Determining Day-to-Day Human Variation in Indirect Calorimetry Using Bayesian Decision Theory

Matthew S. Tenan[1], Addison W. Bohannon[2], Duncan J. Macfarlane[3], Scott E. Crouter[4]

[1]U.S. Army Research Laboratory, Research Triangle Park, NC, USA
[2]U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, USA
[3]Hong Kong University, Hong Kong, China
[4]University of Tennessee at Knoxville, TN, USA

Running Title: Human Variation in Indirect Calorimetry

Keywords: $\dot{V}O_2$, Bayesian, Probability, Individual Differences, Ventilation, $\dot{V}E$, $\dot{V}CO_2$

Total Number of Words (excluding references & figure legends): 3,175

Total Number of References: 21

Corresponding Author Address:

Matthew S. Tenan PhD
U.S. Army Research Laboratory
800 Park Office Drive
Suite #4229
Research Triangle Park, NC 27709
matthew.s.tenan.civ@mail.mil

Subject Area: Human, Environmental & Exercise

New Findings

- **What is the central question of this study?**
  We sought to understand the day-to-day variability of human indirect calorimetry during rest and exercise. Previous work has been unable to separate human day-to-day variability from measurement error and within-trial human variability. We developed models accounting for different levels of human and machine-level variance and compared the probability density functions using total variation distance.
- **What is the main finding and its importance?**
  After accounting for multiple levels of variance, the average human day-to-day variability of $\dot{V}E$, $\dot{V}CO_2$, and $\dot{V}O_2$ are 4.0%, 1.8% and 2.0%, respectively. This work presents a novel method to understand human variability as a whole and directly enhances our understanding of human variance during indirect calorimetry.

Abstract

One of the challenges of precision medicine is understanding when serial measurements taken across days are quantifiably different from each other. Previous work examining indirect calorimetry gas exchange has been unable to effectively separate differential measurement error, within-trial human variance and day-to-day human variance to ascertain how variable humans are across testing sessions. We use previously published reliability data to construct models of indirect calorimetry variance and compare these models with methods arising from Bayesian decision theory. These models are conditional on the data upon which they are derived and assume that error conform to a truncated normal distribution. A serial analysis of modeled probability density functions demonstrated that the average human day-to-day variance in $\dot{V}E$, $\dot{V}CO_2$, and $\dot{V}O_2$ are 4.0%, 1.8% and 2.0%, respectively. However, the average day-to-day variability masks a wide range of non-linear variance across flow rates, particularly for $\dot{V}E$. This is the first report isolating day-to-day human variability in indirect calorimetry gas exchange from other sources of variability. This method can be extended to other physiological tools and an extension of this work facilitates a statistical tool to examine within-trial $\dot{V}O_2$ differences, available in a graphical user interface.

Introduction

In the age of precision medicine, there has been renewed interest in understanding, quantifying and using individual-level variability to improve practice (Collins & Varmus, 2015). The concept of treating or training an individual person and not a group-average is not new; however, our treatment of data in the research domain often does not reflect this individual-level perspective. Furthermore, there is a tendency to see apparent changes across time in an individual person's data and assume that these differences are "real". In order for differences at the individual-level to be assumed "real" they must exceed both the measurement error of the device and any natural variability inherent in the human. Once the individual changes across time are determined to be "real", it is important to consider whether the changes are "meaningful" (Hopkins, 2004; Hecksteden *et al.*, 2015; Hopkins, 2015). The present work deals with the former consideration in the application of indirect calorimetry.

Previous work on gas exchange indirect calorimetry has indicated that human variance in $\dot{V}O_2$ between days ranges from 1.4% to 8.5% (Crouter 2006, Katch 1982, Armstrong & Costill 1985, Pereira 1994, Pereira & Freedson 1997); however, these studies generally use extremely small study samples, do not examine a wide variety of workloads, do not report variance in other common variables of interest (e.g. $\dot{V}CO_2$ and $\dot{V}E$) and may not have adequately separated human variance from device measurement error. Indeed, most studies examining day-to-day variability in human exercise indirect calorimetry actually have variance at three different levels: 1) device measurement error, 2) within-exercise trial human variance, and 3) day-to-day human variance. Using the gold standard Douglas bag method for exercise indirect calorimetry and exercise at five different absolute intensities, Crouter et al. (2006) showed day-to-day total variance in exercise $\dot{V}O_2$, $\dot{V}CO_2$ and $\dot{V}E$ to be 8.5%, 5.3% and 6.0%, respectively, without any adjustment for measurement error. Performing $\dot{V}O_2$ max tests on multiple days using the Douglas bag method and five subjects, Katch et al. (1982) showed total variation to be 5.6% and concluded that ~7.0% of the total variation was attributable to the methodology equipment. Armstrong and Costill (1985) examined 10 subjects and three absolute workloads and indicated that day-to-day human variance in $\dot{V}O_2$ was 2.84-4.32% after subtracting out a constant instrument fluctuation error (0.41% variance). Pereira conducted two studies, one on five subjects at a workload approximating 70% $\dot{V}O_2$ max (Pereira 1994) and one on 15 subjects at 70% (Pereira). They concluded that day-to-day variance in biological $\dot{V}O_2$ was 1.4% and 1.7% after removal of 'technical variation' which was <0.1% and 0.1% in the respective studies (Pereira 1997; Pereira 1994) While all of the previous work examining human- and machine-level variability is helpful to understand general variation, all of these examinations of gas exchange indirect calorimetry kinetics ignore the expiratory flow-based differential measurement error in both mixing chamber (Macfarlane & Wu, 2013; Tenan, 2016) and breath-by-breath systems (Huszczuk & Haouzi, 2016). The actual measurement error is highly device specific (Tenan 2016) and may be variable for reasons of calibration and component part signal drift which create a unique non-linear measurement error across ventilatory volumes. Previous work by Tenan (2016) has explicitly modeled both the machine-level and total human-level variability across varying volumes of gas flow to probabilistically determine if two indirect calorimetry testing periods are different on a day-to-day basis. The innovation proposed by Tenan was in representing the true underlying measure of $\dot{V}E$, $\dot{V}CO_2$, and $\dot{V}O_2$ as a probability density function (PDF) which incorporated measurement uncertainty while also accounting for the effects of differential measurement error. However, the primary weakness in this work is the inability to separate device measurement error and human within-trial variability from day-

to-day human variability. This weakness makes it a vastly over-conservative measure of indirect calorimetry changes on a within-trial basis.

The goal of this manuscript is to examine the amount of day-to-day human variation that exists in indirect calorimetry gas exchange by extending and improving upon the general framework employed by Tenan (2016). Furthermore, an extension of this methodology enables a single-subject, within-trial statistical test to determine the probability that $\dot{V}O_2$, $\dot{V}CO_2$, or $\dot{V}E$ levels have changed with a change in physical workload or other physical perturbation. These new methods are incorporated into a developmental toolbox and website maintained by the authors for general scientific and clinical use (https://tenan.shinyapps.io/gassim/). This knowledge and the dissemination of data products which can apply the tools developed in this work facilitate the ability of researchers, clinicians and coaches to better discern whether changes in single-subject gas exchange indirect calorimetry are due to natural variation or the result of an intervention.

Methods

All computations described in this study are performed in the R programming language (Version 3.4.4), and a series of R packages: dplyr (Version 0.7.4), devtools (Version 1.13.5), geepack (Version 1.2-1), msm (Version 1.6.6), cubature (Version 1.3-11), and ggplot2 (Version 2.2.1). The general flow of analyses performed in this study are as follows:

1. Define the models for overall variability of human gas exchange (human variation [within-trial and between days] plus machine measurement error) in indirect calorimetry ($\dot{V}E$, $\dot{V}O_2$ and $\dot{V}CO_2$).
2. Define the models which include only within-trial variability and machine measurement error in indirect calorimetry.
3. Using the models previously defined in steps 1 and 2, determine the amount of day-to-day variability in human indirect calorimetry which exceeds that attributable to within-trial human variability and machine measurement error.

Data Underlying the Probability Density Function Models

The inclusion for data in the models were:

1. Raw data for the study must be either publicly available or made available by the authors of previously published work.
2. The previously published work must have used the Parvomedics 2400 TrueOne system because machine measurement error is device-specific.
3. The data from the published work needed to conform to one of two designs:
    a. The study data needed to contain variance at only three levels (human variation [within-trial and between days] plus machine measurement error)
    b. The study data needed to contain variance at only two levels (human within-trial variation and machine measurement error)
4. For studies conforming to criteria 3a, the measurements needed to have occurred less than 10 days apart with no intervention in between.

Based on the above criteria, the experimental data used in the present analysis have been previously published and described in two separate studies (Crouter *et al.*, 2006; Macfarlane & Wu, 2013). Thus, it was determined that this work did not require regulatory approval as defined by the Common Rule, whereby existing datasets of de-identified data are excluded from review. Study 1, performed by Crouter et al. (2006), used the ParvoMedics TrueOne 2400 to determine gas exchange variables ($\dot{V}E$, $\dot{V}O_2$, and $\dot{V}CO_2$) while performing a progressive cycling task (rest, 50, 100, 150, 200, and 250 W). Two trials were performed within 48 hours to examine the reliability of $\dot{V}E$, $\dot{V}O_2$ and $\dot{V}CO_2$ across multiple days. Study 2, performed by Macfarlane and Wu (2013), used two ParvoMedics TrueOne 2400 units aligned in parallel (termed "collateral configuration"), so that each unit obtained measurements from the same steady state work rate. This arrangement examines the combination of within-work rate human variance and machine-level variance in $\dot{V}E$, $\dot{V}O_2$ and $\dot{V}CO_2$ metrics during a progressive cycling task (rest, 30, 60, 90, 120 W). The raw data obtained from these studies were used to define a model of uncertainty for each ventilatory metric, as described below. In order to ensure model stability, the derived models did not extrapolate beyond the minimum and maximum values obtained in the original datasets (Table 1).

Determination of Gas Probability Density Function Models and Calculation of Human Variability

The models of variability for each ventilatory metric used in the present paper are consistent with the framework prescribed by Tenan (2016). Briefly, uncertainty in the true measurement ($\dot{V}E$, $\dot{V}O_2$, or $\dot{V}CO_2$) is modeled with a univariate truncated normal distribution with mean given by the observed measurement. To account for the differential measurement error, the standard deviation of the distribution is modeled as a function of the observed measurement. That function results from fitting a third-order polynomial generalized estimating equation regression for the mean and standard deviation of the test-retest measurements of the mean gas variable ($\dot{V}E$, $\dot{V}O_2$, or $\dot{V}CO_2$) obtained in the respective study (Study 1 or Study 2), clustering for the multiple observations on a single subject. Therefore, each measurement yields a distribution of measurements with central tendency influenced by the gas flow rate, consistent with observations (Macfarlane & Wu, 2013). Formally, let $X: \Omega \to R$ be a random variable representing the measurement of the gas flow variable, and let $g: R \to R$ be the generalized estimating equation, a third-order polynomial. Then, the model of variability is a truncated normal random variable $Y: \Omega \to R$ with mean X, variance $g^2(X)$, and support on $[0, \infty)$.

$$Y \sim N\big(X, g(X), [0, \infty)\big)$$

The primary differences between the current model formulation and Tenan (2016) is the use of a generalized estimating equation methodology versus a general linear model as well as the use of a truncated normal distribution instead of a standard normal distribution. In contrast to the general linear model, the generalized estimating equation provides theoretically superior model estimates while controlling for multiple observations. The truncated normal distribution is preferable to a standard normal distribution because $\dot{V}E$, $\dot{V}O_2$, or $\dot{V}CO_2$ measurements are lower bounded by zero. Both of these modifications to the original Tenan (2016) method are included in updates to the Gas.Sim package for the R programming language. Once the models for calculating expected overall variation (i.e. the combined day-to-day human variance, within-trial human variance and machine measurement error) and within-trial human variance plus machine-level variation are constructed, the respective measurements can be compared via the total variation distance (Gibbs & Su, 2002). Formally, let $Y \sim N\big(X, g(X), [0, \infty)\big)$ be a random variable representing day-to-day measurements with PDF f_Y and $Z \sim$

$N\big(X, h(X), [0, \infty)\big)$ be a random variable for within-trial measurements with PDF f_Z, where g and h are the respective generalized estimating equations of Y and Z.

The overlapping coefficient measures the similarity of two random variables (Weitzman, 1970; Inman & Bradley Jr, 1989).

$$s(Y, Z) = \int_0^\infty \min[f_Y(x), f_Z(x)]\, dx$$

In Bayesian decision theory, this quantity defines the minimum decision risk and is interpreted as the probability that any observation of overall measurement could be an observation of within-trial measurement and vice versa. While accounting for the differential measurement error in the measurement device, Tenan (2016) used the overlapping coefficient to determine if two indirect calorimetry testing periods differed on a day-to-day basis. In this case, a high overlapping coefficient would indicate a low likelihood of substantial difference in the given ventilatory measure, and a low overlapping coefficient would indicate that the observed ventilatory measure on one day was unlikely to have been observed on another day (i.e., the observed variability is likely to have resulted from a change in the true underlying ventilatory metric).

In the present study, the goal is to discriminate between the variability introduced by the measurement device and that of the inherent day-to-day variability of humans and to attribute changes in indirect calorimetry to true changes in the ventilatory measure of interest. As the overlapping coefficient takes values in $[0, 1]$ and measures similarity, the complement provides a normalized measure of dissimilarity. Intuitively, it measures the probability that an observation of a random variable cannot be the realization of another random variable, which should aid in disambiguating the effects of day-to-day human variability. The complement of the overlapping coefficient is known as the total variation distance (Gibbs & Su, 2002),

$$d(Y, Z) = \int_0^\infty |f_Y(x) - f_Z(x)|\, dx.$$

The total variation distance defines a valid metric on probability density functions (i.e., non-negative, symmetric, and sub-additive). Thus, it is a probabilistic measure of day-to-day human variability in indirect calorimetry gas exchange apart from other sources of error.

These models can be defined across all gas flow rates to examine the amount of human variation at different work rates as well as determine the mean level of human variation in indirect calorimetry measures on a day-to-day basis. $\dot{V}E$ flow rates from 6 – 60 L/min were assessed in 0.1 L/min increments whereas $\dot{V}O_2$ and $\dot{V}CO_2$ were assessed in the 0.28 – 2.00 L/min and 0.26 – 2.00 L/min ranges in 0.01 L/min increments, respectively. As noted earlier and in Table 1, these limits are defined by the maxima and minima of the two data sets on which the models are based. In practice, the limits are bounded on the low end by the Crouter et al. data and on the high end by the Macfarlane & Wu data.

Extension of Methodology for a Statistical Test of Within-Trial Changes in Human Gas Exchange

The models derived from Macfarlane and Wu (2013) enable the creation of a statistical test for within-trial differences in human gas exchange in line with the framework established by Tenan (2016). Since this model contains only within-trial human variability and machine level differential measurement

error, the proposed procedure can be used to determine the probability that the two different experimental measures from within the same trial arose from the same distribution. Since the differential measurement error is device specific, this methodology is currently limited to the Parvomedics 2400 TrueOne system. A function for this statistical test has been added to the Gas.Sim package for R (https://github.com/TenanATC/gas.sim) as well as an online web application (https://tenan.shinyapps.io/gassim/).

Results

A series of example analyses for human day-to-day variability in $\dot{V}O_2$ at different flow rates is shown in Figure 1. The mean variance in $\dot{V}E$ attributable to human variability is 4.0%, though the amount of variance differs considerably across a range of $\dot{V}E$ flow rates (range: 0.5% – 12.1%) (Figure 2A). The mean variance in $\dot{V}O_2$ and $\dot{V}CO_2$ attributable to human variability is 2.0% (range: <0.1% - 6.6%) and 1.8% (range: 0.1% - 3.1%), respectively, and this variance remains generally constant across flow rates (Figures 2B and 2C).

As described in the methods above, the PDF models derived from Macfarlane and Wu (2013) can be used to develop a single-subject analysis for within-trial differences in gas exchange. This same analytic technique can be compared across a stratum of gas flow rates to examine within-trial differences in gas exchange more broadly. The gross results of these analyses are visualized as a heat map in Figure 3. This visualization demonstrates some small fluctuations in reliability across gas flow rates, particularly at high flow rates, but overall the measurement remains relatively stable. This supports the idea that while there is some level of differential measurement error at the device-level, the majority of differential error may arise from the human.

Discussion

This report uses established methods from probability-based decision theory, epidemiology (differential measurement error) and previously published work in the human physiology literature (Crouter *et al.*, 2006; Macfarlane & Wu, 2013; Tenan, 2016) to determine the variability in human indirect calorimetry gas exchange on a day-to-day basis. Previous research has indicated that day-to-day variance in $\dot{V}O_2$ ranges from 1% to 8.5% (Katch *et al.*, 1982; Armstrong & Costill, 1985; Murgatroyd *et al.*, 1987; Donahoo *et al.*, 2004; Crouter *et al.*, 2006). However, all of the previous literature has either implicitly or explicitly included day-to-day variance with one (Katch *et al.*, 1982; Armstrong & Costill, 1985) or two (Murgatroyd *et al.*, 1987; Donahoo *et al.*, 2004; Crouter *et al.*, 2006) other types of variance (within-exercise trial variance or device measurement error and within-exercise trial variance, respectively). Furthermore, the previous studies which have attempted to account for device measurement error have considered the error to be constant (Katch *et al.*, 1982; Armstrong & Costill, 1985); recent literature has shown that measurement error is not constant and changes with gas flow rate in both mixing chamber (Macfarlane & Wu, 2013) and breath-by-breath systems (Huszczuk & Haouzi, 2016). The present analysis models both measurement variability which includes only device measurement error and within-trial human variance from published data by Macfarlane (2013) and which includes device measurement error, within-trial human variance, and day-to-day human variance from Crouter's

published data (Crouter *et al.*, 2006). The net difference between these two models of probability density functions, as calculated via total variation can be attributed to the day-to-day human variance in indirect calorimetry gas exchange. We can conclude that day-to-day variance in healthy humans for $\dot{V}O_2$ and $\dot{V}CO_2$ are about 2.0% and 1.8%, respectively, and that there is some level of variability across gas flow rates (Figure 2B and 2C). Day-to-day human variance in $\dot{V}E$ is 4.0% on average, but exhibits more extreme and notable non-linearity across workloads/flow rates (Figure 2A).

Since the present methodology eliminates variance at more levels and accounts for differential measurement error, it is not surprising that the ~2% day-to-day variance in $\dot{V}O_2$ is lower than most previous studies (Katch *et al.*, 1982; Armstrong & Costill, 1985; Murgatroyd *et al.*, 1987; Crouter *et al.*, 2006). One practical reason for this low level of daily variance in $\dot{V}O_2$ may be due to the $\dot{V}O_2$ ranges available for use in the present analysis. In order to maintain model stability, it was necessary to only model within the ranges of data available (Table 1). For $\dot{V}O_2$, the daily variance is examined from 0.2 L/min to 2.0 L/min, which is a low to moderate $\dot{V}O_2$ level for many healthy moderately active individuals. It is entirely possible that daily variance in $\dot{V}O_2$ becomes greater at levels approximating $\dot{V}O_2$ max. Interestingly, the highest level of daily variance for ventilatory measures is $\dot{V}E$ and this is also the measure exhibiting the greatest nonlinear variance across work rate (Figure 2A). Since $\dot{V}E$ results from an interaction between tidal volume and breathing frequency, it is interesting to speculate as to how each of these individual values may vary on a daily basis and what those underlying mechanisms may be. For example, breathing frequency is highly influenced by psychological stress/perception of effort (Nicolò *et al.*, 2016; Nicolò *et al.*, 2017); if we assume tidal volume remains relatively constant for a given physical workload, a differential change in psychological stress could result in less reliable $\dot{V}E$ measure on a day-to-day basis. Additional variability that may disproportionally affect $\dot{V}E$, as opposed to $\dot{V}O_2$ and $\dot{V}CO_2$, are sudden transients such as coughing, swallowing or yawning during data collection. Any uncoupling between metabolic demands to complete a given exercise work rate and the components of $\dot{V}E$ may contribute the observed non-linearity in reproducibility.

Notably, an extension of the analysis framework established by Tenan (2016) to understand day-to-day variability in human gas exchange facilitates the development of a statistical test for within-trial single-subject analysis. This statistical test is implemented in the R package Gas.Sim and a graphical user interface implementation (https://tenan.shinyapps.io/gassim/). Across the flow rates available in the current model, there is some degree of consistency in the reliability (Figure 3). The probability density functions for $\dot{V}O_2$ appear to have greater dispersion at higher flow rates, which may be physiological in nature or a result of higher flow rates creating a non-linearity in device accuracy. The non-linearity in reproducibility has implications both for how we interpret changes in primary measures of indirect calorimetry (i.e. $\dot{V}E$, $\dot{V}O_2$, $\dot{V}CO_2$), but also derived measures, such as caloric expenditure (Beck *et al.*, 2018 [In Press]; Tenan *et al.*, 2018 [In Press]). The weakness of the statistical test for within-trial single-subject analysis is the low to moderate flow rates (Table 1) and that it is currently limited to the Parvomedics 2400 TrueOne system. As more data become available at higher workloads or additional systems, support will be added to the Gas.Sim package.

The current analysis framework attributes day-to-day human variability in indirect calorimetry gas exchange to the total variation between PDFs which account for different sources of measurement uncertainty. The total variation distance is one of numerous metrics on probability spaces, and it is applicable in the present report due to its decision-theoretic interpretation. However, metrics on probability spaces is a currently active area of research in statistical inference (Genevay *et al.*, 2016;

Arjovsky *et al.*, 2017), and it is possible that use of an alternative metric for the proposed analysis framework such as the Wasserstein-Kantorovich distance (Genevay *et al.*, 2016) could yield additional insights on the day-to-day variability of indirect calorimetry gas exchange.

The present study proposed a novel analysis framework based on probabilistic decision theory to understand how variable human gas exchange is during exercise apart from other sources of error. In theory, this method can be applied to other variables of physiological interest where measurement variability can be modeled with probability density functions containing different types of variance. Through this analysis, we establish the degree of variability in $\dot{V}E$, $\dot{V}O_2$, and $\dot{V}CO_2$ across workloads and characterize any apparent nonlinearities. Furthermore, this work has a freely-available statistical package for the R programming language and a web-based graphical user interface to facilitate research and clinical use.

References

Arjovsky M, Chintala S & Bottou L (2017). Wasserstein generative adversarial networks. In *Proceedings of 34th International Conference on Machine Learning*, pp. 214-223, Sydney, Australia.

Armstrong LE & Costill DL (1985). Variability of respiration and metabolism: responses to submaximal cycling and running. *Res Q Exerc Sport* **56,** 93-96.

Beck ON, Kipp S, Byrnes WC & Kram R (2018 [In Press]). Use aerobic energy expenditure instead of oxygen uptake to quantify exercise intensity and predict endurance performance. *J Appl Physiol*.

Collins FS & Varmus H (2015). A new initiative on precision medicine. *N Engl J Med* **372,** 793-795.

Crouter SE, Antczak A, Hudak JR, DellaValle DM & Haas JD (2006). Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. *Eur J Appl Physiol* **98,** 139-151.

Donahoo WT, Levine JA & Melanson EL (2004). Variability in energy expenditure and its components. *Curr Opin Clin Nutr Metab Care* **7,** 599-605.

Genevay A, Cuturi M, Peyré G & Bach F (2016). Stochastic optimization for large-scale optimal transport. In *Adv Neural Inf Process Syst*, pp. 3440-3448.

Gibbs AL & Su FE (2002). On choosing and bounding probability metrics. *Int Stat Rev* **70,** 419-435.

Hecksteden A, Kraushaar J, Scharhag-Rosenberger F, Theisen D, Senn S & Meyer T (2015). Individual response to exercise training-a statistical perspective. *J Appl Physiol* **118,** 1450-1459.

Hopkins W (2015). Spreadsheets for analysis of validity and reliability. *Sportscience* **19,** 36-42.

Hopkins WG (2004). How to interpret changes in an athletic performance test. *Sportscience* **8,** 1-7.

Huszczuk A & Haouzi P (2016). On the inaccuracy of breath-by-breath metabolic gas exchange systems. *Resp Physiol Neurobi* **233,** 14-16.

Inman HF & Bradley Jr EL (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun Statist -Theory Meth* **18,** 3851-3874.

Katch VL, Sady SS & Freedson P (1982). Biological variability in maximum aerobic power. *Med Sci Sports Exerc* **14,** 21-25.

Macfarlane D & Wu H (2013). Inter-unit variability in two ParvoMedics TrueOne 2400 automated metabolic gas analysis systems. *Eur J Appl Physiol* **113,** 753-762.

Murgatroyd P, Davies H & Prentice A (1987). Intra-individual variability and measurement noise in estimates of energy expenditure by whole body indirect calorimetry. *Br J Nutr* **58,** 347-356.

Nicolò A, Marcora SM, Bazzucchi I & Sacchetti M (2017). Differential control of respiratory frequency and tidal volume during high-intensity interval training. *Exp Physiol* **102,** 934-949.

Nicolò A, Marcora SM & Sacchetti M (2016). Respiratory frequency is strongly associated with perceived exertion during time trials of different duration. *J Sport Sci* **34,** 1199-1206.

Tenan MS (2016). A statistical method and tool to account for indirect calorimetry differential measurement error in a single-subject analysis. *Front Physiol* **7**.

Tenan MS, Macfarlane D & Hackney AC (2018 [In Press]). Converting oxygen uptake to energy expenditure substantially increases prediction error. *J Appl Physiol*.

Weitzman MS (1970). *Measures of overlap of income distributions of white and Negro families in the United States*, vol. Technical Report No. 22. US Bureau of the Census, Washington, DC.

Competing Interest

None

Author Contributions

The data underlying this work were collected by DJM and SEC. The conception, design and analysis of the current modeling work was performed by MST and AWB. All authors contributed to the drafting of the work and critically revising it for important intellectual content. Furthermore, all authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship and all those who qualify for authorship are listed.

Funding

No extramural funding was received for this work

Table 1. Range of gas metrics in studies underlying models

	Gas Metric	Minimum (L/min)	Maximum (L/min)
Crouter et al. 2006	$\dot{V}E$	7.67	131.30
	$\dot{V}O_2$	0.28	3.86
	$\dot{V}CO_2$	0.26	4.19
Macfarlane and Wu 2013	$\dot{V}E$	6.09	62.57
	$\dot{V}O_2$	0.20	2.07
	$\dot{V}CO_2$	0.17	2.14

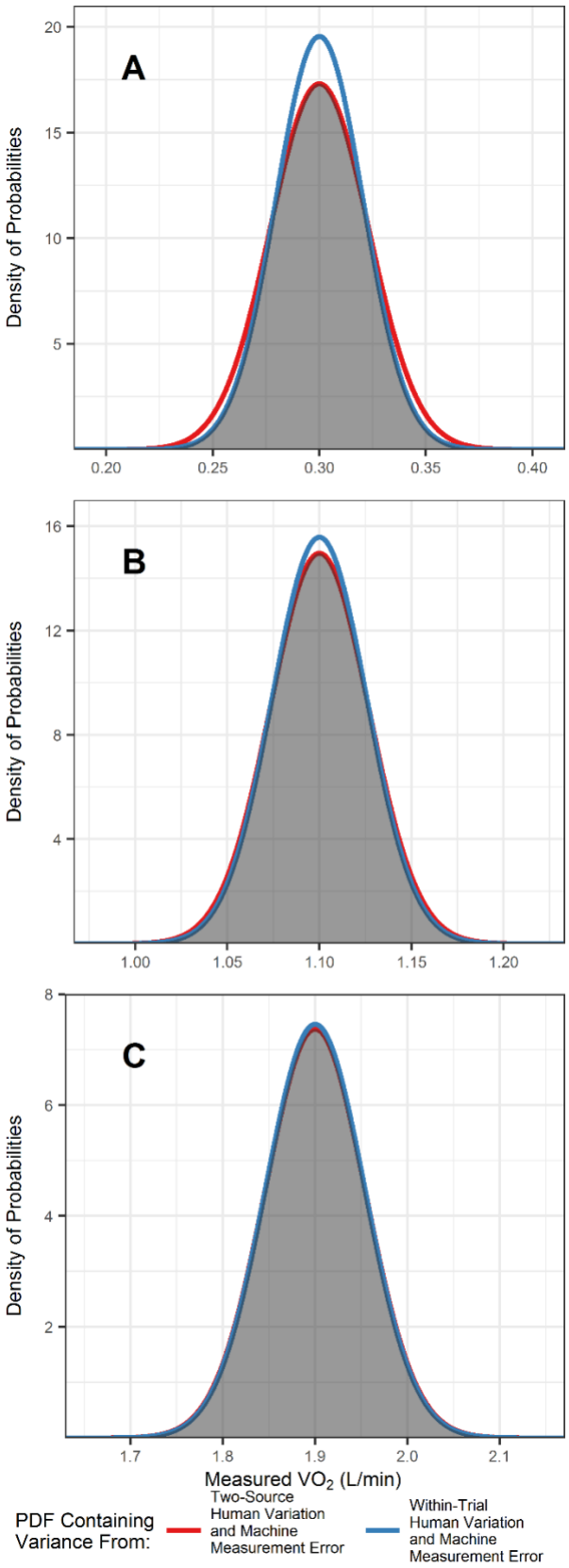

Figure 1. Example of Overlapping Probability Density Functions (PDF) for measurements containing different levels of variance at discrete flow rates. The example $\dot{V}O_2$ flow rates are: 0.3 L/min (A), 1.1 L/min (B), and 1.9 L/min (C).

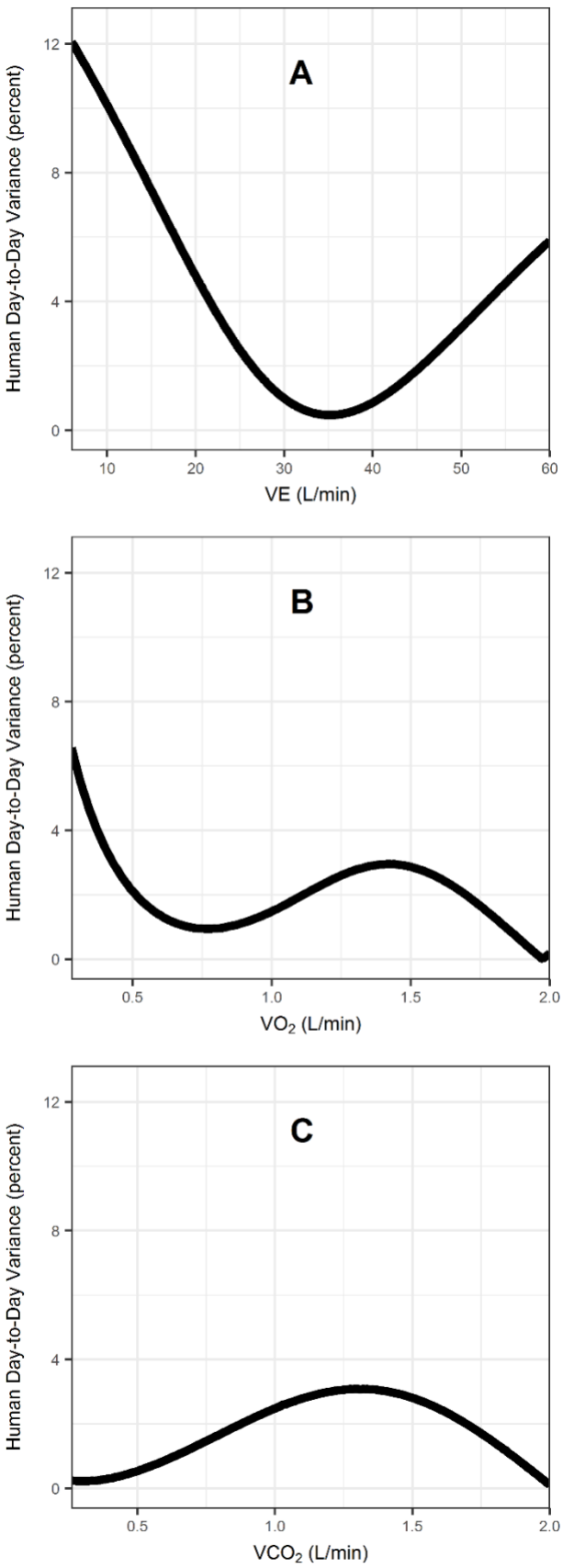

Figure 2. Percent variance in $\dot{V}E$ (plot A), $\dot{V}O_2$ (plot B), and $\dot{V}CO_2$ (plot C) attributable to day-to-day human variability across a range of gas flow rates.

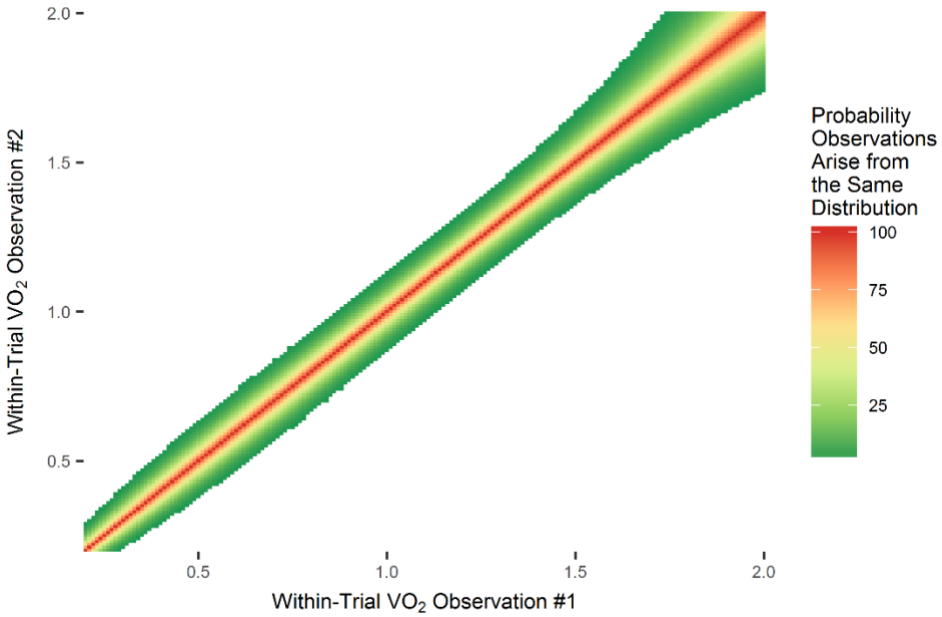

Figure 3. A heat map visualization of within-trial $\dot{V}O_2$ variance across all flow rates from 0.2 – 2.0 L/min for the Parvomedics 2400 TrueOne. All probabilities less than 1% are white.