**Supplemental digital content – Methods and Results**

**Methods**

*Exercise testing*

Symptom-limited incremental exercise tests were performed according to published guidelines (1). 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 FEV1) 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). VE, VO2, VCO2, VE /VO2 and VE /VCO2 were computed using twenty-second averages of breath-by-breath values. Peak VO2 was the highest 20-second mean VO2 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 (3). 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 DLCO measurements were calibrated before each test. Gas analyzers of the CPET system are calibrated daily using a high precision gas cylinder containing 16% O2 and 4% CO2. The O2 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 (4-7).

**Results**

*Comparison of human and computer observers in the determination of VO2VT*

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 (2).

For both human observers, the relationship of VO2VT with the computerized analysis did not differ from linearity, confirming that the data can be used in Passing-Bablok analysis. Using V-slope, VO2VT 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 HRVT 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 VO2VT (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**



**References**

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

2. Bilic-Zulle L. Comparison of methods: Passing and Bablok regression. Biochem Med (Zagreb). 2011;21(1):49-52. PubMed PMID: 22141206.

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

4. Cotes J, Chinn D, Quanjer P, Roca J, Yernault J. 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.

5. 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.

6. 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.

7. Quanjer P, Tammeling G, Cotes J, Pedersen O, Peslin R, Yernault J. 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.