



Criterion validity of the Ekblom-Bak and the Åstrand submaximal test in an elderly population

Daniel Väisänen¹ · Örjan Ekblom¹ · Elin Ekblom-Bak¹ · Eva Andersson^{1,2} · Jonna Nilsson^{1,3} · Maria Ekblom^{1,2}

Received: 25 June 2019 / Accepted: 26 November 2019 / Published online: 9 December 2019
© The Author(s) 2019

Abstract

Purpose The aim of this study was to validate the submaximal Ekblom-Bak test (EB-test) and the Åstrand test (Å-test) for an elderly population.

Methods Participants ($n = 104$), aged 65–75 years, completed a submaximal aerobic test on a cycle ergometer followed by an individually adjusted indirect calorimetry VO_2max test on a treadmill. The HR from the submaximal test was used to estimate VO_2max using both the EB-test and Å-test equations.

Results The correlation between measured and estimated VO_2max using the EB method and Å method in women was $r = 0.64$ and $r = 0.58$, respectively and in men $r = 0.44$ and $r = 0.44$, respectively. In women, the mean difference between estimated and measured VO_2max was -0.02 L min^{-1} (95% CI -0.08 to 0.04) for the EB method and -0.12 L min^{-1} (95% CI -0.22 to -0.02) for the Å method. Corresponding values for men were 0.05 L min^{-1} (95% CI -0.04 to 0.14) and -0.28 L min^{-1} (95% CI -0.42 to -0.14), respectively. However, the EB method was found to overestimate VO_2max in men with low fitness and the Å method was found to underestimate VO_2max in both women and men. For women, the coefficient of variance was 11.1%, when using the EB method and 19.8% when using the Å method. Corresponding values for men were 11.6% and 18.9%, respectively.

Conclusion The submaximal EB-test is valid for estimating VO_2max in elderly women, but not in all elderly men. The Å-test is not valid for estimating VO_2max in the elderly.

Keywords Cardiorespiratory fitness · Elderly · Oxygen uptake · Submaximal test · Validity · Public health

Abbreviations

Å-test Åstrand test

CV Correlation of variance

EB-test Ekblom-Bak test

HR Heart rate

LoA Limits of agreement

PO Power output/work rate

RER Respiratory exchange ratio

RPE Rate of perceived exertion

SEE Standard error of the estimate

VO_2max Maximal oxygen uptake

Communicated by Philip D. Chilibeck.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00421-019-04275-7>) contains supplementary material, which is available to authorized users.

✉ Daniel Väisänen
daniel.vaisanen@gih.se

¹ Åstrand Laboratory of Work Physiology, The Swedish School of Sport and Health Sciences, Box 5626, 114 86 Stockholm, Sweden

² Department of Neuroscience, Karolinska Institute, 171 77 Stockholm, Sweden

³ Aging Research Center, Karolinska Institute, Stockholm University, 171 77 Stockholm, Sweden

Introduction

Cardiorespiratory fitness (CRF) is established as a strong predictor of health (Kodama et al. 2009; Harber et al. 2017). A single measurement of CRF is a stronger predictor for mortality than high blood pressure, smoking, obesity and type 2 diabetes (Myers et al. 2002). CRF typically decreases with age (Betik and Hepple 2008); the rate of decline accelerates at 45 years and is even faster at 65 years (Jackson et al. 2009). Decreased CRF in the elderly can significantly impair functional capacity in everyday life and increase the risk of cardiovascular mortality (Kokkinos et al. 2010). From a

health perspective, it is therefore particularly useful to monitor CRF in the elderly.

The maximal oxygen uptake (VO_2max) test is the gold standard (Fletcher et al. 2001) for measuring CRF. Performing a VO_2max test is time consuming and requires expertise and expensive ventilatory gas-exchange equipment. The test also requires the participant to perform a maximal effort that can be intimidating for some parts of the population. It is especially, challenging for an elderly population prone to abnormal gait (Mahlknecht et al. 2013), impaired balance (Lin and Bhattacharyya 2012), and muscular weakness (Julius et al. 1967). In addition, the elderly are often more apprehensive about performing maximal effort than younger age groups. Since CRF is such an important predictor of health outcomes, increasing its availability may enable identification of elderly individuals with low VO_2max in need of medical care or lifestyle interventions. The American Heart Association has stated that CRF should be used as a clinical evaluation tool (Ross et al. 2016).

Submaximal tests estimate VO_2max based on heart rate response at one or more submaximal work rates (Noonan and Dean 2000). The Åstrand test (Å-test) (Åstrand and Ryhming 1954; Åstrand 1960) is one of the most commonly used submaximal cycle ergometer tests and utilizes the heart rate response to one submaximal work rate. This test has been validated for a population up to 65 years. The validity of this method for individuals older than 65 years is to a large extent unknown.

Another predictive submaximal VO_2max ergometer cycle test is the Ekblom-Bak (EB) test (Ekblom-Bak et al. 2014; Bjorkman et al. 2016). The EB-test consists of exercise at one standardized, low work rate followed by a higher, individually set work rate. This test has been developed and validated in a mixed sample of men and women (aged 20–86 years), with a wide range of VO_2max (ranging from 1.33 to 5.97 L min^{-1}). This test has shown reasonably strong validity (Ekblom-Bak et al. 2014; Bjorkman et al. 2016) with a coefficient of variation (CV) of 9.2% and 8.4% for women and men, respectively and a standard error of the estimate (SEE) of 0.24 L min^{-1} and 0.31 L min^{-1} for women and men, respectively.

Since neither tests has been validated for use in an elderly population of men nor women with a lower VO_2max , the aim of this study was to validate the submaximal EB-test and Å-test in an elderly (> 65 years) population, using

directly measured VO_2max , as a reference. Based on previous research, we speculated that both the EB-test and the Å-test would give valid estimates of VO_2max in an elderly population.

Methods

Participants

Participants were recruited through local newspapers and flyers. Exclusion criteria were severe joint problems, very high blood pressure, or other cardiovascular problems, psychiatric illness or neurological disease. Of the initially screened 170 volunteers, 120 performed the required maximal and submaximal tests for the present study. After being assessed for the VO_2max criteria and the submaximal heart rate criteria for the Å method, a final sample of 104 participants was included (52 women and 52 men, age range 65–75 years with a mean age of 70.6 ± 2.9 years). Prior to undertaking the physical tests, participants answered a single item categorical answer mode questionnaire (Olsson et al. 2016), where they self-rated moderate to vigorous physical activity.

All participants visited the Åstrand Laboratory at the Swedish School of Sport and Health Sciences on one occasion to perform the submaximal and maximal testing. Test duration was approximately 60 min. Each participant provided written consent before the start of the tests. The present study is a part of a larger study, which was approved by the Stockholm ethical committee (2017/1115-31/4). Table 1 shows participant characteristics.

Submaximal and maximal aerobic tests

Participants were instructed to abstain from eating 90 min prior and not to consume caffeine or nicotine 2 hours prior to testing. Furthermore, they were instructed not to perform any heavy or prolonged physical activity the day before or on the day of the test. The test started with participants resting in a seated position for 15 min, where they received oral information about the test procedure. Instructions were also given on how to use Borg's scale of perceived exertion (Borg

Table 1 Anthropometry and physiological characteristics of the study sample, mean \pm standard deviation

	Age (years)	BMI	Body mass (kg)	Height (cm)	Measured VO_2max (L min^{-1})	Measured VO_2max ($\text{mL kg}^{-1} \text{min}^{-1}$)	HRmax (beats/min)
All ($n = 104$)	70.6 ± 2.9	25.5 ± 3.2	75.4 ± 12.8	171.5 ± 8.9	2.36 ± 0.55	31.4 ± 5.3	165 ± 12
Women ($n = 52$)	70.6 ± 2.9	24.6 ± 3.3	66.6 ± 9.9	164.5 ± 5.1	1.89 ± 0.22	28.8 ± 4.6	167 ± 10
Men ($n = 52$)	70.5 ± 2.9	26.4 ± 2.8	83.7 ± 9.2	178.0 ± 6.4	2.82 ± 0.34	34.0 ± 4.6	162 ± 13

1970). Participants were equipped with a heart rate (HR) monitor (Polar model H7, Kempele, Finland) and watch (Polar model M400, Kempele, Finland). The test was initiated at resting heart rate.

Submaximal test

The submaximal test was performed on a mechanically braked cycle ergometer (Monark model 828E, Varberg, Sweden). Participants were instructed to pedal at a cadence of 60 RPM and not speak or adjust position for the duration of the test. Total duration of the test was 8 min, with the initial 4 min performed at a fixed work rate of 0.5 kilo pounds (kp), directly followed by 4 min at a higher individualized work rate that varied between 1–3 kp. The individualized work rate was subjectively chosen by the test leader with regard to gender, body size, training background, and training status. Mean HR was recorded for the final minute of each work rate (calculated as the average of the heart rate recorded at 3:15, 3:30, 3:45, and 4:00). At the high work rate, if the Borg RPE was lower than 12 after the first minute, the load was further increased and the test duration was increased by 1 min.

Maximal test

A maximal incremental treadmill test was performed directly following the submaximal test. A maximal test uses the directly measured gas-exchange during an incremental effort to assess the gas-exchange threshold and hence the oxygen consumption (VO_2max). To measure gas-exchange (O_2 and CO_2), a mask with a flow meter connected to a gas analyzer (Jaeger Oxycon Pro, Hoechberg, Germany) was used. Flow meter volume and gas were calibrated prior to each test using a precision gas mixture ($15.00 \pm 0.01\%$ O_2 and $6.00 \pm 0.01\%$ CO_2 , Air Liquid, Kungsängen, Sweden) and ambient indoor air. All participants wore a safety harness attached to the roof when performing the maximal treadmill test.

Before the maximal test, participants were allowed a short rest (~2 min) and a familiarization/warmup session (~1–2 min) followed by another short rest (~1–2 min). When performing the familiarization/warmup session, participants initially walked at an individualized comfortable speed and 1° inclination and progressed to an individualized higher speed and incline. The maximal incremental VO_2max test started at a comfortable pace and 1° inclination, most times at a walking speed of 3.5–5.0 km/h. The protocol for each participant was individually set, with the aim of reaching respiratory exchange ratio (RER) 1.0 at 5–6 min and RER 1.1 at 6–8 min. A few participants were unable to run due to previous injuries and instead walked at a moderate speed and steep inclination. At the end of the maximal incremental

test, speed and inclination were increased more frequently to ensure a maximal plateau of oxygen consumption (leveling off) was reached. Criteria for an approved test were: a leveling off in VO_2 despite an increase in speed or incline, a $\text{RER} \geq 1.1$, $\text{RPE} \geq 17$, or test duration ≥ 5 min. When the plateau criteria and two of the remaining three criteria were met, the test was accepted as a VO_2max . A small coefficient of variation (2.7%) between a first and second test using the same protocol as mentioned above has previously been reported (Howley et al. 1995; Ekblom-Bak et al. 2014).

Data analysis

The EB-test calculations estimated VO_2max (in L min^{-1}) using the gender-specific equations (equation for women: $\ln \text{VO}_2\text{max} = 1.84390 - 0.00673 (\text{age}) - 0.62578 (\Delta\text{HR}/\Delta\text{PO}) + 0.00175 (\Delta\text{PO}) - 0.00471 (\text{HR at standard work rate})$. Equation for men: $\ln \text{VO}_2\text{max} = 2.04900 - 0.00858 (\text{age}) - 0.90742 (\Delta\text{HR}/\Delta\text{PO}) + 0.00178 (\Delta\text{PO}) - 0.00290 (\text{HR at standard work rate})$) (Bjorkman et al. 2016). The equations use the increase in heart rate (HR) in relation to the increase in work rate (PO), sex, age, and the HR at the lower and higher work rates. For the \dot{V} method, HR was taken from the higher work rate of the EB-test. The \dot{V} -test calculations estimated VO_2max using the \dot{V} strand nomogram and the \dot{V} strand extrapolated age correction factors (with a decreasing factor from 65 years of 0.006 per year).

For relative VO_2max , the absolute L min^{-1} was divided with body weight and multiplied by 1000 to get $\text{ml kg}^{-1} \text{min}^{-1}$. Low fit was seen as the quartile with the lowest measured VO_2max in women and men, respectively. High fit was seen as the quartile with the highest measured VO_2max in women and men, respectively.

Statistical analysis

SPSS Inc (Chicago, III, US) was used for all statistical analyses. Descriptive data are presented as mean \pm SD (range). A Shapiro–Wilk test was used to determine normal distribution, which was present for all tested parameters.

Pearson's correlation coefficient (r) was calculated between the variables estimated from the submaximal test and directly measured during the maximal test. Paired Student's t tests were used to determine differences between measured and estimated VO_2max . To determine whether validity was different for different fitness levels, HR levels, or self-reported physical activity levels, Pearson's or Spearman's correlation coefficients were calculated between the difference of estimated and measured VO_2max and VO_2max , maximal heart rate, and self-rated physical activity. We regarded the Pearson's r and Spearman's ρ as weak (< 0.10),

modest (0.1–0.3), moderate (0.3–0.5), strong (0.5–0.8), or very strong (0.8–1.0). Standard errors of the estimate (SEE) were derived from a linear regression model to show the variation around the regression line. To determine the variation in relation to its mean, we used coefficient of variation (CV), which was calculated using the SD of the differences between measured and estimated VO_2max divided by the mean of the two methods. 95% confidence intervals (95% CI) were calculated for the difference between estimated and measured VO_2max . Limits of Agreement (LoA) was calculated using the equation: mean of the difference between estimated and measured $\text{VO}_2\text{max} \pm 1.96$ multiplied by the SD of difference between the two methods. LoA is expected to include 95% of the differences between the two measurement methods. Significance level was set at $p < 0.05$ and we regarded $0.05 \geq p < 0.1$ as a trend (Curran-Everett and Benos 2004).

Results

Mean measured VO_2max in women was $1.89 \text{ L min}^{-1} \pm 0.22$ and mean estimated VO_2max was $1.88 \pm 0.3 \text{ L min}^{-1}$, using the EB method and $1.76 \text{ L min}^{-1} \pm 0.44$, using the $\dot{V}\text{O}_2$ method. In men, mean measured VO_2max was $2.83 \text{ L min}^{-1} \pm 0.4$ and mean estimated VO_2max was 2.88 ± 0.3 , using the EB method and $2.54 \text{ L min}^{-1} \pm 0.55$, using the $\dot{V}\text{O}_2$ method.

There were no significant differences between estimated and measured VO_2max in neither women nor men for the EB method. However, the $\dot{V}\text{O}_2$ method significantly underestimated women's fitness by -0.12 L min^{-1} (95% CI -0.22 to -0.02) and men's fitness by -0.28 L min^{-1} (95% CI -0.42 to -0.14) (Table 2 and Fig. 1). LoA for women was -0.43 – 0.39 L min^{-1} , when using the EB method and -0.83 – 0.59 L min^{-1} , when using the $\dot{V}\text{O}_2$ method. Corresponding values for men were -0.60 – 0.70 L min^{-1} and -1.28 – 0.71 L min^{-1} , respectively (see online resource 1 to 4).

CV was somewhat lower for the EB method compared to the $\dot{V}\text{O}_2$ method in both women (11.1% vs. 19.8%) and men (11.6% vs. 18.9%), accompanied with a smaller SEE in both women (0.20 L min^{-1} vs. 0.36 L min^{-1}) and men (0.25 L min^{-1} vs. 0.50 L min^{-1}) (Table 2).

Similar tendencies were found for relative estimated VO_2max ($\text{ml kg}^{-1} \text{ min}^{-1}$) as for estimated absolute VO_2max (L min^{-1}) (Table 3). The EB method showed no bias for women ($-0.02 \text{ ml kg}^{-1} \text{ min}^{-1}$, 95% CI -0.08 to 0.04) or men ($0.05 \text{ ml kg}^{-1} \text{ min}^{-1}$, 95% CI -0.04 to 0.14), while the $\dot{V}\text{O}_2$ method significantly underestimated women ($-0.12 \text{ ml kg}^{-1} \text{ min}^{-1}$, 95% CI -0.22 to -0.02) and men ($-0.28 \text{ ml kg}^{-1} \text{ min}^{-1}$, 95% CI -0.42 to -0.14). CV for the EB method and the $\dot{V}\text{O}_2$ method was 11.1% vs. 19.8% for women and 11.6% vs. 18.9% for men, respectively.

In women, the estimated error (difference between estimated and measured VO_2max) for the EB method displayed a trend toward being associated with VO_2max level ($p = 0.051$, $r = -0.27$) and self-rated physical activity level ($p = 0.059$, $\rho = 0.26$). It was also significantly associated with maximal HR level ($p < 0.01$, $r = -0.60$). In men, the estimated error was associated with VO_2max level ($p < 0.01$, $r = -0.67$) and maximal HR level ($p < 0.05$, $r = -0.32$), but not self-rated physical activity level ($p = 0.22$, $\rho = -0.17$). Estimated error for the $\dot{V}\text{O}_2$ method in women was not associated with VO_2max level ($p = 0.50$, $r = -0.10$), but was with maximal HR level ($p < 0.01$, $r = -0.654$) and self-rated physical activity level ($p < 0.05$, $\rho = 0.28$). In men, estimated error was not associated with VO_2max level ($p = 0.17$, $r = -0.19$) or self-rated physical activity level ($p > 0.05$, $\rho = 0.05$), but was associated with maximal HR level ($p < 0.01$, $r = -0.60$).

Discussion

The main finding was that there was good agreement between measured and estimated VO_2max when using the EB method in a population of elderly men and women. In addition, there was good agreement between estimated, using the EB method, and measured VO_2max over the full VO_2max spectrum for women. Low fit men were overestimated and high fit men were partly underestimated using the EB method. The $\dot{V}\text{O}_2$ method significantly underestimated VO_2max in both women and men. Precision expressed as CV and SEE for the EB method was almost half that of the $\dot{V}\text{O}_2$ method in both men and women.

No previous studies have validated the EB or the $\dot{V}\text{O}_2$ -test in a large elderly population (> 65 years). However, there are other submaximal tests commonly used for the elderly and one of them is the 6 min walk test (6MWT). The 6MWT was developed to estimate VO_2max from a single test (Ebbeling et al. 1991). This test has shown varying results in individuals with cardiopulmonary disorders ($r = 0.21$ – 0.70 , mean $r = 0.59$) (Ross et al. 2010) and does not seem to be a valid test for relatively healthy elderly populations (r between estimated and measured VO_2max for women was not significant, men $r = 0.8$) (Andersson et al. 2011). The 6MWT is easy to perform and has a high correlation with measured VO_2max for elderly men but not for elderly women, and the test has a high variability with a relative SEE of $\sim 27\%$ (Ross et al. 2010). The 5 min pyramid test (5MPT) has a strong correlation to measured VO_2max in elderly women ($r = 0.78$) and a very strong correlation for elderly men ($r = 0.98$) (Andersson et al. 2011). However, the 5MPT is a maximal test where factors such as motivation and anaerobic capacity may impact results. In comparison, both the EB-test and $\dot{V}\text{O}_2$ -test are performed submaximally, making them more accessible for populations that are not willing or able to perform

Table 2 Validity of the EB-test and the $\dot{V}O_{2\max}$, absolute $\dot{V}O_{2\max}$, mean \pm standard deviation

	Measured $\dot{V}O_{2\max}$ (L min ⁻¹) mean \pm SD		Estimated $\dot{V}O_{2\max}$ mean \pm SD (L min ⁻¹)		Difference estimated vs. measured $\dot{V}O_{2\max}$ (L min ⁻¹) mean (95% CI)		Correlation coefficient (<i>r</i>)		Coefficient of variation (%)		SEE (L min ⁻¹)	
	EB-test	$\dot{V}O_{2\max}$	EB-test	$\dot{V}O_{2\max}$	EB-test	$\dot{V}O_{2\max}$	EB-test	$\dot{V}O_{2\max}$	EB-test	$\dot{V}O_{2\max}$	EB-test	$\dot{V}O_{2\max}$
All (<i>n</i> = 104)	2.36 \pm 0.55	2.37 \pm 0.57	2.15 \pm 0.63	2.15 \pm 0.63	0.02 (-0.03 to 0.07)	-0.20 (-0.29 to 0.12)	0.88**	0.73**	11.7	19.8	0.27	0.44
Women (<i>n</i> = 52)	1.89 \pm 0.22	1.87 \pm 0.26	1.76 \pm 0.44	1.76 \pm 0.44	-0.02 (-0.08 to 0.04)	-0.12 (-0.22 to -0.02)	0.64**	0.58**	11.1	19.8	0.20	0.36
Quartiles according to absolute $\dot{V}O_{2\max}$ level (L min ⁻¹)												
Q1 (<i>n</i> = 13)	1.61 \pm 0.08	1.64 \pm 0.22	1.45 \pm 0.28	1.45 \pm 0.28	0.03 (-0.10 to 0.17)	-0.15 (-0.33 to 0.02)	0.16	0.04	13.9	18.8	0.23	0.29
Q2 (<i>n</i> = 13)	1.78 \pm 0.05	1.79 \pm 0.14	1.59 \pm 0.21	1.59 \pm 0.21	0.01 (-0.08 to 0.11)	-0.18 (-0.32 to -0.05)	-0.07	-0.16	8.8	13.2	0.15	0.22
Q3 (<i>n</i> = 13)	2.00 \pm 0.04	1.94 \pm 0.21	1.87 \pm 0.35	1.87 \pm 0.35	-0.06 (-0.19 to 0.07)	-0.13 (-0.33 to 0.08)	0.15	0.21	10.6	17.7	0.22	0.36
Q4 (<i>n</i> = 13)	2.16 \pm 0.09	2.10 \pm 0.21	2.14 \pm 0.53	2.14 \pm 0.53	-0.06 (-0.21 to 0.08)	-0.02 (-0.35 to 0.30)	-0.11	0.23	11.2	24.9	0.22	0.56
Men (<i>n</i> = 52)	2.82 \pm 0.34	2.88 \pm 0.27	2.54 \pm 0.55	2.54 \pm 0.55	0.05 (-0.04 to 0.14)	-0.28 (-0.42 to -0.14)	0.44**	0.44**	11.6	18.9	0.25	0.50
Quartiles according to absolute $\dot{V}O_{2\max}$ level (L min ⁻¹)												
Q1 (<i>n</i> = 13)	2.41 \pm 0.12	2.86 \pm 0.20	2.48 \pm 0.45	2.48 \pm 0.45	0.46 (0.35 to 0.57)	0.07 (-0.16 to 0.31)	0.33	0.59	7.4	16.0	0.19	0.38
Q2 (<i>n</i> = 14)	2.71 \pm 0.07	2.67 \pm 0.28	2.21 \pm 0.57	2.21 \pm 0.57	-0.03 (-0.18 to 0.12)	-0.50 (-0.82 to -0.18)	0.38	0.44	9.8	22.7	0.27	0.57
Q3 (<i>n</i> = 12)	2.94 \pm 0.06	2.89 \pm 0.22	2.50 \pm 0.39	2.50 \pm 0.39	-0.05 (-0.19 to 0.09)	-0.44 (-0.68 to -0.20)	0.11	0.21	7.6	13.8	0.23	0.39
Q4 (<i>n</i> = 13)	3.26 \pm 0.21	3.10 \pm 0.22	3.00 \pm 0.50	3.00 \pm 0.50	-0.17 (-0.31 to -0.02)	-0.51 (-0.57 to 0.04)	0.34	0.39	7.7	16.3	0.22	0.51

Quartiles in L min⁻¹, women; < 1.69, 1.69–1.89, 1.89–2.07, 2.07 >; men; < 2.55, 2.55–2.80, 2.80–3.02, 3.02 >

p* < 0.05, *p* < 0.01

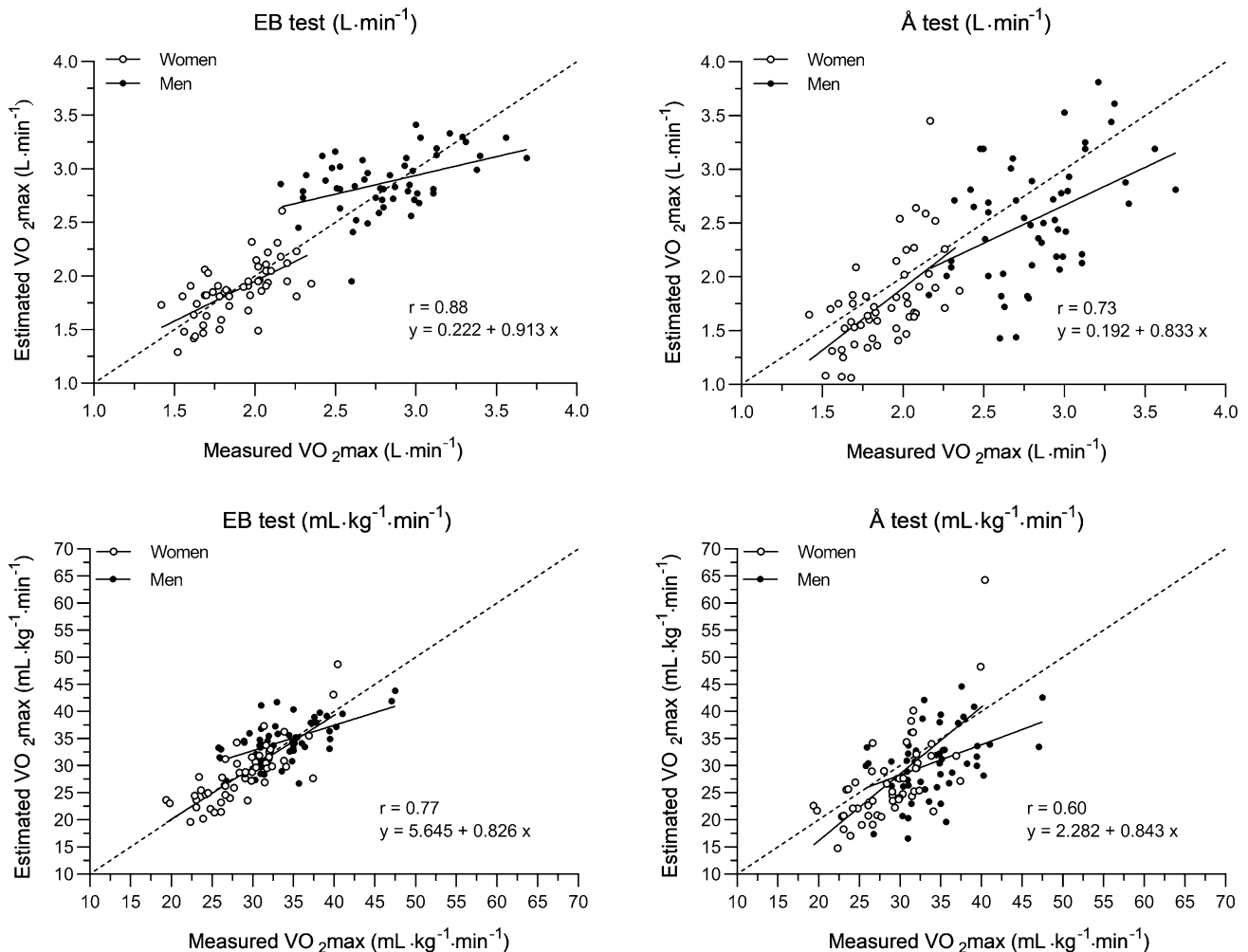


Fig. 1 The submaximal EB method and the Å method vs. absolute (L min⁻¹) and relative (ml kg⁻¹ min⁻¹) measured VO_2max . The correlation coefficients and equations in the figure are for both genders together. Correlation coefficient, r , for absolute estimated VO_2max in women was 0.64 (EB method) and 0.55 (Å method). Correla-

tion coefficient, r , for absolute estimated VO_2max in men was 0.44 (EB method and Å method). Correlation coefficient, r , for relative estimated VO_2max in women was 0.81 (EB method) and 0.70 (Å method). Correlation coefficient, r , for relative estimated VO_2max in men was 0.56 (EB method) and 0.49 (Å method)

maximal effort. The EB method and the Å method had a relative SEE of 11.4% and 20.5%, respectively, in the present study which is better than in the 6MWT.

Although estimated VO_2max using the EB method agreed with measured VO_2max in the overall male and female population, estimated VO_2max for men with low fitness was found to be overestimated. In a previous validation study, including young girls and boys (age 10–15 years), estimated VO_2max in pre-pubertal boys with similar levels of measured VO_2max as the low fitness men in the present study was also found to be significantly overestimated (Bjorkman et al. 2018). However, when the EB equation for estimating VO_2max in women was applied to the pre-pubertal boys in the Bjorkman et al. study, a significantly higher validity and agreement with measured VO_2max was seen. Hence, we reanalyzed the data from the low fitness men using the EB

equation for women. This resulted in a decreased estimation error to non-significant levels [0.04 L min^{-1} (-0.02 – 0.09)] and a subsequent higher correlation in all men ($r = 0.73$ using the EB equation for women, compared to $r = 0.44$ using the EB equation for men). However, it was not possible to distinguish the men with low cardiorespiratory fitness using any other variables in the current study other than measured VO_2max .

We speculate that a decline in testosterone levels with age could affect physiological variables (DeFina et al. 2018; Hosick et al. 2018; Kelsey et al. 2014), which may influence results when using the EB equation for men. Another explanation might be that both pre-pubertal boys and elderly men with low absolute VO_2max are outside or at the lower end of the VO_2max range in the EB sample for men in Bjorkman et al. (2016), indicating a more uncertain estimation

Table 3 Validity of the EB-test and the Å-test, Relative VO₂max, mean ± standard deviation

	Measured VO ₂ max (mL kg ⁻¹ min ⁻¹) mean ± SD		Estimated VO ₂ max (mL kg ⁻¹ min ⁻¹) mean ± SD		Difference estimated vs. measured VO ₂ max (mL kg ⁻¹ min ⁻¹) mean (95% CI)		Correlation coef- ficient (r)		Coefficient of variation (%)		SEE (mL kg ⁻¹ min ⁻¹)	
	EB-test	Å-test	EB-test	Å-test	EB-test	Å-test	EB-test	Å-test	EB-test	Å-test	EB-test	Å-test
All (n = 104)	31.4 ± 5.3		31.6 ± 5.6	28.8 ± 7.5	0.19 (-0.53 to 0.91)	-2.66 (-3.83 to -1.5)	0.77*	0.60*	11.7	20.1	3.59	6.02
Women (n = 52)	28.8 ± 4.6		28.6 ± 5.5	27.0 ± 8.2	-0.25 (-1.16 to 0.66)	-1.80 (-3.46 to -0.15)	0.81*	0.70*	11.4	21.3	3.30	5.89
Quartiles according to absolute VO ₂ max level (L min ⁻¹)												
Q1 (n = 13)	26.5 ± 4.9		26.8 ± 4.6	23.7 ± 4.7	0.30 (-1.87 to 2.46)	-2.89 (-5.81 to 0.02)	0.72**	0.49	13.4	19.2	3.33	4.24
Q2 (n = 13)	27.5 ± 3.3		27.7 ± 4.7	24.8 ± 5.6	0.28 (-1.22 to 1.79)	-2.65 (-4.82 to -0.47)	0.86**	0.80**	9.0	13.7	2.49	3.54
Q3 (n = 13)	29.7 ± 4.5		28.7 ± 4.2	27.6 ± 5.1	-1.07 (-3.15 to 1.02)	-2.14 (-5.25 to 0.97)	0.69**	0.44	11.8	18.0	3.20	4.80
Q4 (n = 13)	31.5 ± 4.4		31.0 ± 7.6	31.9 ± 12.6	-0.51 (-2.70 to 1.67)	0.46 (-4.92 to 5.84)	0.96**	0.89**	11.6	28.1	2.28	5.90
Men (n = 52)	34.0 ± 4.6		34.7 ± 3.9	30.5 ± 6.3	0.63 (-0.50 to 1.76)	-3.51 (-5.21 to -1.81)	0.56*	0.49*	11.8	18.9	3.27	5.82
Quartiles according to absolute VO ₂ max level (L min ⁻¹)												
Q1 (n = 13)	30.2 ± 3.4		35.8 ± 3.2	30.8 ± 4.2	5.62 (4.17 to 7.08)	0.58 (-2.41 to 3.57)	0.73**	0.16	7.3	16.2	2.28	4.35
Q2 (n = 14)	33.8 ± 3.4		33.3 ± 4.1	27.6 ± 7.7	-0.50 (-2.49 to 1.50)	-6.18 (-10.16 to -2.19)	0.59*	0.44	10.3	22.5	3.44	7.19
Q3 (n = 12)	34.7 ± 3.1		34.1 ± 3.2	29.5 ± 4.6	-0.67 (-2.27 to 0.93)	-5.22 (-7.96 to -2.49)	0.68*	0.43	7.3	13.4	2.46	4.36
Q4 (n = 13)	38.4 ± 5.4		35.5 ± 4.7	34.3 ± 6.5	-1.96 (-3.67 to -0.25)	-3.15 (-6.8 to 0.46)	0.85**	0.50	7.8	16.7	2.58	5.84

Quartiles in L min⁻¹, women; < 1.69, 1.69–1.89, 1.89–2.07, 2.07 > ; men; < 2.55, 2.55–2.80, 2.80–3.02, 3.02 >
 *p < 0.05, **p < 0.01

of VO_2max in all low fit men ($< 2.55 \text{ L min}^{-1}$). For clinical practice, the overestimation of elderly men in the present study with low absolute CRF means that elderly men with a general low fitness level and small body stature should be considered at risk of having their VO_2max overestimated when using the EB-test. Therefore, it would be advantageous for this population to use the EB equation for women.

The Å-test was initially developed for 18–30 year old fairly well-trained individuals (Åstrand and Ryhming 1954). Later, an age correction formula was developed to be able to apply the test to a wider range of ages (Åstrand 1960). However, even with the age correction formula, many studies have reported that the Å-test still underestimates VO_2max (Jessup et al. 1977; Jette 1979; Kasch 1984; Hartung et al. 1993), in agreement with the present study on elderly. This could be partly due to the Å-test having been developed with data from both maximal cycle and treadmill tests, since measured VO_2max tends to be lower when using a cycle ergometer compared to a treadmill (Hermansen et al. 1970). Moreover, in 1960, when the age correction factor was developed, the creators raised a concern that it may underestimate the VO_2max of older adults by 10% since they displayed lower lactate levels compared to the young adults in the study (Åstrand 1960). This led to the belief that the older participants may not have reached their true VO_2max . Another important factor that could go toward explaining the underestimation of the Å-test in the present study is that the test was designed to be performed on two occasions to eliminate variables such as nervousness and other factors that affect the absolute submaximal HR level on the first test occasion. The EB and the Å-tests both assume a decline in VO_2max with age, which is usually the case in larger populations (Jackson et al. 2009). A discrepancy between biological and chronological age is a source of error that increases the uncertainty of the EB-test and the Å-test.

In previous studies (Eklom-Bak et al. 2014; Bjorkman et al. 2016), the difference between estimated and measured VO_2max using the EB-test was not dependent on maximal HR level. However in the present study, the difference between estimated and measured VO_2max was found to be dependent on maximal HR for both women and men, when using the EB method. Incidentally, the same result was also seen when using the Å method. The lack of agreement between the current and previous studies is possibly due to the age difference in the sample populations (elderly vs. age mixed), resulting in a higher number of individuals with a low maximal HR in the present study. The present study showed that an individual with low maximal HR has a higher risk of having their VO_2max overestimated. In previous studies using the EB-test, the difference between estimated and measured VO_2max was dependent on VO_2max , as it was in the present study for men, while for women there was a tendency ($p = 0.051$). In other words, elderly men

with low measured absolute VO_2max are at risk of being overestimated and elderly men with high VO_2max are at risk of being underestimated, when using the EB method. The difference between estimated and measured VO_2max had a tendency toward a relationship ($p = 0.059$), for the EB method, and was significantly correlated, for the Å method, with self-rated physical activity for women, but not for men. This suggests that the self-rated single item questionnaire may be useful for further strengthening the female, but not the male, algorithms for the estimation of VO_2max .

Strengths and limitations

The strength of the study is the modality with which the measured VO_2max tests were undertaken, i.e. walking or running on a treadmill. It is a modality that most people are familiar with (Lear et al. 1999). A maximal test on a treadmill utilizes greater recruitment of exercising muscle mass than a VO_2max test performed on a cycle ergometer, where local fatigue in leg musculature could lead up to an 20% lower VO_2max (Myers et al. 1991). On the other hand, when performing a submaximal VO_2max test, it is better to use a cycle ergometer because of the low variability in energy expenditure at a certain work rate between individuals (Eklom and Gjessing 1968).

This study adds to the pool of studies investigating directly measured CRF in an elderly population. Mean VO_2max in the present study sample was similar to an elderly group in a previous study using the EB-test (Bjorkman et al. 2016). In comparison, a large Norwegian study, where a sample ($n = 129$) of > 70 year olds was tested on a treadmill, reported similar VO_2max levels (women $1.85 \pm 0.35 \text{ L min}^{-1}$, men $2.81 \pm 0.5 \text{ L min}^{-1}$) (Loe et al. 2013) to the present study. Other Nordic studies where VO_2max was measured directly in the elderly using a cycle ergometer have shown slightly lower VO_2max values (Andersson et al. 2011; Eriksen et al. 2016). A limitation in the present study could be the self-selection and exclusion criteria resulting in a selected sample of participants. Most likely, this resulted in the present sample being biased toward a higher CRF than would generally be seen in the elderly population in Sweden. The present findings indicate that the EB-test is a good test for the elderly population, but that population-based studies will ultimately be required to ensure generalizability.

Future perspectives

It has been shown that the EB-test can be used to monitor long-term changes in CRF in an age and gender mixed population (Bjorkman 2017). Future intervention studies are

needed to evaluate the ability of the EB-test to also monitor CRF in an elderly population and to identify subtle changes that may result from health promoting interventions such as a physical activity on prescription (Kallings et al. 2008). Another important topic for future research is if the EB-test is affected by certain medicines such as beta-adrenergic blockers and stimulators that affect the function of the cardiorespiratory system. Lastly it would be advantageous to further study the relationship between the EB-test and measured VO_2max in low fit men and thereafter possibly adjust the EB equation for better VO_2max estimation.

Conclusion

The validity of the EB-test in an elderly population was satisfactory both in women and men combined and in women alone but not in men. We found a moderate correlation between the EB method and measured VO_2max in men; however, there was an overestimation of VO_2max in men with low fitness. The moderate correlation and overestimation of VO_2max in low fit men, in contrast to the good correlation and the similarity between estimated and measured VO_2max in women, could be due to a gender difference in the physiological variables that affect VO_2max with increasing age. Alternatively, the EB equation for men is unable to correctly estimate VO_2max in men of all ages with low absolute VO_2max . The Å method significantly underestimated VO_2max in both women and men and had a variability that was almost twice that of the EB method. The current study therefore supports using both submaximal methods for population-based studies aiming to evaluate cardiorespiratory fitness as a predictor of health outcomes. On an individual level, the EB method appears suitable for estimating CRF in elderly women, but has insufficient precision for this purpose in elderly men. While the Å method was more accurate than the EB method at identifying the low fit men, its high variability still suggests that it should not be used alone for identifying individuals in need for lifestyle or medical support in elderly populations.

Acknowledgements Open access funding provided by Swedish School of Sport and Health Sciences (GIH).

Author contributions ME, ÖE, JN contributed to the original study conception, research question of this validation sub-study was written by DV. Data was collected by DV and EA. Analysis and interpretation was performed by DV. The first draft of the manuscript was written by DV, ME, ÖE, EA, and EEB commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This research received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) and ERC Grant no. 617280-REBOOT. This

research was also funded by a grant from the Swedish Research Council (446-2013-7189).

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee in Stockholm (2017/1115-31/4) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andersson EA, Lundahl G, Wecke L, Lindblom I, Nilsson J (2011) Maximal aerobic power versus performance in two aerobic endurance tests among young and old adults. *Gerontology* 57(6):502–512. <https://doi.org/10.1159/000329174>
- Astrand I (1960) Aerobic work capacity in men and women with special reference to age. *Acta Physiol Scand Suppl* 49(169):1–92
- Astrand PO, Ryhming I (1954) A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during sub-maximal work. *J Appl Physiol* 7(2):218–221. <https://doi.org/10.1152/jappl.1954.7.2.218>
- Betik AC, Hepple RT (2008) Determinants of VO_2max decline with aging: an integrated perspective. *Appl Physiol Nutr Metab* 33(1):130–140. <https://doi.org/10.1139/H07-174>
- Bjorkman F (2017) Validity and reliability of a submaximal cycle ergometer test for estimation of maximal oxygen uptake. Pp. 1–126.
- Bjorkman F, Ekblom-Bak E, Ekblom O, Ekblom B (2016) Validity of the revised Ekblom Bak cycle ergometer test in adults. *Eur J Appl Physiol* 116(9):1627–1638. <https://doi.org/10.1007/s00421-016-3412-0>
- Bjorkman F, Eggers A, Stenman A, Bohman T, Ekblom B, Ekblom O (2018) Sex and maturity status affected the validity of a submaximal cycle test in adolescents. *Acta Paediatr* 107(1):126–133. <https://doi.org/10.1111/apa.14080>
- Borg G (1970) Perceived exertion as an indicator of somatic stress. *Scand J Rehabil Med* 2(2):92–98
- Curran-Everett D, Benos DJ (2004) Guidelines for reporting statistics in journals published by the American Physiological Society. *Am J Physiol Regul Integr Comp Physiol* 287(2):R247–249. <https://doi.org/10.1152/ajpregu.00346.2004>
- DeFina LF, Radford NB, Leonard D, Wilson RK, Cooper TC, Clark SM, Willis BL et al (2018) The association of cardiorespiratory fitness, body mass index, and age with testosterone levels at

- screening of healthy men undergoing preventive medical examinations: The Cooper Center Longitudinal Study. *Maturitas* 118:1–6. <https://doi.org/10.1016/j.maturitas.2018.09.004>
- Ebbeling CB, Ward A, Puleo EM, Widrick J, Rippe JM (1991) Development of a single-stage submaximal treadmill walking test. *Med Sci Sports Exerc* 23(8):966–973
- Eklblom B, Gjessing E (1968) Maximal oxygen uptake of the Easter Island population. *J Appl Physiol* 25(2):124–129. <https://doi.org/10.1152/jappl.1968.25.2.124>
- Eklblom-Bak E, Bjorkman F, Hellenius ML, Eklblom B (2014) A new submaximal cycle ergometer test for prediction of VO₂max. *Scand J Med Sci Sports* 24(2):319–326. <https://doi.org/10.1111/sms.12014>
- Eriksen L, Gronbaek M, Helge JW, Tolstrup JS (2016) Cardiorespiratory fitness in 16 025 adults aged 18–91 years and associations with physical activity and sitting time. *Scand J Med Sci Sports* 26(12):1435–1443. <https://doi.org/10.1111/sms.12608>
- Fletcher GF, Balady GJ, Amsterdam EA, Chaitman B, Eckel R, Fleg J, Froelicher VF et al (2001) Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association. *Circulation* 104(14):1694–1740
- Harber MP, Kaminsky LA, Arena R, Blair SN, Franklin BA, Myers J, Ross R (2017) Impact of cardiorespiratory fitness on all-cause and disease-specific mortality: advances since 2009. *Prog Cardiovasc Dis* 60(1):11–20. <https://doi.org/10.1016/j.pcad.2017.03.001>
- Hartung GH, Krock LP, Crandall CG, Bisson RU, Myhre LG (1993) Prediction of maximal oxygen uptake from submaximal exercise testing in aerobically fit and nonfit men. *Aviat Space Environ Med* 64(8):735–740
- Hermansen L, Eklblom B, Saltin B (1970) Cardiac output during submaximal and maximal treadmill and bicycle exercise. *J Appl Physiol* 29(1):82–86. <https://doi.org/10.1152/jappl.1970.29.1.82>
- Hosick PA, Matthews EL, Leigh S (2018) Cardiorespiratory fitness and the relationship between body fat and resting testosterone in men. *Arch Physiol Biochem*. <https://doi.org/10.1080/13813455.2018.1491996>
- Howley ET, Bassett DR Jr, Welch HG (1995) Criteria for maximal oxygen uptake: review and commentary. *Med Sci Sports Exerc* 27(9):1292–1301
- Jackson AS, Sui X, Hebert JR, Church TS, Blair SN (2009) Role of lifestyle and aging on the longitudinal change in cardiorespiratory fitness. *Arch Intern Med* 169(19):1781–1787. <https://doi.org/10.1001/archinternmed.2009.312>
- Jessup GT, Riggs CE, Lambert J, Miller WD (1977) The effect of pedalling speed on the validity of the Astrand-Rhyming aerobic work capacity test. *J Sports Med Phys Fitness* 17(4):367–371
- Jette M (1979) A comparison between predicted VO₂ max from the Astrand procedure and the Canadian home fitness test. *Can J Appl Sport Sci* 4(3):214–218
- Julius S, Amery A, Whitlock LS, Conway J (1967) Influence of age on the hemodynamic response to exercise. *Circulation* 36(2):222–230
- Kallings LV, Leijon M, Hellenius ML, Stahle A (2008) Physical activity on prescription in primary health care: a follow-up of physical activity level and quality of life. *Scand J Med Sci Sports* 18(2):154–161. <https://doi.org/10.1111/j.1600-0838.2007.00678.x>
- Kasch FW (1984) The validity of the Astrand and Sjostrand submaximal tests. *Phys Sportsmed* 12:47–51
- Kelsey TW, Li LQ, Mitchell RT, Whelan A, Anderson RA, Wallace WH (2014) A validated age-related normative model for male total testosterone shows increasing variance but no decline after age 40 years. *PLoS ONE* 9(10):e109346. <https://doi.org/10.1371/journal.pone.0109346>
- Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, Sugawara A et al (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. <https://doi.org/10.1001/jama.2009.681>
- Kokkinos P, Myers J, Faselis C, Panagiotakos DB, Doulmas M, Pittaras A, Manolis A et al (2010) Exercise capacity and mortality in older men: a 20-year follow-up study. *Circulation* 122(8):790–797. <https://doi.org/10.1161/CIRCULATIONAHA.110.938852>
- Lear SA, Brozic A, Myers JN, Ignaszewski A (1999) Exercise stress testing An overview of current guidelines. *Sports Med* 27(5):285–312. <https://doi.org/10.2165/00007256-199927050-00002>
- Lin HW, Bhattacharyya N (2012) Balance disorders in the elderly: epidemiology and functional impact. *Laryngoscope* 122(8):1858–1861. <https://doi.org/10.1002/lary.23376>
- Loe H, Rognmo O, Saltin B, Wisloff U (2013) Aerobic capacity reference data in 3816 healthy men and women 20–90 years. *PLoS ONE* 8(5):e64319. <https://doi.org/10.1371/journal.pone.0064319>
- Mahlknecht P, Kiechl S, Bloem BR, Willeit J, Scherfler C, Gasperi A, Rungger G, Poewe W, Seppi K (2013) Prevalence and burden of gait disorders in elderly men and women aged 60–97 years: a population-based study. *PLoS ONE* 8(7):e69627. <https://doi.org/10.1371/journal.pone.0069627>
- Myers J, Buchanan N, Walsh D, Kraemer M, McAuley P, Hamilton-Wessler M, Froelicher VF (1991) Comparison of the ramp versus standard exercise protocols. *J Am Coll Cardiol* 17(6):1334–1342
- Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE (2002) Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 346(11):793–801. <https://doi.org/10.1056/NEJMoa011858>
- Noonan V, Dean E (2000) Submaximal exercise testing: clinical application and interpretation. *Phys Ther* 80(8):782–807
- Olsson SJ, Eklblom O, Andersson E, Borjesson M, Kallings LV (2016) Categorical answer modes provide superior validity to open answers when asking for level of physical activity: a cross-sectional study. *Scand J Public Health* 44(1):70–76. <https://doi.org/10.1177/1403494815602830>
- Ross RM, Murthy JN, Wollak ID, Jackson AS (2010) The six minute walk test accurately estimates mean peak oxygen uptake. *BMC Pulm Med* 10:31. <https://doi.org/10.1186/1471-2466-10-31>
- Ross R, Blair SN, Arena R, Church TS, Despres JP, Franklin BA, Haskell WL et al (2016) Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement From the American Heart Association. *Circulation* 134(24):e653–e699. <https://doi.org/10.1161/CIR.0000000000000461>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.