JEPonline
Journal of Exercise Physiology**online**

Official Journal of The American Society of Exercise Physiologists (ASEP)

ISSN 1097-9751
An International Electronic Journal
Volume 6 Number 2 May 2003

New Concepts: Equipment Testing and Validation

METHODS USED TO PROCESS DATA FROM INDIRECT CALORIMETRY AND THEIR APPLICATION TO VO_2MAX

ROBERT A. ROBERGS[1] AND ANGUS F. BURNETT[2]

[1] Exercise Physiology Laboratories, University of New Mexico, Albuquerque, NM, USA.; [2] School of Biomedical and Sports Science, Edith Cowan University, Perth, Western Australia.

ABSTRACT

METHODS USED TO PROCESS DATA FROM INDIRECT CALORIMETRY AND THEIR APPLICATION TO VO_2MAX. **Robert A. Robergs, Angus F. Burnett**. **JEP**online. 2003;6(2):44-57. Our purpose was to provide objective evidence in support of recommendations for how to process data in gas exchange indirect calorimetry (GEIC). A computer generated data set, devoid of biological variability and measurement error, was used to assess the error in data processing methods of a) 0.25, 0.5 and 1.0 min time averages, b) a 7 breath average, and c) a 11 breath smoothing (moving average) function. We aligned averaged data to the end and center of the time interval, and also used a breath-by-breath data set from an incremental exercise test to volitional fatigue to show the results of data processing using real data. Aligning time averaged data to the end of the interval under-estimated oxygen consumption (VO_2) (mean±SD, mL/min) by 29.99±5.21 (0.25 min end), 65.09±6.40 (0.5 min end), and 134.20±3.48 (1 min end). Aligning time averaged data to the center of the interval resulted in slight over-estimation of VO_2, with mean±SD (mL/min) errors of -12.61±5.10 (1.0 min), -5.94±3.43 9 (0.5 min), and -3.41±4.01 (0.25 min). Results from the 7 breath averaged data revealed mono-exponential decreases in error with increases in VO_2, ranging from (rest to peak VO_2, mL/min) 67.93 to 11.8 and 0.47 to 0.01, for end and center aligned data, respectively. Smoothing (11 breath) also resulted in mono-exponential decreases in error with increases in VO_2, with a range of residuals from (rest to peak VO_2) 1.08 to 0.03 mL/min. Similar procedures were applied to examples of breath-by-breath data obtained from research subjects, and therefore data containing biological variability. In addition, data processing involving interpolation and filtering was applied to the human subjects acquired data. Averaging VO_2 data from GEIC across 0.5 and 1.0 min time intervals adds to the total error of measurement in GEIC. Where possible, data averaging should occur by breath averaging (~7 breaths) aligned to the center of the interval. Data processing of breath-by-breath data using interpolation and filtering has potential, and required further research to assess validity.

Key Words: Oxygen Consumption, Time Average, Interpolation, Filter, Exercise Physiology

INTRODUCTION

The application of gas exchange indirect calorimetry (GEIC) to the measurement of oxygen consumption (VO_2) and related variables during exercise has application to numerous academic fields and professions. Furthermore, the measure of VO_2 during incremental exercise to volitional exhaustion is a common measure in exercise physiology and clinical fields such as cardiology and pulmonology. These tests are usually conducted to,

a) quantify the maximal rate of VO_2 (VO_2max),
b) indirectly assess the onset of exercise-induced acidosis (ventilation threshold, VT),
c) establish steady state values of VO_2,
d) quantify the rate of change in VO_2 in response to changes in exercise intensity (VO_2 kinetics),
e) assist in the quantification of anaerobic capacity, or
f) supplement additional procedures in the detection of cardiovascular and pulmonary diseases (1-7).

Despite the obvious importance and routine applications of GEIC, there is limited published review of the data processing procedures used for data acquired from GEIC for any end measure. For example, there is a general acceptance that time averaged data can be presented aligned to the end of the averaging interval (8), or that breath-by-breath data should be "smoothed" using a breath average (typically 7 to 11 breaths) (6). Similarly, and as recently reviewed by Howley et al. (8), there exists confusion in the literature over the criteria needed to verify a true VO_2max, and how data should be time averaged to best represent "true" values for VO_2 at a given exercise intensity.

The purpose of this manuscript is to present examples of how to process data from the different methods of GEIC, and quantify the errors inherent in each method of data processing. The examples of GEIC data processing are presented for each specific data collection method of GEIC, and consist of 1) time averaging, 2) breath averaging, 3) breath smoothing (running average), and 4) signal filtering and curve fitting. Such approaches can be applied to both semi-automated and fully automated mixing chamber systems, as well as fully automated breath-by-breath systems.

We have chosen to use the GEIC measurement of VO_2max as the primary example of errors in processing. This decision was based on VO_2max being the most widely performed measurement from GEIC, and being the single variable that has the broadest physiological application to pure, applied and clinical applications of exercise physiology (1-9). However, many items discussed also have application to submaximal values for VO_2 during exercise, such as measurements of the heart rate, ventilatory and lactate thresholds, exercise prescription, and quantification of the accumulated oxygen deficit.

Finally, our findings and recommendations are presented in this "New Concepts" classification based on the need to inform exercise physiologists and related scientist and practitioners of the caution needed when making decisions in how to process data from GEIC. It is also anticipated that based on the content of this manuscript, research can be undertaken to assess the impact of data averaging techniques on data from GEIC using either time averaging or breath-by-breath technologies. Considerable communication among scientists and educators is needed for providing concrete recommendations on how to process data from GEIC. The format of **JEP**online provides a useful venue to further stimulate this need.

The content that follows explains the systems used in GEIC, followed by specific sections on each of the data processing options. Where pertinent, data from prior human subjects' research, previously approved by university Human Subjects' Research Committees, are presented to exemplify differences in data processing. In addition, computer generated data is presented, devoid of measurement error and subject physiological variability, to more clearly compare the magnitude of errors introduced by data processing procedures.

SYSTEMS USED IN GAS EXCHANGE INDIRECT CALORIMETRY

There are numerous systems available for performing GEIC. However, all systems have in common the need to measure ventilation (expired (V_E) or inspired (V_I)), and fractions of oxygen and carbon dioxide in expired air (F_ECO_2 and F_EO_2, respectively). From these measurements, computations are performed for oxygen consumption (VO_2), carbon dioxide production (VCO_2), and the respiratory exchange ratio (RER). These three calculated variables can be used to quantify energy expenditure via body-adjusted constants for the complete combustion of carbohydrate, fat, or carbohydrate-fat mixes.

There are three methods of GEIC; 1) manual Douglas bag, 2) semi-automated and fully automated mixing chamber, and 3) fully automated breath-by-breath. Differences exist within and between methods for the equipment used to quantify ventilation, how values for expired fractions of oxygen and carbon dioxide are derived, and the frequency at which data is acquired and processed for computations for VO_2, VCO_2, and RER. However, the purpose of this manuscript is to focus on data processing procedures and not the strengths and weaknesses of specific methods of GEIC.

PROCESSING DATA FROM EXPIRED GAS ANALYSIS INDIRECT CALORIMETRY

In any data collected repeatedly over time, errors exist that increase the variability of the data and have the potential to alter the true physiological trend of the data. Where the error is random, the errors tend to be distributed above and below “true” values, and thus, researchers often average the data over specific time intervals to minimize the impact of this error on subsequent data processing and interpretation. Such random errors are caused by imprecision in measurements of ventilation and expired gas fractions. However, with methods such as time interval averaging, components of “true” physiological responses can be removed along with random errors. Consequently, there is a risk that data processing can decrease the validity of the measures and calculations of GEIC, which in turn can invalidate the physiological interpretations of the data.

Data Averaging Methods

Data from GEIC can be processed in different ways to result in an average presentation. Typically, data is time averaged, but other methods of averaging are multiple breath and data smoothing (moving average) techniques. The attributes and limitations of these methods of data averaging are discussed below.

Time Averaging

It is reasonable to conclude that time averaging methods of GEIC remain the most commonly used in research of exercise physiology. Furthermore, there seems to be an acceptance that longer time interval averaging is better than smaller time interval averaging, with guidelines still present for preference of 0.5 to 1 min averaging of data (4,8-11). The purpose of time averaging data from GEIC is to decrease the physiological variability in the data (ventilation, muscle blood flow, and metabolism) in addition to the experimental error. The latter component is obviously the most important argument for the time average. However, recent criticisms of time averaging exist (6,14), and it is worthwhile to evaluate the error introduced by both short (<30 s) and long (>30 s) duration time averaging.

Researchers who use time-averaged methods of GEIC typically coincide the averaged data with the top end of the time interval. For example, a 1 min average of data between min 5 and 6 is denoted as min 6 on the time axis. Alternatively, a superficial evaluation of this presentation may lead researchers to assign the averaged data point to the centre of the time interval. The obvious questions are; Does the time average modify the temporal alignment of the unprocessed data? Are potential errors removed when data are aligned to the centre of the time interval? If error is retained, even after altered data alignment, how should data be processed? Do errors in data processing have the potential to decrease the accuracy of VO_2max, other end measures of GEIC, and subsequent data interpretations?

To simplify the process of evaluating the effect of a time average, the data of Figure 1a show computer generated data depicting an increase in VO_2 at the rate of 250 mL/min, starting at 250 mL/min, with no physiological error variability, or end VO_2 plateau. Data are presented with a curvilinear increase in ventilation frequency from 0.2 to 1.22 Hz (Figure 1b) to better represent incremental exercise. These features resulted in 510 data points over 15 min, with a peak VO_2 of 4,501.5 mL/min. All data were derived using Excel™ software.

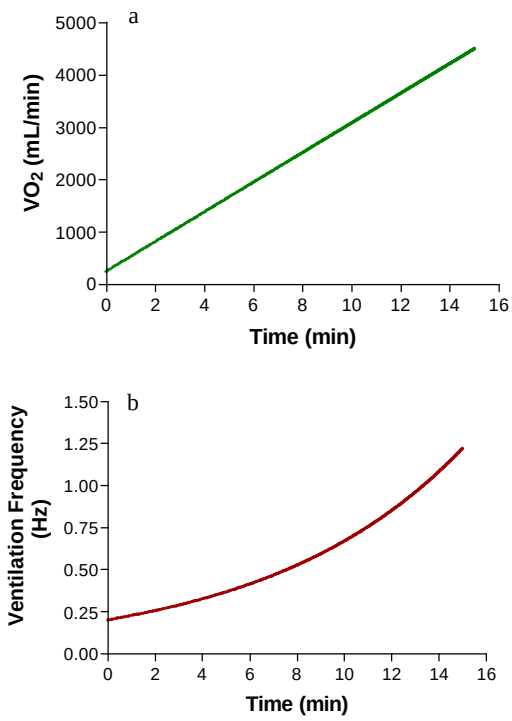

Figure 1. Computer generated data for a) oxygen consumption (VO_2), and b) the change in the frequency of ventilation.

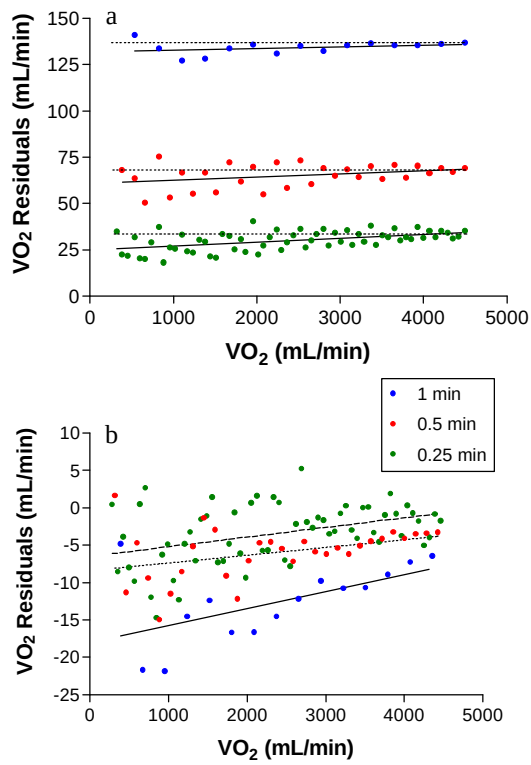

Figure 2. a) Residual errors resulting from aligning time averaged data (0.25, 0.5 and 1.0 min) to a) the top end of the averaging interval, and b) the center of the averaging interval. Data are "true" – averaged values.

The computer generated data were processed into time block averages of 0.25,0.5 and 1.0 min. Figure 2a illustrates the residuals ("true"–averaged) and systematic error of data processing for the 0.25, 0.5 and 1.0 min time averaging, when the average value is aligned at the end of the time averaging interval. For clarification, an "ideal" response would show random errors slightly above and below the "true" values across the physiological range of the variable. The data of Figure 2a are far from ideal. When the average value is aligned at the end of the time averaging interval there is an underestimation of the "true" value, with mean±SD errors across the range of VO_2 for the 0.25, 0.5 and 1.0 min intervals of 30.0±5.2, 65.1±6.4, and 134.2±3.5 mL/min, respectively. For a 75 kg person, such errors compute to relative VO_2 values of 0.4, 0.9, and 1.8 mL/kg/min respectively.

The error introduced by time averaging is due to the incorporation of data in the time average that occur prior to the specific time chosen to align the data. The larger the time average interval, the greater the under-representation of the "true" VO_2. For the 0.25 min data averages, the regression slope is significantly different from zero (2.1 mL/min/L), with trends for the same response for the 0.5 (1.7 mL/min/L) and 1 min (0.8 mL/min/L) data averages. The small systematic trend in the data for this short time average is explained by the increasing number of breaths/interval as exercise intensity increases. For the lower exercise intensities, having one additional breath in the second half of the interval has a more dramatic effect on raising the average VO_2,

and decreasing the error. Conversely, the greater frequency of breaths as exercise intensity and ventilation rate increase makes for a more uniform spread of data and produces an average that better reflects the error (biological and experimental) of the procedure. Finally, the variability in each data set around their respective regression line decreased as VO_2, and the frequency of breathing, increased. The variability in the data results from the mismatching that occurs between data obtained every breath, and subsequent time averaging. When the breathing frequency is low, aligning data points to time intervals yields fluctuating numbers of data points in each average, which produces a staggered response in the data with low values followed by higher values. As breathing frequency increases, the impact of one or two additional breaths on differences between successive time averages decreases in magnitude, producing less variable data.

Figure 2b illustrates the error when the average value is aligned with the center of the time interval. Although these errors are small, they remain systematic. The mean errors across the range of VO_2 for the 0.25, 0.5 and 1.0 min intervals are -3.4±4.0, -5.9±3.4, and -12.6±5.1 mL/min, respectively. This overestimation results from the unequal number of data points in the first and second half of each time interval. Due to the increasing frequency of breathing, there are an increasing number of data points that occur in the second half of the interval, and this inflates the average to the higher end of the VO_2 response. The result is an averaged VO_2 that is larger than the time-centered value of the "true" data. Furthermore, the slopes for these regression lines are all significant at 1.29, 1.04, and 2.25 mL/min/L for the 0.25, 0.5 and 1.0 min data averages, respectively. This is to be expected, for as explained for the slope responses of Figure 2a, the increased frequency of breathing as exercise intensity increases reduces the impact of the unequal nature of the number of breaths in each half of the time interval.

Finally, although the data of Figure 2b appear to have increased variability introduced into the data by the centrally aligned time averaged, this is not true and occurs due to the smaller mean errors and the different scale of the y-axis compared to Figure 2a. In reality, the error of the data as measured by the standard error of the estimate ($S_{y.x}$) ranges between 3.2 and 6.1 mL/min across both data sets. The 0.5 min data set had the largest change in $S_{y.x}$ between the end vs. centrally aligned data of 6.13 and 3.24 mL/min, respectively. Compared to the 0.25 min data averages, the larger error of the 0.5 min end aligned data set results from the a greater influence of the loss in data points than decrease in range of variability.

In summary, for a change in the dependent variable that is best represented by a positive linear function and a progressively increasing frequency of data points, aligning any time average to the top end of the averaging interval will systematically under-represent the "true" value. Aligning any time average to the centre of the averaging interval will systematically over-represent the "true" value. For both methods, the magnitude of the error decreases as the time averaging interval decreases. However, as VO_2 and the frequency of breathing increases, alignment of the data at the end of the averaging interval increases error, while the error of the centrally aligned data is decreased to physiologically negligible values.

The Influence of Breath Averaging on VO_2

The introduction of breath-by-breath systems of GEIC also provided another means to average data; the breath average. However, to clarify the utility of this approach it must be stressed that a breath average can also be applied to mixing chamber methods of GEIC. However, the breath average has not been a feature of GEIC due to the limitations in software used in fully automated commercial and self-developed mixing chamber systems. This is unfortunate, for compared to the data of Figure 2, Figure 3a and b reveal less error in the breath average approach to processing data from GEIC.

Figure 3a and b show the same data used in Figure 2 averaged every 7 breaths. The breath average in this example is also a block average, yet there is always the same number of breaths in each average throughout the data range. When the breath average is aligned to the end of the interval, the error is appreciably less than the block time average (Figure 2a), ranging from 67.93 to 11.8 mL/min. Unlike the error from the time block average, the error from the 7 breath average decreases as VO_2 and the frequency of breathing increases, and the

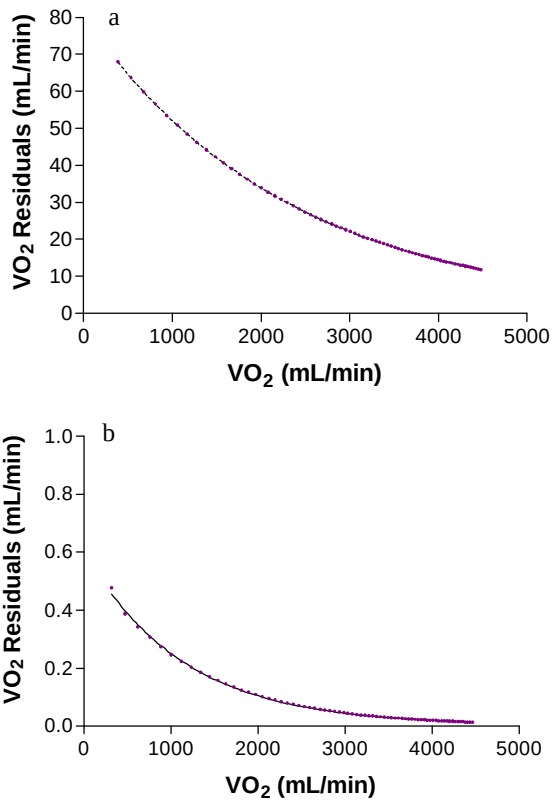

Figure 3. a) Residual errors resulting from aligning 7 breath averaged data to a) the top end of the averaging interval, and b) the center (breath 4) of the averaging interval. Data are "true" – averaged values.

data perfectly fit a mono-exponential decay. The differences between Figure 2a and 3a indicate the contribution to error caused by the unequal breaths in successive intervals in the time block average. The mono-exponential decay of the 7 breath errors is the reciprocal of the increase in breathing frequency programmed into the original data set. This reduction in error is explained by the fact that as the breathing frequency increases, the differences in the VO_2 between each breath decreases, and therefore the error in the mean value to the aligned value must decrease.

When the 7-breath average is aligned to the centre of the interval (breath 4), the absolute errors are reduced to values that for the error of the method of GEIC are meaningless (0.47 to 0.01 mL/min). Interestingly, the errors are also reduced as VO_2 and breathing frequency increase, and also fit a perfect monoexponental decay. The explanation for this pattern is the same as for the 7 breath data aligned to the end of the interval (Figure 3a).

The Influence of Data Smoothing (Moving Average) on VO_2

As with the option for a multiple breath average, the relatively large number of data points obtained from breath-by-breath systems of GEIC have resulted in the use of additional methods of data processing. One such approach is the "smoothing" of data by programming an averaging function that in effect averages a data point at a given time point with a number of points before and after this value. This is a common procedure with data obtained breath-by-breath due to the high frequency of data points, and relatively high variability in data over time. A commonly used number of data points for the smoothing average is 11 (5 breaths before and after) (6).

The data set used for Figures 2 and 3 were also used to show the effects of a 11 breath smoothing function compared to the "true" data, and the resulting residuals are illustrated in Figure 4. The error from smoothing is also systematic, causing a decrease in the VO_2. As with the 7-breath average, the decreases in VO_2 exhibited a perfect mono-exponential decay with increases in VO_2 and breathing frequency. Interestingly, the error introduced by an 11 breath smoothing function is greater than the error from the 7-breath average aligned to the center of the interval.

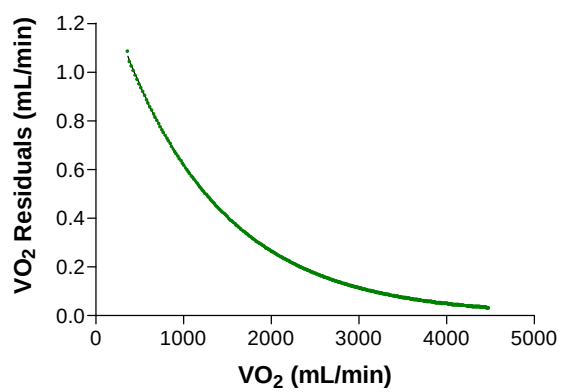

Figure 4. Residual errors resulting from an 11 breath smoothing function. Data are "true" – averaged values.

The utility of the smoothing function is better visualized with real data. The data of Figure 5a represent actual breath-by-breath data (Medical Graphics CPXD) from a cycle ergometer test to VO_2max in an untrained male, with moderate cardiorespiratory and muscular endurance. The

subject and data were from a research study utilizing informed consent procedures approved by the university human research committee. Figure 5b reveals that the frequency of ventilation (data acquisition) increased from 0.2 to 1.3 Hz from rest to VO_2max, which will be discussed later in the section on interpolation and filtering. Note that this data differs to that of Figures 1-4 due to biological variability and measurement error, yet the increasing frequency of ventilation closely resembles that programmed into the earlier theoretical data set (Figure 1b).

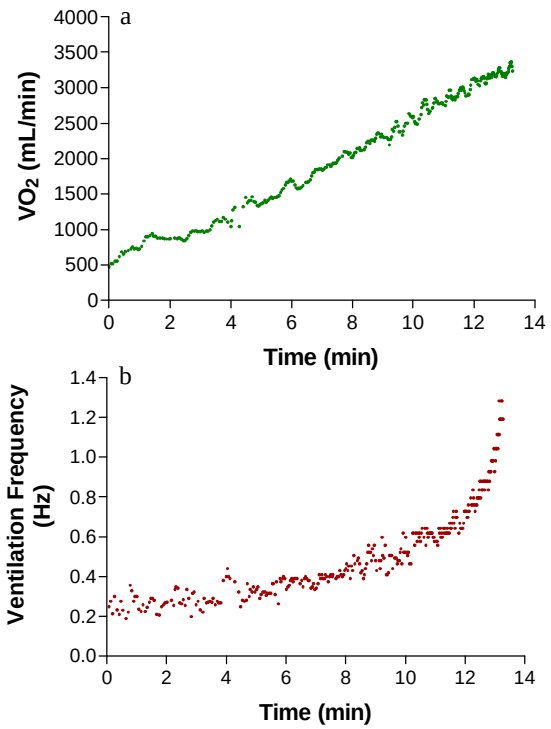

Figure 5. a) The increase in VO_2 to VO_2max in a subject with moderate cardiorespiratory and muscular endurance. b) The increase in the rate of ventilation during the test to VO_2max, causing an increased frequency (decreasing time interval between data points) of breath-by-breath data collection.

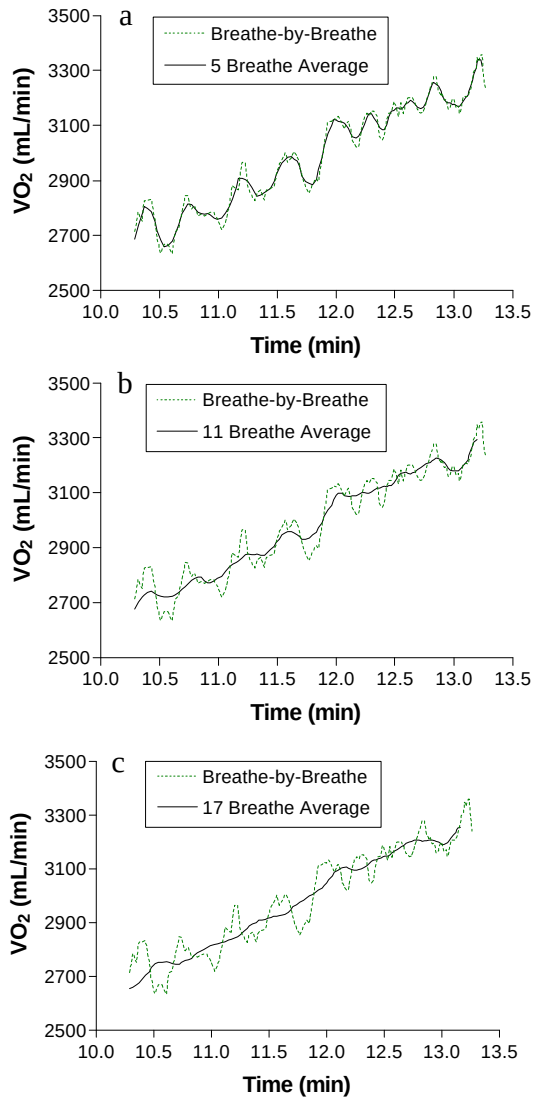

Figure 6. VO_2 data for the last 3 min of the test to VO_2max. Data are presented that are the raw breath-by-breath results compared to smoothing functions centered about the data point for a) 5 breaths, b) 11 breaths, and c) 17 breaths.

Figure 6a-c presents the last 3 min of the data expressed as raw data, as well as smoothed at 5, 11 and 17 data point averages. For the purpose of the detection of VO_2max, increasing the number of data points in the averaging function removes variability, and lowers the peak VO_2 attained near the end of the test. In addition, smoothing decreases the number of data points at the beginning and end of the data set by; (n-1)/2 (where 'n' represents the number of data points in the average), and introduces a systematic alteration of the data as previously shown in Figure 4. A crucial question is whether this method of data smoothing removes real biological variability in the data, thereby altering the true physiological meaning and interpretation of the data. This question will be answered in the next section.

Other Methods Of Data Processing

Although the time average, breath average, and smoothing methods of data processing are prevalent in the research using GEIC, another method of data processing is also available yet not routinely used in GEIC– interpolation and filtering. Comparing data obtained from the different processing functions will reveal the benefits and limitations of the methods.

Interpolation and Filtering

Myers et al. (15) discussed the inherent error and causes of variability within the pulmonary gas exchange variables when performing GEIC using the Medical Graphics system. As with any system used in GEIC, error results from imprecision in the gas analyzers, ventilation modules, and to a smaller extent, the measurement of environmental factors used in the calibration of ventilation. We will explore the magnitude of these errors and their affect on VO_2 later. Nevertheless, as the frequency of data sampling and processing increases, such as progressing from larger to smaller time averaging, or increasing the breathing frequency in the breath-by-breath method of GEIC, there is more potential to show true biological variability in the data if such variability exists. However, and as well explained by Howley et al. (8), the errors inherent in the measurements of V_E, F_EO_2 and F_ECO_2 can produce large errors when converting small time interval sampling to rates expressed relative to 1 min.

When concerned with breath-by-breath data, and as explained previously for the data used in Figures 1-4, as well as Figure 5, there is an increased frequency of ventilation and data point acquisition as exercise intensity increases. This results in data that is non-equispaced, which for technical reasons, complicates the ability to perform more advanced data processing. For example, application of a Butterworth digital filter requires that data be equispaced via a method such as cubic spline interpolation programmed to obtain values at 1 Hz (ie. one breath[data point]/s). The frequency of 1 Hz is used as this represents a typical peak rate of ventilation and data acquisition for the test (ie: at VO_2max). Cubic spline interpolation involves fitting a number of cubic polynomials to the data in points called "knots". Based on these equations, data can be equispaced with very high accuracy. Nevertheless, data are redistributed over a shorter x-axis range, with the magnitude of the altered range being dependent on the extent to which the assigned constant frequency differs to the mean frequency at which the data is acquired.

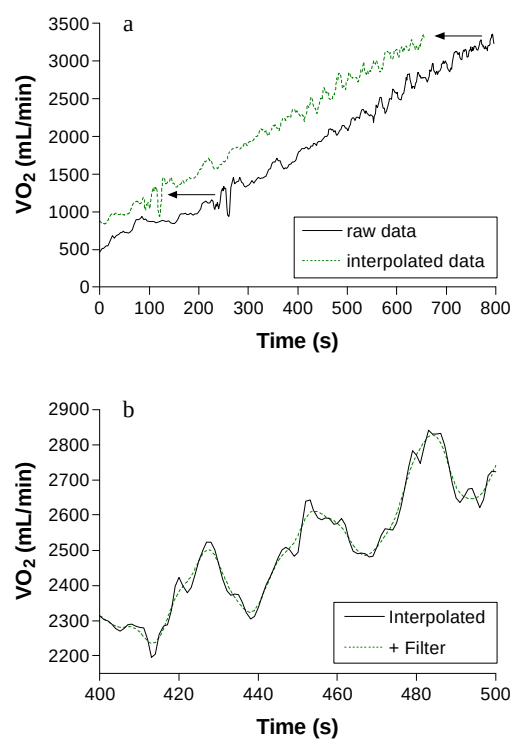

Figure 7. a) The change in raw breath-by-breath data after interpolation at 1 Hz. Although the data set is shorter in time, the interpolation produces an increased number of data points. b) Application of a low pass filter to the interpolated data. Data are of a select region of the data used in Figures 5 and 6 so that the small changes resulting from the filter can be seen.

Although altering the time of data acquisition may appear to change the response of the data, this change is small, irrelevant for purposes of detecting VO_2max, and the benefits of the application of a filter outweigh this change. For example, the data of Figure 5b reveals that the frequency of data acquisition for this test and subject increased from 0.2 to 1.2 Hz from rest to VO_2max. Unlike the data of figures 1-4, this data possesses both measurement error and biological variability. We processed the data with a cubic spline interpolation programmed to obtain values at 1 Hz (Figure 7a). To assess the change in data from using a filter, we used customized software developed in LabView (National Instruments, Austin, Texas) to apply a Butterworth fourth

order low pass digital filter. The cut off frequency of the digital filter was chosen by the method of Winter (16), which examines the residuals over a variety of smoothing frequencies. The optimal frequency is that which maximizes the decrease in error whilst minimizing the loss of "real" signal. In this breath-by-breath data set the optimal smoothing frequency was 0.1013 Hz, being equivalent to a 3 data point smoothing function.

Consequently, for our subject data, using more than three data points in a moving average may remove biological signal. This result will vary between individuals based on the between-subject variability in the ventilation parameters. Nevertheless, for smoothing functions involving greater than 5 breaths (Figure 6a-c), such processing artificially alters the true VO_2 response to reflect a more constant (linear or mono-exponential) function.

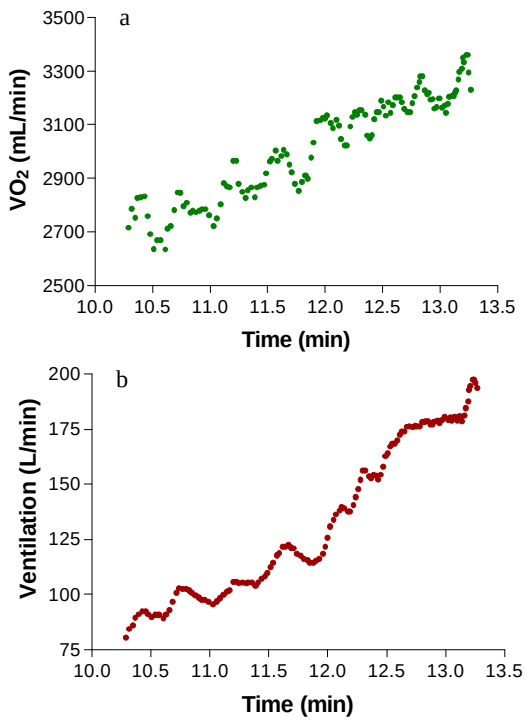

Figure 8. The temporal alignment between the variability in a) VO_2 and b) ventilation for data of the last 3 min of a cycle ergometer test to VO_2max.

The low pass digital filter produced a small decrease in data variability. This small change can be interpreted as evidence for only a small contribution to data variability by non-systematic measurement error. Therefore, the remaining variability, characterized by repeated oscillations in VO_2 that is typical for this subject, and others when using breath-by-breath acquisition, is likely to be true biological variability. The biological cause of these sinusoidal waveforms is almost totally explained by the variability in ventilation (15). To better illustrate the coupling between ventilation and VO_2 in GEIC, the VO_2 and ventilation during the last 3 min of the test are presented in Figure 8. Although the frequency of the sinusoidal changes in VO_2 appear greater than for ventilation, it is clear that the large increases and decreases in VO_2 over time correspond to similar changes in ventilation.

These findings are in direct opposition to the recommendation by Howley et al. (8), where a time average of at least 30 s was stated to be necessary for the detection of VO_2max to nullify any measurement error in data acquisition. We have clearly shown that this recommendation is extremely crude, lowers actual VO_2 values, and thereby underestimates VO_2max. Based on our rationale and presentation of data, we believe that there is no conceptual or technical justification for the 0.5 and 1 min time average of VO_2 data when measuring VO_2max. The obvious questions are; What is the optimal method of data averaging? Which data processing procedure reveals the true VO_2max value? Do these recommendations change depending on the GEIC system used? Prior to answering these questions, we should first explore the potential measurement errors involved in GEIC and mathematically define what is VO_2max.

When Does Error Become Physiologically Meaningful?

The data of Figures 2 to 4 are useful in that they show the error resulting solely from the data processing method. Unfortunately, added error also exists when collecting data from GEIC. As with all experimental data collection, the error in the data can be categorized into components of a) experimental (or measurement), b) experimentor and c) processing. We have already quantified the error of data processing, and experimentor error is impossible to profile (Table 1). However, we can easily assess the potential measurement error of GEIC.

Table 1: Sources of error in GEIC.

Error Category	*Error Component*
Experimental	Environmental measures (barometric pressure, humidity, tem
	Accuracy of ventilation device
	Accuracy of gas analyzers
	Drift in ventilation accuracy
	Drift in gas analyzer accuracy
Experimentor	Recorded and software entered environmental measures
	Ventilation calibration
	Gas analyzer calibration
	For manual methods
	- measurement of ventilation
	- measurement of expired gas fractions
Data Processing	Method of error reduction
	- time averaging
	- breath averaging
	- smoothing
	- interpolation and filtering

As absolute errors depend on the exercise intensity, available data on the magnitude of error is best presented as a relative (%) value. Howley et al. (8) have estimated that the best possible scenario with GEIC applied to the measure of VO_2max is an error of ±2%. It is worthwhile to explore the contributions to this error to further justify this value, or perhaps provide a more valid estimate (Table 2).

Table 2: Examples of error in VO_2 measurements introduced by individual and summed components of potential error during GEIC.

Condition	*VO_2 (L/min)*	*FEO_2*	*$FECO_2$*	*V_E (L/min)*	*P_B (mmHg)*	*P_W (mmHg)*	*T_R (°C)*	*Accumulated* Error VO_2*
Steady state-1	1.35	0.1657	0.0383	30.1	760	9	22.0	
Likely error		0.0002	0.0001	0.5	2	1	0.5	
Likely error (%)		0.12	0.26	1.7	0.26	11.1	2.27	
Corrected VO_2 (L/min)		1.362	1.357	1.380	1.387	1.388	1.388	1.4052
Error (mL/min)		0.012	0.007	0.03	0.037	0.038	0.038	0.055
Error (mL/kg/min)		0.15	0.09	0.40	0.49	0.51	0.51	0.73
% error		0.85	0.51	2.20	2.64	2.77	2.74	3.93
Steady state-2	2.50	0.1689	0.0391	61.3	760	9	22.0	
Likely error		0.0002	0.0001	1.0	2	1	0.5	
Likely error (%)		0.12	0.25	1.6	0.26	11.1	2.27	
Corrected VO_2 (L/min)		2.51	2.50	2.54	2.56	2.56	2.56	2.59
Error (mL/min)		0.012	0.002	0.045	0.058	0.061	0.060	0.091
Error (mL/kg/min)		0.16	0.03	0.60	0.77	0.81	0.80	1.21
% error		0.47	0.10	1.77	2.25	2.38	2.34	3.50
VO_2max	4.53	0.1781	0.0361	151.2	760	9	22.0	
Likely error		0.0002	0.0001	3.0	2	1	0.5	
Likely error (%)		0.11	0.28	2.0	0.26	11.1	2.27	
Corrected VO_2 (L/min)		4.56	4.53	4.63	4.63	4.60	4.64	4.71
Error (mL/min)		0.026	0.003	0.100	0.10	0.07	0.11	0.18
Error (mL/kg/min)		0.35	0.04	1.30	1.39	1.00	1.45	2.40
% error		0.58	0.07	2.10	2.25	1.62	2.34	3.83

*Note that due to the interdependence of some error components, the combined application of errors computed to less than the sum of independent errors ; For an expired volume measurement system, and for environmental conditions of 1 ATM, a room temperature of 22 °C, a relative humidity of 45% and a room air water vapor pressure of 9 mmHg. Data for VO_2 expressed as mL/kg/min are calculated for a 75 kg person.

Table 2 presents data of errors in the three measures of GEIC for different values of VO_2 and the measured determinants of VO_2. Note that the data apply to an expired volume measurement system, and for specific environmental conditions. The data of Table 2 reveal some interesting facts. Realistic errors in expired ventilation (V_E) have the potential to contribute most to errors in VO_2 calculation. For example, a given percentage error in the measurement of V_E produces a similar relative error in the calculation of VO_2. Thus, at VO_2max, when V_E may equal close to 150 L/min, a 2% error equates to 3 L/min, which in turn can be partitioned to an error of approximately 50 mL/breath at a ventilatory frequency of 1 Hz. Thus, this error is realistic for most laboratories and methods of V_E measurement. For a 75 kg person, such an error corresponds to an approximate VO_2 of 1.3 mL/kg/min, which is a sizeable error of measurement for a single component of the VO_2 calculation. Similar explorations of error can be made for all additional components of the VO_2 calculation.

Note once again that absolute errors are proportional to the error in V_E, and therefore exercise intensity. Thus, when concerned with VO_2max, the higher the value the greater the absolute error of the measurement. This fact is also revealed in the data of Table 2.

Another important component of VO_2 calculation is the electronic oxygen analyzer. We based our calculations on a realistic fractional error in the oxygen reading of 0.0002, which can produce an error in VO_2 approximating 0.6%. However, for equal relative error to that of V_E, the oxygen analyzer is the most important component of the system. For example, if there was a 2% error in a FEO_2 of 0.1781, this would be a value of 0.1745, and have calculated VO_2max to be 5.23 L/min instead of 4.53 L/min, a 15.4% error equating to 9.3 mL/kg/min. Thankfully, electronic gas analyzers are exceptionally accurate, and absolute errors of less than 0.0002 (< 0.6%) should be expected during GEIC.

Of even greater meaning from the data of Table 2 are the accumulated error estimates. The accumulated error data were derived from unidirectional changes in each component that resulted in increases in VO_2. As indicated in Table 2, due to the interdependence of some error components, the combined application of errors computed to less than the sum of independent errors. With just a 2% error in V_E, additional V_E error from 0.5 °C and 1 mmHg water vapor pressure deviations required when calibrating ventilation devices, and 0.0002 and 0.0001 errors in F_EO_2 and F_ECO_2, respectively, an overestimation of VO_2max by 3.83% (2.4 mL/kg/min in our example) is likely. Clearly, unless the method used to measure V_E is highly accurate, accurate gas analyzers are used with zero drift during a test, and environmental factors are measured correctly, it is difficult to expect VO_2max measures that are within ±3% of true values. Interestingly, this is close to the ±2% expected error of VO_2max measurement proposed by Howley et al. (8). Although all components of the VO_2 calculation are important, the greatest likely error in measurement lies in ventilation, and all users of GEIC should go to great lengths to ensure that their method of V_E measurement, and measurement of environmental factors used in calibrating ventilation, are as accurate as possible.

Finally, given the sizeable potential measurement error in GEIC, added error from data processing should be avoided. It is wrong to conclude that because of the error of the measurement, errors up to the same relative value (±3%) can be tolerated. The goal of any scientists should be to reduce all error. As such, error from data processing could add to the already likely 3% error resulting from measurement, and therefore should be kept as small as possible.

RECOMMENDATIONS FOR PROCESSING GEIC DATA TO MEASURE VO_2MAX

Now that all aspects of data acquisition and processing in GEIC have been addressed, answers can be provided to earlier questions, and several recommendations can be made. As these answers and recommendations are dependent on the type of GEIC system used, they will be presented organized by GEIC method.

Time Averaged Systems

Data should be processed in as short a time interval as possible. To accomplish this, mouthpiece and tubing dead space must be small, and the delay in gas analyzer response to exhaled air must be short (<5 s). We showed that averaging durations of 15 s result in considerably less error for peak VO_2 values than 0.5 and 1 min averaging, especially when aligned to the center of the time average interval (Figure 2b). If averaging intervals of less than 15 s can be validly obtained by a time averaging system, than researchers should attempt to collect with these durations. However, a simple rule to guide sampling interval durations might be to always keep the interval at least double that of the time delay between ventilation measurement and expired air sampling by the gas analyzers. We provide no data to support this recommendation, and there is no published data available on this topic. Nevertheless, it is prudent to attempt to decrease this delay as longer delay times are associated with increased dead space, slower gas analyzer response times, and in general, a GEIC system that is less sensitive to changes in metabolism.

The data processing and subsequent procedure used to detect VO_2max using time-averaged systems depends on the mathematical definition of VO_2max. A historical bias to such a definition of VO_2max has been that it represents the maximal rate of VO_2 that can be sustained for a minute. Such a definition has within it the assumption that a 1 min data average is an essential component of this measure. Is this a valid definition? We believe that this is not a valid definition as it is not based totally on physiology, and is more a reflection of the historical insensitivity of the GEIC methods dated to the mid-20th century. For example, such a definition implies that all people should be able to attain and retain VO_2max for more than 1 min at the end of an incremental exercise protocol. We do not want to get into the debate over the incidence and detection of a VO_2 plateau. Nevertheless, past research has clearly revealed that the longer the time interval, the more difficult it is to show a plateau at VO_2max (6,14). Thus, it is unrealistic to assume that a 1 min average of VO_2 is a functionally valid mathematical construct to use in the detection of VO_2max.

The next important issue is the error inherent in small time interval data averages, and how to remove this error and improve the validity of using the peak VO_2 as VO_2max. Clearly, the large time interval average is not recommended as a method of removing error, despite comments by Howely et al. (8) that support such an approach. An alternative is to apply a filter, as we did to the breath-by-breath data of Figure 7b. The real breath-by-breath data of Figure 5a have been adjusted to give 0.25 min averages, aligned to the center of the time interval and are presented in Figure 9. For comparison, the results of interpolation (0.25 min intervals) and filtering are also provided. The peak VO_2 of the test was 3251.82 mL/min, and was adjusted down to a value of 3225.08 mL/min after filtering. The use of a filter is a more technically sound approach at removing non-systematic error in VO_2 data compared to relatively large time averages, and should be used to detect the peak VO_2 from GEIC during incremental exercise to VO_2max. Other criteria should then be applied to ascertain if the peak VO_2 is truly VO_2max (3,5,8,10,11,14,17,18).

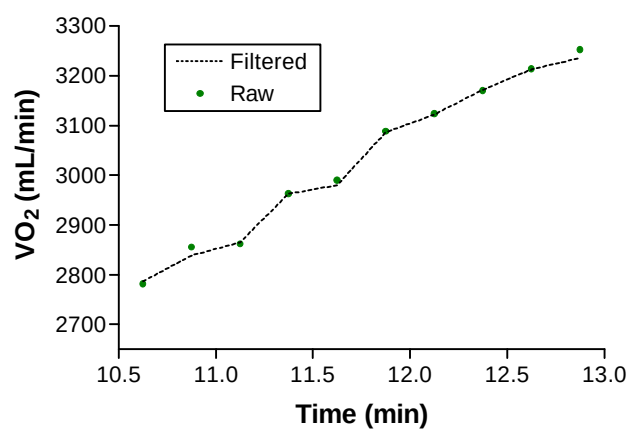

Figure 9. The differences between the 0.25 min time average data aligned to the center of the time interval and adjusted after interpolation and filtering (dashed line).

Finally, and as for all methods of GEIC, the error of data processing is error that is additional to measurement error. As presented in Figures 2-8, time or breath averaging, data smoothing, or filtering all adjust raw data, and when based on our data can result in peak VO_2 data alterations that vary from 3 to 134 mL/min (0.04 to 1.8 mL/kg/min, for a 75 kg person). When compared to the likely measurement error for GEIC of ±3%, these errors appear small. However, as the data processing errors are additional to measurement error, ignoring them can change a total error of measurement to in excess of 5%. Logically, given the potential size of the measurement error, data processing error should be kept negligible.

Breath Averaged Systems

The 7-breath average aligned to the center of the data period was previously shown to be highly accurate, with negligible systematic error (Figure 3b) causing underestimation of VO_2 by < 0.5 mL/min. Obviously, this is a suitable method of data presentation. Further improvement would result from interpolation and filtering, as completed for the 0.25 min data explained above.

Breath-by-Breath Systems

Breath-by-breath data acquisition in GEIC may best be processed by interpolation and filtering. We showed that such procedures reveal that there is a need to remove error in the data, and that this error can be removed by the equivalent of a 3 data point averaging (smoothing) function. Any smoothing function greater that 3 data points has an increased likelihood to remove "true" biological variability in the data, subsequently lowering any average VO_2 value from "true". At this time, we do not want to recommend how to select VO_2max from breath-by-breath data. Research needs to be done to compare values for VO_2max between different strategies of data manipulation for the time region where VO_2max occurs. Such manipulations could involve a time block average, nonlinear curve fitting, statistically testing slope differences from zero for this range of data, or simply detecting the highest data point after smoothing, interpolation and/or filtering.

Data Processing For Other End Measures of GEIC

Although our main application of the examples of data processing in this manuscript has been for the detection of VO_2max, our recommendations have additional application. For example, any time average of VO_2 data from GEIC that has a slope significantly greater than zero will lower the value from what it truly is. Consequently, use of VO_2 for quantifying exercise intensities at metabolic thresholds will be underestimated.

Additional Research

Given the constructive criticism we direct to the science of GEIC, additional research is needed to better understand the correct methods of data processing for specific systems. The recommendations we provide are the best that we can devise at this time, and each recommendation needs to be assessed by research, and for specific end measures (eg. VO_2max, ventilation threshold, oxygen deficit, excess post-exercise VO_2, etc.).

Despite the science of GEIC being more than 100 years old, numerous data processing procedures are still performed that have no clear technical or physiological rationale. This reality is not a characteristic of good science, and obviously detracts from the quality of research that is based on this method. We encourage scientists and educators within exercise physiology and the related disciplines and professions to reach agreement on how to process data from indirect calorimetry. We have provided objective evidence in support of our recommendations, and welcome additional commentary and research on this topic.

Address for correspondence: Robert A. Robergs, Ph.D., FASEP, EPC, Director: Exercise Physiology Laboratories, Exercise Science Program, Department of Physical Performance and Development
Johnson Center, Room B143, The University of New Mexico, Albuquerque, NM 87131-1258
Phone: (505) 277-2658; FAX: (505) 277-9742; Email: rrobergs@unm.edu

---

REFERENCES

1. Bassett DR, Howley ET. Maximal oxygen uptake: "classical" versus "contemporary" viewpoints. ***Med Sci Sports Exerc*** 1977;29(5):591-603.

2. Bassett DR, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. ***Med Sci Sports Exerc*** 2000;32:70-84.
3. Bergh U, Ekblom B, Astrand P. Maximal oxygen uptake "classical" versus "contemporary" viewpoints. ***Med Sci Sports Exerc*** 2000;32:85-88.
4. Mitchell JH, Sproule BJ, Chapman CP. The physiological meaning of the maximal oxygen intake test. ***J Clin Invest*** *1958;*37:538-546.
5. Noakes TD. Maximal oxygen uptake: "classical" versus "contemporary" viewpoints: a rebuttal. ***Med Sci Sports Exerc*** 1998;30:1381-1398.
6. Robergs RA. An exercise physiologists interpretations of the "ugly and creaking edifices" of the VO2max concept. ***JEP****online* 2001;4(1):1-44.
7. Taylor HL, Buskirk EF, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. ***J Appl Physiol*** 1955;8:73-80.
8. Howley ET, Bassett DR, Welch HG. Criteria for maximal oxygen uptake: review and commentary. ***Med Sci Sports Exerc*** 1995;27:1292-1301.
9. Mitchell JH, Blomqvist G. Maximal oxygen uptake. ***New Eng J Med*** 1971;284:1018-1022.
10. Cumming GR, Borysyk LM. Criteria for maximum oxygen uptake in men over 40 in a population survey. ***Med Sci Sports Exerc*** 1972;4:18-20.
11. Duncan GE, Howley ET, Johnson BN. ***Med Sci Sports Exerc*** 1997;29:273-278.
12. Freedson P, Kline G, Porcari J, Hintermeister R, McCarron R, Ross J. Criteria for defining VO_2max: a new approach to an old problem. ***Med Sci Sports Exerc*** 1986;18:S36.
13. Hill AV, Lupton H. Muscular exercise, lactic acid, and the supply and utilization of oxygen. ***Q Q J Med*** 1923;16:135-171.
14. Astorino TA, Robergs RA, Ghiasvand F, Marks D, Burns S. Incidence Of The Oxygen Plateau at VO_2max During Exercise Testing To Volitional Fatigue. ***JEP****online* 2000;3(4):1-12.
15. Myers J, Walsh D, Sullivan M, Froelicher VF. Effect of sampling on variability and plateau in oxygen uptake. ***J Appl Physiol*** 1990;68:404-410.
16. Winter D. **Biomechanics and motor control of human movement**. New York: John Wiley and Sons, Inc., 1991.
17. Myers J, Walsh D, Buchanan N, Froelicher VF. Can maximal cardiopulmonary capacity be recognized by a plateau in oxygen uptake? ***Chest*** 1989;96:1312-1316.
18. Wyndham CH, Strydom NB, Maritz JS, Morrison JF, Peter J, Potgieter ZU. Maximal oxygen intake and maximum HR during strenuous work. ***J Appl Physiol*** 1959;14:927-936.