they found little difference between the effects of smoothing. They did report that four trials were optimal. They also found that combining data before modeling was superior to modeling and then combining the results (i.e., averaging the parameter estimates from different tests, as we did in the present study.) However, they noted that modeling results from tests individually and then averaging the results has been proposed and used (18) and may be statistically more appropriate (19). We note, that this method also allows the investigator to evaluate whether there is a trend in the results. i.e., to determine if responses are changing over time (they were not, in the present study).

In a study similar to the present study, Keir and col-

leagues (9) evaluated the effects of several smoothing techniques applied to data from four identical moderate intensity exercise tests (i.e., assuming that combining data from four tests was requisite for obtaining data with good precision). They concluded that modeling had no effect on the mean values for the parameter estimates but that it did affect the precision of the estimates, as judged by the 95% confidence intervals. Differences between their study and ours include that we used stricter confidence intervals, we tested the effect of performing multiple trials (rather than limiting analyses to data combined from all the trials that were performed), and we used data from heavy intensity exercise, rather than sub-threshold moderate intensity exercise. We fitted data to a two-component model, so that the effect of smoothing and the effect of number of trials was determined for primary phase and slow component parameters. Given the greater complexity of the response and greater number of parameters, compared to studies which used only a mono-exponential response, less precision might be expected around the parameter estimates generated in the present study. The good precision and accuracy that we report may reflect that our participants were familiar with exercise testing while breathing through a mouthpiece.

5.1. Conclusions

The accuracy and precision of estimates of the parameters of the primary and slow phases of the VO_2 response during heavy intensity exercise can be improved by using data from more than one test and by smoothing the data prior to fitting them to an appropriate mathematical model. Depending upon the accuracy required, that is depending upon how close each and every participant's value must be to his or her 'true' value, smoothed data from one or two tests is sufficient to calculate the values that describe the two-parameter VO_2 response profile in heavy intensity cycling exercise.

Acknowledgments

The contribution of Dr. Daniela Pedini to collection of data from the first fourteen research participants and the efforts of all participants made this study possible.

Footnotes

Authors' Contribution: The author was responsible for designing the study, collecting and analyzing the data, and preparing the manuscript.

Conflict of Interests: No conflict of interests.

Ethical Approval: The study was approved by the Institutional Review Board for the Protection of Human Subjects in Research at the University of North Texas.

Funding/Support: No funding.

Informed Consent: Participants participate and provided informed consent.

References

- Whipp BJ, Ward SA, Lamarra N, Davis JA, Wasserman K. Parameters of ventilatory and gas exchange dynamics during exercise. *J Appl Physiol Respir Environ Exerc Physiol.* 1982;52(6):1506-13. doi: [10.1152/jappl.1982.52.6.1506](https://doi.org/10.1152/jappl.1982.52.6.1506). [PubMed: 6809716].
- Krustrup P, Jones AM, Wilkerson DP, Calbet JA, Bangsbo J. Muscular and pulmonary O₂ uptake kinetics during moderate- and high-intensity sub-maximal knee-extensor exercise in humans. *J Physiol.* 2009;587(Pt 8):1843-56. doi: [10.1113/jphysiol.2008.166397](https://doi.org/10.1113/jphysiol.2008.166397). [PubMed: 19255119]. [PubMed Central: PMC2683969].
- Monod H, Scherrer J. The work capacity of a synergic muscular group. *Ergonomics.* 2007;8(3):329-38. doi: [10.1080/00140136508930810](https://doi.org/10.1080/00140136508930810).
- Hughson RL, Orok CJ, Staudt LE. A high velocity treadmill running test to assess endurance running potential. *Int J Sports Med.* 1984;5(1):23-5. doi: [10.1055/s-2008-1025875](https://doi.org/10.1055/s-2008-1025875). [PubMed: 6698679].
- Hill DW. The critical power concept. A review. *Sports Med.* 1993;16(4):237-54. doi: [10.2165/00007256-199316040-00003](https://doi.org/10.2165/00007256-199316040-00003). [PubMed: 8248682].
- Poole DC, Ward SA, Whipp BJ. The effects of training on the metabolic and respiratory profile of high-intensity cycle ergometer exercise. *Eur J Appl Physiol Occup Physiol.* 1990;59(6):421-9. doi: [10.1007/bf02388623](https://doi.org/10.1007/bf02388623). [PubMed: 2303047].
- Gaesser GA, Poole DC. The slow component of oxygen uptake kinetics in humans. In: Holloszy JO, editor. *Exercise and sport sciences reviews.* 24. Baltimore MD: Williams and Wilkins; 1996. p. 35-70. doi: [10.1249/00003677-199600240-00004](https://doi.org/10.1249/00003677-199600240-00004).
- Benson AP, Bowen TS, Ferguson C, Murgatroyd SR, Rossiter HB. Data collection, handling, and fitting strategies to optimize accuracy and precision of oxygen uptake kinetics estimation from breath-by-breath measurements. *J Appl Physiol (1985).* 2017;123(1):227-42. doi: [10.1152/japplphysiol.00988.2016](https://doi.org/10.1152/japplphysiol.00988.2016). [PubMed: 28450551].
- Keir DA, Murias JM, Paterson DH, Kowalchuk JM. Breath-by-breath pulmonary O₂ uptake kinetics: effect of data processing on confidence in estimating model parameters. *Exp Physiol.* 2014;99(11):1511-22. doi: [10.1113/expphysiol.2014.080812](https://doi.org/10.1113/expphysiol.2014.080812). [PubMed: 25063837].
- Lamarra N, Whipp BJ, Ward SA, Wasserman K. Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. *J Appl Physiol (1985).* 1987;62(5):2003-12. doi: [10.1152/jappl.1987.62.5.2003](https://doi.org/10.1152/jappl.1987.62.5.2003). [PubMed: 3110126].
- World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-4. doi: [10.1001/jama.2013.281053](https://doi.org/10.1001/jama.2013.281053). [PubMed: 24141714].
- Hill DW, Cureton KJ, Collins MA. Effect of time of day on perceived exertion at work rates above and below the ventilatory threshold. *Res Q Exerc Sport.* 1989;60(2):127-33. doi: [10.1080/02701367.1989.10607427](https://doi.org/10.1080/02701367.1989.10607427). [PubMed: 2489833].
- Hill DW. Morning-evening differences in response to exhaustive severe-intensity exercise. *Appl Physiol Nutr Metab.* 2014;39(2):248-54. doi: [10.1139/apnm-2013-0140](https://doi.org/10.1139/apnm-2013-0140). [PubMed: 24476482].
- Wasserman K, Whipp BJ, Koyl SN, Beaver WL. Anaerobic threshold and respiratory gas exchange during exercise. *J Appl Physiol.* 1973;35(2):236-43. doi: [10.1152/jappl.1973.35.2.236](https://doi.org/10.1152/jappl.1973.35.2.236). [PubMed: 4723033].

15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-10. [PubMed: [2868172](#)].
16. Krouwer JS. Why Bland-Altman plots should use X, not $(Y+X)/2$ when X is a reference method. *Stat Med*. 2008;27(5):778-80. doi: [10.1002/sim.3086](#). [PubMed: [17907247](#)].
17. Francescato MP, Cettolo V, Bellio R. Confidence intervals for the parameters estimated from simulated O₂ uptake kinetics: effects of different data treatments. *Exp Physiol*. 2014;99(1):187-95. doi: [10.1113/expphysiol.2013.076208](#). [PubMed: [24121286](#)].
18. Lamarra N. Variables, constants, and parameters: Clarifying the system structure. *Med Sci Sports Exerc*. 1990;22(1):88-95. [PubMed: [2304410](#)].
19. Chechile RA. Pooling data versus averaging model fits for some prototypical multinomial processing tree models. *Journal of Mathematical Psychology*. 2009;53(6):562-76. doi: [10.1016/j.jmp.2009.06.005](#).