

Comparison of visual, automatic and semiautomatic methods to determine ventilatory indices in 50 to 60 years old adults

Martin Pühringer , Susanne Ring-Dimitriou , Thomas Stöggl , Bernhard Iglseider & Bernhard Paulweber

To cite this article: Martin Pühringer , Susanne Ring-Dimitriou , Thomas Stöggl , Bernhard Iglseider & Bernhard Paulweber (2020) Comparison of visual, automatic and semiautomatic methods to determine ventilatory indices in 50 to 60 years old adults, Journal of Sports Sciences, 38:6, 692-702, DOI: [10.1080/02640414.2020.1725993](https://doi.org/10.1080/02640414.2020.1725993)

To link to this article: <https://doi.org/10.1080/02640414.2020.1725993>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 24 Feb 2020.



Submit your article to this journal [↗](#)



Article views: 431



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Comparison of visual, automatic and semiautomatic methods to determine ventilatory indices in 50 to 60 years old adults

Martin Pühringer ^a, Susanne Ring-Dimitriou ^a, Thomas Stöggl ^a, Bernhard Iglseider ^b and Bernhard Paulweber^c

^aDepartment of Sport and Exercise Science, University of Salzburg, Salzburg, Austria; ^bDepartment of Geriatric Medicine, Christian-Doppler-Klinik, Paracelsus Private Medical University, Salzburg, Austria; ^cDepartment of Internal Medicine I, Paracelsus Private Medical University, Salzburg, Austria

ABSTRACT

The aim of this study was to compare different methods of detecting ventilatory indices (VI) and to investigate the impact of cardiorespiratory fitness (CRF) level on VI detection. Fifty females and fifty males completed a graded exercise test until volitional exhaustion with continuous gas-exchange measurement. The first and second ventilatory indices (VI-1, VI-2) were detected through different single automatic methods and through a semiautomatic method which combines visual and automatic detection methods. Additionally, the VIs were detected visually by two experts which served as the study specific gold standard. When comparing the semiautomatic method at VI-1 (intraclass correlation coefficients (ICC) 0.88 [0.81, 0.92], Bland-Altman bias \pm limits of agreement (LoA) 55 ± 334 ml O₂ · min⁻¹) and VI-2 (ICC 0.97 [0.96, 0.98], LoA 1 ± 268 ml O₂ · min⁻¹) to the visually detected VI, high levels of agreements and no significant differences were found. This was not the case for any of the other automatic methods. Additionally, we couldn't find any relevant differences regarding the CRF level.

We therefore concluded that the semiautomatic detection method should be used for VI detection, as results are more accurate than in any of the single-automatic methods.

Abbreviations: CPET: cardiopulmonary exercise test; CRF: Cardiorespiratory fitness; VO_{2peak}: peak oxygen uptake; VI-1: first ventilatory indices; VI-2: second ventilatory indices; LoA: Bland-Altman bias \pm limits of agreement; ICC: intraclass correlation coefficient.

ARTICLE HISTORY

Accepted 27 January 2020

KEYWORDS

Aerobic Capacity; cardio Pulmonary; exercise Physiology; exercise Testing

Introduction

Incremental cardiopulmonary exercise testing (CPET) is used in research to provide prognostic values for populations (e.g. VO_{2peak}), to assess functional capacity or impairment, to demarcate exercise-intensity domains, and to design training programs for athletes as well as for therapeutic purposes (Ross, 2003). The gold standard of measuring aerobic power, which serves as an index of cardiorespiratory fitness, is achieved by assessing maximal oxygen uptake (VO_{2max}) (Ross, 2003). However, submaximal indices (e.g. anaerobic threshold, respiratory compensation point) have been shown to provide an alternative to VO_{2max} and have shown their prognostic value in assessing the effectiveness of training, rehabilitation, and classifying the aerobic capacity of individuals (Meyer, Lucía, Earnest, & Kindermann, 2005). Furthermore, when utilising CPET it has been shown that people with a low exercise tolerance may not reach the maximal workload needed to achieve VO_{2max}. Physical symptoms such as hypertension, dyspnoea or skeletal muscle soreness can limit exercise and therefore submaximal indices are highly valuable for interpreting aerobic capacity (Meyer et al., 2005).

Different submaximal ventilatory indices have been identified in previous research. A three-phase model with two ventilatory indices (VI-1 and VI-2; also called ventilatory threshold 1 and 2) is generally applied to detect the changes of metabolism, gas exchange variables, and ventilation during an

incremental CPET (Binder et al., 2008; Meyer et al., 2005; Westhoff et al., 2013).

The VI-1 can be determined by finding the disproportional increase in a carbon dioxide (VCO₂) vs. oxygen uptake (VO₂) plot (Beaver, Wasserman, & Whipp, 1986). The VI-2 can be determined by finding the disproportional increase in a pulmonary ventilation (VE) vs. VCO₂ plot (Wasserman, Whipp, & Davis, 1981). VI-1 and VI-2 can further be determined by the ventilatory equivalent method based on the equivalents for oxygen and carbon dioxide (EQO₂, EQCO₂) or the end-tidal partial pressure of oxygen and carbon dioxide (PETO₂, PETCO₂) (Binder et al., 2008; Meyer et al., 2005; Wasserman, Hansen, Sue, Whipp, & Froelicher, 2011; Westhoff et al., 2013).

As shown in previous studies the ability of interpreters to detect VI is strongly dependent on their experience and the use of a systematic approach. Dolezal et al. (2017) reported agreement levels between interpreters using visual detection methods on VO₂ at VI-1 ranging from ± 195 ml · min⁻¹ (expert) to ± 790 ml · min⁻¹ (novice). Similar levels of agreement (± 100 ml O₂ · min⁻¹ for VI-1 and ± 130 ml O₂ · min⁻¹ for VI-2) between experienced interpreters were reported by Santos and Giannella-Neto (2004). In a recent paper Meyer et al. (2005) mentioned that false detection of VI (e.g. mistakenly identifying VI-2 as VI-1) is still common in clinical exercise testing.

To increase observer reliability of VI detection, automatic methods have been developed, that commonly use piecewise

linear regression models (e.g. Beaver et al., 1986). Hereby an assumption of the total number of “breakpoints (i.e. VI)” must be made beforehand utilising bi-segmental or tri-segmental models. Ekkekakis, Lind, Hall, and Petruzzello (2008) compared nine automatic methods based on linear regression and found significant differences between the methods at VI-1 ranging from $\pm 790 \text{ ml O}_2 \cdot \text{min}^{-1}$ to $\pm 1.730 \text{ ml O}_2 \cdot \text{min}^{-1}$. A difference in VI-1 of $\pm 830 \text{ ml O}_2 \cdot \text{min}^{-1}$ between an automatic method using linear regression models and a visual detection by experienced interpreters was reported by Dolezal et al. (2017). A possible explanation for these large differences is that these linear regression models can identify unreliable VI, especially in subjects with a low or undetectable VI-2 (Sherrill, Anderson, & Swanson, 1990).

Another VI-detection method consists of polynomial regression analysis without the need of assumptions about the shape of the curves or the number of “breakpoints”. Moreover, the shape of the polynomial function and its first and second derivative (slope and curvature of the regression) yield some additional information concerning the characteristics of the curves (Sherrill et al., 1990). Santos and Giannella-Neto (2004) reported good agreement on VO_2 between visual and automatic detection methods (using polynomial regression and calculation of the first and second derivatives of EQO_{2t} , EQCO_{2t} , PETO_{2t} , PETCO_{2t} , VE vs. VO_{2t} , and VE vs. VCO_{2t} curves) ranging from ± 320 to $\pm 430 \text{ ml O}_2 \cdot \text{min}^{-1}$ for VI-1, and ± 360 to $\pm 440 \text{ ml O}_2 \cdot \text{min}^{-1}$ for VI-2 (32 participants aged 17–45 years). Wisén and Wohlfart (2004) used polynomial regression models to automatically determine various VI in VO_2 , VCO_2 , EQO_2 , and EQCO_2 time curves by calculating the first and second derivatives in 19 male participants aged 20–48 years. They had to manually adjust 34% of VI-1 and 40% of VI-2 determinations after a visual inspection of the graphical representations because of implausible values. These high percentages demonstrate the difficulty of automatic VI detection. In their analysis, Wisén and Wohlfart (2004) also detected two patterns of VI related to the cardiorespiratory fitness (CRF) levels (i.e. aerobic power, $\text{VO}_{2\text{peak}}$) of the participants, which may reflect different strategies of individuals to cope with the exercise strain indicated by the onset of lactate accumulation, isocapnic buffering or respiratory compensation. These different patterns may be one of the reasons for implausible values of VI using visual as well as automatic detection methods. Mixing up VI-2 with VI-1 is one of the most common methodological errors in the literature as well as results reported for VI-1 being usually too high (McLellan, 1987). Especially in subjects with low CRF levels, which are often unable to reach VI-2 during exercise testing or which does not show the expected ventilatory response curves, this false detection can occur (Meyer et al., 2005). To account for these different VI patterns and ventilatory response curves, detection methods should be tested on subjects with different CRF levels to insure their accuracy.

To conclude, the results from previous studies suggest that automatic VI-detection still needs to be visually inspected and corrected due to the high number of false VI-detection (Wisén & Wohlfart, 2004). Therefore, automatic detection methods should be further improved because precise identification of VI is essential for its utility in research and clinical practice.

Moreover, regarding an ageing population there is a need for reliable detection methods especially concerning sedentary and elderly subjects. To our knowledge, there are no studies comparing different automatic methods using polynomial regression in people older than 50 years and with different CRF levels.

Therefore, the aims of this study were: 1) to compare different methods (visual, automatic and semiautomatic) used to detect VI-1 and VI-2 in a group of 50–60 year old males and females from the Paracelsus 10.000 study (Salk, 2016), 2) to investigate the impact of the CRF level on the detection of VI-1 and VI-2, and finally 3) to create a recommendation for a systematic VI-detection method with clearly defined confidence levels.

Methods

Participants

One hundred data-sets were randomly drawn from a sample of 520 Paracelsus 10.000 participants who reached volitional exhaustion (Salk, 2016). Hence, our sample used for VI detection consisted of 50 females (55 ± 3 years, 166 ± 7 cm, 68.3 ± 11.2 kg, $\text{VO}_{2\text{peak}}$ range: $16.2\text{--}43.3 \text{ ml} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$) and 50 males (55 ± 3 years, 177 ± 6 cm, 82.9 ± 14.5 kg, $\text{VO}_{2\text{peak}}$ range: $19.0\text{--}45.3 \text{ ml} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$). To test the impact of CRF level on the detection of VI-1 and VI-2 in males and females the investigated subjects were categorised into the following subgroups: (1) CRF-, representing the ≤ 20 . Percentile: $\text{VO}_{2\text{peak}} < 27.0 \text{ ml} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$ in males and $< 21.5 \text{ ml} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$ in females. (2) CRF+, representing the ≥ 60 . Percentile: $\text{VO}_{2\text{peak}} > 34.0 \text{ ml} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$ in males and $> 27.5 \text{ ml} \cdot \text{min}^{-1} \cdot \text{kg}^{-1}$ in females (Rapp, Scharhag, Wagenpfeil, & Scholl, 2018). Thereafter, 25 data-sets per subgroup were selected to analyse the VI-detection (Figure 1). The Paracelsus 10.000 Study conformed to the principles outlined in the Declaration of Helsinki and was approved by the local ethics committee (E1521). All participants gave written informed consent.

Data collection and processing

After a 2-minute stationary phase and a 2-minute warm-up phase at 10 W, each participant performed an incremental cycle ergometer task (ergo select 200P, ergo line GmbH, Bitz, Germany) until volitional exhaustion at a pedalling rate of 60 rpm. The test started at a workload of 40–60 W for females with an increase of 10–15 W every minute until exhaustion, while males started at 50–110 W with an increase of 10–15 W every minute. Starting workload and the consecutive work rates were dependent on sex and body mass to insure that volitional exhaustion was reached after 8 to 12 minutes of test duration (American College of Sports Medicine, 2013). A 5-minute recovery phase at 10 W was completed after test termination.

Attainment of volitional exhaustion was confirmed if at least two of the following criteria were met: (1) a plateau in VO_2 (2) $\text{EQO}_2 > 30$, (3) respiratory exchange ratio (RER) > 1.1 , (4) reaching 90% of age predicted maximum heart rate, (5) pedalling rate below 50 rpm. The exercise testing was terminated if any

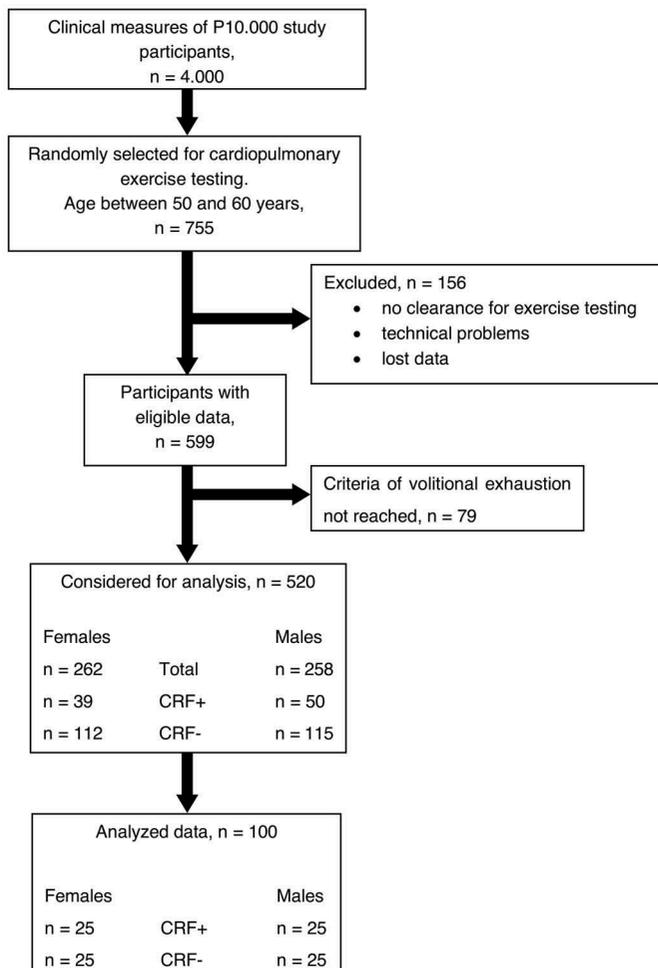


Figure 1. Participant flow. CRF+ = high cardiorespiratory fitness subgroup; CRF- = low cardiorespiratory fitness subgroup.

complications or contraindications occurred (Ross, 2003). Datasets were excluded from further analysis, if the participants could not complete at least five minutes of the incremental exercise test, i.e. if they had not completed at least five load levels.

Continuous respiratory gas analysis and volume measurements were obtained breath-by-breath (Master Screen CPX, Jaeger, Hoechberg, Germany). Data from the stationary, warm-up and recovery phase were excluded from further analysis. 10-second time-based averages were calculated and the mean of the three successive highest 10-second VO_2 -values were taken as $\text{VO}_{2\text{peak}}$. Peak power output (PPO) was determined as the mean power output during the last minute of the exercise test (Merry, Glaister, Howatson, & van Someren, 2016; Robergs & Burnett, 2003).

Visual method for VI detection

The visual detection of VI was conducted by two experts independently. VI-2 was determined by finding the disproportional increase in a VE vs. VCO_2 plot (Wasserman et al., 1981) and VI-1 was determined by finding the disproportional increase in a VCO_2 vs. VO_2 plot (*v*-slope method; Beaver et al., 1986). The ventilatory equivalents method (Meyer et al., 2005) was then

used to verify VI determination. The experts had no prior knowledge of any results or identity of the participants. In the presence of a difference greater $\pm 300 \text{ ml O}_2 \cdot \text{min}^{-1}$ in VI detection between the two experts (Gaskill et al., 2001), a third expert identified the respective VI after which the two closest VO_2 values were used for further analysis. As primary outcome the mean VO_2 of the two detected values was used to denote VI-1_visual and VI-2_visual (Santos & Giannella-Neto, 2004). The visual detection method served as our reference method and was compared to the automatic and semiautomatic detection methods (Dolezal et al., 2017; Novais et al., 2015).

Automatic methods using polynomial regression for VI detection

The VI-1 and VI-2 were automatically detected by fitting a 6th order polynomial (method: linear least squares; robust fitting option: bi-square) to the gas exchange and ventilatory time series of interest. The first four derivatives were further used to calculate local minima (min_E QO_2 , min_PET O_2 , min_EQCO $_2$), local maxima (max_PETCO $_2$), and local maximal positive curvatures (maxcurv_E QO_2 , maxcurv_PET O_2 , maxcurv_EQCO $_2$, maxcurv_VCO $_2/\text{VO}_2$, maxcurv_VE/VCO $_2$) and negative curvatures (maxcurv_PETCO $_2$) of these polynomial functions (Cross, Morris, Schneider, & Sabapathy, 2012; Sherrill et al., 1990; Wisén & Wohlfart, 2004). VI-1 was defined as VO_2 equal to the min_E QO_2 , min_PET O_2 , maxcurv_E QO_2 , maxcurv_PET O_2 and maxcurv_VCO $_2/\text{VO}_2$. Further, VI-2 was defined as VO_2 equal to min_EQCO $_2$, max_PETCO $_2$, maxcurv_EQCO $_2$, maxcurv_PETCO $_2$, and maxcurv_VE/VCO $_2$ (Meyer et al., 2005). An *n*th order polynomial function can reveal a maximum of *n*-1 extrema. The curvature of the fitted 6th order polynomial is a 4th order polynomial function with three possible extrema (i.e. two minima and one maximum, or vice versa). Therefore, in the case of the occurrence of two extrema (two minima or two maxima), VI-1 was defined as the first and VI-2 as the second extrema (Sherrill et al., 1990).

Semiautomatic method for VI detection

A semiautomatic method (semiauto) with a systematic approach using automatic and visual detection (conducted by an expert as depicted in Table 1) methods was developed to detect VI-1 and VI-2. For this purpose, different plots of the ventilatory variables together with the automatically detected VI (as described above) were reviewed and the expert visually chooses the most appropriate value (Figure 2). To avoid false identification of VI-1 as VI-2, the VI-2 was detected first and then shown in the plots to help VI-1 detection. Finally, the expert assigned a level of confidence for VI-1 and VI-2 detection. The confidence levels were defined equal for VI-1 and VI-2 detection as follows (method A and B refer to the description in Table 1):

Confidence Level 1 (two of the following three criteria had to be fulfilled): (1) The difference in the selected time point for the VI between the first (A) and the second (B) method was $\leq 30 \text{ s}$. (2) The selected VI could be clearly identified visually in both methods (A and B). (3) At least one of the automatically

detected VIs was in agreement with the visually selected VI within a limit of ± 30 s.

Confidence Level 2 (two of the following three criteria had to be fulfilled): (1) The difference in the selected time point for the VI between the first (A) and the second (B) method was ≤ 60 s. (2) The selected VI could be clearly identified visually with one of the two methods (A and B). (3) At least one of the automatically detected VIs was in agreement with the visually selected VI within a limit of ± 60 s.

Confidence Level 3: all cases which didn't fulfil the criteria for confidence level 1 and 2 were defined as indeterminable.

The VO_2 values corresponding to the selected time points of VI-1 and VI-2, detected with the different automatic and semi-automatic methods described above, were then obtained by fitting a 6th order polynomial (method: linear least squares; robust fitting option: bi-square) to the VO_2 time series data.

Statistical analysis

Data are given as mean \pm standard deviation. Two-way ANOVA and post-hoc tests with Bonferroni-adjusted α were conducted to compare CRF-subgroups separated by sex (Table 2). The agreements between the experts for visual detection was assessed by calculating intraclass correlation coefficients (ICC [95% confidence interval]) based on a single-rater, absolute-agreement, and a two way mixed effects model (Koo & Li, 2016).

VO_2 values between detection methods and CRF-subgroups were compared separately for VI-1 and VI-2 using two-way repeated-measures ANOVA. Post-hoc tests with Bonferroni-adjusted α were conducted to identify the differences between the detection methods.

The level of absolute agreement between the visual and the other two methods (automatic and semiautomatic) were evaluated by calculating ICC as described above and mean biases $\pm 95\%$ limits of agreement (LoA) according to Bland and Altman (2010) for VI-1, VI-2, and each CRF subgroup.

A statistical power analysis was performed a priori for sample size estimation using G*Power version 3.1.9.2 (Faul, Erdfelder, Lang, & Bucher, 2007). With an assumed Type I error of .05 and a Type II error of .20 for finding a large effect (f value = .4 (Cohen, 1969)) between CRF level groups, the

projected sample size is approximately 26 participants per group. Thus, our proposed sample size of 50 subjects per CRF level group should be more than adequate for the main objective of this study. The level of significance was set at $\alpha < .05$. The statistical analyses were performed using RStudio version 1.1.383 (RStudio Inc., Boston, Massachusetts, USA).

Results

Participant characteristics

Table 2 lists age, anthropometric characteristics and results of CPET for all CRF subgroups. There were no significant differences between the CRF subgroups in relation to age, height, VI-1_visual and VI-2_visual. For body mass and body mass index (BMI) we found significant higher values in participants of the low CRF subgroups. As expected, PPO, VO_{2peak} , peak heart rate (HR_{peak}) and number of one-minute steps to reach volitional exhaustion during CPET ($step_{max}$) were significantly higher in participants of the high CRF subgroups.

Agreement between researchers of the visual detection method

Agreement between the two researchers of the visual detection of 100 cases showed good to excellent reliability for VI-1 (ICC = .93 [.80, .97]) and excellent reliability for VI-2 (ICC = .98 [.98, .99]). Because of the high expert agreement and in accordance with others, the visually detected VI served as our reference method for further comparisons with the automatic and semiautomatic methods (Higa et al., 2007; Novais et al., 2015; Santos & Giannella-Neto, 2004).

Number of detectable cases

Of all cases, VI-1/VI-2 was successfully identified 88%/94% visually, 97%/89% through min_EQO_2 , 96%/85% through min_PETO_2 , 65%/85% through $maxcurv_EQO_2$, 58%/83% through $maxcurv_PETO_2$ and 92%/93% through $maxcurv_VCO_2/VO_2$. The semiautomatic method revealed the highest number of detectable cases in each CRF subgroup for VI-1 (Table 3) and VI-2 (Table 4).

Table 1. Guideline for the semiautomatic VI detection method.

1 st step: VI-2 detection
(A1) The inflection point in the VE vs. VCO_2 plot was visually identified, with the calculated VI-2 ($maxcurv_VE/VCO_2$) shown on the plot for guidance (Figure 2(a)).
(B1) To verify VI-2 selection, $EQCO_2$, $PETCO_2$, EQO_2 , and $PETO_2$ time plots were drawn and the identified inflection point from (A1) as well as the calculated VI-2s (min_EQO_2 , $maxcurv_EQO_2$, max_PETCO_2 and $maxcurv_PETO_2$) were shown on the plots for guidance. Based on the displayed information, a final visual selection of VI-2 was made. If the results of method (A1) and (B1) didn't agree, the result of method (A1) was preferred (Figure 2(b,c)).
(C1) A confidence level for VI-2 selection was given. [#]
2 nd step: VI-1 detection
(A2) The first inflection point in the VCO_2 vs. VO_2 plot was visually identified, with the calculated VI-1 ($maxcurv_VCO_2/VO_2$), the line of identity and the previously selected VI-2 shown on the plot for guidance (Figure 2(d)).
(B2) To verify VI-1 selection, EQO_2 , $PETO_2$, $EQCO_2$ and $PETCO_2$ time series plots were drawn and the identified inflection point from (A2), as well as the calculated VI-1s (min_EQO_2 , $maxcurv_EQO_2$, min_PETO_2 and $maxcurv_PETO_2$) and the previously selected VI-2 (1 st step) were shown on the plots for guidance. Based on the displayed information, a final visual selection of VI-1 was made. If the results of method (A2) and (B2) didn't agree, the result of method (A2) was preferred (Figure 2(e,f)).
(C2) A confidence level for VI-1 selection was given. [#]

[#]The confidence levels were defined equal for both VI-1 and VI-2 detection and are described in the text.

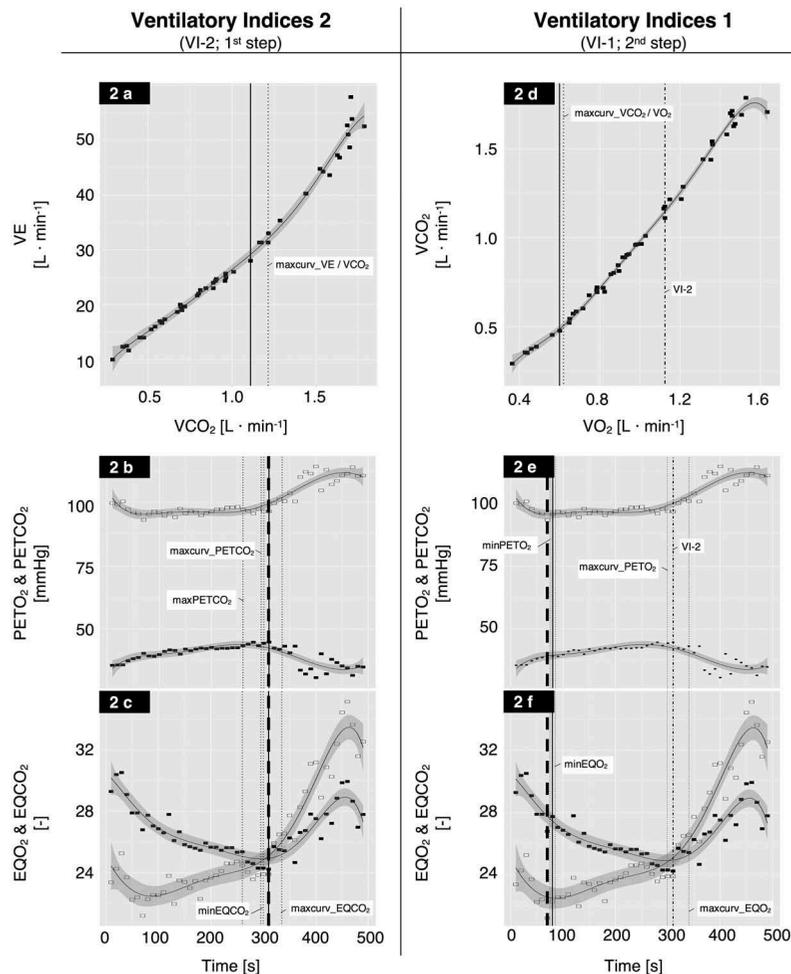


Figure 2. Sample ventilatory indices (VI) determination using the semiautomatic method for one subject. The left panels (a–c) illustrates the first step of the detection procedure (VI-2), the right panels (d–f) illustrates the second step (VI-1). The data-points (open and closed symbols) represent the 10-second time-based averaged breath-by-breath results of the CPET. Open symbols = PETO_2 (2b/2e) and EQO_2 (2c/2f); Closed symbols = PETCO_2 (2b/2e) and EQCO_2 (2c/2f); The solid lines along the data-points represent the fitted 6th degree polynomial with the 95% confidence interval (dark grey area). The dotted vertical lines represent the automatically calculated VIs, and the solid dashed thick lines represent the visually selected VIs. The vertical dashed thick lines in the panels 2b/2c and 2e/2f represent the visually selected VI-2 from panel 2a and the visually selected VI-1 from panel 2d, respectively. The dotted and dashed vertical lines in panels 2d – 2f represent the finally selected VI-2 from the 1st step (panels 2a – 2c).

VO₂ obtained via different detection methods at VI-1 and VI-2

The absolute values (VO_2 in $\text{ml O}_2 \cdot \text{min}^{-1}$) corresponding to VI-1 and VI-2 detected via the different methods are shown in Tables 3 and 4. The estimates for VI-1 ranged from $977 \pm 319 \text{ ml O}_2 \cdot \text{min}^{-1}$ (min_ EQO_2) to $1334 \pm 537 \text{ ml O}_2 \cdot \text{min}^{-1}$ (maxcurv_ PETO_2) and for VI-2 from $1418 \pm 457 \text{ ml O}_2 \cdot \text{min}^{-1}$ (max_ PETCO_2) to $1735 \pm 597 \text{ ml O}_2 \cdot \text{min}^{-1}$ (visual). The CRF- subgroup resulted generally in lower estimates for VI-1 and VI-2 than the CRF+ subgroup.

For VI-1, two-way repeated-measures ANOVA revealed a significant main effect for detection method ($F(6, 246) = 6.6$, $p < .001$, $\eta^2_p = .27$), a significant main effect for CRF level ($F(1, 41) = 4.6$, $p < .01$, $\eta^2_p = .10$) but no significant interaction between the two. For VI-2, two-way repeated-measures ANOVA revealed a significant main effect for detection method ($F(6, 372) = 7.0$, $p < .001$, $\eta^2_p = .26$), a significant main effect for CRF level ($F(1, 62) = 21.8$, $p < .001$, $\eta^2_p = .26$), and a significant interaction between detection method and CRF level ($F(6, 372) = 2.8$, $p < .05$, $\eta^2_p = .04$).

Post hoc testing revealed no significant differences of VO_2 to the visual method for maxcurv_ EQO_2 , maxcurv_ PETO_2 , and maxcurv_ VCO_2/VO_2 . Additionally, no significant differences were found in the semiauto method for VI-1 detection when comparing the total sample or the CRF+ subgroup. In the CRF- subgroup, none of the automatic methods differed significantly to the visual method (Table 3). For VI-2, no significant differences to the visual method were found for maxcurv_ PETCO_2 and semiauto method when comparing the total sample or the CRF subgroups.

Agreement between methods

The highest level of absolute agreement of the visual method was found for the semiauto method at VI-1 (ICC = .88 [.81, .92]) and VI-2 (ICC = .97 [.96, .98]) in the total sample. Similar results were found for both CRF subgroups (see Tables 3 and 4). The lowest levels of agreements were found for maxcurv_ EQO_2 , maxcurv_ PETO_2 and maxcurv_ VCO_2/VO_2 method for VI-1, and

Table 2. Characteristics of cardiopulmonary exercise test (CPET) participants.

Total, n	Total	CRF-	CRF+	p
	100	50	50	
Age, yrs	55 ± 3	56 ± 3	55 ± 3	ns
Height, m	1.7 ± 0.1	1.7 ± 0.1	1.7 ± 0.1	ns
Body mass, kg	76 ± 15	83 ± 15	68 ± 10	< .001
BMI, kg · m ⁻²	25.7 ± 4.3	28.2 ± 4.3	23.2 ± 2.6	< .001
PO _{VI-1} , W · kg ⁻¹	1.1 ± 0.4	0.9 ± 0.2	1.4 ± 0.4	< .001
PO _{VI-2} , W · kg ⁻¹	2.0 ± 0.9	1.4 ± 0.3	2.6 ± 0.8	< .001
PPO, W · kg ⁻¹	2.4 ± 0.8	1.7 ± 0.3	3.1 ± 0.6	< .001
VI-1_visual, %VO _{2peak}	54 ± 10	58 ± 10	51 ± 9	ns
VI-2_visual, %VO _{2peak}	83 ± 10	83 ± 10	82 ± 10	ns
VO _{2peak} , ml · min ⁻¹ · kg ⁻¹	28.0 ± 8.9	20.1 ± 1.9	36.0 ± 5.1	< .001
HR _{VI-1} , min ⁻¹	113 ± 14	114 ± 15	113 ± 14	ns
HR _{VI-1} , %HR _{peak}	71 ± 8	75 ± 6	68 ± 7	< .001
HR _{VI-2} , min ⁻¹	142 ± 18	135 ± 15	149 ± 18	< .001
HR _{VI-2} , %HR _{peak}	89 ± 6	89 ± 6	89 ± 7	ns
HR _{peak} , min ⁻¹	159 ± 16	152 ± 15	166 ± 14	< .001
step _{max}	9 ± 3	7 ± 1	12 ± 2	< .001
Males, n				
Age, yrs	55 ± 3	56 ± 3	55 ± 3	ns
Height, m	1.8 ± 0.1	1.8 ± 0.1	1.8 ± 0.1	ns
Body mass, kg	83 ± 15	92 ± 14	74 ± 8	< .001
BMI, kg · m ⁻²	26.5 ± 4.3	29.0 ± 4.3	23.9 ± 2.5	< .001
PO _{VI-1} , W · kg ⁻¹	1.3 ± 0.4	1.1 ± 0.2	1.6 ± 0.5	< .001
PO _{VI-2} , W · kg ⁻¹	2.3 ± 0.4	1.6 ± 0.2	2.9 ± 0.4	< .001
PPO, W · kg ⁻¹	2.7 ± 0.9	1.9 ± 0.2	3.5 ± 0.4	< .001
VI-1_visual, %VO _{2peak}	55 ± 11	59 ± 9	52 ± 12	ns
VI-2_visual, %VO _{2peak}	85 ± 8	85 ± 8	85 ± 8	ns
VO _{2peak} , ml · min ⁻¹ · kg ⁻¹	29.9 ± 9.2	21.7 ± 1.2	40.2 ± 2.7	< .001
HR _{VI-1} , min ⁻¹	114 ± 16	116 ± 16	112 ± 16	ns
HR _{VI-1} , %HR _{peak}	72 ± 8	76 ± 7	68 ± 8	.005
HR _{VI-2} , min ⁻¹	143 ± 17	138 ± 14	147 ± 19	ns
HR _{VI-2} , %HR _{peak}	89 ± 5	90 ± 5	88 ± 6	ns
HR _{peak} , min ⁻¹	158 ± 17	152 ± 16	166 ± 17	.007
step _{max}	9 ± 3	7 ± 1	11 ± 2	< .001
Females, n				
Age, yrs	55 ± 3	56 ± 3	54 ± 8	ns
Height, m	1.6 ± 0.1	1.6 ± 0.1	1.7 ± 0.2	ns
Body mass, kg	68 ± 11	74 ± 11	67 ± 13	.002
BMI, kg · m ⁻²	25.0 ± 4.2	27.4 ± 4.2	22.8 ± 3.8	< .001
PO _{VI-1} , W · kg ⁻¹	1.0 ± 0.3	0.8 ± 0.1	1.2 ± 0.3	< .001
PO _{VI-2} , W · kg ⁻¹	1.8 ± 0.9	1.2 ± 0.3	2.3 ± 1.0	< .001
PPO, W · kg ⁻¹	2.1 ± 0.6	1.6 ± 0.2	3.0 ± 0.7	< .001
VI-1_visual, %VO _{2peak}	54 ± 9	57 ± 10	50 ± 7	ns
VI-2_visual, %VO _{2peak}	80 ± 12	81 ± 12	80 ± 12	ns
VO _{2peak} , ml · min ⁻¹ · kg ⁻¹	24.0 ± 6.6	18.4 ± 0.8	35.4 ± 6.9	< .001
HR _{VI-1} , min ⁻¹	113 ± 13	112 ± 13	113 ± 13	ns
HR _{VI-1} , %HR _{peak}	71 ± 7	75 ± 6	68 ± 7	.005
HR _{VI-2} , min ⁻¹	142 ± 18	133 ± 15	151 ± 17	.003
HR _{VI-2} , %HR _{peak}	89 ± 7	88 ± 6	90 ± 7	ns
HR _{peak} , min ⁻¹	159 ± 15	152 ± 15	164 ± 25	.002
step _{max}	9 ± 3	7 ± 1	12 ± 2	< .001

Note: Data are presented as mean ± standard deviation after two-way ANOVA; CRF+ = high cardiorespiratory fitness level subgroup; CRF- = low cardiorespiratory fitness level subgroup; BMI = body mass index; PO_{VI-1} and PO_{VI-2} = power output at VI-1 and VI-2 using the visual detection method; PPO = peak power output; VO_{2peak} = peak oxygen consumption; VI-1_visual and VI-2_visual = visually detected ventilatory indices 1 and 2, respectively; HR_{VI-1} and HR_{VI-2} = heart rate at VI-1 and VI-2 using the visual detection method; HR_{peak} = peak heart rate; step_{max} = number of one-minute steps to reach volition exhaustion during CPET. p = significance level between the CRF subgroups; ns = not significant.

max_PETCO₂ and maxcurv_VE/VCO₂ method for VI-2 in the total sample as well as in both CRF subgroups.

The mean biases ± LoA for the visually detected VI are shown in Table 3 and Figure 3 for VI-1 and in Table 4 and Figure 4 for VI-2. Although the mean bias of comparison between the visual and the various automatic methods was small (ranging from

−292 to −216 ml O₂ · min⁻¹), the dispersion of agreement was large with the LoA ranging from ±268 to ±777 ml O₂ · min⁻¹ in the total sample. The smallest LoA were found for min_EQO₂, min_PETO₂ and semiauto at VI-1, and for semiauto at VI-2 in the total sample and in both CRF subgroups. The confidence levels for semiautomatic detection method can be seen in Table 5.

Discussion

It has been shown previously, that changes in gas exchange analysed visually or automatically, and blood lactate accumulations correspond with each other. Therefore, the detection of VI from gas exchange and ventilatory variables is an appropriate non-invasive method to describe metabolic changes during incremental CPET (Binder et al., 2008; Meyer et al., 2005; Westhoff et al., 2013). The purpose of this study was to extend these findings by providing additional information about the accuracy of different methods used to detect VI-1 and VI-2 in a group of 50–60 years old males and females, and to investigate the impact of CRF level on VI detection. Further, we created a recommendation for a systematic VI detection method with clearly defined confidence levels.

Participant characteristics

Body mass and BMI, as well as the results of CPET (PPO, VO_{2peak}, HR_{peak}, step_{max}) showed significant differences between the CRF subgroups (Table 2). Compared to reference values for VO_{2peak} published by Rapp et al. (2018), the CRF- and CRF+ subgroup presented with a mean VO_{2peak} correspond approximately to the 10th and 80th percentile, respectively.

In the current study, the visual detection method (the reference method in this study) revealed a VI-1 of 54 ± 10% of VO_{2peak} and VI-2 of 83 ± 10% of VO_{2peak} in the total sample. No significant differences for VI-1 and VI-2 could be found between the CRF subgroups. These findings are in agreement with other studies using automatic as well as visual detection methods (Gaskill et al., 2001; Meyer et al., 2005; Santos & Giannella-Neto, 2004; Wisén & Wohlfart, 2004).

Number of detectable cases

The semiautomatic method proved to be the detection method with the highest receiver rate compared to any of the automatic methods for VI-1 and VI-2 in the total sample as well as in both CRF subgroups. For VI-1 detection in the CRF- subgroup, the maxcurv_EQO₂ and maxcurv_PETO₂ resulted in detection rates of only 58% and 42%, respectively. Low detection rates of single automatic methods have been reported previously by others (e.g. 66% for VI-1 and 60% for VI-2 (Wisén & Wohlfart, 2004)). Furthermore, an improvement of detection rates by combining different visual detection methods were reported by Gaskill et al. (2001). To our knowledge this is the only study showing an improvement in detection rates by combining automatic and visual detection methods. Though improved levels of detection rates are valuable, the accuracy of a combined detection method should also be high.

Table 3. Mean \pm standard deviation of oxygen uptake (VO_2) at ventilatory indices 1 determined by visual and automatic detection methods. Intraclass correlation coefficients (ICC) and mean bias \pm 95% limits of agreement (LoA) for comparisons between visual and the different automatic detection methods (5 single- and 1 semiautomatic).

Ventilatory indices 1	Total, n = 100	Visual		Automatic		Comparisons		
		n_{dc} [%]	VO_2 [ml \cdot min $^{-1}$]	Methods	n_{dc} [%]	VO_2 [ml \cdot min $^{-1}$]	ICC [95% CI]	LoA [ml \cdot min $^{-1}$]
Ventilatory indices 1	Total, n = 100	88	1099 \pm 359	*min_EQO ₂	97	977 \pm 319	.79 [.43, .90]	-135 \pm 346
				*min_PETO ₂	96	986 \pm 317	.85 [.48, .94]	-124 \pm 282
				Maxcurv_EQO ₂	65	1214 \pm 454	.56 [.35, .72]	129 \pm 681
				Maxcurv_PETO ₂	58	1334 \pm 537	.49 [.20, .69]	216 \pm 777
				Maxcurv_VCO ₂ /VO ₂	92	1022 \pm 364	.63 [.46, .75]	-110 \pm 579
	CRF-, n = 50	88	972 \pm 291	Semiautomatic	97	1183 \pm 399	.88 [.81, .92]	55 \pm 334
				Min_EQO ₂	94	856 \pm 261	.84 [.44, .94]	-107 \pm 247
				Min_PETO ₂	92	865 \pm 273	.87 [.51, .95]	-97 \pm 224
				Maxcurv_EQO ₂	58	1032 \pm 309	.75 [.51, .88]	86 \pm 417
				Maxcurv_PETO ₂	42	989 \pm 263	.75 [.45, .90]	90 \pm 399
	CRF+, n = 50	88	1227 \pm 377	Maxcurv_VCO ₂ /VO ₂	88	887 \pm 268	.73 [.47, .86]	-103 \pm 372
				Semiautomatic	94	993 \pm 277	.91 [.84, .95]	28 \pm 239
				*min_EQO ₂	100	1091 \pm 328	.71 [.29, .87]	-164 \pm 417
				*min_PETO ₂	100	1097 \pm 316	.81 [.34, .93]	-150 \pm 325
				Maxcurv_EQO ₂	72	1361 \pm 501	.41 [.09, .66]	166 \pm 847
			Maxcurv_PETO ₂	74	1531 \pm 556	.34 [.01, .61]	293 \pm 910	
			mAxcurv_VCO ₂ /VO ₂	96	1145 \pm 398	.51 [.26, .70]	-116 \pm 732	
			Semiautomatic	100	1362 \pm 416	.83 [.69, .91]	82 \pm 404	

n_{dc} ... number of detectable cases associated with each method, VO_2 ... mean \pm standard deviation of detectable cases. * indicating significant difference ($p < .05$) in relation to visual detection method by two-way repeated-measures ANOVA with 43 complete cases for total, 27 for CRF+ and 16 for CRF- subgroup.

Table 4. Mean \pm standard deviation of oxygen uptake (VO_2) at ventilatory indices 2 determined with visual and automatic detection methods. Intraclass correlation coefficients (ICC) and mean bias \pm 95% limits of agreement (LoA) for comparisons between visual and the different automatic detection methods (5 single- and 1 semiautomatic).

Ventilatory indices 2	Total, n = 100	visual		automatic		comparisons		
		n_{dc} [%]	VO_2 [ml \cdot min $^{-1}$]	methods	n_{dc} [%]	VO_2 [ml \cdot min $^{-1}$]	ICC [95% CI]	LoA [ml \cdot min $^{-1}$]
Ventilatory indices 2	Total, n = 100	94	1735 \pm 597	*min_EQCO ₂	89	1504 \pm 504	.82 [.41, .92]	-224 \pm 511
				*max_PETCO ₂	85	1418 \pm 457	.74 [.16, .89]	-292 \pm 563
				*maxcurv_EQCO ₂	85	1548 \pm 545	.82 [.71, .89]	-123 \pm 639
				maxcurv_PETCO ₂	83	1689 \pm 581	.85 [.78, .90]	-69 \pm 626
				*maxcurv_VE/VCO ₂	93	1464 \pm 596	.74 [.45, .86]	-247 \pm 727
	CRF-, n = 50	88	1376 \pm 415	Semiautomatic	96	1710 \pm 562	.97 [.96, .98]	1 \pm 268
				*min_EQCO ₂	90	1254 \pm 387	.88 [.53, .96]	-131 \pm 295
				*max_PETCO ₂	90	1233 \pm 394	.87 [.35, .95]	-157 \pm 301
				maxcurv_EQCO ₂	88	1302 \pm 388	.89 [.80, .94]	-56 \pm 372
				maxcurv_PETCO ₂	82	1413 \pm 408	.92 [.84, .96]	0 \pm 339
	CRF+, n = 50	100	2051 \pm 554	*maxcurv_VE/VCO ₂	92	1142 \pm 298	.56 [.07, .79]	-254 \pm 557
				Semiautomatic	94	1401 \pm 388	.96 [.92, .98]	23 \pm 232
				*min_EQCO ₂	88	1760 \pm 485	.69 [.10, .88]	-313 \pm 608
				*max_PETCO ₂	80	1626 \pm 436	.55 [-.08, .83]	-437 \pm 641
				maxcurv_EQCO ₂	82	1813 \pm 568	.70 [.46, .84]	-192 \pm 811
			maxcurv_PETCO ₂	84	1959 \pm 601	.74 [.56, .86]	-134 \pm 790	
			*maxcurv_VE/VCO ₂	94	1779 \pm 646	.69 [.42, .83]	-241 \pm 859	
			Semiautomatic	98	2007 \pm 546	.96 [.93, .98]	-20 \pm 294	

n_{dc} ... number of detectable cases associated with each method, VO_2 ... mean \pm standard deviation of detectable cases. * indicating significant difference ($p < .05$) in relation to visual detection method by two-way repeated-measures ANOVA with 64 complete cases for total, 28 for CRF+ and 36 for CRF- subgroup.

Agreement between the visual and the various automatic methods

For VI-1 detection, we found higher agreement (expressed as ICC and mean bias \pm LoA) between visual, and min_EQO₂ and min_PETO₂ method than for the "maxcurv" detection methods (maxcurv_EQO₂, maxcurv_PETO₂ and max_VCO₂/VO₂) in the total sample as well as in both CRF subgroups. As described by Meyer et al. (2005), the EQO₂ time series can produce two inflection points, a first corresponding to VI-1 and a second caused by hyperventilation at VI-2. When using automatic as well as visual detection methods, the second inflection point could mistakenly be identified as VI-1 (e.g. as a consequence to irregularities in breathing pattern at VI-1 or a greater disproportional increase at VI-2), which is "[...] the single most common

methodological error in the literature" as stated by McLellan (1987). This second inflection point could also be a reason for the slightly smaller inter-rater reliability for visual detection of VI-1 than VI-2. This smaller inter-rater reliability for VI-1 detection is substantiated by the high number of cases (17 out of 100), where the disagreement between the two experts during the visual detection was greater 300 ml O₂ \cdot min $^{-1}$ and therefore had to be verified by a third expert.

For VI-2 detection, agreement was similar throughout all applied automatic methods in the total sample with ICC ranging from .74 to .85, mean biases ranging from -292 to -69 ml O₂ \cdot min $^{-1}$, and LoA ranging from \pm 511 to \pm 727 ml O₂ \cdot min $^{-1}$. Despite the ICC-results have shown moderate to good reliability, the result of the LoA analysis indicates substantial

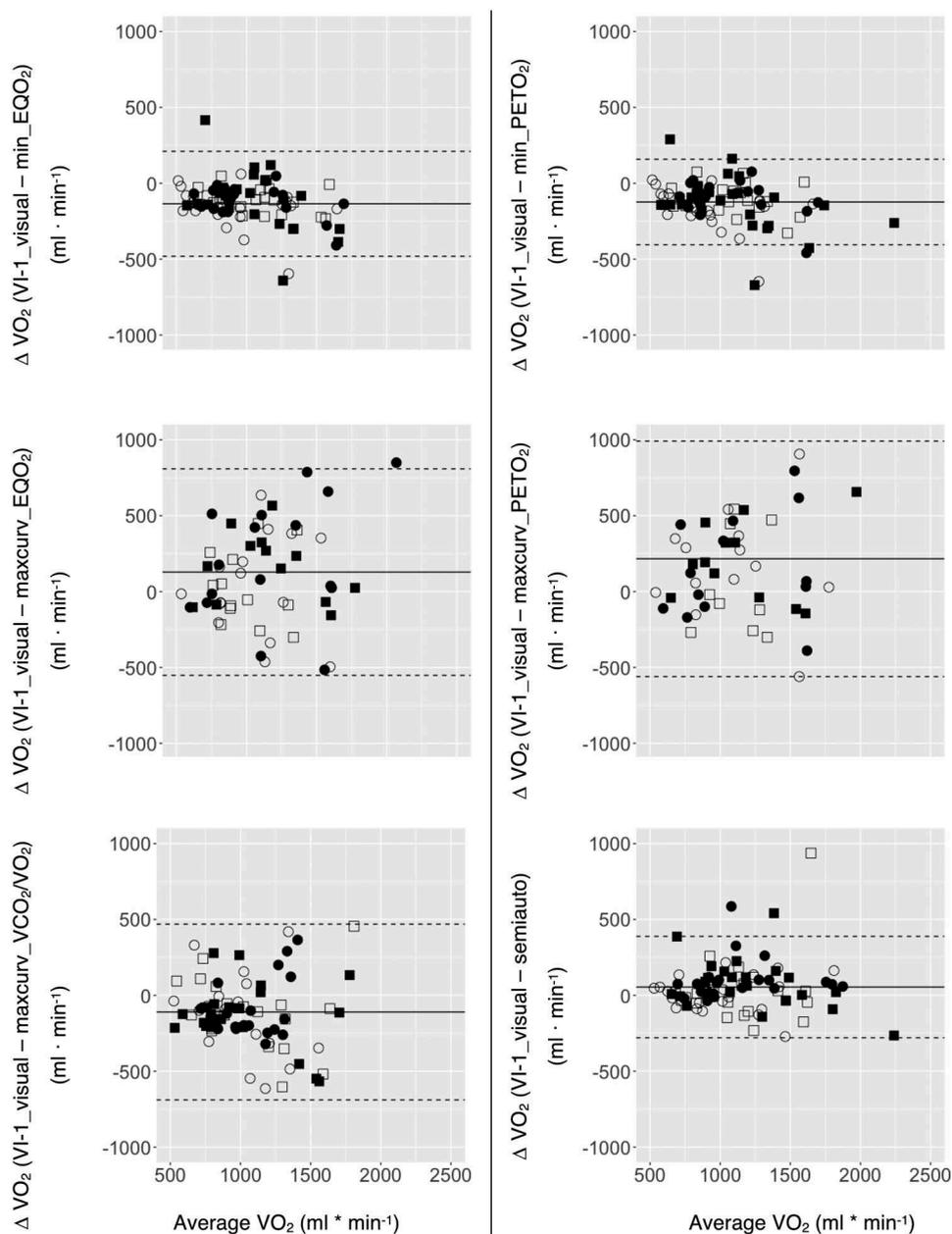


Figure 3. Bland-Altman plots of the difference between the determinations of ventilatory indices 1 (VI-1) by visual, automatic (min_EQO₂, min_PETO₂, maxcurv_EQO₂, maxcurv_PETO₂ and maxcurv_VCO₂/VO₂) and semiautomatic (semiauto) methods vs. the mean of their determinations. The solid horizontal line represents the mean bias between the two methods, and the top and bottom dashed lines represent the 95% limits of agreement [$\pm 1.96 \cdot$ standard deviation]. Open symbols = low cardiorespiratory fitness subgroup; Closed symbols = high cardiorespiratory fitness subgroup; Circles = females; Squares = males.

differences between the visual and the different automatic detection methods. Similar conclusions can be drawn from the analysis of the CRF subgroups for VI-2 detection.

A thorough examination of the Bland-Altman plots for all automatic methods (Figures 3 and 4) revealed similar inter-method differences (y-axis) throughout the full range of their mean estimates (x-axis). It is noteworthy that, for nearly every automatic method we found subjects with values far outside the concordance intervals. This was found for the total sample as well as for both CRF subgroups (empty vs. filled symbols). Therefore, CRF level does not seem to have an impact on VI detection in 50 to 60 years old adults.

Despite the high ICC found for some of the automatic detection methods, taking into account the high LoA and the number of subjects with values outside the concordance intervals, none of the automatic detection methods used in our study can be recommended for VI-1 as well as VI-2 detection in 50 to 60 years old adults.

The semiautomatic detection method

It has been shown by Dolezal et al. (2017), that a systematic approach combining detection methods based on different ventilatory and gas exchange response variables can increase agreement on VO₂ at VI-1 between interpreters. However, the

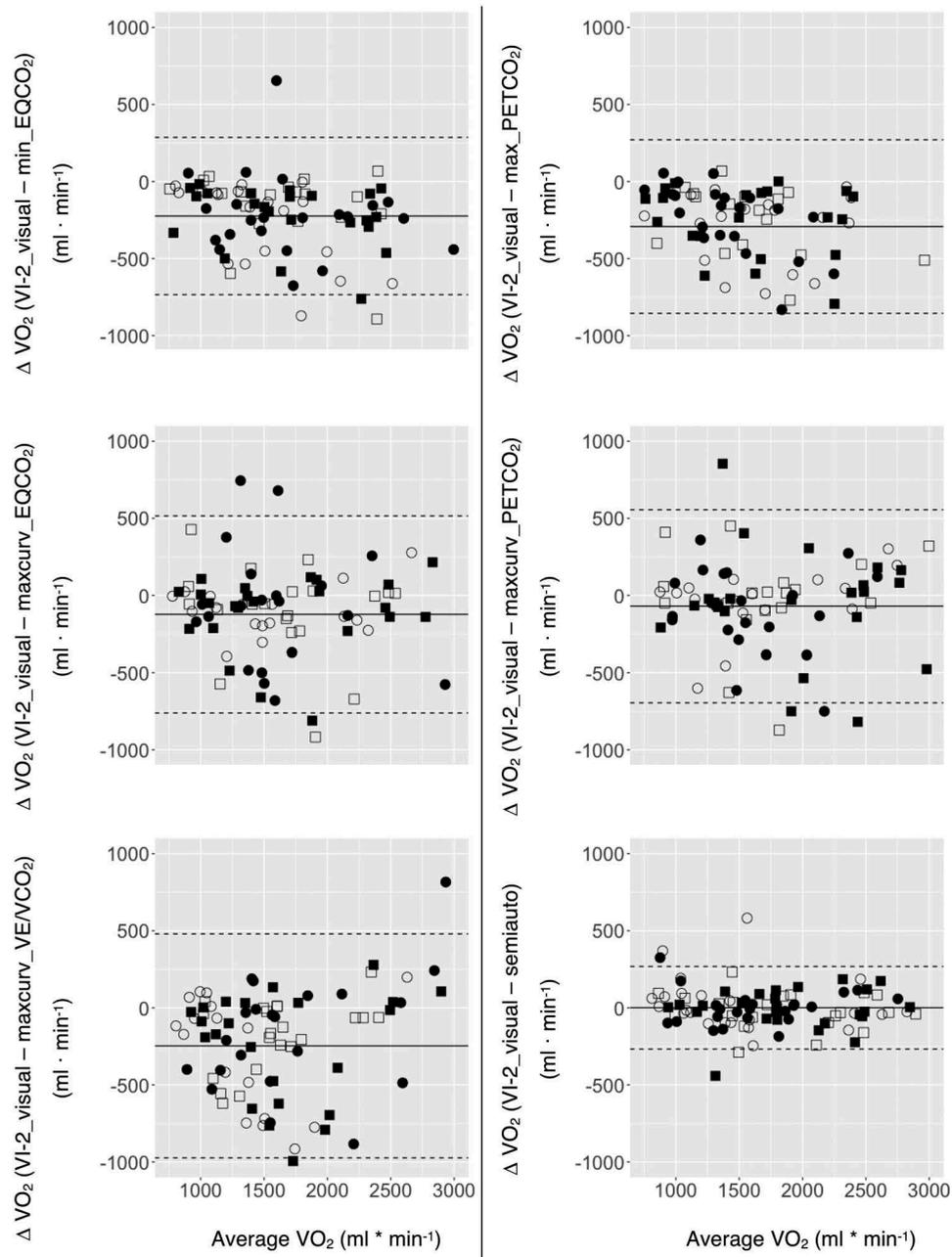


Figure 4. Bland-Altman plots of the difference between the determinations of ventilatory indices 2 (VI-2) by visual, automatic (min_EQCO₂, max_PETCO₂, maxcurv_EQCO₂, maxcurv_PETCO₂ and maxcurv_VE/VCO₂) and semiautomatic (semiauto) methods vs. the mean of their determinations. The solid horizontal line represents the mean bias between the two methods, and the top and bottom dashed lines represent the 95% limits of agreement [$\pm 1.96 \cdot$ standard deviation]. Open symbols = low cardiorespiratory fitness subgroup; Closed symbols = high cardiorespiratory fitness subgroup; Circles = females; Squares = males.

Table 5. Percentage of confidence level rating by the expert using the semiautomatic method for ventilatory indices (VI-1 and VI-2) determination.

Confidence level		1	2	3
VI-1	CRF+	52%	48%	0%
	CRF-	74%	20%	6%
VI-2	CRF+	64%	34%	2%
	CRF-	78%	16%	6%

agreement is strongly dependent on the experience of the interpreters and the characteristics of the subjects investigated. In the referred study CPET data from 10 healthy, recreational active young men (25 ± 5 years) were analysed (Dolezal et al., 2017). Gaskill et al. (2001) used CPET data from three different

populations (athletes, active, and sedentary males and females aged between 15 to 52 years) to show improved agreement on VO₂ at VI-1 up to 11 ± 300 ml O₂ · min⁻¹ by combining three different visual detection methods. We found no studies which looked at combined detection methods for VI-2.

In our study, the semiautomatic detection method yielded the highest agreement for VI-1 and VI-2 in 50 to 60 years old adults, shown by the lowest mean bias \pm LoA and the highest ICC (Tables 3 and 4) in the total sample as well as in both CRF subgroups.

These improvements in inter-method agreement are similar to the studies mentioned above and therefore, this is the first study to show an improved accuracy of VI detection by

combining different automatic procedures along with visual judgement. The high confidence levels rated by the expert for the semiautomatic method (VI-1: level 1 = 63% and level 2 = 34%; VI-2: level 1 = 71% and level 2 = 25%) indicate that the semiautomatic systematic approach provides a useful support for VI detection.

Especially in clinical circumstances, where false VI detection can lead to harmful consequences, e.g. by misclassifying surgical candidates into inappropriate risk categories (Vainshelboim et al., 2017), the semiautomatic method with clearly defined confidence levels can support clinical personal in their decision making. However, further research is needed to improve confidence level definitions for specific populations (e.g. age classes, athletes, sedentary or clinical populations).

The lack of blood lactate measures to validate the visually detected VI can be seen as a limitation. However, previous reports have shown good agreement between visually detected VI based on respiratory and gas exchange measures, and lactate responses (Binder et al., 2008; Meyer et al., 2005; Westhoff et al., 2013). Therefore, it can be considered appropriate using the visually detected VI as our reference method. Because of the high number of not detectable cases in the various automatic methods, only 43/100 complete cases for VI-1 and 64/100 complete cases for VI-2 could be used to calculate the two-way repeated-measures ANOVA. Therefore, results should be interpreted with caution.

Conclusion

The results of this study demonstrate that automatic methods based on polynomial regression are useful for detecting VI in 50 to 60 years old adults. In this study we could not find any differences between the CRF subgroups regarding the accuracy of VI detection. However, as different patterns in the gas exchange and ventilatory response curves may exist, the detection of VI based on a single automatic method can lead to incorrect VI detection and therefore false classification of aerobic capacity or inadequate training recommendations. The proposed semiautomatic detection method using a systematic approach and clearly defined confidence levels seem promising in overcoming these limitations.

Author Contributions Statement

M.P, S.RD and S.T conceived and designed research. M.P conducted experiments, analyzed data and drafted the manuscript. B.I and B.P intellectually contributed the content of the manuscript. All authors read and approved the final version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Martin Pühringer  <http://orcid.org/0000-0002-8139-5213>
 Susanne Ring-Dimitriou  <http://orcid.org/0000-0002-5317-3891>
 Thomas Stöggel  <http://orcid.org/0000-0002-6685-1540>
 Bernhard Iglseider  <http://orcid.org/0000-0002-6132-8670>

References

- American College of Sports Medicine. (2013). *Guidelines for exercise testing and prescription*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Beaver, W. L., Wasserman, K., & Whipp, B. J. (1986). A new method for detecting anaerobic threshold by gas exchange. *Journal of Applied Physiology*, 60, 2020–2027.
- Binder, R. K., Wonisch, M., Corra, U., Cohen-Solal, A., Vanhees, L., Saner, H., & Schmid, J.-P. (2008). Methodological approach to the first and second lactate threshold in incremental cardiopulmonary exercise testing. *European Journal of Cardiovascular Prevention & Rehabilitation*, 15, 726–734.
- Bland, J. M., & Altman, D. G. (2010). Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*, 47, 931–936.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
- Cross, T. J., Morris, N. R., Schneider, D. A., & Sabapathy, S. (2012). Evidence of break-points in breathing pattern at the gas-exchange thresholds during incremental cycling in young, healthy subjects. *European Journal of Applied Physiology*, 112, 1067–1076.
- Dolezal, B. A., Storer, T. W., Neufeld, E. V., Smooke, S., Tseng, C.-H., & Cooper, C. B. (2017). A systematic method to detect the metabolic threshold from gas exchange during incremental exercise. *Journal of Sports Science & Medicine*, 16, 396–406.
- Ekkekakis, P., Lind, E., Hall, E. E., & Petruzzello, S. J. (2008). Do regression-based computer algorithms for determining the ventilatory threshold agree? *Journal of Sports Sciences*, 26, 967–976.
- Faul, F., Erdfelder, E., Lang, A., & Bucher, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Gaskill, S. E., Ruby, B. C., Walker, A. J., Sanchez, O. A., Serfass, R. C., & Leon, A. S. (2001). Validity and reliability of combining three methods to determine ventilatory threshold. *Medicine and Science in Sports and Exercise*, 33, 1841–1848.
- Higa, M. N., Silva, E., Neves, V. F. C., Catai, A. M., Gallo, L. Jr, & Silva de Sá, M. F. (2007). Comparison of anaerobic threshold determined by visual and mathematical methods in healthy women. *Brazilian Journal of Medical and Biological Research*, 40, 501–508.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163.
- McLellan, T. (1987). The anaerobic threshold: Concept and controversy. *Australian Journal of Science and Medicine in Sport*, 19, 3–8.
- Merry, K. L., Glaister, M., Howatson, G., & van Someren, K. (2016). The exercise intensity at maximal oxygen uptake (iVO 2max): Methodological issues and repeatability. *European Journal of Sport Science*, 1391, 1–7.
- Meyer, T., Lucía, A., Earnest, C. P., & Kindermann, W. (2005). A conceptual framework for performance diagnosis and training prescription from submaximal gas exchange parameters—theory and application. *International Journal of Sports Medicine*, 26(Suppl 1), S38–48.
- Novais, L. D., Silva, E., Simões, R. P., Sakabe, D., Martins, L., Oliveira, L., ... Catai, A. (2015). Anaerobic threshold by mathematical model in healthy and post-myocardial infarction men. *International Journal of Sports Medicine*, 37, 112–118.
- Rapp, D., Scharhag, J., Wagenpfeil, S., & Scholl, J. (2018). Reference values for peak oxygen uptake: Cross-sectional analysis of cycle ergometry-based cardiopulmonary exercise tests of 10 090 adult German volunteers from the prevention first registry. *BMJ Open*, 8, 1–11.
- Robergs, R. A., & Burnett, A. F. (2003). Methods used to process data from indirect calorimetry and their application to VO₂max. *Journal Of Exercise Physiology Online / American Society of Exercise Physiologists*, 6, 44–57.
- Ross, R. M. (2003). ATS/ACCP statement on cardiopulmonary exercise testing. *American Journal of Respiratory and Critical Care Medicine*, 167, 211–277.
- Salk. (2016) Paracelsus study. Retrieved from <http://www.salk.at/12103.html>
- Santos, E. L., & Giannela-Neto, A. (2004). Comparison of computerized methods for detecting the ventilatory thresholds. *European Journal of Applied Physiology*, 93, 315–324.

- Sherrill, D. L., Anderson, S. J., & Swanson, G. (1990). Using smoothing splines for detecting ventilatory thresholds. *Medicine & Science in Sports & Exercise*, 22, 684–689.
- Vainshelboim, B., Rao, S., Chan, K., Lima, R. M., Ashley, E. A., & Myers, J. (2017). A comparison of methods for determining the ventilatory threshold: Implications for surgical risk stratification. *Canadian Journal of Anesthesia/Journal Canadien D'anesthésie*, 64, 634–642.
- Wasserman, K., Hansen, J. E., Sue, D. Y., Whipp, B. J., & Froelicher, V. F. (2011). *Principles of exercise testing and interpretation* (5th ed.). Baltimore: Lippincott Williams & Wilkins.
- Wasserman, K., Whipp, B., & Davis, J. A. (1981). Respiratory physiology of exercise: Metabolism, gas exchange, and ventilatory control. *International Review of Physiology*, 23, 149–211.
- Westhoff, M., Rühle, K. H., Greiwing, A., Schomaker, R., Eschenbacher, H., Siepmann, M., & Lehnigk, B. (2013). Positional paper of the German working group “cardiopulmonary exercise testing” to ventilatory and metabolic (lactate) thresholds. *Dtsch Medizinische Wochenschrift*, 138, 275–280.
- Wisén, A. G. M., & Wohlfart, B. (2004). A refined technique for determining the respiratory gas exchange responses to anaerobic metabolism during progressive exercise - repeatability in a group of healthy men. *Clinical Physiology and Functional Imaging*, 24, 1–9.