RESEARCH ARTICLE

Data collection, handling, and fitting strategies to optimize accuracy and precision of oxygen uptake kinetics estimation from breath-by-breath measurements

Alan P. Benson,1,2 T. Scott Bowen,3 Carrie Ferguson,1,2 Scott R. Murgatroyd,4 and Harry B. Rossiter5,1

1School of Biomedical Sciences, University of Leeds, Leeds, United Kingdom; 2Multidisciplinary Cardiovascular Research Centre, University of Leeds, Leeds, United Kingdom; 3Heart Centre, University of Leipzig, Leipzig, Germany; 4Neurosciences Intensive Care Unit, Wexs Neurological Centre, University Hospital Southampton, Southampton, United Kingdom; and 5Rehabilitation Clinical Trials Center, Division of Respiratory and Critical Care Physiology and Medicine, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California

Submitted 4 November 2016; accepted in final form 19 April 2017

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 123: 227–242, 2017. First published April 27, 2017; doi:10.1152/japplphysiol.00988.2016.—Phase 2 pulmonary oxygen uptake kinetics (\(\dot{V}\)O\(_2\)P) reflect muscle oxygen consumption dynamics and are sensitive to changes in state of training or health. This study identified an unbiased method for data collection, handling, and fitting to optimize \(\dot{V}\)O\(_2\)P kinetics estimation. A validated computational model of \(\dot{V}\)O\(_2\)P kinetics and a Monte Carlo approach simulated \(2 \times 10^5\) moderate-intensity transitions using a distribution of metabolic and circulatory parameters spanning normal health. Effects of averaging (interpolation, binning, stacking, or separate fitting of up to 10 transitions) and fitting procedures (biexponential fitting, or \(\phi_2\) isolation by time removal, statistical, or derivative methods followed by monoeponential fitting) on accuracy and precision of \(\dot{V}\)O\(_2\)P kinetics estimation were assessed. The optimal strategy to maximize accuracy and precision of \(\tau\dot{V}\)O\(_2\)P estimation was 1-s interpolation of 4 bouts, ensemble averaged, with the first 20 s of exercise data removed. Contradictory to previous advice, we found optimal fitting procedures removed no more than 20 s of \(\phi_1\) data. Averaging method was less critical: interpolation, binning, and stacking gave similar results, each with greater accuracy compared with analyzing repeated bouts separately. The optimal procedure resulted in \(\phi_2\) \(\tau\dot{V}\)O\(_2\)P estimates for transitions from an unloaded or loaded baseline that averaged 1.97 ± 2.08 and 1.04 ± 2.30 s from true, but were within 2 s of true in only 47–62% of simulations. Optimized 95% confidence intervals for \(\tau\dot{V}\)O\(_2\)P ranged from 4.08 to 4.51 s, suggesting a minimally important difference of ~5 s to determine significant changes in \(\tau\dot{V}\)O\(_2\)P during interventional and comparative studies.

NEW & NOTEWORTHY We identified an unbiased method to maximize accuracy and precision of oxygen uptake kinetics (\(\tau\dot{V}\)O\(_2\)P) estimation. The optimum number of bouts to average was 4; interpolation, binning, and stacking averaging methods gave similar results. Contradictory to previous advice, we found that optimal fitting procedures removed no more than 20 s of phase 1 data. Our data suggest a minimally important difference of ~5 s to determine significant changes in \(\tau\dot{V}\)O\(_2\)P during interventional and comparative studies.

AT THE ONSET of constant power exercise below the lactate threshold (LT) in humans, mitochondrial oxidative phosphorylation and, subsequently, muscle oxygen uptake (\(\dot{V}\)O\(_2\)m) in activated muscle increase in a manner that is an approximate first-order exponential in vivo (2, 22, 48; cf. 30). The kinetics of phase (\(\phi\)) 2 of the pulmonary \(\dot{V}\)O\(_2\) (\(\dot{V}\)O\(_2\)P), characterized by the response time constant (\(\tau\)) from repeated breath-by-breath gas exchange measurements, are commonly used to infer \(\dot{V}\)O\(_2\)m kinetics and provide a noninvasive tool to investigate the control of exercise energetics (27, 41, 46). Fast \(\phi_2\) \(\dot{V}\)O\(_2\)P kinetics reflect effective cardiopulmonary and neuromuscular integration and are associated with high-endurance exercise performance (29, 38, 41), whereas \(\phi_2\) \(\dot{V}\)O\(_2\)P kinetics are slowed in the elderly (1) and with chronic disease (12, 23, 40, 46, 51). In addition, \(\phi_2\) \(\dot{V}\)O\(_2\)P kinetics are sensitive to interventions that influence blood flow distribution and muscle O\(_2\) delivery, muscle metabolism, or muscle recruitment (41, 46), making them a useful prognosticator (49) and method for evaluation of therapeutic benefit (44). Furthermore, the kinetics of \(\phi_1\) of the \(\dot{V}\)O\(_2\)P response (\(\phi_1\) duration and amplitude) are clinically discriminatory (50) and sensitive to age (37). Thus the strong link between \(\dot{V}\)O\(_2\)P kinetics and state of health provides the basis for an inherently attractive, noninvasive, and effort-independent method to characterize the efficacy of the integrated physiological systems response to exercise.

Although there are general guidelines for characterizing \(\dot{V}\)O\(_2\)P kinetics in terms of data collection, processing, and fitting procedures (56), a range of proposals exist for each of these steps (e.g., 10, 14, 19, 20, 26, 33, 39, 58). However, a systematic quantification of the effects of these different procedures on the precision and accuracy of the final \(\phi_1\) duration and amplitude and \(\phi_2\) \(\dot{V}\)O\(_2\)P characterization, as well as a standardization of these procedures, is lacking. This study therefore aimed to identify an unbiased (i.e., free from human error) method for \(\dot{V}\)O\(_2\)P data collection, handling, and fitting that allows the most accurate and precise estimation of \(\dot{V}\)O\(_2\)P kinetics. We identified this optimal criterion by sys-
tematically determining the influences of a range of common and uncommon collection, averaging, and fitting strategies on both the precision and accuracy of $\phi_1$ duration and amplitude and $\phi_2$ $\tau V_{O2P}$ estimation, using a validated cardiopulmonary simulation of exercise gas exchange (8) and a Monte Carlo approach.

THEORETICAL CONSIDERATIONS

The process linking $\dot{V}_{O2P}$ data collection in the laboratory or clinic, to kinetics characterization, is typically undertaken in three distinct steps: 1) data collection, 2) data processing, and 3) data fitting.

**Step 1: data collection.** Strategies employed in this step include identification of the optimal algorithms for calculating breath-by-breath gas exchange to improve signal-to-noise for kinetic fitting (6, 13, 14, 55). Strategies to improve primary $\dot{V}_{O2P}$ data also include the repetition of identical bouts of exercise with the intention of combining and averaging those data in the data processing step (Fig. 1B) (10, 26, 33, 57). The breath-by-breath fluctuations (also referred to as “noise”) inherent in any $\dot{V}_{O2P}$ measurement are uncorrelated (33) and have a Gaussian distribution in adults (although not in children; 42) with the standard deviation (SD) of this distribution ranging from ~30 to 110 ml/min (33, 47), independent of metabolic rate (33). What is less clear, however, is how different signal-to-noise ratios (or, analogously, the number of combined exercise bouts) affect $\dot{V}_{O2P}$ kinetics estimation and, therefore, whether there is an optimal number of exercise bouts required to estimate $\dot{V}_{O2P}$ kinetics to a given level of confidence.

**Step 2: data processing.** After the removal of outlying breaths generated by swallows or coughs or other “mistriggers” of the breath identification algorithms, and unrelated to tidal breathing [typically those breaths more than 3 or 4 SDs from the local mean (33, 57)], the second step involves averaging of the data collected from multiple exercise bouts to obtain a single (processed) $\dot{V}_{O2P}$ signal with a high signal-to-noise ratio, before kinetic characterization. Several averaging techniques are employed (Fig. 1, C–E), the most widely used involving some form of interpolation and/or averaging. Linear interpolation of data before averaging (commonly to 1-s intervals) is necessary to normalize gas exchange sampling frequency, from the nonuniform breath-by-breath sampling, and therefore ensure equal weighting of data among repeated trials (Fig. 1C) (57). Averaging may be in the form of postinterpolation ensemble averaging (56), or by arranging uninterpolated data from all bouts in time (10) before averaging the combined breaths into bins whose size depends on the number of averaged bouts (38) or time (9, 26) (Fig. 1D). This “binning” approach to averaging, although improving the signal-to-noise ratio, may help to maintain the density of the data close to that

--

Fig. 1. Example of data production and processing during a single Monte Carlo iteration from unloaded pedaling. A: for each iteration, model parameters were varied stochastically (see Table 1) and a clean model $\dot{V}_{O2P}$ trace with known kinetic parameters (e.g., $\phi_1$ duration and amplitude, and $\phi_2$ $\tau V_{O2P}$) was produced. Note that this clean model trace varied for each of the 10^4 Monte Carlo iterations. B: the single clean trace was used to produce 10 noisy “experimental” $\dot{V}_{O2P}$ traces (filled circles) with the sampling (breathing) and $\dot{V}_{O2P}$ noise characteristics seen in experimental data. Here, four examples are shown. Although each of the 10 noisy data sets is different, they have identical underlying kinetic parameters (known from the clean model trace shown in A, and shown in these panels as dashed lines). The noisy $\dot{V}_{O2P}$ data sets were processed in one of four ways: interpolation followed by ensemble averaging (C); bin averaging (D); stacking of data sets (E); and fitting of the separate traces before averaging of the resultant fit parameters (not shown). Fits to these processed data were compared with the true underlying kinetic parameters (known from the clean model trace shown in A, and shown in these panels as dashed lines).
at which it was collected (i.e., breathing frequency), and improve the validity of the estimated confidence intervals (21, 38). Despite the general popularity and acceptance of these approaches, several other data processing methods warrant investigation. Recent simulation studies have suggested that simple superimposition of all data from all bouts before fitting can give accurate \( \phi_2 \) \( \tau V_{O2P} \) estimates, with the added simplicity of reducing the requirement for complex data treatments (Fig. 1E) (19). Another alternative averaging approach, and maybe one that is statistically more robust (16) yet is not typically used for estimating \( V_{O2P} \) kinetics, involves fitting the individual exercise bouts, then averaging the resulting fit parameters (32). Kier et al. (26) showed that various stacking, interpolation, and bin or ensemble averaging procedures had essentially no effect on the precision of subsequent \( \tau V_{O2P} \) estimation. It remains unclear, however, how averaging strategies affect both the precision and accuracy of \( V_{O2P} \) kinetics estimation in the context of different numbers of averaged bouts and different approaches to fitting the data.

**Step 3: data fitting.** The third step involves the fitting of the processed \( V_{O2P} \) data to obtain an estimate of the kinetics of \( V_{O2} \). The \( V_{O2P} \) response to a step change in work rate in the moderate-intensity domain consists of an initial “cardiodynamic” phase (largely a result of increased blood flow through the pulmonary circulation; 56) followed by a “fundamental” phase, the kinetics of which closely represent those of \( V_{O2m} \) in young healthy adults (Fig. 1A) (22, 48). This entire response has been described mathematically using a piecewise biexponential equation of the form

\[
V_{O2P}(t) = V_{O2P,base} + A_1[1 - e^{-t/\tau_1}] + H(t)A_2[1 - e^{-(t-TD)/\tau_2}],
\]

\[
H(t) = \begin{cases} 
0, & t < TD \\
1, & t \geq TD 
\end{cases}
\]

where \( t \) is time, \( V_{O2P,base} \) is baseline \( V_{O2P} \), \( A_1 \) and \( A_2 \) are the amplitudes of the first and second phases of the response, \( \tau_1 \) and \( \tau_2 \) are time constants associated with each phase of the response, \( TD \) is a time delay, and \( H(t) \) is the Heaviside step function (cf. 36). Generally, the parameter of most interest is \( \tau_2 \), i.e., \( \phi_2 \tau V_{O2P} \). However, \( \phi_1 \) is a complex physiological construct, influenced by several processes including changes in mixed-venous gas tensions, pulmonary perfusion, and end-expiratory lung volume, which sum to generate a response that often deviates from a monoexponential (15, 55). In addition, there are several practical difficulties when using Eq. 1 to fit \( V_{O2P} \) data: Phase 1 typically contains only a few breaths (typically 5 or 6 in our simulations; see Fig. 1B), and fitting so few data points with the first exponential term in Eq. 1 drastically reduces the confidence of the parameter estimations in that first exponential term. The influence of this potentially unconfident \( \phi_1 \) fit continues into \( \phi_2 \), affecting \( \tau_2 \) (\( \phi_2 \tau V_{O2P} \)) estimation, particularly if the fit to the \( \phi_1 \) data does not reach a steady state before \( \phi_2 \) begins (i.e., at \( t = TD \)). Furthermore, most nonlinear least-squares algorithms used by data-fitting software (the Levenberg-Marquardt algorithm being the standard; 43) require the calculation of derivatives and cannot handle the Heaviside step function in Eq. 1; the parameters \( A_1 \) and \( \tau_1 \) are shared over, and influenced by the data in, the two different subdomains (\( t < TD \) or \( \phi_1 \), and \( t \geq TD \) or \( \phi_2 \)), and the extents of the subdomains themselves are determined by the parameter \( TD \). As such, fitting Eq. 1 is difficult without custom implementation of alternative, potentially less robust, nonlinear fitting algorithms such as direct search methods (35). As the parameter of most interest is the time constant of \( \phi_2 \), an alternative (and the most commonly used) approach is to isolate the \( \phi_2 \) data, then fit these data with a monoexponential equation of the form

\[
V_{O2P}(t) = V_{O2P,base} + A[1 - e^{-(t-TD)/\tau}].
\]

Such a monoexponential equation accurately describes the \( \phi_2 \) \( V_{O2P} \) response to moderate-intensity step exercise (4, 5) and can be handled by most nonlinear least-squares algorithms. If Eq. 2 is used to fit the \( V_{O2P} \) data and obtain an estimate of \( \phi_2 \) \( \tau V_{O2P} \), it is necessary to omit the \( \phi_1 \) data from the fit. The most widely used methods for removing \( \phi_1 \) data are empirically derived time-removal methods, where “at least” the first 20 s of data from the exercise transient are removed before fitting (7, 39, 54, 57). The rationale behind this strategy is that, because \( \phi_1 \) is expected to last less than 20 s and the \( \phi_2 \) \( V_{O2P} \) response is expected to be exponential, starting the fit from any given point past the \( \phi_1 \)–2 transition will yield an identical time constant that represents the underlying \( \phi_2 \) kinetics; whereas starting the fit from any point before the \( \phi_1 \)–2 transition will result in a larger (incorrect) time constant for \( \phi_2 \) (39, 54, 57). However, the \( \phi_2 \) \( V_{O2P} \) response is not truly exponential, but rather is a nonlinear distortion of a monoexponential \( V_{O2m} \) response (3, 5, 8, 25; cf. 18). Thus contrary to \( \tau V_{O2m} \), the \( \tau V_{O2P} \) is not a “true” constant throughout the transient, and fitting an exponential equation from different points in such a nonexponential \( \phi_2 \) will yield varying values for \( \tau V_{O2P} \); progressively larger values as the fit is started from later in \( \phi_2 \) (cf. 8). Such behavior is suggested in the empirical results of Murias et al. (39) where \( \tau V_{O2P} \) becomes larger as the imposed exponential fit is started from later in the exercise transient, at least in older adults. Although \( \tau V_{O2P} \) is influenced by a complex interaction of circulatory and gas exchange responses to exercise, and \( \phi_2 \) \( V_{O2P} \) is not quite exponential, a monoexponential fit of moderate-intensity \( V_{O2P} \) kinetics remains a useful, concise, and effort-independent method to characterize the integrated dynamic responsiveness of cardiopulmonary and neuromuscular health. Nevertheless, it seems crucial that all data contained in the \( \phi_2 \) response, but none of the \( \phi_1 \) data, are fitted to obtain the most accurate characterization of \( V_{O2P} \) kinetics (57). As such, accurate identification of the \( \phi_1 \)–2 transition is paramount.

When using the monoexponential Eq. 2 to fit \( V_{O2P} \) data, human error in selecting the \( \phi_1 \)–2 transition can lead to an unintended bias in \( \tau V_{O2P} \) estimation, and so an ideal, unbiased method for isolating \( \phi_2 \) data for such a fit would be based on either I) identification of some consistent time period (rather than leaving the choice to the individual researcher) at the start of exercise during which data should be removed, or 2) some other information in the data itself that could algorithmically identify the \( \phi_1 \)–2 transition.

Rather than employing empirical time-removal methods, the abrupt change in \( V_{O2P} \) at the \( \phi_1 \)–2 transition may be identifiable from the \( V_{O2P} \) data using either the peak time derivative of the \( V_{O2P} \) data (34) or statistical measures reflecting the best confidence in the fit parameters [e.g., the smallest confidence
interval of the obtained time constant (48)]. Although theoretically sound, in that both methods can identify abrupt changes in a continuous signal, their application to experimental VO$_2$P data may be hindered by the low sampling rate (relative to the duration of $\phi 1$) and noise inherent in those data. Whether the use of derivatives or statistical methods to identify the $\phi 1$–2 transition results in improved $\tau$VO$_2$P estimates over the empirical time-removal methods currently favored remains to be investigated.

Several studies have examined the effects of the different strategies employed in the three steps described above on the confidence of VO$_2$P kinetic parameter estimates using experimental data [e.g., $\phi 1$–2 transition and $\phi 2$ $\tau$VO$_2$P (10, 26, 39, 54)]. However, a limitation of such studies is that the true underlying VO$_2$P kinetic parameters are unknown: such experimental methods can therefore give an indication of the precision of VO$_2$P kinetics estimation but not of its accuracy. Computational approaches using Monte Carlo methods (17) can overcome some of these limitations. For this, a simulation is first used to produce a clean, continuous VO$_2$P trace with known kinetic parameters. This trace is then sampled using simulations of breathing frequency, and Gaussian noise is added (using known characteristics) to produce a data set with similar sampling, noise, and kinetic characteristics as experimentally obtained VO$_2$P data, but where the underlying VO$_2$P kinetic parameters are known (33). In addition, the same clean trace can be randomly resampled and new noise added to produce further noisy data sets (but all with the same underlying kinetic parameters), analogous to obtaining experimental VO$_2$P data during repeated bouts of exercise from a single subject. Thus these Monte Carlo methods allow both the precision and accuracy of VO$_2$P fitting methods to be systematically assessed.

Computational approaches have been previously applied using a simple delayed monoeponential (19, 20) or a biexponential (10, 33) VO$_2$P response generated in silico. However, as the underlying VO$_2$P kinetics do not follow a simple mono- or biexponential time course (3, 5, 8), it is necessary to use a validated simulation of VO$_2$P kinetics that takes into account how circulatory dynamics modulate the monoexponential VO$_2$P response to produce the $\phi 1$ and $\phi 2$ VO$_2$P responses (8). Such computationally produced data sets can therefore contain the influence of normal variation in the steady states and kinetics of, for example, cardiac output, muscle blood flow, and VO$_2$P, to derive a distribution of VO$_2$P characteristics (including $\phi 1$ duration and amplitude, and $\phi 2$ $\tau$VO$_2$P), analogous to collecting experimental VO$_2$P data from a large number of healthy human subjects.

**METHODS**

We used a validated simulation of VO$_2$ and circulatory dynamic interactions during moderate-intensity cycling exercise in humans (8) that accounts for the vascular capacitances and circulatory dynamics that cause a monoexponential VO$_2$P response to manifest at the lungs as a three-phase VO$_2$P response, with a cardiodynamic $\phi 1$, a near-exponential fundamental $\phi 2$, and a steady-state $\phi 3$. The simulation VO$_2$P outputs initially have no noise, so the baseline VO$_2$P steady state, $\phi 1$ duration and amplitude, $\phi 2$ $\tau$VO$_2$P, and $\phi 3$ VO$_2$P steady state for each output are precisely known. This allows quantification of both the accuracy and the precision of subsequent fits to the data.

Data production. The minimum required number of Monte Carlo iterations, $n$, was estimated from the central limit theorem (17) using $n = \left(\frac{z_{\alpha/2} \sigma}{\varepsilon}\right)^2$, where $z_{\alpha/2}$ is the $z$ score associated with significance level $\alpha$, $\sigma$ is the estimated SD of the simulation output, and $\varepsilon$ is the acceptable margin of error for the simulation output (equal to half the required confidence interval). We set $\alpha$ at 0.05 to give $z_{0.025} = 1.96$; it was assumed that the SD of $\phi 2$ $\tau$VO$_2$P (our parameter of interest) produced by stochastic simulations would be 4.3 s [based on the experimental data used to parameterize the simulations (8, 22)]; and the acceptable margin of error was set at 0.1 s (the same as the simulation time resolution). This predicted a minimum iteration number of $n = 7,104$; we therefore performed $10^4$ iterations during the Monte Carlo simulations.

We examined two protocols for a step increase in work rate (WR), both constrained to be within the moderate-intensity exercise domain: the first from unloaded pedaling (UP-WR) and the second from a raised baseline (WR-WR). For each of these two protocols, $10^5$ clean (time resolution = 0.1 s) VO$_2$P simulations, each with different kinetics, were produced (see Fig. 1A for an example). The start of the step increase in WR was set to $t = 0$ s. Simulation input parameters were varied stochastically (43) using distributions taken from the data of Grassi et al. (22) and Benson et al. (8) (Table 1). This provided simulations with normal physiological variation in, for example, baseline VO$_2$, VO$_2$P gain ($\Delta$VO$_2$/$\Delta$W), the relative increase in cardiac output ($\Delta$Q$_m$/$\Delta$VO$_2$), and the kinetics of cardiac output and VO$_2$P ($\tau_{Q_m}$/$\tau_{VO_2}$). Parameter sets that resulted in venous O$_2$ con-

---

**Table 1. Distributions of model input parameters for Monte Carlo simulations**

| Parameters | Mean | SD |
|------------|------|----|
| Arterial O$_2$ concentration, ml O$_2$/100 ml blood | 20.0 | 1.00 |
| Total venous volume, liters$^b$ | 3.07 | 0.61 |
| Baseline VO$_2$P, l/min | 0.87 | 0.08 |
| Fraction of baseline VO$_2$P from muscle$^b$ | 0.57 | 0.11 |
| Baseline Q$_m$, l/min | 8.89 | 0.44 |
| Fraction of baseline Q$_m$ to muscle$^b$ | 0.57 | 0.08 |
| $\Delta$VO$_2$/$\Delta$W, ml·min$^{-1}$·W$^{-1}$ | 9.47 | 0.85 |
| $\Delta$Q$_m$/$\Delta$W, ml·min$^{-1}$·W$^{-1}$ | 11.04 | 1.36 |
| $\Delta$Q$_m$/$\Delta$VO$_2$ | 6.03 | 0.53 |
| $\tau_{Q_m}$/$\tau_{VO_2}$ | 1.08 | 0.08 |

---

$^a$Q denotes blood flow (with $Q_{tot}$ denoting cardiac output). Subscript m denotes muscle compartment; subscript b denotes rest-of-body compartment. Baseline is unloaded pedaling (i.e., 0 W). See Benson et al. (8) for a detailed description of the model. Gaussian distributions were calculated from the data of Grassi et al. (22) and Benson et al. (8). Linear distributions were set for this study. The ratio of the muscle, body, and mixed-venous volumes was maintained as in the default model; only the total venous volume was altered. $^b$The remainder of the baseline VO$_2$P (and $Q_{o_2}$) comes from (and goes to) the body compartment. $^c$To avoid kinetic mismatch between muscle Q and VO$_2$ (as occurs with slow $Q_{o_2}$, but fast VO$_2$, kinetics, that result in muscle O$_2$ concentration dropping to zero), we first set the absolute VO$_2$P value, and then constrained $\tau_{Q_m}$ to be similar to $\tau_{VO_2}$ using this ratio. $^d$Baseline WR for UP-WR simulations was fixed at 0 W. $^e$ΔWR was constrained to be positive, and the final WR in the WR-WR simulations (i.e., baseline WR + ΔWR) was constrained to be no greater than 150 W.
variation of key parameters across bouts (“separate”).

These procedures produced, from the clean simulation output, a trace with the sampling, noise, and kinetic characteristics observed in experimentally collected data (see Fig. 1B for examples). For all $2 \times 10^4$ clean simulations, this sampling and noise procedure was performed 10 times to simulate 10 bouts of exercise repeated by a single subject (see Fig. 1, A and B, for examples). At the end of this Monte Carlo procedure, we therefore had $10^4$ noisy UP-WR data sets, i.e., $10^4$ “subjects,” each with different physiological characteristics, who performed moderate-intensity step exercise from unloaded pedaling: each data set contained 10 noisy traces from separate “exercise bouts,” i.e., each subject performed the same WR protocol 10 times. A further $10^4$ noisy WR-WR data sets, with each data set again containing 10 traces from separate exercise bouts, were produced. Thus a total of $2 \times 10^4$ moderate-intensity “exercise bouts” in $2 \times 10^4$ “subjects” were produced, which sampled the normal variation of key parameters observed in healthy young humans. Note that, despite the sampling and noise procedure used to produce the data, the true underlying kinetic characteristics of any given noisy trace were known from the kinetics of the original clean simulation from which it was produced.

Data processing. Outlying breaths were first removed by fitting Eq. 2 to the noisy traces and removing breaths that lay further than 3 SDs away from the local mean (i.e., outside the 99.7% prediction bands of the fit) (33). For each data set, we used the following data processing techniques, covering a range of commonly used or potentially useful methods, to process up to 10 bouts of noisy data (see Fig. 1 for examples): 1) interpolation of each bout to 1-s intervals before ensemble averaging across bouts (“interpolated”); 2) time alignment of data from the bouts to be averaged, before bin averaging into bins whose size depends on the number of bouts being averaged (“binned”); 3) superimposition, or stacking, of the data from different bouts, with no further interpolation or averaging (“stacked”); 4) fitting of individual bouts (see below) followed by averaging of fit parameters across bouts (“separate”).

Data fitting. For each processed VO$_{2P}$ trace, we fit the biexponential Eq. 1 to the entire φ1 and φ2 data, and used the following strategies for identification of the φ1–2 transition and subsequently fit the monoexponential Eq. 2 to the isolated φ2 data: 1) empirical time-removal methods, where 10, 15, 20, 25, or 30 s of data was removed from the beginning of each processed VO$_{2P}$ trace; 2) use of VO$_{2P}$ time derivatives on both unsmoothed and smoothed (with a moving 5-breadth average) processed data, where the highest derivative of VO$_{2P}$ with respect to time during the first 60 s of exercise was taken as the φ1–2 transition; 3) statistical methods to identify the φ1–2 transition, where a datum was incrementally removed from the beginning of each data set (until 60 s into exercise) and the remaining data were fit using the monoexponential Eq. 2; the reduced chi-squared ($\chi^2_{\text{red}}$), adjusted coefficient of determination ($R^2$), confidence interval for the time constant (CL), and the corrected Akaike information criterion (AICC) were then calculated for each fit (42, 46). The first datum in the fit that returned the minimum statistical value (or maximum for $R^2$) was taken as the identified φ1–2 transition for that statistical method [see Rossiter et al. (48) for an example using CL, to identify the φ1–2 transition]. For each processed trace we therefore obtained 12 fits to the data: one using the biexponential fit to the entire φ1 and φ2 data, and 11 using a monoexponential fit to isolated φ2 data (five using empirical time removal methods, two using VO$_{2P}$-time derivatives, and four using statistical measures). As a control condition, for each processed trace we also fit the true isolated noisy φ2 data with Eq. 2, i.e., the data were fit beginning at the true first “breath” in φ2, known from the clean simulation. Each of these 13 methods provided an estimate of the φ1–2 transition (i.e., TD from Eq. 1 when using the biexponential fit, or the identified first breath in φ2 when using the monoexponential fits) and an estimate of φ2 τVO$_{2P}$ (i.e., $\tau_2$ from fits using Eq. 1, or $\tau$ from fits using Eq. 2). The φ1 amplitude (as a percentage of the steady-state response) was estimated from the value of the fit at the identified φ1–2 transition. Each of the φ1–2 transition, φ1 amplitude, and φ2 τVO$_{2P}$ estimates were then compared with the known true underlying values obtained from the clean simulated VO$_{2P}$ trace. These true values represent the most accurate estimates possible of φ1 and φ2 VO$_{2P}$ kinetics.

Numerical methods and statistical analyses. Details of the model used to produce the clean VO$_{2P}$ data, along with numerical methods, are given in Benson et al. (8). Because of its unique piecewise nature, Eq. 1 was fit using a custom direct search method (35), although this precluded calculation of parameter confidence intervals. Equation 2 was fit using the Levenberg-Marquardt algorithm (43). Values are presented as means ± SD unless otherwise stated. Significant differences between data were tested for using two-sample t-tests, or one-way repeated-measures analysis of variance (ANOVA) with Tukey’s post hoc tests, as appropriate. Significance level was set at $P < 0.05$.

RESULTS

Simulation outputs. Simulation input WR and output VO$_{2P}$ characteristics are summarized in Table 2. Time of the φ1–2 transition was significantly different between UP-WR and WR-WR simulations (19.8 ± 3.4 vs. 15.9 ± 3.4 s, respectively; $P < 0.05$, t-test), as was φ1 amplitude (reported as percentage of the steady-state response: 28.2 ± 8.3% vs. 28.9 ± 7.7%, respectively; $P < 0.05$, t-test) and φ2 τVO$_{2P}$ (22.4 ± 7.2 vs. 25.0 ± 7.2 s, respectively; $P < 0.05$, t-test). These different VO$_{2P}$ characteristics from UP-WR and WR-WR simulations were found.
WR-WR protocols can be explained by the increased baseline cardiac output associated with starting an exercise transition from a raised WR: muscle-to-lung transit time is shortened, reducing δ1 duration (3), and the altered blood flow during the exercise transient modifies the association between muscle and pulmonary VO2 kinetics (8). The Monte Carlo simulation output data (10⁴ clean UP-WR traces and 10⁴ clean WR-WR traces, along with the corresponding 2 × 10⁵ noisy traces, and details of the input and output characteristics for each simulation) are available from the corresponding author upon request.

The results below present in detail the findings for UP-WR simulations. The key differences between UP-WR and WR-WR simulations are then presented. For the sake of brevity, we present only data pertinent to our significant findings.

**Number of averaged exercise bouts.** Figure 2 shows the effects of averaging exercise bouts on the precision and accuracy of δ2 τ̂vo2p estimation (generally the parameter of most interest) during UP-WR simulations. For this example, data from different bouts were interpolated to 1-s intervals then ensemble averaged (see Averaging methods below), and fitting was made beginning at the known first breath in δ2 (i.e., control fits). Qualitatively similar results were found for the other averaging and fitting methods. The mean and SD of the estimated δ2 τ̂vo2p are shown in Fig. 2A, and example distributions of the estimated δ2 τ̂vo2p for 1, 4, and 10 exercise bouts are shown in Fig. 2B. The δ2 τ̂vo2p estimates obtained by averaging 1, 2, or 3 bouts were significantly greater than using 10 bouts (P < 0.05, ANOVA). This indicates that precision and accuracy of δ2 τ̂vo2p estimation is not statistically improved by averaging data from more than four bouts of exercise.

Figure 2, A and B, demonstrates that τ̂vo2p tends to be overestimated on average by ~2 s, irrespective of the number of bouts averaged: mean difference between estimated and true τ̂vo2p was 1.92 ± 4.24 s with 1 bout, 1.68 ± 2.06 s with 4 bouts, and 1.62 ± 1.37 s with 10 bouts. Figure 2C shows the percentage of estimated δ2 τ̂vo2p values that lay within ±2 s of true. Using data from a single exercise bout, the estimated δ2 τ̂vo2p was within 2 s of the true value in only 41.3% of cases. When 4 bouts were averaged, the percentage of estimated values within 2 s of the true value increased to 53.0%, even when the first breath in δ2 is known precisely (see also Data fitting and kinetic characterization). The asymptote of this relationship is 62.0% (Fig. 2C), indicating that the maximum probability of returning a δ2 τ̂vo2p estimate within 2 s of true is 62%, even when the first breath in δ2 is known and no matter how many bouts are averaged.

**Averaging methods.** Figure 3A shows the effects on δ2 τ̂vo2p estimation of the different averaging methods during UP-WR simulations. For the example shown, data from four exercise bouts were averaged and fitting was from the known first breath in δ2 (i.e., control fits). Qualitatively similar results were found for other numbers of averaged bouts and for the other fitting methods. Each averaging method returned significantly different δ2 τ̂vo2p estimates (P < 0.05, ANOVA), although the mean δ2 τ̂vo2p values obtained using the interpolated, binned, and stacked averaging methods were quantitatively very similar, being within 0.1 s of each other (i.e., within the acceptable margin of error set for our Monte Carlo simulations). Mean δ2 τ̂vo2p estimation with the interpolation method was 1.68 ± 2.06 s from true (53.0% of values within ± 2 s of true), compared with 1.76 ± 2.17 s (50.7%) for binned, 1.72 ± 2.13 s (51.4%) for stacked, and 2.04 ± 2.34 s (46.9%) for separate. The distribution of the confidence intervals of the estimated δ2 τ̂vo2p are shown in Fig. 3B. Each averaging method returned a significantly different confidence interval distribution (P < 0.05, ANOVA), although the confidence interval distributions for the binned and stacked averaging methods were quantitatively similar (the difference between the means of these two distributions was 0.14 s).

**Data fitting and kinetic characterization.** Figures 4–6 compare the different methods for estimating the δ1–2 transition (Fig. 4), and the subsequent estimation of δ1 amplitude (Fig. 5) and δ2 τ̂vo2p (Fig. 6), during UP-WR simulations. In Figs. 5 and 6, the distributions of δ1 amplitude and δ2 τ̂vo2p estimates obtained from control fits (i.e., fits from the known first δ2 breath) are shown as dashed curves. The examples shown use data from four bouts averaged using the interpola-

---

**Fig. 2.** Effects of the number of averaged bouts on the precision and accuracy of τ̂vo2p estimation, using control fits (i.e., using the known δ2 data) to interpolated and ensemble averaged UP-WR data: A: mean ± SD difference of the estimated τ̂vo2p from the true value, for 1–10 averaged bouts. Horizontal lines show zero difference (solid) ± 2 s (dashed) from true. n = 10⁴ in each case. *P < 0.05 vs. 10 averaged bouts (ANOVA). B: distributions of the difference between estimated and true τ̂vo2p for 1, 4, and 10 averaged bouts. Vertical lines show zero difference (solid) ± 2 s (dashed) from true. n = 10⁴ in each case. C: percentages of the 10⁴ τ̂vo2p estimates within ±2 s of true, for 1–10 averaged bouts. The solid line is an exponential fit to the data.
tion method, although qualitatively similar results were found for other numbers of averaged bouts and for the other averaging methods. Only removal of the first 20 s of data (Figs. 4B, 5B, and 6B) resulted in the accurate identification of the first breath in $\phi_2$, and $\phi_1$ amplitude and $\phi_2 \tau V\dot{O}_{2P}$ values that were not significantly different from the control fits; all other methods were significantly different from true ($P < 0.05$, ANOVA). Using this empirical 20 s removal method, the identified $\phi_1$–2 transition was within ±2 breaths of true in 99.3% of cases, estimated $\phi_1$ amplitude was within ±5% of true in 32.6% of cases (vs. 34.2% with control fits), and estimated $\phi_2 \tau V\dot{O}_{2P}$ was within ±2 s of true in 46.5% of cases (vs. 53.0% with control fits). Although the biexponential fitting method (Figs. 4A, 5A, and 6A) returned the second best estimates of the $\phi_1$–2 transition (93.8% of estimates within ±2 breaths of true), the overparameterization of the model resulted in less accurate and precise $\phi_2 \tau V\dot{O}_{2P}$ estimates (only 32.0% of estimates within ±2 s of true) than both the empirical 15 s and 25 s removal methods (37.9% and 37.6%, respectively) (Figs. 4B, 5B, and 6B). Interestingly, removal of 15 s of data (i.e., including some $\phi_1$ data in the fit) gave more accurate and precise $\phi_1$ amplitude and $\phi_2 \tau V\dot{O}_{2P}$ estimates than removal of 25 s of data (i.e., excluding the initial portion of $\phi_2$ data). Basing $\phi_1$–2 identification on time-derivative or statistical methods resulted in skewed distributions (Fig. 4, C and D), and $\phi_1$–2 transition, $\phi_1$ amplitude, and $\phi_2 \tau V\dot{O}_{2P}$ values that were furthest from true (Fig. 5, C and D, and Fig. 6, C and D).

Optimal protocol. Having identified that removal of the first 20 s of data, followed by a monoexponential fit to the isolated $\phi_2$ data, was the optimal fitting method for UP-WR transitions, we repeated the previous analyses that were performed on the control, i.e., known $\phi_2$, data (as shown in Figs. 2 and 3) using this empirical 20 s removal fitting method (Fig. 7). Qualitatively, the results were identical, in that four averaged bouts provided no more accuracy and precision than 10 averaged bouts, and the interpolated averaging method gave the most accurate and precise $\phi_1$–2 transition, $\phi_1$ amplitude, and $\phi_2 \tau V\dot{O}_{2P}$ estimates, that were not significantly different from the control fits. Quantitatively, the mean estimate of the $\phi_1$–2 transition was 0.06 ± 0.85 breaths from true, with 99.3% of values within ±2 breaths of true; the mean $\phi_1$ amplitude estimate was 6.63 ± 10.61% from true (vs. 6.65 ± 4.46% from true with control data), with 32.6% of values within ±5% of true (vs. 34.2% with control fits); and the mean $\phi_2 \tau V\dot{O}_{2P}$ estimate was 1.97 ± 2.08 s from true (vs. 1.68 ± 2.06 s from true with control data), with 46.5% of estimates within ±2 s of true (vs. 53.0% with control fits). Again, the binned and stacked averaging methods gave very similar (but slightly less precise and accurate) $\phi_2 \tau V\dot{O}_{2P}$ estimates to the interpolated method: 2.00 ± 2.19 and 1.98 ± 2.16 s from true, respectively. Using the optimal methods, the asymptote of the exponential fit to the proportion of $\phi_2 \tau V\dot{O}_{2P}$ estimates within ±2 s across all numbers of averaged bouts (Fig. 7C) was 51.3%.

WR-WR simulations. The analyses performed for the UP-WR simulations (Figs. 2–7) were repeated for the WR-WR simulations, where “exercise” was initiated from a raised baseline WR between 0 and 100 W. These analyses are summarized in Fig. 8. As with UP-WR simulations, averaging of four bouts (Figs. 8, A–C), using interpolated, binned, or stacked data, optimized $\phi_2 \tau V\dot{O}_{2P}$ estimation while minimizing the number of required bouts (Fig. 8, D and E). However, for WR-WR data, removal of the first 15 or 20 s of data gave statistically similar results to control fits (where the first breath in $\phi_2$ is known), although qualitatively the removal of 15 s of data gave more precise and accurate estimates of $V\dot{O}_{2P}$ kinetics than removing 20 s of data: 97.2% (with 15 s removal) vs. 93.1% (with 20 s removal) of the $\phi_1$–2 transition estimates.

![Fig. 3. Effects of averaging method on the precision and accuracy of $\tau V\dot{O}_{2P}$ estimation, using control fits (i.e., using the known $\phi_2$ data) to four averaged UP-WR bouts.](image-url)
within \pm 2 \text{ breaths of true}; 41.5\% \text{ vs. } 16.9\% \text{ of } \phi_1 \text{ amplitude values within } \pm 5\% \text{ of true; and } 61.9\% \text{ vs. } 57.6\% \text{ of } \phi_2 \tau \dot{V}O_{2P} \text{ values within } \pm 2 \text{ s of true (Fig. 8F). Phase 2 } \tau \dot{V}O_{2P} \text{ estimation was more accurate for WR-WR data than for UP-WR data: using four interpolated and ensemble averaged bouts with } \phi_2 \text{ isolated by removal of the first 15 s of data, the mean difference between estimated and known } \tau \dot{V}O_{2P} \text{ was } 1.04 \pm 2.30 \text{ s (vs. } 1.97 \pm 2.08 \text{ s with the optimal UP-WR analysis; } P < 0.05, \ t\text{-test) and the percentage of values lying within } \pm 2 \text{ s of the true value was } 61.9\% \text{ (vs. } 46.5\% \text{ with UP-WR data). The asymptote of the exponential fit to these data (Fig. 8C) suggested that a maximum of } 75.9\% \text{ of } \phi_2 \tau \dot{V}O_{2P} \text{ values would lie within } \pm 2 \text{ s of the true value (vs. } 51.3\% \text{ for UP-WR data).}

\textit{Minimally important difference.} The optimal collection, handling, and fitting procedures for UP-WR and WR-WR simulations were used to determine the minimally important difference for significant changes in } \tau \dot{V}O_{2P} \text{ during moderate-intensity exercise. Table 3 shows that the 95\% confidence limits of } \tau \dot{V}O_{2P} \text{ estimation narrows from 8.25 to 4.08 s for UP-WR, and from 9.43 to 4.51 s for WR-WR, as the number of bouts averaged is increased from 1 to 4. These data propose a minimal important difference of } \sim 5 \text{ s to detect differences in...
Robustness of Monte Carlo simulations. To confirm the robustness of the Monte Carlo simulations, the entire data production procedure was repeated (i.e., a second set of $10^4$ UP-WR and $10^4$ WR-WR clean simulations was produced, and noise was added to each trace 10 times, to give $2 \times 10^5$ noisy traces) and these data were analyzed as described above. There were no differences in the key findings with this second set of simulations (data not shown). As with the original Monte Carlo data, the output data from this second set of Monte Carlo simulations ($2 \times 10^4$ clean and $2 \times 10^5$ noisy traces, along with simulation input and output characteristics) are available from the corresponding author upon request.

DISCUSSION

We used a validated computational model together with a Monte Carlo approach to produce $2 \times 10^5$ simulated $\dot{V}O_2P$ data sets with similar sampling, noise, and kinetic characteristics as experimentally obtained $\dot{V}O_2P$ data. As the true underlying $V_{O2P}$ kinetic parameters of these data sets were known from the clean simulation traces from which they were produced, we
could assess both the accuracy and the precision of various averaging and fitting procedures on the estimation of $\tau V\dot{O}_2P$, something that is not feasible using experimentally obtained data where the true underlying $\tau V\dot{O}_2P$ is not known. We showed that the optimal data handling steps to give the most accurate and precise estimation of $\tau V\dot{O}_2P$ were linear interpolation with ensemble averaging data from four bouts of exercise, followed by removal of the first 20 s (if exercise was from unloaded pedaling) or 15 s (if exercise was from a raised work rate) of data before monoexponential fitting of the isolated $\phi 2$ data. Variations on the averaging method led to substantially similar results, with the exception that the confidence interval for kinetic estimation was significantly wider for the technique of independently fitting repeats of the same exercise transition (the separate method). This suggests that different data processing techniques currently used among different laboratories is unlikely to substantially influence the derived parameters. However, it is of note that even the optimal procedures that we identified yielded $\tau V\dot{O}_2P$ estimates that were within 2 s of true in just 47% of simulations from unloaded pedaling, rising to only 62% for protocols where exercise started from a raised work rate.

Fig. 6. Effects of fitting methods on the precision and accuracy of $\phi 2 \tau V\dot{O}_2P$ estimation. Shown are distributions of the difference between the estimated and true $\phi 2 \tau V\dot{O}_2P$ for the biexponential fit (A), and empirical (B), statistical (C), and derivative (D) $\phi 2$ isolation methods. The control fit distribution (i.e., using the known $\phi 2$ data) is shown as a dashed curve in each panel. For all panels, vertical lines show zero difference (solid) ± 2 s (dashed) from true, and $n = 10^4$ in each distribution. Note the different scales on the abscissas.


Data collection. The simulated data of exercise transitions either from unloaded pedaling or from a raised work rate spanned a wide range of variable and parameter estimates expected for subLT exercise (Table 2). Simulated duration ranged from 7 to 30 s and was 9% to 72% of the steady-state response in amplitude, and simulated \( \dot{V}O_2P \) spanned 7 to 40 s, across transitions ranging from 50 to 150 W in amplitude, making our findings widely generalizable to the study of moderate-intensity \( \dot{V}O_2P \) kinetics in healthy adults. We showed that averaging data from four exercise bouts optimized accuracy and precision of \( \dot{V}O_2P \) estimation, while minimizing experimental burden, regardless of the averaging or fitting methods subsequently used. Averaging more bouts did not give a significantly more precise or accurate estimation of \( \dot{V}O_2P \).

Some investigators may be willing to accept lower accuracy and precision of \( \dot{V}O_2P \) estimation to reduce the testing burden of four exercise bouts. For example, interpolating and averaging three bouts of UP-WR exercise, and removing 20 s of data to isolate \( \phi_2 \), resulted in \( \dot{V}O_2P \) estimates that were 2.00 ± 2.39 s from true, with 45.0% of these estimations within 2 s of true, a relatively small reduction in accuracy and precision compared with the same data handling method with four exercise bouts (1.97 ± 2.08 s and 46.5%). These differences are associated with an increase in the minimal detectable difference for \( \dot{V}O_2P \), e.g., for use in comparative and interventional studies, from ~5 to ~6 s. The data shown in Table 3 can be used to inform such decisions.

Our four-bout data collection recommendation is only applicable to data that have similar breath-by-breath fluctuation characteristics as the data produced in our simulation studies (68 ± 26 ml/min). Nevertheless, our simulated transitions mimicked very well typical observations using many standard gas exchange measurement approaches. Our findings indicate that to provide more precise estimations of \( \dot{V}O_2P \) from experimental data, strategies should focus not on averaging additional exercise bouts, but on increasing the signal-to-noise ratio in the collected data. These findings echo those of Lamarra et al. (33), who also used a Monte Carlo approach to show that increasing \( \dot{V}O_2P \) noise, expressed as a percentage of the steady-state change in the \( \dot{V}O_2P \) response, increased the confidence intervals for the estimated fit parameters (\( \phi_1 \) duration and \( \phi_2 \) \( \dot{V}O_2P \)). We showed that approaches that increase the signal-to-noise ratio have a substantial effect on precision, but little effect on accuracy, of kinetic estimates. These fluctuations are expected to arise from the interaction of a number of variables, not least the breath-by-breath variations in tidal volume and pulmonary blood flow, within which fluctuation and timing of stroke volume and thoracic pressure changes may variably sum or counteract one another to give rise to fluctuations in gas exchange. Therefore, algorithms for breath-by-breath gas ex-
change measurement that reduce the inherent fluctuation of the data, e.g., by accounting for changes in alveolar gas storage, or by recharacterizing a breath to be equal to a tidal breathing cycle that returns to an identical end-expiratory lung volume (6, 13), would be expected to further reduce the testing burden while maintaining optimal precision and accuracy of kinetic estimates.

Data processing. Although there are many possible methods for data averaging, the four techniques examined in this study (interpolation, binning, stacking, and separate fitting) provide a cross-section of the most commonly used methods. Although we have identified linear interpolation followed by ensemble averaging as the optimal method for averaging data [similar to the findings of Keir et al. (26)], both the breath binning and stacking methods produced quantitatively similar estimates of $\dot{V}O_2p$. As such, researchers who have previously used, or currently use, any of these methods should be confident that their choice of averaging procedure does not unduly influence their estimates of $\dot{V}O_2p$. Although averaging of the exponential fit parameters from separate bouts of exercise offers the

![Fig. 8. Precision and accuracy of $\dot{V}O_2p$ estimation for WR-WR bouts. A–C: effects of the number of averaged bouts, where data processing is by interpolation and ensemble averaging and removal of the first 15 s of data is used to isolate $\dot{V}O_2p$ (see Fig. 2 for explanations; *P < 0.05 vs. 10 averaged bouts, ANOVA). D and E: effects of averaging method on $\dot{V}O_2p$ estimation and the associated confidence interval, using four bouts and where removal of the first 15 s of data is used to isolate $\phi 2$ (see Fig. 3 for explanations). F: effects of empirical $\phi 2$ isolation methods, using four interpolated and ensemble averaged bouts (see Fig. 4 for explanations).](image)

Table 3. Phase 2 $\tau V O_2p$ estimates and confidence intervals for 1-4 averaged UP-WR and WR-WR exercise bouts

| No. of Averaged Bouts | $\phi 2 \cdot \dot{V}O_2p$ Estimation | Percentage of Values within 2 s of True | 95% Confidence Limits, s from mean |
|-----------------------|-------------------------------------|---------------------------------------|-----------------------------------|
|                       | Mean, s from true                   | SD, s                                 |                                   |
| UP-WR simulations ($n = 10^4$) |                                      |                                       |                                   |
| 1                     | 2.21                                | 4.21                                  | 38.27                             | 8.25                              |
| 2                     | 2.03                                | 2.90                                  | 43.32                             | 5.68                              |
| 3                     | 2.00                                | 2.39                                  | 45.00                             | 4.68                              |
| 4                     | 1.97                                | 2.08                                  | 46.50                             | 4.08                              |
| WR-WR simulations ($n = 10^4$) |                                      |                                       |                                   |
| 1                     | 1.33                                | 4.81                                  | 41.82                             | 9.43                              |
| 2                     | 1.15                                | 3.24                                  | 51.61                             | 6.35                              |
| 3                     | 1.07                                | 2.66                                  | 57.52                             | 5.21                              |
| 4                     | 1.04                                | 2.30                                  | 61.91                             | 4.51                              |

Averaging was by linear interpolation to 1-s intervals before ensemble averaging; $\phi 2$ isolation was by removal of the first 20 s or 15 s of data for UP-WR and WR-WR protocols, respectively.
simplicity of avoiding potentially complicated and assumption-laden averaging procedures on large data sets, $\tau \dot{V}O_2P$ estimation using this averaging method reduced accuracy and markedly lessened the confidence in the derived parameter estimates and should therefore be avoided. This likely arose because the influence on $\tau \dot{V}O_2P$ of breath-by-breath fluctuations is nonlinear: large "noise" in the early transient has more influence on $\tau \dot{V}O_2P$ than the same "noise" in the later transient (57). Therefore, data handling approaches that first reduce breath-by-breath fluctuations and then characterize the fit (rather than the other way around) appear to result in more robust parameterization of the kinetics.

Another cautionary note is evident in our data for the interpolation method of averaging. This method appears to return a substantially narrowed confidence interval for $\tau \dot{V}O_2P$ estimation (Figs. 3B, 7E, and 8E). However, because the confidence interval is dependent on the number of samples (i.e., breaths), interpolation artificially increases the sampling frequency of the original data. The interpolation method therefore returns an artificial confidence interval that is more dependent on the characteristics of the interpolation than on the original measurements (21). The true confidence interval of parameter estimation for the interpolation method is likely better reflected in the binned and stacked methods (Fig. 3B), which were substantially similar across all simulations.

Each data processing method investigated resulted in a similar degree of accuracy around the true value, and therefore approaches to data processing should focus on attempts to optimize the confidence of parameter estimation. As with data collection, valid and appropriate processing methods that reduce breath-by-breath fluctuations in the data will result in increased confidence.

Data fitting. We found that empirical time removal methods to isolate the $\phi2$ data for fitting resulted in significantly more accurate and precise estimations of $\tau \dot{V}O_2P$ than either a biexponential fit, or statistical and time-derivative methods to identify the $\phi1\rightarrow2$ transition followed by a monoeponential fit to the isolated $\phi2$ data. The majority of published experimental studies that have quantified the kinetics of $\dot{V}O_2P$ have used such empirical time removal methods (usually removing the first 20 s of data), and so researchers have historically used the $\phi2$ isolation method that we have now shown provides the most accurate and precise estimations of $\tau \dot{V}O_2P$. Furthermore, this empirical time removal approach is far simpler to implement than the biexponential, statistical, or time-derivative methods. Previous recommendations have been to remove at least 20 s of data from the beginning of the data set to completely remove $\phi1$ data, even though some data from the start of $\phi2$ may also be removed (7, 57). However, our results suggest that, somewhat counterintuitively, it is better to include a small amount of data from the end of $\phi1$ in the fitting procedure than exclude data from the start of $\phi2$. This is seen in Figs. 5B and 6B, where $\phi1$ amplitude and $\phi2 \tau \dot{V}O_2P$ estimation for exercise from unloaded pedaling was more precise and accurate when the initial 15 s of data was removed than when the initial 25 s of data was removed (the true $\phi1\rightarrow2$ transition for these data occurred at 19.5 ± 3.3 s). We suggest that this is because the inherent fluctuations in the $\dot{V}O_2P$ data means that including a small amount of $\phi1$ data in the fit has minimal effect on the resultant $\phi1$ amplitude and $\phi2 \tau \dot{V}O_2P$ estimation. The rapidly changing initial portion of $\phi2$ data (which changes rapidly with respect to the breath-by-breath fluctuations at the end of $\phi1$) is key to obtaining accurate and precise estimations. Qualitatively similar results were found for exercise that started from a raised work rate, but here the best $\tau \dot{V}O_2P$ estimation was with the removal of the first 15 s of data (Fig. 8F). This is likely due to the increased baseline work rate elevating cardiac output, which reduces muscle-to-lung blood transit times and, therefore, the cardiodynamic $\phi1$ duration. Nevertheless, the accuracy and precision of $\tau \dot{V}O_2P$ estimation was statistically similar for WR-WR transitions when either 15 or 20 s of data were removed. We therefore recommend that researchers err on the side of caution when isolating $\phi2$ $\dot{V}O_2P$ data and remove no more than 20 s of data to optimize $\tau \dot{V}O_2P$ estimation.

Implications for interpretation of $\phi2 \dot{V}O_2P$ kinetics. There are two significant findings from our simulations that have implications for interpretation of $\phi2 \dot{V}O_2P$ kinetics. First, we found that, on average, $\phi2 \tau \dot{V}O_2P$ was overestimated in all the data collection and handling strategies investigated. This overestimation can be explained, at least in part, by the two-phase $\dot{V}O_2P$ response and the nonexponentiality of $\phi2$ (3, 5, 8, 25; cf. 18). Figure 9 shows the effects on $\phi2 \tau \dot{V}O_2P$ estimation when the nonexponential Eq. 2 is fit to clean simulation output data from different points throughout the $\dot{V}O_2P$ response. If the nonexponential fit is started during $\phi1$ (i.e., from any point before 19.4 s in this example), then the estimated $\phi2 \tau \dot{V}O_2P$ is larger than true, due to the inclusion of some $\phi1$ data in the fit. If the fit is started after the $\phi1\rightarrow2$ transition, then the $\phi2 \tau \dot{V}O_2P$ estimation is also larger than true, becoming larger as the fit is started further from the $\phi1\rightarrow2$ transition, because the underlying $\phi2$ response is not a pure monoexponential; it initially increases more rapidly than a monoexponential before slowing down as it reaches the steady state (8). Only a fit that starts exactly at the $\phi1\rightarrow2$ transition returns the true $\phi2 \tau \dot{V}O_2P$. For these clean simulated data, inaccurate identification of the $\phi1\rightarrow2$ transition by just 2 s can result in a $\phi2 \tau \dot{V}O_2P$ estimation that is 1.6 s larger than the true value; the influence of noise in experimentally obtained data may exacerbate this error. Because of these effects on $\phi2 \tau \dot{V}O_2P$ estimation, when using the identified optimal data processing and fitting procedures, we were only able to estimate $\phi2 \tau \dot{V}O_2P$ to within 2 s of true in 47% of the $10^4$ UP-WR simulations, and in 62% of the $10^4$ WR-WR simulations [2 s represents an effect size of ~10% for a healthy young human, where $\tau \dot{V}O_2P$ is typically ~20 s (45)]. Extrapolating this analysis further, we calculated the 95% confidence limits of our $\tau \dot{V}O_2P$ estimate distributions (as shown in Fig. 7, B and D, and Fig. 8, B, D, and F); $\tau \dot{V}O_2P$ estimates from outside this confidence interval are statistically likely to come from a different distribution/population. These 95% confidence limits, for $\tau \dot{V}O_2P$ estimates using our predetermined optimal data processing and fitting procedures, are ±4.08 and ±4.51 s from the mean, for transitions from unloaded pedaling or a raised work rate, respectively (Table 3). We therefore propose that the minimally important difference for a significant change in $\tau \dot{V}O_2P$, e.g., during interventional and comparative studies, should be 5.0 s. If the number of averaged bouts is reduced from the optimum of four, this minimally important difference should be increased in accordance with the confidence limits shown in Table 3.

The second implication for interpretation of $\tau \dot{V}O_2P$ from our data is to question whether an exponential fit should be used at
We have previously shown that the dynamics and mixing of circulatory compartments between muscle and lung distort the approximately exponential muscle \( \dot{V}_{O_2} \) kinetics into a nonexponential \( \phi_2 \dot{V}_{O_{2P}} \) response at the lung (8). A recent meta-analysis of available data measuring both muscle and lung \( \dot{V}_{O_2} \) kinetics during cycling and knee extension exercise demonstrates a wide variability of \( \tau_{V_{O2}} \) between muscle and lung (27). Some have proposed alternative methods to assess kinetic responses, such as the time to steady state (45). However, such approaches have been demonstrated to be inherently more variable than relying on a method that maximizes the utility of available non-steady-state data (24, 47) and is conceptually flawed on the basis that the time to steady state of a nonexponential process is continually changing (8). Alternative approaches to kinetics estimation using, for example, pseudorandom binary sequence exercise testing and time-series analysis may allow for muscle \( \tau_{V_{O2}} \) to be resolved by alternative methods (24, 31). It remains to be determined whether such methods provide increased accuracy for noninvasive estimation of muscle \( V_{O2} \) kinetic responses compared with \( \phi_2 \tau_{V_{O2}} \) estimation by repeated step transitions. Our simulations here demonstrate that a monoeponential fit to \( \phi_2 \dot{V}_{O_{2P}} \) is a useful and concise method for accurately describing the overall kinetics of the exponential-like pulmonary \( \phi_2 \dot{V}_{O2} \) kinetic response.

**Limitations.** The means and SDs of the parameters used in our Monte Carlo simulations were representative of healthy young adults (8, 22). Quantitatively different results may be found for other populations with different \( \phi_2 \dot{V}_{O_{2P}} \) kinetic parameters, such as the elderly or heart failure patients who have slowed \( V_{O2P} \) kinetics (9, 39). Nevertheless, our main qualitative findings will still be pertinent when collecting, processing, and fitting \( V_{O2P} \) data from these other populations. In particular, our main point regarding optimal data collection and processing methods—that methods should be employed to minimize breath-by-breath fluctuations and that it is essential to include all \( \phi_2 \dot{V}_{O_{2P}} \) data in the fit—will more than likely stand for these populations, as it is still expected that the (potentially slowed) initial portion of \( \phi_2 \dot{V}_{O_{2P}} \) will change rapidly with respect to the noise in the data at the end of \( \phi_1 \).

For populations where individuals are expected to have a reduced cardiac output and slowed cardiac output kinetics, and a concomitant prolongation of \( \phi_1 \) duration compared with young healthy adults [such as heart failure patients (52)], the use of a biexponential fit, or statistical or derivative methods, to automatically identify the \( \phi_1–2 \) transition is inherently attractive. However, our results highlight that the noise in the \( V_{O2P} \) data limit the ability of these methods to correctly identify the \( \phi_1–2 \) transition, reducing the accuracy and precision of subsequent \( \tau_{V_{O2P}} \) estimation. In this study, the empirical time-removal methods (removal of the first 20 s of data for exercise from unloaded pedaling, or 15 s if exercise was started from a raised baseline) were the only methods that gave statistically similar \( \tau_{V_{O2P}} \) estimates to control fits, despite \( \phi_1 \) duration ranging from 7 to 39 s across all simulations. It remains to be determined whether removal of the first 20 s of data results in the most accurate and precise \( \tau_{V_{O2P}} \) estimations for populations where \( \phi_1 \) is prolonged, but it may be necessary to compensate for the prolonged \( \phi_1 \) duration when removing \( \phi_1 \) data from the fitting window.

Only on-transient exercise in the moderate-intensity domain was simulated in this study. It is still to be determined whether the identified optimal fitting procedures will produce the most accurate and precise \( \tau_{V_{O2P}} \) estimations for on-transient data in higher exercise intensity domains where fitting can be complicated by the emergence of a \( V_{O2P} \) slow component (40, 45). Similarly, the applicability of our identified optimal procedures for off-transient data, where cardiac output is expected to be initially elevated and so produce a much shorter \( \phi_1 \), potentially influencing the amount of data that should be removed before fitting, is still to be determined.

---

**Fig. 9.** A: simulated \( \dot{V}_{O_{2P}} \) response to a 100 W UP-WR step using default model parameters [see Ref. 8 for details]. The \( \phi_1–2 \) transition occurs at 19.4 s and the true \( \phi_2 \tau_{V_{O2P}} \) is 16.3 s. B: effects on \( \tau_{V_{O2P}} \) estimation of fitting \( \phi_2 \) \( \dot{V}_{O_{2P}} \) data from different points throughout the clean simulated \( \dot{V}_{O_{2P}} \) response. The vertical dashed line shows the time of the \( \phi_1–2 \) transition, the horizontal dashed line shows the true \( \phi_2 \tau_{V_{O2P}} \).
Conclusions. We used a validated computational model together with a Monte Carlo approach to assess the accuracy and the precision of various averaging and fitting procedures on the estimation of $\dot{V}O_{2P}$ kinetics. Our analyses showed that four bouts of exercise was the optimal number to average to increase accuracy and precision of $\tau V O_{2P}$ estimation. Choice of averaging strategy was not so critical, with interpolation, bin averaging, and stacking all giving quantitatively similar $\tau V O_{2P}$ estimates. The interpolation, binning, and stacking methods did, however, allow more confident parameter estimates when compared with analyzing repeated bouts separately. Data collection and processing strategies should therefore focus on increasing the signal-to-noise ratio in the collected data. Contradictory to previous advice that suggests removal of at least 20 s of data to isolate $\dot{d}V O_{2P}$ before fitting, our analyses show that data fitting procedures should remove no more than 20 s of data, as this provided the most precise and accurate estimates of $\tau V O_{2P}$. Our analyses showed the widely used standard approaches for data collection, processing, and fitting, although often different between laboratories, did not have a substantial effect on the quantitation of $\dot{d}V O_{2P}$ kinetics per se. However, we found that even this optimal procedure yielded $\tau V O_{2P}$ estimates that were within $\pm 2$ s of true in only 47–62% of simulations. Thus we identified the minimally important difference for $\tau V O_{2P}$ for use in interventional and comparative studies to be 5 s.

GRANTS

This work was supported by a Biotechnology and Biological Science Research Council UK research grant (BR/I00162X/1) and a University of Leeds International Research Collaboration Award.

DISCLOSURES

Conflict of interest statement: No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

A.P.B., T.S.B., C.F., S.R.M., and H.B.R. conceived and designed research; A.P.B. performed experiments; A.P.B. analyzed data; A.P.B., T.S.B., C.F., S.R.M., and H.B.R. interpreted results of experiments; A.P.B. prepared figures; A.P.B. drafted manuscript; A.P.B., T.S.B., C.F., S.R.M., and H.B.R. edited and revised manuscript; A.P.B., T.S.B., C.F., S.R.M., and H.B.R. approved final version of manuscript.

REFERENCES

1. Babcock MA, Paterson DH, Cunningham DA, Dickinson JR. Exercise on-transient gas exchange kinetics are slowed as a function of age. Med Sci Sports Exerc 26: 440–446, 1994. doi: 10.1249/00005768-19940400-00007.

2. Bangsbo J, Krustup P, González-Alonso J, Boushel R, Saltin B. Muscle oxygen kinetics at onset of intense dynamic exercise in humans. Am J Physiol Regul Integr Comp Physiol 279: R899–R906, 2000.

3. Barstow TJ, Lamarrna N, Whipp BJ. Modulation of muscle and pulmonary O$_2$ uptakes by circulatory dynamics during exercise. J Appl Physiol (1985) 68: 979–989, 1990.

4. Barstow TJ, Molé PA. Linear and nonlinear characteristics of oxygen uptake kinetics during heavy exercise. J Appl Physiol (1985) 71: 2099–2106, 1991.

5. Barstow TJ, Molé PA. Simulation of pulmonary O$_2$ uptake during exercise transients in humans. J Appl Physiol (1985) 63: 2253–2261, 1987.

6. Beaver WL, Lamarrna N, Wasserman K. Breath-by-breath measurement of true alveolar gas exchange. J Appl Physiol Respir Environ Exerc Physiol 51: 1662–1675, 1981.

7. Bell C, Paterson DH, Kowalchuk JM, Padilla J, Cunningham DA. A comparison of modelling techniques used to characterise oxygen uptake kinetics during the on-transient of exercise. Exp Physiol 86: 667–676, 2001. doi: 10.1113/jphp8602150.

8. Benson AP, Grassi B, Rossiter HB. A validated model of oxygen uptake and circulatory dynamic interactions at exercise onset in humans. J Appl Physiol (1985) 115: 743–755, 2013. doi: 10.1152/japplphysiol.00184.2013.

9. Bowen TS, Cannon DT, Murgatroyd SR, Birch KM, Witte KK, Rossiter HB. The intramuscular contribution to the slow oxygen uptake kinetics during exercise in chronic heart failure is related to the severity of the condition. J Appl Physiol (1985) 112: 378–387, 2012. doi: 10.1152/japplphysiol.00779.2011.

10. Bringard A, Adami A, Moia C, Ferretti G. A new interpolation-free procedure for breath-by-breath analysis of oxygen uptake in exercise transients. Eur J Appl Physiol 114: 1983–1994, 2014. doi: 10.1007/s00424-014-2920-z.

11. Bruce EN. Temporal variations in the pattern of breathing. J Appl Physiol (1985) 80: 1079–1087, 1996.

12. Brunner-La Rocca HP, Weilemann D, Follath F, Schlumpf M, Rickli H, Schalcher C, Maly FE, Candinas R, Kiowski W. Oxygen uptake kinetics during low level exercise in patients with heart failure: relation to neurohormones, peak oxygen consumption, and clinical findings. Heart 81: 121–127, 1999. doi: 10.1136/hrt.81.121.

13. Capelli C, Cautero M, di Prampero PE. New perspectives in breath-by-breath determination of alveolar gas exchange in humans. Pflugers Arch 441: 566–577, 2001. doi: 10.1007/s004240000429.

14. Capelli C, Cautero M, Pogliaghi S. Algorithms, modelling and V$\dot{O}$2 kinetics. Eur J Appl Physiol 111: 331–342, 2011. doi: 10.1007/s00424-010-1396-8.

15. Casaburi R, Daly J, Hansen JE, Effros RM. Abrupt changes in mixed venous blood gas composition after the onset of exercise. J Appl Physiol (1985) 67: 1106–1112, 1989.

16. Chechile RA. Pooling data versus averaging model fits for some prototype time- and breathwise processing models. J Math Psychol 53: 562–576, 2009. doi: 10.1016/j.jmph.2009.06.005.

17. Dunn WL, Shultis JK. Assembling more VO$\dot{2}$ uptake parameters of breath-by-breath pulmonary V$\dot{O}$2P before fitting, our analyses show that data fitting procedures should remove no more than 20 s of data, as this provided the most precise and accurate estimates of $\tau V O_{2P}$. Our analyses showed the widely used standard approaches for data collection, processing, and fitting, although often different between laboratories, did not have a substantial effect on the quantitation of $\dot{d}V O_{2P}$ kinetics per se. However, we found that even this optimal procedure yielded $\tau V O_{2P}$ estimates that were within $\pm 2$ s of true in only 47–62% of simulations. Thus we identified the minimally important difference for $\tau V O_{2P}$ for use in interventional and comparative studies to be 5 s.

18. Francescato MP, Cettolo V, di Prampero PE. Oxygen uptake kinetics at work onset: role of cardiac output and of phosphocreatine breakdown. Respir Physiol Neurobiol 185: 287–295, 2013. doi: 10.1016/j.resp.2012.09.015.

19. Francescato MP, Cettolo V, Bellio R. Assembling more O$_2$ uptake responses: is it possible to merely stack the repeated transitions? Respir Physiol Neurobiol 200: 46–49, 2014. doi: 10.1016/j.resp.2014.06.004.

20. Francescato MP, Cettolo V, Bellio R. Confidence intervals for the parameters estimated from simulated O$_2$ uptake kinetics: effects of different data treatments. Exp Physiol 99: 187–195, 2014. doi: 10.1113/exphphysiol.2013.076208.

21. Francescato MP, Cettolo V, Bellio R. Interpreting the confidence intervals of model parameters of breath-by-breath pulmonary O$_2$ uptake. Exp Physiol 100: 475, 2015. doi: 10.1113/EP085043.

22. Grassi B, Poole DC, Richardson RS, Knight DR, Erickson BK, Wagner PD. Muscle O$_2$ uptake kinetics in humans: implications for metabolic control. J Appl Physiol (1985) 80: 988–998, 1996.

23. Grassi B, Porcelli S, Marzorati M, Lanfranconi F, Vago P, Marconi C, Morandi L. Metabolic myopathies: functional evaluation by analysis of oxygen uptake kinetics. Med Sci Sports Exerc 41: 2120–2127, 2009. doi: 10.1249/MSS.0b013e3181aae96b.

24. Hoffmann U, Drescher U, Benson AP, Rossiter HB, Essfeld D. Skeletal muscle VO$_2$ kinetics from cardio-pulmonary measurements: assessing distortions through $O_2$ transport by means of stochastic work-rate signals and circulatory modelling. Eur J Appl Physiol 113: 1745–1754, 2013. doi: 10.1007/s00424-013-2958-7.

25. Hughson RL. Oxygen uptake kinetics: historical perspective and future directions. Appl Physiol Nutr Metab 34: 840–850, 2009. doi: 10.1113/jhp-089-088.

26. Keir DA, Murias JM, Paterson DH, Kowalchuk JM. Breath-by-breath pulmonary $O_2$ uptake kinetics: effect of data processing on confidence in estimating model parameters. Exp Physiol 99: 1511–1522, 2014. doi: 10.1113/epj09881.000817.
28. Kohl J, Koller EA, Jüger M. Relation between pedalling- and breathing rhythm. Eur J Appl Physiol Occup Physiol 47: 223–237, 1981. doi:10.1007/BF00422468.
29. Kopp K, Bouckaert J, Jones AM. Effects of training status and exercise intensity on phase II VO$_2$ kinetics. Med Sci Sports Exerc 36: 225–232, 2004. doi:10.1249/01.MSS.0000113473.48220.20.
30. Korzeniewski B, Rossiter HB. Each-step activation of oxidative phosphorylation is necessary to explain muscle metabolic kinetic responses to exercise and recovery in humans. J Physiol 593: 5255–5268, 2015. doi:10.1113/JPP271299.
31. Koschate J, Drescher U, Baum K, Eichberg S, Schiffer T, Latsch J, Brixius K, Hoffmann U. Muscular oxygen uptake kinetics in aged adults. Int J Sports Med 37: 516–524, 2016. doi:10.1055/s-0042-101413.
32. Lamarra N, Whipp BJ, Ward SA, Wasserman K. Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. J Appl Physiol (1985) 62: 2003–2012, 1987.
33. Lamarra N, Whipp BJ, Ward SA, Wasserman K. The effect of hyperoxia on the coupling of ventilator and gas-exchange dynamics in response to impulse exercise testing. In: Concepts and Formalizations in the Control of Breathing, edited by Bencherit G, Bacoينer P, Demongeot J, Manchester: UK: Manchester University Press, 1987.
34. Lewis RM, Torczon V, Trosset MW. Direct search methods: then and now. J Comput Appl Math 124: 191–207, 2000. doi:10.1016/S0377-0427(00)00423-4.
35. Mezzani A, Grassi B, Giordano A, Corra` U, Colombo S, Giannuzzi P. Understanding cardiac output kinetics in aged adults. J Appl Physiol Respir Environ Exerc Physiol 99: 2000-2009, 2000.
36. Mezzani A, Grassi B, Giordano A, Corra` U, Colombo S, Giannuzzi P. Age-related prolongation of phase I of VO$_2$-on-kinetics in healthy humans. Am J Physiol Regul Integr Comp Physiol 299: R968–R976, 2010. doi:10.1152/ajpregu.00079.2009.
37. Murgatroyd SR, Ferguson C, Ward SA, Whipp BJ, Rossiter HB. Pulmonary O$_2$ uptake kinetics as a determinant of high-intensity exercise tolerance in humans. J Appl Physiol (1985) 110: 1598–1606, 2011. doi:10.1152/japplphysiol.00459.2010.
38. Nery LE, Wasserman K, Kowalchuk JM, Paterson DH. Influence of phase I duration on phase II VO$_2$ kinetics: methodological considerations. Eur J Appl Physiol Occup Physiol 301: R218–R224, 2011. doi:10.1007/japregu.00069.2011.
39. Newport DM, Ward SA, Swanson GD, Sherrill DL. On the breath-to-breath estimation of gas exchange. J Appl Physiol Respir Environ Exerc Physiol 56: 259–261, 1984.
40. Paterson DH, Gravelle BM, Murias JM, Zerbini L, Pogliaghi S, Paterson DH. Duration of “Phase I” VO$_2$: a comparison of methods used in its estimation and the effects of varying moderate-intensity work rate. Am J Physiol Regul Integr Comp Physiol 304: R238–R247, 2013. doi:10.1152/ajpregu.00029.2012.
41. Paterson DH, Swanson GD, Sherrill DL. On the breath-to-breath estimation of gas exchange. J Appl Physiol Respir Environ Exerc Physiol 56: 259–261, 1984.
42. Petroll CR, Childs DJ, Houghton W, Armstrong N. Breath-to-breath “noise” in the ventilatory and gas exchange responses to exercise. Eur J Appl Physiol Occup Physiol 80: 118–124, 1999. doi:10.1007/s004240050567.
43. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes: The Art of Scientific Computing (3rd ed.). Cambridge, UK: Cambridge University Press, 2007.
44. Puente-Maestu L, Palange P, Casaburi R, Laveneziana P, Maltais F, Neder JA, O’Donnell DE, Onorati P, Porszasz J, Rabinovich R, Rossiter HB, Singh S, Troosters T, Ward S. Use of exercise testing in the evaluation of interventional efficacy: an official ERS statement. Eur Respir J 47: 429–460, 2016. doi:10.1183/13993003.00745-2015.
45. Robergs RA. A critical review of the history of low- to moderate-intensity steady-state VO$_2$ kinetics. Sports Med 44: 641–653, 2014. doi:10.1107/s40279-014-0161-2.
46. Rossiter HB. Exercise: kinetic considerations for gas exchange. Compr Physiol 1: 203–244, 2011. doi:10.1002/cphy.c090010.
47. Rossiter HB, Howe FA, Ward SA, Kowalchuk JM, Griffiths JR, Whipp BJ. Intersample fluctuations in phosphocreatine concentration determined by $^{31}$P-magnetic resonance spectroscopy and parameter estimation of metabolic responses to exercise in humans. J Physiol 528: 359–369, 2000. doi:10.1111/j.1469-7793.2000.01-100359.x.
48. Rossiter HB, Ward SA, Doyle VL, Howe FA, Griffiths JR, Whipp BJ. Inferences from pulmonary O$_2$ uptake with respect to intramuscular [phosphocreatine] kinetics during moderate exercise in humans. J Physiol 518: 921–932, 1999. doi:10.1111/j.1469-7793.1999.9021p.x.
49. Schalcher C, Rickhi L, Brehm M, Weilennman D, Oechslin E, Kiowski W, Brunner-La Rocca HP. Prolonged oxygen uptake kinetics during low-intensity exercise are related to poor prognosis in patients with mild-to-moderate congestive heart failure. Chest 124: 580–586, 2003. doi:10.1378/chest.124.2.580.
50. Siesjesma KE. Oxygen uptake kinetics in response to exercise in patients with pulmonary vascular disease. Am Rev Respir Dis 145: 1052–1057, 1992. doi:10.1164/ajrccm.145.5.1052.
51. Siesjesma KE, Ben-Dov I, Zhang YY, Sullivan C, Wasserman K. Dynamics of oxygen uptake for submaximal exercise and recovery in patients with chronic heart failure. Chest 105: 1693–1700, 1994. doi:10.1378/chest.105.6.1693.
52. Spee RF, Nieemier JvM, Schoots T, Wijn PF, Doevendans PA, Kemps HM. The relation between cardiac output kinetics and skeletal muscle oxygenation during moderate exercise in moderately impaired patients with chronic heart failure. J Appl Physiol (1985) 121: 198–204, 2016. doi:10.1152/ajpregu.00079.2016.
53. Spiess A-N, Neumeyer N. An evaluation of $R^2$ as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. BMC Pharmacol 10: 6, 2010. doi:10.1186/1471-2210-6-6.
54. Spencer MD, Gravelle BM, Murias JM, Zerbini L, Pogliaghi S, Paterson DH. Duration of “Phase I” VO$_2$: a comparison of methods used in its estimation and the effects of varying moderate-intensity work rate. Am J Physiol Regul Integr Comp Physiol 304: R238–R247, 2013. doi:10.1152/ajpregu.00029.2012.