Reliability of the determination of the ventilatory threshold in patients with COPD

presented by

Dr Bruno-Pierre Dubé

in partial fulfillment of the requirements for the degree of

Master of Science (M.Sc) in Exercise Science
at
Concordia University
Montreal, Qc, Canada

October 2016

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By : Bruno-Pierre Dubé

Entitled : Reliability of the determination of the ventilatory threshold in patients with COPD

And submitted in partial fulfillment of the requirements for the degree of

Master of Science (Exercise Science)

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee

______________________________ Chair

______________________________ Examiner

______________________________ Examiner

______________________________ Supervisor

Approved by ____

Graduate Program Director

Date ___________ ___________________________________

Dean of Faculty

Abstract

Purpose

The ventilatory threshold (VT) is a physiological turning point that can be used to guide for exercise prescription, as a tool to monitor response to an intervention and as a prognostic marker, but the presence of respiratory disease may limit the reliability of its measurement. This project aimed to determine the reliability of the assessment of the ventilatory threshold among human and computerized observers, in patients with chronic obstructive pulmonary disease (COPD) and controls.

Methods

VT was identified from incremental exercise testing graphs of 115 subjects (23 controls and 23 in each COPD severity class) by two human observers and a computer analysis, using the V-slope method and the VEM. Agreement between observers for VO_2 at VT (VO_{2VT}) and heart rate at VT (HR_{VT}) were evaluated using intra-class correlation (ICC) for humans and Passing-Bablok regression analysis (human *vs* computer).

Results

For humans, ICCs for VO_{2VT} were higher in controls [0.98 (0.97-0.99) both with V-slope and with VEM] than in COPD patients [0.72 (0.60-0.81) with V-slope and 0.64 (0.50-0.74) with VEM]. Human and computerized values of VO_{2VT} were interchangeable in controls, but not in COPD patients. FEV_1 and peak-ventilation were independent predictors of a lesser reliability of VO_{2VT}. Inter-observer differences in HR_{VT} ranged from 2±1 beats/minute (controls) to 10±3 beats/minute (GOLD 4).

Conclusions

In COPD, the reliability of human estimation of VO_{2VT} is less in than in controls and not interchangeable with a computerized analysis. This should be taken into account when using VT in the clinical and research settings.

Acknowledgements

This work would not have been possible without the important and continuous support that I received from many collaborators, friends and colleagues. I wish to express my appreciation and deepest gratitude to the following persons who in one way or another have contributed in making this project possible.

Dr Véronique Pepin, my supervisor, who supported and guided me through this project. I am grateful for her knowledge, her patience and her ability to provide me with constant encouragement despite my unusual trajectory through the Master's program, which included a whole year abroad to compete my medical formation. This Master's project made me not only grow as a scientist, but also as a person, and this is in no small part thanks to her.

Dr Simon Bacon and **Dr Andreas Bergdahl**, members of my thesis committee, for their suggestions, comments and insightful criticism of this work.

Dr François Beaucage, former member of my thesis committee, for his time and help, and for his role in giving me access to the respiratory physiology data from the Sacré-Coeur hospital.

Dr Myriam Mesbahi, a trusted and respected colleague, who took of her own time and expertise to play a fundamental role in my project.

The **staff of the respiratory physiology laboratory** of Sacré-Coeur hospital, which helped us extract data from the databases and had to tolerate me for many days in their work environment. Thank you!

Contribution of Authors

BPD and VP designed the study. BPD coordinated the study. BPD, MM and FB were responsible for patient screening and selection. BPD and MM performed ventilatory threshold measurements. BPD, MM, FB and VP analyzed the data. BPD and VP wrote the manuscript. All authors had full access to all of the study data, contributed to draft the manuscript and revised it critically for important intellectual content, approved the final version of the manuscript, and take responsibility for the integrity of the data and the accuracy of the data analysis.

Table of contents

1. List of figures ... **1**

2. List of tables ... **2**

3. List of abbreviations ... **3**

4. Theoretical context ... **5**
4.1 COPD ... **5**
4.1.1 Epidemiology of COPD ... 5
4.1.2 Risk factors, pathogenesis and clinical consequences of COPD ... 6
4.1.3 Management of COPD ... 7
4.2 Anaerobic threshold *vs* Ventilatory threshold ... **8**
4.2.1 Anaerobic threshold (AT) ... 8
4.2.2 Ventilatory threshold (VT) ... 9
4.2.2.1 How to identify it ... 9
4.2.2.2 Is there a true relationship between VT and AT? ... 11
4.2.2.3 Use of the VT in clinical practice ... 12
4.3 Variance in the measure of VT ... **16**
4.3.1 Manual measure of VT ... 16
4.3.2 Computerized measure of VT ... 19

5. Rationale and objectives ... **20**
5.1 Rationale ... **20**
5.2 Objectives ... **20**
5.2.1 Primary research objective: ... 20
5.2.2 Secondary research objectives: ... 20

6. Hypotheses ... **21**

7. Article: Reliability of the determination of the ventilatory threshold in patients with COPD **22**
7.1 Abstract ... **23**
7.2 Introduction ... **24**
7.3 Methods ... **25**
7.3.1 Subjects ... 26
7.3.2 Baseline measurements ... 27
7.3.3 Exercise testing ... 27
7.3.4 Ventilatory threshold ... 28
7.3.5 Statistical analyses ... 29
7.4 Results ... **30**
7.4.1 Agreement in the determination of VT for human observers ... 31
7.4.2 Inter-observer differences in HR_{VT} ... 33
7.4.3 Predictors of a larger inter-observer difference in VO_{2VT} ... 33
7.4.4 Internal validity ... 34
7.5 Discussion ... **34**
7.6 Conclusion ... **39**

8. References ... **40**

9. APPENDIX A: Supplemental digital content ... **51**

1. List of figures

Figure 1. Cut-off points of various indices during exercise……………………………9

Figure 2. V-slope method…………………………………………………..……10

Figure 3. Mean blood lactate concentration during exercise when applying a fixed 70-75% VO_2max intensity……………………………………………………….…….13

Figure 4. Bland-Altman procedures for inter-observer differences in VO_{2VT}, when using (A) the V-slope method and (B) the VEM…………………………………..…...32

2. List of tables

Table 1. Baseline subjects characteristics……………………………………………….30

Table 2. Agreement of human observers in the determination of VO_{2VT}……………….31

Table 3. Mean (SD) inter-observer difference in $VO2_{VT}$ (ml/min) according to disease severity……………………………………………………………….……….…31

Table 4. Mean (SD) inter-observer difference in the measure of HR $_{VT}$ (absolute value and %peakHR) using two methods, by COPD severity……………….………………..33

Table 5. Multiple linear regression models for the prediction of a larger inter-observer difference in VO_{2VT} when using V-slope or VEM……………………………….……...34

3. List of abbreviations

Acetyl-CoA	Acetyl coenzyme A
ACCP	American College of Chest Physicians
ACSM	American College of Sports Medicine
ANOVA	Analysis of Variance
AT	Anaerobic Threshold
ATP	Adenoside Triphosphate
ATS	American Thoracic Society
BMI	Body Mass Index
BPM	Beat Per Minute
CCL2	Chemokine (c-c motif) ligand 2
CO_2	Carbon dioxide
COPD	Chronic Obstructive Pulmonary Disease
CPET	CardioPulmonary Exercise Testing
CXCL	Chemokine (c-x-c motif) ligand
D_LCO	Diffusion Capacity of the Lung for Carbon Monoxide
DO_2	Oxygen delivery
EMG	Electromyogram
FEV_1	Forced Expiratory Volume in 1 second
FVC	Forced Vital Capacity
GOLD	Global initiative for Obstructive Lung Disease
HR	Heart Rate
HR_{peak}	Peak Heart Rate
HRR	Heart Rate Reserve
HR_{VT}	Heart Rate at the ventilatory threshold
ICC	Intra-class correlation
IPF	Idiopathic Pulmonary Fibrosis
O_2	Oxygen
P_aCO_2	Arterial partial pressure of carbon dioxide
$P_{ET}CO_2$	End-tidal tension of carbon dioxide
$P_{ET}O_2$	End-tidal tension of oxygen
pK_a	Acid dissociation constant
PR	Pulmonary Rehabilitation
R	Respiratory coefficient
RER	Respiratory Exchange Ratio
SD	Standard Deviation
V_A	Alveolar Ventilation
V_E	Minute-Ventilation
VCO_2	Carbon dioxide production
V_D	Dead space Volume
VEM	Ventilatory Equivalent Method
V_E/VCO_2	Respiratory equivalent for carbon dioxide production
V_E/VO_2	Respiratory equivalent for oxygen uptake
VO_2	Oxygen uptake

VO_{2AT}	Oxygen uptake at the anaerobic threshold
VO_2max	Maximal oxygen uptake
VO_2SL	Symptom-limited oxygen uptake
VO_{2VT}	Oxygen uptake at the ventilatory threshold
V_T	Tidal Volume
VT	Ventilatory Threshold
W	Watt

4. Theoretical context

4.1 COPD

4.1.1 Epidemiology of COPD

Chronic obstructive pulmonary disease (COPD) is a constellation of conditions characterized by persistent expiratory airflow limitation. Among the many reported phenotypes of COPD, the most prevalent remain the emphysema and chronic bronchitis variants. The prevalence of COPD varies widely across geographic regions(1). This is an effect of both technical differences in the assessment of COPD (for example, the use of self-administered questionnaires *vs* objective spirometric values, or the use of a fixed FEV_1/FVC ratio *vs* a lower-limit of normal model)(2) and real differences across countries. As an example, in the largest international study of the prevalence of the disease, Global initiative for Obstructive Lung Disease (GOLD) stage II COPD in women ranged from about 5% in China to 17% in South Africa, and in men from 4% in Mexico to 23% in South Africa(1). In 2012, the age-standardized prevalence of COPD among Canadian men was 3.5% and in women 4.3%, with fluctuations of about 1% across provinces(3).

COPD is the fourth leading cause of death in Canada, being responsible for almost 12 000 deaths in 2013(3). In the United States, COPD has risen to third place in mortality causes since 2008(4), and was one of the few diseases showing an increase in mortality rate between 2008 and 2011(5). Although cigarette smoking showed a slow decline in the last 10 years in Canada(6), the repercussions of the epidemic of cigarette smoking in the 20^{th} century are still evident today, as evidenced by the high prevalence and burden of COPD. As such, although the mortality associated with COPD in men peaked in the mid-1980's in Canada, mortality in women is still slowly increasing, representing the later peak in smoking prevalence in this population(7). These observations, coupled to the fact that COPD is a vastly underdiagnosed disease(8), underline the crucial importance of its early detection to allow for a timely and effective management of both its risk factors and its complications.

Along with being a major economic burden (COPD-related costs approximate 3.94 billion dollars in 2010 in Canada, and are expected to rise to 9.45 billion by 2030)(9), COPD is a cause of significant decrease in quality of life, which is related to the disease severity(10).

4.1.2 Risk factors, pathogenesis and clinical consequences of COPD

Worldwide, cigarette smoking remains the main risk factor associated with the development of COPD(11). In the last decades, evidence for other new determinant conditions, such as passive exposition to second-hand smoke, occupational exposures such as organic and inorganic dusts (i.e. asbestos, gold, cadmium, isocyanates, welding fumes and industrial cotton manufacturing)(12, 13) has increased. A growing concern in developing countries is the threat of indoor pollution (i.e. from biomass cooking and coal heating, especially in poorly ventilated areas), which is a newly identified risk factor for COPD(14, 15).

COPD results from a gene-environment interaction: among with people with the same smoking history, not all will develop the disease(16). The well documented predisposition of patients with alpha-1 antitrypsin (A1AT) deficiency(17) to develop emphysema, even in the absence of cigarette smoking, is an example of this phenomenon. The absence of A1AT, which normally inhibits the tissue-damaging effects of neutrophil elastase, promotes unregulated pulmonary destruction of elastase, which mimicks the effects of cigarette smoke on the lung parenchyma.

Pathologically, COPD is characterized by chronic inflammation of the respiratory tract, mediated by cytokines, chemokines (CCL2, CXCL1-8-9-10), adhesion molecules, inflammatory enzymes and reactive oxygen species(18), expressed by epithelial cells and macrophages in response to an inhaled irritant(19). These mediators cause a local cellular self-perpetuating inflammatory reaction involving fibroblasts, neutrophils, T_c1 lymphocytes and monocytes, which, through proliferation and production of enzymes (neutrophil elastase and matrix metalloproteinase-9)(20) will result in the hallmark pathological changes seen in COPD: small airway fibrosis, alveolar wall destruction (emphysema) and mucus hypersecretion(19). These alterations in the airway will induce the fixed expiratory airflow limitation that is characteristic of COPD(21).

Although chronic cough and sputum production are common symptoms in COPD patients, it is dyspnea, especially on exertion, that is the cardinal clinical finding in these patients(16). Airflow obstruction causes increased work of breathing, increased airway resistance, gas-exchange inefficiency, intrinsic positive end-expiratory airway pressure and dynamic hyperinflation on exertion(21, 22). The resulting shortness of breath is the first step in a downward spiral of breathlessness that includes fear of dyspnea and immobilisation, which itself induces muscle mass wasting (muscle atrophy) and exercise anxiety and intolerance(23). This phenomenon is known to negatively impact prognosis(24) and, coupled to the well-described systemic exercise-limiting effects of COPD(25) (nutritional anomalies and ‘’pulmonary cachexia’’(26), skeletal muscle dysfunction(27), coronary artery disease(28), depression(29), cognitive decline(30) and osteoporosis(31)) make increasing exercise capacity a key goal in the management of COPD patients.

4.1.3 Management of COPD

Current guidelines(16, 32) place emphasis on two goals in the management of COPD: risk reduction (preventing disease progression and exacerbations) and symptoms reduction (relieving dyspnea, improving exercise tolerance and improving health status). Although smoking cessation and pharmacological agents (using inhaled short- and long-acting $beta_2$-agonists and anticholinergics) remain the mainstay of COPD management, few interventions have proven to be as effective as pulmonary rehabilitation (PR) in reaching these goals. PR has clearly been shown to reduce dyspnea, increase exercise tolerance, improve quality of life, decrease healthcare utilisation and exacerbations in patients with even mild COPD(33-35).

Compared with healthy controls however, COPD patients exhibit a wide range of ventilatory and circulatory anomalies during exercise, including a slower adaptation to increasing work (slower time constants for minute-ventilation (V_E), VCO_2 and O_2 pulse), increased V_E with decreased alveolar ventilation (V_A) due to dynamic hyperinflation, increased work of breathing, oxygen desaturation and increase in right ventricular afterload(34, 36, 37), that all lead to marked exercise intolerance. Identifying the optimal intensity and modality of training in these patients, in order to balance the benefits of training with the risk/intolerance of exercising in these subjects, is therefore crucial, but difficult, and is still a matter of debate (34). Although the current PR guidelines(34) recommend endurance training based on the recommendations of the American College of Sports Medicine (i.e. with a goal of at least 60% of maximal work rate, either predicted or evaluated on an incremental exercise test(38), many patients with COPD cannot sustain or comply to this recommendation(39, 40). In light of these observations, and as an introduction to the following section, many studies have shown that an individualized training program based on each patient's level of aerobic fitness (i.e. based on their ventilatory threshold) is both safe and effective in patients with COPD(41-45).

4.2 Anaerobic threshold *vs* Ventilatory threshold

4.2.1 Anaerobic threshold (AT)

During exercise, muscle cells initially derive energy from adenosine triphosphate (ATP) molecules. ATP is the end-product of the aerobic cellular metabolism, in which glucose is sequentially converted to pyruvate and acetyl-CoA, which enters the Krebs cycle to produce ATP through oxidative phosphorylation. This aerobic process is highly efficient (30 ATP molecules per molecule of glucose)(46). During incremental exercise, a point is reached at which oxygen delivery (DO_2) is insufficient to meet the increasing demands of muscle cells. From this point, anaerobic metabolism contribution to energy production increases to further supplement energy production. This process is much less efficient (net gain of 2 ATP per glucose molecule) and results in the accumulation of lactic acid in the blood, as a by-product of pyruvate metabolism(47, 48). Lactic acid, having a very low pK_a, will readily dissociate into lactate ions and protons, which are buffered by serum bicarbonates, producing water and carbon dioxide that can then be excreted by the lungs(49). Although there is controversy as to whether the AT is truly a ''threshold'', (as some data suggests that blood lactate accumulation during exercise occurs continuously, or in a hyperbolic fashion(50, 51)), it is clear that the VO_2 associated with the AT (VO_{2AT}) represents a real metabolic turning point. Indeed, this point in exercise is associated with significant alterations in the ventilatory parameters – which form the premise for the concept of ventilatory threshold.

<u>4.2.2 Ventilatory threshold (VT)</u>

4.2.2.1 How to identify it

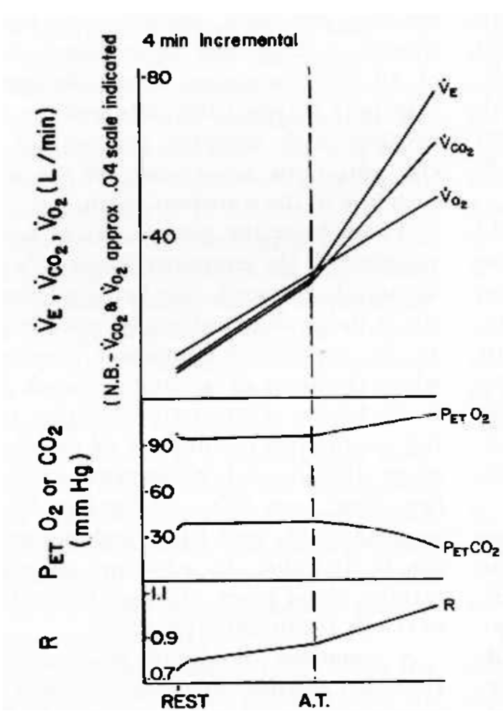

Figure 1. Cut-off points of various indices during exercise – purportedly representing the AT. *From reference 54.*

With the advent of cardiopulmonary exercise testing (CPET) in the 1960's, it was observed that both V_E and VCO_2 showed a break in the linearity of their increase during exercise, at a point approximately corresponding to AT(52-54). The term ''ventilatory'' threshold was introduced to specify that this point in exercise had been identified using ventilation-derived parameters (figure 1), instead on relying directly on the measurement of serum lactic acid.

Although the cut-off points of the respiratory coefficient (R) and V_E were the first described methods of identifying VT, their correlation with serum lactate concentration (AT) was variable.

Caiozzo et al. studied the correlation between different ventilatory indices and lactates values on 16 healthy individuals, and found correlations of 0.88 for V_E and 0.39 for R(55). Green et al. showed that, in 10 healthy participants, the difference between AT and VT (estimated using V_E plotted against VO_2) was large when expressed as power output, reaching 383 kg/min)(56).

Two non-invasive methods were showed to be more reliable in identifying the AT:

1. The respiratory equivalent method (VEM). First described by Reinhard(54), the VEM uses a plot of both the ventilatory equivalent for O_2 (V_E/VO_2) and for CO_2 (V_E/VCO_2) against work rate (W). VT is defined as the first point where there is an increase in V_E/VO_2 without a concomitant change in V_E/VCO_2. On 15 healthy subjects, correlation between the VEM and the AT was 0.94. Caiozzo et al. also evaluated V_E/VCO_2 for VT determination, and found a correlation of 0.93 with AT(55).
2. The V-slope method, originally described by Beaver et al(57). This method uses a plot of VCO_2 against VO_2, with VT being the breaking point in the linearity of their relationship (figure 2). They proposed that the V-slope has the advantage of excluding minute ventilation (V_E) from the graphical representation of the data, therefore truly only considering the metabolic compensation phenomenon (faster increase of VCO_2 relative to VO_2), without interference from the actual ventilatory rate or pattern, as can be seen in patients with hyperventilation syndromes, or COPD. In the original study, when compared with a mean value of VT derived from other methods (VEM, R, $P_{ET}O_2$ and $P_{ET}CO_2$), with AT as a benchmark measure, the mean value of VT using the V-slope method was not different from the composite value, but much more reliable (coefficient of variance 0.023 *vs* 0.127). V-slope analysis was the only method that could identify a VT in all subjects.

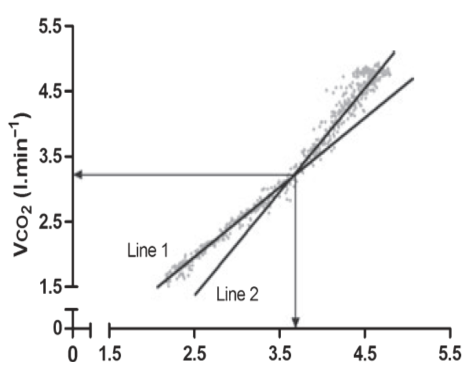

Figure 2. V-slope method. The intersection point of 2 regression lines show the inflection point of the relationship between VCO_2 and VO_2. *From reference 64.*

4.2.2.2 Is there a true relationship between VT and AT?

Although the rationale behind the VT is intuitive when considered under the aforementioned model where the excess CO_2 produced by anaerobic metabolism has to be excreted by the lungs via an increase in minute ventilation and VCO_2 (47, 49, 52, 53, 57, 58), the link of causality between AT and VT has been challenged by some authors. Green et al.(56), Patessio et al.(59) and Gladden et al.(60) all showed that AT and VT occurred at significantly different moments during exercise in healthy subjects. More convincingly, Péronnet and Aguilaniu(61) argued that:

1) It is impossible for any ''excess'' nonmetabolic CO_2 to be produced during exercise, as this would violate the law of mass conservation. The CO_2 that is thought to be created from anaerobic metabolism in the Wasserman model is in fact already present in the blood, in the form of bicarbonates formed from CO_2 during normal aerobic metabolism. Thus, the disproportionate increase in ventilation during exercise cannot be explained by ''new'' CO_2 synthesis.

2) The assertion that VCO_2 (measured at the mouth, as in CPET) determines V_E is wrong, as VCO_2 at the mouth does not equal CO_2 delivery to the lungs (Q_{VCO2} – true CO_2 production). The fact that VCO_2 increases disproportionally during exercise (as seen in the V-slope method) *cannot* be said to represent an increase in CO_2 production, but rather could be due to hyperventilation, with an increase in CO_2 release at the mouth, without change in CO_2 production.

This can be mechanistically shown using the developed equation of alveolar gases(61):

$$\mathrm{VCO2} = \frac{\mathrm{VE} \times \mathrm{PaCO2} \times (1 - \frac{V_D}{V_T})}{\mathrm{K}}$$

It clearly shows that, for a given value of dead space ratio, VCO_2 is *determined* by V_E (and P_aCO_2) – not the other way around. Of note, this argument was mentioned by Wasserman in the past(53), but dismissed on the grounds that the total quantity of CO_2 excreted ''in excess'' of metabolic demands was too significant to be solely attributed to hyperventilation.

Finally, and most importantly, Hagberg et al.(62) conducted a study on 4 patients with McArdle syndrome, an autosomal recessive genetic disease in which patients lack the enzyme glycogen phosphorylase, and therefore are *incapable* of producing lactic acid during exercise. During incremental testing, all patients showed a distinct and disproportionate increase in V_E similar to healthy controls, *despite no change in serum lactate values* and an increase, rather than a decrease, in serum pH.

These observations led to the search for another potential trigger for the VT – other than lactate production(63). Potassium has been implicated as a potential humoral trigger of ventilation during exercise(64). In anesthetised cats, potassium stimulates ventilation through excitation of chemoreceptors in the carotid bodies, and surgical denervation of these receptors prevent this phenomenon(65). In patients with McArdle disease, serum potassium levels track V_E better than serum lactate levels, both during exercise and recovery(66). It has also been shown that VT correlates well with a ''fatigue threshold'' on EMG(67, 68), which leads to the possibility of a higher neural activity controlling ventilation during exercise, possibly in relation to motor unit recruitment(69).

4.2.2.3 Use of the VT in clinical practice

Despite the uncertainties outlined in the last section, there remains little doubt that, independently of its underling mechanism, the VT represent a pivotal point for metabolism during exercise, and is correlated to a wide variety of relevant clinical outcomes.

Work beyond the VT is associated with significantly reduced exercise tolerance(58) and major metabolic changes such as metabolic acidosis, a slowing of VO_2 and VCO_2 kinetics, an increase in oxygen debt, a disproportionate increase in minute ventilation increases compared to metabolic demand(47) and a sharp rise in subjective dyspnea(70). Conversely, exercise performed before VT, in the hypothetical availability of enough substrate, can theoretically be sustained indefinitely(71).

VT is usually expressed either in ml or in %VO_2max predicted, and its normal value varies with age, sex and fitness level, but usually lies between 50-60% of %VO_2max predicted(71). VT is widely regarded as one of the best estimator of overall fitness(71-74), and it is responsive to aerobic training, both in normal subjects and patients with chronic lung disease(75-79).

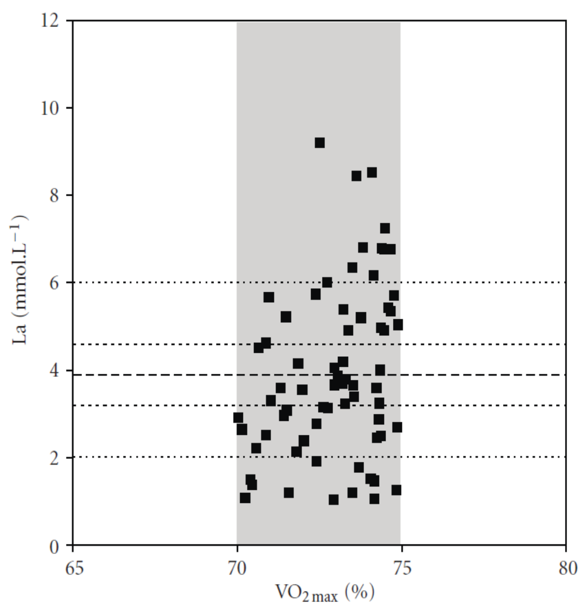

Figure 3. Mean blood lactate concentration during exercise when applying a fixed 70-75% VO_2max intensity. *From reference 81.*

One of the main clinical uses of the VT is for exercise prescription. Although the ACSM suggests heart rate (HR – in percentage of HR reserve or percentage of HR_{peak}) or a percentage of VO_2 reserve as a guide to prescribe exercise(38), evidence suggests that this may not be appropriate for all patients, especially those with heart or lung disease. In 2011, Hofmann and Tschakert(80) published a review paper highlighting the fact that using fixed percentages of either HRR or VO_{2max} to prescribe exercise can result in a wide range of different training intensities, both potentially above or below VT, which could respectively result in exercise inducing undue fatigue and intolerance (above VT) or of too low intensity to provide benefits (below VT). They suggested using an individualised threshold (either VT or AT) to prevent this. Figure 3 is taken from this article and illustrate the wide range of blood lactate values obtained in subjects training at a fixed percentage of their VO_{2max}, representing a wide array of metabolic demand.

In 2000, Zacarias and colleagues(81) studied 26 patients with COPD (mean FEV_1 49% predicted) during incremental exercise on ergocycle, with the goal of evaluating if their heart rate at VT (HR_{VT}) expressed as three different methods (%HR_{peak}, %HR predicted and %HR reserve) fell within the recommended intensity range (+/- 5% of VT). Of note, despite using the V-slope method, they could identify VT in 18 patients only. The HR_{VT} for the three methods corresponded to a wide range of exercise intensities, and to ensure that patients could be trained to +/- 5% of VT, a prescription based on HR alone would have had to be 80-85% of HR_{peak} or 40-45% HRR – which is discordant with current guidelines. The authors concluded that exercise prescription based on HR should be discouraged in COPD patients, highlighting the need for a different marker to guide exercise prescription in these patients.

More recently, Diaz-Buschmann and colleagues(82) studied if using a fixed HR value (either with the Karvonen equation or as %HRR) for exercise prescription in 159 patients on beta-blocker treatment would result in exercise at too low or too high intensity (relative to VT, determined using the VEM). They found that a significant proportion of patients would be exercising significantly passed VT, or way below it, depending on the HR method used, and that, overall, no fixed HR value resulted in a satisfactory training regimen for all patients.

In a study on the use of VT as a guide for exercise prescription in patients with COPD, Vallet and colleagues(41) randomized 20 patients with COPD (mean FEV_1 about 1.8 l) to either an eight week, four times a week active training program at HR_{VT} (using the V-slope method) or usual care. They noted increases of 25% in symptom-limited VO_2 (VO_2SL), 20% in maximal V_E, 19% VO_{2VT} and a decrease in V_E and respiratory rate (RR) for work at 50% and 75% VO_2SL. In another study(42), the same investigators randomized 24 patients with COPD (mean FEV_1 54% and 63% for both arms) to a 4 week, 5 days a week training program prescribed either using an ''individualised'' protocol (HR_{VT} using V-slope) or ''standard'' protocol (using 50% of HRR). The individualised protocol (based of HR_{VT}) resulted in significant increase in VO_2SL (20%, $p < 0.05$), VO_{2VT} (22%, $p < 0.01$) and O_2 pulse (17%, $p < 0.05$). The standard protocol (based on %HRR) resulted in a significant increase in VO_{2VT} (8%, $p < 0.05$), but significantly less so than the individualized group, and no change in VO_2SL or O_2 pulse. Of note, the actual mean HR during training for both groups was *identical* – a finding that highlights the importance of trying to identify the personalized HR_{VT} for each patient rather than aiming for a generic percentage of HR.

Serres and colleagues(43) studied the adaptation of skeletal muscle to training in 8 COPD patients after a short 3-week exercise program, at an intensity corresponding to HR_{VT}, and 6 controls. They showed that, along with a significant increase in VO_2SL and VO_{2VT}, the training group also showed better maximum voluntary contraction (MVC) of the quadriceps (+ 8%, $p < 0.05$), and critical power (+ 39%, $p < 0.05$). They concluded that an individualized training based on HR_{VT} was effective at rapidly increasing peripheral muscle performance in COPD patients.

More recently, Gimenez and colleagues(44) randomized 13 COPD patients (mean FEV_1 1.6 l) to either high-intensity training (1 minute at VO_{2peak} alternating with 4 minutes at VO_{2VT}) or moderate intensity training (40-50 W), 5 days a week for 6 weeks. The high-intensity group showed decreased dyspnea at rest ($p \leq 0.01$), decreased blood lactate levels during exercise ($p < 0.001$), increased VO_2SL, maximal inspiratory and expiratory pressures, V_E and VO_{2VT}, while decreasing V_E/VCO_2 (all $p \leq 0.01$). The moderate-intensity group only improved on the 12-minute walk test. These findings support the use of the VT as a guide to individualize training regimen in patients with COPD.

It is worth noting that, although the 2013 American Thoracic Society (ATS) guidelines on pulmonary rehabilitation recommends using the ACSM framework for exercise prescription, they acknowledge that using standard “high-intensity” (high workload) training may not be tolerable by COPD patients and that, in this context, a training program based on perceived exhaustion (Borg scale rating 4-6) is adequate(34). Coincidentally, this level of perceived exhaustion is well known to correspond to the VO_{2VT}(70).

Other than exercise prescription, the VT has many other clinical uses. It is known to be one of the best predictors of exercise endurance in patients with COPD(84) and an important prognostic marker in heart failure(85-88). In patients with primary pulmonary hypertension, VT (measured using V-slope and VEM) stands out as an independent marker of disease severity(89), and in patients with idiopathic pulmonary fibrosis (IPF), a value of $V_E/VCO_2 \geq 45$ when measured at VT was an independent predictor of the presence of systolic pulmonary hypertension and worse survival(90).

Finally, VT is a significant prognostic marker in the peri-operative context. Older and colleagues(91) showed that, in 187 elderly patients undergoing major abdominal surgery, a pre-operative $VO_{2VT} < 11$ ml/kg/min was associated with a major increase in per-operative mortality rate (18% *vs* 0.8%, $p < 0.001$). Torchio and colleagues(92) studied the outcome of 54 COPD patients after lung-resection surgery and found that a pre-operative VT < 14.5 ml/kg/min could predict severe post-operative complications with a sensitivity of 91.6% and a specificity of 97.6%. Finally, West and colleagues(93) recently showed that, when using multiple physiological parameters to predict complications following colonic resection surgery, a multivariable logistic regression model identified only VO_{2VT} and sex as reliable predictors of complications (with area under curve 0.71).

4.3 Variance in the measure of VT

4.3.1 Manual measure of VT

As mentioned previously, the identification of VT during exercise relies on a manual manipulation by an observer. When using the V-slope method, the observer has to manually draw tangent lines on the graph of VCO_2 vs VO_2 to identify the inflection point in their relationship, and when using the VEM, the precise identification of the point where V_E/VO_2 increases without change in V_E/VCO_2 can be made difficult by the inherent irregularities of the graph(63). Alterations in breathing pattern (i.e. hyperventilation or irregular breathing) are expected to increase the potential difficulty of an observer to identify a precise VT by making the relationships between the different variables less clear. In patients with COPD, these alterations could be expected to be more prominent, as anomalies in breathing pattern are ubiquitous in this disease, especially during exercise(23, 34, 36, 49, 94, 95). Although the V-slope method was initially presented as having the advantage of being independent of V_E(57), evidence suggests that: 1) both in healthy and COPD patients, the V-slope sometimes fails to identify VT(81, 96) and 2) as discussed in section 4.2.2.2, VCO_2 *measured at the mouth* is dependent on V_E(61, 63), and therefore the V-slope method can be expected to be influenced by anomalies in V_E. Thus, the inter-observer variance in the measure of VT has the potential of being large in the presence of ventilatory anomalies – a concerning finding that could pose problem when trying to use the VT for clinical purposes. Despite these implications, relatively few studies have evaluated this question.

Yeh and colleagues(51) studied the inter-observer variance of the VEM on 8 healthy subjects undergoing incremental exercise testing using measures made by 4 experienced physiologists. The mean standard deviation of the VT value for each subject was 8% of VO_{2peak}, equivalent to +/- 289 ml. For one subject, the range between the largest and smallest determination of VO_{2VT} reached 890 ml/min (24% VO_{2peak}). They concluded that their results cast doubt on the ability of the VEM to reliably identify VT.

Gladden and colleagues(60) used nine experienced observers to determine the inter- and intra-observer reliability of the VEM on 24 normal exercise tests. They found a median inter-observer correlation coefficient of 0.70 and an excellent intra-observer correlation (analysis of duplicate tests) of 0.97. Another study from 1987(97) used 6 healthy subjects who each performed CPET six times, and evaluated the capacity of V_E, VCO_2, R and V_E/VO_2 (VEM) to reliably identify VT. The mean observer error in identifying VT from the tests was 24% for V_E, 19% for VCO_2, 29% for R and 15% for the VEM. For the VEM, which stood out as the most precise marker among observers, the error rate corresponded to a mean difference in VO_{2VT} of 625 ml/min.

Shimizu and colleagues(98) studied the variance between three observers in the measure of VT using the V-slope and the VEM in 17 patients with heart disease and six normal controls. The maximal inter-observer difference in VO_{2VT} using both methods was 70 ml/min, and the overall intraclass correlation coefficient for VO_{2VT} among the reviewers was 0.60. V-slope was consistently associated with better agreement between observers.

More recently, Filho and colleagues(99) evaluated the inter-observer variability of the VT during incremental exercise in 14 healthy volunteers. Unusually, they used a single composite measure of VT per observer, derived from the mean of their estimation of V-slope, VEM, V_E x time and R x time. Overall, the mean inter-observer difference in the composite measure of VO_{2VT} was 140 ml (equivalent to 2 ml/kg/min). Another group from the United Kingdom(100) recently evaluated the V-slope method using nine different observers, on 21 incremental exercise tests from patients undergoing pre-operative evaluation. The technical error of measurement across observers (a measure of bias + random error) was 8.1% (i.e 0.9 ml/kg/min).

Finally, the only study investigating the inter-observer reliability of the determination of VT in COPD patients was presented by Belman and colleagues in 1992(101). They investigated the variance in the determination of VT between two observers using V-slope and VEM on 29 COPD patients (mean FEV_1: 40% predicted), on two separate exercise tests. The inter-observer Pearson correlation using VEM was 0.79 for the first test (although analyses were performed only on 11 subjects because VT could not be identified by both observers in 18 cases) and 0.77 for the second (11 patients analysed), and 0.97 for the first test using V-slope (on 9 patients) and 0.98 for the second (on 7 patients). This suggests the V-slope is better than VEM in patients with COPD, although the majority of patients had to be excluded from both analyses (VEM and V-slope) because their VT could not be defined by both observers.

In summary, the available evidence, although heterogeneous and prone to methodological errors, suggests a wide range of inter-observer reliability when evaluating VT. This may be due to differences in the population studied, and in the choice of technique used to estimate VT. These results are difficult to generalize and command additional research, especially in patients with COPD, where the link between inter-observer variance and disease severity and other commanding factors remains to be quantified.

4.3.2 Computerized measure of VT

To try to eliminate random measurement bias from human observers, a few studies have evaluated the value of automatic computerized algorithms for the identification of VT.
Orr and colleagues(102) were the first to propose a computer algorithm, based on the identification of the breaking point in the increase in V_E when plotted against VO_2. The program is instructed to minimize the pooled residual sum of squares when analyzing the best-fit regression model of that graph, and VT is reported as the first break in that model. They compared the automatic VT determination to the mean of VT (estimated using VE as well) by four human observers, on 37 exercise tests, and found a correlation coefficient of 0.94 between the two measures (absolute mean difference 50 ml/min). The study by Gladden(60) mentioned earlier also included an analysis VT by a computer (with the same protocol as Orr and colleagues) and, when compared to the mean of 9 human observers (that used a different method: the VEM) showed poor correlation (0.58).
A study by Solberg and colleagues(103) used 3 computerized algorithm (V-slope, VEM and R) on 12 healthy subjects and found that R had the best correlation to serum lactate concentration (AT). Finally, a recent study by Ekkekakis and colleagues(104) evaluated nine different computer protocols to estimate VT in healthy patients and showed that, although mean correlation between methods was relatively good (0.76-0.81), the absolute differences in VO_{2VT} derived from the different protocols were often larger than 500 ml/min.

Overall, although the initial study by Orr was promising, the advent of many different algorithms for VT detection seems to have complicated, rather than simplified, the question of automatic VT determination. It seems unlikely that an automated analysis of variables as volatile and fluctuating as the ventilatory parameters during exercise will supplant a human analysis, which, although burdened with its share of bias, always leaves place to clinical judgment.

5. Rationale and objectives

5.1 Rationale

The ventilatory threshold is widely accepted as a marker of aerobic fitness. It is commonly used for exercise prescription, assessment of response to an intervention and as a prognostic marker in many diseases. Its measurement relies on the accurate identification of an inflection point in respiratory kinetics during aerobic exercise. Patients with COPD greatly benefit from aerobic training. However, these patients suffer from chronic airflow limitation that may alter their ventilatory kinetics and impair our ability to reliably identify their VT. The exact magnitude of this effect and its relationship to the severity of the disease are yet unknown.

5.2 Objectives

5.2.1 Primary research objective:

To quantify and compare the inter-observer reliability of human observers in determining VT in control subjects and COPD patients, and to compare the performance of the V-slope and VEM methods.

5.2.2 Secondary research objectives:

1. To compare human versus computerized analyses of VT.
2. To determine if the inter-observer variation in the identification of the VT corresponds to a clinically significant difference in the determination of HR_{VT}.

6. Hypotheses

We expected that COPD would have a negative impact on the reliability of the determination of the VT. More specifically, we hypothesized that:

1. The human inter-observer variation in the identification of the VT would be greater for patients with COPD than for healthy controls.

2. The inter-observer reliability would be better with the V-slope method than with the ventilatory-equivalent method, both for patients with COPD and healthy controls.

3. There would be a significant difference in the identification of the VT by clinicians compared with computerized analysis, both for the V-slope method and the VEM.

4. Markers of airway obstruction and COPD severity (severity of airflow obstruction, presence of chronic hypercapnic respiratory failure and presence of significant exercise desaturation) would predict a larger inter-observer variation in the measure of the VT.

5. The inter-observer variation in the measure of the VT would correspond to a clinically significant difference in the measure of HR_{VT} (greater than 5 bpm).

7. Article: Reliability of the determination of the ventilatory threshold in patients with COPD

Primary author

Bruno-Pierre Dubé[1,2]

Co-authors

Myriam Mesbahi[2]

François Beaucage[3]

Véronique Pepin[1,4]

Affiliations:

[1]Department of Exercise Science, Concordia University, Montreal, Canada

[2]Faculty of Medicine, University of Montreal, Montreal, Canada.

[3]Sacré-Cœur Hospital, Montreal, Canada.

[4]Research center, Sacré-Cœur Hospital, Montreal, Canada.

Author contribution:

BPD and VP designed the study. BPD coordinated the study. BPD, MM and FB were responsible for patient screening and selection. BPD and MM performed ventilatory threshold measurements. BPD, MM, FB and VP analyzed the data. BPD and VP wrote the manuscript. All authors had full access to all of the study data, contributed to draft the manuscript and revised it critically for important intellectual content, approved the final version of the manuscript, and take responsibility for the integrity of the data and the accuracy of the data analysis.

Keywords: V-slope, ventilatory equivalent, exercise testing, anaerobic threshold, exercise physiology

Published in: *Medicine and Science in Sports and Exercise* 2016, 48(2):210-8

7.1 Abstract

Purpose

The purpose of this study was to determine the inter-observer reliability of the assessment of the ventilatory threshold (VT) using two methods, in COPD patients and in control subjects.

Methods

VT was identified from incremental exercise testing graphs of 115 subjects (23 controls and 23 in each COPD GOLD class) by two human observers and a computer analysis, using the V-slope method and the ventilatory equivalent method (VEM). Agreement between observers in identifying VO_2 at VT (VO_{2VT}) and heart rate at VT (HR_{VT}) across disease severity groups were evaluated using intra-class correlation (for humans) and Passing-Bablok regression analysis (human *vs* computer).

Results

For human observers, ICCs (95% confidence interval) in determining VO_{2VT} were higher in controls [0.98 (0.97-0.99) both with V-slope and with VEM] than in COPD patients [0.72 (0.60-0.81) with V-slope and 0.64 (0.50-0.74) with VEM]. Passing-Bablok analysis showed that human and computerized determination of VO_{2VT} were interchangeable in controls, but not in COPD patients. FEV_1 and peak minute-ventilation during exercise were the only variables independently associated with greater inter-observer differences in VO_{2VT}. Inter-observer differences in HR_{VT} ranged from 2±1 beats/minute (controls) to 10±3 beats/minute (GOLD 4).

Conclusions

In COPD patients, the reliability of human estimation of VO_{2VT} is less in than in controls and not interchangeable with a computerized analysis. This should be taken into account when using VT for exercise prescription, as a tool to monitor response to an intervention, as a surrogate measure of overall aerobic fitness, or as a prognostic marker in COPD patients.

7.2 Introduction

The concept and definitions of anaerobic and ventilatory thresholds have sparked considerable literature and debate over the years. One of the reasons for this may reside in the lack of a consensual definition of both concepts and the proliferation of terms used to describe them (105). The anaerobic threshold can be defined as the oxygen consumption above which aerobic metabolism is supplemented by anaerobic mechanisms, and after which a progressive increase in blood lactate concentration and metabolic acidosis occur (106). The onset of this pivotal event can be estimated using non-invasive techniques based on the non-linear evolution of carbon dioxide production (VCO_2) and minute-ventilation (V_E) relative to VO_2 during incremental exercise (the so-called ventilatory threshold – VT). In particular, the breaking point in the VCO_2-VO_2 relationship (V-slope method) (57) and the moment at which there is a rise in the ventilatory equivalent for O_2 (V_E/VO_2) without a concomitant rise in ventilatory equivalent for VCO_2 (V_E/VCO_2) (the ventilatory equivalent method – VEM) (54) have been used to identify VT. Exercise above the VT is associated with reduced exercise tolerance, metabolic acidosis, a slowing of oxygen consumption (VO_2) and CO_2 production (VCO_2) kinetics (47), and a sharp rise in dyspnea (70).

In the clinical setting, VT is used as a predictor of overall aerobic fitness and is responsive to training, both in healthy subjects and in patients with chronic diseases (42, 107). It is a prognostic marker in chronic cardio-respiratory diseases (85-87) and in the perioperative period (92, 93). Because of its close relationship with overall exercise tolerance, the VT is also useful as a tool for exercise prescription. Patients with chronic obstructive respiratory disease (COPD) greatly benefit from exercise training, and pulmonary rehabilitation has become a standard of care in the management of these patients (34). Compared with healthy individuals however, patients with COPD exhibit a marked reduction in exercise tolerance caused in part by expiratory flow limitation and dynamic hyperinflation, increased work of breathing, abnormal breathing pattern, high V_D/V_T ratio and gas exchange anomalies (36, 108), and as such, may be unable to tolerate prolonged high-intensity training. Training programs using the VT as a tool to guide exercise intensity have been safely and successfully used in this population (41-44, 109), and additional data suggest that such an ‘’individualised’’ prescription may offer an advantage over ‘’interval-based’’ regimen (80-82).

Both the V-slope and the VEM rely on a manual manipulation by an observer or an automated computerized analysis, and as such are prone to variation and error. The presence of ventilatory and gas exchange anomalies in patients with COPD may further impair the reliable identification of VT using these techniques. A large intra- or inter-observer variation may have consequences when using the VT for exercise prescription, when monitoring response to training or when performing prognostic evaluation in patients undergoing surgery or with heart failure that have concomitant COPD. In healthy subjects, the high intra-observer reliability of the measurement of the VT has already been demonstrated (60), but the inter-observer reliability showed more heterogeneous results (51, 60, 97, 99). In patients with COPD, one study (101) showed acceptable inter-observer variability in the identification of the VT, but was limited by its small sample size and the lack of details regarding the clinical characteristics of the patients included.

We hypothesized that COPD severity would negatively impact the inter-observer reliability of the identification of the VT as determined from the V-slope method and VEM. In accordance, the aims of this study were: 1) to quantify the reliability of human observers in determining VT in control subjects and COPD patients; 2) to compare human versus computerized analyses of VT, and 3) to evaluate if the inter-observer difference in VT identification amounts to a clinically significant difference in the corresponding heart rate (HR_{VT}).

7.3 Methods

This study was based on an analysis of incremental exercise test data from individuals who completed an exercise test in the respiratory physiology laboratory at l’Hôpital du Sacré-Coeur de Montréal. Data from all pulmonary function tests, exercise tests, and blood gas analyses performed since March 2010 are stored in a common database located on a stand-alone computer in the physiology laboratory. Data for both COPD patients and controls were extracted from this database. The study was approved by the institutional ethics committee.

7.3.1 Subjects

A convenience sample of individuals with COPD and controls was selected from the aforementioned database. For patients with COPD, inclusion criteria were: age ≥ 40 years, a history of smoking of at least 20 pack-years, an objective diagnosis of COPD (as assessed by clinical evaluation and a spirometry result showing a post-bronchodilator FEV_1/FVC ratio less than 0.70) and an exercise test duration time of at least 6 minutes. This last criterion was implemented to maximise the chance of observing a VT.

With the assumption that patients with GOLD 4 disease would be less represented in the database, they were selected first. The database was then screened to identify, for each very severe patient, a matching subject amongst all other severity groups and amongst controls. Matching was based on age (+/- 4 years), sex, and body mass index (BMI, +/- 4 kg/m^2). Control subjects were defined as individuals with normal resting pulmonary function tests, normal VO_{2peak} (i.e. ≥ 85% VO_{2max} predicted) and normal cardiorespiratory response to exercise, and were matched to COPD patients for age, sex and BMI (see above). Reasons for referral to CPET in control patients were: unexplained dyspnea on exertion (16 patients), pre-operative evaluation (5 patients) and lung cancer (2 patients).

Subjects were excluded from the study if their medical file suggested clinical disease worsening or a respiratory exacerbation in the four weeks preceding the exercise test, evidence of another condition that could limit exercise performance (asthma, unstable coronary heart disease, heart failure, cancer, symptomatic peripheral vascular disease or significant osteoarthritis), long-term oxygen therapy, or incomplete baseline evaluations.

Twenty-three patients with GOLD 4 disease meeting inclusion criteria and having a suitable match in all other disease severity groups were identified and included in the study (total sample 115 patients). Based on the results of the means and standard deviations of the first 65 patients studied, a sample size of 22 patients per disease subgroup was necessary to identify a difference on 100 ml/min in VO_{2VT} between controls and each COPD group with a power on 80% and α-level of 0.05.

7.3.2 Baseline measurements

Demographic and clinical information were collected from medical files. These include age, sex, BMI, ethnicity, current medication, and self-reported smoking status. Lung function was assessed using spirometry (for expiratory flow rates), body plethysmography (for lung volumes) and single breath-hold technique (for lung diffusion capacity for carbon monoxide). All tests were performed and interpreted according to American Thoracic Society guidelines in a laboratory at sea level.

7.3.3 Exercise testing

Symptom-limited incremental exercise tests were performed according to published guidelines (110). More specifically, tests were realized on an electromagnetically braked cycle ergometer (Ergoline 200, Ergoline, Bitz, Germany), with a protocol including two minutes of rest and a three-minute period of initial unloaded cycling. Load was increased linearly until exhaustion (ramp was individually determined for each patient by the attending physician, based on either previous exercise testing result or expected maximal work rate as estimated by overall physical fitness and/or FEV_1) with the goal of maintaining a cycling speed of 60 revolutions per minute. Breath-by-breath analysis of expired gases was performed using electronic analysis (Jaeger Oxycon Pro, CareFusion, Hoechberg, Germany). V_E, VO_2, VCO_2, V_E/VO_2 and V_E/VCO_2 were computed using twenty-second averages of breath-by-breath values. Peak VO_2 was the highest 20-second mean VO_2 obtained. Patients using beta-blockers were not required to withhold them before performing CPET. Oxygen saturation was monitored using finger or ear pulse oximetry. Exercise capacity was defined as the highest work rate achieved for at least 20 seconds at a rate of at least 50 revolutions per minute. Arterial blood gases were assessed at baseline using a standard blood gas analyzer (ABL800 Flex, Radiometer, Copenhagen, Denmark). Dyspnea and leg fatigue were evaluated at rest and at maximal exercise intensity using the modified 10-point Borg scale (111). Additional information regarding internal quality control can be found in the "Methods" section of the supplemental digital content file (Appendix A).

7.3.4 Ventilatory threshold

For all patients, two manual methods were used to identify the VT: 1) the V-slope method and 2) the VEM. To optimize the validity of the V-slope method, care was taken to ensure that the ranges of VO_2 and VCO_2 in the plots were equal (57), and that the VO_2 scale was adequate to allow a precise identification of VO_{2VT}. In addition, a computer-generated analysis of VT (LABManager, version 5.3.0.4, CardinalHealth, Hoechberg, Germany) was used, again using both the V-slope method and the VEM. Human observers took care to exclude the first minute of exercise from analysis in order to avoid confounding VT with a "pseudo-threshold" (112) sometimes associated with hyperventilation at the onset of exercise. For all subjects, computerized determination of VT was allowed between the onset and the termination of incremental exercise; the warm-up and recovery periods were excluded from analysis by the software. In the event where a VT could not be identified, it was reported as ''undetermined''.

The VT was reported both as the VO_2 at which it occurred (in absolute value) and as the corresponding heart rate (HR_{VT}). Graphs for VT analysis for both methods were extracted from the database by a research assistant unrelated to the study, coded, duplicated, and submitted to two observers (B.P.D and M.M.), who blindly recorded the presence or absence of a VT, its value in millilitres of VO_2, and the corresponding heart rate. The graphs used by the human observers were identical to the ones used for computerized analysis. Precise and identical instructions on how to identify VT using both the V-slope method and the VEM were given to both observers. For the V-slope method, VT was defined as "the breaking point in the line of the graphical representation of VCO_2 against VO_2" (57). For the VEM, VT was defined as "the point where V_E/VO_2 begins to increase while V_E/VCO_2 remains stable, when both are plotted against VO_2" (54).

To test for internal validity, a subsample of 50 graphs drawn randomly from COPD and controls was blindly resubmitted to the two observers for a second VT determination. Both observers were physicians with formal medical training in respiratory medicine, specific training in exercise physiology, but less than 5 years of clinical experience. VT analyses were performed in an independent and blinded manner.

7.3.5 Statistical analyses

Agreement between the human observers in the determination of $VO2_{VT}$ was assessed using intra-class correlations coefficients (ICC 2,1 – two-way random single measure). Reliability using ICC was interpreted according to the following scale: virtually none for ICC ≤ 0.10, slight for ICC 0.11 – 0.40, fair for ICC 0.41 – 0.60, moderate for ICC 0.61 – 0.80 and substantial for ICC ≥ 0.81 (113).

To test whether human and computer analysis of $VO2_{VT}$ are interchangeable, Bland-Altman graphical analysis and Passing-Bablok regression analysis was performed (114). This non-parametric statistical tool allows the estimation of the interchangeability of two analytical methods and of the possible bias between them. It provides a numerical quantification of agreement levels and does not make any assumption about the distributions of the samples of their measurement errors and is non-sensitive to outliers. It does however require that data be continuously distributed and linearly related.

The mean differences between the two human observers' assessment of VO_{2VT} were compared across disease category groups using one-way ANOVA with post-hoc Bonferroni correction.

A stepwise multiple linear regression analysis that included baseline demographic data and pulmonary function and exercise test results was performed to identify independent predictors of a larger inter-observer difference in VO_{2VT}.

The mean inter-observer difference in HR_{VT} for each of the five subgroups was compared using one way ANOVA with post-hoc Bonferroni correction. An empirical threshold of +/- 5 bpm (total range of 10 bpm) was chosen as the cut-off for clinical significance for this parameter, as we believe that when exercise training is based on a target heart rate value, the training heart rate should stay within this limit of the objective.

Intra-observer reliability was assessed using intra-class correlation coefficients. All analyses were performed using SPSS version 21 (Chicago, IL, USA) and MedCalc (MedCalc Software, Ostend, Belgium). In all instances, a p-value of less than 0.05 was considered as the threshold for statistical significance.

7.4 Results

The clinical characteristics of the 115 subjects are summarized in Table 1. Most were males (70%) and sex, age, and BMI were evenly distributed across subgroups, as projected by the recruitment design. Exercise performance evolved as expected with increasing COPD severity, with ventilatory limitation and gas exchange abnormalities becoming prominent in GOLD 3 and 4 patients.

Table 1. Baseline patient characteristics.

	Controls n=23	GOLD 1 n=23	GOLD 2 n=23	GOLD 3 n=23	GOLD 4 n=23
Sex (number of males)	16	16	16	16	16
Age (y)	55 (7)	56 (7)	56 (7)	56 (5)	57 (6)
BMI (kg/m^2)	27 (4)	26 (5)	26 (5)	26 (5)	25 (6)
Beta-blocker use, n	1	6	4	7	7
Pulmonary function tests					
FEV_1/FVC	76 (3)	64 (3)	57 (8)	41 (7)	36 (7)
FEV_1 (l)	3.49 (0.63)	2.65 (0.55)	2.21 (0.59)	1.23 (0.33)	0.95 (0.33)
FEV_1 (% pred.)	111 (15)	86 (6)	70 (7)	39 (6)	28 (2)
FVC (l)	4.57 (0.83)	4.04 (0.85)	3.95 (1.22)	2.96 (0.54)	2.55 (0.67)
FVC (% pred.)	119 (18)	106 (8)	102 (19)	77 (14)	63 (12)
FRC (% pred.)	99 (18)	108 (19)	111 (24)	149 (25)	155 (50)
TLC (% pred.)	113 (14)	107 (10)	110 (18)	119 (18)	112 (37)
RV (% pred.)	93 (16)	104 (22)	115 (29)	179 (38)	188 (69)
D_LCO (% pred.)	92 (10)	73 (18)	69 (18)	52 (12)	43 (10)
Resting hypercapnia (n)	0	0	0	4	5
Incremental exercise test					
Ramp (W/min)	14 (3)	12 (5)	11 (4)	9 (5)	7 (4)
Peak Power (W)	156 (56)	121 (62)	100 (46)	62 (34)	45 (24)
Peak Power (% pred.)	108 (28)	82 (24)	70 (16)	44 (14)	33 (12)
Peak Heart rate (bpm)	157 (14)	138 (21)	137 (22)	124 (14)	119 (11)
Peak Heart rate (% pred.)	92 (8)	82 (12)	81 (14)	74 (8)	71 (6)
Peak VO_2 (l/min)	2.20 (0.67)	1.74 (0.64)	1.63 (0.61)	1.27 (0.37)	1.07 (0.31)
Peak VO_2 (% pred.)	110 (23)	86 (18)	81 (16)	63 (13)	50 (10)
Peak V_E (l/min)	82 (28)	69 (23)	63 (19)	46 (15)	36 (12)
Peak V_E (% pred.)	63 (13)	74 (20)	80 (16)	104 (26)	111 (22)
Exercise desaturation (n)	0	1	1	2	6

All data presented as mean (standard deviation) unless stated otherwise. BMI = body mass index. FEV_1 = Forced expiratory volume in 1 second. FVC = forced vital capacity. FRC = functional residual capacity. TLC = total lung capacity. RV = residual volume. D_LCO = Diffusion capacity of the lung for carbon monoxide. $PaCO_2$ = arterial partial pressure of carbon dioxide. VO_2 = oxygen uptake. V_E = minute-ventilation. Resting hypercapnia = resting PaCO2 $\geq$ 45 mmHg. Exercise desaturation = a decrease in > 4% saturation during exercise.

7.4.1 Agreement in the determination of VT for human observers

There were no instances of "undetermined" VT.

Table 2 shows the agreement in the determination of $VO2_{VT}$ between human observers, assessed using ICC. Overall, reliability between human observers was higher in control subjects than in patients with COPD: in control subjects, ICC was 0.98 with V-slope and 0.98 with VEM, whereas in patients with COPD as a whole, ICC was 0.72 with V-slope and 0.64 with VEM. The 95% confidence intervals of ICCs of the controls and COPD patients were mutually exclusive. There was also a progressive decline in agreement between the two observers with increasing disease severity. In patients with GOLD 4 disease, agreement reached only "slight" levels.

Table 2. Agreement of human observers in the determination of VO_{2VT}.

	V-slope			VEM		
	ICC	95% CI	p	ICC	95% CI	p
Controls	0.98	0.97 – 0.99	< 0.001	0.98	0.97 – 0.99	< 0.001
All COPD	0.72	0.60 – 0.81	< 0.001	0.64	0.50 – 0.74	< 0.001
GOLD 1	0.92	0.83 – 0.96	< 0.001	0.94	0.86 – 0.96	< 0.001
GOLD 2	0.78	0.53 – 0.90	< 0.001	0.70	0.42 – 0.86	< 0.001
GOLD 3	0.68	0.38 – 0.85	< 0.001	0.42	0.03 – 0.70	0.02
GOLD 4	0.35	-0.34 – 0.66	0.04	0.15	-0.23 – 0.51	0.23

p-values refer to individual intra-class correlations. ICC=intraclass correlation coefficient; CI=confidence interval; VEM=ventilatory equivalents method; COPD=chronic obstructive pulmonary disease; GOLD=Global initiative for Obstructive Lung Disease.

ANOVA analysis revealed that the mean absolute differences in the measures of VO_{2VT} using V-slope and VEM were statistically greater in COPD patients compared with controls, and that this difference increased with severity (table 3).

Table 3. Mean (SD) inter-observer difference in $VO2_{VT}$ (ml/min) according to disease severity.

	V-slope	p-value	VEM	p-value
Controls	42 (26)	-	41 (26)	-
All COPD	189 (115)	<0.001	204 (117)	<0.001
GOLD 1	111 (33)	0.12	94 (45)	0.29
GOLD 2	165 (100)	<0.001	194 (103)	0.001
GOLD 3	209 (125)	<0.001	222 (86)	<0.001
GOLD 4	270 (120)	<0.001	307 (112)	<0.001

VEM: ventilatory equivalent method. GOLD: Global initiative for chronic Obstructive Lung Disease. $VO2_{VT}$: oxygen uptake at the ventilatory threshold.

Comparison of human and computer observers in the determination of VO_{2VT}

E-table 1 and e-figure 1 of the supplemental digital content file (Appendix A) describe the results of the Passing-Bablok regression analysis comparing each human observer to the computerized analysis. In short, for both human observers, the relationship of VO_{2VT} with the computerized analysis did not differ from linearity, confirming that the data can be used in Passing-Bablok analysis. Using V-slope, VO_{2VT} values from human observer 1 were interchangeable with computer analysis for controls, but not for patients with COPD. Similar results were obtained using the VEM. In an identical manner, observer 2 was found to be interchangeable with computerized analysis when evaluating controls, but not COPD patients. Additional description of the Passing-Bablok regression analyses can be found in the supplemental digital content file (Appendix A).

Bland-Altman plots for both methods are shown in figure 4, and similarly show that, although most data points remain inside of the limits of agreement, COPD patients generally have greater inter-observer differences and wider dispersion of values than control subjects.

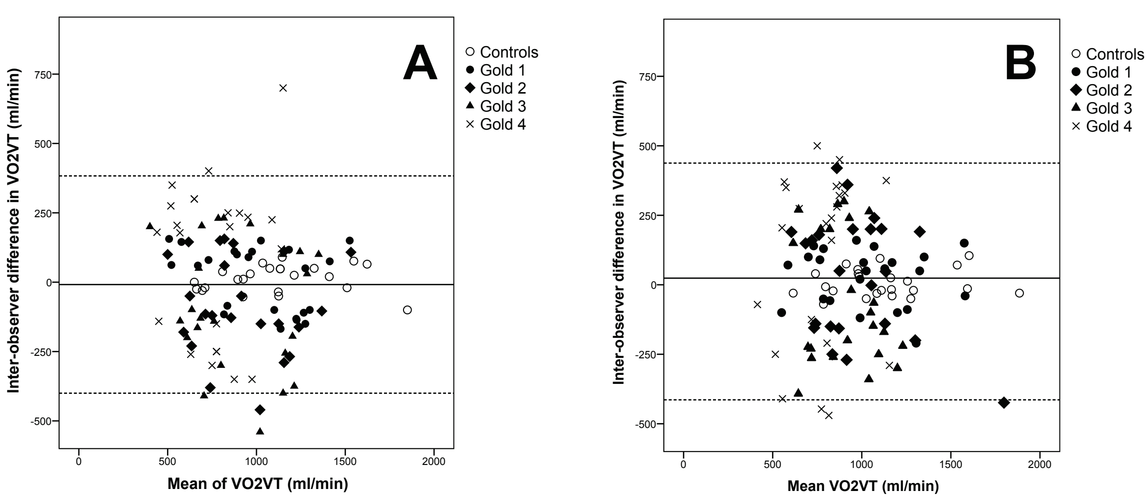

Figure 4. Bland-Altman procedures for inter-observer differences in $VO2_{VT}$, when using (A) the V-slope method and (B) the VEM. The horizontal lines represent the average inter-observer difference in $VO2_{VT}$ (center) and 95% limits of agreement (top and bottom), calculated as: mean difference ± 1.96 standard deviation of the difference.

7.4.2 Inter-observer differences in HR_{VT}

Table 4 summarises the inter-observer differences in the evaluation of HR_{VT} expressed both as absolute values and as a percentage of the peak heart rate attained during incremental exercise testing. Compared with controls using ANOVA, there was a statistically significant gradual increase in the inter-observer difference of HR_{VT} with disease severity. On average, only patients in the most severe COPD subgroup reached the pre-specified threshold of clinical significance (+/- 5 bpm). For each subgroup of patients, there were no significant differences in the mean inter-observer difference in HR_{VT} between patients with and without beta-blockers (complete data can be found in e-table 2 of the supplemental digital content file – Appendix A).

Table 4. Mean (SD) inter-observer difference in the measure of HR $_{VT}$ (absolute value and %peakHR) using two methods, by COPD severity.

	HR_{VT} - beats per minute				**HR_{VT} - %peakHR**			
	V-slope	p	VEM	p	V-slope	p	VEM	p
Controls	2 (1)	-	2 (2)	-	1.3 (0.7	-	1.4 (1.1)	-
All COPD	6 (4)	<0.001	7 (3)	<0.001	4.9 (3.1)	<0.001	5.6 (2.7)	<0.001
GOLD 1	3 (2)	0.99	4 (1)	0.01	2.1 (1.4)	0.99	3.0 (1.0)	0.01
GOLD 2	5 (2)	0.02	6 (2)	<0.001	3.5 (1.9)	0.002	4.6 (1.8)	<0.001
GOLD 3	7 (3)	<0.001	8 (2)	< 0.001	5.0 (2.6)	<0.001	6.3 (1.8)	< 0.001
GOLD 4	10 (3)	<0.001	10 (3)	<0.001	8.3 (2.3)	<0.001	8.6 (2.3)	<0.001

HR_{VT}=heart rate at the ventilatory threshold; VEM=ventilatory equivalent method; GOLD=Global initiative for chronic Obstructive Lung Disease.

7.4.3 Predictors of a larger inter-observer difference in VO_{2VT}

Table 5 describes the results of the stepwise multiple linear regression analysis. With V-slope, FEV_1, %predicted and peak minute-ventilation were the two sole independent predictors of a larger inter-observer difference in VO_{2VT} (R^2=0.41), whereas with VEM, only FEV_1 %predicted reached statistical significance (R^2=0.50).

Table 5. Multiple linear regression models for the prediction of a larger inter-observer difference in VO_{2VT} when using V-slope or VEM.

	V-slope		**VEM**	
	β	p	β	p-value
FEV_1 (%pred)	**-0.431**	**<0.001**	**-0.708**	**<0.001**
PeakV$_E$ (%pred)	**0.268**	**0.007**	0.015	0.87
Age	-0.039	0.60	0.029	0.66
Gender	0.097	0.20	-0.024	0.72
BMI	-0.006	0.94	-0.002	0.97
FEV_1/FVC	-0.098	0.58	-0.071	0.65
FVC (%pred)	0.114	0.42	0.012	0.93
FRC (%pred)	-0.003	0.98	0.066	0.43
TLC (%pred)	0.022	0.76	0.050	0.46
RV (%pred)	0.029	0.78	0.079	0.39
D_LCO (%pred)	-0.014	0.90	0.048	0.64
Hypercapnia	-0.118	0.13	-0.096	0.18
Peak workrate (W)	0.053	0.62	0.046	0.61
Peak HR (% pred)	-0.090	0.33	0.117	0.16
Peak VO2 (% pred)	0.060	0.68	-0.052	0.87
Desaturation	-0.110	0.17	0.108	0.46

VO_{2VT}=Oxygen uptake at ventilatory threshold; VEM=ventilatory equivalents method; FEV_1=Forced expiratory volume in 1 second; V_E=minute-ventilation; BMI=body mass index; FVC=Forced vital capacity; FRC=Functional residual capacity; TLC=Total lung capacity; RV=Residual volume; D_LCO=Diffusion capacity for carbon monoxide; Hypercapnia=Baseline $PaCO_2$ > 45 mmHg; W=Watt; HR=Heart rate; VO_2=oxygen uptake; Desaturation=Decrease of > 4% in saturation during exercise; %pred=percent of predicted value.

7.4.4 Internal validity

Intra-observer ICC measured on a subset of 50 patients showed relatively high reliability throughout the spectrum of disease severity (complete data shown in e-table 3 of the supplemental digital content file – Appendix A). For both observers and for both methods of observation, ICCs across disease severity groups were all higher than 0.81.

7.5 Discussion

To our knowledge, this is the first study to report on a direct evaluation of the reliability of human and computerized identification of VO_{2VT} and HR_{VT} across COPD severity groups. Our main results indicate that 1) reliability of human observers in the determination of VO_{2VT} is lower in patients with COPD than in controls, for both the V-slope and VEM methods; 2) human and computerized analyses of VO_{2VT} are interchangeable in controls, but not in patients with COPD; 3) FEV_1 (percent predicted) is an independent predictor of a larger inter-observer difference in measurement of VO_{2VT}

(with peak minute-ventilation also being significant for V-slope) and 4) compared to controls, the increasing inter-observer disparity in VO_{2VT} assessment in COPD patients corresponds to a gradually larger difference in the estimation of $HR_{VT.}$ These combined findings suggest that the baseline airflow obstruction and subsequent abnormalities in the ventilatory response of patients with COPD during exercise may be a causative factor in the increasing variance of inter-observer assessment of VO_{2VT} (21, 108). This supports a common impression amongst clinicians that, when represented graphically, the ventilatory parameters of patients with COPD produce more irregular and noisy patterns. Coupled with the fact that these patients show higher-than-predicted ventilation for any work rate, these anomalies seem to hinder the precise identification of a breaking point in the kinetics of ventilatory variables.

The available literature on this subject is scarce, especially in patients with COPD, and has produced inconsistent results. Our results for control subjects are similar to those of Gladden et al. (60), who showed that, in healthy volunteers, the intra-observer reliability of the VEM was high (ICC=0.97), that the inter-observer reliability (tested on nine observers) was lower (ICC=0.70) and that agreement between a human observer and a computerized value of VO_{2VT} was only moderate (ICC=0.58). Filho et al. (99) also described similar results in a sample of 14 healthy subjects. In contrast, Garrard et al. (97) showed a higher intra-observer error when assessing VO_{2VT} in healthy subjects. In this study, inter-observer error reached 29 and 24% using plots of the respiratory exchange ratio and $V_{E,}$ but the V-slope and VEM performed better (19% and 15% error, respectively). Yeh et al. (51) described a mean range of 560 ml/min among four observers trying to identify VO_{2VT} using the VEM in healthy subjects. This is much larger than the difference found in our study in control subjects (44 ml/min).

Our results seem in line with those of Belman et al. (101) who studied the intra- and inter-observer reliability of the determination of VT in patients with COPD using the V-slope method and the VEM, on two separate exercise tests. They reported excellent intra-observer reliability for both method (Pearson correlation 0.97 and 0.99 for the two observers) and good inter-observer reliability for all methods (Pearson correlations all higher than 0.74). However, their analysis was performed on a small, uncharacterized subset (n=14 at the maximum) of their overall cohort, which contained subjects with widely variable FEV_1 values. In addition, the use of Pearson correlation to assess agreement between observers is often inappropriate (115). Our study used a larger, matched, well-characterized population and adds the findings of a progressive decline in inter-observer reliability with disease progression and the poor relationship between human and automated analysis in patients with COPD.

The clinical importance of the magnitude of inter-observer differences identified can be put into perspective by comparing it to reported improvements in VO_{2VT} following an exercise-training program. In patients with moderate to severe COPD, prior studies have documented improvements in VO_{2VT} ranging from approximately 83 to 350 ml/min following training (41-44). In our study, for moderate to very severe COPD patients, inter-observer differences in VO_{2VT} ranged from 165 to 270 ml/min using the V-slope and from 194 to 307 ml/min using the VEM. It is therefore likely that inter-observer differences in VT determination have an impact on the evaluation of changes in VO_{2VT} following an exercise-training program. In contrast, agreement for control subjects was much better (less than 50 ml/min difference in VO_{2VT}) suggesting that inter-observer differences in VT play a lesser role in this population (116). In addition, our findings concerning the low inter-observer reliability of the determination of VO_{2VT} should raise caution when using VO_{2VT} or V_E/VCO_2 at VT as a prognostic marker in patients with heart failure or undergoing surgery if these subjects also have concomitant COPD.

Data concerning the reliability of computerized measurements of VT is limited. Most manufacturers of exercise testing equipment provide a unique software algorithm and these different equations have been shown to provide variable estimates of VT, both when using V-slope and the VEM (104). Any comparison of results originating from different software calculations must therefore be made with caution. Our data show that for control subjects, both human observers could be considered interchangeable with computer analysis, which is in line with the results of Santos et al (117). In COPD patients however, human and automatic analysis were not interchangeable owing to significant systematic and proportional differences. This sheds an interesting light on the use of these computerized algorithms in daily practice, and clinicians may want to take it into account when assessing VO_{2VT} using only automated reported values. Indeed, we believe these findings emphasize the need for clinicians to manually confirm any automated measurements of VT.

Our choice of using ± 5 bpm as a threshold for a significant difference in HRVT was mostly empirical. It seems likely, however, that an error in measurement reaching 10 bpm would lead to important differences in the corresponding workrate or VO_2. This estimation is difficult to quantify as the slope of the HR/VO_2 relationship during incremental exercise varies among individuals depending on baseline fitness level, use of negative chronotropic medication or underlying cardiopulmonary disease. A crude estimate of the impact of varying HR values on exercise intensity can be estimated using our cohort as a whole, where peakHR was linearly related to peak workrate. Using this relationship, a difference of 10 bpm in HR corresponded to an approximately 40W difference in workrate, a difference that is arguably clinically significant, especially when considering patients with severe disease.

The optimal training intensity and modality for patients with COPD is an active matter of debate. Although current guidelines on pulmonary rehabilitation suggest using the American College of Sports Medicine framework for exercise prescription, they acknowledge that using standard "high-intensity" training may not be tolerable by COPD patients (118) and that, in this context, a training program based on perceived exhaustion (Borg scale rating 4-6) is adequate (34). Coincidentally, this level of perceived exhaustion is known to correspond to VO_{2VT} (70). Moreover, the safety and efficacy of using fixed percentages of heart rate or VO_{2peak} as training targets has been challenged by recent publications (80, 82). Our data show that the absolute inter-observer difference in HR_{VT} in patients with COPD becomes increasingly large as disease worsens when compared to controls. In this context, if HR_{VT} was used a marker for exercise intensity prescription in patients with very severe COPD, this target could translate into an unacceptably large array of actual training intensity, which could respectively result in exercise inducing undue fatigue and intolerance (above VT) or of too low intensity to provide benefits (below VT). In the other subgroups of patients, the inter-observer difference in HR_{VT} was less important, and therefore less likely to negatively impact a training regimen.

This study has several limitations. First, VT measurement was performed on retrospectively collected data and as such there is a chance of selection bias. We tried to limit this effect by matching subjects on several relevant clinical parameters. Second, the external validity of the study is impaired by our choice of including only patients having performed at least six minutes on incremental exercise testing, a duration which might not be routinely sustained by the most severe patients with COPD. We believe this criterion is valid in a proof-of-concept framework, with the goal of maximizing the attainment of a true VT, but needs to be taken into account when generalizing these results to a wider COPD population. Third, we recognize that the relative lack of experience of the observers (< 5 years) may be of concern. Hansen et al. (119) showed that when measuring VT in patients with pulmonary hypertension, the agreement between two experienced observers was better than the one between inexperienced observers. However, the overall difference in agreement was small: while experienced observers had a mean difference of 20 ml/min in their measure of VO_{2VT} between them, less experienced observers had a mean difference of 60 ml/min. The clinical relevance of such a small difference is unclear. The fact that our two observers strongly agreed with each other in patients with milder disease, maintained relatively high intra-observer reliability is reassuring.

Fifth, the choice of using the V-slope and the VEM to measure VT was based on the abundance of their use in the literature and the available data regarding their reliability in healthy subjects. Current guidelines suggest the use of either techniques when measuring VT (110). Although other methods to assess VT have been reported, they are often lacking a standardized definition, scarcely used or known to relate closely to the V-slope or the VEM (i.e. changes in $P_{ET}O_2$ and $P_{ET}CO_2$ *vs* time, V_E, VO_2, VCO_2 or RER *vs* work rate, V_E *vs* VCO_2, heart rate inflection point) (120). Therefore, we believe that the choice of using these two methods is representative of common clinical practice and allows a more thorough comparison the available literature. Sixth, our choice of reporting data using 20-second averages of breath-by-breath data could be criticized. This parameter was chosen in accordance with current guidelines concerning the reporting of data during CPET(110) and as it allows, in our opinion, a balance between "noise" and the clear representation of respiratory kinetics. Whether the use of different time-averaging intervals could further influence the detection of VT requires further studies. Finally, the differences in ramp increment rate between subgroups are an expected finding, and whether this may have had an impact on the determination of VO_{2VT} is unclear. However, studies have reported the lack of significant differences in the determination of VO_{2VT} between ramp increments of 7 to 23 W/min in patients with heart failure (121), and between increments of 20 to 50 W/min in young healthy subjects (122).

7.6 Conclusion

In conclusion, results from the present study show that the agreement between human observers in the determination of VT in patients with COPD is lower than in controls, that human and computer analyses of $VO2_{VT}$ are not interchangeable in these patients and that these findings are directly related to the severity of airflow obstruction. Furthermore, the decline in precision in the identification of VO_{2VT} corresponds to an increasing variability when evaluating HR_{VT}. Clinicians should be aware of the discrepancy between software and human identification of VT when reporting automated values of VO_{2VT}, and these findings should be taken into account when using VT for exercise prescription, as a tool to monitor response to an intervention, or as a prognostic marker in patients with COPD.

8. References

1. Mannino DM, Buist S. Global burden of COPD: risk factors, prevalence, and future trends. The Lancet. 2007;370:765-73.
2. Raherison C, Girodet P. Epidemiology of COPD. European Respiratory Review. 2009;18(114):213-21.
3. Canada S. Age-standardized rates, chronic obstructive pulmonary disease (copd) , both sexes, Canada, provinces and territories 2013 [cited 2017 January 6]. Available from: http://www5.statcan.gc.ca/cansim/a26?lang=eng&id=1020561.
4. Heron M. Deaths: Leading Causes for 2008. Natl Vital Stat Rep. 2012;60(6):1-95.
5. Hoyert DL, Xu J. Deaths: Preliminary Data for 2011. Natl Vital Stat Rep. 2012;61(6):1-52.
6. Health Canada. Canadian Tobacco Use Monitoring Survey (CTUMS) 1999-2012 2012 [cited 2014 January 3rd]. Available from: http://www.hc-sc.gc.ca/hc-ps/tobac-tabac/research-recherche/stat/ctums-esutc_2012-eng.php.
7. Statistics Canada. Tables for Mortality: Causes of death, 1981-2007, by sex 2012 [cited 2014 January 4]. Available from: http://www.statcan.gc.ca/pub/91-209-x/2011001/article/11525/tbl/tbl-eng.htm - a1.
8. Hill K, Goldstein RS, Guyatt GH*, et al.* Prevalence and underdiagnosis of chronic obstructive pulmonary disease among patients at risk in primary care. Can Med Assoc J. 2010;182(7):673-8.
9. Hermus G, Stonebridge C, Goldfarb D*, et al.* Cost Risk Analysis for Chronic Lung Disease in Canada. The Conference Board of Canada, 2012 February. Report No.
10. Ståhl E, Lindberg A, Jansson S-A*, et al.* Health-related quality of life is related to COPD disease severity. Health and quality of life outcomes. 2005;3(56):1-8.
11. Center For Disease Control And Prevention. Surgeon General's Report: The Health Consequences of Smoking—50 Years of Progress. U.S. Department of Health and Human Services, 2014.
12. Pirozzi C, Scholand M. Smoking cessation and environmental hygiene. Med Clin North Am. 2012;96(4):849–67.

13. Devereux G. ABC of chronic obstructive pulmonary disease : Definition, epidemiology, and risk factors. British Journal of Medicine. 2006;332.

14. Orozco-Levi M, Garcia-Aymerich J, Villar J*, et al.* Wood smoke exposure and risk of chronic obstructive pulmonary disease. Eur Respir J. 2006;27:542-6.

15. Smith KR, Mehta S, Maeusezahl-Feuz M. Indoor air pollution from household use of solid fuels. Comparative quantification of health risks: Global and regional burden of disease attribution to selected major risk factors. Geneva: World Health Organisation; 2004.

16. Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2013.

17. Stoller J, Aboussouan L. Alpha-1 antitrypsin deficiency. The Lancet. 2005;365:2225-36.

18. Rahman I. Oxydative stress in the pathogenesis of chronic pulmonary obstructive disease: cellular and molecular mechanisms. Cell Biochem Biophys. 2005;43:167-88.

19. Barnes PJ. Immunology of asthma and chronic obstructive pulmonary disease. Nature Reviews Immunology. 2008;8:183-92.

20. Keatings VM, Collins PD, Scott DM*, et al.* Differences in interleukin‑8 and tumor necrosis factor‑α in induced sputum from patients with chronic obstructive pulmonary disease or asthma. American Journal of Respiratory and Critical Care. 1996;153:530-4.

21. Loring SH, Garcia-Jacques M, Malhotra A. Pulmonary characteristics in COPD and mechanisms of increased work of breathing. J Appl Physiol. 2009;107:309-14.

22. Hyatt R. Expiratory flow limitation. J Appl Physiol. 1983;55(1):1-8.

23. O'Donnell D, Banzett R, Carrieri-Kohlman V*, et al.* Pathophysiology of dyspnea in chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society. 2007;4:145–68.

24. Oga T, Nishimura K, Tsukino M*, et al.* Analysis of the factors related to mortality in chronic obstructive pulmonary disease: Role of exercise capacity and health status. American Journal of Respiratory and Critical Care. 2003;167:544-9.

25. Agusti AGN, Noguera A, Sauleda J*, et al.* Systemic effects of chronic obstructive pulmonary disease. Eur Respir J. 2003;21:347-60.

26. Schols A, Soeters P, Dingemans A*, et al.* Prevalence and characteristics of nutritional depletion in patients with stable COPD eligible for pulmonary rehabilitation. American Review of Respiratory Diseases. 1993;147:1151-6.
27. Bernard S, Blanc PLE, Whittom F*, et al.* Peripheral Muscle Weakness in Patients with Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care. 1998;158:629-34.
28. Sin D, Man S. Chronic obstructive pulmonary disease as a risk factor for cardiovascular morbidity and mortality. Proceedings of the American Thoracic Society. 2005;2(1):8.
29. Light R, Merrill E, Despars J*, et al.* Prevalence of depression and anxiety in patients with COPD. Relationship to functional capacity. Chest. 1985;87(1):35.
30. Hung W, Wisnivesky J, Siu A*, et al.* Cognitive decline among patients with chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care. 2009;180(2):134.
31. Ferguson G, Calverley P, Anderson J*, et al.* Prevalence and progression of osteoporosis in patients with COPD: results from the TOwards a Revolution in COPD Health study. Chest. 2009;136(6):1456.
32. O'Donnell D, Aaron S, Bourbeau J. Canadian Thoracic Society recommendations for management of chronic obstructive pulmonary disease – 2007 update. Can Respir J. 2007;14(SB):5B-32B.
33. Nici L, Donner C, Wouters E*, et al.* American Thoracic Society/European Respiratory Society Statement on Pulmonary Rehabilitation. American Journal of Respiratory and Critical Care. 2006;173:1390-413.
34. Spruit MA, Singh SJ, Garvey C*, et al.* An Official American Thoracic Society/EuropeanRespiratory Society Statement: Key Concepts and Advances in Pulmonary Rehabilitation. Am J Respir Crit Care Med. 2013;188(8):e13-e64.
35. Lacasse Y, Goldstein R, Lasserson T*, et al.* Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2009(4).
36. Nery LE, Wasserman K, Huntsman DJ*, et al.* Ventilatory and gas exchange during exercise in chronic airway obstruction. J Appl Physiol. 1982;53(6):1594-602.

37. Sietsema K. Cardiovascular limitations in chronic pulmonary disease. Med Sci Sports Exerc. 2001;33 (Suppl 7):S656-S61.

38. ACSM's guidelines for exercise testing and prescription. Philadelphia: American College of Sports Medicine, 2009.

39. Maltais F, LeBlanc P, Jobin J*, et al.* Intensity of training and physiologic adaptation in patients with chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care. 1997;155:555-61.

40. Probst V, Troosters T, Pitta F*, et al.* Cardiopulmonary stress during exercise training in patients with COPD. Eur Respir J. 2006;27:1110-8.

41. Vallet G, Varray A, Fontaine JL*, et al.* Value of individualized rehabilitation at the ventilatory threshold level in moderately severe chronic obstructive pulmonary disease. Rev Mal Respir. 1994;11(5):493-501.

42. Vallet G, Ahmaïdi S, Serres I*, et al.* Comparison of two training programmes in chronic airway limitation patients: standardized versus individualized protocols. Eur Respir J. 1997;10:114-22.

43. Serres I, Varray A, Vallet G*, et al.* Improved Skeletal Muscle Performance After Individualized Exercise Training in Patients With Chronic Obstructive Pulmonary Disease. J Cardiopulm Rehabil. 1997;17(4):232-8.

44. Gimenez M, Servera E, Vergara P*, et al.* Endurance Training in Patients With Chronic Obstructive Pulmonary Disease: A Comparison of High Versus Moderate Intensity. Arch Phys Med Rehabil. 2000;81:102-9.

45. Préfaut C, Varray A, Vallet G. Pathophysiological basis of exercise training in patients with chronic obstructive lung disease. European Respiratory Review. 1995;5(25):27-32.

46. Rich P. The molecular machinery of Keilin's respiratory chain. Biochem Soc Trans. 2003;31:1095-105.

47. Wasserman K. Determinants and detection of anaerobic threshold and consequences of exercise above it. Circulation. 1987;76 (Suppl VI):VI-29.

48. Hill A, Long C, Lupton H. Muscular exercise, lactic acid, and the supply and utilization of oxygen. Part VI The oxygen debt at the end of exercise. Proceedings of the Royal Society B. 1924;97:127.

49. Wasserman K, Whipp B, Koyal S*, et al.* Anaerobic threshold and respiratory gas exchange during exercise. J Appl Physiol. 1973;35(2):236-43.
50. Hughson R, Weisiger K, Swanson G. Blood lactate concentration increases as a continuous function in progressive exercise. J Appl Physiol. 1987;62:1975–81.
51. Yeh MP, Gardner RM, Adams TD*, et al.* "Anaerobic threshold": problems of determination and validation. J Appl Physiol. 1983;55(4):1178-86.
52. Wasserman K, McIlroy M. Detecting the threshold of anaerobic metabolism in cardiac patients during exercise. Am J Cardiol. 1964;14(6):844-52.
53. Wasserman K, Whipp BJ. Exercise Physiology in Health and Disease. Am Rev Respir Dis. 1975;112:219-49.
54. Reinhard U, Muller PH, Schmulling R-M. Determination of the anaerobic threshold by the ventilation equivalent in normal individuals. Respiration. 1979;38:36-42.
55. Caiozzo VJ, Davis JA, Ellis JF*, et al.* A comparison of gas exchange indices used to detect the anaerobic threshold. J Appl Physiol. 1982;53(5):1184-9.
56. Green HJ, Hughson RL, Orr GW*, et al.* Anaerobic threshold, blood lactate, and muscle metabolites in progressive exercise. J Appl Physiol. 1983;54(4):1032-8.
57. Beaver WL, Wasserman K, Whipp BJ. A new method for detecting anaerobic threshold by gas exchange. J Appl Physiol. 1986;60(6):2020-7.
58. Sullivan C, Casaburi R, Storer T*, et al.* Non-invasive prediction of blood lactate response to constant power outputs from incremental exercise tests. Eur J Appl Physiol Occup Physiol. 1995;71(349-354):349.
59. Patessio A, Casaburi R, Carone M*, et al.* Comparison of Gas Exchange, Lactate, and Lactic Acidosis Thresholds in Patients with Chronic Obstructive Pulmonary Disease. Am Rev Respir Dis. 1993;148:622-6.
60. Gladden LB, Yates JW, Stremel RW*, et al.* Gas exchange and lactate anaerobic thresholds: inter- and intraevaluator agreement. J Appl Physiol. 1985;58(6):2082-9.
61. Péronnet F, Aguilaniu B. Lactic acid buffering, nonmetabolic CO2 and exercise hyperventilation: A critical reappraisal. Respir Physiol Neurobiol. 2006;150:4-18.
62. Hagberg JM, Coyle EF, Carroll JE*, et al.* Exercise hyperventilation in patients with McArdle's disease. J Appl Physiol. 1982;52(4):991-4.

63. 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.

64. Paterson D. Potassium and ventilation in exercise. J Appl Physiol. 1992;72:811-20.

65. McLoughlin P, Linton R, Band D. Effects of potassium and lactic acid on ventilation in anesthetized cats. Respir Physiol. 1994;1994:171-9.

66. Paterson D, Friedland J, Bascom D. Changes in arterial K+ and ventilation during exercise in normal subjects and subjects with McArdle's syndrome. J Physiol. 1990;429:339-48.

67. Hanninen O, Airaksinen O, Karipohja M. On-line determination of anaerobic threshold with rms-EMG. Biomedica Biochimica Acta. 1989;48:493-503.

68. Vitasalo J, Luhtanen P, Rahkila P. Electromyographic activity related to aerobic and inaerobic threshold in ergometer bicycling. Acta Physiol Scand. 1985;124:287-93.

69. Helal J, Guezennec C, Goubel F. The aerobic-anaerobic transition: reexarnination of the threshold concept including an electromyographic approach. Eur J Appl Physiol. 1987;56:643-9.

70. Zamunér AR, Moreno MA, Camargo TM*, et al.* Assessment of subjective perceived exertion at the anaerobic threshold with the Borg CR-10 scale. J Sports Sci Med. 2011;10:130-6.

71. ATS/ACCP Statement on Cardiopulmonary Exercise Testing. American Journal of Respiratory and Critical Care. 2001;117:211-77.

72. Weltmana A, Katchb V, Sadyb S*, et al.* Onset of Metabolic Acidosis (Anaerobic Threshold) as a Criterion Measure of Submaximum Fitness. Research Quarterly American Alliance for Health, Physical Education and Recreation. 1978;49(2):218-27.

73. Tanaka K, Matsuura Y, Matsuzaka A*, et al.* A longitudinal assessment of anaerobic threshold and distance-running performance. Medicine and Science in Sports and Medicine. 1984;16(3):278-82.

74. Brandon L. Physiological factors associated with middle distance running performance. Sports Med. 1995;19(4):268-77.

75. Casaburi R. Physiologic responses to training. Clin Chest Med. 1994;15:215-27.

76. Davis JA, Frank MH, Whipp B*, et al.* Anaerobic threshold alterations caused by endurance training in middle-aged men. J Appl Physiol. 1979;46(6):1039-46.

77. Casaburi R, Patessio A, Loll F*, et al.* Reductions in Exercise Lactic Acidosis and Ventilation as a Result of Exercise Training in Patients with Obstructive Lung Disease. American Review of Respiratory Diseases. 1991;143:9-18.

78. Gruber W, Orenstein DM, Braumann KM*, et al.* Interval exercise training in cystic fibrosis—Effects on exercise capacity in severely affected adults. Journal of Cystic Fibrosis. 2014;13:86-91.

79. Ahmaidi S, Masse-Biron J, Adam B*, et al.* Effects of interval training at the ventilatory threshold on clinical and cardiorespiratory responses in elderly humans. Eur J Appl Physiol. 1998;78:170-6.

80. Hofmann P, Tschakert G. Special Needs to Prescribe Exercise Intensity for Scientific Studies. Cardiol Res Pract. 2011;2011:1-10.

81. Zacarias E, Neder JA, Cendom S*, et al.* Heart Rate at the Estimated Lactate Threshold in Patients With Chronic Obstructive Pulmonary Disease: Effects on the Target Intensity for Dynamic Exercise Training. J Cardiopulm Rehabil. 2000;20(6):369-76.

82. Diaz-Buschmann I, Jaureguizar KV, Calero MJ*, et al.* Programming exercise intensity in patients on beta-blocker treatment: the importance of choosing an appropriate method. Eur J Prev Cardiol. 2013:1-7.

83. Meyer T, Gorge G, Schwaab B*, et al.* An alternative approach for exercise prescription and efficacy testing in patients with chronic heart failure: A randomized controlled training study. Am Heart J. 2005;149(5):926.e1-.e7.

84. Vivodtzev I, Gagnon P, Pepin Vr*, et al.* Physiological Correlates of Endurance Time Variability during Constant-Workrate Cycling Exercise in Patients with COPD. PLoS One. 2011;6(2):1-8.

85. Myers J, Gullestad L, Vagelos R*, et al.* Clinical, hemodynamic, and cardiopulmonary exercise test determinants of survival in patients referred for evaluation of heart failure. Ann Intern Med. 1998;129:286-93.

86. Gitt A, Wasserman K, Kilkowski C. Exercise anaerobic threshold and ventilatory efficiency identify heart failure patients for high risk of early death. Circulation. 2002;106:3079-84.

87. Piepoli MF. Exercise Tolerance Measurements in Pulmonary Vascular Diseases and Chronic Heart Failure. Respiration. 2009;77:121-7.

88. Ramos RP, Alencar MCN, Treptow E*, et al.* Clinical Usefulness of Response Profiles to Rapidly Incremental Cardiopulmonary Exercise Testing. Pulm Med. 2013;2013:1-25.

89. Sun XG, Hansen JE, Oudiz RJ*, et al.* Exercise pathophysiology in patients with primary pulmonary hypertension. Circulation. 2001;104(4):429-35.

90. Plas MVD, Kan CV, Wells AU*, et al.* Pulmonary vascular limitation to exercise and survival in idiopathic pulmonary fibrosis. Respirology. 2013.

91. Older P, Smith R, Courtney P*, et al.* Preoperative Evaluation of Cardiac Failure and lschemia in Elderly Patients by Cardiopulmonary Exercise Testing. Chest. 1993;104:701-4.

92. Torchio R, Gulotta C, Parvis M*, et al.* Gas exchange threshold as a predictor of severe postoperative complications after lung resection in mild-to-moderate chronic obstructive pulmonary disease. Monaldi Arch Chest Dis. 1998;53(2):127-33.

93. 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. 2013:1-7.

94. Palange P, Ward SA, Carlsen K-H*, et al.* Recommendations on the use of exercise testing in clinical practice. European Journal of Respirology. 2009;29:185-209.

95. Ries AL, Bauldoff GS, Carlin BW*, et al.* Pulmonary Rehabilitation Joint ACCP/AACVPR Evidence-Based Clinical Practice Guidelines. Chest. 2007;131:4-42.

96. Maltais F, Leblanc P, Simard C*, et al.* Skeletal Muscle Adaptation to Endurance Training in Patients with Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care. 1996;154:442-7.

97. Garrard C, Das R. Sources of error and variability in the determination of anaerobic threshold in healthy humans. Respiration. 1987;51:137-45.

98. Shimizu M, Myers J, Buchanan N*, et al.* The ventilatory threshold: Method, protocol, and evaluator agreement. Am Heart J. 1991;122(2):509-16.

99. Filho P, Pompeu F, Silva APRdSe. Accuracy of VO2max and anaerobic threshold determination. Rev Bras Med Esporte. 2005;11(4):162-5.

100. Sinclair RCF, Danjoux GR, Goodridge V*, et al.* Determination of the anaerobic threshold in the pre-operative assessment clinic: inter-observer measurement error. Anaesthesia. 2009;64:1192-5.

101. Belman MJ, Epstein LJ, Doornbos D*, et al.* Noninvasive Determinations of the Anaerobic Threshold ; Reliability and Validity in Patients with COPD. Chest. 1992;102:1028-34.

102. Orr G, Green HJ, Hughson RL*, et al.* A computer ventilatory linear regression model to determine anaerobic threshold. J Appl Physiol. 1982;52(5):1349-52.

103. Solberg G, Robstad B, Skjønsberg OH*, et al.* Respiratory gas exchange indices for estimating the anaerobic threshold. Journal of Sports Science and Medicine. 2005;4:29-36.

104. Ekkekakis P, Lind E, Hall EE*, et al.* Do regression-based computer algorithms for determining the ventilatory threshold agree? J Sports Sci. 2008;26(9):967-76.

105. Svedahl K, MacIntosh BR. Anaerobic threshold: the concept and methods of measurement. Can J Appl Physiol. 2003;28(2):299-323.

106. Wasserman K, Whipp BJ, Koyl SN*, et al.* Anaerobic threshold and respiratory gas exchange during exercise. J Appl Physiol. 1973;35(2):236-43.

107. Péronnet F, Thibault G, Rhodes E*, et al.* Correlation between ventilatory threshold and endurance capability in marathon runners. Med Sci Sports Exerc. 1987;19(6):610-5.

108. Frisk B, Espehaug B, Hardie JA*, et al.* Airway obstruction, dynamic hyperinflation, and breathing pattern during incremental exercise in COPD patients. Physiol Rep. 2014;2(2):1-8.

109. Casaburi R, Patessio A, Ioli F*, et al.* Reductions in exercise lactic acidosis and ventilation as a result of exercise training in patients with obstructive lung disease. Am Rev Respir Dis. 1991;143(1):9-18.

110. ATS/ACCP Statement on Cardiopulmonary Exercise Testing. Am J Respir Crit Care Med. 2003;117:211-77.

111. Borg G. Psychophysical bases of perceived exertion. Med Sci Sports Exerc. 1982;14(5):377-81.

112. Ozcelik O, Ward SA, Whipp BJ. Effect of altered body CO2 stores on pulmonary gas exchange dynamics during incremental exercise in humans. Exp Physiol. 1999;84(5):999-1011.

113. Shrout P. Measurement reliability and agreement in psychiatry. Stat Methods Med Res. 1998;7:301-17.

114. Passing H, Bablok. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem. 1983;21(11):709-20.

115. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

116. Poole DC, Gaesser GA. Response of ventilatory and lactate thresholds to continuous and interval training. Journal of applied physiology (Bethesda, Md : 1985). 1985;58(4):1115-21.

117. Santos EL, Giannella-Neto A. Comparison of computerized methods for detecting the ventilatory thresholds. Eur J Appl Physiol. 2004;93(3):315-24.

118. Maltais F, Leblanc P, Jobin J*, et al.* Intensity of Training and Physiologic Adaptation in Patients with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 1997;155:555-56.

119. Hansen JE, Sun X-G, Yasunobu Y*, et al.* Reproducibility of Cardiopulmonary Exercise Measurements in Patients With Pulmonary Arterial Hypertension. Chest. 2004;126:816-24.

120. Binder RK, Wonisch M, Corra U*, et al.* Methodological approach to the first and second lactate threshold in incremental cardiopulmonary exercise testing. Eur J Cardiovasc Prev Rehabil. 2008;15:726-34.

121. Agostoni P, Bianchi M, Moraschi A*, et al.* Work-rate affects cardiopulmonary exercise test results in heart failure. Eur J Heart Fail. 2005;7(4):498-504.

122. Davis JA, Whipp BJ, Lamarra N*, et al.* Effect of ramp slope on determination of aerobic parameters from the ramp exercise test. Med Sci Sports Exerc. 1982;14(5):339-43.
123. 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;159:179-87.
124. Quanjer P, Tammeling G, Cotes J*, et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J. 1993;Mar(16):5-40.
125. Cotes J, Chinn D, Quanjer P*, et al.* Standardization of the measurement of transfer factor (diffusing capacity). Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J. 1993;Mar(16):41-52.
126. Koch B, Schäper C, Ittermann T. Reference values for cardiopulmonary exercise testing in healthy volunteers: the SHIP study. Eur Respir J. 2009;33:389-97.
127. Bilic-Zulle L. Comparison of methods: Passing and Bablok regression. Biochem Med (Zagreb). 2011;21(1):49-52.

9. APPENDIX A: Supplemental digital content

Methods

Exercise testing

Symptom-limited incremental exercise tests were performed according to published guidelines (110). More specifically, tests were realized on an electromagnetically braked cycle ergometer (Ergoline 200, Ergoline, Bitz, Germany), with a protocol including two minutes of rest and a three-minute period of initial unloaded cycling. Load was increased linearly until exhaustion (ramp was individually determined for each patient by the attending physician, based on either previous exercise testing result or expected maximal work rate as estimated by overall physical fitness and/or FEV_1) with the goal of maintaining a cycling speed of 60 revolutions per minute. Breath-by-breath analysis of expired gases was performed using electronic analysis (Jaeger Oxycon Pro, CareFusion, Hoechberg, Germany). V_E, VO_2, VCO_2, V_E/VO_2 and V_E/VCO_2 were computed using twenty-second averages of breath-by-breath values. Peak VO_2 was the highest 20-second mean VO_2 obtained. Patients using beta-blockers were not required to stop them prior to CPET. Oxygen saturation was monitored using finger or ear pulse oximetry. Exercise capacity was defined as the highest work rate achieved for at least 20 seconds at a rate of at least 50 revolutions per minute. Arterial blood gases were assessed at baseline using a standard blood gas analyzer (ABL800 Flex, Radiometer, Copenhagen, Denmark). Dyspnea and leg fatigue were evaluated at rest and at maximal exercise intensity using the modified 10-point Borg scale (111). The pneumotachograph and plethysmography box used for the tests was calibrated twice a day, while the turbine flow sensor of the CPET system was calibrated daily. Gas analyzers for D_LCO measurements were calibrated before each test. Gas analyzers of the CPET system are calibrated daily using a high precision gas cylinder containing 16% O_2 and 4% CO_2. The O_2 cell of the CPET system is changed every 18 months, or as soon as gas calibration becomes unstable. The cycle ergometer is calibrated using a standard procedure once a year. In the period from which the study's tests were performed (2010-2014), a total of seven technicians operated the system in rotations. Six of these seven operators were present during the whole 4-year period. Every operator received training by the same head technician of the laboratory, ensuring homogeneity. A structured and clear protocol was implemented for the realization of incremental exercise testing. This protocol was approved by the physician in charge of the pulmonary function test laboratory and is based on the latest ATS/ACCP guidelines. Every technician operating the system is familiar with the protocol, which is easily and readily available in written form in the exercise-testing laboratory. Every morning, the head technician of the laboratory reviewed the resting and exercise tests from the day before to ensure internal quality and conformity with the protocols. In the event of an error, the technician responsible for the test was informed, ensuring continuous training and retroaction. Physicians supervising the test based their evaluation of the ramp increment on the following parameters: the predicted maximal workrate was divided by 10 with the goal of reaching a test duration of 8-12 minutes. The resulting ramp was adjusted based on the physician's judgment based on either a previous exercise test performed in our institution or lung function, as described. Reference values for spirometry, lung function, diffusing capacity and exercise testing were taken from standard sources (123-126).

Results

Comparison of human and computer observers in the determination of VO_{2VT}

E-table 1 and e-figure 1 describe the results of the Passing-Bablok regression analysis comparing each human observer to the computerized analysis. This technique generates a regression equation in the form "y=a + bx", where "a" is the regression line's intercept and "b" its slope. Each of these variables is associated to a 95% confidence interval that will explain if their value differ from zero for intercept and value one for slope only by chance. If the 95% CI for the intercept includes "0", it can be concluded that there is no significant difference between obtained intercept value and value zero and there is no constant difference between two methods. In the same manner, if the 95% CI for "slope" includes "1", it can be concluded that there is no significant difference between obtained slope value and value one and there is no proportional difference between two methods. In such case we can assume both analytical methods of measurement can be used interchangeably. In addition, Passing-Bablok regression requires both variables to be linearly related. The test therefore evaluates if a significant deviation from linearity is present before beginning analysis (127).

For both human observers, the relationship of VO_{2VT} with the computerized analysis did not differ from linearity, confirming that the data can be used in Passing-Bablok analysis. Using V-slope, VO_{2VT} values from human observer 1 were interchangeable with computer analysis for controls (the 95% CI for Intercept and Slope include "0" and "1" respectively) but not for patients with COPD (95% CI for Intercept and Slope do not include "0" and "1", respectively). Similar results were obtained using the VEM. In an identical manner, observer 2 was found to be interchangeable with computerized analysis when evaluating controls, but not COPD patients.

Internal validity

Intra-observer ICC measured on a subset of 50 patients showed relatively high reliability throughout the spectrum of disease severity (see e-table 3). For both observers and for both methods of observation, ICCs across disease severity groups were all higher than 0.81. Although ICCs remained high across disease subgroups, there is a small tendency for agreement between observers to get lower with disease progression.

e-table 1

e-table 1. Passing-Bablok regression analysis comparing computer analysis to each human observers.

		V-slope						VEM					
		Intercept A	95% CI	Slope B	95% CI	Deviation from linearity?	p	Intercept A	95% CI	Slope B	95% CI	Deviation from linearity?	p
Controls	Observer 1 *vs* computer analysis	149	-546 – 675	0.99	0.53 – 1.77	No	0.89	-313	-1744 – 554	1.41	0.57 – 2.73	No	0.42
	Observer 2 *vs* computer analysis	299	-424 – 729	0.89	0.47-1.61	No	0.78	-282	-1652 – 603	1.43	0.56 – 2.52	No	0.42
COPD	Observer 1 *vs* computer analysis	-391	-678 – -161	1.43	1.14 –1.80	No	0.13	-697	-1153 – -431	1.71	1.36 – 2.16	No	0.08
	Observer 2 *vs* computer analysis	-481	-731 – -275	1.60	1.34 –1.91	No	0.64	-1008	-1498 – -594	1.99	1.53 – 2.53	No	0.33

e-table 2

e-table 2. Mean (SD) inter-observer difference in HR_{VT} according to the use of beta-blockers

	V-slope			VEM		
	BB	No BB	p	BB	No BB	p
Controls*	2 (-)	2 (1)	0.94	2 (-)	2 (2)	0.96
All COPD	7 (4)	6 (3)	0.41	7 (4)	7 (3)	0.95
GOLD 1	3 (1)	3 (2)	0.73	4 (2)	4 (1)	0.74
GOLD 2	3 (1)	5 (2)	0.06	7 (1)	6 (2)	0.52
GOLD 3	7 (2)	6 (3)	0.48	7 (3)	8 (2)	0.15
GOLD 4	11 (3)	9 (2)	0.29	10 (4)	10 (2)	0.64

Data presented as mean (standard deviation).
P values refer to comparisons between BB and No BB for each severity subgroup, using independent-samples t-tests.
*=only 1 subject with BB in this group.
HRVT=heart rate at the ventilatory threshold; BB=beta-blockers; VEM=ventilatory equivalent method; COPD=chronic obstructive pulmonary disease; GOLD=Global initiative for Obstructive Lung Disease.

e-table 3

e-table 3 Intra-observer reliability in the determination of the $VO2_{VT}$ (ml/min) using two methods, on a subset of 50 patients.

	Intra-class correlation - observer 1		**Intra-class correlation - observer 2**	
	V-slope	VEM	V-slope	VEM
Controls	0.99	0.99	0.99	0.99
All COPD	0.92	0.91	0.92	0.89
GOLD 1	0.95	0.96	0.99	0.94
GOLD 2	0.89	0.90	0.93	0.84
GOLD 3	0.91	0.87	0.90	0.86
GOLD 4	0.86	0.82	0.83	0.81

GOLD = Global initiative for Obstructive Lung Disease. HRVT = heart rate at the ventilatory threshold. VEM = ventilatory equivalent method. VO2VT = oxygen uptake at the ventilatory threshold.

e-figure 1

V-slope **VEM**

Controls COPD Controls COPD

Observer 1 *vs* computer analysis

Observer 2 *vs* computer analysis