Title

Variability in Cardiopulmonary Exercise Testing Biologic Controls.

Journal

Respiratory Care, 68(1)

Authors

Haverkamp, Hans

Gooding, Thomas

Collingridge, Dave

et al.

Publication Date

2023

DOI

10.4187/respcare.10022

Peer reviewed

Variability in Cardiopulmonary Exercise Testing Biologic Controls

Thomas W DeCato, Hans C Haverkamp, Thomas Gooding, Dave S Collingridge, and Matthew J Hegewald

BACKGROUND: Cardiopulmonary exercise testing is an increasingly common test and is considered the accepted standard for assessing exercise capacity. Quantifying variability is important to assess the instrument for quality control purposes. Though guidelines recommend biologic control testing, there are minimal data on how to do it. We sought to describe variability for oxygen consumption ($\dot{V}_{O_2}$), carbon dioxide production ($\dot{V}_{CO_2}$), and minute ventilation ($\dot{V}_E$) at various work rates under steady-state conditions in multiple subjects over a 1-y period to provide a practical approach to assess and perform biologic control testing. METHODS: We performed a single-center, prospective study with 4 healthy subjects, 2 men and 2 women. Subjects performed constant work rate exercise tests for 6 min each at 25–100 W intervals on a computer-controlled cycle ergometer. Data were averaged over the last 120 s at each work rate to reflect stepwise steady-state conditions. Descriptive statistics, including the mean, median, range, SD, and coefficient of variation (CoV) are reported for each individual across the 4 work rates and all repetitions. As these data were normative, z-scores were utilized, and a value greater than ± 1.96 z-scores was used to define significant test variability. RESULTS: Subjects performed 16–39 biocontrol studies over 1-y. The mean CoV for all subjects in $\dot{V}_{O_2}$ was 6.59%, $\dot{V}_{CO_2}$ was 6.41%, and $\dot{V}_E$ was 6.32%. The ± 1.96 z-scores corresponded to a 9.4–18.1% change in $\dot{V}_{O_2}$, a 9.6–18.1% change in $\dot{V}_{CO_2}$, and a 9–21.5% change in $\dot{V}_E$ across the 4 workloads. CONCLUSIONS: We report long-term variability for steady-state measurement of $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$ obtained during biocontrol testing. Utilizing ± 1.96 z-scores allows one to determine if a result exceeds expected variability, which may warrant investigation of the instrument. *Key words: cardiopulmonary exercise testing; CPET; biocontrol; quality control.* [Respir Care 2023;68(1):38–43. © 2023 Daedalus Enterprises]

Introduction

Cardiopulmonary exercise testing (CPET) is a commonly used and clinically useful pulmonary laboratory test.[1,2] It is considered the accepted standard for assessing functional exercise capacity and for evaluating otherwise unexplained exertional dyspnea among other indications. Quantifying long-term (at least 1-y) variability of CPET parameters is important in instrument quality control. The accepted premise of physiologic validation, or biologic control testing (biocontrols), is that physiological responses to exercise in healthy individuals are highly repeatable. Assuming stable physiology of biocontrols, changes in CPET parameters that exceed normal variability suggest that an instrument is not providing precise results and that technical factors need to be addressed. Instrument measurement errors may result in spurious results leading to misdiagnosis, incorrect conclusions, unnecessary diagnostic evaluation, and even inappropriate therapy. There is limited information regarding long-term intra-individual variability of CPET parameters in biocontrols and the optimal method for assessing such variability.[3,4] The studies that are available utilized a treadmill, as opposed to a cycle ergometer that allows work rate to be more accurately quantified. The 2003 American Thoracic Society/American College of Chest Physicians (ATS/ACCP) statement on CPET recommends "physiologic/biologic validation, in which a healthy member of the laboratory staff, consuming a stable diet, performs a constant work rate test at varying workloads (eg, 50 and 100 or 150 W) at regular intervals depending on machine use."[5]

The purposes of this study were 2-fold: (1) to describe the normal intra-individual variability for oxygen consumption ($\dot{V}_{O_2}$), carbon dioxide production ($\dot{V}_{CO_2}$), and minute ventilation ($\dot{V}_E$) at several work rates under steady-state conditions over a 1-y time period; and (2) use this variability to provide a practical approach for performing biocontrol testing in a clinical CPET laboratory. The goal of this study was not to describe how well the instrument reflected true values (accuracy) but how close measurements were to each other

(precision). Subject variability is also referred to as repeatability, with low variability reflecting high repeatability.

Methods

Study Design and Subjects

We performed a single-center, prospective study with 4 healthy subjects, 2 men and 2 women, ranging in age from 34–56 y at the time of study initiation. All subjects had normal pulmonary function tests including spirometry and diffusing capacity for carbon monoxide, using National Health and Nutrition Examination Survey III and Crapo reference values, respectively, and normal symptom-limited maximal CPETs using Hansen reference values.[6-8] The Intermountain Healthcare Institutional Review Board provided a waiver of informed consent because this study involved normal laboratory quality control measures. No outside funding was provided for this study.

Testing Procedure and Data Collection

CPET was performed using a computer-controlled cycle ergometer with a Vyaire Vmax system (software version 28–7; model 229N E, Vyaire, Mettawa, Illinois). The cycle ergometer was calibrated for accuracy by the manufacturer prior to this study. The mass flow sensor and gas analyzers were calibrated before each test following manufacturer's recommendations and met current standards for accuracy, reproducibility, and response time.[5,9] Oxygen gas analyzer calibration was performed using 3-point calibration utilizing room air; primary standard gas mixture with 26.0% oxygen and balance nitrogen; and certified standard gas mixture with 16.0% oxygen, 4.0% carbon dioxide, and balance nitrogen. The barometric pressure and temperature were measured by the instrument and ranged from 642–670 mm Hg and 19.5–22.7°C, respectively. The subjects performed 4 constant work rate exercise tests for 6 min each at 25-W intervals (25–100 W). Subjects were instructed to avoid exercise and have a consistent light breakfast on all testing days. Our overall goal was for each subject to perform 2–3 tests per month over 1-y. Testing was not performed if the subject reported illness or injury. Testing was performed at the same time of day ± 2 h. A Hans Rudolph 7450 Series Silicone V2i Oro-Nasal Mask (Hans Rudolph, Shawnee, Kansas) was worn during testing. Each test was reviewed for evidence of air leak or other technical irregularities. Data were collected on a breath-by-breath basis and averaged over the last 120 s of each workload to reflect steady-state conditions. Each test was also reviewed for evidence that the anaerobic threshold had been attained or exceeded. Anaerobic threshold was exceeded if the $\dot{V}_{O_2}$ or $\dot{V}_{CO_2}$ versus time tracing was not linear (slope > 0). Because this appeared consistently in subject #4 at the 100-W work rate and represented a non–steady-state condition, these data were excluded.

QUICK LOOK

Current knowledge

Cardiopulmonary exercise testing (CPET) is commonly used and considered the accepted standard for assessing functional exercise capacity. However, there is limited information on long-term CPET biocontrol variability and how to best implement and perform biocontrol testing in a busy clinical pulmonary/exercise laboratory.

What this paper contributes to our knowledge

We show that long-term variability in oxygen consumption ($\dot{V}_{O_2}$), carbon dioxide production ($\dot{V}_{CO_2}$), and minute ventilation ($\dot{V}_E$) biocontrols was similar to that as described in other settings. The use of ± 5% for $\dot{V}_{O_2}$, ± 6% for $\dot{V}_{CO_2}$, and ± 5.5% for $\dot{V}_E$ may be too restrictive when performing biocontrol testing. Utilizing ± 1.96 z-scores is an alternate method that appears practical in a busy clinical laboratory.

Statistical Analysis

Descriptive statistics, including the mean, median, range, SD, and coefficient of variation (CoV), are reported for each individual at each work rate for $\dot{V}_{O_2}$,

Dr DeCato is affiliated with Division of Respiratory and Critical Care Physiology and Medicine and The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, California; and Department of Medical Education and Clinical Sciences, Elson S. Floyd College of Medicine, Washington State University, Spokane, Washington. Dr Haverkamp and Mr Gooding are affiliated with Department of Nutrition and Exercise Physiology, Elson S. Floyd College of Medicine, Washington State University, Spokane, Washington. Dr Collingridge is affiliated with Pulmonary and Critical Care Medicine, Intermountain Medical Center, Murray, Utah. Dr Hegewald is affiliated with Pulmonary and Critical Care Medicine, Intermountain Medical Center, Murray, Utah; and Division of Respiratory, Critical Care, and Occupational Pulmonary Medicine, University of Utah, Salt Lake City, Utah.

The authors have disclosed no conflicts of interest.

Drs DeCato and Hegewald presented a version of this paper at the ATS International Conference 2019, held in Dallas, Texas, May 17–22, 2019.

Supplementary material related to this paper is available at http://www.rcjournal.com.

Correspondence: Thomas W DeCato MD, Harbor-UCLA Medical Center, 1000 W. Carson Street, CDCRC Box 402, Torrance, CA 90509. Email: tdecato@ucla.edu.

DOI: 10.4187/respcare.10022

Table 1. Demographics and Mean $\dot{V}_{O_2}$ in L/min and mL/kg/min for Each Subject

	1	2	3	4
Age, y	56	55	34	36
Sex	M	F	M	F
Height, cm	193	158	173	168
Weight, kg	86.4	47.6	80	56
BMI, kg/m^2	23.2	19.8	26.7	19.8
Measures, no.	39	22	28	16
Mean $\dot{V}_{O_2}$, L/min				
25 W	0.75	0.51	0.68	0.61
50 W	0.96	0.75	0.92	0.88
75 W	1.23	1.04	1.20	1.15
100 W	1.51	1.37	1.49	*
Mean $\dot{V}_{O_2}$, mL/kg/min				
25 W	8.7	10.7	8.5	10.9
50 W	11.1	15.8	11.5	15.7
75 W	14.2	21.9	15	20.5
100 W	17.5	28.8	18.6	*

*These data were not included, as they were above the anaerobic threshold.

$\dot{V}_{CO_2}$, and $\dot{V}_E$. Average CoVs for $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$ across the 4 work rates are also reported to further describe the repeatability of these measures. Following Rozali and Wah's[10] recommendation, we evaluated normality of our measures with 3 approaches: (1) boxplots of z-scores, (2) skew and kurtosis, and (3) Shapiro-Wilk test of normality. Our criterion for establishing normality involved a combination of all 3 approaches. If the absolute value of the ratios of skew and kurtosis to their standard errors were less than 1.96, Shapiro-Wilk scores were nonsignificant, and boxplots resembled normality (ie, medians were near the mean horizontal line [Z-score = 0] and boxplot shape was approximately symmetrical), we concluded that the distribution for a particular measure was normally distributed. Because repeated biocontrol testing data satisfied all 3 normality criteria (Table 1 and Supplementary Fig. 1; see related supplementary material at http://www.rcjournal.com.), we were able to use z-scores, refer them to the standard normal distribution, and rely on a z-score threshold of $\pm$ 1.96 to identify significant test variability. According to standard normal distribution logic, z-scores $\pm$ 1.96 encompass 95% of observations and are commonly used to define observations within normal range.

A z-score measures the number of SD a data point lies above or below the mean. When doing biologic control testing, z-scores are calculated as testing value minus sample mean, divided by the sample SD ($[x - \overline{X}]$ /SD). Finally, for $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$ values greater than $\pm$ 1.96 z-scores from the mean, we calculated percent change from the original mean. Percent change values estimated how much each measure must deviate from its mean to be considered anomalous.

Results

Subject #1 performed 39 tests; subject #2 performed 22; subject #3 performed 28, and subject #4 performed 16 during the study period. The subjects age range was 34–56 y, and body mass index was 19.8–26.7 kg/m^2 (see Table 1). The mean $\dot{V}_{O_2}$ in L/min and mL/kg/min is also reported in Table 1. Descriptive statistics are reported in Table 2 for each subject across the 4 workloads for $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$. There were fewer measures for $\dot{V}_E$ as it was not recorded in the first several tests. For all 4 subjects, we found limited variability in the first 8 test scores (data not shown) for $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$. This limited variability led to our decision to start computing z-scores after 8 tests. The mean CoV for all subjects was 6.59% for $\dot{V}_{O_2}$, 6.41% for $\dot{V}_{CO_2}$, and 6.32% for $\dot{V}_E$ (excluding subject #4 from the 100-W work rate). The percentage change values from the original mean for $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E > 1.96$ z-scores were variable. Table 3 shows that for subject #1 the percentage change ranges from 11.1–15.9%. Corresponding values for subject #2 were 9.4–14.1%, subject #3 were 8.7–15.9%, and subject #4 were 11.5–21.5% (see related supplementary material at http://www.rcjournal.com).

Discussion

This is the first study, to our knowledge, that describes long-term variability in important CPET variables, $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_{E,}$ at several work rates with a cycle ergometer using a biocontrol protocol as suggested in the ATS/ACCP statement on CPET.[5] Because the data satisfied criteria for being normally distributed, we suggest the use of z-scores to describe biocontrol variability. A value greater than $\pm$ 1.96 z-scores signals an instrument may not be providing precise results and should prompt further investigation.

There is relatively little consensus about the correct or best way to perform biocontrol testing, although it is an essential component for monitoring the precision of testing instruments. The 2003 ATS/ACCP statement on CPET remains a definitive document on this topic and notes, "Subsequent steady state values for $\dot{V}_E$, $\dot{V}_{O_2}$, and $\dot{V}_{CO_2}$ are then compared with the database and values outside the 95% CI for that individual should engender a thorough systemwide reassessment."[5] This recommendation appears to come from a single study utilizing 35 biocontrol tests performed in one person over a 2.5-y period on a treadmill.[4] The authors concluded that variations in $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and heart rate should be $< 5\%$, whereas the variability in $\dot{V}_E$ should be $< 7\%$.

The ATS Pulmonary Function Laboratory Management and Procedure Manual does comment on CPET quality control (chapter 19) with little mention of evidence base other than to reference the 2010 Clinician's Guide to Cardiopulmonary Exercise Testing in Adults, a scientific statement piece from the American Heart Association.[9,11] In the latter, the authors

Table 2. Descriptive Statistics of the 4 Subjects Across the 4 Work Rates for $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$

$\dot{V}_{O_2}$ Measures	Subject			
	1 (no. = 39)	2 (no. = 22)	3 (no. = 28)	4 (no. = 16)
25 W				
Mean (SD)	0.75 (0.06)	0.51 (0.03)	0.68 (0.05)	0.61 (0.05)
Median (min–max)	0.74 (0.63–0.86)	0.51 (0.45–0.55)	0.67 (0.60–0.80)	0.63 (0.50–0.69)
CoV, %	7.5	5.47	7.29	7.82
50 W				
Mean (SD)	0.96 (0.07)	0.75 (0.04)	0.92 (0.07)	0.88 (0.06)
Median (min–max)	0.97 (0.83–1.08)	0.75 (0.65–0.82)	0.94 (0.80–1.03)	0.88 (0.77–0.98)
CoV, %	6.82	5.07	7.4	6.51
75 W				
Mean (SD)	1.23 (0.08)	1.04 (0.06)	1.20 (0.10)	1.15 (0.08)
Median (min–max)	1.21 (1.05–1.39)	1.02 (0.92–1.15)	1.21 (1.05–1.38)	1.15 (1.04–1.30)
CoV, %	6.74	5.6	8.11	6.95
100 W				
Mean (SD)	1.51 (0.09)	1.37 (0.07)	1.49 (0.11)	†
Median (min–max)	1.51 (1.33–1.73)	1.36 (1.27–1.49)	1.50 (1.29–1.67)	
CoV, %	6.19	4.76	7.23	
$\dot{V}_{CO_2}$ Measures	Subject			
	1 (no. = 39)	2 (no. = 22)	3 (no. = 28)	4 (no. = 16)
25 W				
Mean (SD)	0.60 (0.05)	0.42 (0.03)	0.58 (0.04)	0.55 (0.05)
Median (min–max)	0.61 (0.49–0.70)	0.42 (0.36–0.48)	0.57 (0.51–0.66)	0.55 (0.44–0.65)
CoV, %	8.08	7.17	7.34	9.24
50 W				
Mean (SD)	0.81 (0.06)	0.65 (0.04)	0.82 (0.05)	0.82 (0.05)
Median (min–max)	0.80 (0.69–0.94)	0.65 (0.59–0.73)	0.83 (0.69–0.90)	0.82 (0.75–0.95)
CoV, %	7.53	5.72	6.18	5.87
75 W				
Mean (SD)	1.07 (0.07)	0.94 (0.05)	1.08 (0.07)	1.10 (0.07)
Median (min–max)	1.06 (0.92–1.19)	0.94 (0.86–1.03)	1.08 (0.97–1.22)	1.10 (0.96–1.26)
CoV, %	6.21	5.44	6.79	6.32
100 W				
Mean (SD)	1.33 (0.08)	1.29 (0.07)	1.35 (0.07)	†
Median (min–max)	1.32 (1.21–1.47)	1.28 (1.17–1.44)	1.36 (1.23–1.47)	
CoV, %	5.65	5	4.91	
$\dot{V}_E$ Measures	Subject			
	1 (no. = 30)	2 (no. = 16)	3 (no. = 20)	4 (no. = 14)
25 W				
Mean (SD)	23.84 (1.58)	15.15 (0.85)	19.68 (1.09)	19.89 (2.19)
Median (min–max)	23.90 (20.40–26.60)	15.35 (13.50–16.30)	19.75 (17.40–21.70)	19.45 (15.20–23.80)
CoV, %	6.63	5.63	5.53	10.99
50 W				
Mean (SD)	30.03 (1.79)	21.33 (1.12)	25.86 (1.39)	27.44 (2.07)
Median (min–max)	29.90 (26.70–34.40)	21.45 (19.20–23.30)	26.10 (22.30–28.40)	27 (24.50–31.40)
CoV, %	5.85	5.27	5.36	7.54
75 W				
Mean (SD)	37.77 (2.19)	30.26 (2.08)	32.45 (1.49)	35.01 (3.10)
Median (min–max)	37.55 (33.20–43.40)	29.60 (27.50–34.80)	32.50 (29.80–35.30)	35.20 (29.10–41.70)
CoV, %	5.81	6.86	4.58	8.84
100 W				
Mean (SD)	45.73 (2.63)	42.44 (2.51)	39.30 (1.74)	†
Median (min–max)	45.35 (40.00–53.30)	41.55 (37.90–47.00)	39.10 (36.50–42.80)	
CoV, %	5.74	5.92	4.42	

Values are in L/min unless otherwise specified.
† These data were not included, as they were above the anaerobic threshold.
CoV = coefficient of variation.

Table 3. Subject #1 Values and Percentage Change From the Mean in $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$ at a Given Work Rate

Measure	Mean (L/min)	SD (L/min)	Original Scale Values at Z ± 1.96* (L/min)	Percent Change From Mean†
$\dot{V}_{O_2}$				
25 W	0.75	0.06	0.64 and 0.86	14.7
50 W	0.96	0.07	0.83 and 1.09	13.4
75 W	1.23	0.08	1.06 and 1.38	13.2
100 W	1.51	0.09	1.33 and 1.70	12.1
$\dot{V}_{CO_2}$				
25 W	0.60	0.05	0.51 and 0.70	15.9
50 W	0.81	0.06	0.69 and 0.93	14.8
75 W	1.07	0.07	0.94 and 1.20	12.1
100 W	1.33	0.08	1.18 and 1.48	11.1
$\dot{V}_E$				
25 W	23.84	1.58	20.74 and 26.93	13
50 W	30.03	1.79	26.53 and 33.53	11.7
75 W	37.77	2.19	33.47 and 42.07	11.4
100 W	45.73	2.63	40.59 and 50.88	11.3

* Calculated as (± 1.96 * SD) + mean.
† Calculated as ((raw score at Z ± 1.96 – mean)/mean)*100.

note limits of variation of ± 5% for $\dot{V}_{O_2}$, ± 6% for $\dot{V}_{CO_2}$, and ± 5.5% for $\dot{V}_E$. Elsewhere in this manual (chapter 5, quality control), Westgard's rules are discussed as originally described in 1977.[11,12] These rules utilizing "warning" and "out of control" conditions are not straightforward to implement in the case of CPET where there are several parameters being evaluated at once across different workloads compared to spirometry or diffusing capacity biocontrols. Otherwise, there is little discussion of, or reference to, Westgard's rules when it comes to quality control in the clinical pulmonary laboratory in general but also for CPET specifically as this is not mentioned in either of the documents by the ATS/ACCP or American Heart Association.[5,9] The recently published Association for Respiratory Technology and Physiology Statement on Cardiopulmonary Exercise Testing 2021 does attempt to address some of these issues, and whereas they recommend use of Westgard's rules, they acknowledge that these may not inform accurately on errors due to "subject variability."[13]

A more contemporary study by Porszasz et al[3] described quality control in the context of multi-center clinical trials utilizing treadmills where within-lab CoV at 20 W and 70 W, respectively, was 8.5% and 5.8% for $\dot{V}_{O_2}$, 9.2% and 7.2% for $\dot{V}_{CO_2}$, and 8.3% and 6.3% for $\dot{V}_E$. Prior data from our laboratory assessed short-term variability of CPET parameters during an incremental ramp test on a cycle ergometer in healthy subjects and also reported similar variability with a CoV of 4.9% in peak $\dot{V}_{O_2}$, 10.4% in $\dot{V}_{O_2}$ at anerobic threshold, 7.4% in peak $\dot{V}_E$, and 11.0% in $\dot{V}_E$ at anaerobic threshold.[14] These studies provide similar estimates of variability. Our findings suggest that the use of ± 5% for $\dot{V}_{O_2}$, ± 6% for $\dot{V}_{CO_2}$, and ± 5.5% for $\dot{V}_E$ may be too restrictive, suggesting a problem with an instrument, when in fact there is normal variability.[3,5,9]

Performing biologic controls is a time-consuming process, and the question of how often this testing should be performed still lacks an evidence base but may depend on how frequently tests are performed in the lab. Generally, it is recommended that biocontrol testing be performed 1–2 times per month.[3] The optimum, or sufficient, number of workloads tested during biocontrol testing is also not known. However, we provide a practical and feasible approach to performing CPET biocontrol testing in the clinical pulmonary or exercise laboratory. Because the number of scores available for computing the sample mean and SD will be small in the early stages of testing, these values are prone to sampling error; consequently, z-scores from very small samples may be unreliable estimates of machine function. However, as more data from a control subject are added, error is minimized and confidence in the z-scores to accurately represent machine status increases.

As mentioned, we found that the first 8 tests scores were relatively stable for each subject (ie, there was no profound visible variability in CPET z-scores). Therefore, we suggest that once stability has been achieved in 8 tests the mean and SD for those tests can be used to calculate z-scores moving forward. If there is extreme variability in the first 8 tests, testing should be halted, the machine and/or individual evaluated, and new testing started once the source of variability has been identified and removed. A spreadsheet for entering test scores and computing z-scores has been created and can be found in the online supplemental material (see related supplementary material at http://www.rcjournal.com). Data can be entered into the spreadsheet with a z-score for $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$ then automatically reported. As biocontrol testing is performed past 8 tests, the mean and SD are updated. After a year, we suggest using the prior 12 months of data to calculate the mean and SD going forward.

Our study does have limitations. (1) All data were collected in a single-center utilizing a single instrument. It is possible that variability differs among instruments. However, there are multiple studies describing similar variability during repeated incremental ramp tests on different instruments.[14,15] We studied only 4 subjects, and they performed a variable number of tests ranging from 16–39. However, our subjects showed similar variability, and we found that the SD remained stable after the initial 8 tests. Diet and exercise for each subject was not strictly monitored, though consistency was endorsed. Additionally, the CoV for parameters reported is very similar to prior published data as noted.[3,14] (2) It is important to note that we did not evaluate the accuracy of $\dot{V}_{O_2}$ (or $\dot{V}_{CO_2}$) but rather

the precision of measurements while utilizing manufacturer recommended quality assurance measures. (3) We averaged the last 2 min of data recorded at each work rate. This may not actually be the segment best representative of $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$ due to transient changes that can occur, and it may be preferable to pick the most representative data segment after 3 min have passed. We opted for this approach to facilitate a practical method to implement in busy clinical labs. (4) Finally, whereas it has been suggested biocontrol testing be performed at levels below an individual's anaerobic threshold, this may not provide complete information on an instrument as data may be more variable at higher $\dot{V}_E$ (ie, above anaerobic threshold). However, the CoV for $\dot{V}_E$ reported in our study is similar to the CoV in peak $\dot{V}_E$ performed during an incremental ramp protocol testing in healthy subjects.[14]

Conclusions

We report long-term variability of steady-state $\dot{V}_{O_2}$, $\dot{V}_{CO_2}$, and $\dot{V}_E$ obtained during biocontrol testing. Utilizing a threshold of 1.96 z-scores allows one to readily determine if a biocontrol result is abnormal and may warrant instrument investigation. Performing biocontrol testing can be time consuming for an active clinical pulmonary or exercise laboratory. To alleviate the burden of testing, we provide a practical, data-driven approach that can be added to a laboratory quality control program.

REFERENCES

1. Albouaini K, Egred M, Alahmar A, Wright DJ. Cardiopulmonary exercise testing and its application. Heart 2007;93(10):1285-1292.
2. Wasserman K. Principles of exercise testing and interpretation: including pathophysiology and clinical applications. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2012.
3. Porszasz J, Blonshine S, Cao R, Paden HA, Casaburi R, Rossiter HB. Biological quality control for cardiopulmonary exercise testing in multicenter clinical trials. BMC Pulm Med 2016;16(1):13.
4. Revill SM, Morgan MD. Biological quality control for exercise testing. Thorax 2000;55(1):63-66.
5. ATS/ACCP Statement on cardiopulmonary exercise testing. Am J Respir Crit Care Med 2003;167(2):211-277.
6. Crapo RO, Morris AH, Gardner RM. Reference values for pulmonary tissue volume, membrane diffusing capacity, and pulmonary capillary blood volume. Bull Eur Physiopathol Respir 1982;18 (6):893-899.
7. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999;159(1):179-187.
8. Hansen JE, Sue DY, Wasserman K. Predicted values for clinical exercise testing. Am Rev Respir Dis 1984;129(2 Pt 2):S49-S55.
9. Balady GJ, Arena R, Sietsema K, Myers J, Coke L, Fletcher GF, et al; Interdisciplinary Council on Quality of Care and Outcomes Research. Clinician's guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. Circulation 2010;122(2):191-225.
10. Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests. Journal of Statistical Modeling and Analytics 2011;2(1):21-33.
11. Wanger J, Mottram C. ATS pulmonary function laboratory management and procedure manual. New York: American Thoracic Society; 2016.
12. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier CH. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. Clin Chem 1977;23(10):1857-1867.
13. Pritchard A, Burns P, Correia J, Jamieson P, Moxon P, Purvis J, et al. ARTP statement on cardiopulmonary exercise testing. BMJ Open Respir Res 2021;2021(1):8.
14. DeCato TW, Bradley SM, Wilson EL, Hegewald MJ. Repeatability and meaningful change of CPET parameters in healthy subjects. Med Sci Sports Exerc 2018;50(3):589-595.
15. Skinner JS, Wilmore KM, Jaskolska A, Jaskolski A, Daw EW, Rice T, et al. Reproducibility of maximal exercise test data in the HERITAGE family study. Med Sci Sports Exerc 1999;31(11):1623-1628.