Validity of the Actiheart step test for the estimation of maximum oxygen consumption in endurance athletes and healthy controls

Juliane Heydenreich[1,2], Yves Schutz[3], Bengt Kayser[2,+,*] & Katarina Melzer[1,+]

1 Swiss Federal Institute of Sport, Magglingen, Switzerland

2 Institute of Sport Sciences, University of Lausanne, Lausanne, Switzerland

3 Department of Physiology, University of Fribourg, Fribourg, Switzerland

+ Shared last authorship

* Corresponding author: Prof. Dr. Bengt Kayser, Institute of Sport Sciences, University of Lausanne, Synathlon Uni-Centre, 1015 Lausanne, Switzerland
Tel: +41 (0)21 692 37 95, Fax: +41 (0)21 692 32 93
Email: bengt.kayser@unil.ch

ORIGINAL ARTICLE

ABSTRACT

Submaximal step tests are often used for estimation of maximum oxygen consumption (VO_2max) in humans. The validity of the Actiheart step test for VO_2max estimation was not fully studied yet. Therefore, purpose of the study was to estimate VO_2max using the Actiheart step test and to compare the data with measured VO_2max in endurance trained athletes (ATH) and healthy non-athletes (CON). 68 ATH (54% men, 28.0±5.4 yrs, 20.9±1.7 kg·m^{-2}) and 63 CON (46% men, 27.6±5.1 yrs, 22.1±1.7 kg·m^{-2}) performed the Actiheart step test and a spiroergometry for assessment of VO_2max. In addition, resting metabolic rate (RMR; indirect calorimetry), maximum heart rate (HRmax; heart rate monitoring system during spiroergometry), and sleeping heart rate (SHR; Actiheart 6-day long term measurement) were determined. Validity of two different Actiheart software entry modes was assessed: (1) AHraw (estimated RMR [Schofield] and HRmax [Tanaka], SHR = 70 bpm) and (2) AHcomplete (measured RMR, HRmax, and SHR). Validity was investigated using linear regression (R^2 and standard error of the estimate (SEE)) and repeated-measures ANOVA with a Bonferroni post-hoc correction. The level of significance was set to α=0.05.

VO_2max estimated by AHraw was significant related to measured VO_2max in women CON (R^2=0.22; $p<0.05$), whereas when VO_2max was estimated by AHcomplete the relation was significant in women ATH and CON, and in men CON (R^2=0.17-0.24; $p<0.05$). AHraw significantly underestimated VO_2max in the total sample by 8% (51.4±10.2 vs. 55.9±7.6 ml·kg^{-1}·min^{-1}; $p<0.0001$), whereas no significant difference between AHcomplete and the criterion method was found (57.0±11.1 vs. 55.9±7.6 ml·kg^{-1}·min^{-1}; $p=0.26$).

The Actiheart step test is an acceptable tool for the estimation of VO_{2max} if an error within 8% can be tolerated. However, accuracy of the VO_2max prediction is much improved when entering measured variables, such as RMR, SHR, and HRmax, into the software.

Keywords:
Maximum oxygen consumption – athletes – cardiorespiratory fitness – exercise testing – metabolic equivalent

Introduction

Maximum oxygen consumption (VO_2max), a key indicator of cardiorespiratory fitness, is used in both athletic and health settings, as a determinant of physical performance (Bassett & Howley, 2000) or as a predictor of health risk and longevity (Kodama et al., 2009). VO_2max-testing is usually performed on a treadmill or a cycle ergometer while oxygen uptake and expired carbon-dioxide are measured with a respiratory gas analyzer. The workload is progressively increased until the participant reaches volitional exhaustion (after a minimum of 5 min (treadmill) or 7 min (cycle), up to 26 min (Midgley, Bentley, Luttikholt, McNaughton, & Millet, 2008)). However, maximum exercise tests are time consuming, expensive and depend on physiological expertise (Björkman, Ekblom-Bak, Ekblom, & Ekblom, 2016). In addition, some individuals from the general population cannot achieve the maximal effort required for the determination of VO_2max. Furthermore, there are contraindications for maximum exercise tests (American Thoracic Society & American College of Chest Physicians, 2003). These contraindications include uncontrolled asthma, syncope, acute myocardial infarction, and respiratory failure. Therefore, several indirect methods to estimate VO_2max have been developed, where VO_2max is predicted from submaximal exercise results. Most of these tests use the linear relationship between heart rate (HR) and oxygen uptake (VO_2). The VO_2max is then predicted by extrapolation of the submaximal values to an estimated maximum heart rate (HRmax). Submaximal tests thus allow evaluation of cardiorespiratory fitness in a population, in which the direct determination of VO_2max is not possible (Björkman et al., 2016).

The Actiheart is a lightweight (10 g), waterproof combined HR and movement sensor (accelerometer) designed to non-invasively assess daily physical activity levels (Cambridge Neurotechnology Ltd., Papworth, UK). The Actiheart was shown to give accurate estimations of activity energy expenditure against indirect calorimetry during a wide range of activities in men and women in both laboratory (Thompson, Batterham, Bock, Robson, & Stokes, 2006) and field (Crouter, Churilla, & Bassett, 2008) settings. A good level of agreement between Actiheart and doubly labelled water measured daily total energy expenditure was found in adult men and women (Brage et al., 2015), in children and adolescents (Butte et al., 2010), and in lean and overweight men of various fitness levels (Villars et al., 2012). The Actiheart needs to be individually calibrated for each person with a standard step test, a built-in function of the Actiheart software. During stepping, the relationships between actimetry and HR vs. work rate are assessed. The step test also yields estimated VO_2max values, obtained by extrapolating the HR vs. work rate regression line to the estimated or measured HRmax for the individual. No studies assessed the validity of the VO_2max estimation using the Actiheart step test. The purpose of the study was to estimate VO_2max using the Actiheart step test in AHraw and AHcomplete modes (see "Methods") and to compare the results with measured VO_2max over a range of aerobic capacities. The results were expected to fill a

gap in the knowledge regarding the practical use and precision of the Actiheart modes for estimating VO_2max in comparison to the VO_2max ergometer measurements for athletic and non-athletic populations.

Methods

Participants

We recruited by advertisement 68 competitive endurance athletes (ATH; 31 women, 37 men; regular endurance training volume $\geq$ 300 min $\cdot$ wk^{-1} and participation in competitions) and 63 healthy, non-endurance-trained nonsmoking controls (CON; 34 women, 29 men; max. 150 min $\cdot$ wk^{-1} moderate endurance training). Most but not all participants were workers, students or athletes located on the campus of the Swiss Federal Institute of Sports Magglingen. All participants had to be weight-stable (< 2 kg of weight difference in the last 3 months) with a Body Mass Index (BMI) between 18.5 and 25 kg $\cdot$ m^{-2}, and aged between 18 and 40 years. Participants were excluded if they were smoking, pregnant, lactating, dieting, suffering from metabolic disease and/or eating disorders, or taking medication (apart from contraceptives). Athletes were excluded from the study if they had changed their training habits within the last four weeks before the experiments (e.g., due to injury or disease).

All experimental procedures were approved by the Regional Ethics Committee in Berne, Switzerland (KEK-number 090/15), and the study was carried out according to the recommendations of the Helsinki Declaration. Written informed consent of the participants was obtained before any testing.

Study design

The participants were recruited by advertisement. They came to the testing center on two separate testing days, having refrained from strenuous physical activity $\geq$ 24 h. After providing written informed consent, on the first day the participants completed, in a fasted state ($\geq$ 12 h absence of any food or fluid intake, $\geq$ 36 h absence of alcohol or caffeine intake) measurements in the following order: 1) anthropometry and body composition; 2) resting metabolic rate (RMR); and 3) a step test using the in-built function of the Actiheart. One week after the first measurement day, the participants did an incremental exercise test (VO_2max) in a non-fasted state. In the days between the two testing days the participants wore an Actiheart for at least 7 days. All tests were carried out in Magglingen (Switzerland) at an altitude of 950 m.

Estimation of VO_2max with the Actiheart

The Actiheart was clipped onto two standard ECG electrodes (3M^{TM} Red Dot^{TM} Electrode 2560; 3M Health Care, St. Paul, USA) on the chest of the participant according to manufacturer's instructions and worn day and night (Brage et al., 2006). The de-

vice was individually calibrated for each participant using the standard step test, a built-in function of the Actiheart software version 4.0.92 (Cambridge Neurotechnology Ltd., Papworth, UK). In short, the participants stepped up and down a 195 mm high step for eight minutes while stepping rate was ramped linearly from 15 step cycles $\cdot$ min^{-1} (one step cycle is "up, up, down, down") to 33 step cycles $\cdot$ min^{-1} at the end (rate of change 2.5 body lifts $\cdot$ min^{-1}). Participants were advised to stop the test if they felt uncomfortable. After the end of the test, participants were requested to stand still for two minutes with the Actiheart still fitted while not speaking during which recovery HR was assessed. The mechanical power of the step test (mass-specific lift work rate) was calculated as 9.81 m $\cdot$ s^{-2} x step height (m) x lift frequency (number of body weight lifts $\cdot$ min^{-1}) and expressed in J $\cdot$ kg$^{-1}\cdot$ min^{-1}. The Actiheart software uses linear regression to model the relationship between work rate and HR during stepping. To estimate the VO$_2$max, the regression line was extrapolated to the estimated HRmax for each participant. The predicted VO$_2$max in L $\cdot$ kg$^{-1}\cdot$ min^{-1} was then calculated as a maximum mechanical power in J $\cdot$ kg$^{-1}\cdot$ min^{-1} divided by a factor of 20.2 (assuming an invariant net efficiency of 20%). The HRmax was defined in two ways: (1) estimated by use of the Tanaka equation (Tanaka, Monahan, & Seals, 2001); and (2) measured during a maximum exercise test (see section "Measurement of VO$_2$max").

Total energy expenditure (TEE) and activity energy expenditure were estimated by analyzing 6 full-day (24 h) recordings of HR and body movement with a 15-second averaging epoch setting. During the recording period participants were requested to continue their habitual life routine and physical activities. Sleeping heart rate (SHR) was measured as a mean of the 120 lowest HR recordings during 24 h, the SHR for a full six-day period was determined as the mean value of the six SHR recordings. TEE was calculated as sum of RMR (measured by indirect calorimetry, see section "Resting metabolic rate"), activity energy expenditure, and diet-induced thermogenesis (estimated as 10% of TEE (Manore & Thompson, 2010)). Physical activity level (PAL) was calculated as TEE/RMR.

For the validity testing, two inbuilt scenarios of the Actiheart step test were looked at: (1) "AHraw". In this mode SHR is set to 70 bpm, RMR is estimated by the Schofield equation (Schofield, 1985), and HRmax is estimated by use of the Tanaka formula (Tanaka et al., 2001); and (2) "AHcomplete". In this mode SHR, measured during the long-term recordings of the Actiheart, RMR, measured using indirect calorimetry, and HRmax, measured during a maximum exercise test, are manually entered into the Actiheart software.

Measurement of VO$_2$max

Before exercise testing each participant filled out the German (Marti, Villiger, Hintermann, & Lerch, 1998) or French (Société canadienne de physiologie de l'exercice, 2002) version of the Physical Activity Readiness Questionnaire (PAR-Q). Only if all items were answered with "no", participants were allowed to start the exercise testing. The test was performed on a treadmill (women: model mercury; men: model venus, h/p/Cosmos Sports & Medical GmbH, Traunstein, Germany). Treadmill inclines were set at 4° throughout the test. After a 5-min warm-up jog, non-athletic participants begun running at 7 km $\cdot$ h^{-1}, whereas participants from the athlete group started at 9 km $\cdot$ h^{-1}. The speed was increased by 1 km $\cdot$ h^{-1} every minute for the first 3 minutes of the test and thereafter by 0.5 km $\cdot$ h^{-1} every 30 s until exhaustion (Steiner & Wehrlin, 2011). Gas exchange was measured breath-by-breath with an open-circuit system (Quark CPET; COSMED Srl, Rome, Italy), which was calibrated before each test according to manufacturer's instructions. VO$_2$ data was processed using 10-second time averages. VO$_2$max was determined as the highest 30-second VO$_2$ average for the test (Robergs, Dwyer, & Astorino, 2010). Heart rate was continuously registered with a wireless HR monitoring system (model SZ990; COSMED Srl, Rome, Italy). If the primary criteria of a plateau in oxygen uptake (defined as an increase of VO$_2$ < 2.1 mL $\cdot$ kg$^{-1}\cdot$ min^{-1}) was not reached by the participant ($n = 4$), then the secondary criteria of a RQ value $\geq$ 1.10, and a HR close ($\pm$ 10 bpm) to the age-predicted HRmax (Tanaka et al., 2001) were used to determine whether the participant reached maximal effort and thus VO$_2$max. VO$_2$max-tests were carried out at a mean room temperature of 21.8 $\pm$ 1.0°C, a humidity of 39.8 $\pm$ 10.1%, and an air pressure of 914 $\pm$ 7 hpa. In general, a temperature range of 20 to 22°C in a cool, dry environment (< 50% humidity) is considered comfortable for exercise testing (Myers et al., 2009).

Anthropometric data and body composition

Body weight was measured to the nearest 0.1 kg on a calibrated beam scale (Seca 877, Seca, Hamburg, Germany) and body height was measured to the nearest 0.5 cm with a height rod (Seca 213, Seca, Hamburg, Germany), with the participants in light clothing and without shoes.

Body composition was assessed using Lunar iDXA (GE Healthcare, Madison, WI, USA). Calibration of the iDXA was performed and checked on a daily basis before testing using a calibration phantom. Participants voided their bladder before the scan. The participants were in underwear and all metal artefacts were removed. During measurement, participants were in supine position on the scanning table with their ankles and legs fixed using supports. Arms were positioned to the side with the palms flat on the table. Participants were requested to stay still during the measurement. Whole body scans were performed according to the manufacturer's instructions, and adipose tissue mass, lean tissue mass, and bone mineral content were derived (enCore software v. 11.10; GE Healthcare, Madison, WI, USA). Total body composition estimates with the Lunar iDXA have been reported to be excellent in other studies (Carver, Christou, & Andersen, 2013; Hind, Oldroyd, & Truscott, 2011).

Resting metabolic rate

RMR was assessed by indirect calorimetry using a ventilated hood system (Quark CPET; COSMED Srl, Rome, Italy). Calibrations of the gas analyzer and flowmeter were carried out before each test according to manufacturer's instructions. After acclimatization and relaxing for 30 min on a bed, a ventilated hood was placed over the participant's head and measurements were started. VO_2 and carbon dioxide production (VCO_2) were measured for 30 min at 10-second intervals with participants remaining motionless in supine position in a thermoneutral environment (20 - 25 °C (Compher, Frankenfield, Keim, Roth-Yousey, & Evidence Analysis Working Group, 2006)). The first 5 min were eliminated as acclimatization artifact. From the remaining 25 min an interval of 5 consecutive minutes with a < 10% coefficient of variation of VO_2 and VCO_2 was considered as steady state. VO_2 and VCO_2 were then used to calculate RMR using the abbreviated Weir equation (Weir, 1949). Since respiratory quotient (RQ, defined as the ratio of VCO_2 and VO_2) measures at rest < 0.70 and > 1.00 suggest protocol violations or inaccurate gas measurement (Compher et al., 2006), participants with RQ-values outside of the plausible range should be excluded from data analysis ($n = 1$). The obtained RQ for all measurements was 0.76 ± 0.04 and the mean coefficients of variation of VO_2 and VCO_2 were $3.8 \pm 1.5\%$ and $5.0 \pm 1.9\%$. RMR measurements took place at a mean temperature of $21.9 \pm 1.2°C$, $40.0 \pm 10.2\%$ humidity, and an air pressure of 914 ± 7 hpa.

Statistics

Statistical analyses were performed using the statistical software SPSS version 24 for MS-Windows (IBM Corp., Chicago, IL, USA). Mean values and standard deviations (SD) were calculated and data was checked for normality using the Shapiro-Wilk-test. All parameters were normal distributed with the exception of age, body fat (%), fat-free mass (kg), RMR (kcal · d^{-1}), VO_2max (L · min^{-1}), recovery HR, and PAL. Group differences were tested by independent t-tests and Mann-Whitney-U-tests ($\alpha = 0.05$). The validity of the Actiheart step test was first investigated using Pearson's Product moment correlation analysis. The correlation coefficients (r) were classified according to Cohen (Cohen, 1988). An r between 0.10-0.29 was considered a small, between 0.30-0.49 a moderate and between 0.50-1.0 a strong association. The data were further analyzed using a repeated-measures ANOVA with a Bonferroni post-hoc correction to compare measured VO_2max, estimated VO_2max by AHraw, and estimated VO_2max by AHcomplete. In addition, the mean absolute error (MAE) and the mean absolute percentage error (MAPE) of AHraw and AHcomplete compared to the criterion measure were computed. Since no standardized threshold exists for high or low MAPE, we considered a MAPE $\geq 10\%$ as an indicator of inaccuracy as suggested by other authors (Boudreaux et al., 2018; Lee, Kim, & Welk, 2014; Nelson, Kaminsky, Dickin, & Montoye, 2016; Roos, Taube, Beeler, & Wyss,

2017). The R^2 and standard error of the estimate (SEE) were calculated by linear regression, where measured VO_2max was entered as dependent variable and estimated VO_2max as independent variable. Levels of agreement between Actiheart and measured values were further expressed as mean difference with Limits of Agreement (mean difference $\pm$ 1.96 SD) (Bland & Altman, 1999).

Results

Participants

Data of in total 50 women (24 ATH and 26 CON) and 41 men (22 ATH and 19 CON) with valid Actiheart, VO_2max, and RMR measurements were included into the analysis. The Table 1 presents the anthropometric data, body composition, RMR and VO_2max of the participants. The VO_2max ranged from 39.1 – 65.4 ml · kg^{-1} · min^{-1} in women and 42.8 – 78.4 ml · kg^{-1} · min^{-1} in men.

Validity of the Actiheart step test for estimation of VO_2max

The Pearson's correlation analyses revealed moderately correlated data of AHraw and the criterion method only in women CON ($r = 0.46$, $p < 0.05$), whereas the correlation between AHcomplete and measured VO_2max was significant in women ATH, women CON, and men CON ($r = 0.41 – 0.49$, all $p < 0.05$; Table 2), and r was interpreted as moderate association. The MAE, MAPE, and SEE for all groups are presented in Table 2.
In women and men ATH the AHraw significantly underestimated VO_2max ($p < 0.05$), whereas in the CON groups no significant differences between estimated and measured VO_2max were found. For AHcomplete, VO_2max was significantly underestimated in women ATH ($p = 0.03$), and overestimated in men CON ($p = 0.03$), whereas no significant differences were obtained in women CON and men ATH. When looking at the total sample, significant differences between AHraw and the measured VO_2max (51.4 ± 10.2 vs. 55.9 ± 7.6 ml · kg^{-1} · min^{-1}; $p < 0.0001$) were observed, whereas the difference between AHcomplete and VO_2max was non-significant (57.0 ± 11.1 vs. 55.9 ± 7.6 ml · kg^{-1} · min^{-1}; $p = 0.26$). The range of the MAPE across all groups was 11.4 – 17.7% and 10.8 – 14.7% for AHraw and AHcomplete, respectively (Table 2). The absolute difference between measured and estimated VO_2max by use of AHraw was smaller than 5 ml · kg^{-1} · min^{-1} in 29% ($n = 7$), 54% ($n = 14$), 27% ($n = 6$), and 58% ($n = 11$) in women ATH, women CON, men ATH, and men CON, respectively. When VO_2max was estimated by AHcomplete, the absolute difference between measured and estimated VO_2max was smaller than 5 ml · kg^{-1} · min^{-1} in 42% ($n = 10$), 39% ($n = 10$), 50% ($n = 11$), and 37% ($n = 7$) of women ATH, women CON, men ATH, and men CON, respectively.
The results of the linear regression analyses are presented in Table 3. In women CON a statistically significant relationship between VO_2max estimated by AHraw and the measured val-

Table 1: Participant characteristics of the men and women athletes (ATH) and controls (CON).

	Women		Men	
	ATH (*n* = 24)	CON (*n* = 26)	ATH (*n* = 22)	CON (*n* = 19)
Age (years)	30.0 ± 5.5	27.4 ± 4.9	25.8 ± 4.4	27.1 ± 5.4
Body mass (kg)	57.0 ± 5.9 [a]	61.6 ± 6.6	69.9 ± 6.8	72.4 ± 7.3
Height (cm)	168 ± 5	168 ± 7	180 ± 4	179 ± 7
BMI ($kg \cdot m^{-2}$)	20.3 ± 1.6 [b]	21.8 ± 1.6	21.6 ± 1.6	22.5 ± 1.7
Body fat (%)	23.3 ± 4.8 [a]	26.0 ± 4.8	14.1 ± 4.1 [a]	17.0 ± 5.0
Fat-free mass (kg)	44.4 ± 4.3	45.7 ± 5.2	60.9 ± 6.6	60.8 ± 6.7
RMR				
$kcal \cdot d^{-1}$	1436 ± 121	1500 ± 153	1868 ± 162	1835 ± 257
$kcal \cdot kg^{-1} \cdot d^{-1}$	25.4 ± 2.5	24.4 ± 1.4	26.8 ± 1.6 [a]	25.3 ± 2.5
$kcal \cdot kg^{-1} \cdot h^{-1}$	1.06 ± 0.10	1.02 ± 0.06	1.12 ± 0.07 [a]	1.06 ± 0.10
VO_2max				
$L \cdot min^{-1}$	3.2 ± 0.3 [a]	3.0 ± 0.4	4.5 ± 0.4 [a]	4.1 ± 0.6
$ml \cdot kg^{-1} \cdot min^{-1}$	56.1 ± 4.5 [c]	48.4 ± 4.4	64.0 ± 6.2 [c]	56.6 ± 5.0
Maximum metabolic equivalent of task	15.2 ± 1.4 [c]	13.7 ± 1.2	16.3 ± 1.2 [a]	15.3 ± 1.4
HRmax (bpm)	181 ± 9 [a]	186 ± 7	189 ± 8	189 ± 7
SHR (bpm)	49 ± 7 [a]	53 ± 6	48 ± 6	50 ± 7
Recovery HR (bpm)	87 ± 14	90 ± 17	85 ± 11	85 ± 12
PAL	2.1 ± 0.3 [b]	1.9 ± 0.2	1.9 ± 0.3	1.7 ± 0.2

Legend: Results are expressed as mean ± standard deviation. [a] significantly different from CON of the same sex ($p < 0.05$), [b] significantly different from CON of the same sex ($p < 0.01$), [c] significantly different from CON of the same sex ($p < 0.0001$)

Table 2: Concurrent validity (tested mode vs. criterion measure) of the two Actiheart modes to estimate VO_2max in men and women athletes (ATH) and controls (CON).

	Women		Men	
	ATH (*n*=24)	CON (*n*=26)	ATH (*n*=22)	CON (*n*=19)
r value				
AHraw	0.225	0.463 [a]	-0.069	0.431
AHcomplete	0.410 [a]	0.488 [a]	0.235	0.480 [a]
MAE [$ml \cdot kg^{-1} \cdot min^{-1}$]				
AHraw	8.8 ± 9.2	2.5 ± 7.2	5.2 ± 10.0	1.0 ± 9.1
AHcomplete	3.0 ± 6.5	-1.2 ± 8.0	-1.5 ± 10.0	-5.4 ± 9.8
MAPE [%]				
AHraw	17.7 ± 13.7	11.4 ± 10.4	13.6 ± 8.6	11.4 ± 10.6
AHcomplete	10.8 ± 6.5	13.6 ± 8.9	11.1 ± 10.5	14.7 ± 12.2
SEE [$ml \cdot kg^{-1} \cdot min^{-1}$]				
AHraw	4.47	3.98	6.31	4.66
AHcomplete	4.19	3.91	6.15	4.53

Legend: Results are expressed as mean ± standard deviation. [a] correlations significant at $p < 0.05$

Table 3: Parameters of linear regression for maximum oxygen consumption (VO$_2$max; ml · kg^{-1} · min^{-1}) in men and women athletes (ATH) and controls (CON). Measured VO$_2$max was entered as dependent variable and estimated VO$_2$max by Actiheart was the independent variable.

	Women		Men	
	ATH (*n*=24)	CON (*n*=26)	ATH (*n*=22)	CON (*n*=19)
R^2				
AHraw	0.051	0.215	0.005	0.186
AHcomplete	0.168	0.238	0.055	0.230
p-value				
AHraw	0.291	**0.017**	0.760	0.065
AHcomplete	**0.047**	**0.011**	0.293	**0.038**
Slope				
AHraw	0.11	0.251	-0.057	0.216
AHcomplete	0.266	0.234	0.153	0.217
Intercept				
AHraw	50.914	36.881	67.394	44.632
AHcomplete	42.004	36.782	54.011	43.183

Legend: Significant *p*-values are highlighted in bold.

ues was detected ($p < 0.05$). The relationship between VO$_2$max estimated by AHcomplete and the criterion method was significant in women ATH and CON and in men CON ($p < 0.05$).
In Figure 1, the data of AHraw and AHcomplete and the reference method are presented using Bland-Altman plots. For the total sample, the mean bias (± 1.96 SD) was 4.5 (-13.6; 22.6) ml · kg^{-1} · min^{-1} for AHraw, whereas for the AHcomplete no systematic errors were observed. Likewise, no systematic errors were found in the single groups for both AHraw and AHcomplete, except for women ATH when VO$_2$max was estimated by AHraw ($p < 0.0001$; Figure 1).

Discussion

We evaluated the validity of the Actiheart step test for the estimation of VO$_2$max in men and women along a range of aerobic fitness levels, comparing two different data entry modes, AHraw (estimated SHR, HRmax and RMR) and AHcomplete (measured SHR, HRmax and RMR). Based on MAPE, the Actiheart step test is not an acceptable method for estimation of VO$_2$max.
There exist several submaximal step tests for the estimation of VO$_2$max. The most common are the Chester step test, the STEP tool protocol, the modified YMCA 3-minute step test, and the Åstrand-Rhyming step test (Bennett, Parfitt, Davison, & Eston, 2016). The *R^2* of previous reported step tests for the estimation of VO$_2$max ranged between 0.22 and 0.88 (Francis & Brasher, 1992; Francis & Culpepper, 1989; Hansen, Jacobs, Thijs, Dendale, & Claes, 2016; Knight, Stuckey, & Petrella, 2014; McArdle, Katch, Pechar, Jacobson, & Ruck, 1972; Perroni, Cortis, Minganti, Cignitti, & Capranica, 2013; Petrella, Koval, Cunningham, &

Paterson, 2001; Santo & Golding, 2003; Sykes & Roberts, 2004; Webb, Vehrs, George, & Hager, 2014). In our study we found a significant relationship between estimated and measured VO$_2$max for the total sample, with *R^2* 0.24 for AHraw and 0.36 for AHcomplete. However, when dividing the groups by sex and experimental group (ATH and CON), the *R^2* ranged from 0.01 to 0.22 for AHraw (with being significant only for women CON), and 0.06 to 0.24 for AHcomplete (significant for women ATH and CON, and men CON).
Compared to other studies the absolute measures of agreement showed similar or slightly better validity of the Actiheart step test for the estimation of VO$_2$max (SEE: 3.98 – 6.31 ml · kg^{-1} · min^{-1} for AHraw and 3.91 – 6.15 ml · kg^{-1} · min^{-1} for AHcomplete). Previous studies reported a SEE of 6.9 – 8.76 ml · kg^{-1} · min^{-1} for the modified YMCA 3-minute step test in healthy men and women (Santo & Golding, 2003) and 3.9 ml · kg^{-1} · min^{-1} for the Chester step test in healthy adults (Sykes & Roberts, 2004). In well-trained men a SEE of 0.28 L · min^{-1} was detected for the Åstrand-Rhyming step test (Åstrand & Ryhming, 1954). Assuming a body mass of 70 kg (not reported in that study) this would amount to a SEE of 3.6 ml · kg^{-1} · min^{-1}.
In the present study, we found a systematic bias of 4.5 ml · kg^{-1} · min^{-1} for the total sample for AHraw, indicating that estimated VO$_2$max was significantly lower compared to the measured VO$_2$max values. When VO$_2$max was estimated by AHcomplete, no systematic errors were observed (-1.1 ml · kg^{-1} · min^{-1}, non-significant). Knight et al. (2014) observed a systematic bias of -6.4 ml · kg^{-1} · min^{-1} for the STEP tool protocol in 40 healthy men and women, with higher VO$_2$max values in the predictive vs. maximal test. For the Chester step test a mean systematic bias of 2.8 ml · kg^{-1} · min^{-1} was reported in university students,
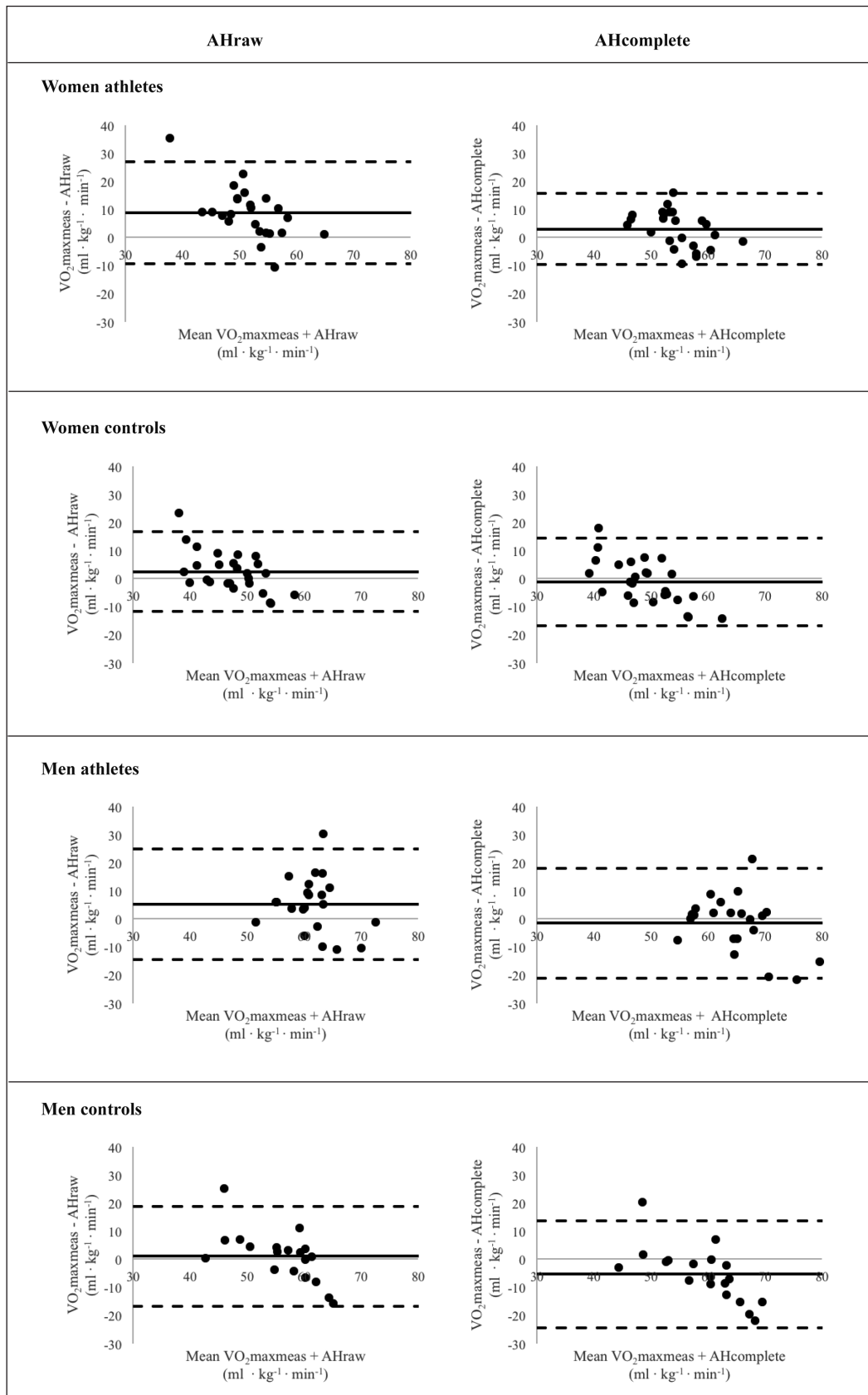
Figure 1: Bland-Altman plots of AHraw and AHcomplete and the reference method (VO$_2$maxmeas) for men and women athletes and controls.

indicating that predicted VO_2max was significantly lower compared to the measured VO_2max during a maximal treadmill test (Buckley, Sim, Eston, Hession, & Fox, 2004). We also found a significant negative relationship between the mean of measured and estimated VO_2max and the difference between both methods, indicating that with increasing VO_2max the difference between actual and estimated VO_2max increased (r = -0.33 for AHraw and r = -0.45 for AHcomplete, $p < 0.01$). This means that for participants with very high (such as endurance athletes) or very low aerobic capacity (untrained) the difference between estimated VO_2max by the Actiheart step test and measured VO_2max is larger. In participants with high aerobic capacity the Actiheart step test thus *overestimates* actual VO_2max, whereas in participants with low aerobic capacity VO_2max is *underestimated*, independent of which entry mode is used.

In our study, the absolute difference between measured and estimated VO_2max was smaller than 5 ml $\cdot$ kg^{-1} $\cdot$ min^{-1} in 41.8% of the participants for both AHraw and AHcomplete. However, when VO_2max was estimated by use of AHraw, only 28.3% (n = 13) of the athletes had a smaller than 5 ml $\cdot$ kg^{-1} $\cdot$ min^{-1} absolute difference between measured and estimated VO_2max, whereas when VO_2max was estimated by AHcomplete 45.7% (n = 21) of the participants were within 5 ml $\cdot$ kg^{-1} $\cdot$ min^{-1}. These results indicate that in endurance trained participants the AHcomplete for estimating VO_2max is more accurate than the AHraw, whereas in persons with a good VO_2max the AHraw setting is sufficient to predict aerobic capacity.

We assume that the underlying equations of the manufacturer were originally developed for the estimation of VO_2max of the general population with a limited range of aerobic capacity and not especially for endurance athletes nor for persons with low aerobic capacity. Based on our results it would thus seem necessary to adapt these equations in order to provide more valid VO_2max values for a broader range of aerobic capacities.

Strengths and limitations

A strength of the study is the number of included men and women covering the upper range of aerobic capacities. In addition, body composition, RMR, and long-term Actiheart data (such as SHR and PAL) were available for all participants and could be entered into the Actiheart software in order to evaluate the effect on the estimation of VO_2max. We also put effort into measuring "real" VO_2max (and not VO_2peak) by applying current criteria for maximal exercise testing (Scharhag-Rosenberger 2010, Midgley et al. 2008), including the recommendation that the duration of VO_2max tests should last between 5 and 26 minutes (Midgley et al., 2008). The specific maximal exercise test in our study was performed on a motorized treadmill and the chosen protocol for the determination of VO_2max was previously shown to induce exhaustion in athletic and non-athletic persons after 5 - 9 min (Steiner & Wehrlin, 2011).

A common issue in studies involving maximal exercise testing is the selection bias, often including fitter individuals than the general population. In a recent study examining the aerobic capacity in the Swiss working population VO_2max values of 33 and 45 ml $\cdot$ kg^{-1} $\cdot$ min^{-1} for women and men were reported (Mundwiler et al., 2017). In our control group, the mean values for VO_2max were 48.4 and 56.6 ml $\cdot$ kg^{-1} $\cdot$ min^{-1} for women and men, indicating a higher fitness level in the included participants compared to the general population.

Step tests, in general, are prone to error. For example, performing a step test requires the person's ability to maintain a certain stepping tempo and technique. Alterations in stepping technique can affect mechanical efficiency and therefore physiological responses in HR and oxygen consumption (Bennett et al., 2016). In addition, performing a step test with a pre-defined fixed step height, might introduce potential error since leg length, step length, body mass, and morphology vary between individuals and these differences might result in different individual physiological responses.

The mode AHcomplete requires the measurement of RMR, a long-term assessment of SHR, and a maximum exercise test for determination of HRmax. These measurements have higher financial and time costs, and require the same or even more expertise compared to the direct assessment of VO_2max. However, HRmax can be also measured using a maximum exercise test without performing a spiroergometry, which makes its application possible in settings where the equipment for a respiratory gas analysis is not available. Another limitation of the AHcomplete mode is that performing a maximum exercise test (with or without respiratory gas analysis) is not possible in every population, which limits the application of this mode. In summary, the mode AHcomplete is only applicable in settings without respiratory gaz analysis equipment and in subjects who can perform a maximum exercise test for determination of HRmax. We did not verify the validity of an Actiheart mode where RMR and SHR are entered as measured variables, and HRmax is estimated. However, using an estimated HRmax as endpoint for the estimation of VO_2max, such as for AHraw, might introduce error since it was shown that estimations of HRmax can have a considerable prediction error (Tanaka et al., 2001). Another potential source of error is the reliance on the linear relationship between VO_2 and power output, as shown by Åstrand and Rodahl (1970). However, Zoladz, Rademaker and Sargeant (1995) found a non-linear relationship between VO_2 and power output, which may affect the prediction of VO_2max by use of submaximal tests. Although the non-linearity might occur predominantly at high intensities above the anaerobic threshold (Majerczak et al., 2012; Zoladz, Szkutnik, Majerczak, & Duda, 1998) this underlines the importance of choosing an individual work rate which is not too high. In addition, it was shown that there is not always a linear relationship between HR and VO_2 (Buckley et al., 2004).

Conclusion

Based on MAPE, in young adults with good to superior aerobic capacity the Actiheart step test was not acceptable for estimation of VO_2max. The mode AHraw significantly underestimated VO_2max in men and women endurance trained athletes, and AHcomplete significantly underestimated VO_2max in women athletes. In endurance trained participants one should manually enter RMR, HRmax, and SHR into the Actiheart software in order to increase the accuracy of the VO_2max prediction. Areas of future investigation include the repetition of analyses of this study when performing the Actiheart step test with other step heights, evaluation of the reliability of the step test, inclusion of older and more sedentary persons and use in a clinical setting, and assessing factors such as HR and VO_2 during the Actiheart step test to improve the estimation of VO_2max.

Acknowledgements

Funding

Competing Interests

The authors have declared that no competing interests exist.

Data Availability Statement

All relevant data are included in the paper. Further data can be made available upon request.

Abbreviations

ATH = athletes, BMI = body mass index, CON = controls, HR = heart rate, HR_{max} = maximum heart rate, MAE = mean absolute error, MAPE = mean absolute percentage error, PAL = physical activity level, RMR = resting metabolic rate, RQ = respiratory quotient, SEE = standard error of the estimate, SHR = sleeping heart rate, TEE = total energy expenditure, VCO_2 = expired carbon dioxide, VO_2 = oxygen consumption, VO_2max = maximum oxygen consumption

References

American Thoracic Society, & American College of Chest Physicians. (2003). ATS/ACCP Statement on cardiopulmonary exercise testing. *American Journal of Respiratory and Critical Care Medicine*, *167* (2), 211–277. Doi: 10.1164/rccm.167.2.211

Åstrand, P. O., & Rodahl, K. (1970). *Textbook of work physiology*. New York: McGraw-Hill.

Åstrand, P. O., & Ryhming, I. (1954). A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during sub-maximal work. *Journal of Applied Physiology*, *7* (2), 218–221. Doi: 10.1152/jappl.1954.7.2.218

Bassett, D. R., & Howley, E. T. (2000). Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine and Science in Sports and Exercise*, *32* (1), 70–84.

Bennett, H., Parfitt, G., Davison, K., & Eston, R. (2016). Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. *Sports Medicine*, *46* (5), 737–750. Doi: 10.1007/s40279-015-0445-1

Björkman, F., Ekblom-Bak, E., Ekblom, Ö., & Ekblom, B. (2016). Validity of the revised Ekblom Bak cycle ergometer test in adults. *European Journal of Applied Physiology*, *116* (9), 1627–1638. Doi: 10.1007/s00421-016-3412-0

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8* (2), 135–160. Doi: 10.1177/096228029900800204

Boudreaux, B. D., Hebert, E. P., Hollander, D. B., Williams, B. M., Cormier, C. L., Naquin, M. R., … Kraemer, R. R. (2018). Validity of wearable activity monitors during cycling and resistance exercise. *Medicine and Science in Sports and Exercise*, *50* (3), 624–633. Doi: 10.1249/MSS.0000000000001471

Brage, S., Brage, N., Ekelund, U., Luan, J., Franks, P. W., Froberg, K., & Wareham, N. J. (2006). Effect of combined movement and heart rate monitor placement on physical activity estimates during treadmill locomotion and free-living. *European Journal of Applied Physiology*, *96* (5), 517–524. Doi: 10.1007/s00421-005-0112-6

Brage, S., Westgate, K., Franks, P. W., Stegle, O., Wright, A., Ekelund, U., & Wareham, N. J. (2015). Estimation of free-living energy expenditure by heart rate and movement sensing: a doubly-labelled water study. *PloS One*, *10* (9), e0137206. Doi: 10.1371/journal.pone.0137206

Buckley, J. P., Sim, J., Eston, R. G., Hession, R., & Fox, R. (2004). Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise. *British Journal of Sports Medicine*, *38* (2), 197–205.

Butte, N. F., Wong, W. W., Adolph, A. L., Puyau, M. R., Vohra, F. A., & Zakeri, I. F. (2010). Validation of cross-sectional time series and multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents using doubly labeled water. *The Journal of Nutrition*, *140* (8), 1516–1523. Doi: 10.3945/jn.109.120162

Carver, T. E., Christou, N. V., & Andersen, R. E. (2013). In vivo precision of the GE iDXA for the assessment of total body composition and fat distribution in severely obese patients. *Obesity (Silver Spring, Md.)*, *21* (7), 1367–1369. Doi: 10.1002/oby.20323

Cohen, J. (1988). *Statistical Power Analysis for Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Compher, C., Frankenfield, D., Keim, N., Roth-Yousey, L., & Evidence Analysis Working Group. (2006). Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. *Journal of the American Dietetic Association*, *106* (6), 881–903. Doi: 10.1016/j.jada.2006.02.009

Crouter, S. E., Churilla, J. R., & Bassett, D. R. (2008). Accuracy of the Actiheart for the assessment of energy expenditure in adults. *European Journal of Clinical Nutrition*, *62* (6), 704–711. Doi: 10.1038/sj.ejcn.1602766

Francis, K., & Brasher, J. (1992). A height-adjusted step test for predicting maximal oxygen consumption in males. *The Journal of Sports Medicine and Physical Fitness*, *32* (3), 282–287.

Francis, K., & Culpepper, M. (1989). Height-adjusted, rate-specific, single-stage step test for predicting maximal oxygen consumption. *Southern Medical Journal*, *82* (5), 602–606.

Hansen, D., Jacobs, N., Thijs, H., Dendale, P., & Claes, N. (2016). Validation of a single-stage fixed-rate step test for the prediction of maximal oxygen uptake in healthy adults. *Clinical Physiology and Functional Imaging*, *36* (5), 401–406. Doi: 10.1111/cpf.12243

Hind, K., Oldroyd, B., & Truscott, J. G. (2011). In vivo precision of the GE Lunar iDXA densitometer for the measurement of total body composition and fat distribution in adults. *European Journal of Clinical Nutrition*, *65* (1), 140–142. Doi: 10.1038/ejcn.2010.190

Knight, E., Stuckey, M. I., & Petrella, R. J. (2014). Validation of the step test and exercise prescription tool for adults. *Canadian Journal of Diabetes*, *38* (3), 164–171. Doi: 10.1016/j.jcjd.2014.03.007

Kodama, S., Saito, K., Tanaka, S., Maki, M., Yachi, Y., Asumi, M., … Sone, H. (2009). Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA*, *301* (19), 2024–2035. Doi: 10.1001/jama.2009.681

Lee, J.-M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, *46* (9), 1840–1848. Doi: 10.1249/MSS.0000000000000287

Majerczak, J., Korostynski, M., Nieckarz, Z., Szkutnik, Z., Duda, K., & Zoladz, J. A. (2012). Endurance training decreases the non-linearity in the oxygen uptake-power output relationship in humans: Endurance training and oxygen uptake-power output relationship. *Experimental Physiology*, *97* (3), 386–399. Doi: 10.1113/expphysiol.2011.062992

Manore, M. M., & Thompson, J. L. (2010). Energy requirements of the athlete: assessment and evidence of energy efficiency. In L. M. Burke & V. Deakin (Eds.), *Clinical Sports Nutrition* (pp. 96–115). North Ryde: McGraw-Hill Australia Pty Ltd.

Marti, B., Villiger, B., Hintermann, M., & Lerch, R. (1998). Plötzlicher Herztod beim Sport: sinnvolle Vorsorgeuntersuchungen und Präventionsmassnahmen. *Schweizerische Zeitschrift für «Sportmedizin und Sporttraumatologie»*, *46* (2), 83–85.

McArdle, W. D., Katch, F. I., Pechar, G. S., Jacobson, L., & Ruck, S. (1972). Reliability and interrelationships between maximal oxygen intake, physical work capacity and step-test scores in college women. *Medicine and Science in Sports*, *4* (4), 182–186.

Midgley, A. W., Bentley, D. J., Luttikholt, H., McNaughton, L. R., & Millet, G. P. (2008). Challenging a dogma of exercise physiology: does an incremental exercise test for valid VO2max determination really need to last between 8 and 12 minutes? *Sports Medicine*, *38* (6), 441–447.

Mundwiler, J., Schüpbach, U., Dieterle, T., Leuppi, J. D., Schmidt-Trucksäss, A., Wolfer, D. P., … Brighenti-Zogg, S. (2017). Association of occupational and leisure-time physical activity with aerobic capacity in a working population. *PloS One*, *12* (1), e0168683. Doi: 10.1371/journal.pone.0168683

Myers, J., Arena, R., Franklin, B., Pina, I., Kraus, W. E., McInnis, K., … Balady, G.J. (2009). Recommendations for clinical exercise laboratories: a scientific statement from the American heart association. *Circulation*, *119* (24), 3144–3161. Doi: 10.1161/CIRCULATIONAHA.109.192520

Nelson, M. B., Kaminsky, L. A., Dickin, D. C., & Montoye, A. H. K. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Medicine and Science in Sports and Exercise*, *48* (8), 1619–1628. Doi: 10.1249/MSS.0000000000000933

Perroni, F., Cortis, C., Minganti, C., Cignitti, L., & Capranica, L. (2013). Maximal oxygen uptake of Italian firefighters: laboratory vs. field evaluations. *Sport Sciences for Health*, *9* (2), 31–35. Doi: 10.1007/s11332-013-0142-0

Petrella, R. J., Koval, J. J., Cunningham, D. A., & Paterson, D. H. (2001). A self-paced step test to predict aerobic fitness in older adults in the primary care clinic. *Journal of the American Geriatrics Society*, *49* (5), 632–638.

Robergs, R. A., Dwyer, D., & Astorino, T. (2010). Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Medicine*, *40* (2), 95–111. Doi: 10.2165/11319670-000000000-00000

Roos, L., Taube, W., Beeler, N., & Wyss, T. (2017). Validity of sports watches when estimating energy expenditure during running. *BMC Sports Science, Medicine & Rehabilitation*, *9*, 22. Doi: 10.1186/s13102-017-0089-6

Santo, A. S., & Golding, L. A. (2003). Predicting maximum oxygen uptake from a modified 3-minute step test. *Research Quarterly for Exercise and Sport*, *74* (1), 110–115. Doi: 10.1080/02701367.2003.10609070

Scharhag-Rosenberger, F. (2010). Spiroergometrie zur Ausdauerleistungsdiagnostik. *Deutsche Zeitschrift für Sportmedizin, 61* (6), 146-147.

Schofield, W. N. (1985). Predicting basal metabolic rate, new standards and review of previous work. *Human Nutrition. Clinical Nutrition*, *39 Suppl 1*, 5–41.

Société canadienne de physiologie de l'exercice. (2002). *Questionnaire sur l'aptitude à l'activité physique (Q-AAP)*. Retrieved 25 August 2016 from http://www.csep.ca/CMFiles/publications/parq/Q-AAP.pdf

Steiner, T., & Wehrlin, J. P. (2011). Does hemoglobin mass increase from age 16 to 21 and 28 in elite endurance athletes? *Medicine and Science in Sports and Exercise*, *43* (9), 1735–1743. Doi: 10.1249/MSS.0b013e3182118760

Sykes, K., & Roberts, A. (2004). The Chester step test—a simple yet effective tool for the prediction of aerobic capacity. *Physiotherapy*, *90* (4), 183–188. Doi: 10.1016/j.physio.2004.03.008

Tanaka, H., Monahan, K. D., & Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, *37* (1), 153–156.

Thompson, D., Batterham, A. M., Bock, S., Robson, C., & Stokes, K. (2006). Assessment of low-to-moderate intensity physical activity thermogenesis in young adults using synchronized heart rate and accelerometry with branched-equation modeling. *The Journal of Nutrition*, *136* (4), 1037–1042. Doi: 10.1093/jn/136.4.1037

Villars, C., Bergouignan, A., Dugas, J., Antoun, E., Schoeller, D. A., Roth, H., … Simon, C. (2012). Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men. *Journal of Applied Physiology*, *113* (11), 1763–1771. Doi: 10.1152/japplphysiol.01413.2011

Webb, C., Vehrs, P. R., George, J. D., & Hager, R. (2014). Estimating VO$_2$max using a personalized step test. *Measurement in Physical Education and Exercise Science*, *18* (3), 184–197. Doi: 10.1080/1091367X.2014.912985

Weir, J. B. D. B. (1949). New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of Physiology*, *109* (1–2), 1–9.

Zoladz, J. A., Rademaker, A. C., & Sargeant, A. J. (1995). Non-linear relationship between O$_2$ uptake and power output at high intensities of exercise in humans. *The Journal of Physiology*, *488 (Pt 1)*, 211–217.

Zoladz, J. A., Szkutnik, Z., Majerczak, J., & Duda, K. (1998). Detection of the change point in oxygen uptake during an incremental exercise test using recursive residuals: relationship to the plasma lactate accumulation and blood acid base balance. *European Journal of Applied Physiology and Occupational Physiology*, *78* (4), 369–377.