A comparison of methods for determining the ventilatory threshold: implications for surgical risk stratification

Comparaison des méthodes de détermination des seuils ventilatoires: implications pour la stratification du risque chirurgical

Baruch Vainshelboim, PhD · Shravan Rao, MD · Khin Chan, MD · Ricardo M. Lima, PhD · Euan A. Ashley, MD · Jonathan Myers, PhD

Received: 22 August 2016 / Revised: 16 January 2017 / Accepted: 13 March 2017 / Published online: 5 April 2017
© Canadian Anesthesiologists' Society 2017

Abstract

Purpose The ventilatory threshold (VT) is an objective physiological marker of the capacity of aerobic endurance that has good prognostic applications in preoperative settings. Nevertheless, determining the VT can be challenging due to physiological and methodological issues, especially in evaluating surgical risk. The purpose of the current study was to compare different methods of determining VT and to highlight the implications for assessing perioperative risk.

Methods Our study entailed analysis of 445 treadmill cardiopulmonary exercise tests from 140 presurgical candidates with an aortic abdominal aneurysm (≥ 3.0 to ≤ 5.0 cm) and a mean (standard deviation [SD]) age of 72 (8) yr. We used three methods to determine the VT in 328 comparable tests, namely, self-detected metabolic system (MS), experts' visual (V) readings, and software using a log-log transformation (LLT) of ventilation vs oxygen uptake. Differences and agreement between the three methods were assessed using analysis of variance (ANOVA), coefficient of variation (CV), typical error limits of agreement (LoA), and interclass correlation coefficients (ICC).

Results Overall, ANOVA revealed significant differences between the methods [$MS = 14.1$ (4.3) $mLO_2 \cdot kg^{-1} \cdot min^{-1}$; $V = 14.6$ (4.4) $mLO_2 \cdot kg^{-1} \cdot min^{-1}$; and $LLT = 12.3$ (3.3) $mLO_2 \cdot kg^{-1} \cdot min^{-1}$; $P < 0.001$]. The assessment of agreement between methods provided the following results: $ICC = 0.85$; 95% confidence interval (CI), 0.82 to 0.87; $P < 0.001$; typical error, 2.1–2.8 $mLO_2 \cdot kg^{-1} \cdot min^{-1}$; and, 95% LoA and CV ranged from 43 to 55% and 15.9 to 19.6%, respectively.

Conclusions The results show clinically significant variations between the methods and underscore the challenges of determining VT for perioperative risk stratification. The findings highlight the importance of meticulous evaluation of VT for predicting surgical outcomes. Future studies should address the prognostic perioperative utility of computed mathematical models combined with an expert's review. This trial was registered at ClinicalTrials.gov, identifier: NCT00349947.

Résumé

Objectif Le seuil ventilatoire (SV) est un marqueur physiologique objectif de la capacité d'endurance aérobie. Ce marqueur a de bonnes applications diagnostiques dans les contextes préopératoires. Toutefois, la détermination du SV peut être difficile en raison de problèmes physiologiques ou méthodologiques, particulièrement en ce qui touche l'évaluation du risque chirurgical. L'objectif de cette étude était de comparer différentes méthodes de détermination du SV et de mettre en lumière les implications de chacune quant à l'évaluation du risque périopératoire.

Méthode Notre étude comprenait l'analyse de 445 tapis d'effort cardiopulmonaire de 140 candidats préchirurgicaux souffrant d'un anévrisme de l'aorte

B. Vainshelboim, PhD (✉) · S. Rao, MD · K. Chan, MD · R. M. Lima, PhD · J. Myers, PhD
Division of Cardiology, Veterans Affairs Palo Alto Health Care System/Stanford University, 111C, 3801 Miranda Ave, Palo Alto, CA 94304, USA
e-mail: baruch.v1981@gmail.com

E. A. Ashley, MD
Stanford Center for Inherited Cardiovascular Disease, Stanford University, Stanford, CA, USA

abdominale ($\geq 3,0$ à $\leq 5,0$ cm) et d'un âge moyen (écart type [ET]) de 72 (8) ans. Nous avons utilisé trois méthodes pour déterminer le SV dans 328 tests comparables, soit le système métabolique (SM) auto-déecté, les relevés visuels (V) des experts, et un logiciel utilisant une transformation bilogarithmique de la ventilation vs la consommation d'oxygène (LLT). Les différences et les concordances entre les trois méthodes ont été évaluées à l'aide d'une analyse de la variance (ANOVA), du coefficient de variation (CV), de l'erreur type vs les limites d'agrément (LoA), et des coefficients de corrélation interclasse (CCI).

Résultats Globalement, l'ANOVA a révélé des différences significatives entre les méthodes [SM = 14,1 (4,3) $\text{mLO}_2\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$; V = 14,6 (4,4) $\text{mLO}_2\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$; et LLT = 12,3 (3,3) $\text{mLO}_2\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$; $P < 0,001$]. L'évaluation de l'agrément entre les méthodes a donné les résultats suivants : CCI = 0,85; intervalle de confiance (IC) 95 %, 0,82 à 0,87; $P < 0,001$; erreur type, 2,1–2,8 $\text{mLO}_2\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$; et, 95 % LoA et CV allaient de 43 à 55 % et de 15,9 à 19,6 %, respectivement.

Conclusion Les résultats montrent des variations significatives d'un point de vue clinique entre les diverses méthodes et soulignent les difficultés rencontrées pour déterminer le SV afin de stratifier le risque périopératoire. Ces résultats mettent en lumière l'importance de l'évaluation méticuleuse du SV pour prédire les pronostics chirurgicaux. De futures études devraient examiner l'utilité périopératoire pronostique de modèles mathématiques calculés combinés à un examen par un expert. Cette étude est enregistrée au ClinicalTrials.gov, identifiant : NCT00349947.

It is estimated that more than 320 million major surgical procedures are performed each year worldwide and this number is increasing.¹ Based on United States data, perioperative mortality rates of 1.3–3.2% were reported.^{2,3} The European Society of Intensive Care Medicine and the European Society of Anaesthesiology reported 4% mortality and 16–18% morbidity rates in patients admitted to noncardiac surgery, suggesting the importance of implementing measures to improve surgical outcomes.^{4,5} As life expectancy increases, a growing proportion of patients are likely to be high-risk elderly patients with multiple comorbidities who present risk stratification challenges to both anesthesiologists and surgeons.⁵

The ventilatory threshold (VT)—derived from a cardiopulmonary exercise test (CPET)—has been widely used in assessing both athletes and patients with cardio pulmonary disease, and it has recently been shown to be a good prognostic marker in the context of surgical

outcomes.^{5,6,8–12} Ventilatory threshold values were associated with morbidity, mortality, surgery complications, and length of hospital stay.^{5,12–17} The VT is defined as an exercise intensity beyond which the increase in ventilation becomes disproportionate to the increase in power output or oxygen uptake during progressive exercise.^{11,18} This measure provides important information regarding the major physiological systems involved in aerobic exercise and constitutes a determinant of an individual's functional capabilities and cardiorespiratory performance.^{6,8–11,18}

The VT is an effort-independent variable that can be determined during submaximal exercise, and thus, it is considered safer than maximal exercise for particular patients. In pre-and postsurgical settings, this may be advantageous relative to peak VO_2 , which requires maximal effort.^{5,12,17} Nevertheless, determining the VT can be challenging in clinical conditions for several reasons. Many patients exhibit an abnormal ventilatory pattern, making it difficult to discern a clear breakpoint; the exercise protocol can influence the VT; and even experienced reviewers frequently differ in terms of detecting the VT.^{18–20} Computerized metabolic systems use varying software programs to detect the VT automatically, but the algorithms used are often unknown. Furthermore, most of these systems will determine a VT when a VT may not exist or may not be discernible to a clinician. Studies have shown that VTs defined using different mathematical algorithms for the same patient vary significantly.^{20,21} We have empirically observed that visual interpretation of the VT often differs significantly from the automatic interpretation derived from a metabolic system. Accordingly, we developed a computer program that chooses a deflection point in ventilation based on the least mean square error of the fit between the log of oxygen uptake and the log of ventilation.²² Given the caveats related to determining the VT, guidelines on CPET suggest confirming the VT visually using one of several methods.^{6,8,9}

Because the VT has been increasingly applied to assess surgical outcomes,^{5,12–17,23} a reliable determination of this point is critically important in evaluating perioperative prognosis. In the current study, we sought to quantify agreement between methods of determining the VT. We compared visual readings by experienced reviewers, automated determination by a metabolic system, and the application of a computer-derived program (developed by the authors) which employs the point with the least mean square error of the fit between the log of oxygen uptake and the log of ventilation. The latter has been suggested as the optimal mathematical technique to define a breakpoint between two regression lines and has been applied to interpret the VT.^{22,24,25} Our overall objective was to provide insight into the reliability of common methods for determining the VT and to highlight the implications for assessing surgical outcomes.

Methods

Subjects and study design

The current study is a secondary analysis of a previously reported clinical trial (Clinical Trials identifier: NCT00349947).²⁶ The study entailed analysis of 445 treadmill CPETs from 140 presurgical candidates participating in a clinical trial at the Veterans Affairs Palo Alto Health Care System. The study participants each had a small aortic abdominal aneurysm (AAA) (≥ 3.0 to ≤ 5.0 cm) and a mean (standard deviation [SD]) age of 72 (8) yr. Patients were assessed at baseline and after three, 12, 24, and 36 months. Recruitment procedures and all study-related activities were reviewed and approved in advance by the Institutional Review Board at Stanford University (approved February 2016). Written informed consent was obtained from all participants.²⁶

Cardiopulmonary exercise testing

Symptom-limited exercise testing was performed using an individualized ramp treadmill protocol such that test duration was targeted to fall within a range of 8–12 min as previously recommended.^{6,27,28} Standardized medical examinations were performed before testing, and medications were continued as prescribed. A 12-lead electrocardiogram was obtained at rest, each minute during exercise, and for at least eight minutes during recovery. Blood pressure was measured at rest, every other minute during exercise, and at one, two, five, and eight minutes during recovery or until symptoms, electrocardiogram changes, and blood pressure stabilized. In the absence of clinical indications for stopping the exercise test, participants were encouraged to exercise until volitional fatigue, and the Borg 6–20 scale was used to quantify effort.²⁸ Cardiopulmonary responses were obtained using a COSMED Quark CPET metabolic system (Rome, Italy) that was calibrated in a standard fashion before each test. Minute ventilation, body temperature pressure, saturated (V_E); oxygen uptake, standard temperature pressure, dry (VO_2); carbon dioxide production, standard temperature pressure, dry (VCO_2); and other CPET variables were acquired breath by breath, reported in rolling ten-second intervals, and averaged over 30 sec.

Ventilatory threshold determination

We used all three methods to conduct 445 tests and determined 328 VTs to use in our analysis. The metabolic system method employed an automated program based on the V-slope method.⁶ The visual VT was determined by two reviewers who were highly experienced in CPET interpretation. The reviewers were blinded to the other

methods of determination, to any patient identification, and to the VT determined by the other reviewer. The reviewers considered a plot of the V-slope and the ventilatory equivalents for O_2 and CO_2 (V_E/VO_2 and V_E/VCO_2 , respectively) for determining the VT, using rolling averages of 30-sec data listed every ten seconds. In case of disagreement between the reviewers, the VT was considered indeterminate by consensus.^{6,10} The third method involved a log-log transformation (LLT) of VO_2 and V_E using a validated program developed previously by our group.²² The program is based on an exponential plus constant model describing the relationship between gas exchange and blood lactate accumulation, previously described by Hughson *et al.*,²⁵ and a modification of the log-log transformation model reported by Beaver *et al.*²⁴ The program uses the log-log threshold model, plotting transformed V_E vs transformed VO_2 data in which the VT is chosen as a division point between regression lines determined by those segments yielding the lowest mean square error, as described by Lundberg *et al.*²⁹

Statistical analysis

Descriptive statistics are presented as mean (SD). Comparisons between methods were performed using one-way analysis of variance (ANOVA) for repeated measures, and *post hoc* testing was performed using the Bonferroni method. Reliability of the VT readings was assessed using the typical error limits of agreement (LoA), expressed as both absolute values and as a percentage of the mean VT (LoA%), coefficient of variation (CV), intraclass correlation coefficient (ICC), and Pearson's correlations.^{19,30} The ICC values were classified as < 0.40 = poor; 0.40 – 0.59 = fair; 0.60 – 0.74 = good; and 0.75 – 1.0 = excellent.³¹ Additionally, the degree of concordance between the methods was evaluated by Bland-Altman plots.³² The typical error was calculated as the standard deviation of the differences between each pair of VT methods divided by the square root of 2. The CV was calculated as the typical error divided by the mean of both values expressed as a percentage. The 95% LoA was calculated as the standard deviation of the difference times 1.96. The LoA was also expressed relative to the mean and expressed as a percentage.³⁰ The statistical analyses were performed using SPSS® v.17 software (Chicago, IL, USA), with the significance level set at $P < 0.05$.

Results

Demographic, clinical, and physiological characteristics of the subjects are summarized in Table 1. In general, the sample was Caucasian (81%) and male (92%) with a mean

age of 72 (8) yr. Most subjects had hypertension (75%) and a previous smoking history (81%). The average peak VO_2 of the sample was $19.4(6.8) \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, with a mean peak respiratory exchange ratio of 1.1 (0.10) (Table 1).

The VT was detected in 96% using the metabolic system, 84% using visual observation (the two reviewers agreed that a VT occurred in 84% of the tests), and 76% using the log-log transformation program. The VT was determined in 328 tests by all three methods that were used for the analysis. Overall, ANOVA revealed significant differences between the methods [$\text{MS} = 14.1(4.3) \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$; $\text{V} = 14.6(4.4) \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$; and $\text{LLT} = 12.3(3.3) \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$; $P < 0.001$]. Further analysis using Bonferroni *post hoc* method showed significant differences only between MS vs LLT and between V vs LLT. Mean differences between the methods were as follows: MS vs V, $-0.50(4) \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$; $P = 0.105$; MS vs LLT, $1.8(2.9) \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$; $P < 0.001$; and, V vs LLT, $2.3(3.1) \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$; $P < 0.001$) (Table 2).

The reliability and agreement tests between the methods of VT determination showed that the typical error ranged from $2.1\text{--}2.8 \text{ mL O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$; ICC was 0.85 (95% confidence interval, 0.82 to 0.87; $P < 0.001$), and the 95% CVs ranged from 15.9 to 19.6%. The 95% LoA as a percentage of the mean ranged from 43% and 55%, with significant correlations between the methods ranging from $r = 0.59\text{--}0.73$ (Table 2, Figs 1–3).

Discussion

In the current study, we aimed to quantify agreement between three methods of determining VT in patients with small, presurgical AAA and to highlight its implications for perioperative risk assessment. We observed significant differences between the methods, although agreement was evident (Figs 1–3, Table 2). The CVs, Bland-Altman plots, and 95% LoA confirmed a relatively large variation between the methods, although moderately high correlations were observed (Figs 1–3, Table 2). These findings have clinical significance, particularly in preoperative settings among severely deconditioned patients when a maximal exercise test is not advisable and the VT is used as a primary efficacy or prognostic endpoint. In light of the growing use of the VT in the context of surgical risk,^{5,12} these results also underscore the importance of meticulous evaluation of this physiological point when applying the VT in the context of perioperative risk stratification. The results may also underscore the need for future studies that address a combination of methods and use experienced interpretation (e.g., a clinical exercise

Table 1 Demographics and clinical and physiological characteristics of the studied population

Demographics	<i>n</i> = 140 (328 tests)
Age (yr)	72 (8)
Sex male, <i>n</i> (%)	129 (92%)
Sex female, <i>n</i> (%)	11 (8%)
<i>Race</i>	
Caucasian <i>n</i> (%)	113 (81%)
Black <i>n</i> (%)	6 (4%)
Hispanic <i>n</i> (%)	9 (6%)
Asian <i>n</i> (%)	5 (4%)
Other <i>n</i> (%)	20 (15%)
BMI (index)	28 (4)
Clinical History	
Small Aortic Aneurysm Disease <i>n</i> (%)	140 (100%)
Coronary Artery Disease <i>n</i> (%)	39 (28%)
Hypertension <i>n</i> (%)	105 (75%)
Peripheral Vascular Disease <i>n</i> (%)	16 (22%)
Diabetes <i>n</i> (%)	29 (21%)
Current Smokers <i>n</i> (%)	22 (16%)
Previous Smokers <i>n</i> (%)	113 (81%)
Pack/Years	32.8 (29)
Medications	
Ace Inhibitors/ Angiotensin Receptor Blocker <i>n</i> (%)	13 (9%)
Beta blockers <i>n</i> (%)	65 (47%)
Statins <i>n</i> (%)	113 (81%)
Calcium Channel Blockers <i>n</i> (%)	68 (49%)
Cardiopulmonary exercise test	
<i>Pretest resting conditions</i>	
HR (beats·min ⁻¹)	70 (12)
SBP (mmHg)	133 (19)
DBP (mmHg)	77 (11)
<i>Peak values</i>	
HR (beats·min ⁻¹)	127 (24)
SBP (mmHg)	181 (30)
DBP (mmHg)	82 (20)
Minute Ventilation (L·min ⁻¹)	59 (17)
Respiratory exchange ratio	1.1 (0.1)
Rating of perceived exertion (6–20)	18.6 (1.4)
Oxygen consumption (mL·kg ⁻¹ ·min ⁻¹)	19.4 (6.8)
Test duration (sec)	524 (227)

Data are presented as means (standard deviation) and absolute numbers with % of the group for categorical variables

BMI body mass index, *DBP* diastolic blood pressure, *HR* heart rate, *SBP* systolic blood pressure

physiologist) along with computed mathematical algorithms for accuracy of predicting surgical outcomes.

The current study shows the challenges associated with determining the VT from a clinical perspective. This novel

Table 2 Inter-methods agreement between ventilatory threshold determinations

Inter-methods Agreement	Mean (SD) Between the Methods ($\text{mLO}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$)	Mean Difference (95% CI) ($\text{mLO}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$)	Typical Error	95% LoA ($\text{mLO}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$)	LoA as % of Mean	CV
Metabolic System vs Visual	14.3 (3.9)	-0.5 (-1 to 0.06)	2.8	± 7.8	55%	19.6%
Visual vs Log-log Transformation	13.4 (3.6)	2.3 (1.9 to 2.7)	2.2	± 6.1	46%	16.4%
Metabolic System vs Log-log Transformation	13.2 (3.5)	1.8 (1.4 to 2.1)	2.1	± 5.7	43%	15.9%
Average of 3 Methods	13.7 (3.5)	1.2 (-2.2 to 4.1)	1.4	± 3.9	29%	10.2%

SD = Standard Deviation; CI = confidence interval; CV = coefficient of variation; LoA = level of agreement

Fig. 1 Bland-Altman plot and correlation coefficient between the visual and metabolic systems for determining the ventilatory threshold

study addresses the methodological issues involved in determining VT in a large sample of presurgical elderly individuals. By using this approach, we attempted to resolve several existing gaps in the literature and to highlight concerns with determining VT in the context of preoperative risk stratification. The inclusion of a relatively large sample of presurgical AAA subjects overcomes one of the limitations in previous studies, that is, comparing different methods of determining VT among young healthy

subjects.^{11,20–22,33–37} In general, our findings are consistent with previous reports showing fairly good agreement between different methods of determining VT.^{11,19,22,33–35,37} In this regard, several studies have shown moderately high correlation ($r = 0.59–0.98$) between computed mathematical models, and visual readings of the VT, lactate responses, or both.^{11,19,22,33–35,37} Nevertheless, these relatively high associations between methods are contrasted by several other studies reporting poor agreement.^{20,21,38} For instance,

Fig. 2 Bland-Altman plot and correlation coefficient between the visual and log-log transformation methods for determining the ventilatory threshold

Crescêncio *et al.*²¹ reported that, despite a good correlation ($r = 0.82$) between visual and linear-linear models, significant underestimation was shown in the mathematical model when compared with other models in 23 healthy males.²¹ Ekkekakis *et al.*²⁰ similarly showed large variations between computerized methods of determining VT in two different samples of 30 young healthy volunteers.²⁰ Dickstein *et al.*³⁸ also reported significant differences between a computerized regression model, visual assessment, and lactate measures in 30 males with documented myocardial infarction.³⁸ Taken together, these findings highlight the existing debate regarding an optimal method for determining VT and the need for further research in this area.

Our observations of the different methods for determining VT have practical implications. Depending on the method used, we detected a mean difference of 0.5–2.3 $\text{mLO}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ and variation ranging from 15.9 to 19.6% (Table 2). Previous studies assessing variation in the VT between experienced readers similarly showed roughly 20% interobserver variability, 17–25% intra-observer variability, and 21% variability between centres.^{19,39} In practical terms, this degree of variability is potentially significant when considering outcomes in perioperative

assessments.^{5,11,13,15,22,39–41} In two recent reviews, Levett *et al.*^{5,12} summarized studies assessing the association between preoperative VT and postoperative outcomes. These studies among different surgery candidates (including major intra-abdominal, colon, rectal, hepatobiliary, hepatic resection, liver transplant, upper gastrointestinal, bariatric, cystectomy, and AAA repair procedures) showed that values below 9–12 $\text{mLO}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ were associated with a significantly increased risk for postoperative mortality, morbidity, surgical complications, and prolonged hospital stay.^{5,12} The variation between the methods for determining VT observed in the present study could potentially misclassify surgical candidates into inappropriate risk categories, particularly among patients with borderline perioperative risk presenting VT values ranging, for example, from 10 to 13 $\text{mLO}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$. The current results call attention to the variability between methods determining the VT and underscore that this should be considered when applying the VT in clinical settings, particularly in the context of surgical decisions.

Our findings suggest that, although there is agreement between methods of determining VT, differences and relatively large variations were observed with both statistical and clinical significance (Table 2, Figs 1–3). These may be

Fig. 3 Bland-Altman plot and correlation coefficient between the metabolic system and log-log transformation for determining the ventilatory threshold

due in part to the fact that computed and mathematical algorithms lack subjective judgement by an experienced reviewer. Such judgement may be valuable in many clinical circumstances, particularly when the ventilatory response is curvilinear and there is no clear breakpoint evident in V_E or VCO_2 .¹⁸ Nevertheless, visual interpretation is also subject to inter- and intra-observer variation.^{11,19} Determining the lactate threshold by measuring blood lactate responses directly is a gold standard, but this is rarely employed in clinical practice due to the need for frequent blood sampling and the associated costs. Moreover, similar to ventilatory methods to derive a threshold, direct measures using blood sampling have also been the topic of a great deal of debate.^{11,18} Due to the lack of a gold standard for determining the VT, our results suggest that a combination of mathematically computed algorithms, along with visual adjusting by individuals experienced in CPET, can improve the accuracy of VT determination. Nevertheless, this needs to be ascertained in future studies comparing a combined VT determination method with blood lactate kinetics.

The strengths of the current study include a relatively large sample size (328 comparable tests), which significantly exceeds previous studies that have generally included only several dozen subjects.^{22,33,34,36,37} An additional strength

includes the assessment of elderly presurgical AAA patients with multiple comorbidities. These patients are likely to be a reasonable reflection of many surgical candidates, although further studies are needed to determine the application of our findings to other conditions. The VT values we observed ($12.3\text{--}14.6\text{ mL O}_2\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) using treadmill testing are comparable with VTs ($10\text{--}12\text{ mL O}_2\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) observed using a cycle ergometer test among surgical candidates with high perioperative risk. This reflects the 10–25% higher values typical for treadmill testing vs the cycle ergometer.^{5,12,28} Finally, we used a broad statistical analysis to assess both differences and level of agreement between methods to gain more insight into the determination of the VT and its implications for surgical risk.

Our study also has several limitations. First, we lacked validation of ventilatory assessments with blood lactate measures. Nevertheless, previous reports have shown good agreement between ventilatory responses and lactate responses,^{11,22,33,35} providing reasonable confidence for the use of ventilatory-derived surrogates for lactate accumulation. Secondly, our sample consisted largely of male subjects (92%), thus generalization of the results to females is unknown. Finally, our study provides objective assessment of differences and agreement between common

methods of determining VT and their consideration for perioperative risk stratification. Nevertheless, the results do not suggest the utility of using one method over another to predict perioperative outcomes. This needs to be ascertained in future studies assessing preoperative VT and its association with surgical outcomes.

In summary, we observed significant differences and variation between the VT determined by a metabolic system, visual reading by experienced clinical researchers, and a log-log transformation program, although agreement between the methods was observed. These data underscore the challenges associated with reliably determining the VT, and they emphasize the importance of meticulous evaluation in interpreting VT for clinical decisions in perioperative settings. Future research is warranted addressing the combination of mathematical-computed models with visual review by individuals experienced in CPET when applying the VT for surgical outcomes, particularly when the VT is used as a primary endpoint.

Conflicts of interest All authors declare that there is no conflict of interest.

Editorial responsibility This submission was handled by Dr. Gregory L. Bryson, Deputy Editor-in-Chief, *Canadian Journal of Anesthesia*.

Author's contributions All authors made a significant contribution to the conception and design of the manuscript, analysis and interpretation of data, drafting the article, and/or revising it critically for important intellectual content. *Jonathan Myers* was the principal investigator and was involved in critically revising the manuscript. *Jonathan Myers* and *Baruch Vainshelboim* were involved in the conception and design of the study. *Baruch Vainshelboim*, *Shravan Rao*, *Khin Chan*, *Euan A. Ashley*, and *Jonathan Myers* were involved in data acquisition. *Baruch Vainshelboim* was involved in statistical analysis and interpretation and in writing, drafting, revising and submitting the article. *Shravan Rao*, *Khin Chan*, *Ricardo M. Lima*, and *Euan A. Ashley* were involved in data analysis. *Ricardo M. Lima*, *Euan A. Ashley*, and *Jonathan Myers* were involved in data interpretation. *Ricardo M. Lima* and *Euan A. Ashley* were involved in drafting and critical review of manuscript.

Funding The study was funded by the National Heart, Lung, and Blood Institute.

References

1. Rose J, Weiser TG, Hider P, Wilson L, Gruen RL, Bickler SW. Estimated need for surgery worldwide based on prevalence of diseases: a modelling strategy for the WHO Global Health Estimate. *Lancet Glob Health* 2015; 3 Suppl 2: S13-20.
2. Watters DA, Hollands MJ, Gruen RL, et al. Perioperative mortality rate (POMR): a global indicator of access to safe surgery and anaesthesia. *World J Surg* 2015; 39: 856-64.
3. Khuri SF, Henderson WG, DePalma RG, et al. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Ann Surg* 2005; 242: 326-41; discussion 341-3.
4. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012; 380: 1059-65.
5. Levett DZ, Grocott MP. Cardiopulmonary exercise testing for risk prediction in major abdominal surgery. *Anesthesiol Clin* 2015; 33: 1-16.
6. American Thoracic Society; American College of Chest Physicians. ATS/ACCP Statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med* 2003; 167: 211-77.
7. Balady GJ, Arena R, Sietsema K, et al. Clinician's guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. *Circulation* 2010; 122: 191-225.
8. Guazzi M, Adams V, Conraads V, et al. EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. *Circulation* 2012; 126: 2261-74.
9. Wasserman K, Hansen JE, Sue DY, et al. Principles of Exercise Testing and Interpretation : Including Pathophysiology and Clinical Applications. 5th ed. Philadelphia: Wolters Kluwer/ Lippincott Williams & Wilkins; 2012 .
10. Svedahl K, MacIntosh BR. Anaerobic threshold: the concept and methods of measurement. *Can J Appl Physiol* 2003; 28: 299-323.
11. Levett DZ, Grocott MP. Cardiopulmonary exercise testing, prehabilitation, and Enhanced Recovery After Surgery (ERAS). *Can J Anesth* 2015; 62: 131-42.
12. Smith TB, Stonell C, Purkayastha S, Paraskevas P. Cardiopulmonary exercise testing as a risk assessment method in non cardio-pulmonary surgery: a systematic review. *Anaesthesia* 2009; 64: 883-93.
13. West MA, Lythgoe D, Barben CP, et al. Cardiopulmonary exercise variables are associated with postoperative morbidity after major colonic surgery: a prospective blinded observational study. *Br J Anaesth* 2014; 112: 665-71.
14. Older P, Hall A, Hader R. Cardiopulmonary exercise testing as a screening test for perioperative management of major surgery in the elderly. *Chest* 1999; 116: 355-62.
15. Hartley RA, Pichel AC, Grant SW, et al. Preoperative cardiopulmonary exercise testing and risk of early mortality following abdominal aortic aneurysm repair. *Br J Surg* 2012; 99: 1539-46.
16. Ridgway ZA, Howell SJ. Cardiopulmonary exercise testing: a review of methods and applications in surgical patients. *Eur J Anaesthesiol* 2010; 27: 858-65.
17. Hopker JG, Jobson SA, Pandit JJ. Controversies in the physiological basis of the 'anaerobic threshold' and their implications for clinical cardiopulmonary exercise testing. *Anaesthesia* 2011; 66: 111-23.
18. Myers J, Goldsmith RL, Keteyian SJ, et al. The ventilatory anaerobic threshold in heart failure: a multicenter evaluation of reliability. *J Card Fail* 2010; 16: 76-83.
19. Ekkkekakis 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.
20. Crescencio JC, Martins LE, Murta LO Jr, et al. Measurement of anaerobic threshold during dynamic exercise in healthy subjects: comparison among visual analysis and mathematical models. *Comput Cardiol* 2003; 30: 801-4.
21. Myers J, Walsh D, Buchanan N, McAuley P, Bowes E, Froelicher V. Increase in blood lactate during ramp exercise: comparison of continuous and threshold models. *Med Sci Sports Exerc* 1994; 26: 1413-9.
22. Tolchard S, Angell J, Pyke M, et al. Cardiopulmonary reserve as determined by cardiopulmonary exercise testing correlates with

- length of stay and predicts complications after radical cystectomy. *BJU Int* 2015; 115: 554-61.
23. Beaver WL, Wasserman K, Whipp BJ. Improved detection of lactate threshold during exercise using a log-log transformation. *J Appl Physiol* (1985) 1985; 59: 1936-40.
 24. Hughson RL, Weisiger KH, Swanson GD. Blood lactate concentration increases as a continuous function in progressive exercise. *J Appl Physiol* (1985) 1987; 62: 1975-81.
 25. Myers J, McElrath M, Jaffe A, et al. A randomized trial of exercise training in abdominal aortic aneurysm disease. *Med Sci Sports Exerc* 2014; 46: 2-9.
 26. Myers J, Buchanan N, Walsh D, et al. Comparison of the ramp versus standard exercise protocols. *J Am Coll Cardiol* 1991; 17: 1334-42.
 27. Pescatello LS, American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription, 9th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health; 2013.
 28. Lundberg MA, Hughson RL, Weisiger KH, Jones RH, Swanson GD. Computerized estimation of lactate threshold. *Comput Biomed Res* 1986; 19: 481-6.
 29. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000; 30: 1-15.
 30. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol* 2012; 8: 23-34.
 31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-10.
 32. Yamamoto Y, Miyashita M, Hughson RL, Tamura S, Shinohara M, Mutoh Y. The ventilatory threshold gives maximal lactate steady state. *Eur J Appl Physiol Occup Physiol* 1991; 63: 55-9.
 33. Orr GW, Green HJ, Hughson RL, Bennett GW. A computer linear regression model to determine ventilatory anaerobic threshold. *J Appl Physiol Respir Environ Exerc Physiol* 1982; 52: 1349-52.
 34. Gaskill SE, Ruby BC, Walker AJ, Sanchez OA, Serfass RC, Leon AS. Validity and reliability of combining three methods to determine ventilatory threshold. *Med Sci Sports Exerc* 2001; 33: 1841-8.
 35. Higa MN, Silva E, Neves VF, Catai AM, Gallo L Jr, Silva de Sa MF. Comparison of anaerobic threshold determined by visual and mathematical methods in healthy women. *Braz J Med Biol Res* 2007; 40: 501-8.
 36. Novais LD, Silva E, Simoes RP, et al. Anaerobic threshold by mathematical model in healthy and post-myocardial infarction men. *Int J Sports Med* 2016; 37: 112-8.
 37. Dickstein K, Barvik S, Aarsland T, Snapinn S, Karlsson J. A comparison of methodologies in detection of the anaerobic threshold. *Circulation* 1990; 81(1 Suppl): II38-46.
 38. Yeh MP, Gardner RM, Adams TD, Yanowitz FG, Crapo RO. "Anaerobic threshold": problems of determination and validation. *J Appl Physiol Respir Environ Exerc Physiol* 1983; 55: 1178-86.
 39. Grant SW, Hickey GL, Wisely NA, et al. Cardiopulmonary exercise testing and survival after elective abdominal aortic aneurysm repair. *Br J Anaesth* 2015; 114: 430-6.
 40. Dunne DF, Jones RP, Lythgoe DT, et al. Cardiopulmonary exercise testing before liver surgery. *J Surg Oncol* 2014; 110: 439-44.
 41. Junejo MA, Mason JM, Sheen AJ, et al. Cardiopulmonary exercise testing for preoperative risk assessment before pancreaticoduodenectomy for cancer. *Ann Surg Oncol* 2014; 21: 1929-36.