Article:
Benson, AP orcid.org/0000-0003-4679-9842, Bowen, TS, Ferguson, C orcid.org/0000-0001-5235-1505 et al. (2 more authors) (2017) Data collection, handling and fitting strategies to optimize accuracy and precision of oxygen uptake kinetics estimation from breath-by-breath measurements. Journal of Applied Physiology, 123 (1). pp. 227-242. ISSN 8750-7587

https://doi.org/10.1152/japplphysiol.00988.2016

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. Rossiter**[5,1]

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

RUNNING HEAD

Strategies to optimize $\dot{V}O_2$ kinetics estimation

ADDRESS FOR CORRESPONDENCE

Alan P. Benson

School of Biomedical Sciences

University of Leeds

Leeds

LS2 9JT

United Kingdom

E-mail: a.p.benson@leeds.ac.uk

ABSTRACT

Phase 2 pulmonary oxygen uptake kinetics ($\phi 2\ \tau\dot{V}O_{2P}$) 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 $\dot{V}O_{2P}$ kinetics estimation. A validated computational model of $\dot{V}O_{2P}$ kinetics and a Monte Carlo approach simulated 2 x 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 (bi-exponential fitting, or $\phi 2$ isolation by time removal, statistical or derivative methods followed by mono-exponential fitting) on accuracy and precision of $\dot{V}O_{2P}$ kinetics estimation were assessed. The optimal strategy to maximize accuracy and precision of $\tau\dot{V}O_{2P}$ 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 to analyzing repeated bouts separately. The optimal procedure resulted in $\phi 2\ \tau\dot{V}O_{2P}$ 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 $\tau\dot{V}O_{2P}$ ranged from 4.08-4.51 s, suggesting a minimally important difference of ~5 s to determine significant changes in $\tau\dot{V}O_{2P}$ during interventional and comparative studies.

NEW & NOTEWORTHY

We identified an unbiased method to maximize accuracy and precision of oxygen uptake kinetics ($\tau\dot{V}O_{2P}$) estimation. The optimum number of bouts to average was four; interpolation, bin 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 $\tau\dot{V}O_{2P}$ during interventional and comparative studies.

KEYWORDS

Oxygen uptake kinetics; Accuracy and precision; Data handling; Computational modeling.

INTRODUCTION

At the onset of constant power exercise below the lactate threshold (LT) in humans, mitochondrial oxidative phosphorylation and, subsequently, muscle oxygen uptake ($\dot{V}O_{2m}$) 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 (ϕ) 2 of the pulmonary $\dot{V}O_2$ ($\dot{V}O_{2P}$), characterized by the response time constant (τ) from repeated breath-by-breath gas exchange measurements, are commonly used to infer $\dot{V}O_{2m}$ kinetics and provide a non-invasive tool to investigate the control of exercise energetics (27, 41, 46). Fast ϕ2 $\dot{V}O_{2P}$ kinetics reflect effective cardiopulmonary and neuromuscular integration, and are associated with high endurance exercise performance (29, 38, 41), whereas ϕ2 $\dot{V}O_{2P}$ kinetics are slowed in the elderly (1) and with chronic disease (12, 23, 40, 46, 51). In addition, ϕ2 $\dot{V}O_{2P}$ 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 ϕ1 of the $\dot{V}O_{2P}$ response (ϕ1 duration and amplitude) are clinically discriminatory (50) and sensitive to age (37). Thus, the strong link between $\dot{V}O_{2P}$ kinetics and state of health provides the basis for an inherently attractive, non-invasive and effort-independent method to characterize the efficacy of the integrated physiologic systems response to exercise.

While there are general guidelines for characterizing $\dot{V}O_{2P}$ 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 ϕ1 duration and

amplitude and $\phi 2$ $\tau\dot{V}O_{2P}$ 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 $\dot{V}O_{2P}$ data collection, handling and fitting that allows the most accurate and precise estimation of $\dot{V}O_{2P}$ kinetics. We identified this optimal criterion by systematically 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 $\phi 2$ $\tau\dot{V}O_{2P}$ estimation, using a validated cardiopulmonary simulation of exercise gas exchange (8) and a Monte Carlo approach.

THEORETICAL CONSIDERATIONS

The process linking $\dot{V}O_{2P}$ data collection in the laboratory or clinic, to kinetics characterization, is typically undertaken in three distinct steps: (i) data collection, (ii) data processing, and (iii) 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 $\dot{V}O_{2P}$ 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. 1*B*) (10, 26, 33, 57). The breath-by-breath fluctuations (also referred to as "noise") inherent in any $\dot{V}O_{2P}$ 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 approximately 30 to 110 $ml.min^{-1}$ (33, 47),

105 independent of metabolic rate (33). What is less clear, however, is how different signal-to-
106 noise ratios (or, analogously, the number of combined exercise bouts) affect $\dot{V}O_{2P}$ kinetics
107 estimation and, therefore, whether there is an optimal number of exercise bouts required to
108 estimate $\dot{V}O_{2P}$ kinetics to a given level of confidence.
109
110 *Step 2 – data processing:* After the removal of outlying breaths generated by swallows or
111 coughs or other 'mistriggers' of the breath identification algorithms, and unrelated to tidal
112 breathing [typically those breaths more than 3 or 4 SDs from the local mean (33, 57)], the
113 second step involves averaging of the data collected from multiple exercise bouts to obtain a
114 single (processed) $\dot{V}O_{2P}$ signal with a high signal-to-noise ratio, prior to kinetic
115 characterization. Several averaging techniques are employed (Fig. 1*C*-*E*), the most widely-
116 used involving some form of interpolation and/or averaging. Linear interpolation of data prior
117 to averaging (commonly to 1 s intervals) is necessary to normalize gas exchange sampling
118 frequency, from the non-uniform breath-by-breath sampling, and therefore ensure equal
119 weighting of data among repeated trials (Fig. 1*C*) (57). Averaging may be in the form of post-
120 interpolation ensemble averaging (56), or by arranging un-interpolated data from all bouts in
121 time (10) before averaging the combined breaths into bins whose size depends on the number
122 of averaged bouts (38) or time (9, 26) (Fig. 1*D*). This "binning" approach to averaging, while
123 improving the signal-to-noise ratio, may help to maintain the density of the data close to that
124 at which it was collected (i.e. breathing frequency), and improve the validity of the estimated
125 confidence intervals (21, 38). Despite the general popularity and acceptance of these
126 approaches, several other data processing methods warrant investigation. Recent simulation
127 studies have suggested that simple superimposition of all data from all bouts before fitting
128 can give accurate $\phi 2$ $\tau\dot{V}O_{2P}$ estimates, with the added simplicity of reducing the requirement
129 for complex data treatments (Fig. 1*E*) (19). Another alternative averaging approach, and

maybe one that is statistically more robust (16) yet is not typically used for estimating $\dot{V}O_{2P}$ 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 $\tau\dot{V}O_{2P}$ estimation. It remains unclear, though, how averaging strategies affect both the precision and accuracy of $\dot{V}O_{2P}$ 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 $\dot{V}O_{2P}$ data in order to obtain an estimate of the kinetics of $\dot{V}O_{2P}$. The $\dot{V}O_{2P}$ 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 $\dot{V}O_{2m}$ in young healthy adults (Fig. 1*A*) (22, 48). This entire response has been described mathematically using a piecewise bi-exponential equation of the form

$$\dot{V}O_{2P}(t) = \dot{V}O_{2P,base} + A_1\left[1 - e^{-t/\tau_1}\right] + H(t)A_2\left[1 - e^{-(t-TD)/\tau_2}\right],$$
$$H(t) = \begin{cases} 0, & t < TD, \\ 1, & t \geq TD, \end{cases} \quad (1)$$

where t is time, $\dot{V}O_{2P,base}$ is baseline $\dot{V}O_{2P}$, A_1 and A_2 are the amplitudes of the first and second phases of the response, τ_1 and τ_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 τ_2, i.e. $\phi 2$ $\tau\dot{V}O_{2P}$. 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 mono-exponential (15, 55). In addition, there are several practical

difficulties when using Equation (1) to fit $\dot{V}O_{2P}$ data: Phase 1 typically contains only a few breaths (typically 5 or 6 in our simulations; see Fig. 1*B*), and fitting so few data points with the first exponential term in Equation (1) drastically reduces the confidence of the parameter estimations in that first exponential term. The influence of this potentially unconfident ϕ1 fit continues into ϕ2, affecting τ_2 (ϕ2 $\tau\dot{V}O_{2P}$) estimation, particularly if the fit to the ϕ1 data does not reach a steady-state before ϕ2 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 Equation (1); the parameters A_1 and τ_1 are shared over, and influenced by the data in, the two different sub-domains ($t < TD$ or ϕ1, and $t \geq TD$ or ϕ2), and the extents of the sub-domains themselves are determined by the parameter *TD*. As such, fitting Equation (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 ϕ2, an alternative (and the most commonly used) approach is to isolate the ϕ2 data then fit these data with a mono-exponential equation of the form

$$\dot{V}O_{2P}(t) = \dot{V}O_{2P,base} + A\left[1 - e^{-(t-TD)/\tau}\right]. \quad (2)$$

Such a mono-exponential equation accurately describes the ϕ2 $\dot{V}O_{2P}$ response to moderate intensity step exercise (4, 5) and can be handled by most nonlinear least squares algorithms. If Equation (2) is used to fit the $\dot{V}O_{2P}$ data and obtain an estimate of ϕ2 $\tau\dot{V}O_{2P}$, it is necessary to omit the ϕ1 data from the fit. The most widely-used methods for removing ϕ1 data are empirically-derived time-removal methods, where "at least" the first 20 s of data from the exercise transient are removed prior to fitting (7, 39, 54, 57). The rationale behind this strategy is that, because ϕ1 is expected to last less than 20 s and the ϕ2 $\dot{V}O_{2P}$ response is expected to be exponential, starting the fit from any given point past the ϕ1-2 transition will yield an identical time constant that represents the underlying ϕ2 kinetics; whereas starting

176 the fit from any point before the ϕ1-2 transition will result in a larger (incorrect) time
177 constant for ϕ2 (39, 54, 57). However, the ϕ2 $\dot{V}O_{2P}$ response is not truly exponential, but
178 rather is a non-linear distortion of a mono-exponential $\dot{V}O_{2m}$ response (3, 5, 8, 25; cf. 18).
179 Thus, contrary to $\tau\dot{V}O_{2m}$, the $\tau\dot{V}O_{2P}$ is not a "true" constant throughout the transient, and
180 fitting an exponential equation from different points in such a non-exponential ϕ2 will yield
181 varying values for $\tau\dot{V}O_{2P}$; progressively larger values as the fit is started from later in ϕ2 (cf.
182 8). Such behavior is suggested in the empirical results of Murias et al. (39) where $\tau\dot{V}O_{2P}$
183 becomes larger as the imposed exponential fit is started from later in the exercise transient, at
184 least in older adults. Although $\tau\dot{V}O_{2P}$ is influenced by a complex interaction of circulatory
185 and gas exchange responses to exercise, and ϕ2 $\dot{V}O_{2P}$ is not quite exponential, a mono-
186 exponential fit of moderate intensity $\dot{V}O_{2P}$ kinetics remains a useful, concise and effort-
187 independent method to characterize the integrated dynamic responsiveness of
188 cardiopulmonary and neuromuscular health. Nevertheless, it seems crucial that *all data*
189 contained in the ϕ2 response, but *none* of the ϕ1 data, are fitted in order to obtain the most
190 accurate characterization of $\dot{V}O_{2P}$ kinetics (57). As such, accurate identification of the ϕ1-2
191 transition is paramount.
192
193 When using the mono-exponential Equation (2) to fit $\dot{V}O_{2P}$ data, human error in selecting the
194 ϕ1-2 transition can lead to an unintended bias in $\tau\dot{V}O_{2P}$ estimation, and so an ideal, unbiased
195 method for isolating ϕ2 data for such a fit would be based on either (i) identification of some
196 *consistent* time period (rather than leaving the choice to the individual researcher) at the start
197 of exercise during which data should be removed, or (ii) some other information in the data
198 itself that could algorithmically identify the ϕ1-2 transition.
199

Rather than employing empirical time-removal methods, the abrupt change in $\dot{V}O_{2P}$ at the ϕ1-2 transition may be identifiable from the $\dot{V}O_{2P}$ data using either the peak time-derivative of the $\dot{V}O_{2P}$ 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 $\dot{V}O_{2P}$ data may be hindered by the low sampling rate (relative to the duration of ϕ1) and noise inherent in those data. Whether the use of derivatives or statistical methods to identify the ϕ1-2 transition results in improved $\tau\dot{V}O_{2P}$ 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 $\dot{V}O_{2P}$ kinetic parameter estimates using experimental data [e.g. ϕ1-2 transition and ϕ2 $\tau\dot{V}O_{2P}$; (10, 26, 39, 54)]. However, a limitation of such studies is that the true underlying $\dot{V}O_{2P}$ kinetic parameters are unknown: such experimental methods can therefore give an indication of the precision of $\dot{V}O_{2P}$ 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 $\dot{V}O_{2P}$ 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 dataset with similar sampling, noise and kinetic characteristics as experimentally-obtained $\dot{V}O_{2P}$ data, but where the underlying $\dot{V}O_{2P}$ kinetic parameters are known (33). In addition, the same clean trace can be randomly resampled and new noise added to produce further noisy datasets (but all with the same underlying kinetic parameters), analogous to obtaining experimental $\dot{V}O_{2P}$ data during repeated bouts of exercise from a

single subject. Thus, these Monte Carlo methods allow both the precision and accuracy of $\dot{V}O_{2P}$ fitting methods to be systematically assessed.

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

METHODS

We used a validated simulation of $\dot{V}O_2$ and circulatory dynamic interactions during moderate intensity cycling exercise in humans (8) that accounts for the vascular capacitances and circulatory dynamics that cause a mono-exponential $\dot{V}O_{2m}$ response to manifest at the lungs as a three-phase $\dot{V}O_{2P}$ response, with a cardiodynamic $\phi1$, a near-exponential fundamental $\phi2$, and a steady-state $\phi3$. The simulation $\dot{V}O_{2P}$ outputs initially have no noise, so the baseline $\dot{V}O_{2P}$ steady-state, $\phi1$ duration and amplitude, $\phi2$ $\tau\dot{V}O_{2P}$, and $\phi3$ $\dot{V}O_{2P}$ 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 $n = \left(z_{\alpha/2}\,\sigma/\varepsilon\right)^2$, where $z_{\alpha/2}$ is the z score associated with significance level α, σ is the estimated SD of the simulation output, and ε is the acceptable margin of error for the simulation output (equal to half the required confidence interval). We set α at 0.05 to give $z_{\alpha/2} = 1.96$, it was assumed that the SD of $\phi2$ $\tau\dot{V}O_{2P}$ (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 = 7104; 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, 10^4 clean

(time resolution = 0.1 s) $\dot{V}O_{2P}$ simulations, each with different kinetics, were produced (see Fig. 1*A* 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 physiologic variation in, for example, baseline $\dot{V}O_{2P}$, $\dot{V}O_{2P}$ gain ($\Delta\dot{V}O_{2P}/\Delta W$), the relative increase in cardiac output ($\Delta\dot{Q}_m/\Delta\dot{V}O_{2m}$), and the kinetics of cardiac output and $\dot{V}O_{2m}$ ($\tau\dot{Q}_m/\tau\dot{V}O_{2m}$). Parameter sets that resulted in venous O_2 concentration dropping to zero at any point during the simulated exercise transient were discarded, and a new parameter set was generated.

Each of these 2×10^4 clean traces (one set of UP-WR, and one set of WR-WR simulations) was then sampled at a variable breathing frequency. The sampling interval was based on the relationship between breathing frequency (bf) and $\dot{V}O_{2P}$ in data collected during moderate intensity exercise in our laboratory, and was given by $\mathrm{bf}(t) = 8 \times \dot{V}O_{2P}(t) + 8$. Gaussian noise with an SD of $0.25 \times \mathrm{bf}(t)$ was subsequently added to this interval (11, 28), with the noise constrained to be no greater than 2 SDs to avoid unphysiologically-large intervals between sampled "breaths".

We then added Gaussian $\dot{V}O_{2P}$ noise to each "breath": the SD of this noise distribution was randomly sampled for each clean trace from a Gaussian distribution with a mean of 67.96 $ml.min^{-1}$ and an SD of 25.54 $ml.min^{-1}$ [calculated from the individual values reported in Lamarra et al. (33) and Rossiter et al. (47); n = 22], with the obtained value constrained to be within 2 SD of the mean, to avoid datasets that were unphysiologically noisy.

These procedures produced, from the clean simulation output, a trace with the sampling, noise and kinetic characteristics observed in experimentally-collected data (see Fig. 1*B* for examples). For all 2 x 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-B* for examples). At the end of this Monte Carlo procedure, we therefore had 10^4 noisy UP-WR datasets, i.e. 10^4 "subjects", each with different physiological characteristics, who performed moderate intensity step exercise from unloaded pedaling: each dataset 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 datasets, with each dataset again containing 10 traces from separate exercise bouts, were produced. Thus, a total of 2 x 10^5 simulated moderate-intensity "exercise bouts" in 2 x 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 Equation (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 dataset, 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): (i) Interpolation of each bout to 1-s intervals before ensemble averaging across bouts ("interpolated"); (ii) 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"); (iii) Superimposition, or stacking, of the data from different bouts, with no further interpolation or averaging

312 ("stacked"); (iv) Fitting of individual bouts (see below) followed by averaging of fit
313 parameters across bouts ("separate").
314
315 *Data fitting:* For each processed $\dot{V}O_{2P}$ trace, we fit the bi-exponential Equation (1) to the
316 entire ϕ1 and ϕ2 data, and used the following strategies for identification of the ϕ1-2
317 transition and subsequently fit the mono-exponential Equation (2) to the isolated ϕ2 data: (i)
318 Empirical time-removal methods, where 10, 15, 20, 25 or 30 s of data were removed from the
319 beginning of each processed $\dot{V}O_{2P}$ trace. (ii) Use of $\dot{V}O_{2P}$ time derivatives on both
320 unsmoothed and smoothed (with a moving 5-breath average) processed data, where the
321 highest derivative of $\dot{V}O_{2P}$ with respect to time during the first 60 s of exercise was taken as
322 the ϕ1-2 transition. (iii) Statistical methods to identify the ϕ1-2 transition, where a datum was
323 incrementally removed from the beginning of each dataset (until 60 s into exercise) and the
324 remaining data were fit using the mono-exponential Equation (2); the reduced chi-squared
325 (χ^2_{red}), adjusted coefficient of determination ($\bar{R}^2$), confidence interval for the time constant
326 (CI_τ) and the corrected Akaike information criterion (AICc) were then calculated for each fit
327 (42, 46); the first datum in the fit that returned the minimum statistical value (or maximum
328 for $\bar{R}^2$) was taken as the identified ϕ1–2 transition for that statistical method; See Rossiter et
329 al. (48) for an example using CI_τ to identify the ϕ1–2 transition. For each processed trace we
330 therefore obtained 12 fits to the data: one using the bi-exponential fit to the entire ϕ1 and ϕ2
331 data, and 11 using a mono-exponential fit to isolated ϕ2 data (five using empirical time
332 removal methods, two using $\dot{V}O_{2P}$ time derivatives, and four using statistical measures). As a
333 control condition, for each processed trace we also fit the true isolated noisy ϕ2 data with
334 Equation (2), i.e. the data were fit beginning at the true first "breath" in ϕ2, known from the
335 clean simulation. Each of these 13 methods provided an estimate of the ϕ1–2 transition [i.e.
336 *TD* from Equation (1) when using the bi-exponential fit, or the identified first breath in ϕ2

when using the mono-exponential fits] and an estimate of $\phi 2\ \tau\dot{V}O_{2P}$ [i.e. τ_2 from fits using Equation (1), or τ from fits using Equation (2)]. The $\phi 1$ amplitude (as a percentage of the steady-state response) was estimated from the value of the fit at the identified $\phi 1$–2 transition. Each of the $\phi 1$-2 transition, $\phi 1$ amplitude and $\phi 2\ \tau\dot{V}O_{2P}$ estimates were then compared to the known true underlying values obtained from the clean simulated $\dot{V}O_{2P}$ trace. These true values represent the most accurate estimates possible of $\phi 1$ and $\phi 2\ \dot{V}O_{2P}$ kinetics.

Numerical methods and statistical analyses: Details of the model used to produce the clean $\dot{V}O_{2P}$ data, along with numerical methods, are given in Benson et al. (8). Because of its unique piecewise nature, Equation (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 mean ± 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 $P < 0.05$.

352 **RESULTS**

353

354 *Simulation outputs:* Simulation input WR and output $\dot{V}O_{2P}$ characteristics are summarized in
355 Table 2. Time of the ϕ1-2 transition was significantly different between UP-WR and WR-
356 WR simulations (19.8 ± 3.4 s vs. 15.9 ± 3.4 s, respectively; $P < 0.05$, *t*-test), as was ϕ1
357 amplitude (reported as percentage of the steady-state response: 28.2 ± 8.3 % vs. 28.9 ± 7.7 %,
358 respectively; $P < 0.05$, *t*-test) and ϕ2 $\tau\dot{V}O_{2P}$ (22.4 ± 7.2 s vs. 25.0 ± 7.2 s, respectively; $P <$
359 0.05, *t*-test). These different $\dot{V}O_{2P}$ characteristics from UP-WR and WR-WR protocols can be
360 explained by the increased baseline cardiac output associated with starting an exercise
361 transition from a raised WR: muscle-to-lung transit time is shortened, reducing ϕ1 duration
362 (3), and the altered blood flow during the exercise transient modifies the association between
363 muscle and pulmonary $\dot{V}O_2$ kinetics (8). The Monte Carlo simulation output data (10^4 clean
364 UP-WR traces and 10^4 clean WR-WR traces, along with the corresponding 2 x 10^5 noisy
365 traces, and details of the input and output characteristics for each simulation) are available
366 from the corresponding author upon request.

367

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

371

372 *Number of averaged exercise bouts:* Figure 2 shows the effects of averaging exercise bouts
373 on the precision and accuracy of ϕ2 $\tau\dot{V}O_{2P}$ estimation (generally the parameter of most
374 interest) during UP-WR simulations. For this example, data from different bouts were
375 interpolated to 1-s intervals then ensemble averaged (see "Averaging methods" below), and
376 fitting was made beginning at the known first breath in ϕ2 (i.e. control fits). Qualitatively

similar results were found for the other averaging and fitting methods. The mean and SD of the estimated $\phi 2\ \tau\dot{V}O_{2P}$ are shown in Fig. 2*A*, and example distributions of the estimated $\phi 2\ \tau\dot{V}O_{2P}$ for 1, 4 and 10 exercise bouts are shown in Fig. 2*B*. The $\phi 2\ \tau\dot{V}O_{2P}$ estimates obtained by averaging 1, 2 or 3 bouts were significantly greater than using 10 bouts ($P < 0.05$, ANOVA; there was no difference when averaging 4-9 bouts; Fig. 2*A*). This indicates that precision and accuracy of $\phi 2\ \tau\dot{V}O_{2P}$ estimation is not statistically improved by averaging data from more than four bouts of exercise.

Figs. 2*A* and 2*B* demonstrate that $\tau\dot{V}O_{2P}$ tends to be overestimated on average by ~2 s, irrespective of the number of bouts averaged: mean difference between estimated and true $\tau\dot{V}O_{2P}$ 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 2*C* shows the percentage of estimated $\phi 2\ \tau\dot{V}O_{2P}$ values that lay within ± 2 s of true. Using data from a single exercise bout, the estimated $\phi 2\ \tau\dot{V}O_{2P}$ 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 $\phi 2$ is known precisely (see also "Data fitting and kinetic characterization" below). The asymptote of this relationship is 62.0% (Fig. 2*C*), indicating that the maximum probability of returning a $\phi 2\ \tau\dot{V}O_{2P}$ estimate within 2 s of true is 62%, even when the first breath in $\phi 2$ is known and no matter how many bouts are averaged.

Averaging methods: Figure 3*A* shows the effects on $\phi 2\ \tau\dot{V}O_{2P}$ 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 $\phi 2$ (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 $\phi 2\ \tau\dot{V}O_{2P}$

estimates ($P < 0.05$, ANOVA), although the mean $\phi 2\ \tau\dot{V}O_{2P}$ 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 $\phi 2\ \tau\dot{V}O_{2P}$ estimation with the interpolation method was 1.68 ± 2.06 s from true (53.0% of values within ± 2 s of true), compared to 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 $\phi 2\ \tau\dot{V}O_{2P}$ are shown in Fig. 3*B*. 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 to 6 compare the different methods for estimating the ϕ1-2 transition (Fig. 4), and the subsequent estimation of ϕ1 amplitude (Fig. 5) and $\phi 2\ \tau\dot{V}O_{2P}$ (Fig. 6), during UP-WR simulations. In Figs. 5 and 6, the distributions of ϕ1 amplitude and $\phi 2\ \tau\dot{V}O_{2P}$ estimates obtained from control fits (i.e. fits from the known first ϕ2 breath) are shown as dashed curves. The examples shown use data from four bouts averaged using the interpolation 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 (Panel *B* in Figs. 4-6) resulted in the accurate identification of the first breath in ϕ2, and ϕ1 amplitude and $\phi 2\ \tau\dot{V}O_{2P}$ 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 ϕ1-2 transition was within ±2 breaths of true in 99.3% of cases, estimated ϕ1 amplitude was within ±5% of true in 32.6% of cases (vs. 34.2% with control fits), and estimated $\phi 2\ \tau\dot{V}O_{2P}$ was within ±2 s of true in 46.5% of cases (vs.

53.0% with control fits). Although the bi-exponential fitting method (Panel *A* in Figs. 4-6) returned the second best estimates of the ϕ1-2 transition (93.8% of estimates within ±2 breaths of true), the over-parameterization of the model resulted in less accurate and precise ϕ2 $\tau\dot{V}O_{2P}$ 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) (Panel *B* in Figs. 4-6). Interestingly, removal of 15 s of data (i.e. including some ϕ1 data in the fit) gave more accurate and precise ϕ1 amplitude and ϕ2 $\tau\dot{V}O_{2P}$ estimates than removal of 25 s of data (i.e. excluding the initial portion of ϕ2 data). Basing ϕ1-2 identification on time-derivative or statistical methods resulted in skewed distributions (Fig. 4*C*,*D*), and ϕ1-2 transition, ϕ1 amplitude and ϕ2 $\tau\dot{V}O_{2P}$ values that were furthest from true (Figs. 5*C*,*D* & 6*C*,*D*).

Optimal protocol: Having identified that removal of the first 20 s of data, followed by a mono-exponential fit to the isolated ϕ2 data, was the optimal fitting method for UP-WR transitions, we repeated the previous analyses that were performed on the control, i.e. known ϕ2, 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 ϕ1-2 transition, ϕ1 amplitude and ϕ2 $\tau\dot{V}O_{2P}$ estimates, that were not significantly different to the control fits. Quantitatively, the mean estimate of the ϕ1-2 transition was 0.06 ± 0.85 breaths from true, with 99.3% of values within ±2 breaths of true; the mean ϕ1 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 ϕ2 $\tau\dot{V}O_{2P}$ 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) ϕ2 $\tau\dot{V}O_{2P}$ estimates to the interpolated method: 2.00 ± 2.19 s and 1.98 ± 2.16 s from true, respectively. Using the optimal methods, the asymptote of the exponential fit to the proportion of ϕ2 $\tau\dot{V}O_{2P}$ estimates within ± 2 s across all numbers of averaged bouts (Fig. 7*C*) 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 (Fig. 8*A*-*C*), using interpolated, binned or stacked data, optimized ϕ2 $\tau\dot{V}O_{2P}$ estimation while minimizing the number of required bouts (Fig. 8*D*-*E*). However, for WR-WR data, removal of the first 15 s or 20 s of data gave statistically similar results to control fits (where the first breath in ϕ2 is known), although quantitatively the removal of 15 s of data gave more precise and accurate estimates of $\dot{V}O_{2P}$ kinetics than removing 20 s of data: 97.2% (with 15 s removal) vs. 93.1% (with 20 s removal) of the ϕ1-2 transition estimates within ±2 breaths of true; 41.5% vs. 16.9% of ϕ1 amplitude values within ±5% of true; and 61.9% vs. 57.6% of ϕ2 $\tau\dot{V}O_{2P}$ values within ±2 s of true (Fig. 8*F*). Phase 2 $\tau\dot{V}O_{2P}$ estimation was more accurate for WR-WR data than for UP-WR data: using four interpolated and ensemble averaged bouts with ϕ2 isolated by removal of the first 15 s of data, the mean difference between estimated and known $\tau\dot{V}O_{2P}$ was 1.04 ± 2.30 s (vs. 1.97 ± 2.08 s with the optimal UP-WR analysis; $P < 0.05$, *t*-test) and the percentage of values lying within ± 2 s of the true value was 61.9% (vs. 46.5% with UP-WR data). The asymptote of the exponential fit to these data (Fig. 8*C*) suggested that a maximum of 75.9% of ϕ2 $\tau\dot{V}O_{2P}$ values would lie within ± 2 s of the true value (vs. 51.3% for UP-WR data).

476 *Minimally important difference*: The optimal collection, handling and fitting procedures for
477 UP-WR and WR-WR simulations were used to determine the minimally important difference
478 for significant changes in $\tau\dot{V}O_{2P}$ during moderate intensity exercise. Table 3 shows that the
479 95% confidence limits of $\tau\dot{V}O_{2P}$ estimation narrows from 8.25 s to 4.08 s for UP-WR, and
480 from 9.43 s to 4.51 s for WR-WR, as the number of bouts averaged is increased from 1 to 4.
481 These data propose a minimal important difference of ~5 s to detect differences in $\tau\dot{V}O_{2P}$
482 among groups or within individuals for comparative or interventional studies.
483
484 *Robustness of Monte Carlo simulations:* To confirm the robustness of the Monte Carlo
485 simulations, the entire data production procedure was repeated (i.e. a second set of 10^4 UP-
486 WR and 10^4 WR-WR clean simulations was produced, and noise was added to each trace 10
487 times, to give 2 x 10^5 noisy traces) and these data were analyzed as described above. There
488 were no differences in the key findings with this second set of simulations (data not shown).
489 As with the original Monte Carlo data, the output data from this second set of Monte Carlo
490 simulations (2 x 10^4 clean and 2 x 10^5 noisy traces, along with simulation input and output
491 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 x 10^5 simulated $\dot{V}O_{2P}$ datasets with similar sampling, noise and kinetic characteristics as experimentally-obtained $\dot{V}O_{2P}$ data. As the true underlying $\dot{V}O_{2P}$ kinetic parameters of these datasets 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 $\tau\dot{V}O_{2P}$; something that is not feasible using experimentally-obtained data where the true underlying $\tau\dot{V}O_{2P}$ is not known. We showed that the optimal data handling steps to give the most accurate and precise estimation of $\tau\dot{V}O_{2P}$ 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 mono-exponential 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 $\tau\dot{V}O_{2P}$ 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.

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 sub-LT exercise (Table 2). Simulated $\phi 1$ duration ranged from 7 s to 30 s and was 9% to

72% of the steady-state response in amplitude, and simulated $\phi 2$ $\tau\dot{V}O_{2P}$ spanned approximately 7 s to 40 s, across transitions ranging from 50 W 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 $\tau\dot{V}O_{2P}$ 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 $\tau\dot{V}O_{2P}$. Some investigators may be willing to accept lower accuracy and precision in $\tau\dot{V}O_{2P}$ estimation in order 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 $\tau\dot{V}O_{2P}$ estimations 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 to 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 $\tau\dot{V}O_{2P}$, e.g. for use in comparative and interventional studies, from ~5 s to ~6 s. The data shown in Table 3 can be used to inform such decisions.

Our 4-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 $ml.min^{-1}$). Nevertheless, our simulated transitions mimicked very well typical observations using many standard gas exchange measurement approaches. Our findings indicate that in order to provide more precise estimations of $\tau\dot{V}O_{2P}$ 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. (32), who also used a Monte Carlo approach to show that increasing $\dot{V}O_{2P}$ noise, expressed as a

542 percentage of the steady-state change in the $\dot{V}O_{2P}$ response, increased the confidence
543 intervals for the estimated fit parameters (ϕ1 duration and ϕ2 $\tau\dot{V}O_{2P}$). We showed that
544 approaches that increase the signal-to-noise ratio have a substantial effect on precision, but
545 little effect on accuracy, of kinetic estimates. These fluctuations are expected to arise from
546 the interaction of a number of variables, not least the breath-by-breath variations in tidal
547 volume and pulmonary blood flow, within which fluctuation and timing of stroke volume and
548 thoracic pressure changes may variably sum or counteract one another to give rise to
549 fluctuations in gas exchange. Therefore, algorithms for breath-by-breath gas exchange
550 measurement that reduce the inherent fluctuation of the data, e.g. by accounting for changes
551 in alveolar gas storage, or by re-characterizing a breath to be equal to a tidal breathing cycle
552 that returns to an identical end-expiratory lung volume (6, 13), would be expected to further
553 reduce the testing burden while maintaining optimal precision and accuracy of kinetic
554 estimates.
555
556 *Data processing:* Although there are many possible methods for data averaging, the four
557 techniques examined in this study (interpolation, binning, stacking, and separate fitting)
558 provide a cross-section of the most commonly used methods. Although we have identified
559 linear interpolation followed by ensemble averaging as the optimal method for averaging data
560 [similar to the findings of Keir et al. (26)], both the breath binning and stacking methods
561 produced quantitatively similar estimates of $\tau\dot{V}O_{2P}$. As such, researchers who have
562 previously used, or currently use, any of these methods should be confident that their choice
563 of averaging procedure does not unduly influence their estimates of $\tau\dot{V}O_{2P}$. While averaging
564 of the exponential fit parameters from separate bouts of exercise offers the simplicity of
565 avoiding potentially complicated and assumption-laden averaging procedures on large
566 datasets, $\tau\dot{V}O_{2P}$ 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 non-linear: 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 (Figure 3*B*, 7*E* and 8*E*). 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 3*B*), 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 $\phi 2$ data for fitting resulted in significantly more accurate and precise estimations of $\tau\dot{V}O_{2P}$ than either a bi-

592 exponential fit, or statistical and time-derivative methods to identify the ϕ1-2 transition
593 followed by a mono-exponential fit to the isolated ϕ2 data. The majority of published
594 experimental studies that have quantified the kinetics of $\dot{V}O_{2P}$ have used such empirical time
595 removal methods (usually removing the first 20 s of data), and so researchers have
596 historically used the ϕ2 isolation method that we have now shown provides the most accurate
597 and precise estimations of $\tau\dot{V}O_{2P}$. Furthermore, this empirical time removal approach is far
598 simpler to implement than the bi-exponential, statistical or time-derivative methods. Previous
599 recommendations have been to remove *at least* 20 s of data from the beginning of the dataset
600 in order to completely remove ϕ1 data, even though some data from the start of ϕ2 may also
601 be removed (7, 57). However, our results suggest that, somewhat counter-intuitively, it is
602 better to include a small amount of data from the end of ϕ1 in the fitting procedure than
603 exclude data from the start of ϕ2. This is seen in Figs. 5*B* and 6*B*, where ϕ1 amplitude and ϕ2
604 $\tau\dot{V}O_{2P}$ estimation for exercise from unloaded pedaling was more precise and accurate when
605 the initial 15 s of data were removed than when the initial 25 s of data were removed (the true
606 ϕ1-2 transition for these data occurred at 19.5 ± 3.3 s). We suggest that this is because the
607 inherent fluctuations in the $\dot{V}O_{2P}$ data means that including a small amount of ϕ1 data in the
608 fit has minimal effect on the resultant ϕ1 amplitude and ϕ2 $\tau\dot{V}O_{2P}$ estimation. The rapidly
609 changing initial portion of ϕ2 data (which changes rapidly with respect to the breath-by-
610 breath fluctuations at the end of ϕ1) is key to obtaining accurate and precise estimations.
611 Qualitatively similar results were found for exercise that started from a raised work rate, but
612 here the best $\tau\dot{V}O_{2P}$ estimation was with the removal of the first 15 s of data (Fig. 8*F*). This is
613 likely due to the increased baseline work rate elevating cardiac output, which reduces muscle-
614 to-lung blood transit times and, therefore, the cardiodynamic ϕ1 duration. Nevertheless, the
615 accuracy and precision of $\tau\dot{V}O_{2P}$ estimation was statistically similar for WR-WR transitions
616 when either 15 s or 20 s of data were removed. We therefore recommend that researchers err

617 on the side of caution when isolating ϕ2 $\dot{V}O_{2P}$ data and remove *no more than* 20 s of data to
618 optimize $\tau\dot{V}O_{2P}$ estimation.
619
620 *Implications for interpretation of ϕ2 $\dot{V}O_{2P}$ kinetics:* There are two significant findings from
621 our simulations that have implications for interpretation of ϕ2 $\dot{V}O_{2P}$ kinetics. Firstly, we
622 found that, on average, ϕ2 $\tau\dot{V}O_{2P}$ was overestimated in all the data collection and handling
623 strategies investigated. This overestimation can be explained, at least in part, by the two-
624 phase $\dot{V}O_{2P}$ response and the non-exponentiality of ϕ2 (3, 5, 8, 25; cf. 18). Figure 9 shows
625 the effects on ϕ2 $\tau\dot{V}O_{2P}$ estimation when the mono-exponential Equation (2) is fit to clean
626 simulation output data from different points throughout the $\dot{V}O_{2P}$ response. If the mono-
627 exponential fit is started during ϕ1 (i.e. from any point before 19.4 s in this example) then the
628 estimated ϕ2 $\tau\dot{V}O_{2P}$ is larger than true, due to the inclusion of some ϕ1 data in the fit. If the
629 fit is started after the ϕ1-2 transition, then the ϕ2 $\tau\dot{V}O_{2P}$ estimation is also larger than true,
630 becoming larger as the fit is started further from the ϕ1-2 transition, because the underlying
631 ϕ2 response is not a pure mono-exponential; it initially increases more rapidly than a mono-
632 exponential before slowing down as it reaches the steady-state (8). Only a fit that starts
633 exactly at the ϕ1-2 transition returns the true ϕ2 $\tau\dot{V}O_{2P}$. For these clean simulated data,
634 inaccurate identification of the ϕ1-2 transition by just 2 s can result in a ϕ2 $\tau\dot{V}O_{2P}$ estimation
635 that is 1.6 s larger than the true value; the influence of noise in experimentally-obtained data
636 may exacerbate this error. Because of these effects on ϕ2 $\tau\dot{V}O_{2P}$ estimation, when using the
637 identified optimal data processing and fitting procedures we were only able to estimate ϕ2
638 $\tau\dot{V}O_{2P}$ to within 2 s of true in 47% of the 10^4 UP-WR simulations, and in 62% of the 10^4
639 WR-WR simulations [2 s represents an effect size of ~10% for a healthy young human,
640 where $\tau\dot{V}O_{2P}$ is typically ~20 s (45)]. Extrapolating this analysis further, we calculated the
641 95% confidence limits of our $\tau\dot{V}O_{2P}$ estimate distributions (as shown in Figs. 7*B* and *D*, and

642 Figs. 8*B*, *D* and *F*); $\tau\dot{V}O_{2P}$ estimates from outside this confidence interval are statistically
643 likely to come from a different distribution/population. These 95% confidence limits, for
644 $\tau\dot{V}O_{2P}$ estimates using our predetermined optimal data processing and fitting procedures, are
645 ± 4.08 s and ± 4.51 s from the mean, for transitions from unloaded pedaling or a raised work
646 rate respectively (Table 3). We therefore propose that the minimally important difference for
647 a significant change in $\tau\dot{V}O_{2P}$, e.g. during interventional and comparative studies, should be
648 5.0 s. If the number of averaged bouts is reduced from the optimum of four, this minimally
649 important difference should be increased in accordance with the confidence limits shown in
650 Table 3.
651
652 The second implication for interpretation of $\tau\dot{V}O_{2P}$ from our data is to question whether an
653 exponential fit should be used at all. We have previously shown that the dynamics and
654 mixing of circulatory compartments between muscle and lung distort the approximately-
655 exponential muscle $\dot{V}O_2$ kinetics into a non-exponential $\phi 2$ $\dot{V}O_{2P}$ response at the lung (8). A
656 recent meta-analysis of available data measuring both muscle and lung $\dot{V}O_2$ kinetics during
657 cycling and knee extension exercise demonstrates a wide variability of $\tau\dot{V}O_2$ between muscle
658 and lung (27). Some have proposed alternative methods to assess kinetic responses, such as
659 the time to steady state (45). However, such approaches have been demonstrated to be both
660 inherently more variable than relying on a method that maximizes the utility of available non-
661 steady-state data (24, 47) and is conceptually flawed on the basis that the time to steady state
662 of a non-exponential process is continually changing (8). Alternative approaches to kinetics
663 estimation using, for example, pseudorandom binary sequence exercise testing and time-
664 series analysis may allow for muscle $\tau\dot{V}O_2$ to be resolved by alternative methods (24, 31). It
665 remains to be determined whether such methods provide increased accuracy for non-invasive
666 estimation of muscle $\dot{V}O_2$ kinetic responses compared with $\phi 2$ $\tau\dot{V}O_2$ estimation by repeated

667 step transitions. Our simulations here demonstrate that a mono-exponential fit to $\phi 2\ \dot{V}O_{2P}$ is a
668 useful and concise method for accurately describing the overall kinetics of the exponential-
669 like pulmonary $\phi 2\ \dot{V}O_2$ kinetic response.
670
671 *Limitations:* The means and SDs of the parameters used in our Monte Carlo simulations were
672 representative of healthy young adults (8, 22). Quantitatively different results may be found
673 for other populations with different $\phi 2\ \dot{V}O_{2P}$ kinetic parameters, such as the elderly or heart
674 failure patients who have slowed $\dot{V}O_{2P}$ kinetics (9, 39). Nevertheless, our main qualitative
675 findings will still be pertinent when collecting, processing and fitting $\dot{V}O_{2P}$ data from these
676 other populations. In particular, our main point regarding optimal data collection and
677 processing methods – that methods should be employed to minimize breath-by-breath
678 fluctuations and that it is essential to include all $\phi 2\ \dot{V}O_{2P}$ data in the fit – will more than
679 likely stand for these populations, as it is still expected that the (potentially slowed) initial
680 portion of $\phi 2\ \dot{V}O_{2P}$ will change rapidly with respect to the noise in the data at the end of $\phi 1$.
681
682 For populations where individuals are expected to have a reduced cardiac output and slowed
683 cardiac output kinetics, and a concomitant prolongation of $\phi 1$ duration compared to young
684 healthy adults [such as heart failure patients (52)], the use of a bi-exponential fit, or statistical
685 or derivative methods, to automatically identify the $\phi 1$-2 transition is inherently attractive.
686 However, our results highlight that the noise in the $\dot{V}O_{2P}$ data limit the ability of these
687 methods to correctly identify the $\phi 1$-2 transition, reducing the accuracy and precision of
688 subsequent $\tau\dot{V}O_{2P}$ estimation. In this study, the empirical time-removal methods (removal of
689 the first 20 s of data for exercise from unloaded pedaling, or 15 s if exercise was started from
690 a raised baseline) were the only methods that gave statistically similar $\tau\dot{V}O_{2P}$ estimates to
691 control fits, despite $\phi 1$ duration ranging from 7 s to 39 s across all simulations. It remains to

692 be determined whether removal of the first 20 s of data results in the most accurate and
693 precise $\tau\dot{V}O_{2P}$ estimations for populations where $\phi1$ is prolonged, but it may be necessary to
694 compensate for the prolonged $\phi1$ duration when removing $\phi1$ data from the fitting window.
695
696 Only on-transient exercise in the moderate intensity domain was simulated in this study. It is
697 still to be determined whether the identified optimal fitting procedures will produce the most
698 accurate and precise $\tau\dot{V}O_{2P}$ estimations for on-transient data in higher exercise intensity
699 domains where fitting can be complicated by the emergence of a $\dot{V}O_{2P}$ slow component (40,
700 45). Similarly, the applicability of our identified optimal procedures for off-transient data,
701 where cardiac output is expected to be initially elevated and so produce a much shorter $\phi1$,
702 potentially influencing the amount of data that should be removed before fitting, is still to be
703 determined.
704
705 **CONCLUSIONS**
706
707 We used a validated computational model together with a Monte Carlo approach to assess the
708 accuracy and the precision of various averaging and fitting procedures on the estimation of
709 $\dot{V}O_{2P}$ kinetics. Our analyses showed that four bouts of exercise was the optimal number to
710 average in order to increase accuracy and precision of $\tau\dot{V}O_{2P}$ estimation. Choice of averaging
711 strategy was not so critical, with interpolation, bin averaging and stacking all giving
712 quantitatively similar $\tau\dot{V}O_{2P}$ estimates. The interpolation, binning and stacking methods did,
713 however, allow more confident parameter estimates when compared to analyzing repeated
714 bouts separately. Data collection and processing strategies should therefore focus on
715 increasing the signal-to-noise ratio in the collected data. Contradictory to previous advice that
716 suggests removal of *at least* 20 s of data to isolate $\phi2$ $\dot{V}O_{2P}$ before fitting, our analyses show

717 that data fitting procedures should remove *no more than* 20 s of data, as this provided the
718 most precise and accurate estimates of $\tau\dot{V}O_{2P}$. Our analyses showed the widely used standard
719 approaches for data collection, processing and fitting, while often different between
720 laboratories, did not have a substantial effect on the quantitation of $\phi 2$ $\dot{V}O_{2P}$ kinetics *per se*.
721 However, we found that even this optimal procedure yielded $\tau\dot{V}O_{2P}$ estimates that were
722 within ± 2 s of true in only 47-62% of simulations. Thus, we identified the minimally
723 important difference for $\tau\dot{V}O_{2P}$ 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

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

AUTHOR CONTRIBUTIONS

All authors conceived and designed the study; A.P.B. carried out the simulations and processed and analyzed the collected data; all authors interpreted the results of the simulations; A.P.B. prepared the figures and the first draft of the manuscript; all authors edited and approved the final version of the manuscript.

737 **REFERENCES**

738

739 1. **Babcock MA, Paterson DH, Cunningham DA, Dickinson JR.** Exercise on-transient
740 gas exchange kinetics are slowed as a function of age. *Med Sci Sports Exerc* 26: 440-
741 446, 1994.
742 2. **Bangsbo J, Krustrup P, González-Alonso J, Boushel R, Saltin B.** Muscle oxygen
743 kinetics at onset of intense dynamic exercise in humans. *Am J Physiol Regul Integr*
744 *Comp Physiol* 279: R899-R906, 2000.
745 3. **Barstow TJ, Lamarra N, Whipp BJ.** Modulation of muscle and pulmonary O_2 uptakes
746 by circulatory dynamics during exercise. *J Appl Physiol* 68: 979-989, 1990.
747 4. **Barstow TJ, Molé PA.** Linear and nonlinear characteristics of oxygen uptake kinetics
748 during heavy exercise. *J Appl Physiol* 71: 2099-2106, 1991.
749 5. **Barstow TJ, Molé PA.** Simulation of pulmonary O_2 uptake during exercise transients in
750 humans. *J Appl Physiol* 63: 2253-2261, 1987.
751 6. **Beaver WL, Lamarra N, Wasserman K.** Breath-by-breath measurement of true
752 alveolar gas exchange. *J Appl Physiol* 51: 1662-1675, 1981.
753 7. **Bell C, Paterson DH, Kowalchuk JM, Padilla J, Cunningham DA.** A comparison of
754 modelling techniques used to characterise oxygen uptake kinetics during the on-transient
755 of exercise. *Exp Physiol* 86.5: 667-676, 2001.
756 8. **Benson AP, Grassi B, Rossiter HB.** A validated model of oxygen uptake and
757 circulatory dynamic interactions at exercise onset in humans. *J Appl Physiol* 115: 743-
758 755, 2013.
759 9. **Bowen TS, Cannon DT, Murgatroyd SR, Birch KM, Witte KK, Rossiter HB.** The
760 intramuscular contribution to the slow oxygen uptake kinetics during exercise in chronic

761 heart failure is related to the severity of the condition. *J Appl Physiol* 112: 378-387,
762 2012.
763 10. **Bringard A, Adami A, Moia C, Ferretti G.** A new interpolation-free procedure for
764 breath-by-breath analysis of oxygen uptake in exercise transients. *Eur J Appl Physiol*
765 114: 1983-1994, 2014.
766 11. **Bruce EN.** Temporal variations of breathing pattern. *J Appl Physiol* 80: 1079-1087,
767 1996.
768 12. **Brunner-La Rocca HP, Weilenmann D, Follath F, Schlumpf M, Rickli H, Schalcher**
769 **C, Maly FE, Candinas R, Kiowski W.** Oxygen uptake kinetics during low level
770 exercise in patients with heart failure: relation to neurohormones, peak oxygen
771 consumption, and clinical findings. *Heart* 81: 121-127, 1999.
772 13. **Capelli C, Cautero M, di Prampero PE.** New perspectives in breath-by-breath
773 determination of alveolar gas exchange in humans. *Pflugers Arch* 441: 566-577, 2001.
774 14. **Capelli C, Cautero M, Pogliaghi S.** Algorithms, modelling and VO_2 kinetics. *Eur J*
775 *Appl Physiol* 111: 331-342, 2011.
776 15. **Casaburi R, Daly J, Hansen JE, Effros RM.** Abrupt changes in mixed venous blood
777 gas composition after the onset of exercise. *J Appl Physiol* 67: 1106-1112, 1989.
778 16. **Chechile RA.** Pooling data versus averaging model fits for some prototypical
779 multinomial processing tree models. *J Math Psychol* 53: 562-576, 2009.
780 17. **Dunn WL, Shultis JK.** *Exploring Monte Carlo Methods.* Amsterdam, The Netherlands:
781 Elsevier, 2012.
782 18. **Francescato MP, Cettolo V, di Prampero PE.** Oxygen uptake kinetics at work onset:
783 role of cardiac output and of phosphocreatine breakdown. *Respir Physiol Neurobiol* 185:
784 287-295, 2013.

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.

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.

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.

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* 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.

24. **Hoffmann U, Drescher U, Benson AP, Rossiter HB, Essfeld D.** Skeletal muscle $\dot{V}O_2$ 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.

25. **Hughson RL.** Oxygen uptake kinetics: historical perspective and future directions. *Appl Physiol Nutr Metab* 34: 840-850, 2009.

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.

27. **Koga S, Rossiter HB, Heinonen I, Musch TI, Poole DC.** Dynamic heterogeneity of exercising muscle blood flow and O_2 utilization. *Med Sci Sports Exerc* 46: 860-876, 2014.

28. **Kohl J, Koller EA, Jäger M.** Relation between pedaling- and breathing rhythm. *Eur J Appl Physiol* 47: 223-237, 1981.

29. **Koppo K, Bouckaert J, Jones AM.** Effects of training status and exercise intensity on phase II $\dot{V}O_2$ kinetics. *Med Sci Sports Exerc* 36: 225-232, 2004.

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.

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.

32. **Lamarra N.** Variables, constants, and parameters: clarifying the system structure. *Med Sci Sports Exerc* 22: 88-95, 1990.

33. **Lamarra N, Whipp BJ, Ward SA, Wasserman K.** Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. *J Appl Physiol* 62: 2003-2012, 1987.

34. **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 Benchetrit G, Baconnier P, Demongeot J. Manchester, UK: Manchester University Press, 1987.

35. **Lewis RM, Torczon V, Trosset MW.** Direct search methods: then and now. *J Comput Appl Math* 124: 191-207, 2000.

36. **Ma S, Rossiter HB, Barstow TJ, Casaburi R, Porszasz J.** Clarifying the equation for modeling of $\dot{V}O_2$ kinetics above the lactate threshold. *J Appl Physiol* 109: 1283-1284, 2010.

37. **Mezzani A, Grassi B, Giordano A, Corrà U, Colombo S, Giannuzzi P.** Age-related prolongation of phase I of $\dot{V}O_2$ on-kinetics in healthy humans. *Am J Physiol* 299: R968-R976, 2010.

38. **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* 110: 1598-1606, 2011.

39. **Murias JM, Spencer MD, Kowalchuk JM, Paterson DH.** Influence of phase I duration on phase II $\dot{V}O_2$ kinetics parameter estimates in older and young adults. *Am J Physiol Regul Integr Comp Physiol* 301: R218-R224, 2011

40. **Nery LE, Wasserman K, Andrews JD, Huntsman DJ, Hansen JE, Whipp BJ.** Ventilatory and gas exchange kinetics during exercise in chronic airways obstruction. *J Appl Physiol* 53: 1594-1602, 1982.

41. **Poole DC, Jones AM.** Oxygen uptake kinetics. *Compr Physiol* 2: 933-996, 2012.

42. **Potter CR, Childs DJ, Houghton W, Armstrong N.** Breath-to-breath "noise" in the ventilatory and gas exchange responses of children to exercise. *Eur J Appl Physiol Occup Physiol* 80: 118-124, 1999.

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: 429-60, 2016.

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.

46. **Rossiter HB.** Exercise: kinetic considerations for gas exchange. *Compr Physiol* 1: 203-244, 2011.

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.

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.

49. **Schalcher C, Rickli H, Brehm M, Weilenmann 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.

50. **Sietsema KE.** Oxygen uptake kinetics in response to exercise in patients with pulmonary vascular disease. *Am Rev Respir Dis* 145: 1052-1057, 1992.

51. **Sietsema 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.

52. **Spee RF, Niemeijer VM, 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* 121: 198-204, 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.

54. **Spencer MD, Granvelle BMR, Murias JM, Zerbini L, Pogliaghi S, Paterson DH.** Duration of "Phase I" $\dot{V}O_{2p}$: 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.

55. **Swanson GD, Sherrill DL.** On the breath-to-breath estimation of gas exchange. *J Appl Physiol* 56: 259-261, 1984.

56. **Whipp BJ, Ward SA, Lamarra N, Davis JA, Wasserman K.** Parameters of ventilatory and gas exchange dynamics during exercise. *J Appl Physiol* 52: 1506-1513, 1982.

57. **Whipp BJ, Rossiter HB.** The kinetics of oxygen uptake: physiological inferences from the parameters. In: *Oxygen Uptake Kinetics in Sport, Exercise and Medicine*, edited by Jones AM, Poole DC. London, UK: Routledge, 2005.

58. **Wüst RC, Aliverti A, Capelli C, Kayser B.** Breath-by-breath changes of lung oxygen stores at rest and during exercise in humans. *Respir Physiol Neurobiol* 164: 291-299, 2008.

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

Parameters with Gaussian distributions:	**Mean**	**SD**
Arterial O_2 concentration (ml O_2/100 ml blood)	20.0	1.00
Total venous volume (l) [a]	3.07	0.61
Baseline $\dot{V}O_{2P}$ ($l.min^{-1}$)	0.87	0.08
Fraction of baseline $\dot{V}O_{2P}$ from muscle [b]	0.57	0.11
Baseline $\dot{Q}_{tot}$ ($l.min^{-1}$)	8.89	0.44
Fraction of baseline $\dot{Q}_{tot}$ to muscle [b]	0.57	0.08
$\Delta\dot{V}O_{2P}/\Delta W$ ($ml.min^{-1}.W^{-1}$)	9.47	0.85
$\Delta\dot{V}O_{2m}/\Delta W$ ($ml.min^{-1}.W^{-1}$)	11.04	1.36
$\Delta\dot{Q}_m/\Delta\dot{V}O_{2m}$	6.03	0.53
$\tau\dot{Q}_m/\tau\dot{V}O_{2m}$ [c]	1.08	0.08
Parameters with linear distributions:	**Minimum**	**Maximum**
$\tau\dot{V}O_{2m}$ (s)	15.0	40.0
Baseline WR (for WR-WR simulations only; W) [d]	0.0	100.0
ΔWR (W) [e]	50.0	150.0
Parameters with other dependencies:		
$\Delta\dot{V}O_{2b}/\Delta W = \Delta\dot{V}O_{2P}/\Delta W - \Delta\dot{V}O_{2m}/\Delta W$		
$\Delta\dot{Q}_b/\Delta\dot{V}O_{2b} = \Delta\dot{Q}_m/\Delta\dot{V}O_{2m}$		
$\tau\dot{V}O_{2b} = \tau\dot{V}O_{2m}$		
$\tau\dot{Q}_b = \tau\dot{Q}_m$		

$\dot{Q}$ denotes blood flow (with $\dot{Q}_{tot}$ denoting cardiac output). The subscript 'm' denotes muscle compartment, the subscript 'b' denotes rest-of-body compartment. Baseline is unloaded pedaling (i.e. 0 W). [a]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 $\dot{V}O_{2P}$ (and $\dot{Q}_{tot}$) comes from (and goes to) the body compartment. [c]To avoid kinetic mismatch between muscle $\dot{Q}$ and $\dot{V}O_2$ (as occurs with slow $\dot{Q}_m$ but fast $\dot{V}O_{2m}$ kinetics, that result in muscle O_2 concentration dropping to zero), we first set the absolute $\tau\dot{V}O_{2m}$ value, and then constrained $\tau\dot{Q}_m$ to be similar to $\tau\dot{V}O_{2m}$ 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

910 WR in the WR-WR simulations (i.e. baseline WR + ΔWR) was constrained to be no greater
911 than 150 W.

912 **Table 2.** Monte Carlo simulation input WR and output $\dot{V}O_{2P}$ characteristics

	Mean	SD	Range
UP-WR simulations ($n = 10^4$):			
Baseline WR (W)	0.0	0.0	0.0 – 0.0
ΔWR (W) *	97.3	28.4	50.0 – 150.0
$\dot{V}O_{2P}$ ϕ1 duration (s)	19.8	3.4	10.8 – 31.1
$\dot{V}O_{2P}$ ϕ1 amplitude (% of steady-state response)	28.2	8.3	9.4 – 71.8
ϕ2 $\tau\dot{V}O_{2P}$ (s)	22.4	7.2	7.3 – 38.8
WR-WR simulations ($n = 10^4$):			
Baseline WR (W)	47.3	28.6	0.0 – 100.0
ΔWR (W) *	75.6	22.0	50.0 – 150.0
$\dot{V}O_{2P}$ ϕ1 duration (s)	15.9	3.4	7.4 – 29.4
$\dot{V}O_{2P}$ ϕ1 amplitude (% of steady-state response)	28.9	7.7	11.0 – 63.3
ϕ2 $\tau\dot{V}O_{2P}$ (s)	25.0	7.2	8.4 – 40.3

913 *ΔWR was constrained to be positive and at least 50 W, with the final WR (i.e. ΔWR in the
914 UP-WR simulations, and baseline WR + ΔWR in the WR-WR simulations) constrained to be
915 no greater than 150 W.

Table 3. Phase 2 $\tau\dot{V}O_{2P}$ estimates and confidence intervals for 1-4 averaged UP-WR and WR-WR exercise bouts. 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.

	Number of averaged bouts	$\phi 2$ $\tau\dot{V}O_{2P}$ 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

FIGURE LEGENDS

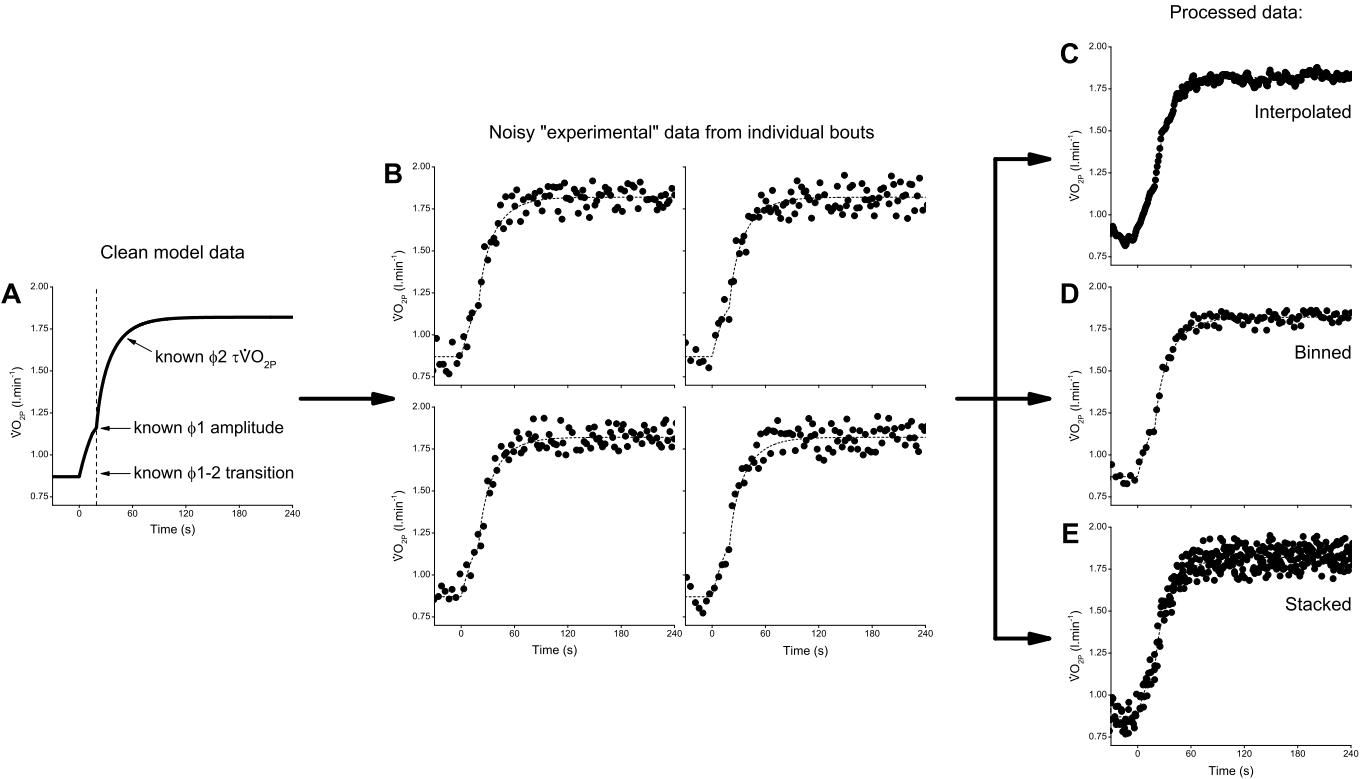
Fig. 1. Example of data production and processing during a single Monte Carlo iteration from unloaded pedaling. *A*: for each iteration, model parameters were varied stochastically (see Table 1) and a clean model $\dot{V}O_{2P}$ trace with known kinetic parameters (e.g. $\phi 1$ duration and amplitude, and $\phi 2$ $\tau\dot{V}O_{2P}$) 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" $\dot{V}O_{2P}$ traces (filled circles) with the sampling (breathing) and $\dot{V}O_{2P}$ noise characteristics seen in experimental data. Here, four examples are shown. Although each of the 10 noisy datasets is different, they have identical underlying kinetic parameters (known from the clean model trace shown in panel *A*, and shown in these panels as dashed lines). The noisy $\dot{V}O_{2P}$ datasets were processed in one of four ways: *C*: interpolation followed by ensemble averaging; *D*: bin averaging; *E*: stacking of datasets; and fitting of the separate traces before averaging of the resultant fit parameters (not shown). Fits to these processed data were compared to the true underlying kinetic parameters (known from the clean model trace shown in panel *A*, and shown in these panels as dashed lines).

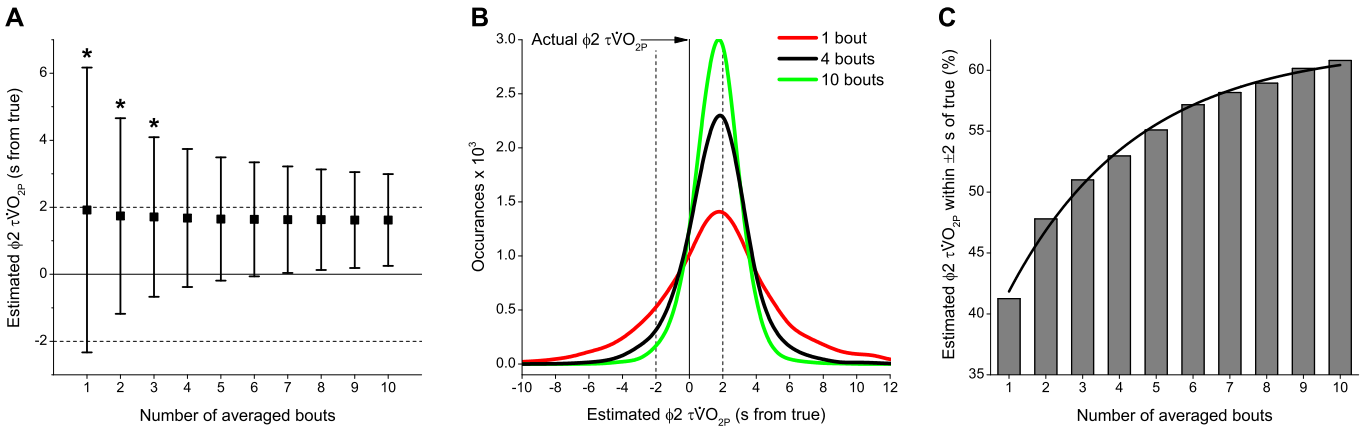
Fig 2. Effects of the number of averaged bouts on the precision and accuracy of $\tau\dot{V}O_{2P}$ estimation, using control fits (i.e. using the known $\phi 2$ data) to interpolated and ensemble averaged UP-WR data. *A*: mean ± SD difference of the estimated $\tau\dot{V}O_{2P}$ from the true value, for 1-10 averaged bouts. Horizontal lines show zero difference (solid) ± 2 s (dashed) from true. $n = 10^4$ in each case. * = $P < 0.05$ vs. 10 averaged bouts (ANOVA). *B*: distributions of the difference between estimated and true $\tau\dot{V}O_{2P}$ for 1, 4 and 10 averaged bouts. Vertical lines show zero difference (solid) ± 2 s (dashed) from true. $n = 10^4$ in each case. *C*:

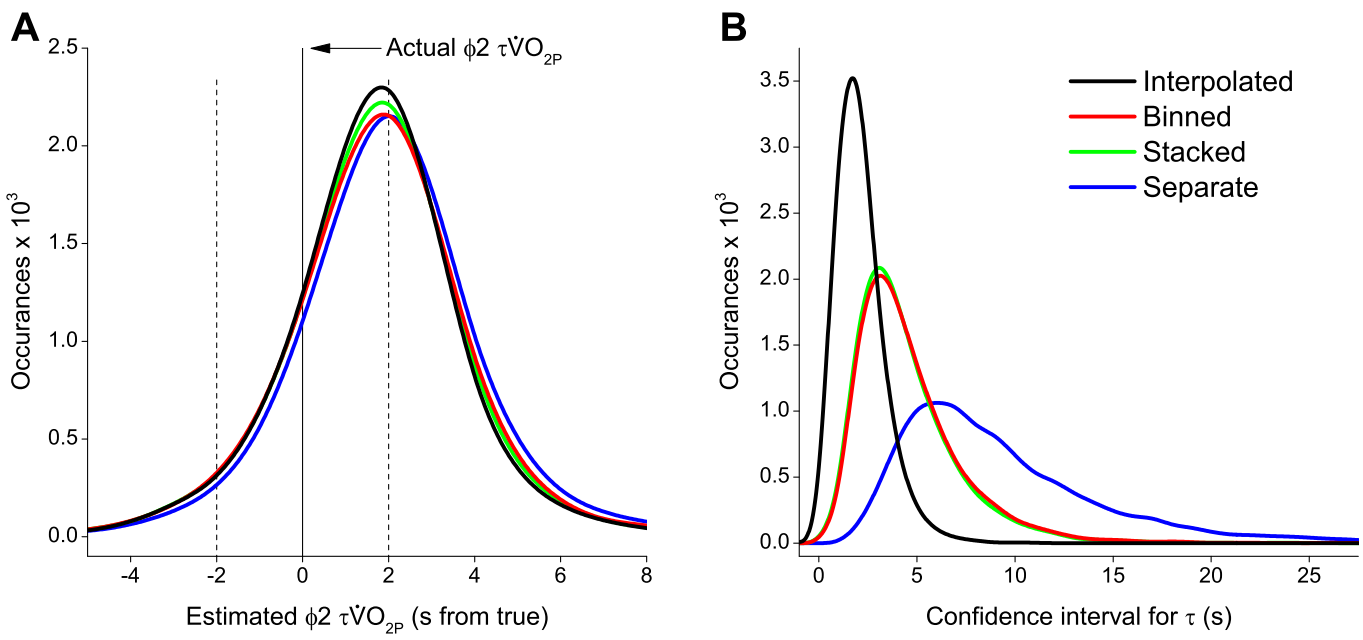
943 percentages of the 10^4 $\tau\dot{V}O_{2P}$ estimates within ± 2 s of true, for 1-10 averaged bouts. The
944 solid line is an exponential fit to the data.
945
946 Fig 3. Effects of averaging method on the precision and accuracy of $\tau\dot{V}O_{2P}$ estimation, using
947 control fits (i.e. using the known ϕ2 data) to four averaged UP-WR bouts. *A*: distributions of
948 the difference between estimated and true $\tau\dot{V}O_{2P}$ for the four different averaging methods.
949 Vertical lines show zero difference (solid) ± 2 s (dashed) from true. $n = 10^4$ in each case. B:
950 distributions of the confidence interval of the fitted τ for the four different averaging
951 methods. $n = 10^4$ in each case.
952
953 Fig 4. Effects of fitting methods on the precision and accuracy of ϕ1-2 transition
954 identification. Shown are distributions of the difference between the estimated and true ϕ1-2
955 transition for the bi-exponential fit (*A*), and empirical (*B*), statistical (*C*) and derivative (*D*) ϕ2
956 isolation methods. For all panels, vertical lines show zero difference (solid) ± 2 breaths
957 (dashed) from true, and $n = 10^4$ in each distribution. Note the different scales on the
958 abscissas.
959
960 Fig 5. Effects of fitting methods on the precision and accuracy of ϕ1 amplitude estimation.
961 Shown are distributions of the difference between the estimated and true ϕ1 amplitude for the
962 bi-exponential fit (*A*), and empirical (*B*), statistical (*C*) and derivative (*D*) ϕ2 isolation
963 methods. The control fit distribution (i.e. using the known ϕ2 data) is shown as a dashed
964 curve in each panel. For all panels, vertical lines show zero difference (solid) ± 5% (dashed)
965 from true, and $n = 10^4$ in each distribution. Note the different scales on the abscissas.
966

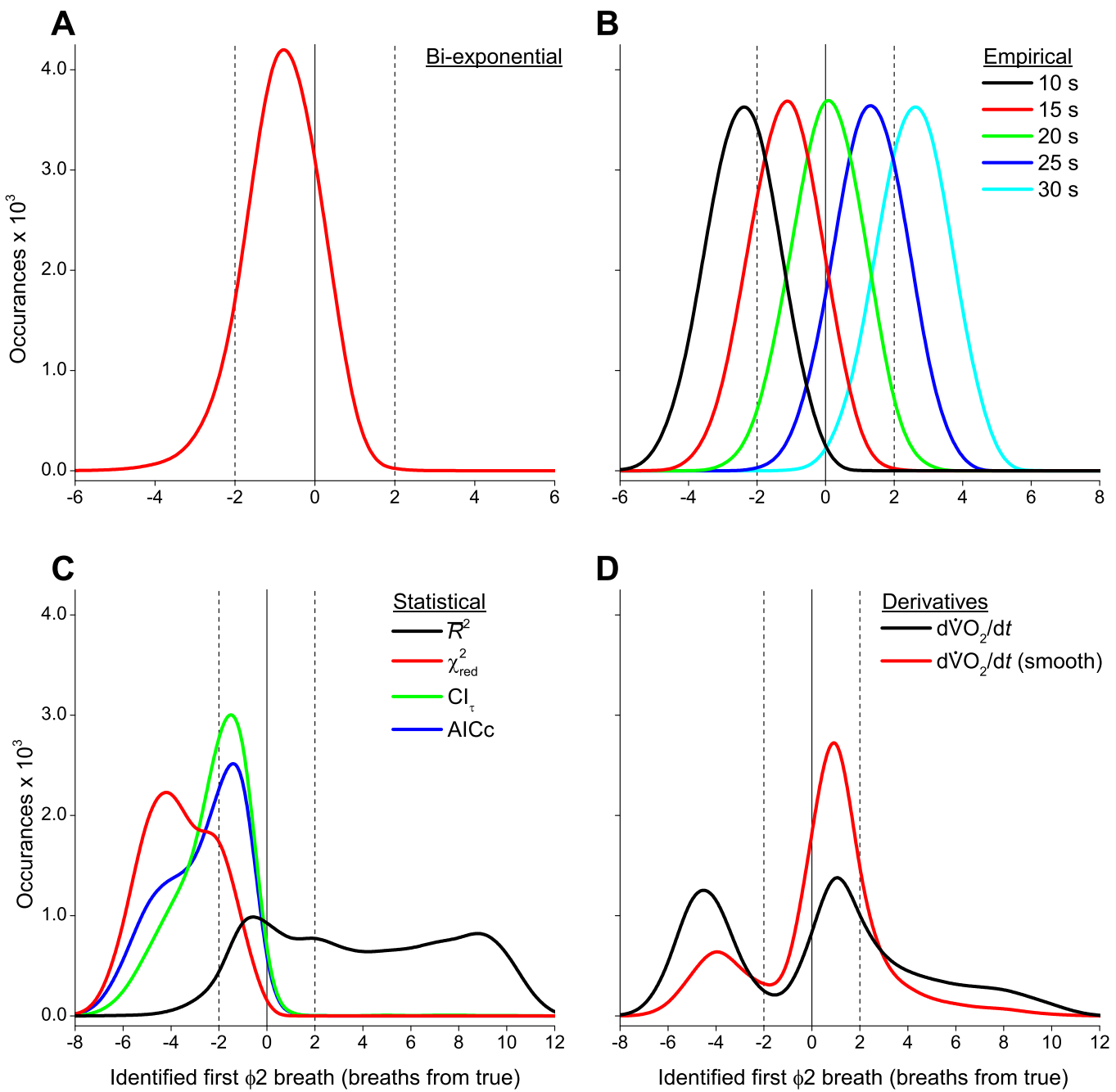
967 Fig 6. Effects of fitting methods on the precision and accuracy of ϕ2 $\tau\dot{V}O_{2P}$ estimation.
968 Shown are distributions of the difference between the estimated and true ϕ2 $\tau\dot{V}O_{2P}$ for the bi-
969 exponential fit (*A*), and empirical (*B*), statistical (*C*) and derivative (*D*) ϕ2 isolation methods.
970 The control fit distribution (i.e. using the known ϕ2 data) is shown as a dashed curve in each
971 panel. For all panels, vertical lines show zero difference (solid) ± 2 s (dashed) from true, and
972 $n = 10^4$ in each distribution. Note the different scales on the abscissas.
973
974 Fig 7. Precision and accuracy of $\tau\dot{V}O_{2P}$ estimation for UP-WR bouts when removal of the
975 first 20 s of data is used to isolate ϕ2. *A*-*C*: effects of the number of averaged bouts, where
976 data processing is by interpolation and ensemble averaging (see Fig. 2 for explanations). *D*-*E*:
977 effects of averaging method on $\tau\dot{V}O_{2P}$ estimation and the associated confidence interval,
978 using four averaged bouts (see Fig. 3 for explanations).
979
980 Fig 8. Precision and accuracy of $\tau\dot{V}O_{2P}$ estimation for WR-WR bouts. *A*-*C*: effects of the
981 number of averaged bouts, where data processing is by interpolation and ensemble averaging
982 and removal of the first 15 s of data is used to isolate ϕ2 (see Fig. 2 for explanations). *D*-*E*:
983 effects of averaging method on $\tau\dot{V}O_{2P}$ estimation and the associated confidence interval,
984 using four bouts and where removal of the first 15 s of data is used to isolate ϕ2 (see Fig. 3
985 for explanations). *F*: effects of empirical ϕ2 isolation methods, using four interpolated and
986 ensemble averaged bouts (see Fig. 4 for explanations).
987
988 Fig. 9. *A*: Simulated $\dot{V}O_{2P}$ response to a 100 W UP-WR step using default model parameters
989 [see (8) for details]. The ϕ1-2 transition occurs at 19.4 s and the true ϕ2 $\tau\dot{V}O_{2P}$ is 16.3 s. *B*:
990 Effects on $\tau\dot{V}O_{2P}$ estimation of fitting the mono-exponential Equation (2) starting from

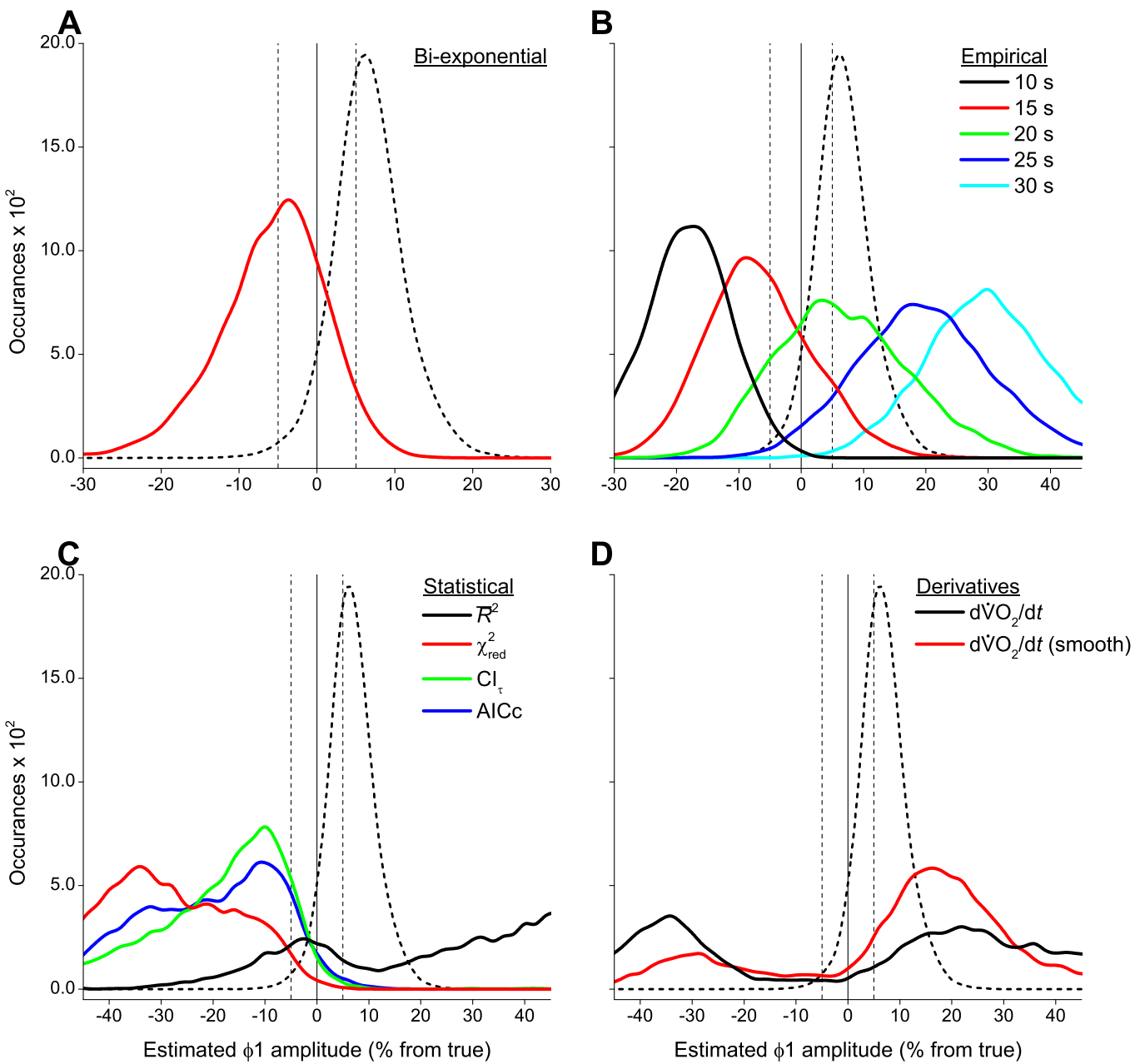
different points throughout the clean simulated $\dot{V}O_{2P}$ response. The vertical dashed line shows the time of the ϕ1-2 transition; the horizontal dashed line shows the true ϕ2 $\tau\dot{V}O_{2P}$.

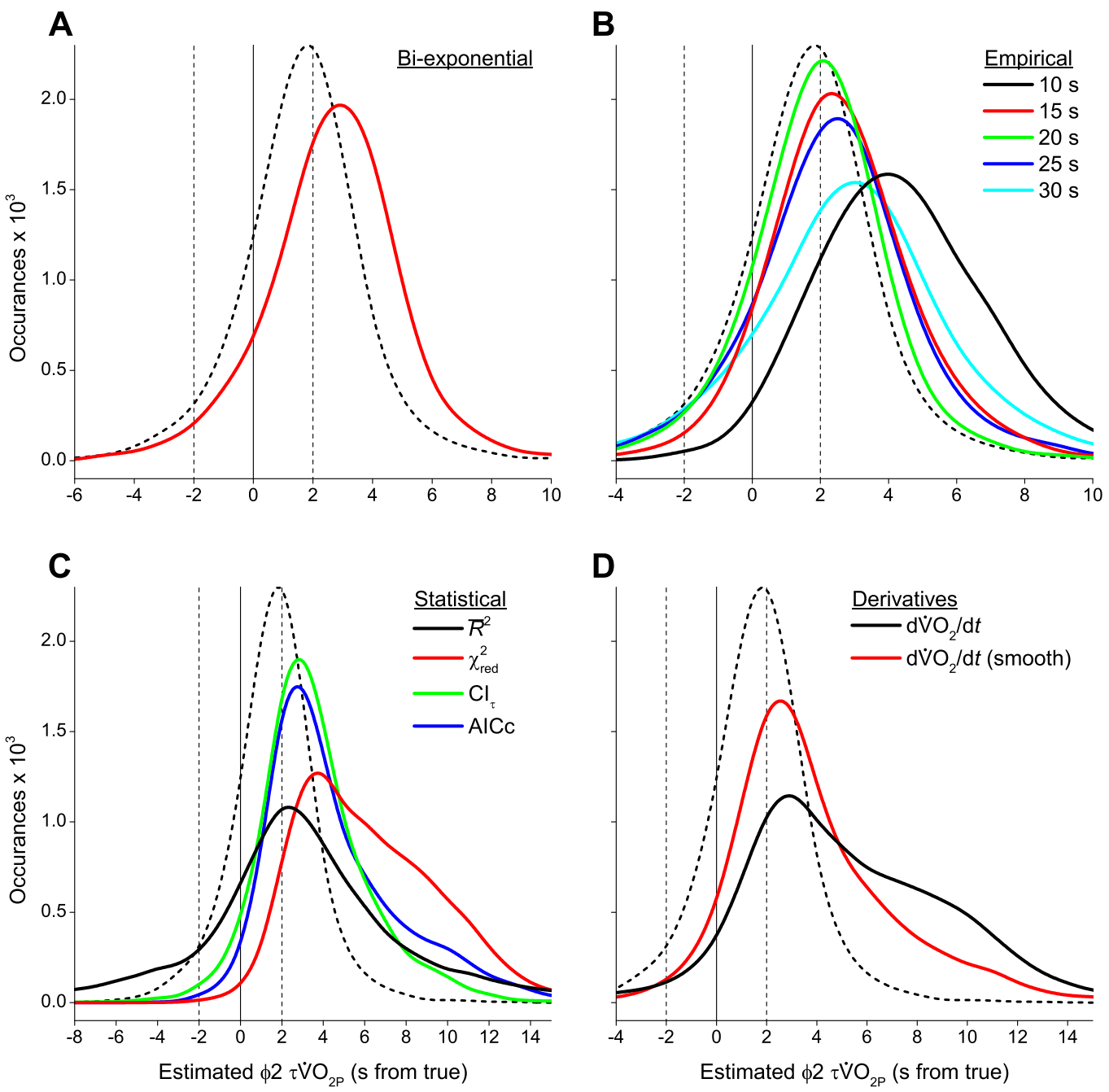

Clean model data
A
known ϕ2 τV̇O2P
known ϕ1 amplitude
known ϕ1-2 transition
V̇O2P (l.min-1)
Time (s)
Noisy "experimental" data from individual bouts
B
Processed data:
C
Interpolated
D
Binned
E
Stacked

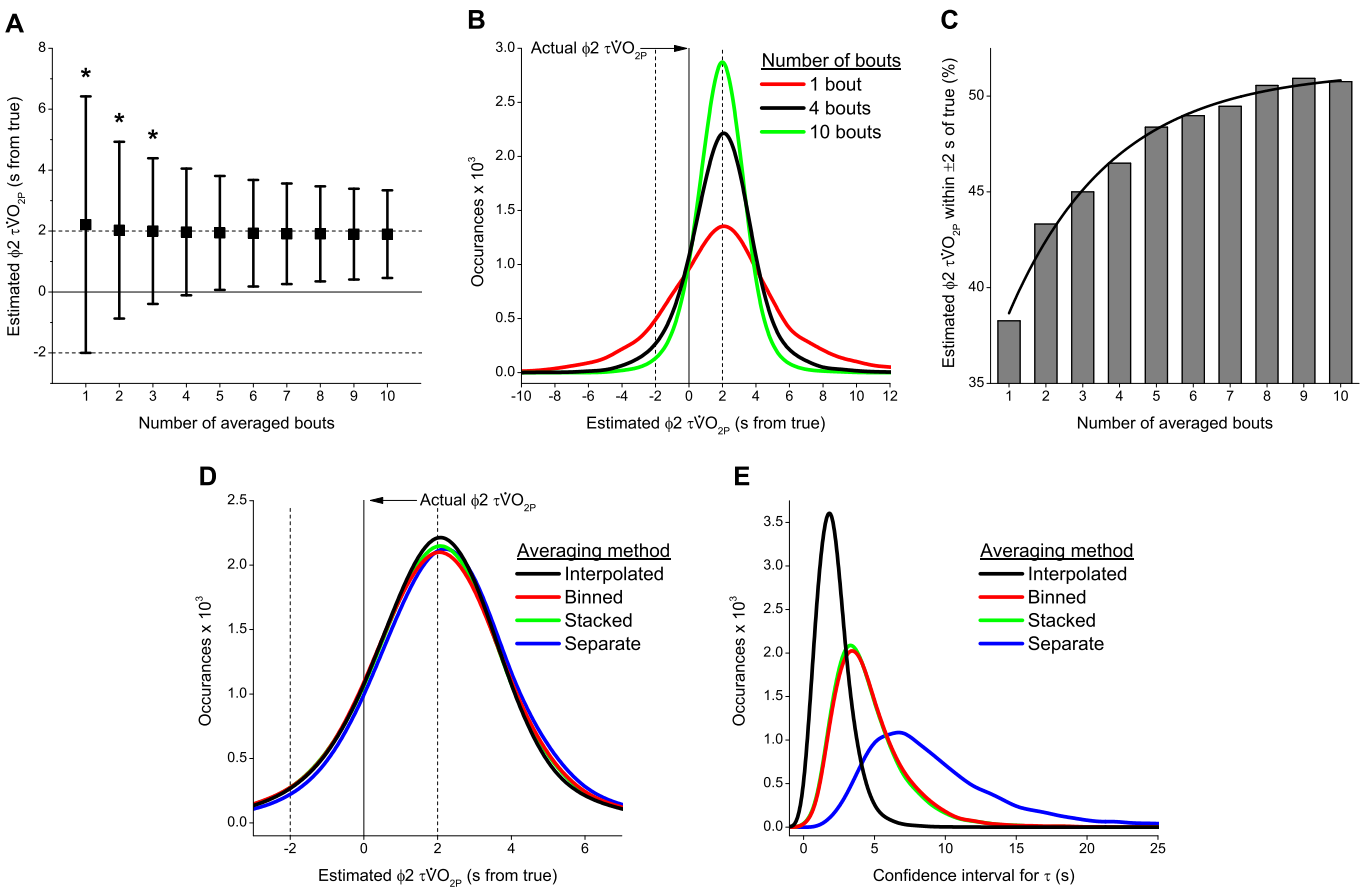

A
Estimated ϕ2 τVO2P (s from true)
Number of averaged bouts
B
Actual ϕ2 τVO2P
1 bout
4 bouts
10 bouts
Occurances x 10^3
Estimated ϕ2 τVO2P (s from true)
C
Estimated ϕ2 τVO2P within ±2 s of true (%)
Number of averaged bouts

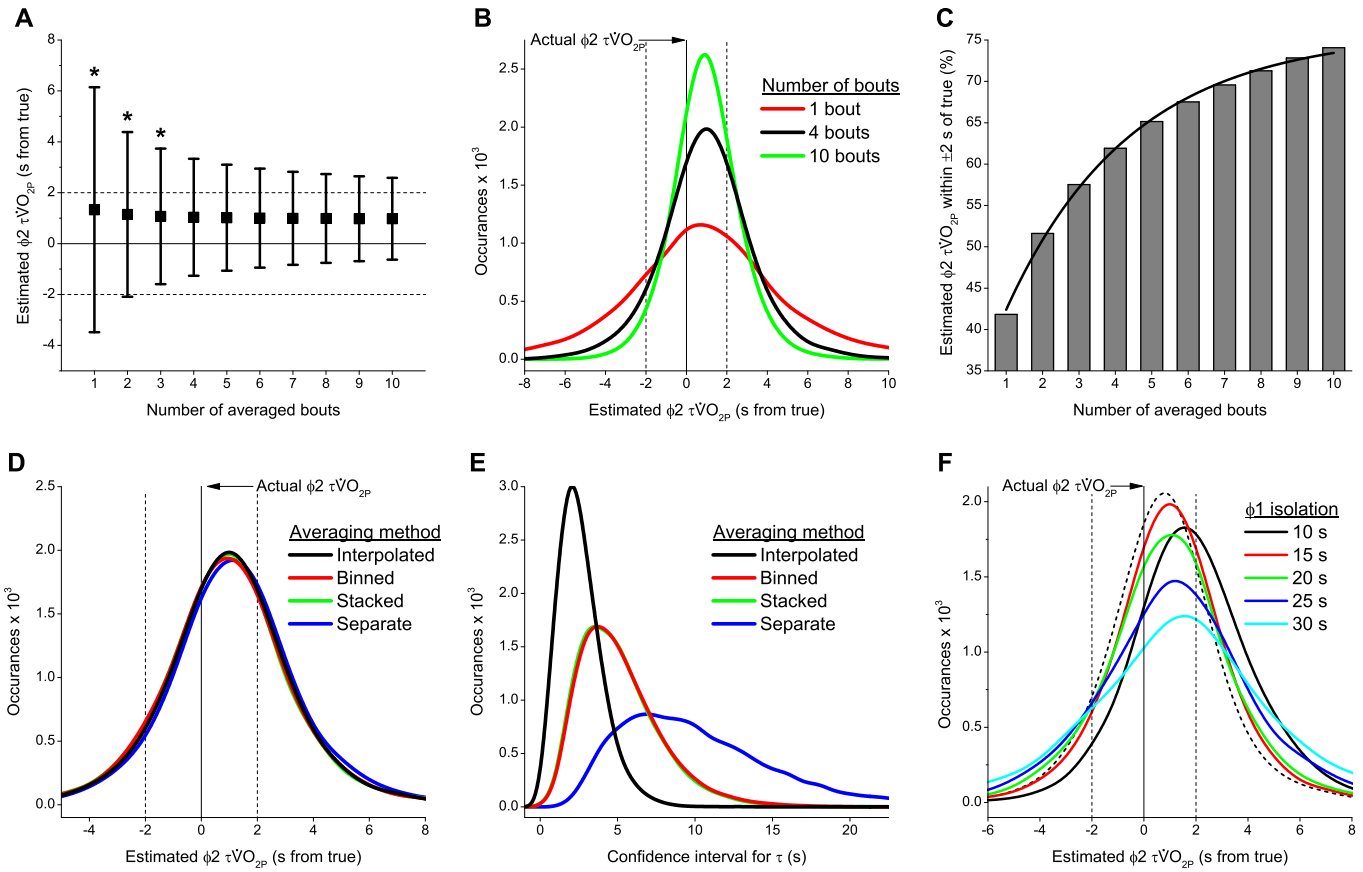

A
Actual ϕ2 τV̇O2P
Occurances x 10^3
2.5
2.0
1.5
1.0
0.5
0.0
-4
-2
0
2
4
6
8
Estimated ϕ2 τV̇O2P (s from true)
B
Interpolated
Binned
Stacked
Separate
Occurances x 10^3
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Confidence interval for τ (s)

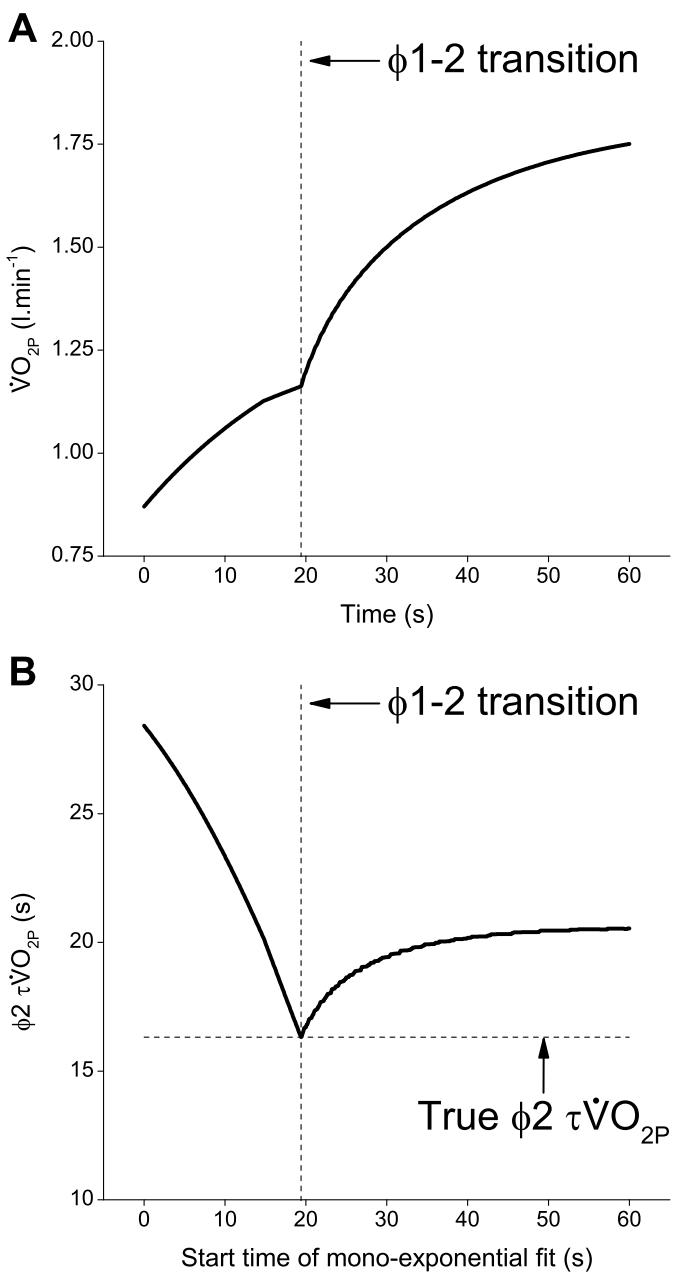

A
Bi-exponential
Occurances x 10^3
4.0
3.0
2.0
1.0
0.0
-6
-4
-2
0
2
4
6
B
Empirical
10 s
15 s
20 s
25 s
30 s
-6
-4
-2
0
2
4
6
8
C
Statistical
R^2
χ^2_red
CI_τ
AICc
Occurances x 10^3
4.0
3.0
2.0
1.0
0.0
-8
-6
-4
-2
0
2
4
6
8
10
12
Identified first ϕ2 breath (breaths from true)
D
Derivatives
dV̇O2/dt
dV̇O2/dt (smooth)
-8
-6
-4
-2
0
2
4
6
8
10
12
Identified first ϕ2 breath (breaths from true)


A
Bi-exponential
B
Empirical
10 s
15 s
20 s
25 s
30 s
C
Statistical
R^2
χ^2_{red}
CI_τ
AICc
D
Derivatives
$d\dot{V}O_2/dt$
$d\dot{V}O_2/dt$ (smooth)
Occurances x 10^2
Estimated ϕ1 amplitude (% from true)


A
Bi-exponential
Occurances x 10^3
2.0
1.5
1.0
0.5
0.0
-6
-4
-2
0
2
4
6
8
10
B
Empirical
10 s
15 s
20 s
25 s
30 s
-4
-2
0
2
4
6
8
10
C
Statistical
R^2
χ^2_red
CI_τ
AICc
Occurances x 10^3
2.0
1.5
1.0
0.5
0.0
-8
-6
-4
-2
0
2
4
6
8
10
12
14
Estimated ϕ2 τV̇O2P (s from true)
D
Derivatives
dV̇O2/dt
dV̇O2/dt (smooth)
-4
-2
0
2
4
6
8
10
12
14
Estimated ϕ2 τV̇O2P (s from true)


A
Estimated ϕ2 τV̇O2P (s from true)
Number of averaged bouts
B
Actual ϕ2 τV̇O2P
Number of bouts
1 bout
4 bouts
10 bouts
Occurances x 10^3
Estimated ϕ2 τV̇O2P (s from true)
C
Estimated ϕ2 τV̇O2P within ±2 s of true (%)
Number of averaged bouts
D
Actual ϕ2 τV̇O2P
Averaging method
Interpolated
Binned
Stacked
Separate
Occurances x 10^3
Estimated ϕ2 τV̇O2P (s from true)
E
Averaging method
Interpolated
Binned
Stacked
Separate
Occurances x 10^3
Confidence interval for τ (s)


A
Estimated ϕ2 τ̂VO2P (s from true)
Number of averaged bouts
B
Actual ϕ2 τ̂VO2P
Number of bouts
1 bout
4 bouts
10 bouts
Occurances x 10^3
Estimated ϕ2 τ̂VO2P (s from true)
C
Estimated ϕ2 τ̂VO2P within ±2 s of true (%)
Number of averaged bouts
D
Actual ϕ2 τ̂VO2P
Averaging method
Interpolated
Binned
Stacked
Separate
Occurances x 10^3
Estimated ϕ2 τ̂VO2P (s from true)
E
Averaging method
Interpolated
Binned
Stacked
Separate
Occurances x 10^3
Confidence interval for τ (s)
F
Actual ϕ2 τ̂VO2P
ϕ1 isolation
10 s
15 s
20 s
25 s
30 s
Occurances x 10^3
Estimated ϕ2 τ̂VO2P (s from true)


A
2.00
1.75
1.50
1.25
1.00
0.75
$\dot{V}O_{2P}$ (l.min^{-1})
ϕ1-2 transition
0
10
20
30
40
50
60
Time (s)
B
30
25
20
15
10
ϕ2 τ$\dot{V}O_{2P}$ (s)
ϕ1-2 transition
True ϕ2 τ$\dot{V}O_{2P}$
0
10
20
30
40
50
60
Start time of mono-exponential fit (s)