The Ventilatory Anaerobic Threshold in Heart Failure: A Multicenter Evaluation of Reliability

JONATHAN MYERS, PhD,[1] ROCHELLE L. GOLDSMITH, PhD,[2] STEVEN J. KETEYIAN, PhD,[3] CLINTON A. BRAWNER, MS,[3] DEIRDRE A. BRAZIL, BS,[2] HEATHER ALDRED, PhD,[3] JONATHAN K. EHRMAN, PhD,[3] AND DANIEL BURKHOFF, MD[2]

Palo Alto, California; New York, New York; Detroit, Michigan

ABSTRACT

Background: The ventilatory threshold (VT) is usually determined by visual assessment of the point where the rate of elimination of carbon dioxide (VCO_2) increases nonlinearly with respect to oxygen uptake (VO_2) (the V-Slope method). We quantified the reliability of VT determination using data from a multicenter study in patients with heart failure.

Methods and Results: The Fix-Heart Failure-5 study of cardiac contractility modulation enrolled 428 patients from 50 centers in the United States. Cardiopulmonary exercise tests were performed at baseline and 12, 24, and 50 weeks after randomization, which provided 1679 tests. The VT was determined from each test in a core laboratory by 2 independent readers. VT could not be determined for 276 tests (16.4% indeterminate). Inter-observer variability (quantified by the 95% limits of agreement, LoA, expressed as a percent of the mean value) was 20.2% between the 2 readers, with a coefficient of variation (CV) of 7.3%. Intra-observer variability was assessed by resubmitting (blinded) 179 tests to the same readers; the LoA was 24.7% for reader 1 and 16.9% for reader 2, with CVs of 6.1 and 8.9%, respectively. Ninety-one tests were submitted to 2 additional readers at a second core lab. Inter-observer variability in the second lab was 26.7% with a CV of 9.6%. Inter-laboratory variability was 21.4%, with a CV of 7.7%.

Conclusions: Inter-observer, intra-observer, and inter-site variation in determining the VT should be considered when using the VT as an end point in clinical trials of heart failure. (*J Cardiac Fail 2010;16:76–83*)

Key Words: Cardiopulmonary exercise test, heart failure, anaerobic threshold, peak VO_2.

In the 1920s, Nobel laureate A.V. Hill identified a link between blood lactate, muscle fatigue, and respiratory gas exchange during exercise.[1] The phrase *anaerobic threshold* was applied by Wasserman and colleagues in 1964[2] to denote the point during progressive exercise in which lactate accumulation caused a nonlinear increase in ventilation. The increase in blood lactate concentration during exercise is thought to cause a nonlinear increase in ventilation as a result of bicarbonate buffering of excess hydrogen ions from lactate in the blood and consequent production of carbon dioxide. The resulting hyperventilatory response has been commonly termed the *ventilatory threshold*, or VT. Although the mechanism underlying the VT has been disputed,[3,4] this parameter has been widely used over many years as an index of cardiorespiratory performance in patients with cardiovascular or pulmonary disease.[3,5] The conventional interpretation is that an increased oxygen uptake at the VT after an intervention implies that an increase in oxygen supply to the working muscle has occurred. Thus, this parameter has been applied in many clinical circumstances, including use as an index of the functional state of patients with chronic heart failure (CHF), the development of individualized exercise prescriptions, a measure of the response to exercise training, risk stratification, and to assess the effectiveness of various pharmacologic or device interventions to quantify cardiac performance.[3,5,6–10]

In addition to the mechanism underlying the VT, controversy has arisen over the years regarding whether the VT represents a true ''threshold'' versus a pattern that is mathematically continuous[11–13]; the most appropriate method for detection[14–16]; or the reliability and reproducibility of choosing the VT.[14–18] Although modern metabolic systems assign a VT value automatically at the end of each test,

From the [1]VA Palo Alto Health Care System/Stanford University, Palo Alto CA; [2]Columbia University, New York Presbyterian Medical Center, New York, NY and [3]Henry Ford Hospital, Detroit, MI.

Manuscript received May 25, 2009; revised manuscript received July 7, 2009; revised manuscript accepted July 13, 2009.

The authors have no conflicts of interest.

Reprint requests: Jonathan Myers, PhD, VA Palo Alto Health Care System, Cardiology 111C, 3801 Miranda Ave, Palo Alto CA 94304. Tel: (650) 493-5000 x64661; Fax: (650) 852-3473; E-mail: drj993@aol.com

1071-9164/$ - see front matter

doi:10.1016/j.cardfail.2009.08.009

these systems use numerous mathematical algorithms to define the VT; thus, the point chosen can vary significantly for the same patient.[18] Guidelines on cardiopulmonary exercise testing have therefore suggested that the VT be confirmed visually using one of several methods.[5] Because there is increased interest in using the VT as a primary or secondary end point in multicenter trials among patients with CHF, it is important to know how reliable and reproducible this index can be determined in such a setting. Typically, there is a tacit assumption that the VT, whether determined objectively by an algorithm from the metabolic system, or visually, is reliable, unbiased, and valid, but the methods employed to detect the VT are often not specified. If the VT is to be used to determine the efficacy of a given therapy, valid and unbiased interpretation of the VT is critical. Although there have been a number of efforts to assess inter-reader reproducibility within a given center,[14,16,17] most of these studies involved a relatively small number of subjects that were free of cardiovascular disease. Clearly, much less is known regarding the reliability of the VT when used to evaluate patients enrolled in a clinical trial that involves numerous participating centers. Therefore, few data exist that can be used to assist in trial design and to develop sample size estimates when the VT is used as an efficacy parameter in patients with CHF.

The purpose of the current study was to assess intra- and inter-reader agreement of VT readings within and between different core laboratories. Data from a recently completed, relatively large multicenter CHF study formed the basis for the analysis. The objective was to provide guidance in terms of sample size estimates when using the VT as an end point in a multicenter study involving patients with CHF.

Methods

Subjects

Cardiopulmonary exercise test (CPX) data were obtained from 428 subjects with CHF participating in a multicenter trial of the effects of cardiac contractility modulation. The details of the protocol, device implantation procedure, primary and secondary end points, and statistical analysis plan have been detailed previously.[19] In brief, the study included subjects ≥18 years old with ejection fraction ≤35%, with New York Heart Association Class III or IV symptoms despite medical treatment with angiotensin-converting enzyme inhibitor and/or angiotensin receptor blocker and β-blockers for at least 3 months. Subjects were required to have a baseline peak oxygen uptake (VO_2) ≥9 mL $O_2 \cdot kg \cdot min$ who were in normal sinus rhythm and not indicated for a cardiac resynchronization therapy device (ie, QRS duration <130 ms). Unless there were extenuating circumstances, subjects were required to have an implantable cardiac defibrillator. Subjects were excluded if they were hospitalized within 30 days of enrollment, were inotropic dependent, had >8900 premature ventricular contractions per 24 hours on a baseline Holter monitor, had permanent atrial fibrillation, had a myocardial infarction within 90 days, had percutaneous coronary intervention within 30 days, or had coronary artery bypass surgery within 90 days of enrollment. CPX data were obtained at baseline and weeks 12, 24, and 50 after device implant (±2 weeks). The CPX test data were sent to a central core laboratory (Core Lab) for analysis. The Core Lab was directed and managed at Columbia University (New York City) and involved 2 independent, experienced CPX test readers (reader 1, reader 2) located at Henry Ford Hospital (Detroit). All subjects underwent screening, evaluation for study eligibility and CPX testing in accordance with a standardized protocol (detailed in the following section).

Exercise Testing

Symptom limited maximal exercise tests were performed on a treadmill using a Continuous Workload Modified Naughton Protocol.[20] All subjects were requested to abstain from eating or smoking at least 3 hours before the test. Ventilatory oxygen uptake was measured using a metabolic system that was calibrated prior to each test. A 12-lead electrocardiogram was monitored continuously and recorded every minute. Blood pressure was recorded manually every 2 minutes. All subjects were encouraged to provide a maximal effort, and the Borg 0 to 10 perceived exertion scale was used to quantify effort.[21]

The raw metabolic data were sent to the Core Lab for analysis. The Core Lab determined all metabolic measurements, including oxygen uptake, carbon dioxide production, minute ventilation, and respiratory exchange ratio. After this basic analysis, tests were evaluated by the 2 independent readers (readers 1 and 2) for determination of VT. The 2 readers were blinded to patient identification, study treatment, follow-up time, and the results of the other reader. The final VT captured in the database for the cardiac contractility modulation trial was the average of these 2 readers, if the VT from these 2 readings differed by less than 10% or 150mL. However, if the difference between the readers was greater than these specified limits, then the data were evaluated by a third blinded reader and the average for the 2 readers that were most similar was captured in the database. If the VT from 2 of the 3 was not within 10% or 150 mL, then the VT was considered indeterminate.

A diagram of the data flow and evaluations by the different readers is shown in Fig. 1. For the study as a whole, 1679 CPX tests were initially analyzed for all intervals of the study. Inter-reader reliability was evaluated by comparing the VT values determined by the 2 initial readers. Intra-reader reliability was tested on a sample of 179 tests; these tests were deidentified and provided a second time to the same 2 readers. The first and second reads were compared for each reader. To test inter-core laboratory reliability, 91 of the CPX tests were sent to a second Core Lab, located at the Veterans Administration Palo Alto Health Care System. Two additional independent test readers (reader 3, reader 4) evaluated these same 91 tests. This provided another opportunity to compare inter-reader reliability (by comparing results of reader 3 with those of reader 4) as well as to evaluate inter-laboratory variability. For the latter purpose, each test was assigned a final value of VT from each core lab, which was defined as the average VT values of readers 1 and 2 for Core Lab 1, and the average of readers 3 and 4 for Core Lab 2. Tests were excluded from this evaluation if any of the readers declared the test indeterminate. The number of tests considered indeterminate and the concordance of indeterminate reads between readers was evaluated separately.

Readers were instructed to determine the VO_2 at the VT using the V-slope method,[22] which is based on a visual examination of the regression between rate of elimination of carbon dioxide

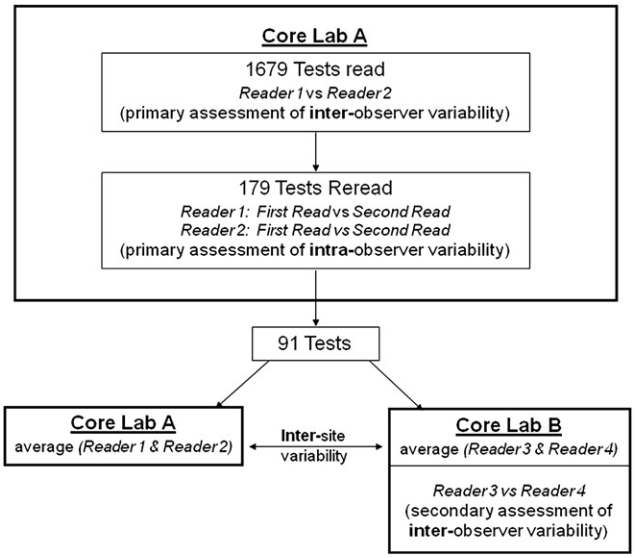

Fig. 1. Diagram of data flow and analyses.

(VCO_2) and VO_2 (Fig. 2A). In this method, a line of identity is derived from a plot of VCO_2 versus VO_2 (both in units of mL O_2/min). A best fit line parallel to the line of identity is then drawn through the VCO_2-VO_2 data points. The point at which the VCO_2 departs from the line (ie, the point at which VCO_2 begins increasing more rapidly than VO_2) is taken as the "V-Slope" VT. In the event where no inflection point could be identified using the V-slope method or if the inflection point was ambiguous, then the VO_2 at the point where there was a systematic increase in the ventilatory equivalent for VO_2 (VE/VO_2) without a concomitant increase in the ventilatory equivalent for VCO_2 (VE/VCO_2) was taken as the VT. However, if a VCO_2-VO_2 data point could not be identified with confidence using either of these methods, the reader declared that the VT was indeterminate. For example, if the gas exchange data were too noisy or oscillatory (Fig. 2C and 2D) or if the test was too submaximal and thus an inflection point could not be detected (Fig. 2E and 2F), the VT was considered indeterminate.

Statistical Analysis

Descriptive statistics are presented as mean±SD. Comparisons between observers were performed using 1-way analysis of variance; post-hoc testing was performed using the Bonferroni method. Reliability of VT readings was assessed by the typical error, limits of agreement expressed both in absolute values (LoA) and as a percentage of the mean VT (LoA%), and the coefficient of variation. NCSS software (Kayesville, UT) was used for all statistical analyses. We calculated the mean and standard deviation (SD_d) of the differences between each pair of VT reads being compared. The typical error (TE) was calculated as the SD_d divided by the square root of 2. The coefficient of variation of the TE (CV_{TE}) was calculated as the TE divided by the mean of both reads multiplied by 100 and expressed as a percentage. The 95% LoA were calculated as the SD_d times 1.96. The LoA was also expressed relative to the mean of both reads as a percent.

Results

Baseline demographic and clinical characteristics of the subjects are summarized in Table 1. These characteristics are typical of heart failure studies that are enrolling patients

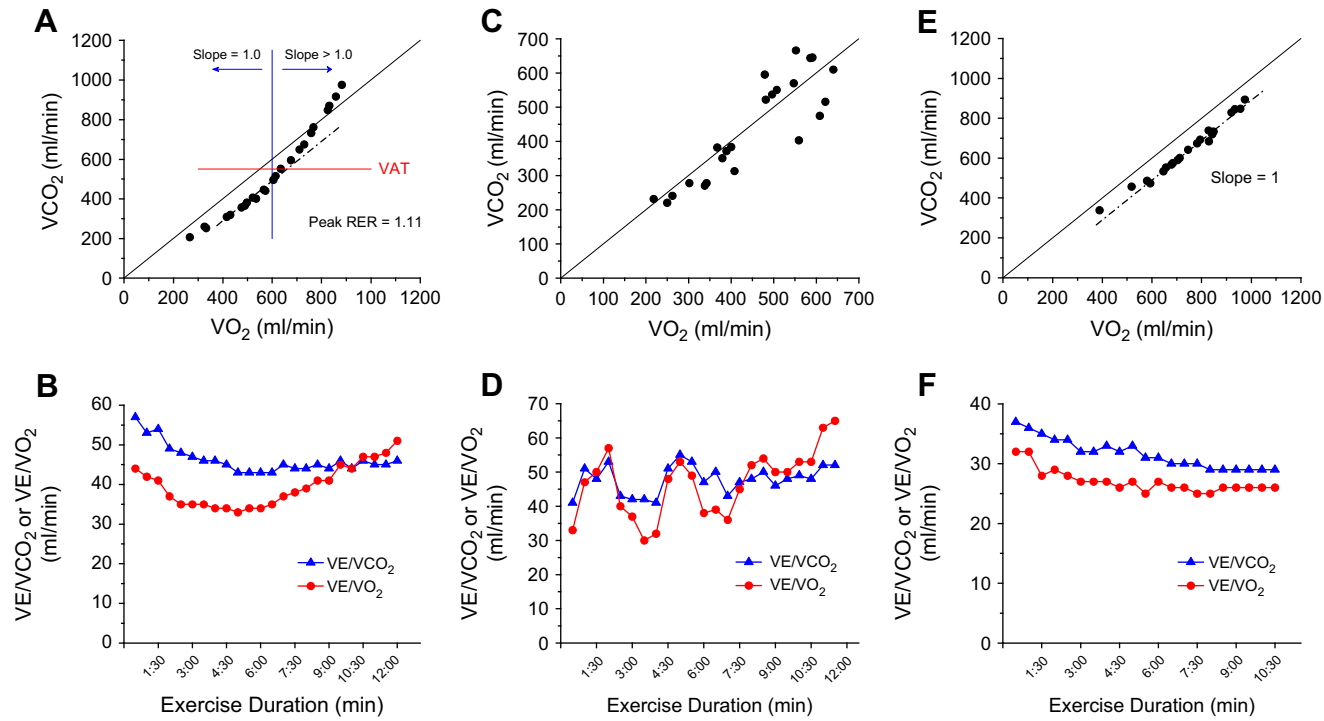

Fig. 2. Graphical depiction of the V-slope method for determining VO_2 at the ventilatory threshold along with the corresponding graphs of the ventilatory equivalents. (A, B) Ventilatory threshold (VT) was determined to be 635 mL/min by both readers. (C, D) VT indeterminate because of noisy data. (E, F) VT indeterminate owing to the submaximal nature of the test. Note: VCO_2 did not depart from line of identity and ventilatory equivalent (VE)/VO_2 did not increase.

Table 1. Demographic and Clinical Characteristics

Characteristic	n=428
Age	58±12
Male	309 (72%)
Weight (kg)	92.3±22.8
Height (cm)	172.9±9.7
Body mass index (kg/m^2)	30.7±6.8
Congestive heart failure etiology	
Ischemic	280 (65%)
Other	148 (35%)
History of diabetes	193 (45%)
Resting heart rate	74±13
Systolic blood pressure (mm Hg)	116±19
QRS duration (ms)	101.7±14.4
Left ventricular ejection fraction (%)	26.0±6.6
Left ventricular end-diastolic diameter (mm)	62.8±8.9
Peak VO_2 (mL·kg·min)	14.8±3.0
Peak respiratory exchange ratio	1.13±0.10
6-minute hall walk (m)	323.8±88.8
Minnesota Living with Heart Failure Questionnaire	59.0 22.8
New York Heart Association	
Class II	1 (0.2%)
Class III	379 (88.6%)
Class IV	48 (11.2%)
Treatments	
Angiotensin-converting enzyme inhibitor or angiotensin receptor blocker	389 (90.9%)
β-Blocker	400 (93.5%)
Aldosterone inhibitor	197 (46.0%)
Implantable cardiac defibrillator	408 (95.3%)

Values are either mean±SD or number of patients (percent of patients).

having ejection fraction ≤35% with New York Heart Association Class III or IV symptomatic heart failure despite optimal medical therapy. Two notable enrollment biases were that the study enrolled patients with QRS durations less than 130 ms and excluded patients with atrial fibrillation.

Table 2 summarizes all of the indexes of inter- and intra-observer and inter-Core Lab agreements for the determination of the VT. For the main database that included 1679 tests, readers 1 and 2 could both assign a VT value for 1403 (83.6%) of the tests. However, in 276 (16.4%) of the tests, either reader 1 or reader 2 declared the test indeterminate; 126 of these tests were declared indeterminate by both readers resulting in a 45.7% concordance in declaring a test indeterminate. Overall, however, reader 1 declared 186 (11%) indeterminate compared with 216 (12.9%) tests by reader 2.

For tests having VT reads from both readers, VT averaged 1015±310 mL/min for reader 1 and 1028±312 mL/min for reader 2, a difference which was statistically significant (paired *t*-test, $P < .0001$). The correlation and Bland and Altman plots for the main database are shown in Fig. 3. Although the readings were highly correlated ($r=0.94$) and the regression line was close to the line of identity ($y=66+0.95x$) the Bland and Altman analysis revealed that the 95% LoA were ±207 mL/min around a statistically significant bias of −13 mL/min ($P < .0001$), which amounts to an LoA% of 20.2% and a CV_{TE} of 7.3%. The VT values from the 2 readers were within 5% of each other for 709 (50.5%) of the tests and within 10% of each other for 1045 (74.5%) of the tests.

From among the subset of 179 tests that were reread at Core Lab 1, reader 1 could assign a value for VT in 165 (92.2%) of the tests and reader 2 could assign a value for VT in 156 (87.2%) of the tests; importantly, for both readers, all tests designated as indeterminate on the first read were also designated as indeterminate on the second read. The correlations and Bland and Altman graphs comparing first and second reads for these 2 readers are shown in Fig. 4. Although first and second reads were highly correlated ($r=0.93$ and 0.94 for readers 1 and 2, respectively), *intra*observer variability as quantified by the LoA% was 24.6% for reader 1 and 16.2% for reader 2, with CV_{TE} for readers 1 and 2 of 8.9% and 5.8%, respectively.

From among the 91 tests that were sent to the second Core Lab, readers 3 and 4 could assign VT values for 86 (94.5%) and 85 (93.4%) of the tests, respectively (though not the same tests in all cases). The LoA% for VT values between these two readers was 26.7%, and the CV_{TE} was 9.6%.

The final VT value for each core lab was defined as the average value of the 2 readers. The correlations and Bland and Altman graphs for Core Labs 1 and 2 are shown in Fig. 5. Although the correlation was high, the 95% LoA (∼200 mL/min), LoA% (21.4%), and CV_{TE} (7.7%) were similar to the comparisons discussed previously.

Table 2. Inter- and Intra-Observer and Intra-Site Agreement for Determination of the VO_2 at Ventilatory Threshold (mL/min)

	n	% Readable[§]	Mean of Readers 1 and 2	Difference	Typical Error	95% LOA	LOA as % of Mean	CV
Inter-observer agreemnet								
Reader #1 vs. Reader #2	1679	83.6%	1022±306	13±105*	75	± 207	20.2%	7.3%
Reader #3 vs. Reader #4	91	91.2%	926±257	23±126	89	± 247	26.7%	9.6%
Intra-observer agreement								
Reader #1	179	92.2%	951±284	95±106*	84	± 234	24.6%	8.9%
Reader #2	179	87.2%	1012±316	49±68	59	± 164	16.2%	5.8%
Final Determination[†]	179	91.6%	987±309	47±58	48	± 133	13.5%	4.9%
Inter-site agreement								
Site 1 vs. Site 2	91	85.7%	935±257	18±102	72	± 200	21.4%	7.7%

LOA, limits of agreement; CV, coefficient of variation.
Mean ± standard deviation.
*$P < .001$.
[§]Ventilatory threshold determined during both reads.
[†]The final determination was based on a 3 reviewer system as described in the Methods.

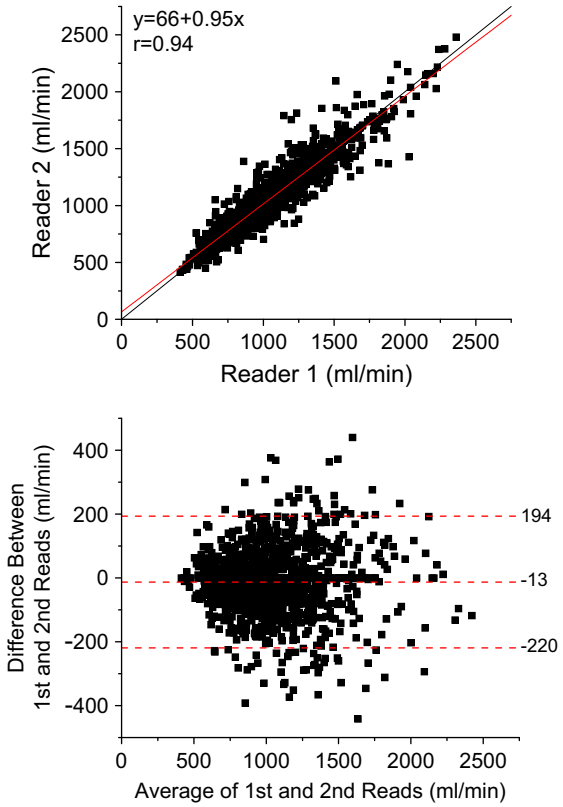

Fig. 3. Correlation coefficients and Bland and Altman plots between readers 1 and 2 from the main database.

The final VT (3-reviewer system) captured into the database for the trial evaluating the cardiac contractility modulation device yielded an LoA% of 13.5% and a CV_{TE} of 4.9%. This 3-reviewer system identified the VT in 95.6% of subjects.

Discussion

The VT has been widely used over the last several decades to assess cardiopulmonary performance in both athletes and patients with cardiovascular disease. Because the VT is thought to reflect the accumulation of lactate in the blood, the point at which it is reached during exercise has numerous physiological consequences, including the onset of metabolic acidosis, impaired glycolysis, reduced exercise tolerance, and altered oxygen kinetics. Thus, the VT is an important parameter to be documented when conducting exercise testing in patients with cardiovascular disease. The VT has been of particular interest in patients with CHF because the early accumulation of lactate in the blood leads to hyperventilation, which impairs the ability to sustain activity. Indeed, dyspnea on exertion is one of the hallmarks of CHF, and when dyspnea is the reason for stopping exercise, the risk for adverse events doubles.[23] One advantage of the VT that is commonly cited is the fact that, unlike peak VO_2, it is independent of motivation. Increasingly, the VT has been applied as a primary or secondary end point in clinical trials in patients with CHF, including exercise training, pharmacologic intervention, ventricular resynchronization, and ventricular reduction surgery.[24–27]

Although the VT has been widely applied in patients with CHF, little attention has been paid to how it is determined and the reliability of its determination between different observers. Although modern metabolic systems usually interpret the VT automatically, a wide variety of algorithms are used; thus, guidelines on exercise testing suggest that this point be overread visually by at least 1 experienced observer.[5] Although this in theory provides a measure of quality control, it also makes the VT determination subjective. In any measurement that is subjective, *inter*-observer agreement is critical if a clinician is to have confidence that a measurement is valid and consistent when laboratories collect the same or similar information. In addition, acceptable reliability of a measure (*intra*-observer agreement) must be confirmed before it can be considered to have valid clinical application. Although high reliability does not ensure accuracy, it quantifies the degree to which a measurement produces the same result each time it is performed. In the current study, inter-observer reliability and inter-laboratory agreement of the VT determined by highly experienced readers was assessed using several closely related and standardized indices: the typical error, limits of agreement, LoA%, and coefficient of variation. The typical error is an expression of the random variation across measurements for a given subject; it is defined simply as the standard deviation of the observations for a given patient. The coefficient of variation is a measure of the relative variation, or variation relative to the size of the mean for the observers expressed as a percentage. The limits of agreement are, in effect, the 95% confidence limits within which the VT readings fall. The LoA% will parallel the coefficient of variation in that it represents the limits of agreement expressed as a percentage of the mean VT.

We observed that although mean differences between readers were reasonably small (eg, 13 mL/min for the main data base of 1493 subjects) (Table 2) and the inter- and intra-observer readings were highly correlated ($r > 0.90$, Fig. 2, 3, 4), the extent of the variation within and between readers differed based on the parameter used to express it. Importantly, the within- and between-observer typical errors were in the range of 70 to 90 mL, resulting in limits of agreement in the order of 200 mL, and coefficients of variation in the order of 5% to 10%. In practical terms, this suggests that a given VT is likely to vary by this magnitude if the same exercise test data are evaluated either by different reviewers or by the same reviewer on a different day. The LoA% ranged from 17% to 27%, suggesting the potential for an unacceptably high variability in about 1 in 5 patients. Thus, significant error in interpreting the VT in particular patients will occur even with experienced reviewers. It should be noted, however, that the LoA for the final VT (3-reviewer system) for the study was 121 mL and associated with an LoA% of 13%, reflecting an unacceptably high variability in roughly 1 in 8 subjects. The improved variability with the 3-reviewer system suggests

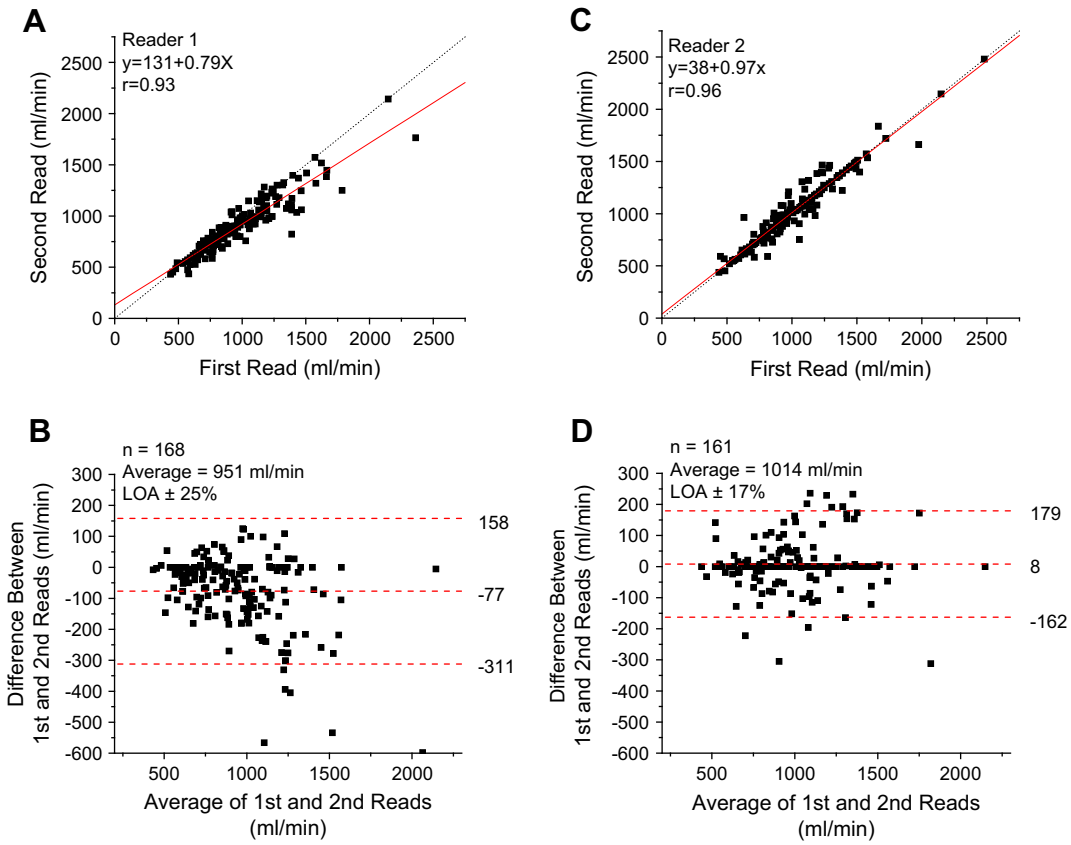

Fig. 4. Correlation coefficients and Bland and Altman graphs comparing ventilatory threshold (VT) determinations between readers 1 and 2.

that this approach should be considered if the VT is to be used in other clinical trials. These differences should also be considered in the context of the reported day-to-day *biological* variability of the VT, which has been reported to differ widely among previous studies.[28–30]

Many previous studies have reported on the reproducibility of the VT using subjective or objective criteria and demonstrated it to have acceptable reproducibility. These studies have generally reported correlation coefficients or intra-class correlation coefficients in the range of 0.70 to 0.90.[15,16,31,32] However, few of these studies have described intra- or inter-*observer* variability, and few studies have focused specifically on patients with CHF. In the current multicenter study, the coefficient of variation was 5% to 10%, quite similar to the 5% to 12% reported by Simonton et al[31] in their single-site study involving both patients with CHF and normal subjects. Sullivan et al[17] studied inter-observer variation of the VT in a group of 14 patients limited during exercise by angina and reported a coefficient of variation of 7%. Shimizu et al[16] reported a coefficient of variation between 3 observers of 17% in a mixed group of patients with CHF and coronary artery disease. Yeh et al[15] evaluated inter-observer variation among 4 VT readers in a group of normal subjects, and reported a coefficient of variation of 34%, which represented an 8% variation between reviewers when the VT was expressed as a percentage of peak VO_2. Similarly, Gladden et al[14] observed that mean differences between any 2 evaluators (of 9) ranged from 3% to 16% of peak VO_2. Clearly, these studies differ in part because of differences in how the VT was defined, the number of reviewers, and the definition of agreement, but together with the current results, they suggest that the degree of variation between reviewers must be considered when applying the VT to evaluate patients with cardiovascular disease.

Cardiopulmonary exercise test in which the VT is designated to be ''indeterminate'' have been one issue limiting its clinical application. We found that up to 16% of tests were designated as indeterminate by the observers. This is typical of previous studies, in which 5% to 20% of patients were excluded from analyses because the VT point was designated to be indeterminate.[16,17,33] Cohen-Solal et al[32] reported a higher indeterminate rate, ranging between 12% and 46% depending on the method used for assessing the VT. The ventilatory equivalent (VE/VO_2) method, one that is widely used and suggested to best reflect lactate accumulation,[34] had the highest indeterminate rate of 46%. In contrast, Shimizu et al[16] found that the VE/VO_2 method had the lowest indeterminate rate (6%, compared with 7% and 12% using the V-slope and partial pressure of end-tidal carbon dioxide methods, respectively). The degree of *concordance* (tests in which both reviewers agreed on whether a test was indeterminate) in the present study is also noteworthy. For the first reader from Core Lab #1, all tests designated as indeterminate on the first read were also

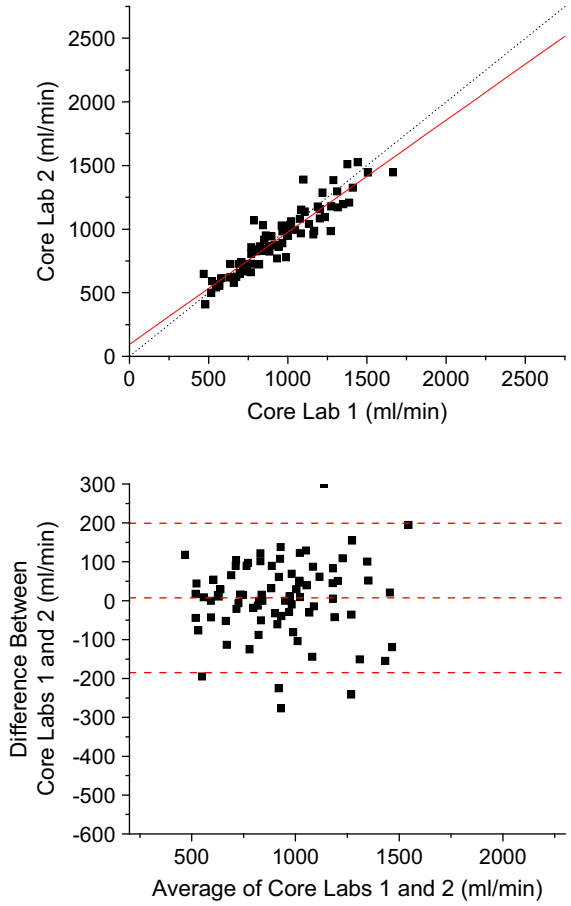

Fig. 5. Correlation coefficients and Bland and Altman graphs comparing readings between Core Labs 1 and 2.

designated as indeterminate on the second read; however, such was not the case for reader #2. Overall, the concordance rate was 45.7%; thus, a test designated as indeterminate by 1 observer was only designated as indeterminate by another reviewer or repeat reading half the time. Because 16.4% of tests overall were designated as indeterminate by 1 or more reviewers, the implication of these findings is that when estimating sample size for clinical trials and the VT is an end point, the indeterminate rate must be considered.

One solution to the issue of reliability of visual VT readings is to employ objective methods to determine this point. Metabolic systems commonly print ventilatory responses graphically and determine the VT automatically; this provides convenience and objectivity. However, automating the determination of the VT is also subject to error. First, most systems base the determination of the VT on a single regression-based computer algorithm; the complexities of the computational problems related to these algorithms are not likely to be appreciated by the user. Different metabolic systems also use different algorithms to define the VT, and comparisons between these algorithms have been shown to vary considerably when determining the VT.[18,35] Moreover, although these objective methods facilitate convenience, they do not ensure accuracy. In addition, a metabolic system may "force" the presence of a VT when experienced observers will judge the point to be indeterminate. Metabolic systems also differ in their ability to filter noise, and are dependent on both the definition used and how the data are sampled. Finally, studies that have compared visual methods with a computer algorithm, such as the V-slope, have shown poor agreement.[35]

Summary

Since the 1960s, the VT has been a popular, simply understood, and useful parameter to assess the functional status of patients with cardiovascular disease.[3,6,36] The concept that a breakpoint in ventilation identifies the onset of inadequate oxygen supply to the exercising muscle and can be determined easily and noninvasively has been attractive to clinicians for obvious reasons. However, the mechanisms responsible for blood lactate increase, its relation to the pattern of the ventilatory response to exercise, and the ability to reliably and reproducibly choose a "threshold" continue to be topics of debate. In the present study we found that inter-observer, intra-observer, and inter-site agreement of the VT ranged between 5% and 10%, with the LoA% as high as 25% of the mean VT. However, to minimize the coefficient of variation and LoA%, our data emphasize the importance of using a third reviewer to adjudicate differences between 2 reviewers determining the VT. These variations should be taken into account when considering VT as an end point in clinical trials, as well as for the assessment of functional status in routine clinical practice and for exercise training and prescription in patients with heart failure.

References

1. Hill AV, Lupton H. Muscular exercise, lactic acid, and the supply and utilization of oxygen. Q J Med 1923;16:135–71.
2. Wasserman K, McElroy MB. Detecting the threshold of anaerobic metabolism in cardiac patients during exercise. Am J Cardiol 1964;14: 844–52.
3. Myers J, Ashley E. Dangerous curves. A perspective on exercise, lactate, and the anaerobic threshold. Chest 1997;111:787–95.
4. Ohira Y, Tabata I. Muscle metabolism during exercise: anaerobic threshold does not exist. Ann Physiol Anthropol 1992;11:319–23.
5. Balady G, Arena R, Coke L, Fletcher G, Forman D, Franklin B, et al. A clinician's guide to cardiopulmonary exercise testing. A Scientific Statement from the American Heart Association Subcommittee on Exercise, Prevention, and Rehabilitation. Circulation. In press.
6. Svedahl K. MacIntosh BR. Anaerobic threshold: the concept and methods of measurement. Can J Appl Physiol 2003;28:299–323.
7. Wasserman K, Sun XG, Hansen JE. Effect of biventricular pacing on the exercise pathophysiology of heart failure. Chest 2007;132: 250–61.
8. Myers J, Gullestad L, Vagelos R, Do D, Bellin D, Ross H, et al. Clinical hemodynamic, and cardiopulmonary exercise test determinants of survival in patients referred for evaluation of heart failure. Ann Intern Med 1998;129:286–92.
9. Dougherty CM, Glenny R, Kudenchuk PJ. Aerobic exercise improves fitness and heart rate variability after an implantable cardioverter defibrillator. J Cardiopulm Rehabil Prev 2008;28:307–11.

10. Wonisch M, Hofmann P, Fruhwald FM, Kraxner W, Hodl R, Pokan R, et al. Influence of beta-blocker use on percentage of target heart rate exercise prescription. Eur J Cardiovasc Prev Rehabil 2003;10:296–301.
11. Hughson RL, Weisiger KH, Swanson GD. Blood lactate concentration increases as a continuous function during progressive exercise. J Appl Physiol 1987;62:1975–81.
12. Myers J, Walsh D, Buchanan N, McAuley P, Bowes E, Froelicher VF. Increase in blood lactate during ramp exercise: comparison of continuous versus threshold models. Med Sci Sports Exerc 1994;26:1413–9.
13. Dennis SC, Noakes TD, Bosch AN. Ventilation and blood lactate increase exponentially during incremental exercise. J Sports Sci 1992; 10:437–49.
14. Gladden LB, Yates JW, Stremel RW, Stamford BA. Gas exchange and lactate anaerobic thresholds: inter- and intra-evaluator agreement. J Appl Physiol 1985;58:2082–9.
15. Yeh MP, Gardner RM, Adams TD, Yanowitz FG, Crapo RO. "Anaerobic threshold": problems of determination and validation. J Appl Physiol 1983;55:1178–86.
16. Shimizu M, Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, et al. The ventilatory threshold: method, protocol, and evaluator agreement. Am Heart J 1991;122:509–16.
17. Sullivan M, Genter F, Savvides M, Roberts M, Myers J, Froelicher VF. The reproducibility of hemodynamic, electrocardiographic, and gas exchange data during treadmill exercise in patients with stable angina pectoris. Chest 1984;86:375–81.
18. Ekkekakis P, Lind E, Hall EE, Petruzzello SJ. Do regression-based computer algorithms for determining the ventilatory threshold agree? J Sports Sci 2008;26:967–76.
19. Abraham WT, Burkhoff D, Nademanee K, Carson P, Bourge R, Ellenbogen KA, et al. A randomized controlled trial to evaluate the safety and efficacy of cardiac contractility modulation in patients with systolic heart failure: rationale, design, and baseline patient characteristics. Am Heart J 2008;156:641–8.
20. Goldsmith RL, Trufant JW, Mueller GM, Arwady MA, Burkhoff DB. Optimizing exercise protocols for heart failure trials. J Amer Coll Cardiol 2004;44:169A.
21. Borg GAV. Borg's perceived exertion and pain scales. Champaign, IL: Human Kinetics; 1998.
22. Beaver W, Wasserman K, Whipp BJ. A new method for detecting anaerobic threshold by gas exchange. J Appl Physiol 1986;60:2020–7.
23. Chase P, Arena R, Myers J, Abella J, Peberdy MA, Guazai M, Kenjale A, et al. Prognostic usefulness of dyspnea versus fatigue as reason for exercise test termination in patients with heart failure. Am J Cardiol 2008;102:879–82.
24. Laveneiana P, O'Donnel DE, Ofir D, Agostoni P, Padletti L, Riccardi G, et al. Effect of biventricular pacing on ventilatory and perceptual responses to exercise in patients with stable chronic angina. J Appl Physiol 2009; Feb 26 (e-pub ahead of print).
25. Curnier D, Galinier M, Pathak A, Fourcade J, Bousquet M, Senard JM, et al. Rehabilitation of patients with congestive heart failure with or without beta blockade therapy. J Card Fail 2001;7:241–8.
26. Van Laethem C, Van De Veire N, De Backer G, Bihija S, Seghers T, et al. Response of the oxygen uptake efficiency slope to exercise training in patients with chronic heart failure. Eur J Heart Fail 2007;9: 625–9.
27. Frigiola A, Tsang V, Bull C, Coats L, Khambadkone S, Derrick G, et al. Biventricular response after pulmonary valve replacement for right ventricular outflow tract dysfunction: is age a predictor of outcome? Circulation 2008;118:S182–90.
28. Kothmann E, Danjoux G, Owen SJ, Turley AJ, Batterham AM. Reliability of the anaerobic threshold in cardiopulmonary exercise testing of patients with abdominal aortic aneurysms. Anaesthesia 2009;64: 9–13.
29. Garrard CS, Das R. Sources of error and variability in the determination of the anaerobic threshold in humans. Respiration 1987;51: 137–45.
30. Thin AG, Kelly GE, Daly L, McLoughlin P. Repeated measurement of the gas exchange threshold: relative size of measurement and biological variabilities. Comp Biol Med 2005;35:703–16.
31. Simonton CA, Higginbotham MB, Cobb FR. The ventilatory threshold: quantitative analysis of reproducibility and relation to arterial lactate concentration in normal subjects and in patients which chronic congestive heart failure. Am J Cardiol 1988;62:100–7.
32. Cohen-Solal A, Zannad F, Kayanakis JG, Aupetit JF, Kolsky H. Multicentre study of the determination of peak oxygen uptake and ventilatory threshold during bicycle exercise in chronic heart failure. Comparison of graphical methods, interobserver variability and influence of the exercise protocol. The VO2 French Study Group. Eur Heart J 1991;10:1055–63.
33. Bensimhon DR, Leifer ES, Ellis SJ, et al. for the HF-ACTION Trial Investigators. Reproducibility of peak oxygen uptake and other cardiopulmonary exercise testing parameters in patients with heart failure. Am J Cardiol 2008;102:712–7.
34. Caiozzo VJ, Davis JA, Ellis JF, Vandagriff R, Prietto CA, McMaster WD. A comparison of gas exchange indices used to detect the anaerobic threshold. J Appl Physiol 1982;53:1184–9.
35. Crescencio JC, Martins LEB, Murta LO, Antloga CM, Kozuki RT, Santos MDB, et al. Measurement of anaerobic threshold during dynamic exercise in healthy subjects: comparison among visual analysis and mathematical models. Comp Cardiol 2003;30:801–4.
36. Wasserman K, Beaver WL, Whipp BJ. Gas exchange theory and the lactic acidosis (anaerobic) threshold. Circulation 1990;81(II):14–30.