Avhandlingsserie för
Gymnastik- och idrottshögskolan

Nr 11

VALIDITY AND RELIABILITY OF A SUBMAXIMAL CYCLE ERGOMETER TEST FOR ESTIMATION OF MAXIMAL OXYGEN UPTAKE

THE SWEDISH SCHOOL OF SPORT AND HEALTH SCIENCES

Validity and reliability of a submaximal cycle ergometer test for estimation of maximal oxygen uptake

Frida Maria Eleonora Björkman

Gymnastik- och idrottshögskolan 2017
ISBN 978-91-983151-2-7

Tryckeri: Universitetsservice US-AB, Stockholm 2017
Distributör: Gymnastik- och idrottshögskolan

To be born free is an accident.
To live free is a privilege.
To die free is a responsibility.

ABSTRACT

Maximal oxygen uptake (VO_2max) is the highest obtained rate of oxygen consumption during a physically intense dynamic whole-body activity. VO_2max is an important factor for many types of physical performance, as well as a strong independent predictor of health and longevity. Thus, it is important to have accurate and precise methods for assessment of VO_2max.

A direct measurement of VO_2max is often conducted via indirect calorimetry during maximal exercise. The demand for maximal effort from an individual, along with the need for laboratory equipment, makes direct measurements unsuitable in the general, non-athlete population. There are also a number of contraindications that limit the possibility to conduct direct measurements of VO_2max in many settings. Instead, several other exercise tests have been developed in order to facilitate the procedure of determination and evaluation of cardiorespiratory fitness in different populations. These tests can be either of submaximal or maximal character. Commonly used work modes are stepping, walking, and cycling. The overall aim of this thesis was to describe the background to, and the development of, submaximal cycle ergometer tests for estimation of VO_2max.

The present thesis focuses on the validity and reliability of a new submaximal cycle ergometer test – the Ekblom-Bak test (EB test). The first study described the test procedure for the new cycle ergometer test and the creation of an accompanying mathematical model (prediction equation) for estimation of VO_2max. The development of the test and its associated prediction equation was continued in study II, while it was further validated in adults and adolescents in study II and IV. Study III examined the ability to use a submaximal cycle ergometer test in order to detect changes in VO_2max over time.

The EB test comprises of 8 minutes of continuous cycling – 4 minutes at 0.5 kp, followed by 4 minutes at a higher, individually chosen work rate – with a pedalling rate of 60 revolutions per minute. The test measures the change in HR (ΔHR) between the two different work rates (ΔPO), and the variable ΔHR/ΔPO was obtained and linked to measured VO_2max. In study I, the validity and reliability of the EB test and the associated prediction equation was tested in a mixed population with regard to sex, age, and physical activity status. The subjects performed repeated submaximal cycle ergometer tests and maximal running tests for direct determination of VO_2max (reference value). There was a strong correlation between estimated and measured VO_2max, with an adjusted R^2 of 0.82 and a corresponding coefficient of variation (CV) of 9.3%. Although there was a relatively high precision in the estimation of VO_2max by the prediction equation, it was evident that individuals with high VO_2max were underestimated and individuals with low VO_2max were overestimated. This issue was further addressed in study II.

In study II, the size of the study population was increased, in order to broaden the valid range and evaluate the use of sex-specific prediction equations. The estimation error was slightly decreased, and the sex-specific prediction equations resulted in an adjusted R^2 of 0.91 and a CV of 8.7% in the whole group. The new models were also evaluated in a cross-validation group, where the adjusted R^2 was 0.90 and CV 9.4%.

The relation between the estimation error and changes in VO_2max over time was investigated in study III. Follow-up tests were conducted in 35 subjects, in order to examine the conformity between changes in measured and estimated VO_2max over a time-span of 5 to 8 years. Results showed a moderate correlation between change in measured VO_2max and change in estimated VO_2max (r = 0.75). Changes in body mass or changes in work efficiency did not relate to the change in assessment error.

In study IV, the aim was to determine the applicability and validity of the EB test in pre-pubertal and pubertal adolescents. Medical examinations and assessment of sexual maturity (according to the stages of Tanner) were performed in addition to the physical tests. The included subjects (n = 50) were 10 to 15 years old and in Tanner stages I–IV. The measurement error (the difference between measured and estimated VO_2max) was related to maturity in boys, but not in girls. The measurement error decreased for the whole group when the equation developed for women was used for the boys in Tanner I and II. This modification in the calculations of VO_2max resulted in an adjusted R^2 of 0.83 and SEE 0.23 L/min. Hence, the most accurate prediction of VO_2max from the EB test is generated if the test result is accompanied by ratings of sexual maturity in adolescents. Analysis of the test-retest values showed no significant change in estimated VO_2max from repeated tests within two weeks of each other.

In summary, the EB test proved to be a reliable and valid test throughout a wide range of ages (20 to 85 years) and fitness levels (1.33 to 3.94 L/min in women, and 1.67 to 5.97 L/min in men). The test was also found to be useful and reasonably valid for determination of VO_2max in pre-pubertal and pubertal adolescents, preferably after adjustment for sexual maturity status in boys. Furthermore, it was shown that the EB test captured fairly well an actual change in VO_2max during a period of 5 to 8 years. However, it is still unknown whether the test has an acceptable sensitivity for detection of a training-induced increase in VO_2max. Further studies are needed to evaluate if the test can be used in diseased individuals with or without different medications. The EB test can be used in health-related clinical settings, sports and fitness clubs.

LIST OF SCIENTIFIC PAPERS

I. Elin Ekblom-Bak, **Frida Björkman**, Maj-Lis Hellenius and Björn Ekblom. A new submaximal cycle ergometer test for prediction of VO_2max. *Scand J Med Sci Sports*. 24: 319-326, 2014.

II. **Frida Björkman**, Elin Ekblom-Bak, Örjan Ekblom and Björn Ekblom. Validity of the revised Ekblom Bak cycle ergometer test in adults. *Eur J Appl Physiol*. 116: 1627-1638, 2016.

III. **Frida Björkman**, Tony Bohman, Elin Ekblom-Bak, Örjan Ekblom. The ability of a submaximal cycle ergometer test to detect changes in VO_2max. Submitted manuscript.

IV. **Frida Björkman**, Andrea Eggers, Adam Stenman, Tony Bohman, Björn Ekblom and Örjan Ekblom. Sex and maturity status affected validity of a submaximal cycle test in adolescents. *Acta Paediatr*. Sep 19, 2017. Published ahead of print.

CONTENTS

ABBREVIATIONS

ANOVA	Analysis of variance
a-$\bar{v}$ O_2 diff	Arteriovenous oxygen difference
CV	Coefficient of variation
EB_{2012}	The Ekblom-Bak test with the original prediction equation
EB_{new}	The Ekblom-Bak test with the sex-specific prediction equations
IQR	Interquartile range
L/min	Liters per minute
mL/kg/min	Milliliters per kilogram of body mass per minute
HR	Heart rate
VCO_2	Carbon dioxide produced
VO_2	Oxygen uptake
V_E	Minute ventilation
Q	Cardiac output
RMR	Resting metabolic rate
RER	Respiratory exchange ratio
rpm	Revolutions per minute
SEE	Standard error of estimate
SV	Stroke volume
*v*max	Maximal velocity
VO_2max	Maximal oxygen uptake
VO_2peak	Peak oxygen uptake
W_{max}	Maximal watt production

1 INTRODUCTION

Human physical performance is determined by many factors. These factors can be summarised in three main domains: energy turnover, neuromuscular function and physiological function. The energy domain includes two central parts – aerobic and anaerobic energy turnover. The aerobic energy turnover includes the maximal aerobic power, mainly determined by the oxygen delivery from the central circulation and the oxygen turnover in muscles and other organs. The anaerobic (oxygen-independent) energy turnover is dominated by processes in the skeletal muscles.

The ability to perform a sustained physical exercise is, to a large extend, dependent on "aerobic fitness" (also known as "cardiorespiratory fitness"). Aerobic fitness refers to the capability of the human body to adjust to a physical work that has a duration of more than a few minutes. Thereby, aerobic fitness is an expression of the complete function of the pulmonary, circulatory and muscular systems, and the total energy turnover by the aerobic systems. Aerobic fitness is often assessed as maximal oxygen uptake (VO_2max). VO_2max is the highest obtained rate of oxygen (O_2) consumption during dynamic, high intensity exercise that involves large amounts of muscle mass (1, 2). In the scientific context, measurements of aerobic capacity have been implemented since the early 20th century.

The nomenclature in literature is often inconsistent with regard to the closely related concepts of VO_2max and VO_2peak. In the present thesis, the definition of VO_2max is the highest physiologically attainable value, whereas VO_2peak is the highest VO_2 observed under specific circumstances (3). Aerobic fitness is often discussed as a key factor for performance. VO_2max is a capacity measure in sports, but also an important parameter and lifestyle indicator in general healthcare. The increasing interest in different measurements of physical capacity and cardiorespiratory fitness in large populations, related to health in a broad sense, may be due to the more sedentary lifestyle and the arising consequences of modern behaviors. The knowledge of the importance of cardiorespiratory fitness is greater today than 50 years ago.

1.1 The concept of VO_2max

In the early 20th century, Hill & Lupton introduced the term "maximal oxygen intake" in sport science, through a series of experiment with measured O_2 uptake during different exercise modes. They described the occurrence of a "levelling off" in O_2 at maximal work rates, and clarified the relation between aerobic and anaerobic metabolism at maximal exercise (4, 5). In the 1920s, also Liljestrand determined O_2 uptake in athletes from a wide variety of sports, and more or less coined the concept of a “maximal oxygen intake” in human beings (6). However, it is unclear whether the true maximal value was achieved in these experiments.

Whole-body oxygen consumption (VO_2) is calculated as the product of cardiac output (Q) and arteriovenous O_2 difference ($a\text{-}\bar{v}\ O_2$ diff), where Q is the product of stroke volume (SV) and heart rate (HR). This principle is called the Fick equation, named after the cardiovascular physiologist Adolph Fick. The equation is:

$$VO_2 = Q \times a\text{-}\bar{v}\ O_2 \text{ diff}$$

or

$$VO_2 = HR \times SV \times a\text{-}\bar{v}\ O_2 \text{ diff}$$

VO_2max is a measure that varies considerably between individuals, due to differences in genetics, training history and exercise habits. More than 50% of the inter-individual variations in VO_2max are determined by heredity (7), and also the trainability of VO_2max includes a significant genetic component (8). It is relatively well accepted that there is an individual upper limit for VO_2max, and this individual “peak” is also highly determined by genetics. However, even if VO_2max has plateaued, improvements in performance can be achieved through other factors and physiological functions (for example adaptions in the skeletal muscles).

When it comes to physical performance, high aerobic capacity is important for athletes in many different sports. The highest VO_2max values are found in elite athletes in endurance sports. For example, rowing is a non-weight bearing activity where male athletes have absolute values well above 6.5 L/min, corresponding to a relative value of circa 70 mL/kg/min (9). The highest relative values (> 80 mL/kg/min) are found in athletes in weight bearing sports, such as cross-country skiing, long-distance running and cycling (10-12). Values for female elite athletes are generally 10–20% lower (12, 13).

In contrast to the elite athlete values, a brief overview of population-based data from three Scandinavian countries is presented in table 1. The untrained middle-aged men and women (i.e. 40–59-years olds below the 20^{th} percentile) in the study by Eriksen et al. had values as low as 31–33 and 25–27 mL/kg/min, respectively (14). It is not unusual that untrained and elderly individuals displays values around 20 mL/kg/min. Consequently, those individuals may reach as much as 85% of their total aerobic capacity while walking. Hence, their low VO_2max can significantly limit their activities of daily living. Furthermore, Ekblom et al. found that only 3% of the cohort had a VO_2max above 4.5 L/min (15). However, it is worth noticing that the VO_2max values in this stud>y were estimated from the Åstrand test (15). In contrast, the values in the study by Eriksen et al. were assessed from the results in a maximal cycle test (14), and Loe et al. conducted direct measurements of VO_2max from an incremental tests on treadmill (16).

Although there are some differences between the Scandinavian countries that are represented in table 1, it is evident that the VO_2max declines with age. The mean decrease in VO_2max is approximately ~7 % over a period of 10 years (14, 16). It is also worth mentioning that these results on fitness level in the Scandinavian countries are not directly comparable, due to a number of factors. With respect to the overall topic of this thesis, the main focus regarding the data collection in these three studies is *1)* the differences in test procedure: submaximal versus maximal tests, *2)* different work modes: cycling versus running, and *3)* the divergent methods for determination of VO_2max: estimation from the Åstrand nomogram, direct measurement, and assessment from a maximal exercise test. Furthermore, there are probably differences in the validity and reliability of the tests. The impact of the above mentioned sources of error, or influencing factors, will be discussed in the upcoming sections in the present thesis.

Table 1. *A brief overview of population based data of VO_2max values in men and women in different age-groups. Values are mean (SD), with exception for the values from Ekblom et al., that are expressed as median (95% CI).*

Study	Age	**Women**		**Men**	
		VO_2max		VO_2max	
		L/min	mL/kg//min	L/min	mL/kg/min
Ekblom 2007	20–29	2.6 (2.4–2.8)	40.7 (35.0–44.8)	3.3 (2.9–3.6)	39.5 (36.7–43.6)
	30–39	2.5 (2.3–2.7)	36.6 (32.5–42.2)	3.0 (2.6–3.2)	35.7 (31.6–40.9)
	40–49	2.4 (2.0–2.6)	34.4 (29.8–37.9)	2.5 (2.3–2.8)	30.4 (27.8–33.2)
	50–59	1.9 (1.8–2.1)	28.1 (25.1–31.1)	2.5 (2.3–2.7)	29.6 (27.9–31.2)
	60–65	1.7 (1.6–2.1)	25.0 (21.1–31.6)	2.2 (1.8–2.4)	26.6 (23.1–28.9)
Loe 2013	20–29	2.78 (0.46)	43.0 (7.7)	4.32 (0.71)	54.4 (8.4)
	30–39	2.75 (0.48)	40.0 (6.8)	4.22 (0.63)	49.1 (7.7)
	40–49	2.65 (0.44)	38.4 (6.9)	4.03 (0.61)	47.2 (7.7)
	50–59	2.36 (0.37)	34.4 (5.7)	3.65 (0.59)	42.6 (7.4)
	60–69	2.16 (0.33)	31.1 (5.1)	3.30 (0.55)	39.2 (6.7)
Eriksen 2013	18–29	-	35.6 (5.5)	-	43.4 (6.6)
	30–39	-	33.1 (5.7)	-	40.0 (6.5)
	40–49	-	32.1 (5.6)	-	38.9 (6.5)
	50–59	-	29.8 (5.1)	-	36.4 (6.2)
	60–69	-	26.5 (4.4)	-	33.2 (5.2)

Up until today, there is no consensus regarding the optimal training regimen in order to enhance aerobic fitness and VO_2max. Very untrained individuals can increase their

VO_2max with almost any aerobic exercise that lasts for at least 20–30 minutes. For experienced endurance athletes, there are higher demands on the intensity and duration of the exercise to achieve a further increase in VO_2max. One popular type of interval session is 4×4 min (with 4 min rest in between), where the accumulated time on very high intensity (90–100% of VO_2max) is 10–15 min (17). Another type of extremely high intensity exercise is 6–8×30 seconds, at a work rate corresponding to 150–250% of velocity VO_2max (*v*VO_2max, the lowest work rate that elicits VO_2max). These high intensity intervals has been shown to induce the same acute effect in mitochondrial gene expression as the more time-consuming session of 3×20 min at ~90% of *v*VO_2max (18). The optimal volume and amount of high intensity training for elite athletes, as well as the exact time and length of the intervals, is still not known (19, 20).

Apart from the obvious performance related aspects, VO_2max is also an important and independent predictor of cardiovascular health, longevity and mortality (21-25). For example, Blair and colleagues presented a number of interesting results from a longitudinal study, initiated in the 1970s. About 13000 men and women, with more than 8 years of follow-up, were categorised in groups of low, moderate, or high cardio-respiratory fitness, based on performance on a graded treadmill exercise test. Higher fitness appeared to delay all-cause mortality through lower rates of cardiovascular disease, also after corrections for confounding factors such as age, smoking, blood pressure, and cholesterol- and glucose levels. Blair et al. also noticed that the lowest acceptable value of aerobic capacity in order to avoid an increased risk for cardiovascular events and premature death was around 35 mL/kg/min in men and 32 mL/kg/min in women (21). However, a performance based test may be unsuitable in a mixed population, which is further discussed in section 1.3.1. Additionally, further studies are needed to verify these limits. Anyhow, if these limits are applied in the Swedish population (see table 1), it is apparent that around half of the population of middle-aged men probably have a fitness level that is low enough to generate negative consequences on the general health.

The independent relation of cardiorespiratory fitness to cardiovascular disease and all-cause mortality has been confirmed in several studies (22-24). Furthermore, also other health related variables are connected to fitness level, such as metabolic risk and type 2 diabetes (26, 27). In children, cardiorespiratory fitness is associated to cardiovascular and metabolic function (28-31) and mental health (32), and cognitive functions may be affected and correlated to VO_2max in both adults and children (33-36).

1.2 Measuring and determining VO_2max

A direct measurement of VO_2max is conducted via indirect calorimetry. In addition to VO_2, the direct measurements also include data on minute ventilation (V_E), and the values of VO_2 and VCO_2 gives an indirect measure of substrate utilisation.

1.2.1 Methods for measuring VO_2max

One of the earliest experiments with collection of expired air in air-sealed bags was performed by Prout in 1813 (37). Later in the 1800s, the German scientist Nathan Zuntz developed a device for ventilatory measurements in humans (38). However, it was in the early 20th century that ventilatory measurements in humans were conducted more frequently. The Douglas bag method (39) became the first, widely used, technique to measure expired air and thereby allowing calculation of the consumed oxygen during exercise. The method is still regarded to be the golden standard for measurements of steady-state respiratory gas exchange (40). However, the method includes manual handling to a great extent. For example, it is important that the start and stop of air sampling into the Douglas bags are exactly at the end of an expiration, and that the sampling time for each bag is carefully monitored and noted. This is easily done during rest and low work rates, but at maximal exercise, with sampling time < 30s and breath frequency > 90–100 breaths per minute, the observed volumes may differ 2-3% depending on whether the valve is closed during an inspiration or an expiration.

The expired air is most often collected in large, air-sealed bags (85–150 L), usually made from PVC material. The “Douglas bags” are slightly permeable to the external air, so it is important to perform the analysis of the gases in in close connection with the experiments. The total volume of the collected expired air is analysed in a spirometer. A small aliquot of the expired gas is analysed for temperature, pressure, and fractional concentrations of expired oxygen and expired carbon dioxide (41).

The Douglas bag method is time-consuming and requires advanced equipment and careful preparations. The collection of expired air and the concomitant gas analysis places high demands on the user, who has great control over the complete procedure. On the other hand, the Douglas bag method has limitations in sampling frequency, duration and measurement resolution. The apparatus can impose additional airway resistance because of resistance in the respiratory valves and long, narrow tubing into the Douglas bags. Especially in the early years of ventilatory measurements, this caused a restriction in air flow during maximal ventilations, which can be noted as considerably lower maximal V_E during maximal work in many studies. Since 1971, the WHO guidelines of air flow (resistance of respiratory valves < 5cmH_2O at a flow of 300 L/min, and tubing < 1.5cmH_2O at a flow of 300 L/min) have minimised these reactive effects (40). Furthermore, the collection of expired air only allows a total volume measurement, and breath-by-breath data cannot be obtained. Thereby, extremely rapid changes in ventilation or VO_2 cannot be studied (42). With regard to validity and reliability, repeated exercise testing have shown a test–retest variance of ~4–5% for measurements of VO_2max (40), and as low as 1.5 % for measurements of steady-state oxygen uptake (43). However, this only holds true in standardised laboratory settings, where all the equipment are carefully handled and operated by experienced physiologists and scientists.

Using the same principle of air sampling and gas analysis, the technique for indirect calorimetry has been modernised and computerised during the last decades (41). This development has resulted in portable and automated on-line gas analysis systems, and a markedly increased efficiency of the gas analysis procedure (42). Some of the modern metabolic measurement systems can also be used during outdoor activities (41).

The computerised metabolic systems have built-in volume- or flow sensing devices, as well as rapid and very accurate CO_2- and O_2-analysers. There are different types of flow-sensing spirometers used in automated metabolic systems, where the turbines are most commonly used. Some turbines have displayed linearity problems (caused by friction and inertia of the vane). These difficulties are seen at low flow rates ("lag-before-start") and at high flow rates ("spin-after-stop"), but these problems are relatively less important for measurements of total volume (41). With regard to the sampling of the expired air, some ergospirometry systems use mixing chambers, usually with a fixed volume of 5–8 L. It is also possible to conduct breath-by-breath measurements, which allows rapid and detailed analysis of respiratory gas exchange kinetics (44). However, the sampling interval can have a major impact on gas exchange data during exercise. Both the variability and the maximal values of VO_2 have been shown to be higher when shorter sampling intervals are used (45), implying that it is important to pay attention to the measurement resolution and time averaging technique in a VO_2max measurement. It has been suggested that a 15- to 20-s average or a 5- to 8-breath rolling average may be used, as they produce similar variability but allow a high degree of precision (41).

The accuracy and precision among the commercially available computerised ergospirometry systems varies considerably. A commonly used computerised ergospirometry system is the Oxycon Pro (Erich Jaeger GmbH, Hoechberg, Germany), with a reported CV of 1.2% compared to the Douglas bag technique (46). Carter & Jeukendrup studied the validity and reliability of three on-line systems (Oxycon Alpha, Oxycon Pro and Pulmolab EX670) with that of Douglas bags. They used both a metabolic simulator and human subjects, who were cycling at 100 or 150 W at three occasions. Oxycon Alpha and Douglas bags produced similar respiratory values over all levels of ventilation, while the Oxycon Pro tended to slightly overestimated values at the higher ventilations. The Pulmolab produced large overestimations at all ventilations for VCO_2, whilst values for V_E and VO_2 were slightly underestimated at higher ventilations (up to 7.5% from expectations). With regard to the reliability, the CV for VO_2 and VCO_2 measured using Douglas bags, Oxycon Pro and Oxycon Alpha were 3.3–5.1%, 4.7–7.0% and 4.5–6.3%, respectively, whilst that for the Pulmolab was highly variable (26.8–45.8%). The exercise tests supported the results from the study with the metabolic simulators, leading to the conclusion that Oxycon Pro and Oxycon Alpha are valid and reliable on-line systems for the measurement of parameters of respiration. This was later confirmed in a study by Rietjens et al., where twelve highly trained subjects performed an incremental cycle ergometer test. There were strong correlations between the Oxycon Pro and the Douglas bag measurements for of the ventilatory variables (r^2 for V_E = 0.996, VO_2 = 0.957 and VCO_2 = 0.98), and analysis of validity demonstrated high precision (47). In conclusion, Jaeger Oxycon Pro is an accurate system for measurements of metabolic parameters during low intensity as well as during maximal intensity exercise, requiring oxygen uptakes up to 6 L/min.

1.2.2 Choice of work mode

One important aspect in the VO_2max testing procedure is the choice of work mode. VO_2max can only be achieved during a dynamic whole-body exercise at sea level. It is widely accepted that the limiting factor for VO_2max is the ability of the cardiorespiratory system (i.e. the heart and blood) to transport O_2 to the muscles, not the O_2 uptake in muscle mitochondria (1, 4). The total O_2 uptake is dependent on the load on the muscles, as well as the mass of muscles involved in the work (48). Hence, a measurement of an individual's "true" VO_2max requires a work mode that involves a large muscle mass, in order to stress the central circulatory system to its maximum and avoid local/muscular fatigue before the physiological VO_2max is achieved. As mentioned before, VO_2max is the physiological peak in O_2 uptake, while VO_2peak not necessarily reflects a maximal value. VO_2peak is rather a limit to the subject's exercise tolerance during specific conditions (3).

Maximal work performed with only the arm muscles (arm-cycling) generates approximately 70% of the VO_2 uptake that can be reached with maximal leg work on a cycle ergometer, with inter-individual variations. The mathematical sum of the maximal VO_2 uptakes for the two exercise modalities is therefore often considerably higher than the actual "true" whole-body VO_2max (2, 48). The absence of a significant difference between the achieved VO_2max for running, ordinary cycling and the combined arm + leg-work, is a strong support for the theory that the capability of the heart muscle sets the upper limit for VO_2max (48). In most VO_2max testing situations, it is reasonable to attempt to get as close as possible to involvement of all parts of the body in order to reach a "true" maximal value. Some examples of work modes are specially designed "whole body" ergometer cycles, with combined arm-cranking and pedalling with the legs, cross-country skiing and running, which all produces similarly high VO_2max values (48). In a normal population, a highly demanding and technical activity, such as simultaneously arm- and leg cycling or skiing, is unsuitable for an exercise test. Therefore, the most commonly used work modes are running and cycling.

Cycling and horizontal running can generate approximately the same VO_2peak (49). However, it has been shown that a higher VO_2 can be achieved through uphill running ($\geq 3°$) compare to horizontal running on a treadmill (2, 12, 50-52). This aspect is important to consider when a maximal test protocol is designed, whereas a majority of the population (with exception of elite runners) are capable to reach their highest VO_2max if some incline is added to the protocol. Furthermore, it is worth noticing that uphill running produces about 6–7% higher maximal VO_2 values than cycling (53-55), so the correct term for a maximal value from cycling is ought to be VO_2peak.

Cycling is a popular work mode for maximal exercise testing, and tests on a cycle ergometer may be more suitable for subjects that are unaccustomed or unable to run/walk at high intensities. In some situations, local fatigue in the quadriceps may occur before the maximal stress of the cardiovascular and respiratory systems has been reached (55). This aspect is of particular importance if subjects are untrained and of less importance if subjects are familiar with high intensity cycling. Furthermore, the cycling can be performed with varying cadence, or pedaling rate, and a test can be conducted with a standardised or self-selected pedaling speed. There are great demands on the working muscle when a maximal effort is achieved with low cadence (i.e. low speed pedaling against a high resistance). In the early experiments with cycle ergometry, it was generally believed that a pedalling rate of 50–60 revolutions per minute (rpm) was the best cadence to use in tests of both energy efficiency on submaximal work rates, performance, and VO_2max (12, 56, 57). In 1967, Eckermann & Millahn came to the conclusion that the optimal pedalling rate was 45 rpm at power outputs of 100 and 150 W. Out of that, they even draw the conclusion that the choice of cadence when cycling should not be left to the subject, because subjects invariably chose to pedal too fast (58). Another example of maximal tests with rather low cadence is found in the experiments by Åstrand & Åstrand et al., where the maximal cycle tests were conducted with a pedalling rate of 50 or 60 rpm, "according to the subjects wishes" (59). The most economical pedalling frequency is probably somewhere in the range of 50–60 rpm (60, 61), and one study has shown that the most efficient pedal rate increased from 42 rpm at 40 W to about 60 rpm at 327 W (62). However, the optimal pedalling rate for achievement of the highest VO_2max is probably within the range of 80–90 rpm (63). With regard to power output, experienced cyclists seem to have higher optimal pedalling rates, and also higher self-selected pedalling rates, than unaccustomed subject (64, 65). These findings may be explained by the fact that the freely chosen cadence in unaccustomed subjects is more closely related to variables that minimise muscle strain and mechanical load than those associated with minimising metabolic economy (65). It has also been shown that elderly (~65 years old) cyclists often have a lower freely chosen cadence during a maximal test, compared to younger (~25 years old) cyclists (66). Possible explanations for the divergent results from previous studies are the wide variety in cycling experience in the tested subjects (for example, untrained compared to elite-cyclists, young or old subjects, etc.) and in outcome measures (peak power, time to exhaustion, efficiency or achievement of VO_2max). The research on the effect of pedalling rate on different test protocols, and the influence on measurements of VO_2max, is far from complete.

In certain groups of subjects, the choice of work mode may be based on sport specificity (67). Among elite athletes, a high level of sport specific training may enhance the body's ability to perform a maximal effort in other work modes than for example inclined running. Therefore, the specificity in the choice of work mode is a more important issue for athletes than for the general population. Furthermore, the specificity in testing is of great importance when VO_2max/VO_2peak is measured as a performance indicator, and/or when the measurement is used for evaluation of training programmes. In athletes, other work modes (for example rowing or cross country skiing) may generate VO_2max value that not significantly differs from the values achieved during running, while a normal person may experience technical difficulties in the execution of the activity itself. However, the ability to generate the highest VO_2max values from sport specific activities in elite athletes only applies to work modes that involve a reasonable amount of working muscles, at ground level. Other work modalities, for example swimming, does not generate enough stress on the cardiovascular and circulatory systems to produce a true, whole-body VO_2max value (48), not even in highly trained swimmers. For example, it has been reported that elite swimmers have a mean VO_2peak during flume swimming that corresponds to 94% of the VO_2max achieved during running (68).

1.2.3 Test procedures and test protocols

Another important aspect in maximal exercise testing is the preparation procedure before the maximal effort, and the design of the test protocol. It is not possible to determine a specific procedure that generate the most accurate VO_2max values, due to different aims, conditions and physical status of the tested subjects. However, some guidelines have been purposed to optimise the possibility to reach a true VO_2max via a continuous graded exercise test. One example is found in a review by Howley et al. These guidelines include five minutes of warm-up at 65–70% of VO_2max, followed by a brief rest. The recommended intensity for test protocol is a starting level at 60–70% of VO_2max, with an increase in load of approximately 5% of VO_2max every minute (which equals a total test time of about 8–10 min). Furthermore, the gas collection time (for example 30 s) and criteria for verification of VO_2max shall be stated before test (69). For details about measurement and sampling, see section 1.2.1, and VO_2max criteria are described in section 1.2.5.

The function of “warm-up” before a maximal exercise test was highlighted in the 1930s. In an early study by Nielsen & Hansen, two young male subjects reached their highest VO_2 values with a test protocol comprised of five minutes of “warm-up” at a rather high intensity (285 W) immediately followed by an all-out effort at 333 W (70). Some decades later, Per-Olof Åstrand conducted several maximal exercise tests for his thesis “Experimental studies of physical working capacity in relation to sex and age” (49). However, the procedure that preceded the maximal tests are rather insufficiently described. In just a couple of sentences, the author states that “for several reasons, it is desirable that experiments with a very high working intensity are not started directly from the resting state; the subject ought to have some “conditioning” activity first” (49). Thereafter, it is concluded that the experiments by Nielsen & Hansen (70) advocate the use of a lower work intensity before a maximal work is performed, but it is also stated that “in the writer’s experiments, the conditioning work had a lower intensity compared to the work of Nielsen and Hansen” (49). From that description, one might conclude that the subjects performed five minutes of conditioning activity at an intensity below 285 W. In later studies, Irma Åstrand conducted numerous maximal exercise tests that were preceded by 2–3 submaximal levels (6 min at each stage). The submaximal loads clearly constitutes a conditioning activity, however, the resting periods between the submaximal and maximal loads are not clearly described (71). Others have used less than five minutes of unloaded pedalling before the start of a VO_2max test on a cycle ergometer, with rapid and frequent increases in work rate, and an all-out effort > 5 min (53). Four minutes of unloaded cycling is a rather sparse warm-up prior to an exercise effort of that intensity and magnitude.

The fundamental meaning of a “pre-conditioning” exercise phase before a maximal effort is linked to VO_2 kinetics, i.e. the slope and timing of the increased VO_2 consumption at the start of physical work. The VO_2 kinetics are an important aspect to take into account to be able to create properly designed test protocols. During the transition from resting state to heavy exercise, the increase in VO_2 uptake can be divided into three phases (72). In phase 1 (initial component), there is an increase in blood flow through the alveoli, and the increase in VO_2 primarily reflects changes in venous O_2 stores. Phase 1 has different characteristics depending upon whether the work is started from a resting state (results in an abrupt response) or from a mild baseline activity (which results in a relatively slow response). In phase 2 (primary component), the further increase in VO_2 is arising from increased tissue oxidation. In phase 3 (slow component) the VO_2 reaches a steady-state condition, i.e. the O_2 supply is sufficiently corresponding to the O_2 demands from the working body (72, 73). The steady-state condition is mostly pronounced at moderate intensity (below the lactate threshold). During more intense or maximal exercise, VO_2 continues to rise for several minutes until either a delayed steady-state is achieved, or exercise is ended (73, 74). Furthermore, during extremely heavy exercise an additional slow component of $\dot{V}O_2$ is superimposed upon the fast component response such that $\dot{V}O_2$ rises above the expected $\dot{V}O_2$ requirement for the external workload (75).

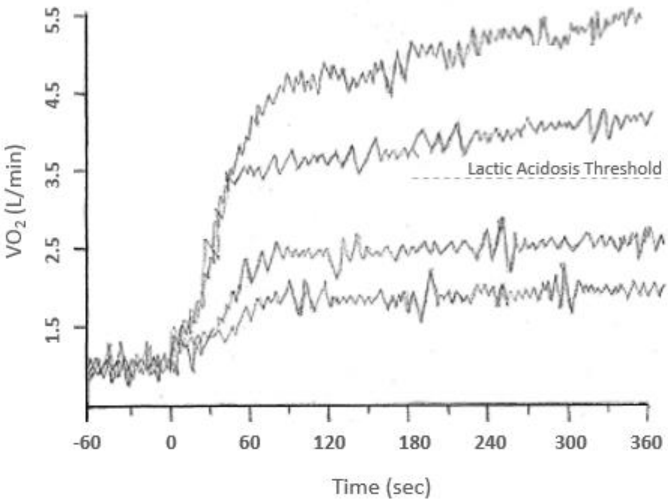

Figure 1. *Schematic illustration of the VO_2 response in one subject at four different work rates. Dashed line denotes the lactic acidosis threshold. The figure is based on data from Barstow et al. (73).*

Figure 1 above shows the VO_2 responses for four work rates in one subject. The work starts at 0 sec. Phase 1 is clearly visible for each exercise intensity. The two lower curves (work rates below the lactate threshold) shows a rapid attainment of steady state, and phase 3 is achieved within 120 seconds. The two upper lines shows a continuous rise in VO_2 and a delayed or absent steady state, i.e. there is an continuing upward trend of the VO_2 (73).

A majority of previous research on VO_2 kinetics have been conducted during cycle ergometer exercise. However, the very detailed aspects of VO_2 kinetics are somewhat different in running compared to cycling, and the VO_2 response to other work modes are not directly comparable to the kinetics while cycling. For example, the time constant (τ) for the phase 2 pulmonary VO_2 response to exercise tends to be shorter and the amplitude of the VO_2 slow component tends to be smaller, in running compared to cycling exercise (76, 77). Possible explanation for these findings can be related to the greater component of eccentric muscle action during running, which facilitates the venous return of the blood to the heart. Furthermore, the faster VO_2 kinetics in phase 2 in running compared to cycling might be linked to greater habitual use of the ambulatory musculature (78).

In order to achieve a true VO_2max value, i.e. stress the cardiovascular and respiratory functions to its very upper limit, the test has to be designed with great caution. For example, it is important to choose a suitable starting intensity and decide the frequency and load for the increases in work rate for a graded exercise test and thereby a suitable test duration. There are a number of different categories of test designs, primarily the one load tests, continuous test (graded exercise test with increments every 15, 30 or 60 s or non-stop ramp tests with 2–5 min stages) and discontinuous tests (various increment and stages with short breaks in between). A test of maximal exercise capacity with simultaneous measurement of VO_2max is almost always continued until subject's voluntary exhaustion. Another alternative – which is relatively uncommon in the context of scientific and athletic exercise testing – is that the test is terminated when specific pre-determined criteria for maximal effort have been fulfilled.

In the early experiments by Åstrand, all subjects performed repeated tests on one fixed work rate. The initial work rates were of submaximal character, and the procedure was repeated within a couple of days with an increase in work rate of 1–2 km/h or ~50 W. The experiments continued until the work rate was high enough to fatigue the subject within four to six minutes. All running tests were conducted with an elevation of 1° for adult subjects, while subjects younger than 20 years old ran on a horizontal treadmill. The protocols for cycling were performed with a pedalling rate of 50 rpm, with exception for the final period during the maximal effort, where subjects sometimes failed to keep the cadence according to the metronome but where verbally encouraged to keep up the highest cadence possible (49). A similar, standardised procedure for treadmill tests has been proposed by Taylor et al. (50).

In 1968, Shepard et al. recommended a continuous test with 2-min increments of work rate, beginning at 90–100% of predicted VO_2max. This test protocol had a shorter duration but produced the same VO_2max as a discontinuous test procedure (55). However, up to the early 1980s, the use of short duration (less than 7 min) test protocols was relatively sparse. In 1981, Whipp et al. conducted several cycle tests with different test protocols: a constant-load test, a 5-min incremental stage test (25 W/increment per stage), a 1-min incremental test (15 W/increment) and a continuous ramp test (slope = 50 W/min). The shortest test duration was present in the ramp test (4–8 min before an all-out effort was achieved), while no significant difference between the protocols were demonstrated for VO_2max (79). Later, Buchfuhrer and colleagues studied various work rate increments in cycle and treadmill tests. They found that the highest VO_2max was achieved with moderate increase in intensity (30 W/min in trained men). Test with greater increases in work rate gave a rather short total work time (7–8 min), and resulted in significantly lower maximal VO_2 value compared to tests with a duration of 8–17 min. The recommendation from the authors was to select a work rate increment that generates ~10 min of work before exhaustion occurs (54).

Another study examined the frequency in the increases in exercise intensity, with protocols that comprised of increases in work rate (16.3 W or 0.8 km/h) every 15 s or every minute. The short duration protocol (15 s increases) resulted in test time < 5 min, while the latter protocol yielded mean test times around 15 min. There were no significant differences between the protocols with regard to VO_2max, maximal V_E, respiratory rate, tidal volume, oxygen pulse, and peak expiratory flow rate (53). Later, Davis et al. studied five different test protocols with a total test duration of 6–12 min, and found no significant difference for the achieved maximal values for VO_2, HR and V_E (80). In 1987, Levler et al. presented similar values of VO_2max from a discontinuous test (started at a work rate of 70 W, and increased by 35 W every second minute, with two minutes of rest between stages), a continuous test (started at a work rate of 70 W with increase in work rate by 35 W/min), and "a jump-max test" (JMT). However, the discontinuous test was performed during 29.3 ± 1.54 min and the JMT was performed during 9.1 ± 0.32 min (81). Finally, even a protocol with 2-min stages of gradually increasing self-selected paces has been shown to produce the same VO_2max as a "traditional" continuous protocol in recreationally trained men (82).

As evident from the information above, many different procedures and protocol can be used for a VO_2max test. The standardised tests protocols can be useful in homogenous groups and when there are high demands on reproducibility. In other situations, for example exercise testing in very heterogeneous populations, protocols with individually determined exercise levels might be more useful. With respect to the design of test protocol, it is important to decide if the test shall start on a relatively high or a lower intensity (depending on preceding conditioning activity and the motivation and endurance of the subject). There are certainly people that are able to achieve their VO_2max value with almost no preceding activity, but most people are capable of higher maximal efforts when the hard work is performed after some preparatory activity. Furthermore, it is crucial to choose an appropriate rate of increase in intensity, with small or large steps in the workload increments. Furthermore, physiological factors like age, training status, sport specificity, muscle mass, and muscle fibre type may influence the VO_2 kinetics, and thereby also affect the choice of procedure and test protocol.

1.2.4 Verification tests

An additional procedure to verify that a true VO_2max has been reached is to conduct a verification test. This test can be conducted on a separate day, or at the same day as the first VO_2max test. The theory behind the verification tests is to stress the body at a rate of work that is higher than the highest level attained in the first maximal test. That level above VO_2max is often referred to as "supramaximal". The supramaximal work rate can verify if the VO_2 consumption can increase even further and reach a higher level than produced at the initial maximal test.

Earlier studies in adults have investigated the concept of a supramaximal test as verification test. The supramaximal tests for these subjects have been conducted in the forms of an all-out effort on a higher work rate than the subjects previously had achieved in a graded maximal exercise test (83-86). No mean difference have been found between an incremented VO_2max test and a supramaximal test conducted on two separate days (83, 84), within hours (83, 86) or even minutes (85). Most studies show no significant mean difference between an incremental VO_2max test and a supramaximal verification test. However, some subjects attain higher VO_2 values in a verification test compared to an incremented VO_2max test (85), which may reflect that the first incremental test did not elicit a true VO_2max value. For example, Scharhag-Rosenberger et al. studied an intermittent incremental protocol, where 3 min running (starting at 6 or 8 km/h, with increments of 2 km/h) were performed to the subject's voluntary exhaustion. The verification test was performed 10 min after the incremental test, and comprised of 1 min running at 60% of the maximal velocity (*v*max), followed by running at 110% of *v*max until voluntary exhaustion. The 110%-effort was meant to elicit an exercise-time of at least 2 min. If the subject reached a VO_2peak > 5.5% higher than the incremental test value, another 10 min rest period was followed by an additional supramaximal run at 115% of *v*max. Furthermore, all subject performed a second supramaximal test on a separate day, where 5 min warm up was followed by the same supramaximal protocol. Results showed that 34 out of 40 subjects satisfied the verification criterion on the first test day, while 15% achieved a higher VO_2 in the verification test. In these subjects, the higher VO_2max from the verification test was also confirmed on the second test day (85).

The findings of equivalent incremental and supramaximal VO_2max in adults are also present in studies with children. Among the previous research of verification tests in children, some of the studies have separated the incremental test and the supramaximal test with a week (87, 88), one study used maximal work on a cycle ergometer to define VO_2peak (89), and one study evaluated the question in children with spina bifida (90). Rowland conducted a study in adolescents (10–13 years old), who performed four tests with a week in between each test. The first test was a progressive treadmill test with a modified Bruce protocol, with an initial speed of 8 km/h and 2.5% elevation. Three out of nine subjects demonstrated a plateau in VO_2max during this test. The supramaximal tests started off with 3 min at the same first level as the progressive test, followed by a work rate at 2.5%, 5%, and 7.5% higher than the final load in the first test, respectively. The supramaximal tests failed to further increase the mean VO_2peak above the values from the progressive test. Hence, the author suggested that the achieved VO_2 values from the progressive test reflected a "true" maximal value (88).

There are no studies in children, up to date, where a supramaximal test on a treadmill has been conducted a few minutes after an incremental VO_2max test. There is an ongoing debate whether or not VO_2max can be assessed in children, and further studies of different verification test procedures are needed to increase the understanding of children's maximal capacity. With regard to methodology, it may be unsuitable to use the Douglas bag method for measurements of VO_2max in youths, due to their relatively low V_E and infrequent occurrence of a plateau in VO_2. The method requires a bag-filling of at least ~50 L of expired air for an accurate determination of volume and gas concentrations. This might be difficult to achieve in those children and adolescents that not produces any pronounced levelling-off at VO_2max.

1.2.5 Criteria for determination of VO_2max or VO_2peak

Lastly, one central and critical issue of concern in the present thesis is the criteria for acceptance of VO_2max. How do you know that someone has reached the final upper limit of VO_2 consumption? How to decide what is a true maximal effort? Throughout the history of maximal exercise testing, researchers have used a wide variety of criteria for determination of VO_2max (69).

The strongest and most used criteria is the presence of a "plateau" in VO_2max, i.e. a levelling off in VO_2 uptake despite an increase in work rate. This concept and criteria is fist mentioned by Hill & Lupton in the early 1900s (4, 5). A commonly used limit for levelling off is a change in VO_2 (ΔVO_2) at VO_2max ≤ 150 mL/min. However, this limit was originally developed from discontinuous treadmill testing (91), and the VO_2 measurements were done with the Douglas bag method. The development of the VO_2max test procedure and protocols in the later years has initiated a debate about the concept of "levelling off" (92). Not all subjects demonstrate a plateau in VO_2 at the end of a continuous exercise test. In previous studies, the occurrence of a plateau in VO_2max has been reported to be anywhere around 90–100% and as low as ≤ 50% (69). Also, the technological development, with higher measurement resolution, has resulted in a situation where researchers sometimes use more stringent plateau criteria, for example ΔVO_2 at VO_2max < 50 mL/min (50). However, a plateau might be absent in a maximal test, but this does not necessarily mean that the subject has failed to reach a "true" VO_2max (93, 94), since it is possible that the work was discontinued at the point of VO_2max.

The ability to detect a plateau in VO_2max is influenced by the study design and choice of test protocol. Duncan and co-workers found that a plateau could be detected with similar incidence in continuous (50%) and discontinuous (60%) test protocols (93). During a discontinuous test, it is generally accepted that a subject has to complete 3–5 min at each stage in order to achieve sufficient measurements of VO_2 uptake. In some cases, a subject may reach VO_2max during the second minute on a supramaximal stage. If fatigue occurs in less than 3–4 min, this data point would not be used in the graph of VO_2 versus work rate. Consequently, a plateau will be absent although VO_2max has been reached (1).

Furthermore, the appearance of a levelling off in VO_2 versus work rate is influenced by measurement resolution, sampling time and the averaging procedure of the collected VO_2 values. It has been shown that shorter sampling intervals (breath-by-breath with 11-breath moving average, and time-averaged into 15 s) results in a plateau in all subjects, while averaging over 30 s and one minute only resulted in a plateau in 57% and

8% of the subjects, respectively (95). In contrast, the high resolution and sampling frequency has also permitted detailed analysis of the slope of the change in VO_2 with a consistent increase in work rate. The slope of that change has considerable variability, supporting the argument that a plateau in VO_2max (defined as the slope of a VO_2 sample at peak exercise that does not differ significantly from a slope of zero) is not a solely reliable marker for maximal effort (45).

The meaning of the levelling off-criteria also has to be judged for the specific test situation. For example, elderly and people who are unaccustomed to intense exercise experiences more difficulties in achieving a levelling off than well-trained athletes (69). Also, pre-pubertal children have limited ability to achieve a plateau in VO_2max, for several reasons (87, 96) – see further in section 1.6.3. As mentioned before, the term "VO_2peak" can be used when there are any doubts concerning the achievement of a true maximal value. This term is frequently used in studies with children.

With respect to all of the above mentioned reasons, the plateau in VO_2 should not be used as the exclusive criterion for achievement of VO_2max. Instead, a number of supporting criteria have been purposed in order to verify a maximal effort. The maximal strain can be indicated through psychological criteria (an acceptably high rate of perceived exertion) and the maximal exercise can be verified with different circulatory criteria, such as achievement of a certain peak HR and ventilation (1, 69). It is also common to use metabolic criteria, which include a respiratory exchange ratio (RER) >1.15 and blood lactic acid level >8–9 mM (49). Also the secondary criteria have to be chosen and modified to suit the subjects that are tested, since many values in different physiological variables differs among children and adults, and young and old people (97).

1.3 Tests for estimation of VO_2max

A direct measurement of VO_2max is time consuming, expensive and requires a laboratory environment. Instead, there are a several tests for indirect estimation of VO_2max.

1.3.1 Maximal tests with estimation of VO_2max based on performance results

Among those tests, the maximal exercise tests can be used in situations where indirect calorimetry is missing, but subjects are healthy and capable of intense exercise. These tests are based on statistical correlations between, for example, time to cover a certain distance (walking/running) and VO_2max. These relationships can be relatively strong, but it is important to take into account that these tests are largely influenced by capacities other than VO_2max, for example anaerobic processes, motivation, tactics, previous experiences (learning effect), running economy. Those capacities can easily be enhanced with specific training, which produces significant increases in test performance without any actual change in VO_2max (98).

However, the tests are easy to set up and administer for a large group of people at the same time, the tests are commonly used in sport- and school settings. Popular tests are the 1-mile test, the Cooper test, and the Yo-Yo Intermittent Recovery test (Yo-Yo IR test). The tests ability to predict actual VO_2max have varying accuracy, normally around r = 0.50 to 0.70 (98-101). The VO_2max from the Yo-Yo test can be estimated with the following formulas (98):

Yo-Yo IR1 test:
VO_2max (mL/kg/min) = IR1 distance (m) × 0.0084 + 36.4
Yo-Yo IR2 test:
VO_2max (mL/kg/min) = IR2 distance (m) × 0.0136 + 45.3

As reported by Bangsbo et al., the strong influence of other capacities than VO_2max result in a situation where two subjects with a measured VO_2max of 53 mL/kg/min can have a range of performance from 1450 to 2600m in the Yo-Yo IR1 test (98).

The Balke test protocol (102) was originally developed for testing of physical fitness in military personnel. The protocol is sometimes used with some modifications, but the standard procedure for men is a constant speed of 5.28 km/h (3.3 mph), with the starting level at 0% incline. After the first minute, the incline is set to 2% for one minute, and thereafter the incline is increased by 1% every minute. If the subject manage to exercise for 25 min, the incline is maintained at 25% and the speed is increased with 0.32 km/h (0.2 mph) until voluntary exhaustion. For women the treadmill speed is set at 4.83 km/h (3.0 mph), starting at 0% inclination, and increased by 2.5% every third minute. The test is a performance test, resulting in a test score based on test time in minutes. However, the test time can also be used in prediction equations for estimation of VO_2max (103, 104). Another exercise stress test is the Bruce protocol, originally developed for evaluation of patients with coronary heart disease. The test is started at a low speed and 10% incline, and the intensity is increased in consecutive 3-min stages until subjects voluntary exhaustion (105). Washburn & Montoye examined the possibility do determine VO_2max from a treadmill protocol in boys and men, 10–39 years old. Regressions equations for each subject were constructed from submaximal HR and VO_2 uptake on the submaximal treadmill stages. The regression equation were extrapolated to the subject's estimated maximal HR. The VO_2max prediction equations by Margarita (106), Maritz (107), and Åstrand-Ryhming nomogram with the age correcting factors (108, 109), was used for estimation of VO_2max. The subjects were divided in six age-groups, from the ages of 10–14 and up to 35–39 years old. The correlation between measured and estimated VO_2max varied among the age-groups, from r = 0.50 to r = 0.84. All methods generated rather poor individual predictions, as indicated by large standard deviations; from 0.28 to 0.53 L/min. None of the methods for estimation of VO_2max provided superior validity across all age groups, but all methods over-estimated VO_2max in 10 to 14-year old boys (110).

All maximal tests, with either direct measurements or estimations of VO_2max, triggers maximal physical strain in the subject. These tests are thereby difficult and inappropriate to conduct in a large proportion of the normal mixed population. The demand of a fully maximal effort is unsuitable or unachievable in many situations, such as when testing older people, patients with orthopedic diagnoses, obesity or people unaccustomed to running or other intense whole-body exercises. The application of these tests in clinical use or in large population-based surveys, may therefore result in a drop-out of individuals who are un-fit, un-experienced, unsecure or uncomfortable with the maximal exhaustion. In accordance, there is a risk for under-achievement in relation to true VO_2max, which in turn may result in a systematic bias in the collected data. It is also worth mentioning that performance is a main outcome from all maximal tests. Hence, the value depends on several other capacities besides VO_2max, including anaerobic capacity, motivation, tactics, endurance and experience. The physical performance, or work tolerance, is an interesting measure, but it is not an equivalent to VO_2max.

1.3.2 Submaximal tests with estimation of VO_2max based on HR response

An estimation of VO_2max can also be done from submaximal work, which may be a suitable option in situations where the maximal tests are inappropriate. During submaximal conditions, a higher work efficiency in the heart muscle is achieved through full SV and lower HR, rather than low SV and higher HR. Complete, or almost complete SV, is reached around 50% of VO_2max (3, 111). Thereafter, the SV remains almost unchanged through higher work rates while HR increases proportionally, in order to meet the oxygen demands of the working muscles.

The estimations of VO_2max from submaximal exercise testing are often based on the linear relationship between steady-state HR at a given work rate and an estimated maximal HR/VO_2 value. It is well-known that subjects with a high aerobic capacity (high VO_2max) has a lower HR at a fixed submaximal rate of work than those with a lower aerobic capacity. This is a consequence of the fact that the cardiovascular system and its components adapt to an increase, or a decrease, in aerobic capacity. The main regulatory changes are found in cardiac muscle (increases in heart size and contractile function lead to decreases in submaximal HR), accompanied with subsequent changes in SV, Q, blood and blood flow. Hence, the function of the HR based tests are linked to the integrated result of all of the above mentioned physiological events that influences the regulation of submaximal HR responses. The observed value of a HR (beats/min) is the “end point” in a series of adaptations of the circulatory system. The number of heart beats is linked to the size of the cardiac muscle, as well as blood volume and autonomic regulation. The body has several ways to adapt to aerobic exercise over a prolonged period of time, with the overall goal to enhance the ability to transport oxygen to the working muscles.

It is known that endurance training will induce hypertrophy of the heart muscle, so that the heart will be stronger and capable to deliver larger amounts of blood with each heartbeat (112). In addition, endurance training also increases blood volume. An increase in the production of red blood cells is accompanied by a significantly greater increase in plasma volume. Hence, haematocrit and the concentration of haemoglobin remains unchanged, or is slightly lowered. Irrespectively of that, an enlarged blood volume contributes to a higher Q, since Q = SV × HR. An absolute Q can be achieved with a lower number of beats/min, if the SV is increased. Consequently, a well-trained heart has the capacity to deliver the same Q with a lower HR, compared to the heart of an untrained person. These physiological adaptions of the circulatory system after a period of aerobic training leads to a decreased HR in the resting state, as well as during submaximal exercise (113).

The reason for the different aerobic capacity of an untrained and a well-trained individual is mainly explained by the higher SV in the well-trained person – both at rest and during submaximal and maximal exercise. An untrained individual have a SV of approximately 50–60 mL/beat at rest, and a maximal SV of 90-110 mL/beat. The corresponding values for an endurance athlete at elite level are 90–110 mL/beat at rest and a maximum of 190-200 mL/beat, respectively (114)

The maximal HR is largely unrelated to training status. Both an elite athlete and an untrained adult can have a maximal HR of for example 195 beats/min. However, the maximal Q for the athlete, at a HR of 195 beats/min, might be ~37 L/min (195×190 mL = 37 L/min). In comparison, an untrained individual will probably have a Q of approximately 20 L/min at the same absolute HR (195×105 mL = 20 L/min.). Consequently, the heart of a very well-trained athlete can manage to deliver more than 35 L of oxygenated blood to the working muscles each minute (115). On the other hand, the maximal capacity of an untrained person can be as low as ~20 L/min. This is a main explanatory factor to the difference in aerobic fitness between untrained and well-trained individuals, and it explains the training-related differences in HR response to a given submaximal exercise. The increased capacity from each heartbeat is furthermore utilized in a more efficient way, through local adaptions in the peripheral structures (i.e. enhanced ability for the muscles to use the delivered oxygen).

The control of the heartbeats is almost completely involuntary regulated. The baseline heart rate rhythm is initiated in the sinoatrial node (SA node), a group of cells in the wall of the right atrium of the heart (116). These cells have the ability to spontaneously produce an electrical impulse (action potential) that causes the contraction of the heart muscle (i.e. an activity in the heart muscle that ultimately leads to a heartbeat). The SA node is under influence of both parasympathetic and sympathetic postganglionic fibres.

The HR frequency is around 100 beats/min when the SA node is unaffected by any nervous or hormonal influences. Hence, in the resting state, there is considerably more parasympathetic stimuli to the SA node, as evident by the observation that the resting HR usually is somewhere between 50–70 beats/min. When the demands on the circulatory systems increases, as a consequence of increased physical work, the HR is up-regulated via intensified sympathetic stimuli and a down-regulated parasympathetic drive. This has been experimentally shown in a study by Ekblom et al. in 1972 (117). The neurons in connection with the SA node is acting via release of different neurotransmitters: acetylcholine (a parasympathetic neurotransmitter) and norepinephrine (a sympathetic neurotransmitter). The HR is also regulated via circulating hormones, for example epinephrine, that acts on the same beta-adrenergic receptors in the heart as the norepinephrine released from neurons.

Moreover, the HR is affected by fluctuations in different physiological functions during the day and night, variations known as circadian rhythm. Previous reports on the influence of circadian rhythm on submaximal and maximal cycle exercise is somewhat inconsistent. However, the presence of a day and night-effect on HR has been reported (118, 119). Furthermore, it has been reported that the ΔVO_2 for a given W and V_E remained unchanged throughout all times of the day and night, and that the influence of the day and night variations seem to be related to the standardization of preparations for the subject (120). Hence, the variation from the circadian rhythm can possibly be minimized with rigid standardization before test. It has also been reported that there is no influence of the circadian rhythm on gross efficiency as long as the exercise is conducted at steady-state condition below a RER of 1.0 (121).

With regard to work mode, the HR based submaximal tests are often executed in cycling (discussed in detail in section 1.4), walking (122), or by stepping up and down a bench at a fixed rate for a few minutes. Some examples of step tests are the Harvard step test (123), the Chester step test (124), and a modified Harvard step test (49, 125) with estimation of VO_2max from the Åstrand-Ryhming nomogram (108). The different step tests have varying correlation to directly measured VO_2max, with values ranging from r = 0.47 to r = 0.92 (126). The prediction errors in step tests may be partly explained by the inter-individual variations in work efficiency, as well as influence of body mass and body composition.

Taken together, most of the submaximal tests are relatively easy to conduct, without any expensive or complicated equipment. A large proportion of the population can perform the requested work, and the tests are free from performance based and competitive aspects.

1.4 Submaximal cycle ergometer tests

In cycling, the weight-bearing demands on the body are lower than for walking and stepping, and the variations in work efficiency are relatively small. This has resulted in development of submaximal exercise tests for estimation of VO_2max. The most precise assumptions are generated through cycle ergometry, since that exercise mode is less dependent on factors like, for example, athletic skills and coordination. Thereby, cycling has relatively small inter-individual variations in work efficiency (127, 128) and the mechanical efficiency is rather constant in mixed populations, even in populations that have never used or even seen a bicycle or cycle ergometer before (128). Furthermore, the work mode is easily performed by people with overweight or some kind of limitations in the capability to walk, step or run (*i.e.*, weight baring activities). The above mentioned advantages for cycling has led to the development of a number of submaximal cycle ergometer tests for estimation of VO_2max. Those tests have been developed based on the assumption of a rather linear relationship between HR and power output (PO) up to maximum, and that VO_2 can be estimated from PO with acceptable precision (108). Thereby, it is also possible to estimate VO_2 from HR during submaximal cycling.

1.4.1 Multi-staged tests

A number of previously described cycle tests are multistage tests with or without breaks in between stages. One of the early described cycle tests is the physical working capacity (PWC) test, or “Sjöstrand test”, from 1948 (129). At this point, researches were aware of the physiological bases of HR as a relatively linear function of VO_2 and workload. By the time of the construction of the PWC test, in the 1940s, it was generally believed that the cardiac output was at its peak around a HR of 170–180 beats/min. This finding was based on the observations that the highest load that could be applied without any signs of insufficiency of the respiratory or circulatory organs usually corresponded to a HR of 170–180 beats/min in adults. A HR of about 180 beats/min was thereby determined to be “the critical pulse level” (127, 129), and this level was thereby considered to be the upper limit for the body to work with adequate supply from the circulatory and respiratory organs.

The original test was conducted in three consecutive 10-min stages at 300, 600, 900 (and in some cases 1200 kpm) kpm, with a pedalling rate of 50 rpm. HR was noted at minute 2 and 8 of each work rate, respectively, except for some of the highest loads that exhausted the some of the subjects in only 4-6 min. If a subject exercised on a work rate that resulted in an increase in HR above 10 beats/min between minute two and eight, or had reached an absolute HR > 175 beats/min, this workload was regarded to be the level for in-sufficient respiratory and circulatory response, i.e., the previously completed level was the highest tolerable work intensity (129). The multiple workloads are described to

”/…/ ascertain the load at which the pulse rate starts rising rapidly or the pulse level reaches or exceeds the critical value where the heart minute volume decreases with rising pulse frequency” (129). In other words, the test is roughly designed to find the work load that no longer result in a steady-state HR for ~10 min of exercise.

The experiments by Sjöstrand were conducted in 20 middle-aged men with respiratory conditions. In later experiments of physical working capacity, Wahlund studied a more heterogeneous group of subjects: 189 work-men at an ore smelting work (with a high incidence of respiratory conditions), 188 men on military services (with experiences of heart- or respiratory problems), and 26 healthy subjects (127). All subjects were men, with a rather wide range with regard to age (19-66 years old) and training status (including cardiac patients and athletic subjects). The previously mentioned observations of a “critical pulse level” in HR (129) also led to the conclusion that the rate limiting factor for cardiac output was the time of diastole, which was believed to be too short for adequate filling of the heart when HR reached beyond 170–180 beats/min (127). Hence, by studying the individual HR response to physical work and find “the critical pulse level”, an estimate of the maximal cardiac output could be done. In the experiments by Wahlund, this upper limit was set to a HR of 170 beats/min (PWC_{170}), and work rates that resulted in a HR above 170 was regarded to be an “overload” (127).

The test procedure was a modification of the test conducted by Sjöstrand; the standardised protocol started at 300 or 600 kpm/min, with increase in work rate by 300 kpm/min. The time for each stage was 6.5 min, and the cadence was set to ~60 rpm. In the experimental procedure, all subjects were exercised to exhaustion or until completion of the workload of 1200 kmp/min. Furthermore, it is mentioned that if the chosen work rates in a test fails to elicit a HR of 170, this point could be estimated from the individual relationship between HR and work rate. Wahlund concludes that a test with series of loads produces more detailed information of an individual’s working capacity, compared to using only one workload. Furthermore, it is also stated that “maximal working capacity for practical use should be defined as the maximum working intensity consistent with steady state” (127).

The procedure of determination of the exact individual correlation between HR and work rate has led to further developments of the Sjöstrand test. For example, the submaximal characteristics has been emphasized and the duration of test stages has been shortened. Today, the commonly used variants of the PWC-test often has work rate is 5-min stages, with 3 min rest in between. The test also been developed in variants such as PWC_{195}, and the rest between stages is sometimes shorter than 3 min, and the work might be conducted with 2-min stages, 3-min stages, or 6-min stages (130, 131). Test results can be derived from prediction equations or graphic determination, where the linear relationship between increase in HR and the increase in work rate is extrapolated up to a HR of 170 beats/min, i.e., the “maximal” work capacity. The test is originally not developed for an indirect calculation of VO_2max, but rather a test of work tolerance. Since the Sjöstrand test, albeit being based on questionable physiological paradigms, combines work rate and HR, the data produced may be used to predict VO_2max. In the later years, the PWC- and Sjöstrand test has been evaluated as tests for estimation of VO_2max (132, 133). Kasch evaluated the Sjöstrand test against directly measured VO_2max from a maximal cycle ergometer test, and found a correlation of 0.55, and a standard error of 12%. The test underestimated VO_2max by 18%, resulting in a mean estimated value of 32.1 mL/min compared to the measured value of 38.8 mL/min, respectively in the validation study (133). The test has also been used in children and adolescents, see section 1.6.4.

With regard to the methodology and terminology, the PWC tests might be seen as a maximal test, due to the fact that the test was developed with the notion that the heart was unable to sufficiently work at higher intensities. Furthermore, one of the main reasons for prediction error is the lack of adjustments for the lower maximal HR and narrower HR range in elderly individuals.

Another test with multiple work rates is the YMCA-test, originally described in *The Y's Way to Physical Fitness* (134). The test was created to be performed up to a maximal level, but the multiple submaximal work rates is often used to estimate VO_2max based on the relationship between HR and work rate. The test is executed on a cycle ergometer, with a pedaling rate of 50 rpm and 3 min work at each load. The test is started on 25 or 50 watts, followed by continuous cycling through 3 or 4 higher work rates. The submaximal version of the test is terminated when the subject reaches 75-80% of maximal HR (220 – age). The estimations of VO_2max is calculated from the higher work-rates. Beekley et al. conducted a cross-validation of the YMCA test in a mixed population (20-54 years old) and found a standard error of estimate (SEE) of 9.8 mL/kg/min for men and 4.4 mL/kg/min for women, with an overestimation of 5.4% in men and 11.8% in women. Correlation between estimated and measured VO_2max was $r = 0.90$ in women (highly influenced by one outstanding high value, affecting the linear correlation) and $r = 0.63$ in men (135).

The Mankato submaximal exercise test (MSET) is a two-point test that consists of two 3-min stages at levels corresponding to approximately 35% and 65% (for inactive subjects) to 70% (for highly active e subjects) of VO_2max, respectively. One study has compared the MSET test and the YMCA test, showing no significant difference between MSET and measured VO_2max, whereas the YMCA test resulted in a significant underestimation of 4 mL/kg/min. Furthermore, the MSET correlated moderately to the measured VO_2max, with an ICC of 0.73, and a CV of 11.3%. The corresponding values for the YMCA test was ICC = 0.29 and CV of 15.1%, respectively (136). One of the main reasons for prediction error with the MEST is the influence of an estimated max

HR in the VO_2max calculation formulas. The formulas are found in an online Excel spreadsheet, available for public use at http://links.lww.com/MSS/A574.

1.4.2 Single stage test

In contrast to the multi-stage test, there are also one-point test. The most commonly used submaximal cycle test with one work rate is the Åstrand test, developed in Sweden during the 1950s and 1960s (108). The physiological basis for the test is the linear relationship between HR and VO_2, and a known relative oxygen consumption at a HR of 128 and 138 for men and women, respectively, as well as the relatively static work efficiency, where the energy cost/VO_2 consumption for produced watts during cycling is rather similar among different subjects (49). The background to the test was that extended research had developed the understanding of the human work physiology. The previously described theoretical background to the PWC tests (for details, see 1.4.1) was challenged and refuted. In the 1950s, Åstrand raised the question regarding HR response as a measure of both physical condition and working capacity by making a distinction between the two forms of "capacities" in the human body. According to Åstrand, "working capacity is a synthesis of aerobic and anaerobic capacity, mechanical efficiency, and physical condition; physical condition states how the circulation, respiration, muscles, etc., are fit to hard work of long duration" (49). With respect to that distinction, the calculated PWC tests are indeed related to maximal SV, but it is no measure of effectiveness or maximal power or rate of work output.

In the work of Åstrand, it was found that 50% of VO_2max corresponded to a HR of 128 beats/min for men and 138 beats/min for women, respectively (108). On basis of that, it may be concluded that if the 50% level was determined, it could easily be extrapolated (i.e. doubled) to generate a prediction of an individual's VO_2max. These observations were made on a homogenous sample of 16 female and 17 male young, relatively fit subjects.

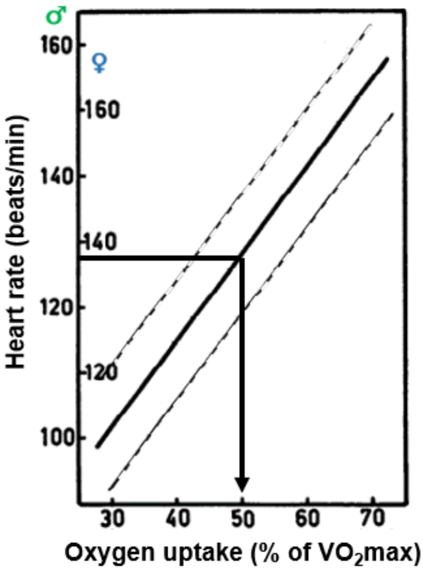

Figure 2. *Relation between percentage (%) of maximal oxygen uptake (VO_2max) and heart rate (beats/min) after about 6 min of work on a cycle ergometer. The left part of the y-axis is the heart rate values for men (♂). The right part of the y-axis is the heart rate values for women (♀). Dashed lines denotes the standard deviation. The illustration is an adjusted version of a previous published figure by Åstrand et al. (108).*

The first nomogram for estimation of VO_2max from the Åstrand test was based on the early findings that are illustrated in figure 2. The Åstrand test consists of 5-6 minutes of cycling one work rate, with a pedaling rate of 50 rpm. The work rate is chosen to correspond to 50–70% of VO_2max. HR is measured throughout the test, and the steady-state HR during the last minute is recorded. Originally, test results were derived from the Åstrand-Ryhming nomogram from 1954. The best test results was obtained with HR from 125 to 170 beats/min, however, the accepted values for HR and work rate in the nomogram ranged from 120–170 and 400–1300 kpm/min, respectively. VO_2max could be estimated between 1.6 and 5.4 L/min (108). However, in personal communication with the present research group, P.O. Åstrand stated that the publication of the first nomogram was an "academic ploy", and that they regretted the publication, because the individual variances were too large.

One of the reason for the misclassification with the Åstrand test is the variation i maximal HR with age. The first significant attempt to correct for this was presented by Irma Åstrand in 1960 (109). It was known that the maximal HR decreased with increasing age, thus narrowing the HR span and thereby affecting the HR response to a standardised work rate. The studies by I. Åstrand were conducted in subjects 20-65 years of age. The difference between the estimated VO_2max from the Åstrand test and measured

VO_2max in % of the measured value was calculated for each subject, and these differences were plotted in diagrams as a function of age (71).

The results revealed that the prediction error for subjects around 30 years old were approximately 10%, for subjects of 40 years of age 20%, etc., and the linear relationship showed that people became more and more overestimated with increasing age. Based on these findings, I. Åstrand created sex-specific age correction factor, with which the estimated VO_2max should be multiplied. For example, 20-years old women has a factor of 1.0, 30-year olds 0.9, 40 year olds 0.82, etc. (71). The number of subjects in the higher ages are still relatively low. Anyway, the work by I. Åstrand resulted in an updated version of the nomogram, which in total is based on a considerably larger and more heterogeneous sample (i.e., men and women, 20 to 65 years old, with varied physical status) than the original nomogram from 1954. Even though the age-correcting factors were applied to the nomogram, the standard error of the method was about 10% in young and well-trained individuals and 15% in moderately trained individuals of different ages. With the age-correcting factors applied, 5% of the subjects who are estimated to have a VO_2max of 3 L/min actually lies below 2.1 L/min, or above 3.9 L/min (113).

Further investigations of age-correction of the estimation of VO_2max were done by Von Döbeln, Åstrand and Bergström later in the 1960s (137). Subjects were 84 men (30-70 years old, mean VO_2max 2.69 ± 0.45 L/min) that cycled on multiple consecutive submaximal work rates, and thereafter performed a maximal cycle test for direct determination of VO_2max. Basically, the only thing that is equal to the procedure of the Åstrand test is that the calculations by Von Döbeln et al are based on HR and work rate from one 6-min stage of submaximal cycling. The experiments and the mathematical treatment of data resulted in an equation, where VO_2max was calculated from age, work rate, and HR from 6 min of cycling. The objective of the study was to find a prediction that was as precise as possible, with the use of as few variables as possible. It was concluded that measures of body size not contributed to any further enhancement in the precision of the prediction equation. In the experimental group, the SEE for estimations of VO_2max from the Åstrand-Ryhming nomogram was 12.7%, a values that was described as "definitely higher" than the results from the best equations presented by von Döbeln and colleagues. When the age-correcting factors by I. Åstrand (71) was applied, the systematic difference between measured and estimated values was 0.17 L/min. This difference disappeared when the age correcting factor by Von Döbeln et al. was used (137).

The Åstrand test has been evaluated by many other researchers. A rather good agreement between measured and estimated VO_2max from the Åstrand test was found by Teräslinna et al. The correlation between measured VO_2max, from a maximal cycle test, and predicted VO_2max from the original nomogram was $r = 0.69$. The use of the nomogram with age correction factors resulted in a correlation of $r = 0.92$ (138).

Other investigators have shown that the Åstrand test underestimates VO_2max. Jette et al. found that the Åstrand test with the age-correcting factors underestimated the measured treadmill VO_2 max, particularly in men. The mean VO_2max measured on the treadmill for males and females combined was 34.6 + 6.0 mL/kg/min while the Åstrand test predicted a mean VO_2max of 29.6 + 6.5 mL/kg/min. The correlation was r = 0.47 for the whole sample, with r = 0.54 for men and r = 0.59 for women, respectively. The CV for the Åstrand test in the whole sample was 15.3% (139). Also, Kasch (1984) investigated the validity of the Åstrand test in men between 30 and 66 years of age. The correlation coefficient for the Åstrand test with or without the age-correcting factors were $r = 0.58$ and $r = 0.46$, respectively. The age-corrected predictions underestimated VO_2 max by 21% (133). Furthermore, Cink et al. examined the validity of the two age correction factors (109, 137) in 40 young men. The measured VO_2max (L/min) from a maximal cycle ergometer test and the estimated VO_2max with the age correction of I. Åstrand had a correlation of r = 0.76, and SEE 0.42 L/min. The von Döbeln factors resulted in a correlation of r = 0.77, and SEE 0.41 L/min (140). Conclusively, in this population, the validity of the estimated VO_2max with the two different age correction factors was almost identical. Over the years, a number of other studies have also attempted improve the precision of the Åstrand test (141, 142) or to facilitate the method to find the individual optimal work rate to use for the test (143, 144).

In large populations, the mean value of estimated VO_2max from the Åstrand test correlates well to measured VO_2max (15). On individual level, the coefficient of variance (CV) for the test has been shown to vary between 15% and 18% (109, 145). One of the main sources for the relatively large prediction error on an individual level is that the method includes an estimation of maximal HR (or more accurate, HR range), a physiological trait that has large individual variations. A one-point test like the Åstrand test is also sensitive for all external factors that may affect the single HR response, for example ambient temperature, caffeine, nicotine, and heavy meals before test, nervousness and dehydration, as well as the intra-individual variation from age-predicted maximal HR, variations in mechanical efficiency, etc. Apart from age related and inter-individual variations in maximal HR, the circadian rhythm may influence the precision of the Åstrand test. This has been discussed by Reilly et al., who concluded that the observed variations over the day and night in their subjects (10 young athletes) could result in an error of 350 mL/min in the estimated value from the Åstrand test and nomogram (146).

The accuracy of the Åstrand test is highly dependent on the fact that the test is executed with all required prerequisites in order to diminish the effect of these factors. Furthermore, it may be valuable to conduct a first familiarisation test, which allows the subject to get used to the procedure. This will limit the influence of the sympathetic nervous system (i.e. nervousness and tension) on the submaximal HR response (147). Another hold back for the Åstrand test and the results from the associated nomograms is that a large number of commonly recorded HR-work rate combinations fall outside the nomogram, and thus leads to loss of data.

With regard to the built-in error of differences in maximal HR among individuals, and as a consequence of aging, this issue was probably masked, and also enhanced, by the fact that original nomogram was built on a very small and homogenous group (33 well-trained men and women, 18-30 years old, with a mean maximal HR of 195 beats/min). It is logical to assume that the estimation error increases when the nomogram is applied on completely different populations with different characteristics. On a group level, the Swedish population in the 1940s where to a greater extent comprised of lean, physically active and relatively fit individuals, whereas a greater proportion of the population today tends to be more inactive and overweight (148). The significance of the inter-individual differences in HR is also underlined in the first article by Åstrand, saying that the HR response to one submaximal work rate (and that the correlation of the slope of a HR – VO_2 uptake curve is determined by the subject's aerobic capacity) actually gives a rather rough estimation the aerobic capacity (108). Furthermore, the actual relationship between the cardiac output and VO_2 uptake was not fully understood when the original nomogram was developed, and the same held true for the variations of SV and a-$\overline{v}$ O_2 diff in response to heavier intensities and maximal workload.

Lastly, one explanations to the discrepancies between studies with regard to validity and reliability of the Åstrand test is the procedure for the direct measurement of VO_2max. The nomogram are developed to correspond to VO_2max measured from low-cadence cycling, and in some cases horizontal running. Whether these test procedure generates an actual VO_2max, or rather a VO_2peak, can be discussed. It is actually mentioned by the researcher that the similar VO_2max for cycling and running was found in male subjects, whereas females exhibited a significantly lower VO_2max in the maximal cycling compared to running test (49). Other studies of validity and reliability for the Åstrand test have used a variety of different cycle or treadmill test protocols, and applied varied criteria for the determination of VO_2max.

1.4.3 Two-point tests – the theoretical background to the EB test

It is evident that the estimates of VO_2max from one work rate, or generalisations of physiological fundaments and/or linearity, can explain/can cause large amounts of the prediction error, especially on individual level. A promising way to circumvent that is to incorporate the individual characteristics and responses to an exercise stress test in the VO_2max prediction equation. The theory behind the model for the Ekblom-Bak test (EB test) is the well-known fact that a trained person with higher VO_2max exhibits a less pronounced increase in HR during a fixed increase in work rate, compared to a less trained individual with lower VO_2max.

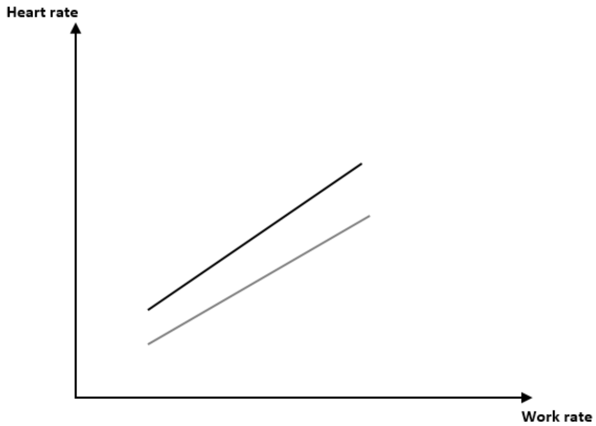

Figure 3. *Schematic illustration of two different heart rate responses to a given increase in workload (upper line = untrained individual, lower line = well-trained individual).*

The hypothesis in the present thesis is that the use of the delta HR (ΔHR) response between a low and an individually chosen higher work rate (the Δ power output, ΔPO), rather than the HR response to one work rate, will enhance validity in estimations of VO_2max in mixed populations. For example, the use of ΔHR in the prediction equation might make the test less sensitive for individual variations in maximal HR, and thereby produce higher validity through lower random variation.

One previous study has used this ΔHR theory in a similar way. In 1986, Legge & Banister developed a nomogram that was based on the linear relationship between VO_2 and ΔHR. The ΔHR was defined as the elevation of exercise HR above the HR from zero-load pedalling at 90 rpm. The participants in the study were young men, divided in groups of trained (n = 15) and untrained (n = 10) subjects. Two separate regression lines were developed for trained and untrained individuals. Furthermore, the predictive validity was examined in a mixed group of subjects with different physical activity level (n = 14). External validation in five subjects showed a significantly better correlation between measured and estimated VO_2max by the Legge-Banister nomogram than the results from the Åstrand nomogram with age correction factors, with r = 0.98 and r = 0.80, respectively (149).

For the EB test, all individuals are supposed to work at the same starting workload, a low work rate of 0.5 kp and 60 rpm (i.e. ~30 W, standard work rate). That is tolerable for almost the whole population. An untrained individual with low aerobic capacity (low VO_2max) also has a relatively high resting HR, and this individual will not reach especially high on the VO_2max scale. A well-trained individual has a lower resting HR, and also a lower HR at the standard work rate. The prediction equation with ΔHR/ΔPO is thereby based on the HR-reserve rather than the relation between a single HR-response and an estimated, or known, maximal HR. For example, if the ΔPO is 100W, an untrained individual may achieve a ΔHF of 50 beats/min, while a well-trained person has a ΔHR of 40 beats/min. The lower quotient of the ΔHR/ΔPO is thereby representing a higher aerobic fitness, i.e. a higher VO_2max. Furthermore, the individual characteristic of this relationship between HR response and increased workload may contribute to enhanced preciseness and lower individual variation, compared to many other submaximal exercise tests and their corresponding prediction equations.

Submaximal exercise tests are useful in many different settings, even if the estimations of VO_2max have a number of limitations. Possible sources of error is the individual variation in moving economy (especially during weight-bearing activities like step- or walk tests), that may lead to variations in HR response at a given work rate. Furthermore, many submaximal tests are based on extrapolation to the maximal values of HR and VO_2. However, there are a rather asymptotic relationship between HR and VO_2 near maximum, and the variability in age-predicted maximal HR has a CV of 10% or even more (110). Also, the estimation of VO_2max from tests that use one absolute HR value at one work rate can be seriously impacted by the variability of submaximal HR due to variations in ambient temperature, nervousness, or emotions, as well as intra-individual variability of VO_2/HR at a given work rate.

1.5 Ability to detect changes in VO_2max trough submaximal exercise tests

The ability of a submaximal test to detect changes in VO_2max is crucial in order to make the test useful in for example longitudinal studies or before and after training interventions. A precise and accurate test with the capacity to follow individual changes in VO_2max is of importance both for individual health care and guidance, as well as in scientific contexts. For example, longitudinal studies may use submaximal cycle tests for estimation of VO_2max because these tests are easy to administer to large groups of people. However, even if a test is reasonably valid at one occasion, it does not necessarily mean that the accuracy (the magnitude of the measurement error) is equal for an individual through training-induced or longitudinal changes in VO_2max.

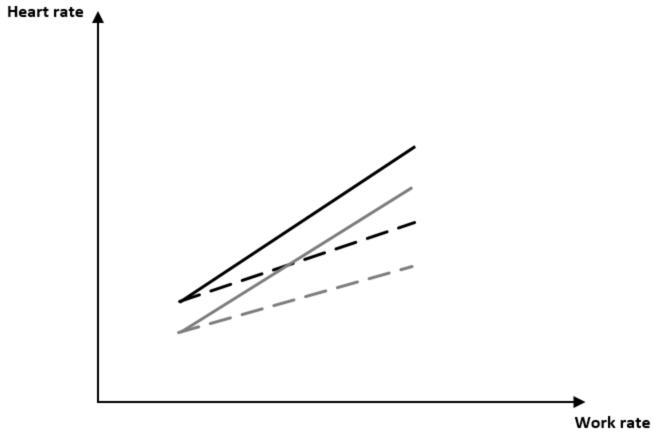

Figure 4. *Variations in physiological response to a two-point test. The upper black line illustrates the heart rate response (ΔHR) to a low and a higher rate of work at baseline. The lower grey line illustrates a change in position of the slope. The dashed black line illustrates a change in the angle of the slope. The dashed grey line illustrates a change in both position and angel of the slope.*

Up to date, it is largely unknown if any submaximal cycle ergometer test have the ability to detect a change in actual VO_2max after a training intervention, or as a consequence of time. With regard to the EB test, the variable ΔHR – described in detail in section 1.4.3 – may vary with changes in VO_2max in a number of ways (see figure 4).

1.5.1 Capture short-term changes in VO_2max

One of the fundamental principles behind the HR-based exercise tests is that HR on a single submaximal workload is lower in well-trained subjects than in less trained subjects. Regular aerobic training results in an increase in SV and a larger a-$\bar{v}$ O_2 diff, and these factors lead to an enhanced oxygen transportation and an increased VO_2max. The higher SV in a trained heart also means that a given cardiac output can be achieved with less heartbeats. Furthermore, a decrease in HR at a given workload has been observed to be related to positive training adaptation. Thus, submaximal exercise HR may be an efficient method of assessing cardiac autonomic activity and tracking changes in maximal aerobic capacity (150-152). It is reasonable to hypothesise that the difference in HR (ΔHR) between two work rates will be less pronounced. Increase in blood volume (with subsequent increase in peripheral resistance, blood pressure and after-load) may counteract and/or hide these changes, while measured (*i.e.* estimated) with a submaximal one- or two-point test. Furthermore, an altered a-$\bar{v}$ O_2 diff, either as a consequence of training or aging, may also alter the relation between VO_2 and HR.

With regard to the natural variations in submaximal HR, Davies et al. (153) have showed that a higher intensity work during a submaximal cycle test resulted in intra-individual variations in HR of 2%, while intra-individual variations at lower intensities were higher and ranged from 3% to 8% when using the Åstrand test (108). This can partly be explained by the fact that HR at lower levels of submaximal work might be affected by influences from the sympathetic nervous system, so that factors such as fear, excitement, and emotional stress elevates the HR (147). This may cause an elevation of HR at a submaximal work rate, without any change in actual VO_2peak. Furthermore, submaximal HR might be affected by medications. For example, treatment with chemotherapy induces autonomic dysfunctions (154) which may influence HR response to submaximal work, especially at lower levels of exercise. This has also been shown in relation to submaximal exercise testing, for example in one study by Mijwel et al. (155).

The use for the submaximal tests is in the evaluation of physical exercise programmes for rehabilitation of patients with different diagnosis is another important area of use. In a situation where maximal exercise testing is not practically possible to conduct, or undesirable from a patient safety perspective, submaximal exercise testing provides an alternative way of estimating VO_2max. Some training studies have examined the association between changes in submaximal and maximal exercise outcomes in cancer survivors. May et al (156) included cancer survivors that were not taking any medications that influences HR. All subjects participated in a training programme during 12 weeks. VO_2peak was measured during a graded exercise test on a cycle ergometer, and the submaximal test comprised of 10 minutes cycling at an intensity corresponding to 50% of peak Power Output. VO_2peak increased significantly from pre- to post-intervention. However, the changes in submaximal HR at a fixed work rate (approx. 50% of Power Output) had a rather weak and non-significant relation to the actual change in measured VO_2max ($r = -0.15$, $p = 0.08$). Furthermore, the results revealed that only submaximal HR > 140 beats/min were associated with changes in VO_2peak. However, the conclusion was that a submaximal test should be conducted at a work rate above 50% of peak power output (> 140 beats/min) to increase the correlation between the change in actual VO_2peak and change in a submaximal test (156). In the present study, there was a weak but significant correlation between HR at standard work rate and change in measured VO_2max ($r = -0.42$, $p = 0.02$). The correlation was somewhat stronger at the high work rate, where the ($r = -0.57$, $p < 0.01$) for the 25 subject that exercised at the same work rate at baseline and follow-up, respectively.

Another study evaluated an 18-week training program and analysed pre- and post-values of VO_2peak and submaximal HR at 50%, 60%, and 70% of peak power output in 36 cancer survivors. The HR at different submaximal workloads did not decrease in their participants from pre-intervention to post-intervention, whereas VO_2peak and peak power output improved significantly (157). A possible explanation of these opposite findings might be the submaximal testing protocol they used: the test started at 50 % of peak power output and was increased 10 % every 3 minutes, sampling the HR during the last 15 seconds of each stage. A duration of 3 minutes might be too short in order for a deconditioned cancer patient to achieve a true steady-state, which is needed for a valid monitoring of a HR response to submaximal exercise.

1.5.2 Capture long-term changes in VO_2max

Apart from training related alterations in HR, a reduction in submaximal HR at exercise can also occur due to other reasons, primarily the age-related decline in maximal HR (71). Apart from changes in training status, and aging, there are other intra-individual variables that can change considerably over time, such as body mass and muscular mass, neuromuscular function, vascularization, etc. Submaximal tests are often used in larger-scale settings (even epidemiological studies) because of the difficulties to conduct direct measurements of VO_2max in large groups of people. However, the ability of the submaximal tests to capture longitudinal (> 4 years) changes in aerobic capacity is seldom studied. To study such longitudinal changes in VO_2max is not quite the same thing as experimentally increase VO_2max by a short period of intense aerobic training, since the former type of study has to take age-related changes into account.

With regard to the commonly used Åstrand test, it is worth mentioning that the test has been shown to have a rather poor ability to capture long-term changes in cardiorespiratory fitness. In a previous work, Irma Åstrand conducted a 20-year follow-up study with cardiorespiratory fitness tests in healthy adults (34 women and 31 men). The females was between 20-25 years old at the time for the baseline tests, and 41-46 years old at follow-up. The corresponding values for men was 20-33 and 41-54 years old, respectively. All subjects were physically fit and engaged in studies to become physical education teachers when the first tests were conducted, and the majority were physically active also at follow-up. With regard to body mass, the females were on average 2.5 kg lighter in 1970 than in 1949, while the men were 2.5 kg heavier. However, there was large individual variations in the weight gain or weight loss, with numbers ranging from -10.3 to +15.7 kg. The mean change in maximal HR was -15 beats/min for women and -12 beats/min for men. This decrease was significant on group level, yet some subjects had a rather unchanged maximal HR, while some subjects displayed as much as 25-30 beats slower maximal HR. With regard to VO_2max, all subject demonstrated lower values in 1970 than in 1949, with a mean decrease of 22% in women (from 2.83 ± 0.27 to 2.20 ± 0.29 L/min; values are mean ± SD) and 20% in men (from 4.08 ± 0.37 to 3.28 ± 0.45 L/min), respectively.

Results from the cycle tests showed that changes in actual VO_2max could not be identified by the Åstrand test. Mean HR at a given submaximal workload was higher at follow-up, but with large inter-individual differences. For example, 43 subjects had very similar mechanical efficiency ($VO_2 \leq 0.11$ L/min during submaximal workload) in 1949 and 1970, respectively. However, some of those subjects displayed a corresponding HR that was higher at follow-up, while others exhibited a lower corresponding HR. The explanation for this differences in submaximal HR response could not be related to difference in VO_2max (there was no correlation between decrease in maximal HR and decline in VO_2max) or max HR (i.e., there was no correlation between change in maximal HR and HR at submaximal work rate). There was almost no difference in efficiency, measured as VO_2 uptake for a certain workload. The only exception from this observation was for the work rate of 200 W, where the men had a higher VO_2 in 1970 compared to 1949 (2.84 L/min compared to 2.67 L/min, respectively). In summary, Åstrand et al. concluded that changes in actual VO_2max could not be identified by the Åstrand test and the corresponding nomogram (59). A contributing factor for this was the inter-individual variation in changes in maximal HR over time.

1.6 Exercise testing in children and adolescents

Cardiorespiratory fitness has been linked to various aspects of well-being in youth, for example cardiovascular health (31, 158, 159), metabolic syndrome markers (30, 160), insulin sensitivity (28), mental health and depression (32, 161). Physical fitness may also have a positive influence on cognitive skills and academic achievement (33, 34). Thereby, valid and reliable submaximal tests for monitoring of VO_2max in children and adolescents are beneficial for both individuals and society.

1.6.1 Cardiovascular and circulatory function

The circulatory responses to exercise in children and adults have some minor differences, and the cardiovascular functions in children during growth are not yet fully understood. It is known that pre-pubertal children have larger blood vessels in relative terms in relation to the size of the heart muscle, which produces a less pronounced increase in blood pressure during incremental work (162, 163). Children have smaller heart muscle/kg body mass, and consequently, a lower SV (and lower Q) than adults (164). This is largely compensated with higher HR and higher a-$\bar{v}$ O_2 diff, in order to supply the oxygen demands of the body. In line with the higher a-$\bar{v}$ O_2 diff, children display a much more efficient peripheral O_2 uptake in working muscles than adults (163, 165, 166), due to a more efficient redistribution of blood during work. As a consequence of the faster O_2 extraction in the muscle, pre-pubertal children also have faster a VO_2 kinetic than adults (167).

At the onset of work, SV has a rapid and linear increase from rest to easy/moderate intensity, where it remains rather stable regardless of further increases in workload. This curve linear response in SV displays more or less the same pattern in children as in adults (164, 166, 168, 169). However, it has also been indicated that the plateau in SV may appear somewhat earlier in children than in adults, approximately at 35-40% of VO_2max compared to 40-60% of VO_2max in adults (170). Also, one previous study indicates that the lower blood pressure during exercise is related to stature (171).

The maximal HR are about the same in boys and girls (96, 172), while SV tend to be higher in boys (172). Also, boys generally have a higher VO_2max than girls (96, 173). There are no significant differences between sexes with regard to peak RER and maximal blood lactate levels, but children displays significantly lower values then adults (174). With regard to submaximal exercise intensities, boys usually display lower HR and higher SV than girls, at any given workload. A possible explanation to these findings is that boys in general have bigger heart muscles than girls (172, 175).

1.6.2 Influence of growth and maturity

Boys and girls display more circulatory and physiological similarities before they enter puberty. For example, Marta et al. conducted a training study in adolescent boys and girls, and found no significant differences in the training induced changes in strength and endurance (VO_2max estimated by the Beep-test) in boys and girls in Tanner stage I or II, respectively (176). Furthermore, there are greater maturity related differences between boys and grown up males than between girls and grown up females. For example, the differences in VO_2max between pre-pubertal girls and boys are not of the same magnitude as the difference in VO_2max between adult men and women (96, 173).

Absolute aerobic capacity, expressed as VO_2max in L/min, usually increase relatively linear with age in boys, and a similar trend is seen in girls (177, 178). In absolute terms, 15-year old boys have on average approximately twice as high VO_2max as 9-year olds, while girls improve their absolute VO_2max ~60% between the same ages (178). Hence, the effects of maturation on the relationship between VO_2 and body size differ between boys and girls (179, 180). However, it is important to remember that there are often larger variations within gender or age groups, than between them.

With regard to the commonly used relative VO_2max value, expressed as mL/kg/min, boys tend to increase it with age while girls tend to decrease it (178). This is partly explained by the maturity related development in body composition, where boys gradually have a higher proportion of muscle mass, while girls exhibit an increased body fat percentage in the later stages of puberty. Although it is common to describe differences and changes in cardiorespiratory fitness with VO_2max in mL/kg/min, it has been shown to be a less accurate way of quantifying growth-related changes in VO_2max in children and adolescents (181). Other methods for removing the influence of body size on VO_2max is allometric scaling (182, 183). With the use of proper scaling-factors, i.e. when VO_2 is correctly normalised for differences in body size, the correlation between

VO_2max and age disappears and it is only older boys (individuals that are soon physically developed to men) that displays significantly higher values than other adolescents. Furthermore, older girls have significantly lower values than younger boys (183).

Children's cardiorespiratory system and capacity are affected of other factors than changes in body size, such as hormonal influences and development of muscle mass (184). This phenomenon further complicates the interpretation of test results in children, regardless of scaling method. It should be emphasised that the cardiovascular and respiratory systems in children and adolescents adapt to the change in body size during growth, and the responses to increased workload is optimised for all ages, i.e., for all body sizes (185).

With respect to the above mentioned observations, it is uncertain whether measurements and/or estimations of VO_2max can be interpreted in the same way in children as in adults. Furthermore, it has also been suggested that the main differences in VO_2max before puberty can be related to genetic factors (186). Gains in performance in endurance events may be achieved through improved submaximal exercise economy, qualitative changes in oxygen delivery not indicated by VO_2max, or the development of non-aerobic factors as speed, strength, and agility.

1.6.3 Determination of VO_2max/VO_2peak

The criterion of a levelling off in VO_2 despite increased workload is a popular criteria for determination of VO_2max in adults. Recordings of VO_2 consumption during incremental exercise test until voluntary exhaustion can be conducted in children in the same manner as in different groups of adult subjects. However, the "levelling off" in a maximal test may occur even more seldom in youth (87), compared to the divergent reports of the appearance of the phenomena in adults of $\leq$ 50% up to 90-100% (69).

In the experiments by Åstrand in the early 1950s, a "levelling off" was found in 70 of 140 tests in school children (108). However, it is unknown (but highly likely) that a number of these subjects were well beyond the pre-pubertal state, and thereby able to perform a relatively high degree of anaerobic work. Rowland & Cunningham found that about ~30% of children (7-10 years old) demonstrated a plateau. There were no significant differences in VO_2peak, maximal HR, maximal RER, or performance in anaerobic exercise tests, between the children whom reached a plateau, and those who did not (88). Additionally, evidence from supramaximal verification tests supports the idea that also children with no evident plateau from a maximal test still may have reached their upper limit of VO_2 utilisation (88, 89). Thereby, the use of the "leveling-off" criteria, has been questioned in children (87, 96), and a number of child-specific supporting criteria has been developed. These supporting criteria have lower cut-off values for RER (~ 0.95 to $\leq$ 1.0) and blood lactate levels, and higher limits for acceptable peak HR, but it is still problematic to conform specific criteria to such a heterogeneous group as children and adolescents during growth (88, 89, 96). Because young subjects may be fatigued just as the highest tolerable VO_2 has been reached, and thereby only exhibits a peak in VO_2 from a maximal exercise test, it is rather common that the literature refer to a value from maximal exercise testing in children and adolescents as a "peak".

1.6.4 Submaximal cycle tests in youth

Commercially available exercise tests must be easy to conduct and safe to perform. These aspects are especially important when testing unaccustomed and/or vulnerable individuals and young people. Thereby, submaximal exercise tests are suitable to use in children and adolescents. Another advantage of submaximal tests in youth is the elimination of a maximal and/or performance aspect, which means that the subject cannot "fail" or perform bad, as in many other tests. Furthermore, a valid and precise test, that not requires maximal effort and/or laboratory equipment, can be used in both epidemiological and clinical studies. However, it is still largely unknown to which extent submaximal tests, developed for adults, are applicable in pre-pubertal and maturing children and adolescents. The above-mentioned circulatory differences between children and adults may affect the validity for submaximal tests that are developed for adults. The observation of a different response in SV, for example, may lead to that an increase in work rate generates a somewhat less increase in HR, resulting in an overestimation from a submaximal one-point test like the Åstrand test.

The Åstrand test has been evaluated in children. In one study, results showed a 12% underestimation of VO_2max in 10-12 years old boys, with a correlation coefficient of $r = 0.60$ ($p < 0.05$) when expressed in L/min and $r = 0.55$ ($p < 0.05$) when expressed in mL/kg/min. Furthermore, the test falsely indicated an enhanced VO_2max after a training period, although no actual change in VO_2max was present (187). One possible explanation to this might be the familiarisation with the test after the training period, resulting in lower HR on the submaximal work rates and thereby a higher estimated VO_2max in the nomogram. Others have reported that submaximal cycle ergometer exercise tests generally overestimated VO_2max Also, previous studies have developed child-specific regression models for estimation of VO_2max from the Åstrand test, for example the Woynarowska model ($r = 0.82$ in girls and 0.52 in boys, respectively) (188) and the Binyildiz model ($r = 0.70$ in 11-13 year old boys) (189). When Binyildiz et al. included height and age in the analysis, r was improved to 0.80, with a standard error (SD) of ± 18% of the mean (189). In a comparative study, Ekblom evaluated the Åstrand-Ryhming nomogram and the two above mentioned regression models in 62 children aged 11-12 years. Low mean misclassification was found for the Woynarowska regression model and Åstrand-Ryhming nomogram (14 mL/min and 23 mL/min, respectively).

The Åstrand value was adjusted to an age of 12 years, while the unadjusted value resulted in an over-estimation of 346 mL/min. The tests had moderately to high correlation with measured VO_2max (r = 0.81 and 0.73, respectively), and a high standard error of estimate (SEE 398 mL/min and 340 mL/min, respectively). The Binyildiz method displayed a high correlation with measured VO_2max (r = 0.87) and relatively low SEE (298 mL/min), but a large mean underestimation (660 mL/min). Furthermore, all methods underestimated VO_2max in well-trained children (190).

Also, different forms of the PWC test have been validated in the pediatric population. The PWC tests predict VO_2max with varying accuracy (r = 0.46-0.87) in children, depending on sex, age, amount of minutes cycled at each stage, and if VO_2max is predicted with absolute or relative values (130, 131, 191-193). For example, Wells et al. found that the significant correlation between measured and estimated VO_2max in teenaged well-trained boys and girls differed between 0.87 and 0.56, respectively (193). Furthermore, Rowland et al. found that the PWC_{170} test had the best precision when VO_2 was expressed in L/min (r = 0.71 and 0.70 for girls and boys, respectively) (192).

In summary, the circulatory functions are more alike in pre-pubertal boys and girls than in adult men and women, and the morphological and physiological changes during puberty are more pronounced in boys than in girls. The influence of pubertal development on the circulatory responses to submaximal and maximal aerobic work is partly unknown, and more research is needed to clarify the role of maturity on cardiovascular responses to submaximal exercise. It is reasonable to believe that any differences in validity are related to growth and sexual maturity, but up today, knowledge in this field is sparse. Based on the above mentioned growth-related differences between boys/men and girls/women, it is possible that pre-pubertal boys has a circulatory response to submaximal work that has more similarities with woman than fully grown men. In that case, for example, the validity for a sex-specific prediction equation would be higher if results for pre-pubertal boys were calculated with the female model.

1.7 Conceptual issues regarding cycle ergometry

The use of ergometers to produce work during exercise tests in laboratory environments dates back to the eighteenth century. In 1896, Bouny constructed the first bicycle ergometer by putting a mechanical brake directly on the rear wheel on an ordinary bicycle, which was lifted from the floor (194). The more precise control of the work rate came with the stationary ergometer with a pendulum with a known weight, an innovation made by von Döbeln in 1954 (195). In order to create a general context for the concept of submaximal exercise testing, this section will include a brief historical and methodological background, as well as some additional experimental cycle ergometry data. Bicycle ergometers of varying construction and price are available.

The cycle ergometer used for the tests in the present thesis, the Monark 828E (Monark Exercise AB, Vansbro, Sweden), is a modification of a construction von Döbeln (195), with the technical assistance of Mr. Hagelin. Monark Exercise AB has produced and developed cycle ergometers since the early 1900s. The ergometer 828E by Monark has a stable frame of a solid steel tube. The flywheel is large and well-balanced (20 kg) and the total weight of the ergometer is 52 kg. The highest tolerable user weight is 250 kg.

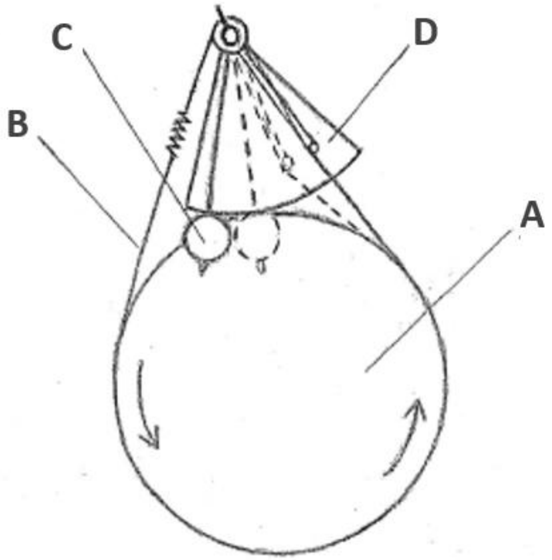

Figure 5. *The central parts of the flywheel construction of a Monark cycle ergometer.*

The main piece of the 828E cycle ergometer construction is shown in figure 5. The main piece is the front wheel (A) which is mechanically braked by a belt (B) running around the rim. Both ends of the belt are connected to a rotating cylinder, to which a pendulum (C in figure 5) is fixed. The belt can be stretched or loosened by the use of load-regulating hand wheel, i.e. the braking power (kp) is set by adjustment of the belt tension. The braking power from the pendulum can be read off on a scale (D in figure 5), where 1 kp is the force acting on the mass of 1 kg at normal acceleration of gravity; 100 kpm/min = 723 foot-ponds/min = 16.35 watts (nominal work rate). However, to obtain the actual work rate to be performed, an additional eight percent to correct for frictional losses at the chain and drive train has to be included (196).

1.7.1 Operational aspects

The use of cycle ergometry in exercise testing requires that certain conditions are fulfilled. For example, it is of great importance that the device is properly handled, from general service to standard calibration. However, it is rather common that important operational aspects – such as the calibration routines for the cycle ergometer – is described very sparsely (or not at all) in different contexts of submaximal cycle testing.

Another possible source of error when using mechanically braked ergometers is the manual handling of the adjustments of the breaking force. The ability to reproduce an exact work rate is dependent on the carefulness and precision from the operating person. Furthermore, the resulting effect from the cycling work against a known breaking force is also dependent on the pedalling rate, which is generated of the subject himself. The resisting force from 2.0 kp will result in a work rate of 118 W if the subject delivers exactly 60 rpm. However, the work rate will be only 114 W if the cadence drops to 58 rpm, and it will increase to 122 W with a cadence of 62 rpm (excluding friction losses). Hence, it is of great importance to carefully monitor the cadence during testing, in both clinical and scientific contexts. For example, the use of a metronome may facilitate the maintenance of the correct cadence. These “built-in” sources for errors are seldom discussed in the current literature.

Yet another operational aspect is the role of the surface where the cycle ergometer is operated. Still, some uncertainty remains with regard to the influence of soft grounds (carpeted floor, rubber floor) compared to steady and stable grounds (hardwood floor, stone floor). The different surfaces may alter the load from the pendulum and influence the work rate. The current recommendation is that the cycle ergometer should stand on a level, firm foundation. Furthermore, the calibration should be done on the same type of ground and at the same place where the ergometer shall be operated. Calibration – with adjustment/zeroing of the scale – shall be performed with the subject seated on the ergometer, without touching the pedals (196). Another aspect that has to be taken into account is the function of the cycle ergometer on a slightly inclined surface. This should be avoided (196).

In addition to the above mentioned aspects of cycle ergometry, it is also possible that difference in body mass may influence the function of the ergometer – i.e. that the calibration and the load (kp) is affected differently by an individual with a body mass of 40 kg compared to a body mass of 100 kg. Up to date, the eventual impact of total body mass on the cycle ergometer is largely unresolved.

1.7.2 Body position and pedals

The influence of body position during stationary cycling attracted attention in the 1970s. Faria et al. investigated the physiological responses to an almost horizontal position during an incremental exercise test on a cycle ergometer. The dropped position generated significantly higher VO_2 at a submaximal intensity around 84% of VO_2max, compared to the upright position. Also, VO_2max, maximal work output, and maximal V_E was higher in the dropped position (197). Alterations in mechanical efficiency or ventilation were suggested as possible explanations to these findings. Later studies have presented contradictory results, with no significant differences between VO_2max for upright and dropped body position during cycling (198, 199). In trained cyclists, physiological responses like VO_2, RER and HR has been reported to be unaffected by different body position during submaximal workloads (200), while other researchers have reported higher ventilatory variables and RPE with a dropped compared to an upright body position (201). Another position during cycling is the aero position, but that feature of competitive cycling will not be further discussed. However, studies of different seat-tube angle variation during steady-state cycling with aero bars have showed that only extreme combinations of seat-tube and trunk angles are associated with higher steady-state cardiorespiratory measures when external power output is kept constant (202). Furthermore, it has been shown that no combination of seat-tube and trunk angle can be considered optimal of experienced cyclists. The most economical work (i.e., lowest VO_2 cost for a given workload) was seen when the cyclists rode with ergometer geometries that most closely replicated the lower-limb kinematics of the subjects riding their own bicycles (203).

Another aspect that may have influence on the efficiency while cycling is the pedals. One of the first mentioned concerns with the "foot fixed" bicycle pedal versus ordinary bicycle pedals, and possible influence on efficiency (i.e., VO_2 uptake) on submaximal workloads, can be found in a study by Sutton et al. 1988. The research group conducted a series of experiments with different workloads on successively increasing during a simulated ascent of Mt. Everest over 40 days. In contrast to many previous studies, Sutton et al. reported a reduced VO_2 at submaximal workloads at simulated altitudes. However, after the first trials in the decompression chamber (sea-level conditions), the cycle ergometer was modified on request from some of the subjects. Toe clips were added, and the handlebars and saddle were changed to a "racing" version. The authors reflect upon their findings by adding the following sentence in the method section: "These modifications may have improved the "efficiency" of subsequent ergometry at altitude, i.e., lowered the O_2 uptake (VO_2) for any given work load (204).

Later, Ostler et al. tested this assertion. They aimed to establish if the efficiency during cycle ergometry could be reduced by at least 8% (a value taken from the study by Sutton et al. 1988) when using pedals fitted with toe-clips and straps compared with flat pedals (i.e. fitted with neither clips, nor straps). The tested flat pedals were the blue pedals (Exerfit PD-E100, Shimano, Japan) that is found on numerous Monark cycle ergometers like the model 824E and 828E. These pedals were modified by removal of the standard Monark rubber foot straps, giving these pedals a resultant mass of 388 g each. The toe-clip pedals were bicycle pedals with plastic toe-clips and toe straps (total mass 236 g per pedal) similar to those supplied by Monark Exercise AB. The cycling was performed in 5-min stages of 60, 120, 180, and 240 W, with a pedalling rate of 60 rpm.

Measurements of VO_2 were done with the Douglas Bag method. Toe-clip pedals did not significantly reduce VO_2 or HR compared to the flat pedals (205).

In the study by Ostler et al., the observed VO_2 values at 60 and 120 W with flat pedals were within 4% of those predicted by the American College of Sports Medicine, 2000 (206), while the higher workloads resulted in 9-12% higher values than expected. Compared with the VO_2 predictions of Åstrand & Rodahl (113) for 180 and 240 W (2.55 and 3.33 L/min, respectively), the corresponding values in the study by Ostler et al. were ~6.5% higher (205). The measured VO_2 for flat pedals at 60 and 120 W (1.26 and 1.85 L/min, respectively) were 5% and 3% higher than those reported by Sutton et al. at sea level at the same power outputs (1.20 and 1.80 L/min, respectively) (204, 205).

Overall, research on the influence of various types of pedals on gross efficiency is sparse.

1.7.3 The influence of body size on work efficiency (O_2 consumption)

It is known that differences in body size influence the VO_2 consumption during standardised cycle exercise. Early studies by von Döbeln and coworkers investigated the effect of body size on the energy cost for exercise (both at standard work and at maximal work). In these studies, the researchers noticed that variations in body size introduces problem regarding comparisons between individuals (207). The energy consumption in relation to a given cycle workload correlates to body mass to various degrees (208-211).

Furthermore, there are support to the idea that the increased VO_2 cost with increasing body mass is a consequence of the difference in the cost of moving the “leg mass” in the pedalling movement (51, 208, 210). In line with the previous findings that work efficiency is dependent on body mass, it has also been shown that the predictive power in estimations of VO_2max from a submaximal work rate can be enhanced by adding body mass to the prediction equation (209).

While it is well understood that individuals with differing body sizes have differing energy costs for a given workload, the development of reliable scaling methods in order to adjust for this is an ongoing concern. There is no current consensus in the field, although it is common to describe differences and changes in cardiorespiratory fitness with VO_2max in mL/kg/min. However, that is a less accurate way of quantifying for example growth-related changes in VO_2max in children and adolescents (179, 181). Other methods for removing the influence of body size on VO_2max are allometric scaling. The scaling of VO_2max is often performed using different functions of body mass, for example the full body mass or raised to the 0.67 (predicted from the theory of geometric similarity), or as a function for fat free body mass or fat free mass of the legs (181). Allometric scaling is a way to even out the effect of body size found in each sample, which therefore is unique for each study. However, all currently used forms includes L/min as a nominator, which is the reason for reporting this in the present paper. Hence, all VO_2 values from the experiments within the present thesis are expressed as L/min, and thereby not corrected or adjusted for any size-dependent differences in VO_2max. This approach places focus on the total energy expenditure and oxygen consumption, rather than adding analyses on scaling.

Differences in body size and body composition are interesting aspect to considerate in relation to the development of submaximal exercise tests. It is useful to have knowledge about the population where a test has been developed and validated. During the past decades, a rapid change in mean body mass, and also height and body composition, has been seen in people in many societies. For example, the Åstrand test (108) was developed in a homogenous group of young, lean and fit students in Sweden. The validity may be affected when the test is applied in other groups, both concerning fitness level and body size. However, the definite influence of differences in body mass on energy expenditure during cycling, and in relation to submaximal cycle tests, needs to be further investigated.

1.7.4 Effect of cadence on work efficiency (O_2 consumption)

It is well established that the pedaling rate (cadence) has an influence on work efficiency (O_2 consumption) and HR at submaximal work rates. As already mentioned (see section 1.2.2), it has been historically believed that a pedalling rate of 50-60 rpm was the best cadence to use in tests of both energy efficiency on submaximal work rates and performance and VO_2max (12, 56-58). It has been repeatedly shown that the most economical pedalling frequency during submaximal cycling is in the range of 50-60 rpm (60, 61). Furthermore, it has been reported that there are no significant differences in mechanical efficiency between 50 and 60 rpm (212, 213).

With the above said, the freely chosen cadence still varies between individuals. For example, one study examined the responses to workloads corresponding to 40, 50, and 60% of maximal power output, with fixed cadences of 50, 65, 80, 95, 110 rpm, and in one trial with freely chose cadence. These non-experienced cyclists had a self-selected freely chose cadence of 70-90 rpm (65), which is higher than the most economical pedalling frequency of 50–60 rpm, that optimize efficiency, economy and ratings of perceived exertion (60, 61, 214). Competitive cyclists tend to select cadences that are even higher (> 90 rpm) during training and racing. Moreover, during submaximal cycling, trained cyclists tend to prefer a cadences that are higher than those that are energetically optimal, resulting in an excess energy expenditure of approximately 5% (215). For these subjects, it may be more important to adopt a cadence that minimise muscle strain and mechanical load, rather than minimising metabolic economy (65). Furthermore, Sacchetti et al. have shown that young cyclists (24.3 ± 5.3 years old) reached their peak efficiency (21.2% ± 1.9%) at 60 rpm, while the peak efficiency (18.3% ± 0.6%) in older cyclists (65.6 ± 2.8 years old) was observed already at 40 rpm (and was not different from that at 60 rpm). Furthermore, the decline in gross efficiency (calculated as the ratio of external work and energy expenditure, i.e. VO_2) with an increase in cadence was significantly more pronounced in older than in young cyclists (66).

1.7.5 Test standardisations

There are a number of aspects to consider with regard to standardisation before, and during, submaximal exercise testing. First of all, it is important that the subject is in a rested state at the time for test (196). The amount of the regulations in training and physical activity before test may vary between individuals. The pre-test conditions shall still be considered and monitored, especially in a test-retest situation. The fundamentals for the recommendations regarding training and activity before test is that the circulatory system has to be in a homeostatic/balanced state, and the working muscles must be well rested and free from previous exercise stresses and lactic acid accumulations.

Another vital aspects to adjust before a submaximal test is food intake, because of the influence on blood flow, HR, blood pressure and cardiac output. The blood flow to the digestive system increases after eating (216). As a consequence of this re-direction in blood flow, there is an increase in HR response to submaximal work (217). The usual recommendations is that subjects can have a light meal ≥ 2 hours before test, or avoid all food consumption within 3 hours before test. Regulations of alcohol consumption may also be appropriate, to avoid that a subject suffers from negative alcohol related issues on test day (such as hangover and altered fluid balance).

The effect of each medication on exercise response and the medication's side effects should be known to the person administering the test (218). For example the β-blocking agents attenuate normal HR and blood pressure (BP) responses to exercise and contribute to fatigue in some people. Furthermore, medications such as bronchodilators and analgesics have peak effect times, which makes it important to ensure that these medications are at peak effect during the test and that this effect is replicated on subsequent test (219). The same principle also applies to the use of nicotine in various forms, since it is well known that both smoking (220) and snuff use (221) have acute effects on submaximal HR. Thus, the regulation/standardisation of the use of nicotine in various forms before submaximal tests is highly important. For example, P-O Åstrand stated that “the subject should not smoke for the last 30 minutes prior to the commencement of the test” (196). Irrespectively of the time span chosen, it is important to be consequent regarding the regulation of the use of nicotine products before tests, especially when it comes to repeated tests. Also, caffeine and coffee intake may need to be restricted before test, due to its central-stimulating effects and influence on HR.

1.8 Observations regarding previous exercise testing research

There are a diversity of areas that need to be covered and considered in order to seize a general understanding of the field of exercise testing. However, there are also a number of issues that complicates the interpretation of the results from previous research.

1.8.1 Insufficient descriptions of test standardisations

The pre-test standardisations and operation procedures (calibration of equipment, execution of test, and more) are often incompletely described in the literature. For example, a submaximal exercise test may be influenced of the timing of food intake (see section 1.7.5), so this aspect has to be addressed in relation to the testing. It is also vital to have information about any restrictions in physical activity and training, smoking and snuff use before tests. The inclusion- and exclusion criteria are not always stated in the previous literature, for example the use of medications that influences HR at submaximal workloads. Furthermore, it is often far from clear whether or not any smokers or snuff users have been included, and if smoking- or snuff-habits were regulated or standardised before test. All of the above-mentioned aspects make it hard to scientifically judge the value of previous studies, and it is also difficult to make fair comparisons amongst previous research on submaximal cycle tests.

1.8.2 Lack of information about the VO_2max estimations

From an overview of previous studies on submaximal exercise tests, it is sometimes unclear what prediction equation that has been used to estimate VO_2max. For example, some studies have examined the validity of the PWC- and YMCA-tests, without giving the exact method for calculations of VO_2max (133, 135). Furthermore, references to original studies do not always contain this information. With regard to the Åstrand test, the test manual recommends exclusion of a first test in order to reduce the effect of stress on the submaximal HR. However, this familiarisation procedure is absent in most studies of validity and reliability for the Åstrand test. It is not always clearly descripted which nomogram that has been used, and whether any correcting factors (for example age, maximal HR) have been applied. Furthermore, different types of ergometers have been used, and these ergometers have probably differed in precision (141, 149). It is important to be aware of the above-mentioned uncertainties when comparing and evaluating the existing literature in the field of submaximal exercise tests.

1.8.3 Differences in methods and procedure

It is useful to have sufficient knowledge of previously used methods for determination of VO_2max, in order to fully understand the historical development of submaximal exercise tests. Hence, to be able to understand and interpret the current literature and previous research on submaximal exercise tests, it is also essential to have basic knowledge about different procedures for direct measurements of VO_2max. All studies uses different arrangements with regard to work mode (for example cycle, horizontal- or uphill running), test protocol, measurement devices, criteria for determination of VO_2max, etc. These aspects have to be taken into account when discussing validity and reliability of a submaximal test. For example, the Åstrand test has been evaluated against reference values (i.e. measured VO_2max) from both treadmill running (139, 143, 222) and cycling (133, 140, 141, 223), which may have influence on the results (see point 1.2.2).

Furthermore, the measurement technology has great influence on the ability to collect reliable ventilatory values during exercise. For example, the issue with reactive effects in the early measurements with the Douglas bag method may have caused low values for maximal V_E, and could have hindered the subject to perform at their absolute maximal level. In the later years, these effects have been minimised. The first Douglas bags also experienced problems with gas diffusion, which may have yielded incorrect VO_2 values (40). The modern computerised apparatuses have more flexible systems for air collection, connected to a comfortable face masks instead of the mouthpiece and nose-clips that are used for Douglas bag measurements.

The development of the bag collecting method also took for granted that the composition of the expired air was representative of the gas exchange that was occurring in the active tissues at the time of gas collection (40). This is also the case, as long as the measurements of metabolic cost are conducted during low-intensity, steady-state conditions. However, the Douglas bag technique have also been used for VO_2 measurements during progressive exercise up to maximal intensity. At the time for occurrence of VO_2max, as well as during anaerobic work, it is most likely a substantial discrepancies between the metabolic activity of the tissues and the composition of the expired air that is collected (40). It is understandable that studies with the Douglas bag technique may present somewhat lower maximal values, compared to the computerised systems with high resolution of the ventilatory measurements, and frequent sampling times. This is especially true for measurements in populations where the subjects are unaccustomed to high intensity exercise. Taken together, the modernisation of measurement equipment and the easier measurement procedures, may have the ability to generate sometimes higher maximal values in certain groups of subjects.

Also, the criteria for achievement of VO_2max also vary considerably among studies. The use of the levelling off criteria as a “golden standard” has been debated, and the supporting criteria/secondary criteria are far from standardised. The reported values can vary from mean of the highest 15 s to 60 s, a wide range that probably has influence of the actual value of VO_2max. These aspects in VO_2max testing further complicate the understanding and interpretation of previous research findings.

Another scientific field that has been through a radical technical development is the HR determinations (224). For example, the measurements by Åstrand in the 1950s were done by palpation over the ictus or the arteria carotis. When tests were conducted on treadmill, the time for the 10 or 15 first beats was taken directly when the subject had stopped exercising. When the subjects were cycling, HR was counted during work (49). The early recommendations for the determination of HR during the Åstrand test was to take the time for 30 heart beats, and then convert that number to beats/min. At this time, technological measurements of HR could be done in the laboratory environment, with the use of a visoscope connected to an electrocardiograph (ECG). However, these electrocardiograms could not be easily collected in large population, and could not be used in field. The development of portable devices for measurement of HR escalated during the 1950s, and the Holter-monitor was the first portable electrocardiocorder with semi-automatic methods for the rapid analysis of the resulting voluminous data (225). It was not until the late 1970s that first wireless HR monitor was invented, and the first wireless HR monitor using electric field data transfer (the Sport Tester PE 2000) was introduced in retail sales in 1983 (226). The introduction of the user-friendly and wireless HR monitors, in contrast to the breakable and cumbersome Holter-monitors, was a revolution in the field of exercise science and training. The modern high-quality devices for analyse of HR data has inevitably enhanced the precision in determination of HR during exercise, both at submaximal and maximal levels.

Lastly, a few studies have doubtful study set-up and test procedures. For example, there is no guarantee that a study is performed in a randomised order, as evident in a work by Faristher et al. They examined different test protocols for determination of VO_2max, and these subjects performed two maximal exercise protocols with one hour rest in-between, in a non-randomised order. A very short-duration and high intensity protocol with frequent and rapid increases in work rate was followed by another all-out effort with increases in work rate every minute, i.e. a test with prolonged duration up to the maximal level. It is not unlikely that such study design may have hindered a higher VO_2max value to occur from the second protocol, due to exhaustion from the first test. Furthermore, 2/3 criteria's for achievement of VO_2max were used as verification, so that the commonly used "plateau in VO_2" may have been ascent. Also, the reported VO_2max values were the highest 15 s (53).

2 AIMS

The overall objective of this thesis was to develop and validate a new submaximal cycle ergometer test for estimation of VO_2max. The specific aims were as follows;

- To construct a new cycle ergometer test for the estimation of VO_2max in adults
- To further develop and cross-validate a new submaximal cycle test and its associated prediction equations for estimation of VO_2max
- To examine the ability of a submaximal cycle ergometer test to detect changes in VO_2max over time
- To determine the applicability and validity of a submaximal cycle ergometer test in pre-pubertal and pubertal adolescents

3 METHODS

3.1 Subjects

All participants in the present research program were recruited via public announcement and word-of-mouth in the region of Stockholm, Sweden. All procedures in the studies were performed in accordance with the principles outlined in the Declaration of Helsinki, and were approved by the Regional Ethics Review Board in Stockholm, Sweden (ref. no. 2008/384-31 for study I and II, 2017/784-31/1 for study III, and 2016/175-31/2 for study IV). All volunteers were provided with written information about the full procedure, including the information that they, at any given point and without any specific reason, could terminate their participation in the study. All subjects signed a written consent at their first visit at the laboratory. When subjects were underage, the parents or legal guardians received written and verbal information about the study, and also signed a written consent. The total number of included subjects in each study, as well as re-recruitment and cross-inclusion of subjects, is shown in figure 6.

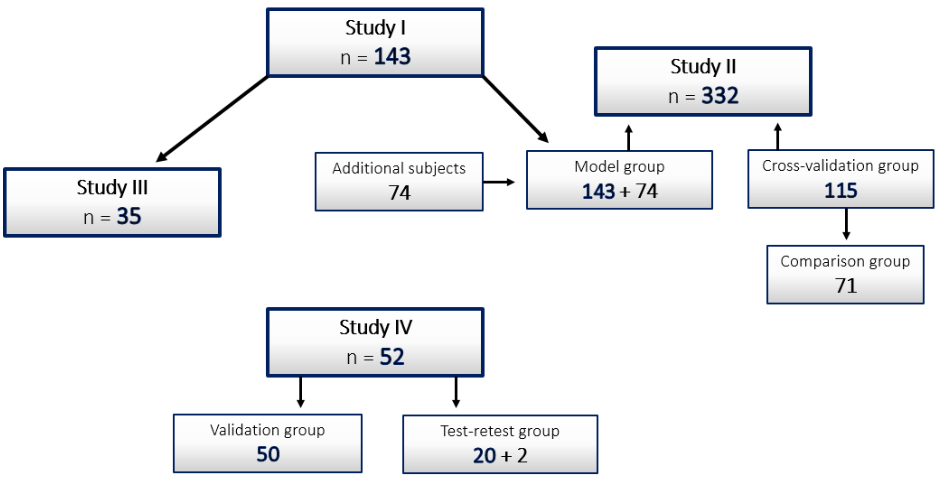

Figure 6. *Schematic model of connections between subjects in the different studies.*

In study I, II and III, the participants were adults with a wide range in age and physical fitness. In study IV, most of the subjects were active or very active adolescents. All included subjects were healthy, with no known diseases that limited the physical work capacity. They also stated themselves as free from ongoing illnesses and infections on the test day. This is a criterion of importance when the exercise testing procedure causes stress on the physiological functions and requires a well-functioning body. Furthermore, an exclusion criteria for all studies was medication that could influence the relationship between HR and VO_2 and the HR response to submaximal exercise. Examples of such medications are β-blocking or β-stimulating agents, known to have influence on HR and the circulatory response to submaximal as well as maximal exercise (117, 227, 228). Furthermore, an additional exclusion criteria in study II were smoking and snuff use.

Study I included men and women in different ages, with a self-reported activity status that ranged from inactive to highly active (145). Six subjects were excluded, due to extreme overweight (n = 1), in-valid cycle tests (n = 2), medication with β-blocking or β-stimulating agents (n = 2), and abnormal changes in the physiological response to the repeated submaximal cycling tests (n = 1).

In study II, a total of 74 subjects were pooled to the model group from study I (see figure 6). These additional subjects had age- and VO_2max values outside the valid range for the prediction equation from study I (145). The additional subjects were 16 men with VO_2max above 4.5 L/min, 11 men with low VO_2max, 43 elderly individuals (≥ 60 years old), and four subjects with the highest and lowest recorded values for age and VO_2max, respectively. The pooled sample was used to create the sex-specific prediction equations (also referred to as EB_{new}) for the EB test. The sex-specific prediction equations were also cross-validated in an external sample. A total of nine volunteers were excluded from the analysis in study II, due to failure to achieve the requirements for acceptance of test (five participants failed the VO_2max test and four participants had non-valid EB test). The corresponding values in the cross-validation group were four excluded volunteers (two with non-valid VO_2max test and two with non-valid EB test).

The subjects in study III were recruited among the individuals that originally participated in the data collection for study I. Hence, these subjects were tested for the first time between the years of 2009 to 2012. Approximately 100 individuals were contacted via email, and a few more received the information about the follow-up tests from colleagues and friends. Three volunteers were excluded, due to onset of medication with β-blocking agents, serious illness, and failure to achieve a valid VO_2max test. All other volunteers were included. The self-reported physical activity status varied from “definitively less trained and/or active” to “definitely more well-trained”.

Inclusion criteria in study IV were pre-pubertal and mid-pubertal boys and girls, 10 to 15 years old. In total, 64 subjects visited the laboratory for the physical exercise tests. Out of those, 50 boys and girls completed the full medical examination including sexual maturity ratings, and provided full data on validity. These subjects where thereby included in the validation group. Furthermore, a total of 22 of the 64 subjects agreed to visit the laboratory for another test session within two weeks, and those subjects where included in the test-retest group (note: two of the boys in the test-retest group lacked maturity ratings, but were still included in the test-retest analysis). Nine of the 64 participants were not assessed for maturity status, and were thereby excluded from the final analysis in the validation group. Furthermore, a total of five subjects were excluded from all analysis: three of these subjects were excluded because they did not meet the criteria for achieving a valid VO_2max, one participant was excluded because of failure to keep the cadence in the submaximal cycling test, and one adolescent was evaluated to be in Tanner stage V.

Table 2. *The subject characteristics for all participants in the four studies. Values are mean (SD), with exception for the value of VO_2max in study III.*

	Sex	*n* =	Age (years)	Height (cm)	Body mass (kg)	VO_2max (L/min)
Study I	Men Women	65 78	41 (12)	173 (9)	70.3 (11.4)	3.2 (0.7)
Study II (model group)	Men Women	117 100	47 (16)	174 (10)	72.9 (12.0)	3.2 (0.9)
Study II (cross-validation group)	Men Women	60 55	41 (17)	174 (9)	73.6 (15.6)	3.4 (1.0)
Study II (comparison group)	Men Women	27 44	37 (13)	172 (9)	70.5 (15.1)	3.2 (0.7)
Study III	Boys Girls	29 23	12 (1)	155 (10)	42.2 (8.8)	2.2 (0.5)*
Study IV (first test)	Men Women	21 14	43 (11)	177 (10)	73.1 (12.3)	3.6 (0.9)
Study IV (follow-up test)	Men Women	21 14	48 (11)	177 (10)	73.5 (12.3)	3.6 (0.9)

**Median and IQR, due to a skewed data distribution.*

Unfortunately, some people in the target group for submaximal exercise tests (older, obese, and people with ongoing medical conditions) are not suitable subjects in the studies in this research project. For example, all people with conditions affecting the heart and central circulation were excluded. A detailed description of contraindications for exercise testing is found in "ACC/AHA 2002 Guideline Update for Exercise Testing" (229). Even though an individual may not exhibited any absolute contraindications for exercise testing (for example unstable angina, uncontrolled cardiac dysrhythmias, or acute systematic infection with fever, body aches, or swollen lymph glands), it is outside the scope of this thesis to examine the validity of the test in diseased and/or medicated populations.

3.2 Equipment

3.2.1 The cycle ergometer

In all four studies, a manually braked pendulum ergometer cycle (828E, Monark Exercise AB, Vansbro, Sweden) was used (see figure 7). The ergometer was operated on a firm and level surface. Zeroing of the scale was performed before each test, while the subject sat on the saddle with the feet on the frame between the pedals and hands on the handle bars. The saddle height and handlebar were adjusted to suit the subject, as long as the preferred settings were within reasonable limits. However, all subjects were advised to adopt an upright position, with the saddle adjusted so that the knees were slightly flexed when the pedals were in the lowest position. The pedals on the ergometer were the original blue Monark 828E pedals, with the accompanying rubber foot straps. All subjects cycled with regular shoes for training or running.

Figure 7. *The Monark 828E cycle ergometer. The photo is published with permission from Monark Exercise AB, Vansbro, Sweden.*

3.2.2 Heart rate measurements

Heart rate was measured with a HR monitor, with a HR sensor chest strap. The HR monitors were model F4 or RS400 (Polar Electro Oy, Kempele, Finland). Measurements of HR were recorded in 5 s epochs. The two models are based on the same technology and are comparable.

3.2.3 Oxycon Pro

All ventilatory measurements were conducted with Oxycon Pro (Erich Jaeger GmbH, Hoechberg, Germany). Before each test session, ambient conditions were measured with the portable humidity and temperature instrument HygroPalm 0 (Rotronic, Crawley, West Sussex, UK). Calibration was performed with the built-in, automated calibration procedures, including control of low and high flow and gas analyse calibration with a high-precision gas mixture of 15.00% O_2 and 6.00% CO_2 (Air Liquide AB, Kungsängen, Sweden), and ambient indoor air. The calibrations were performed in the same environmental settings as for the experiments.

The Oxycon Pro offers air collection with both breath-by-breath and mixing chamber mode. The breath-by-breath method delivers continuously measured expired gas concentrations and ventilation, just outside the mouth (46). Previous studies have reported that the Oxycon Pro is an accurate system for measurements of metabolic parameters, with a CV of around 5% over a wide range of workloads and O_2 uptakes (42, 47). The use of Oxycon Pro with mixing chamber may yield even lower CV, $\leq$ 1.4% (46). However, the latter validation study used a procedure of serial measurements with the Douglas bag method and the Oxycon Pro. In contrast, the studies of the breath-by-breath method used repeated measures, which also includes biological variations.

The ventilatory measurements were conducted with the breath-by-breath method for most of the test sessions, since the equipment is slightly more comfortable for the subject during maximal running. Furthermore, the mixing chamber may be unsuitable to use during test sessions with many consecutive tests, due to the need of constant dryness in the tube and the mixing chamber (46).

Figure 8. *The 7450 Series V2 Mask™ (Hans Rudolph, inc), connected to the flow meter for breath-by-breath measurements with the Oxycon Pro.*

Measurements of gas exchange and ventilatory variables were averaged in 5 s epochs for the maximal tests, and in 10 s epochs for the submaximal tests. All data from the measurements of VO_2 during the submaximal work rates are derived from mean values between the third and fourth minute (i.e. six consecutive 10 s epochs) of work. This time is sufficient to elicit steady-state values at submaximal work rates, as previous research have shown that VO_2 increases exponentially from rest to a steady-state value in about 2–3 min (72, 230). For all maximal tests, VO_2max was referred to as the mean of the highest consecutive values over 30 seconds (i.e. six consecutive 5 s epochs).

3.2.4 RPE scale

Perceived exertion was rated using Borg's scale for rating of perceived exertion (RPE) scale. The scale is a 15-point category scale (ranging from 6 to 20), constructed to increase linearly with increased intensity in workload (231). The scale values are from 6 to 20, values that denotes HR ranging from 60–200 beats/min. This construction also leads to a practical situation where 13 on the scale would match a HR of approximately 130 beats/min, for subjects between 30-50 years of age. Hence, the ratings are highly correlated to exercise HR (231, 232).

The RPE scale has been shown to be a useful tool to assess perceived exertion during exercise testing, and for prescriptions and adjustment of exercise intensity in sports and rehabilitation. The scale numbers are supported by a brief description of the corresponding subjective strain at that level, for example the term "no exertion at all" (rating 6), "somewhat hard (rating 13) and "maximal exertion" (rating 20). The Borg's RPE scale was introduced and described to the participants. Furthermore, the subjects were asked to rate their perceived exertion in the resting state, and/or in some made up situation (for example "which rate would you have during an easy walk in the park"), in order to fully understand the use of the scale and the ratings.

3.3 Pre-test procedures

The basic pre-test procedures were the same for all studies. All participants were asked to refrain from vigorous physical activity the day before and on the test day, and to not consume a heavy meal less than three hours before the test. All participants filled out formulas about health history, illnesses and medication status. Height (to the nearest 0.1 cm) and body mass (to the nearest 0.1 kg) were measured with the participant wearing light clothing and no shoes. All test sessions were preceded by a brief health check with measurement of resting HR and resting blood pressure.

3.3.1 Tanner classification

In study IV, all participants underwent a more extensive medical examination, conducted by a medical doctor or nurse. This examination also included visual assessment of pubertal maturity with the five-point Tanner scale (233). The determinations of pubertal maturity allowed Tanner stage classification of all subjects, according to the indices developed by Tanner (233-235). Classifications were made for pubic hair rating (for both boys and girls), genitalia rating (boys), and breast rating (girls). As pubic hair rating was the only common value for all subjects, this value was used for further analyses. All children were able to choose for themselves if a male or female nurse/doctor should conduct the examination, and legal guardians were invited to accompany the child if they desired.

3.3.2 Familiarisation tests

In study I, all participants visited the laboratory on two occasions, on the same weekday and at the same time, to perform submaximal and maximal testing. The tests were conducted with one week in-between. A subsample (n = 49) visited the laboratory only once, and without measuring actual VO_2 during the submaximal test. However, analysis showed no significant difference in age (men), height, body mass, measured VO_2max, and maximal HR between this subsample and the rest of the study population.

In the thesis of Ekblom-Bak, it is stated that analysis of data from the repeated VO_2max tests in study I revealed that not all participants fulfilled the criteria for achievement of VO_2max during both test sessions. Furthermore, 59 participants with two acceptable measurements were analysed with regard to variation between the first and second test session. There was no significant difference between the achieved VO_2max values (3.49 L/min and 3.50 L/min, respectively). The CV was 2.7%, which is regarded to be a low value, and in concordance with internal laboratory measurements and other previous reports (69). Therefore, the first accepted VO_2max value was used in the analysis. Also study IV comprised of two sessions. The main reason for that was the aim to examine test-retest reliability, but it also made it possible to evaluate the potential need for a familiarisation session when using the EB test in adolescents.

3.4 Submaximal cycle ergometer tests

The physical tests were performed in laboratory environment with a normal ambient climate. Before the submaximal cycling commenced, all participants had at least 10 min of calm activities and 10 min of relaxation in conjunction with the resting measurements.

3.4.1 Development and validation of a new cycle test

The cycle tests in study I were conducted with two different versions of the test protocol, due to experimental reasons. After reviewing a low number of subjects, we set the protocol to include two work rates. The first stage was 0.5 kp and 60 rpm – also referred to as “standard work rate”. The standard work rate was immediately followed by two higher and individually chosen work rates, conducted at a pedalling rate of 60 rpm. Later, the protocol for most of the subjects comprised of continuous cycling with a pedal frequency of 60 rpm, where the first load was the standard work rate (0.5 kp), followed by two higher submaximal work rates. The higher rates of work were individually chosen to obtain an RPE of 12–13 on the Borg’s scale for the first rate, and 14–15 for the second rate. Tests resulting in RPE above 16 were excluded due to suspicion of effects of anaerobic energy production on HR and mechanical efficiency.

In study II, all participants visited the laboratory on one occasion and performed two consecutive submaximal workloads on the cycle ergometer, i.e., the EB test. The test is described in detail in point 3.4.2. With regard to the test protocols in study III, some of the subjects were unable to exercise at the same high work rate as in the baseline test (5–8 years earlier). Before the start of the follow-up test, all subjects were asked about their current physical status. The subjects who self-reported a great change in physical fitness were assessed to cycle on an adjusted high work rate, compared to the first test. In total, 71% of the subjects cycled at the same high work rate at baseline and follow-up, 11% (4/35) cycled at a work rate that was 0.5 or 1.0 kp lower, and 17% (6/35) cycled at a work rate that was 0.5 or 1.0 kp higher.

In study IV, all subjects and their legal guardians were introduced to all procedures and the cycling on the ergometer. Individual adjustments of saddle height and handle bar were done for each adolescent. This was followed by a brief session (1-3 min) of unloaded cycling at 60 rpm, so the subject got familiar with the cadence of the EB test. Thereafter, the subject were introduced to the Borg’s RPE scale, and equipped with a child-specific facemask, connected to the on-line metabolic system. All adolescents performed an EB test, as described in 3.4.2. However, all subjects were smaller and younger than the population that the test is originally developed for. Therefore, all participants initially cycled at 1.0 kp on the second (high) work rate. Only a few individuals were strong enough for an additional increase in work load.

The HR and VO_2 were measured continuously during the physical tests.

3.4.2 Description of the EB test

The EB test is a two-point submaximal cycle test, comprised of two different work rates. The test is conducted with a pedaling rate of 60 rpm. The first work rate is 4 min at a standard work rate of 0.5 kp, which is equal to 29 W when 1 W = 6.116 kpm/min (without taking into account the additional work of 8% due to frictional losses at the chain and drive train. The standard work rate is immediately followed by 4 min of cycling on a higher work rate (145). The higher work rate is individually chosen to induce a steady-state HR above 120 beats/min (or, above 110 beats/min for older subjects) and a perceived rate of exertion of ≈ 14 according to the Borg RPE scale (231), corresponding to 50-70% of maximal capacity. After the first minute of the higher work rate, subjects are asked to state their level of perceived exertion on the Borg scale and the increase in HR is controlled. If the work rate seems to be too easy (≤ RPE 11), an increase of 0.5 kp is done and the subject cycles for another 4 min on the new work rate. HR is measured continuously during the test, and mean HR is calculated from the last minute on each work rate, respectively. The EB test is developed and validated for use on a mechanically braked cycle ergometer, like the model 828E (Monark Exercise AB, Vansbro, Sweden) in figure 7. It is unclear to which extent a change of ergometer type may affect the validity of the test. A point-by-point instruction for the test procedure is given below.

1. Calibrate the ergometer according to standard procedures. The 828E shall be used, and calibrated, and the ergometer should stand on a level, firm foundation. With the subject mounted, but not touching the pedals, adjust the "0" mark on the scale so that it coincides with the mark on the pendulum weight. This setting must be made accurately if the load is to be precisely set.
2. Ensure that the subject has followed conventional pre-test standardisations (comments on this is found in section 1.7.5).
3. Adjust seat and handlebar to suit the subject, as long as the settings are within reasonable limits. The most comfortable position, and in the case of very heavy work the most effective one, is the saddle height that gives a slight bend of the knee-joint in the lower position when the subject has the front part of his foot on the pedal (i.e. with the front part of the knee straight above the tip of the foot).
4. Introduce Borg's scale for ratings of perceived exertion (the RPE scale). Make sure that the subject has understood the scale by asking about the rating of perceived exertion while the subject is sitting inactive on the ergometer.
5. Before the test, estimate a suitable higher work rate to allow the individual to reach a heart rate in the range 120-150 beats/min (for individuals < 50 years) and 110–140 beats/min (for individuals ≥ 50 years), respectively, aiming at a rated perceived exertion of ≈ 14 according to the RPE scale.
6. Work is started with a slack brake belt. Thereafter the belt should be stretched with the aid of the hand wheel until the required work load is obtained. Start standard work rate pedalling for 4 min at 60 rpm and resistance of 0.5 kp. Check each minute that both pedalling speed and resistance are kept constant.

7. Measure average heart rate during the 4th min by taking notes of the heart rate at four occasions (3.15, 3.30, 3.45, and 4.00) and average these.
8. Increase resistance to the higher individual work rate (point 5 above). Check each minute that both pedalling speed and resistance are kept constant.
9. Ask for the subjects RPE during the 2nd min at the higher work rate.
10. If RPE is < 10, increase resistance with 1 kp and repeat point 9. If the RPE is 10-11, increase with 0.5 kp and repeat point 9. With a RPE of 12–16, maintain the work rate and go to point 11. If the subject points at a RPE of 17 or higher, stop the test and let the subject rest for 20 min before performing a new test at a lower rate. However, it is preferable to cease testing and perform the test another day.
11. Measure average heart rate during the 4th min at the higher work rate by taking notes of the heart rate at four occasions (3.15, 3.30, 3.45, and 4.00) and average these.
12. After completion of the test, ask for the RPE for the 4th min at the higher rate.

The variables age, sex, HR and work rates are used in the equations for prediction of VO_2max.

3.4.3 The Åstrand test

For comparative reasons, the HR from one of the higher work rates in study I was also applied to the Åstrand nomogram, and adjusted with associated age-correction factors (109). The Åstrand test was originally based on a pedalling rate of 50 rpm. However, Jessup et al. found no difference in predicted VO_2max by the Åstrand test comparing 50 rpm (correlation to measured VO_2max of 0.64) and 80 rpm (correlation of 0.63), respectively (223). Negligible differences in mechanical efficiency between 50 and 60 rpm are also reported (212, 213). In line with the test prerequisites, a submaximal pulse rate within the range of 120 to 170 beats/min was accepted. However, for some of the older participants with low maximal HR, a HR of ≥ 110 beats/min was accepted, as they reported a RPE of 14-15 on the RPE scale even at these levels. In total, 12 participants were excluded from the Åstrand test analysis, as their HR during the submaximal test was too low (< 110 beats/min).

The same procedure was applied in study II and III, i.e. the HR from the high work rate in the EB test was applied to the Åstrand nomogram. In study II, 44 of the subjects were excluded from the Åstrand analysis because their HR during the submaximal test was too high (> 170 beats/min) or too low (< 120 beats/min), or because the subject had a measured HR that was outside the valid range for estimated VO_2max from the coincident work rate. For the same reasons, 15 subjects were excluded from the Åstrand analysis in study III. Lastly, eight adolescents were excluded from the Åstrand analysis in study IV, due to a submaximal HR that was either too high or too low, and in one case the estimated VO_2max (< 1.5 L/min) were outside the range of valid VO_2max-values in the nomogram.

3.5 Maximal treadmill running tests

All reference values, i.e. the direct measurement of VO_2max, were derived from maximal running tests. The reason for that was the aim to capture the true VO_2max of an individual, and compare this value to the estimations from a submaximal cycle test. During the submaximal test, an estimation of VO_2max is based on the regulation of the circulatory response to a given workload. Even though there might be some differences for running and cycling with regard to the absolute values of VO_2 at a certain HR, this has little importance for the overall predictive capability of the cycle test. Thereby, a graded treadmill running test to voluntary exhaustion was rational to use, in order to elicit a whole body maximal exhaustion in different subjects from a mixed population. The protocols for the VO_2max test included increases in both speed and incline, in order to avoid extremely high velocities and facilitate whole-body work with large muscle groups. The inclined running, or inclined brisk walking, is an activity that most people are somewhat familiar with. The overall aim with the maximal tests was to come as close as possible to a levelling off, caused by maximal cardio-respiratory stress. Direct measurements of ventilatory variables were conducted with a computerised metabolic system, and HR was registered with a HR monitor. See point 3.2 for detailed descriptions of the equipment.

3.5.1 Introduction and warm up

All treadmill-running tests were preceded by at least five minutes of warm up on the treadmill, to familiarise the subject with the work mode and equipment. All running tests were conducted on a treadmill from Rodby Electronics (Vansbro, Sweden). In studies involving children, older and/or untrained people, some subjects experienced their first session of treadmill running. The pre-test warm up procedure, with a careful and meticulous introduction to treadmill running, was of great importance. The procedure was also a possibility to assess the subjects' capacity, and design an individual test protocol. All test protocols were aimed to elicit a test-duration of 6-10 minutes, before subjective exhaustion. After the initial warm-up, all subjects were instructed how to grab the handlebars and jump off the treadmill. This safety procedure was practiced until the subject felt comfortable. Furthermore, a number of high intensity work bouts of approximately 30-45 s were included in the late phase of the warm-up, in order to prepare the subject for the high intensity of the maximal test. Lastly, the velocity were decreased to a comfortable walking or jogging pace, and the elevation was increased to 3-6°, so the subject could practice how to jump off the elevated treadmill. If needed, the warm up was followed by a brief rest (≤ 5 min), sometimes including intake of small amounts of water, some pre-conditioning stretching, or other individual preparation that the subject wished for.

3.5.2 Incremental tests

All participants performed an incremental treadmill test for determination of VO_2max test. The test started off with 1° incline and the same speed as the subject felt comfortable with during the warm up (usually a speed corresponding to approximately 60-65% of VO_2max). Then, speed and/or incline were increased every minute until volitional exhaustion. All test protocols were individually designed with regard to starting level, as well as size and frequency of the increases in workload. The workloads were selected to elicit a total exercise time of 5-8 min before exhaustion occurred. All subjects were given extensive verbal encouragement in order to attain their maximal level and achieve their "true" VO_2max. After completion of the test, subjects were asked to rate their perceived exertion for breathing, legs and whole-body exhaustion, using the Borg RPE scale (231).

Predetermined criteria for acceptance of the test were used to determine whether or not the subject attained a "true" VO_2max during the incremental test, see point 3.6.2. The methods and procedures for the incremental tests were the same for both adults and adolescents, whereas the tests of the adolescents were judged with somewhat different criteria.

In total, 10 of the 35 participants in study III were tested with different work rates at the second step of the test at the follow-up test 2017, compared to their first test 5 to 8 years earlier. This was due to individual changes in physical fitness, which in many cases made the old test protocol unsuitable to the subject. Hence, all protocols in study III were designed with respect to the current physical status of the subject. The acceptance of a test as VO_2max was done with the same criteria as for earlier tests, and the mean values of VO_2max were calculated according to the same procedures.

3.5.3 Supramaximal tests

In study IV, the first incremental running test was followed by approximately 10 minutes of rest. Thereafter, an additional treadmill test was performed. The second test was a supramaximal test, where the supramaximal phases comprised of a speed and incline 1 km/h or 1 degree above the intensity where the previous treadmill test was terminated. The main purpose of the double tests was to increase the possibility for the child to achieve a true VO_2max value.

Two different protocols for the supramaximal exercise testing were examined. All supramaximal tests started with 2 minutes warm-up at the same speed as was used at the first level (first minute) of the incremental test. Then, in one of the protocols (supra I), the intensity was directly increased to a supramaximal load. The other protocol (supra II) comprised of an intermediate stage with two minutes of running at a moderate "threshold-intensity", immediately followed by the supramaximal load. In both protocols, the supramaximal load was set to one level above (in incline by 1º, or in speed by 1 km/h) the level where the subject stopped at the incremental test (236). In total, 45 of the 52 included adolescents in study IV agreed to perform the supramaximal test within 10 min after the incremental test. The supra I-protocol was performed by 24 subjects, and the supra II-protocol was used for 21 subjects.

3.6 Analysis of data

3.6.1 Submaximal tests

In study I, the values of submaximal HR, VO_2, RER and VE were analysed as the mean value from four consecutive 15 s epochs. The mean values for the ventilatory variables in study II, III and IV were derived from six consecutive 10 s epochs, and the HR values were the observed and noted HR at 3:15, 3:30, 3:45, and 4:00 min at each work rate (as described in section 3.4.2).

3.6.2 Maximal tests

The reported maximal HR is the highest observed value during the maximal test, regardless of the time point for the occurrence of this HR (i.e. the highest observed HR must not co-exist with the highest achieved VO_2max). In study I, maximal HR was calculated from the mean values during the highest 30 s registered. In study II–IV, the maximal HR was referred to as the peak value (the highest registered HR for a single 5 s epoch).

Criteria for acceptance of the VO_2max measurement were levelling off of VO_2 despite an increase in speed or incline, a RER > 1.1, RPE above 16, work time above 6 min, supported by a maximal HR within ± 15 beats/min from age-predicted maximal HR (113). A test was accepted as VO_2max when a minimum of three out of the five criteria was achieved. The methods and procedures for the incremental tests were the same for both adults and adolescents. Due to the controversy regarding the ability in children to reach a plateau in VO_2max, somewhat different criteria were used for determination of the maximal values in children. The criteria for achievement of VO_2max from the incremental test in adolescents were a maximal HR > 190, RER > 0.95, Borg rating > 17 in at least one variable, running time > 5 min, and obvious signs of exhaustive/maximal effort. The test was only accepted if all of the first four variables were fulfilled.

The adolescents also performed an additional verification test of VO_2max. Criteria for acceptance for the supramaximal test were a maximal HR ± 4 beats/min compared to the incremental test, RER > 0.95, ratings on the Borg scale >17 in at least one variable, and VO_2peak within 0.2 L/min compared to the incremental test. A supramaximal test was accepted if three of the four variables were confirmed. In case a subject had a supramaximal HR or VO_2 that was higher than the above-mentioned criteria's, the incremental test were dismissed, and the highest recorded VO_2 value from the supramaximal test substituted the incremental VO_2peak value. In both adults and children, VO_2max was defined as the highest recorded oxygen uptake during 30 consecutive seconds/the mean of the six highest consecutive 5 s epochs were regarded to be the maximal value.

3.6.3 Statistical analysis

The assumption of normality were tested for all variables, using the skewness test for normality. Parametric statistical analysis were employed to calculate the means and standard deviation (SD). The continuous descriptive characteristics are presented as means ± SD, if not some other is stated. The descriptive statistics for VO_2max in study IV are presented as median and 25^{th}–75^{th} percentiles (or median and interquartile range, IQR), because the values in L/min were skewed. For comparison between values ("two cases with a single variable") the Students´ paired t-test was used. Pearson's coefficient of correlation (r) and R^2 were used to analyse the association between variables. The R^2 – the determination coefficient – is a measure of how close the data are to the fitted regression line. In general, the higher the R^2, the better the model fits the data. Adjusted R^2 was used when comparing the explanatory power of regression models that contained different numbers of predictors.

Repeated measures ANOVA were used to test differences between changes over time in study I and in study III. The standard errors of the estimate (SEE) were obtained by linear regression (enter mode), by entering measured VO_2max as the dependent variable and estimated VO_2max as the independent variable. The SEE is a measure of the accuracy of predictions made with a regression line. A large SEE is the same thing as a lot of variability in the outcome variable, i.e. measurement in different samples would result in different mean values. A small standard error would mean that the population is more uniform, so your sample mean is likely to be close to the population mean. In study III and IV, measurement errors were evaluated as mean differences and limits of agreement (LoA) and/or measures of systematic error (mean- or median difference, 95% CI). The LoA shows the interval where 95% of all the individual values will be found, or expressed differently, the interval within which a new individual observation will appear with 95% probability.

In study I and II, a multiple linear regression with forward selection was used to include independent variables in the VO_2max estimation models, in order of significance (probability of F = 0.05 for entry, and 0.10 for removal). The final models was checked for equal variance. CV was calculated as the ratio between the standard deviation of the difference between estimated and measured VO_2max, and the mean measured VO_2max.

That makes the CV is a measure of relative variability, and the value is given in percentage (%). A low CV is a sign of smaller variability for the examined variable.

For the reliability analysis, paired sample t-tests were used to compare the absolute HR and ΔHR for the first and second occasion at the standard and higher work rate, respectively. Correlation analysis was used to analyse the influence of maximum HR, deviation from age-predicted maximum HR (220 beats/min - age), and VO_2max for the difference between measured and estimated VO_2max.

In study III, bivariate correlations were performed for all individual variables of the EB test prediction equation, as well as for sex, change in maximal HR (ΔHRmax), change in body mass, change in HR and change in VO_2 at standard work rate, respectively. Linear regression models with enter method were used to identify significant predictors for the EB test measurement error. Correlation analysis was also used to evaluate the association between the error in the estimated VO_2max (i.e. the difference between measured and estimated VO_2max) from the first and the follow-up test, respectively.

The statistical analyses in study I were conducted using version 19.0 of the SPSS software package (SPSS Inc., Chicago, Illinois, USA) and version 12.2.1.0 of MedCalc. The SPSS statistical software version 21.0 (SPSS Inc., Chicago, IL, USA) was used for the statistical analyses in study II. Study III and IV, statistical analyses were performed with SPSS statistical software version 24.0 (SPSS Inc, Chicago, Illinois, USA). Statistical significance was set at $p < 0.05$ for all analyses, in not otherwise stated.

4 RESULTS

An overview of the validity and reliability for the EB test in study I–IV is presented in table 3.

Table 3. *Summary of validity and reliability for the Ekblom-Bak test. The measurement error is calculated as estimated – measured VO_2max in L/min. Correlations are for VO_2max in L/min.*

Study	Subjects	Measurement error Mean (SD)	Correlation to measured VO_2max	Other analyses
I	143 adults 21–64 years	-	r = 0.91, adjusted R^2 = 0.82	CV = 9.3%, SEE = 0.30 L/min, test–retest: CV 6.2%
II	115 adults[*] 20–86 years	0.02 (0.32)	r = 0.95, adjusted R^2 = 0.90	CV = 9.4%, SEE = 0.30 L/min, LoA = -0.64 to 0.61
III	50 youths[**] 10–15 years	0.11 (0.35)	r = 0.86, adjusted R^2 = 0.73	CV = 14.7% SEE = 0.29 L/min, LoA -0.59 to 0.76 test–retest: CV 3.1%
IV	35 adults[†] 29–72 years	-0.17 (0.34)	r = 0.93, adjusted R^2 = 0.86	CV = 9.4 % SEE = 0.33 L/min, LoA = -0.83 to 0.50

[*] *Values from the cross-validation group.* [**] *Values from the validation group.*[†] *Values from the follow-up test.*

4.1 Study I and II

A new submaximal cycle ergometer test for estimation of VO_2max was presented in study I, and associated prediction equations were presented in study I and II. All data used for model construction and calculation on test validity were collected from the first test session in these studies. However, in study I, the submaximal test on the firsts visit was not valid for 15 participants, due to reported RPE > 16 (n = 4), illness (n = 4), failure to follow the standardised procedure before the test (n = 3), and abnormal ambient climate during the test (n = 4). Hence, for these participants, the data from the second test occasion was used. In study II, 24 of 217 subjects in the model group had their estimated VO_2max from the second test occasion, while all the others were calculated from their first test.

4.1.1 Model construction (Study I)

The variable ΔHR was calculated as the difference between HR at the higher, individually chosen work rate and HR at standard work rate, i.e. HR high - HR standard. To standardise ΔHR, and make it comparable between individuals cycling at different final work rates, it had to be related to the increased VO_2 demand (ΔVO_2). However, due to technical hitches with direct VO_2 measurements in clinical and practical use, the test prediction equation includes the delta work rate (delta power output, ΔPO) as a proxy for ΔVO_2. The ΔPO was calculated as the difference between the higher work rate and the standard work rate, and the individually measured ΔVO_2 was plotted against ΔPO. Additionally, a ΔHR/ΔPO score was calculated for each participant. Further details about the variables included in the model construction can be found in the thesis "Physical activity, cardiorespiratory fitness, and abdominal obesity in relation to cardiovascular disease risk – epidemiological studies" (237) and in paper I (145).

4.1.2 Prediction equations (Study I and II)

The prediction equations for the EB test were developed with a multiple linear regression model, where the forward method was used to include independent variables, in order of significance, into the VO_2max estimation model (probability of $F = 0.05$ for entry, and 0.10 for removal).

The original prediction equation for estimation of VO_2max (also referred to as EB_{2012}) included the variables ΔHR/ΔPO, sex and age as independent variables:

$$VO_2max = 4.98196 - 2.88618(\Delta HR/\Delta PO) + 0.65015(sex) - 0.01712(age)$$

where codes for sex male = 1 and female = 0.

Analyses showed that the theory behind the estimation from ΔHR, and further the variable ΔHR/ΔPO, was a good prognostic tool for low or high VO_2max. The higher VO_2max, the lower ΔHR and ΔHR/ΔPO. Hence, the slope of individual relation between HR and PO is lower for a subject with high VO_2max than for a subject with low VO_2max. The correlation between VO_2max and ΔHF/ΔPO is shown in figure 9.

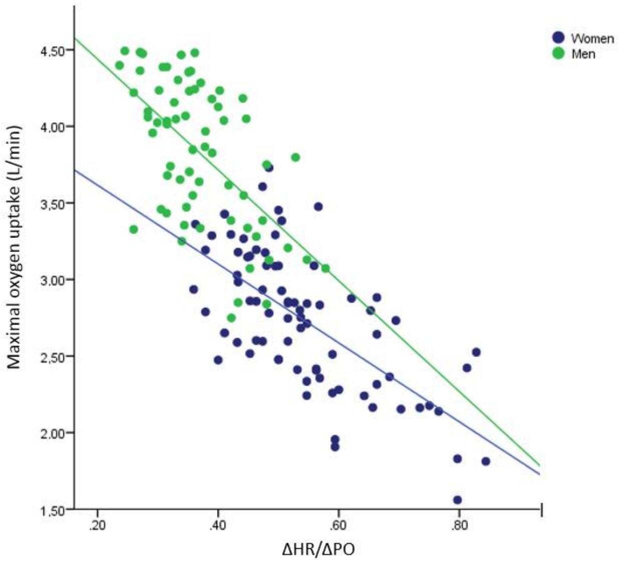

Figure 9. *Sex-specific relations between ΔHR/ΔPO and VO_2max. Data from study I.*

Furthermore, the correlations within each sex were stronger than the correlation for the whole group of subjects (see figure 9), and the prediction error increased with higher VO_2max values. The EB_{2012} model underestimated subjects with very high VO_2max and overestimated those with lower VO_2max, especially men (145). It was not possible to create sex-specific prediction equations because the representation of each sex was too low in some of the age-categories. Based on these shortcomings of the original model, there was a need for a new sex-specific algorithm for estimation of VO_2max from the EB test. The sex-specific prediction equations were presented in study II.

The first step in the modifications of the model was the inclusion of more subjects in the study population. Thereby, it was possible to create sex-specific prediction equations, and enhance the valid range of age and VO_2max. The sex-specific prediction equations for the EB test were developed with a multiple linear regression, where the forward method was used to include independent variables, in order of significance, into the VO_2max estimation model (probability of $F = 0.05$ for entry, and 0.10 for removal). Dependent variable was lnVO_2max, and the independent variables were age, ΔHF/ΔPO, ΔPO and HR at standard work rate. The reason for inclusion of the HR at standard work rate and ΔPO was a high correlation to measured VO_2max. The use of lnVO_2max was motivated through the better fit (see figure 10) and avoiding extrapolation to infinity.

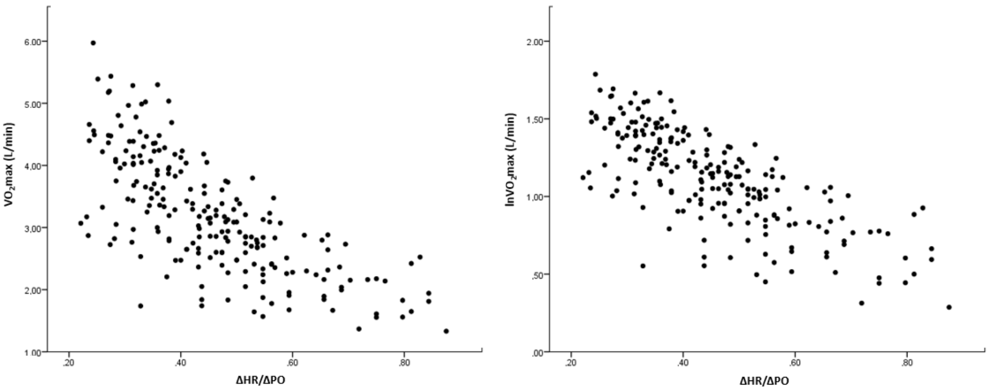

Figure 10. *Relation between ΔHR/ΔPO and VO_2max (left) and $lnVO_2max$ (right). The linear relation was slightly higher for $lnVO_2max$, but no formal interaction test was performed. Data from the cross-validation group in study II.*

The final sex-specific regression models included age, change in HR per unit change in power (ΔHR/ΔPO), the difference in work rates (ΔPO) and HR at standard work rate as independent variables. The EB_{new} prediction equations for men and women, respectively, were:

<u>Men</u>: $lnVO_2max$ = 2.04900 - 0.00858(age) - 0.90742(ΔHR/ΔPO) + 0.00178(ΔPO) - 0.00290(HR at standard work rate)

<u>Women</u>: $lnVO_2max$ = 1.84390 - 0.00673(age) - 0.62578(ΔHR/ΔPO) + 0.00175(ΔPO) - 0.00471(HR at standard work rate)

4.1.3 Validation and cross-validation (Study I and II)

In study I, the mean measured and estimated VO_2max were both 3.23 L/min (SD 0.72 and 0.65, respectively). The internal validation revealed an over-all standard deviation (SD) of 0.30 L/min (0.33 L/min for men and 0.28 L/min for women) for the difference between measured and estimated VO_2max. The coefficient of variance (CV) for the EB_{2012} was 8.5% for men and 10.3% for women, respectively.

In study II, the values in the cross-validation group were a measured VO_2max of 3.37 (0.97), and an estimated VO2max of 3.39 (1.02) L/min, respectively. The mean difference between estimated and measured value was 0.02 L/min (95% CI -0.04 to 0.08). The male subjects were overestimated by 0.11 L/min (95% CI 0.02 to 0.20) and the females were underestimated by -0.09 L/min (95% CI -0.16 to -0.01). The CV for the EB_{new} was 8.3% for men and 10% for women, respectively.

Further, we compared the precision for the EB_{2012} and the EB_{new} prediction equations in a sub-sample of 71 subjects from the cross-validation group. In this sub-sample, the mean (SD) measured VO_2max was 3.17 (0.67) L/min. The corresponding estimated values were 3.14 (0.71) L/min for the EB_{new}, and 3.30 (0.66) L/min for the EB_{2012}, respectively. Hence, the calculations of estimated VO_2max with the EB_{2012} prediction equation resulted in a group mean difference of 0.12 L/min (95% CI 0.04 to 0.21**)** between estimated and measured values. The CV for the entire sample was 10.9%. In comparison, the EB_{new} resulted in a group mean difference of -0.03 L/min (95% CI -0.10 to 0.04), with a CV of 9.4%.

There was a correlation between the estimation error and VO_2max level (especially among men). However, this correlation was lower when the estimations were derived from the sex-specific prediction equations (see figure 11). This reduction may be partly explained by the inclusion of two new variables (ΔPO and HR at standard work rate) in the prediction equation.

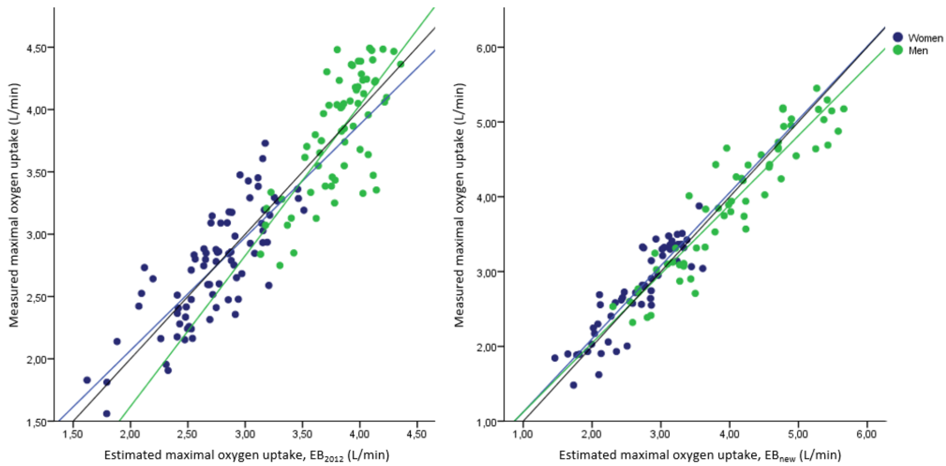

Figure 11. *The relations between measured maximal oxygen uptake and estimated maximal oxygen uptake from the EB_{2012} prediction equation (left) and the sex-specific EB_{new} prediction equations (right). Data in the left picture is from study I (n = 143) and data in the right picture is from the cross-validation sample in study II (n = 115).*

With regard to external validity and generalizability, the first prediction equation was tested for validity in a mixed population of men and women, 21–65 years old with VO_2max values from 1.56 to 4.49 L/min (1.56 to 3.73 L/min for women and 2.75 to 4.49 L/min for men, respectively). The validity for the sex-specific prediction equations from study II were tested in women from 21 to 86 years old, with a VO_2max of 1.33–3.94 L/min (18.9–61.9 mL/kg/min), and in men from 20 to 84 years old, with a VO_2max of 1.67–5.97 L/min (23.5–76.4 mL/kg/min) respectively.

4.1.4 Reliability (Study I)

The test-retest analyses were performed on the data from 57 participants with two valid submaximal tests, at the same workload on both the first and second occasion. The test–retest CV was 6.2%. The absolute HR at both standard and higher work rate was significantly higher for the first occasion compared with the second occasion. However, mean ΔHR for the first occasion and mean ΔHR for the second occasion were correlated with no mean difference. The reliability analysis revealed a high agreement between estimated and measured VO_2max at both test sessions, and no mean difference in the estimated VO_2max values. In general, the HR response was higher at both work rates during the first test compared to the second. However, this did not influence the predictions of VO_2max because the ΔHR was similar at both occasions.

4.1.5 Comparison of the validity for the Åstrand test and the EB test

Results from the comparison group in study II revealed that there were larger systematic errors and variations in the VO_2max estimation from the Åstrand test, compared to the EB test. The adjusted R^2 for the Åstrand test was 0.50. Mean estimation error was -0.07 (95% CI -0.21 to 0.06) L/min, which was significantly higher compared to the results from the EB test with the sex-specific prediction equations (-0.03 L/min, 95% CI -0.10 to 0.04). The SEE for the absolute VO_2max was 0.50 L/min, and the SEE for relative VO_2max values were 5.6 mL/kg/min.

The CV for the Åstrand test in the total sample was 18.1%. Consequently, for subjects with a measured VO_2max of 3.0 L/min, 95 out of 100 tests would generate an estimated VO_2max within ± 0.9 L/min. For the EB test, the CV is ~9%, and with a measured VO_2max of 3.0 L/min, 95% of the observations would instead be expected to fall within ± 0.55 L for the same subjects. With lower values for VO_2max, the margin of error will be smaller. For subjects with an actual VO_2max of 2.0 L/min, VO_2max can be predicted within ± 0.36 L by the EB test, and ± 0.71 L by the Åstrand test, respectively.

In study I, the test-retest CV for the Åstrand test was 9.8%, and the corresponding CV for the EB test was 6.2%. Reliability analysis revealed a high agreement of estimated VO_2max between the first and second occasion, but with a significantly higher mean estimated VO_2max from the Åstrand test for the second occasion.

As mentioned in the introduction, the theory behind the Åstrand test is that 50% of VO_2max is achieved at a HR of 128 beats/min for men, and 138 beats/min for women. This observation held true in the sample of young, healthy subjects that formed the basis for the development of the Åstrand test. However, this relation is highly dependent on age and concomitant age-related variation in maximal HR. Figure 12 is based on data from 18 men and 18 women (20–56 years old, VO_2max 1.96–5.3 L/min), randomly selected among all 20–60 year olds in study II. The figure shows the relationship between maximal HR and the actual % of VO_2max at an absolute HR of 128/138 beats/min.

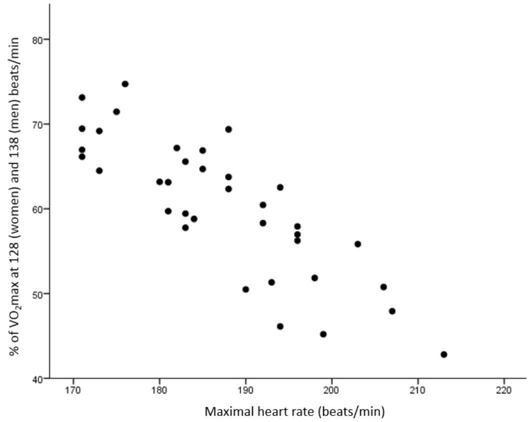

Figure 12. *The relation between maximal HR and the proportion of VO_2max at a HR of 128 beats/min in women and 138 beats/min in men, respectively*.

There is a high correlation between actual % och VO_2max at a HR of 128/138 and individual maximal HR. The first nomogram for the Åstrand test was based on data from a group of young subjects with mean maximal HR of 195 (SD 10) beats/min. The homogeneity of the tested subjects consequently resulted in a poor validity when the test and the nomogram was later applied in other populations.

Furthermore, the deviances from predicted maximal HR (220 - age) was correlated to the measurement error for the age-adjusted Åstrand test ($r = 0.57$, $p < 0.001$). The corresponding correlation for the EB test was considerably lower, and non-significant ($r = 0.22$, $p = 0.10$). Hence, the deviance from the predicted maximal HR is less related to the estimation error for the EB test, compared to the Åstrand test. The reason for this is probably related to the different theoretical bases for the two tests.

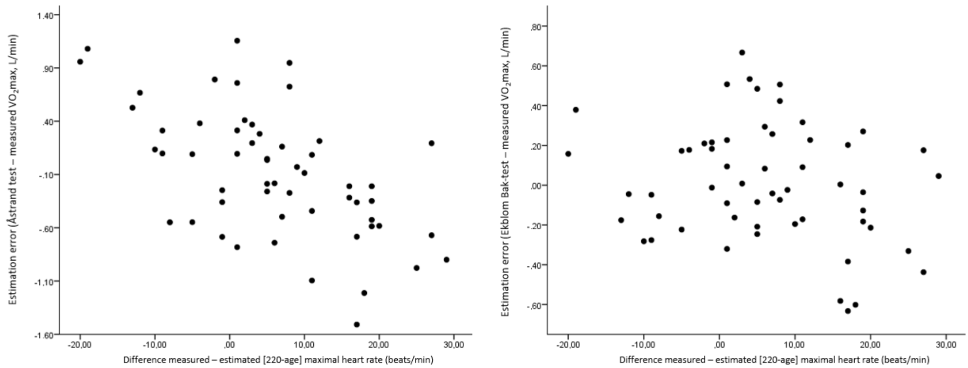

Figure 13. *The relations between the maximal heart rate (HR) deviance (difference between measured and estimated [220-age] maximal HR) and measure error for the Åstrand test (left) and Ekblom-Bak test (right). Data from the comparison group in study II; the Ekblom-Bak test results are derived with the sex-specific prediction equations.*

4.1.6 Additional analysis: the impact of body size on oxygen consumption

In theory, the energy expended for a given rate of work on an ergometer vary only to a small degree between individuals. This is thought to be due to a small variation between individuals, and the fact that a non-weight bearing activity like cycling limits the effect of differences in body mass. However, as pointed out above, when assessing oxygen consumption during a standardised rate of work among individuals of different sex, with variations in body size and fitness-level, it is obvious that the theoretical assumption is not fully correct. Previous studies have found that body size influence energy expenditure on a given cycle ergometer work rate (208, 209).

From the material in the present thesis, we analysed the role of body mass on gross O_2 cost for cycling at a submaximal work rate of 0.5 kp (32 W) and 2.0 kp (127 W). The relation between body mass and O_2 consumption is shown in figure 14.

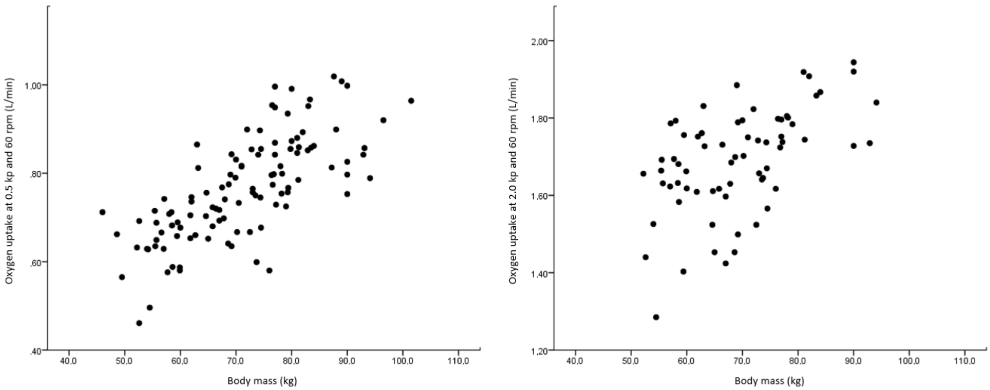

Figure 14. *The relations between body mass (kg) and oxygen consumption at a work rate of 32 W (left) and 127 W (right). Data from the model group in study II.*

Using our data from study II, we applied a linear regression with oxygen consumption at 0.5 kp as dependent variable and body mass, measured VO_2max and sex as independents (n = 110). Results showed 57.5 mL/min higher oxygen consumption among men. For each kilo body mass, oxygen consumption rose by 5.1 mL/min. Also, for each mL/kg/min of higher VO_2max, oxygen consumption rose by 2.8 mL/min. These three independents explained 62% (i.e. adjusted $R^2 = 0.62$) of the variation between individuals in oxygen consumption at a standard work rate of 0.5 kp. When analysing the oxygen consumption at 2.0 kp (n = 68), very similar values were obtained for body mass (6.1 mL/min) and VO_2max (2.1 mL/min), but slightly lower for sex (41.4 mL/min). The adjusted R^2 at 2.0 kp was lower, 0.32.

Table 4. *Predictors of oxygen consumption at 32 W and 127 W. Data show the results from linear regression models, oxygen consumption as dependent variable.*

Independent variables	B at 32 W*	*p-value*	B at 127 W*	*p-value*
Constant	236.3	*0.003*	1148.8	*>0.001*
Body mass (per kg)	5.1	*<0.001*	6.1	*0.005*
VO_2max (per mL/kg/min)	2.8	*0.002*	2.1	*0.31*
Sex (men = 1, women = 0)	57.5	*0.01*	41.4	*0.33*

**32 W, generated from cycling at a cadence of 60 rpm, with a resistance of 0.5 kp.*
***127 W, generated from cycling at a cadence of 60 rpm with a resistance of 2.0 kp.*

Based on data in the present thesis, we analysed relations between measurement error (the difference between estimated and measured VO_2-values) and body size and submaximal O_2 consumption. We found strong correlations between measurement error for the Åstrand test and body size (body mass and stature r≈0.45), as well as for VO_2 at 0.5 and 2.0 kp (r = 0.54 and r = 0.45, respectively). However, no relations between error for the EB test and body size or submaximal oxygen consumption were found (all r<0.14). This indicates that the EB test does not rely fully on the assumptions of a uniform energy expenditure during work on a cycle ergometer, perhaps due to the use of a difference in HR, rather than one single observation.

When focusing on body mass alone, we observed a strong bivariate correlation (r = 0.73, $p < 0.01$) between body mass and oxygen consumption at 0.5 kp (n = 110). The correlation at 2.0 kp was r = 0.54, $p < 0.01$ (n = 68). Since oxygen consumption during exercise is a mixture of activity related energy expenditure and resting metabolic rate (RMR), is logical to assume that a subtracting RMR from the gross oxygen consumption would lead to a less pronounced relation between body size and oxygen consumption on an ergometer. Hence, the role of individual differences in RMR was investigated, in order to gather important information to the discussion about work/activity-related (i.e. the energy cost from the cycling) versus total (i.e. RMR + cycle work) energy expenditure. We applied the Harris–Benedict equations revised by Mifflin St Jeor Equation (238) for estimating RMR and subtracted the found value from the oxygen consumption during standardised work at 0.5 and 2.0 kp. (i.e. 32 W and 127 W, respectively). The strong correlation between oxygen consumption at 32 W and body mass was only slightly attenuated after adjustment for RMR (see figure 15). A possibility is that the equation for estimating RMR lacks validity. However, the fact that the correlation still was significant is interesting. This may indicate that either external or internal efficacy varies between small and large individuals.

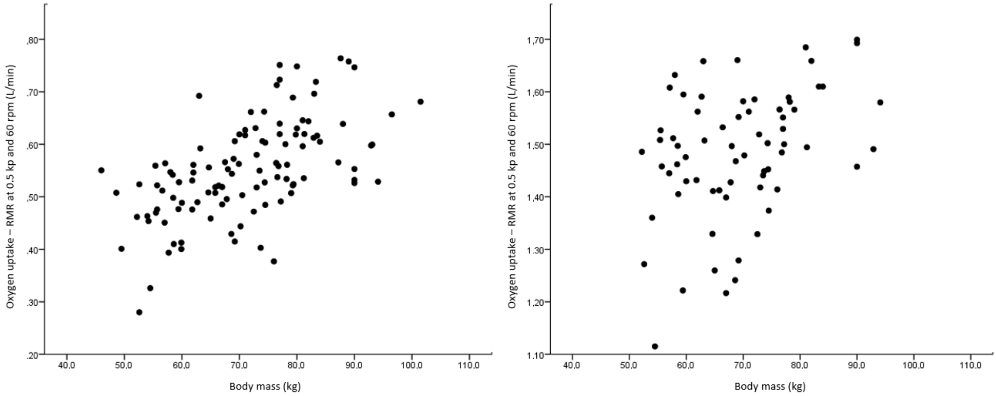

Figure 15. *The relations between body mass (kg) and oxygen consumption at 32 W (left) and 127 W (right), with estimated resting metabolic rate subtracted.*

Another way to briefly investigate the effect of body size is to induce an experimental "weight gain" by adding external load to the legs of subjects who cycle at different work rates, in order to analyse any effect on energy expenditure. Seven subjects cycled at 0.5 and 2.0 kp, with a cadence of 60 rpm. External mass was applied via leg weights of totally 5 kg, applied through a 2.5 kg weight on the left thigh, and a 2.5 kg weight on the right thigh, respectively.

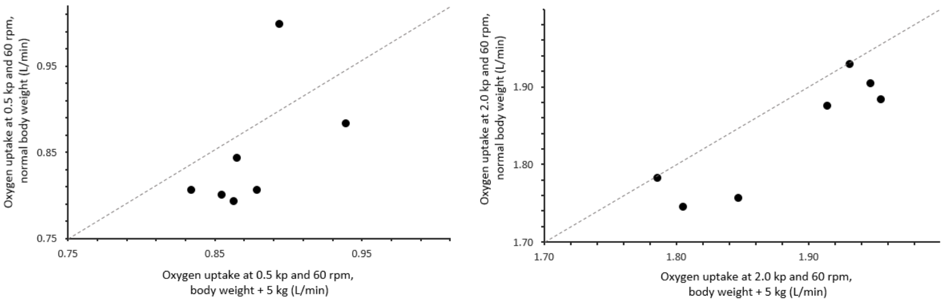

Figure 16. *Oxygen consumption during cycling with and without extra weight added on the legs. Work load of 0.5 kp to the left and 2.0 kp to the right, respectively.*

Results from the pilot experiments showed that the O_2 consumption was altered with extra weight on the legs. The reasons for this tendency to a change in gross efficiency is unknown.

4.2 Study III

Study III examined the conformity between the measured and estimated changes in VO_2max over a time-span of 5 to 8 years. This quality of a test is crucial for the test to be used successfully in health check-ups as well as longitudinal studies. The study also gave additional information to the theoretical background of the EB test, where it was unclear if the test would be able to detect a longitudinal change in VO_2max.

4.2.1 Changes in sample characteristics

The mean time between tests was 2329 days (6.4 years, ranging from 5.2 to 8.1 years). Mean age was 42.8 (11.1) at baseline and 48.4 (11.0) at follow-up, respectively. The females were 24–56 years old at baseline, and 29–63 years old at follow-up. The corresponding values for men were 21–65 and 29–72 years old, respectively. None of the subjects had a resting systolic blood pressure above 150, or a resting diastolic blood pressure above 100 mmHg. There was no significant change in stature. Women had a non-significant increase in mean body mass, from 61.3 to 62.4 kg ($p = 0.08$), ranging from -3 kg to +5.3 kg. Mean body mass for men was almost unchanged, 81.0 kg at baseline and 80.8 kg at follow-up, respectively. The individual changes ranged from -4.5 to +5.1 kg. There was no significant differences in measured VO_2max, while there was a significant decline in maximal HR, from 184.8 (14.6) to 181.0 (14.4) beats/min.

4.2.2 Changes in HR

The mean HR at the standard work rate was unchanged between tests, but with large individual differences. Ten subjects had a decrease of 10 beats/min or more (the greatest decrease was 23 beats/min) and three subjects had ≥ 8 beats/min (the greatest increase was 12 beats/min) in the last compared to the first test. Maximal HR was significantly lower at follow-up, with an average drop of 3.8 beats/min (-2.7 beats/min for women and -4.5 beats/min for men, with no significant difference between sexes). However, 42% of the subjects had an unchanged maximal HR (±2 beats/min), while the maximal HR had dropped by as much as 10–13 beats/min in four subjects.

There was no significant difference in time between the first and the follow-up tests for the subjects who exercised at the same high work rate at both occasions, and those who exercised at a different work rates. The mean HR for subjects that exercised on the same high work rate during both tests was significantly lower at follow-up, with values of 133.3 (14.6) and 129.6 (13.9) beats/min, respectively. The corresponding values for the group with a different high work rate was 123.5 (18.8) and 123.2 (10.3) beats/min.

4.2.3 Changes in VO_2 uptake

There was no significant change in mean measured VO_2max (3.65 and 3.64 L/min, respectively). The estimated VO_2max was significantly lower at follow-up, with values of 3.56 (0.93) at baseline and 3.47 (0.91) at follow-up ($p = 0.05$). Furthermore, the mean difference between estimated and measured VO_2max were -0.09 (0.31) at baseline and -0.17 (0.34) L/min at follow-up, and this difference was significant ($p = 0.018$). There were no significant differences in work efficiency. The VO_2 at standard work rate was 0.82 (0.11) at the first test, and 0.84 (0.09) L/min at follow-up, and 2.03 (SD 0.61) and 2.07 (SD 0.61) L/min for the 18 subjects that had VO_2 measurements from the same higher individually chosen work rate at both baseline and follow-up, respectively.

4.2.4 Conformity between changes

Linear regression analysis showed a significant unstandardized B coefficient of 0.88 (95% CI: 0.61 to 1.15), indicating that for each litre of change in measured VO_2max, the estimated value changed 0.88 L. The mean change in the error was -0.08 L/min. LoA were -0.43 to 0.28 L/min. Thus, at the follow-up test, subjects were slightly more underestimated than at the first test. The corresponding values expressed as mL/kg/min were -1.15, with LoA -6.17 to 3.87 mL/kg/min

Expressed as dichotomous change (increases in measured VO_2max versus decrease in measured VO_2max), those who increased their measured VO_2max had an underestimated change (0.17 L/min). In total, 8 of the 35 subjects had a change in measured VO_2max that was more than +0.20 L/min. The subjects who decreased in measured VO_2max had a non-significant change in error (0.01 L/min).

At follow-up, the mean estimated VO_2max was slightly underestimated compared to the measured values. Furthermore, the subjects that were tested with the same high work rate displayed a significantly larger error at the second test (mean error -0.05 ± 0.31 L/min at the first test and -0.15 ± 0.35 L/min at the second test, respectively). The group with changes in the higher work rate displayed almost identical error at both test occasions (-0.20 ± 0.30 and -0.20 ± 0.31 L/min, respectively). However, repeated measures ANOVA revealed that there was no significant difference between groups.

The strong correlation between changes in measured and changes in VO_2max values indicates that the measurement errors (i.e. the difference between measured and estimated VO_2max) seem to be rather constant over time. This means that anyone being overestimated at the first test will also be so at the second test. This was also verified with a strong correlation ($r = 0.84$) between the errors at the first test and at the follow-up test.

4.2.5 Internal and external factors with association to the prediction error

To examine if slope and/or position changed with changes in VO_2max, bivariate correlations were performed between change in measured VO_2max and the individual variables in the EB test prediction equation. With the exception of increased age ($r = 0.23$), all other variables included in the EB test prediction equation had a significant ($p < 0.05$) bivariate correlation to change in measured VO_2max (change in $\Delta HR/\Delta PO$ $r = 0.27$, change in ΔPO $r = 0.43$, and change in HR at standard work rate $r = 0.42$). These relations remained significant when tested mutually in a linear regression (standardised beta coefficient -0.44, 0.42 and -0.35, respectively). Results indicate that changes in both the slope and it´s position were related to changes in measured VO_2max.

Apart from changes in EB test result (estimated VO_2max using the full formula), ΔHRmax was a significant predictor for change in measured VO_2max (B = 0.941, $p < 0.001$ for the EB test value and B = 0.016, p = 0.028 for ΔHRmax respectively). Adding ΔHRmax to EB test result in a linear regression with measured VO_2max as dependent variable, increased precision slightly (r = 0.80 and SEE 0.17 L/min). Sex, changes in HR as standard work rate, changes in VO_2 at standard work rate, and change in body mass, were not associated to changes in measured VO_2max, when controlling for change in the EB test result (estimated VO_2max value). Changes in error was not correlated to changes in body mass (r = -0.06, p = 0.72) or changes in VO_2 at the standard work rate (r = 0.00, p = 0.99). The same results was observed when using post values for estimated values as dependent and pre values along with potential correlates as independents.

4.2.6 Additional analysis: results from the Åstrand test at baseline and follow-up

Analysis of the age-corrected values from the Åstrand test revealed that there were no significant correlation between the changes in measured and estimated values (r = 0.37 and p = 0.10, and SEE 0.22 L/min). The mean change in error was -0.08 (0.46) L/min, and LoA were -0.98 to 0.82 L/min. There was a significant correlation of r = 0.72 between baseline and follow-up error for the Åstrand test with the age correction factors, and 0.63 when the estimated values were corrected for maximal HR. Due to the relatively small sample, no further analyses were done.

4.2.7 Choice of work rate for repeated test

With regard to the choice of work rate, the EB test does not seem to be heavily influenced by the relative intensity (% of maximal HR) in the higher work rate (within reasonable limits, i.e. the work rate shall always be chosen to correspond to a subjective intensity 12–14 on the RPE scale). However, Davies et al. have showed that a higher work intensity during a submaximal cycle test yielded intra-individual variations in HR of about 2%, while the intra-individual variations at lower intensities were 3–8% (18). In line with that observation, we performed analyses on our data plotting HR at the high work rate against measurement error. The rectified difference was only weakly correlated to the measurement error (r = 0.20, p = 0.02), and we therefore conclude that tests with a high HR down to 100 beats/min can be used for analysis. However, in figure 17, it is also obvious that the absolute measurement error is lower for tests conducted at over 140 beats/min for the high work rate, indicating that this would be an optimal intensity for testing. However, the absolute error did not differ significantly between tests with HR over and under 140 beats/min at the high work rate (Mann-Whitney U-test p = 0.6), indicating that the effect of this factor is limited.

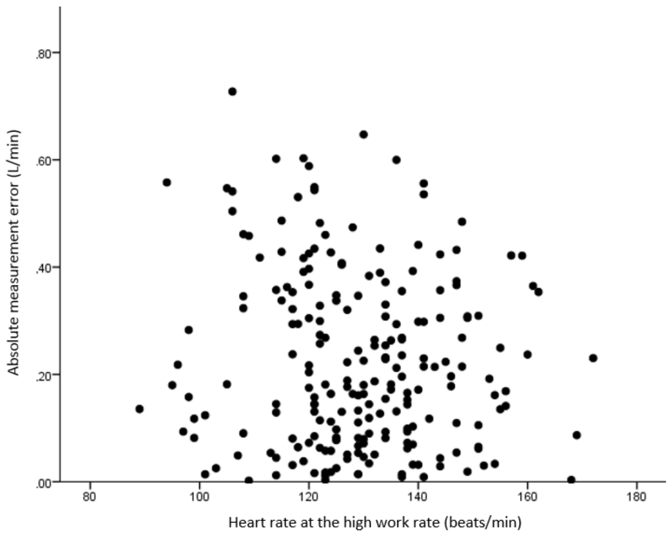

Figure 17. *The correlation between rectified measurement error (estimated - measured VO_2max) and HR at the higher, individually chosen work rate from the EB test.*

In the original publication of the Åstrand test, P-O Åstrand and colleagues stated that a higher load gave a better prediction (108). This observation was later confirmed in work of Irma Åstrand (71). However, regarding the EB test, data showed that there was only a weak correlation between measurement error and HR at the high work rate. Hence, there is no need to exclude a test due to low HR at the second work rate (see figure 17). This is an advantage for the EB test, compare to the Åstrand test, where the nomogram only includes HR above 120 beats/min.

4.3 Study IV

In study IV, the validity and reliability of the EB test was examined in children and adolescents. Furthermore, circulatory and/or maturity-related factors for prediction errors were identified.

4.3.1 Validity

The mean difference between estimated and measured VO_2max was 0.09 L/min, r = 0.86 and SEE 0.29 L/min. Sex and maturity status, as well as the variable ΔHR/ΔPO in the prediction equation, were significant predictors for the estimation error. The largest overestimation was seen in pre-pubertal boys in Tanner stage I and II (see figure 18). The precision of the EB test was enhanced when the data from boys in Tanner stages I and II were re-calculated using the prediction equation developed for women (mean difference -0.05 L/min, r = 0.92 and SEE 0.23 L/min in the entire sample). Prediction error decreased in boys, but not in girls, with increasing pubertal maturity.

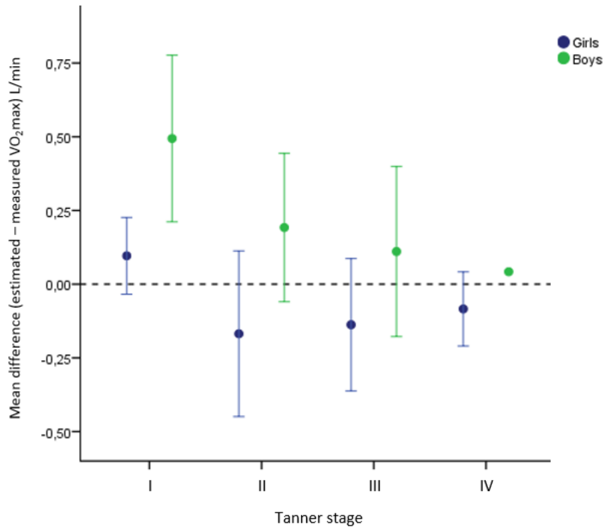

Figure 18. *Mean differences between estimated and measured values of VO2max for the different Tanner stages (I–IV). Error bars denote 95% confidence interval.*

4.3.2 Reliability

Analysis of the test-retest reliability for the EB test in children showed a non-significant mean difference of -0.08 (0.28) L/min, a LoA of -0.63 to 0.47 L/min and a correlation coefficient of 0.87 (adjusted R^2 0.74) between the two test occasions. SEE was 0.28 L/min. Coefficient of variation (CV) for test-retest was low, 3.1%. In comparison, the CV for measured VO_2max was 2.7% in the 14 children who performed repeated maximal tests.

4.3.3 Additional analysis: the Åstrand test

A total of 39 adolescents (17 girls, 22 boys) had test results that allowed for estimation of VO_2max from the Åstrand nomogram with the associated age-correction factors. The median estimated VO_2max was 2.12 (2.00 to 2.45) L/min compared to the median measured VO_2max of 2.20 (1.99 to 2.50) L/min, and there was a significant correlation between these values (r = 0.77 and adjusted R^2 = 0.59). The mean difference between estimated and measured VO_2max was -0.08 (0.43) L/min. The SEE was 0.33 L/min, and LoA were -0.73 to 0.57 L/min. Furthermore, the difference between estimated and measured values was correlated to body mass (r = -0.48). The results from the Åstrand test were also analysed separately for boys and girls, and the values are given in table 5.

Table 5. *The validity of the Åstrand test for boys and girls, respectively*.

	Girls (*n* = 17)	Boys (*n* = 22)
Mean difference, estimated - measured VO_2max (L/min)*	0.09 (0.18)	-0.22 (0.36)
Adjusted R^2	0.70	0.63
Limits of agreement (L/min)	-0.26 to 0.44	-0.93 to 0.49
SEE (L/min)	0.16	0.37

Values are mean (SD).

4.3.3 Additional analysis: Supramaximal (verification) test

In total, 45 subjects performed a supramaximal test after the maximal incremental test. There were no significant differences (with regard to age, height, body mass and incremental VO_2max) between the subjects that performed the supra I- and supra II-protocol. There was a strong significant correlation between the values from the maximal and the supramaximal tests (r = 0.992 and adjusted R^2 0.986).

Table 6. *VO_2 values from the incremental and supramaximal tests. Median (25^{th}–75^{th} percentiles). Unpublished data from study IV.*

	n =	Incremental test, VO_2 (L/min)	Supramaximal test, VO_2 (L/min)	*p-value**
All	45	2.14 (1.94–2.49)	2.17 (1.95–2.56)	*0.04*
Supra I-protocol	24	2.22 (1.96–2.56)	2.26 (1.96–2.61)	*0.004*
Supra II-protocol	21	2.10 (1.94–2.46)	2.05 (1.94–2.42)	*0.69*

Wilcoxon test.

The maximal VO_2 values from the different test protocols are presented in table 5. In the whole group, the supramaximal test values were about 1.5% higher than the incremental values. However, this difference was regarded as small and it was concluded that the incremental tests reflected a true VO_2max. Furthermore, the protocols were also analysed separately. The supra I-protocol produced a higher VO_2max than the incremental test, and this difference was significant. The mean time for the supra I tests was 246 (31) s, and the corresponding value for the supra II-tests was 310 (25) s for supra II. Hence, the latter protocol resulted in a shorter time of running at the supramaximal workload, due to the inclusion of two minutes of threshold running in that protocol.

5 METHODOLOGICAL CONSIDERATIONS

5.1 Instrumental aspects

The cycle ergometer was calibrated according to previously described procedure. The work load was carefully monitored throughout the whole period of cycling. However, it is worth mentioning that the use of the 828E cycle ergometer always contains a small amount of uncertainty, due to the manual handling of the devise. First of all, the ergometer has to be operated on a flat and firm ground. The laboratory environment where the experiments were conducted provided such conditions. The advanced calibration with standard weight of 4 kg was performed as soon as the ergometer had been moved other than rolling and at least once every three months. The zeroing of the scale was performed for every new test. The pendulum was monitored regularly during all cycling, and adjusted if needed. It is worth noticing that there are some differences between a manually adjusted and mechanically braked ergometer – like the model 828E – and other cycle ergometers (for example different models with electronically adjusted resistance). As mentioned in the introduction, frictional losses in the belt and chain is taken into account in the variables that are used in the EB test prediction equations. The importance of considering the difference between nominal work rate and the work rate performed by the subject was also highlighted by Åstrand; “this added load must be considered when comparing work load, oxygen uptake, pulse rate relationships on the Monark and calibrated electronically braked ergometers” (196).

The validity of the Oxycon Pro system and the stability of its measurements have been reported to be high in previous studies (46, 47). We have maintained the unit used for the data collection in this thesis according to the manufacturer’s instructions. Regular services have been performed, where calibration has been a part. At these service calibrations, a maximal error of 3% is reportedly being accepted. However, the validity of this unit has not been tested separately against Douglas bags or a mechanical lung.

5.2 Test standardizations

The procedure before and during each test has been described in detail above. All tests were proceeded by standard routines with regard to calibration of the equipment that was used. All tests were conducted in a laboratory environment. Furthermore, some general advice were given to all participants in the studies included in the present thesis. These advice comprised of regulations of exercise as well as regulations of intake of food and beverage. All participants were told to refrain from exercise training the day before test and on test day. They were also instructed to commute to the lab by train, car, or easy walking. However, it cannot be guaranteed that all subjects followed these guidelines. We do now that almost all subjects had at least 10 min of supine rest before each test, and that any abnormalities in the measured resting values were immediately monitored and evaluated by medical personnel.

Furthermore, all subjects were advised to not ingest any bigger meal within 3 hours before test. It would have been possible to restrict the subjects even more, for example with standardized meals before each test. However, we chose to conduct all test sessions more like the practical/clinical reality, where all tested subjects more or less have a normal "every day" life before test. We believe that this approach was most useful in the studies included in the present thesis. Furthermore, volunteers with medications such as β-blocking or β-stimulating agents were excluded, as well as smokers. Snuff users was included in study I and III, with the regulation that no snuff should be used within 3 hours before test. Study II and IV did not include any known snuff users.

With regard to the practical execution of the test, some aspects have certain influence on the standardisation of the test. One aspect that influenced the actual work rate in the EB test was the subjects' ability to keep the correct cadence. In the studies in the present thesis, a range of 59 to 61 was accepted during the last minute of cycling on each work rate, with an acceptance for short moments with a pedalling rate of 58 or 62 rpm. However, all subjects were carefully guided into a steady pedalling rate during the first minutes of the test, and it was very unusual that participants happened to pedal in 58 or 62 rpm during the latter stages (≥ 2 min of cycling) of the test. A few participants (mainly children) were excluded because they failed to achieve a steady-state pedalling rate mostly within 59–61 rpm. All participants were also instructed to focus on the cycling and maintain quiet during test, in order to minimise the external influences on HR and VO_2.

Lastly, it is worth mentioning that the Åstrand test values were generated from the same test session as the EB test values. Hence, a majority of the Åstrand test values are derived from the results from the first test session, i.e. without any familiarisation test. This approach was chosen because we decided that it generated the most accurate comparison between tests, with respect to the actual practical/clinical use. However, it is probably a disadvantage for the Åstrand test, because it has been reported that the prediction error decreases with repeated tests (239). Nevertheless, we believe that the advantage of the EB test is the possibility to use the test values from the first test session, without any familiarisation test (145).

5.3 Design of the test for determination of VO_2max

We would like to address some issues with regard to the direct measurements of VO_2max. The direct measurements of VO_2max have been conducted during slightly inclined running. This procedure was used in all four studies in the present thesis. The choice of work mode was based on the assumption that the slightly inclined running would generate the highest possible value of VO_2max, i.e. reflect the “true” maximal capacity from whole-body work. Hence, the estimations of VO_2max from the EB test is built upon measured values from the activity “inclined running”. It is worth noticing that many other studies of submaximal exercise tests have used maximal cycling as work mode for the reference values (i.e. the directly measured VO_2max). For example, Åstrand et al. used cycling for the direct measurements of VO_2max in the experiments during the 1950s. Hence, it might be possible to achieve the same VO_2max from cycling and running in certain groups of subjects, while other groups exhibits difficulties in achieving a “true” VO_2max from a maximal effort in cycling. Maximal cycling put high demands on working leg muscles, and generates a high level of local muscle strain. Elderly people with a high degree of sarcopenia may be muscularly exhausted before the central circulation has reached its actual peak.

The aim of this test is not to predict maximal oxygen consumption during cycling, but as close as possible to the true VO_2max. Therefore, we chose running as work mode for the VO_2max reference value. Another interesting question is why there is no mean difference (underestimation) when using the Åstrand test (which was developed using a cycle ergometer maximal test as reference) to predict a maximal O_2 value obtained from a treadmill test. One possible explanation is that the subjects in the original study by Åstrand did not differ in VO_2 obtained with cycle ergometer and treadmill (the entire group did not differ, but in stratified analyses, women showed slightly lower ergometer-based VO_2max, compared to values obtained on a treadmill). This might be due to the fact that those subjects were highly trained and therefore were able to reach a high VO_2 on a cycle ergometer. However, we believe that for most people, there is a risk for local fatigue in the leg muscles before a true cardiovascular maximum has been reached.

Also, the criteria for achievement of VO_2max also vary considerably among studies. The maximal test in the present thesis have used the levelling off criteria as a sign of achieved VO_2max in all adult subjects, with additional supporting criteria. However, it is necessary to evaluate the choice of criteria against the tested specific population that is tested including the definition of VO_2-levelling off, cut off values for RER and maximal HR.

6 DISCUSSION

The submaximal cycle ergometer tests are used in a variety of contexts, including health related clinical work, as well as fitness oriented measurements at sports or fitness clubs. However, the wide spread use of these tests also warrants a thorough analysis of their optimal use, regarding for example how to perform them, when not to use them, how to understand the results, and what to understand about their limitations. This thesis was focused on the validity and reliability of the EB test. In addition, other existing tests were discussed regarding differences in theoretical background and validity. Furthermore, the possible sources of error that can influence the results from submaximal cycle tests have been discussed.

6.1 The development of the test

A series of events in research history have contributed to the construction and development of the EB test. This hold true for the test itself, the associated measurement devices, as well as the need for exercise testing in different contexts in society. The first ergometer cycle tests were developed on mostly athletic populations, and the early tests and measurements were used in a century when the cardiac function was not fully understood. Later, the technological development made it possible to conduct direct measurements of cardiac output, which gave new insights into the function of the heart during maximal work. Also the measurements of O_2 uptake and HR became relatively easy to conduct. At the same time, other fields of research – medicine, public health, etc. – highlighted the strong association between physical fitness and several health parameters including overall mortality (21). Today, the correlation between aerobic capacity (VO_2max) and risk for disease is well established.

The EB test is an ergometer cycle test that comprises of a non-maximal and relatively easy test procedure. The hypothesis behind the development of the test and the associated prediction equation was that the use of the delta HR (ΔHR) response between a lower standard and an individually chosen higher work rate, rather than the HR response to one work rate, would enhance validity in the estimations of VO_2max. For example, the use of ΔHR in the prediction equation might make the test less sensitive for individual variations in maximal HR, and thereby produce higher validity through lower random variation. The studies in the present thesis have shown that the EB test and its corresponding prediction equations result in an estimation of VO_2max that is more accurate than the estimations from many other indirect tests for estimation of VO_2max, for example step tests, walk tests, and other cycle tests. Up until today, one of the most commonly used tests is the Åstrand test. The first reports of validity for the EB test, in a healthy mixed population, showed nearly half of the estimation error for the latter test compared to the former. The CV for the EB test has been shown to be around 9%, while the precision for the Åstrand test has been reported to be between 15 to 18.1% (108, 145). Furthermore, our reliability analysis of test-retest date imply that the EB test can be used without the need of any familiarisation test. The ability to produce a valid and reliable result at the first test occasion is an advantage for a submaximal exercise test.

One important aspect in the future development of the submaximal exercise tests in general, and the EB test in particular, is the use of the test in other populations. For example, illness and medications plays a crucial role in society today, and the use of a submaximal test in these groups of people.

6.2 The construction of prediction equations for estimation of VO_2max

The knowledge about the relation between HR and work rate has an important role in the development of the EB test. Furthermore, it was shown that the theoretical background to the Åstrand test nomogram (the relation between a HR of 128/138 and 50% of VO_2max, see section 1.4.2 for details) was rather inaccurate in a large proportion of a heterogeneous population. The error from the above mentioned relation was mainly dependant on lowered maximal HR with increased age and therefore age-correction factors were introduced. We believed that a test with corresponding prediction equations based on change (Δ) in different aspects of HR and work rate would lower the estimation error from submaximal cycle testing. Hence, we developed a two-point test, with Δ-values as important variables in the prediction equations for estimation of VO_2max.

Consequently, it has been shown that the results from the EB test are unrelated to variations in maximal HR. This is further discussed in detail in a previous thesis (237). The ΔHR variable which reduces the impact of variability in absolute submaximal HR response due to ambient temperature, nervousness, or emotions as well as work efficiency. This further reduces the need of a first familiarisation test, a procedure that is needed to generate accurate results from a HR-based one-point test like the Åstrand test. Higher HR and a lower mechanical efficiency at the first occasion are common features in series testing, and are mainly due to excitement, anxiety, and/or tension among the participants (147). This was confirmed by reliability analysis in study I, showing that although the absolute HR at both standard and higher work rate was significantly higher for the first occasion compared with the second occasion one week later, there was no difference in either ΔHR or estimated VO_2max between the two occasions. Furthermore, variations in mechanical efficiency could also affect the HR response to one single submaximal work rate. For example, it has been shown that there are intra-individual variations in basal metabolic rate (BMR) and oxygen consumption at higher work rates, due to variation in mitochondrial oxygen affinity (p50mito). The p50mito has been shown to be negatively associated ($R^2 = 0.52$) with mass-specific BMR as well as efficiency (i.e. oxygen cost) during exercise at 50% of VO_2peak (240). This aspect in relation to the EB test has been discussed elsewhere (237).

6.3 The possibility to detect changes in VO_2max with the EB test

The EB test was also found to have a rather good precision over time, at least over 5 to 8 years. This aspect of a test is highly important, since submaximal cycle ergometer tests are frequently used to follow individuals over prolonged times. Even though the EB test has been shown to be a rather good tool for estimation and evaluation of aerobic capacity (i.e. VO_2max), over a shorter time frame, it is still unknown whether or not the test has the ability to accurately detect changes in VO_2max over an even more extended time period, or as a consequence of an aerobic training program. Expressed as dichotomous change (i.e. those subjects that showed increases in measured VO_2max versus those that showed decreases in measured VO_2max). The subjects who decreased in measured VO_2max had a non-significant change in error (mean 0.01 L/min). In contrast, the subjects that displayed an increase in measured VO_2max had an underestimated change, i.e. their actual increase in VO_2max was not correctly captured with the EB test. One possible explanation to this may be that the subjects who increased their VO_2max with age actually displayed a contradictory development in aerobic capacity, compared to the expected development (where the expected age-related decline in VO_2max is approximately 7% per decade). The EB test prediction equations includes the variable "age", which could have had a slight influence on the estimation error for the subjects that had increased their VO_2max over the timespan of 5–8 years.

For comparative reasons, we also evaluated the change in age-corrected VO_2max estimated by the Åstrand test (with associated age-correction factors). The change in estimated values were not correlated to the change in measured VO_2max, hence our results were showing low ability of the Åstrand test to detect change in actual VO_2max over time. Furthermore, it was shown that the methodological and theoretical differences between the Åstrand test and the EB test influences the ability to follow a change in VO_2max over time.

6.4 Submaximal cycle tests in children

In study IV, we examined the validity and reliability for submaximal cycle tests in children and adolescents. We found that the EB test and Åstrand test differed in prediction errors. Overall, the Åstrand test resulted in a larger mean misclassification and larger random variation compared to the EB test. One possible explanation is that Åstrand is a one-point test, developed on the relationship between HR and work rate in an adult population. The higher HR at a standard submaximal VO_2 in children interferes with the ability of correct predictions and more frequently leads to losses in data due to the fact that the obtained HR – work rate combinations are not covered in the nomogram. It is interesting to note that body mass was related to the difference between estimated Åstrand values and measured VO_2max, but not correlated to the difference between estimated EB test values and measured VO_2max. The observation regarding the influence of body mass was also seen in adults, and these findings are discussed in the section "Further increases in precision and validity of submaximal cycle tests".

When analysing the data for boys and girls separately, the pattern of maturity related changes were greater in boys than in girls. The girls did not display as great maturity induced changes in O_2pulse nor in RQ at maximal workload and HR at standard work rate (data not shown) as the boys did. Boys, on the other hand, showed larger differences from pre-pubertal (Tanner I and II) to mid- and late-pubertal (Tanner III and IV) stage. Regarding the EB test, this led to an enhanced precision for the more mature boys while the opposite was seen for the Åstrand test, where mid- and late-pubertal boys became more misclassified. The reason for this is not known. However, in line with the earlier mentioned observations in the present study, no such pronounced maturity-induces changes in VO_2max predications from neither the EB test nor the Åstrand test was seen for girls across the different Tanner stages.

Another aspect to address in the measurement of VO_2max in children is the normalization to body size. The VO_2max can be reported in absolute values (L/min), or in relative values. When using the relative terms, the L/min value is divided by a factor related to body size, such as body mass, stature or body surface. The most commonly used denominator is body mass, which often result in an incorrect adjustment of the differences in VO_2 among children and adolescents. Instead, the body mass raised to the power of 2/3, 3/4, or similar values, seem more correct, since the relation between VO_2 and body size in these cases falls closer to zero. However, no consensus has been reached in this area. The relative values can also be generated through allometric scaling, i.e. removing the influence of body size on the VO_2max value on basis of the present data – hence, the scaling values differ between data set. In the present thesis, and in the concomitant research articles, we have decided to only report the VO_2 values as L/min. This leaves the choice of the most appropriate form of scaling to the researchers using the EB test in children and adolescents. The coefficient for scaling for the study population in paper IV was 0.914.

6.5 Further increases in precision and validity of submaximal cycle tests

The possibility to further increase the validity of submaximal cycle tests is not infinite. At some point, it has to be deemed that it is unnecessary to try to enhance the estimation even further. The usefulness of a test is dependent on a certain amount of user-friendliness, a reasonably wide range of validity, etc. Furthermore, there are always external and internal factors that cannot be controlled for via the practical test setup and/or the variables included in the prediction equation for VO_2max. In cases where minute alterations or differences in VO_2 is of interest, submaximal tests are not sufficient. An example of this is training studies in laboratory settings. In these kinds of studies, the direct measurement is superior to submaximal tests.

During the development of the submaximal cycle tests, different aspects has been found to be relatively easy, but yet particularly important, with regard to the most important factors that influences the precision of a submaximal test. First of all, the HR response to exercise is different in different age groups, mainly due to an age-related decline in maximal HR. This is of great importance for the submaximal tests that are based on a HR - work rate relationship. The impact of age- and maximal HR differences was investigated by Åstrand et al. and some adjustments was proposed, and presented in form of different correction factors. However, it may still be possible to further develop the Åstrand test and its associated nomogram. For example, the nomogram has been revised by Åstrand et al., presenting a nomogram based on a wider age-range and an increased number of subjects (71). Also other research groups have proposed corrections to the nomogram, for example von Döbeln et al. (137), and Woynarowska et al. (188) and Binyildiz et al. have developed child-specific regression models for the Åstrand test.

Another possible way to eliminate some of the error caused by the maximal HR estimation is to include a maximal test for direct determination of the subjects' individual value, and thereafter adjust the predicted VO_2max value from the nomogram to the actual (known) maximal HR. This procedure may be applicable in young and healthy subjects, for example athletes.

It has been shown previously that energy consumption in relation to a given cycle workload correlates to body mass to various degrees (208-210), and that the predictive power in estimations of VO_2max from a submaximal work rate can be enhanced by adding body mass to the prediction equation (209). In the present thesis, we investigated the effects of a change of the mean body mass in the population, by adding the relation between body mass and O_2 consumption to a VO_2max estimation. This manoeuvre was based on the finding that variations in body mass was related to measurement error for the Åstrand test. This observation in the present material is in contrast to the conclusions by von Döbeln et al. They concluded that measures of body size not contributed to any further enhancement in the precision of the prediction equation for the Åstrand test (137). However, we believe that a correction factor may be proposed and applied to one-point tests, for example the Åstrand test. Such factor may reduce the impact of the variations in measurement error due to body mass. A correction may be applied in the nomogram itself (by up regulating the "work rate" based on body mass) or as an adjustment made to the obtained value (similar to the one used for variations in HR max). From the material in study II in the present thesis, we found similar values for increase in oxygen consumption with increasing body mass at both 32 W and 127 W (between 4 and 5 mL/min per kilo). Furthermore, it is worth mentioning that the estimation error from the EB test was unaffected by the above mentioned relation between body mass and efficiency.

Apart from the enhanced model group in study II (which allowed the development of sex-specific prediction equations), we have concluded that the ability to further improve the precision of the EB test by including even more subjects is rather limited (237). The usefulness of the EB test is its dependent on a rather easy-administered, cheap and quick test procedure. Perhaps, the precision error cannot be reduced much more in mixed population samples without compromises in the "simplicity" of the test procedure.

6.6 Gender differences

To consider differences between genders is an essential aspect in all areas of physiological research. With regard to the gender aspects in the present research program, the aim is to include both men and women (boys and girls) in all studies. For example, the EB cross-validation study (study II) made it possible to create sex-specific prediction equations for estimation of VO_2max, which in turn made the relatively greater overestimation for untrained male subjects (as observed in study I) less pronounced.

In study I, we found a similar increase in VO_2 with increased work rate for men and women, so no gender-specific ΔPO factors were needed. However, the original prediction equation (EB_{2012}) showed a tendency towards systematic underestimation among well-trained subjects. Hence, this observation also supported the idea to broaden the model group (partly by increasing/including more subjects in all sub-groups of sex, age and fitness level) to be able to create gender specific prediction equations.

Furthermore, study IV included both boys and girls, and we observed that the estimation error was dependent on sex and maturity status. In a clinical setting, it might be valuable to use the submaximal EB test in youngsters, due to the absence of maximal exertion and competitive elements. It has been repeatedly shown that boys usually have a greater proportion of high intensity physical activities and sports, and usually displays a higher contribution to physical education (178, 241). Therefore, a maximal or/and performance based test for classification of aerobic fitness might disfavour a large proportion of the girls. In case there is a need for determination of aerobic fitness (VO_2max), a valid and precise submaximal test might be a good alternative/option to the commonly used maximal performance tests (Beep/Yo-Yo test and Cooper test, for example). The ability to reach a greater proportion of the pupils in school might be increased, if the "measurement" and/or testing of physical capacity is conducted in other settings than during the physical education lessons.

6.7 Strengths and limitations

Strengths of the EB test include the consistent high accuracy and precision in the external cross-validation sample, as well as in the different sub-groups (men and women, different age-groups and a wide range of VO_2max levels). This is of particular interest with regard to the fact that the sex-specific prediction equations are based on a relatively large and heterogeneous sample, with a wide variation in age and VO_2max. We therefore believe that the test now is suitable for most non-diseased individuals. The use of sex specific prediction equations, rather than controlling for sex within an equation, has undoubtedly contributed to better precision.

The subjects in study I, II and IV were recruited after public announcement and word of mouth in the region of Stockholm, Sweden. A common issue in studies involving maximal testing is the selection bias, often including more fit individuals than in the general population. This may limit the accuracy of the prediction equation in the general population. This was mainly seen in study I and IV, where the volunteering participants had a somewhat higher physical capacity and VO_2max than the general population.

In study I, we included subjects of different sex, age and activity level, since we hypothesised that those aspects could influence the HR - VO_2 relationship. We excluded all subjects with medications or illnesses that could influence the above mentioned relationship. The final study population was well mixed with respect to sex and age, while there was an underrepresentation of inactive men with VO_2max < 2.5 L/min. Hence, the EB_{2012} prediction equation was not validated for men with poor physical fitness.

One way to analyse the external validity of study II is to compare the decrease in VO_2max with age between the model group and previously reported values from general population samples. The age-related decline per decade in measured VO_2max for the model group sample in study II was -9.1%, compared to -6.5% in 10 973 men and women, where VO_2max was obtained from maximal testing on cycle ergometer (14) and -6.9% in 3678 men and women, where VO_2max was obtained from treadmill testing (16). Although variations are present, there is a relatively linear decline in VO_2max from approximately 30 years of age and up to 65 years of age. Differences in the magnitude of the decrease may be dependent on the recruitment procedure and test methodology. The small difference between our study II and the other studies that are mentioned above (14, 16) may be due to a higher mean fitness among the young participants in our study. Even if this is a sign of a small bias, the implication of this observation may be regarded as low. Moreover, a submaximal test only estimates VO_2max based on variables obtained during the submaximal exercise. Hence, individuals who deviate in physiological characteristics from the individuals included in the model group, for example with an extremely high or low aerobic fitness, exceptional work efficiency or abnormal HR response, may obtain an estimated VO_2max further from their actual VO_2max than expected. The increased number of subjects in study II made it possible to include individuals with VO_2max > 4.5 L/min. Furthermore, older people and subjects with low VO_2max were specifically recruited, in order to enhance the valid range for the EB test.

All study designs and analysis have limitations, but recognition of these will strengthen the interpretation of the data obtained as well as improve the quality of future work. In paper I, we based analyses on a limited sample size and distribution. Further, no cross-validation was performed. We performed the test-retest assessment at the same time of day, which limited the influence of circadian variations in HR. Study II was based on a larger and more heterogeneous sample and cross-validation was included (strengths), but only one form of equation was tested (not non-linear or spline regressions). In study III, we observed a non-significant group differences in measured VO_2max, but significant group differences in estimated VO_2max. Also, this was not a "training intervention" under controlled circumstances. With regard to study IV, we could only include few individuals in Tanner IV. Hence, the limited number of subjects in the higher Tanner stages made it impossible to conduct detailed analyses of maturity-related and gender specific differences in estimations errors. Also, measurements of body segments mass or heart volume would have improved the analyses. As in study I, we performed the test-retest assessment at the same time of day, which limited the influence of diurnal variations in HR. Furthermore, almost all of the volunteering adolescents were Caucasian.

During all physical tests, the ambient conditions at the laboratory included variations in carbon dioxide concentrations and no correction for this could be made. The extent to which this may have influenced the results is unknown, but since the variations may be assumed to have been random, it may be to a limited degree regarding results on group level.

6.8 Ethical considerations

The ability to use an easy administered, non-expensive and accurate submaximal test for estimation of VO_2max is important in many different health- and training settings. It is my opinion that the strong influence of physical fitness (in terms of aerobic capacity) on general health makes it rather unethical to administer a test with lack of validity and/or reliability.

With regard to the gathered understanding for this specific topic in healthcare services, the overall profit of a precise, correct and subjective measure is of great importance. The tests in the present research program are of both submaximal and maximal character, where the physical and psychological stress of the maximal work may cause some transitory discomfort (especially for subjects unaccustomed to exhaustive exercise). However, the most ethical approach is to implement accurate and profound developmental work with the submaximal test, even if it may involve some short-term discomfort for a limited number of voluntary research subjects. It would be unethical to settle with an incomplete and inaccurate test and test-procedure.

In health care settings, it is valuable to be able to communicate information and advice about health and lifestyle choices through balanced opinions. This may be gathered through objective measures of a capacity (for example cardiorespiratory fitness) instead of any other type of behavior (for example a lifestyle questionnaire). Questions about behavioral aspects of health and wellbeing (such as smoking, alcohol, exercise, etc.) may be perceived as more intrusive and an attack on the integrity, compared to a valid and reliable objective measure on current fitness level. Additionally, as mentioned in the introduction, confounding variables like smoking, alcohol and cholesterol values have been shown to have a less pronounced significance for premature mortality, compared to the overall importance of physical fitness.

When the research is involving children, the procedures must be completed with extra considerations and sensitive treatment. Close and careful communication with the legal guardians is also of great importance, when working with under-aged subjects. There are many arguments for justification of testing in children. One important aspect is the need for a valid and reliable submaximal test for children and adolescents. Today, there is a current predominance of performance-based and maximal tests in conjunction with fitness testing in this population. The performance-based test with high intensity efforts often contains a pronounced learning effect, and in many situations also a competitive element with comparisons easily made between participants (the same is true also in adults, but the performance based all-out tests are more common in school- and sport settings than in health screenings for adults). Moreover, the all-out effort may cause great nervousness and tension, especially in untrained children and those unaccustomed to intense exercise. Moreover, many of these performance based tests are carried out in a group setting, making it even more socially difficult for an unaccustomed or apprehensive child. It is not ethical to conduct such tests, where the children in greatest need of an objective value of physical fitness, also are the least keen to participate. With a custom submaximal test, without any direct judgmental or competitive aspects, healthcare professionals might have a better chance to address young people who are at risk for poor health due to low fitness.

6.9 Future directions

The use of HR based submaximal exercise tests will probably be even more widespread in the future, mainly in different contexts in health care settings and fitness clubs. Some of the reasons for that is the general use of HR monitors nowadays, the increasing business of holistic training and clinical health care, and the escalating scientific interest in the important relations between aerobic fitness/VO_2max and health. We believe the EB test to be suitable in such situations as mentioned above. However, there are still a number of areas that needs to be further investigated to increase the usefulness and applicability of the test.

One area of interest is the influence of different medications on submaximal HR, and how this may affect the results from the test. The EB test prediction equations is based on ΔHR, and the ΔHR relation to power output, whereas medications, such as β-blocking agents, may alter these relationships. It would be of great use to test the validity and reliability for the EB test in subjects who ingest a β-blocking agent. There is a current need for an accurate and precise fitness test also applicable in individuals who take medications for common diseases like elevated blood pressure or minor cardiac dysfunctions. It is crucial to be able to do a quick and easy measurement of fitness level in these populations, in order to facilitate appropriate risk assessment and subsequent health advice. Future research should focus on the feasibility of the EB test in clinical populations. Also, the effect of the β-stimulating agents should be investigated, due to the widespread use of these medications in the general population as well as in athletes.

Another scope of interest is the ability of the EB test to follow a long-term change in VO_2max, for example over a period of 20–30 years. The use of the test in this situation has not yet been evaluated. Another area of interest is the possibility to detect a training-induced increase in VO_2max, for example after a training intervention. The EB test has been shown to have a reasonably good validity, and a rather good precision, for re-tests up to 8 years from the first test occasion. However, we found that individuals that increased their VO_2max over time (5–8 years) were more misclassified than those who had an unchanged or slightly lowered value over time. It is still unknown whether or not the EB test has an acceptable sensitivity for detection and follow-up of training-induced increases in VO_2max.

6.10 Conclusions

The completion of the studies in the present thesis will give great opportunities for general use of the test in many different settings in the society and health care. Furthermore, it is important to be able to have a scientific approach to an escalating problem with increased sedentary behavior and subsequent lifestyle diseases. In line with that, valid and reliable submaximal tests of physical fitness is meaningful in both healthcare settings as well as in epidemiological and clinical studies. Taken together, the use of a scientifically tested tool to measure and evaluate physical fitness, is one of the most ethical approaches to initiate constructive public health advice.

7 SAMMANFATTNING

Aerob kondition (aerob kapacitet) är en viktig aspekt för prestationsförmåga, likväl som för hälsa och välbefinnande. Både genetik och träning styr vår kondition, och på befolkningsnivå är det stora individuella skillnader i kondition. De senaste 20–25 åren har sambandet mellan kondition och olika typer av ohälsa och sjukdom blivit allt mer vedertaget. Därmed har även intresset för att mätning och klassificering av aerob kapacitet ökat. Ett sätt att bedöma graden av kondition är att mäta maximal syreupptagningsförmåga (VO_2max). En direkt mätning av VO_2max hos en individ görs med laboratorieapparatur, under en maximal fysisk insats i en arbetsform som involverar stora delar av muskelmassan (ex. löpning, skidåkning eller cykling).

I många praktiska sammanhang är det inte lämpliga eller möjliga att genomföra direkta mätningar av VO_2max. Därför har det utvecklats ett antal olika submaximala tester för att estimera VO_2max. Dessa tester kan t.ex. utföras i form av gång-test, step-test eller test på cykelergometer. Många av dessa tester bygger baseras på en observerad hjärtfrekvens vid en viss given arbetsbelastning, och en känd relation mellan hjärtfrekvens och syreupptagning (VO_2).

För att ett submaximalt cykeltest ska kunna användas på ett givande sätt krävs det att testets validitet och reliabilitet är känd. Det är inte ovanligt att tester används på ett sätt som är utanför det vetenskapligt beprövade området. Till exempel är det till stor del okänt huruvida en förändring i VO_2max kan detekteras med hjälp av submaximala cykeltest. Ett submaximalt test kan exempelvis komma att användas i större longitudinella studier eller vid upprepade hälsokontroller på vårdcentraler och inom företagshälsovården. Tidigare forskning har visat att det populära Åstrandtestet har låg precision när det gäller att följa en förändring i VO_2max.

Vidare är det av intresse att undersöka om ett submaximalt arbetsprov utformat för vuxna även kan vara användbart på barn och ungdomar. God kondition är viktigt redan i unga år, och det är således intressant att enkelt kunna bedöma den aeroba kapaciteten även hos yngre individer. Däremot är det inte alltid lämpligt att direkt applicera ett test för vuxna på barn och ungdomar, på grund av vissa skillnader i cirkulatorisk respons till arbete. Således kan ett icke-korrigerat test för vuxna ge en större felskattning vid användning på barn och ungdomar.

Sammanfattningsvis kan sägas att det finns ett stort behov av enkla och icke-maximala mätningar av aerob kondition, samtidigt som forskningen gällande validitet och reliabilitet i olika grupper av befolkningen relativt bristfällig. Ett test har sällan överlägsen validitet i alla olika grupper, sett till exempelvis ålder, träningsgrad, olika sjukdomar och medicineringar, med mera. Syftet med föreliggande avhandling är att utveckla ett nytt submaximalt cykeltest för estimering av VO_2max, samt undersöka testets validitet och reliabilitet i olika grupper. Avhandlingen har fokuserat på följande punkter:

- att utveckla och validering av ett nytt submaximalt cykeltest för bestämning av kondition – VO_2max – hos vuxna (studie I och II),
- att vidareutveckla beräkningsmodellerna för skattning av VO_2max hos vuxna (delstudie II) samt hos barn och ungdomar (studie IV),
- att undersöka huruvida ett submaximalt cykeltest kan användas för att skatta förändring i VO_2max (studie III).

I studie I presenterades ett nytt submaximalt cykeltest för estimering av VO_2max, Ekblom-Bak testet (EB-testet). Män och kvinnor i blandade åldrar (n = 143) genomförde ett submaximalt cykeltest, och ett stegrande maximalt test på löpband för direkt bestämning VO_2max. Testerna genomfördes vid två tillfällen, med en veckas mellanrum. Resultaten visade att VO_2max kunde estimeras med god precision på gruppnivå, och med rimligt god precision på individnivå. Däremot ökade prediktionsfelet med ökad kondition: de med låg kondition överskattades (främst män) och de med hög kondition underskattades. Reliabilitetsanalys visade att det inte var någon signifikant skillnad mellan resultaten från det första och andra testtillfälle, vilket gör det möjligt att använda EB-testet utan att utföra något förtest eller tillvänjningstillfälle.

Delstudie II genomfördes med samma metod och procedur som studie I, med de rekryterade försökspersonerna hämtades främst ur grupper utanför testets tidigare valida spann. Den utbyggda modellgruppen utgjordes av 217 försökspersoner. Vidare genomfördes även en korsvalidering (n = 115) av de vidareutvecklade och könsspecifika beräkningsmodellerna. Försökspersonerna i korsvalideringsgruppen genomförde först ett EB-test för estimering av VO_2max, och därefter ett maximalt löpbandstest för bestämning VO_2max.

I studie III gjordes uppföljningstester på 35 försökspersoner som hade genomfört ett EB-test och ett maximalt test 5–8 år tidigare. Analyser gjordes av samstämmigheten mellan förändringar i estimerat och uppmätt VO_2max. Resultaten visade att förändringarna korrelerade till hög grad och att felskattningarna på individnivå var begränsade.

I den avslutande delstudien fick pojkar och flickor i åldrarna 10–15 år utföra ett EB-test och ett maximalt test på löpband för direkt mätning av VO_2max. Testerna genomfördes vid två olika testtillfällen med en veckas mellanrum. För att studera om validiteten var kopplad till biologisk mognad så genomfördes även en läkarundersökning med bestämning av pubertal mognadsgrad, enligt Tannerskalans olika steg. Resultaten visade att EB-testet har rimlig validitet hos barn före och under puberteten, samt att precisionen ökar när prepubertala pojkar analyseras med den ekvation som är framtagen för kvinnor.

8 ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to the **Åstrand Laboratory of Work Physiology**. I am so grateful for the opportunity to conduct my research at a laboratory with a long tradition of research in exercise physiology. It has been a delight to work and belong to the Åstrand Laboratory. It is not only a workplace – it is a second home, a place to grow, and a place full of friends and colleagues. Some people deserves an extra hearty thought. Thank you:

Örjan Ekblom, my main supervisor, for giving me a second chance when I needed it the most. Without you, there would have been no dissertation. Thank you for always having my back. Thank you for all your patience and understanding. You provide so much energy and passion to the research process, even in times of trouble and doubts. Last, but not least, you can make me laugh – even in my most depressed moments.

Björn Ekblom, a never-ending source of scientific knowledge and new ideas. I can literally say that the studies in my thesis would not have happened if it wasn't for you. You are an amazing professor emeritus, always ready to assist in an experiment, read a manuscript, or consult in a tricky research question. Also, thank you for being a supportive and inspiring supervisor during my work with my master thesis.

Elin Ekblom-Bak, my "scientific sister" and cycle test role model, you are an important source of knowledge and inspiration. You have always trusted me and my abilities, and you have helped me to develop into an independent and responsible PhD student and researcher.

Tony Bohman, my co-supervisor, for your support and guidance. You have contributed with advice and thoughtful reflection during my later years as a PhD student.

PO Åstrand, deceased professor emeritus, and a true pioneer in the field of submaximal cycle ergometer testing. Thank you for inspiration, and thank you for all the important work that you did in the field of exercise physiology.

Filip Larsen, also known as **Dr. Spenat,** our own nitro vegan and ultra-endurance machine. You do your own thing, and you do it with great success. You are the crazy scientist, and it is awesome! Thank you for being my friend and colleague for so many years.

Gustav Olsson, we have walked side by side through the academy for 12 years now. I have liked you and appreciated your company from day one. You are more than a former class mate and colleague, you are a friend for life. You know me – and you understand me – better than most people do. You have been my primary source of respite and ventilation for so many years. Thanks for bearing with me! And by the way, I will always call all my fish "Beta Bengt" 3, 4, 5, and so on.

Mikael Mattson, my former supervisor, adventure racing friend and endurance phenomenon. Always "on the go" and always jetlagged. I am so happy and thankful for

your efforts in introducing me to work physiology and research. And it was a pleasure to borrow your apartment in Menlo Park during summer 2016. Thank you!

Marcus Moberg and **William Apró,** my roommates for a great couple of years. The two of you have been a true inspiration when it comes to determination and diligence. Keep up the good work. And Marcus; no one can execute a session on the crosstrainer like you do!

Daniele Cardinale, the Italian whirlwind at the laboratory. I will never forget your hyperoxi research, the pain from cycling 3x8 minutes at VO_2max, and all the blood and equipment and scientists that were involved in your epic experiments. Your determination and focus will lead to great (scientific) success in the future.

Sebastian Edman and **Emil Bojsen-Møller,** newly recruited PhD students and more important, the most loyal gym buddies I have ever had. Thanks for putting up with me, always cheering for new records in the bench press, and often adding a laugh to our never-ending workouts.

Andrea Eggers, "ninjaslaven". You are the best Master Student co-worker I have ever met! I wish you a great career in science and research. You are such a strong woman.

Mikael Salomonsson Flockhart, thank you for all the good times at work. You have been inspiring company during many painful training sessions at the cycle ergometers at GIH. You are truly devoted to training, and research. Lastly, you have the ability to discuss with me even in my darkest moments, without being hated. That is very unusual!

Marjan Pontén, lab manager and human being. You are always helpful and caring, and the door to your office is open for anyone who needs a quick chat or deep talk. The laboratory would not be the same without you.

The BMC laboratory, Maria Ekblom and **Olga Tarassova,** for helpful and friendly advice and invaluable help with figures and formatting for my research articles.

Niklas Psilander, **Karin Söderlund, Eva Blomstrand, Peter Schantz, Eva Andersson, Kerstin Hamrin** and **Kristjan Oddson –** we have worked in the same laboratory for so many years. Thank you all for good times! And a special thanks to **Jane Salier Eriksson** and **Kate Bolam,** who have helped me with my English writing.

Johnny Nilsson, for your inspiring energy and ingenuity, and **Kicki Hård af Segerstad, Anna Tidén** and **Elis Weslien** for the good times in "terränglådan".

Alexander Ovendal and **Mårten Fredriksson** for the good times at **LTIV** – the place to be.

Adam Stenman, MD, PhD, and Ironman, for your expertise and help during the work with study IV. You are so clever and educated, but still funny and kind. I wish you all the best in the future!

Torbjörn Helge, Manne Godhe, Linnéa Eriksson, Johan Petersson, for all the help with the tests in study I and II, and **Catherine Fahlen,** for your kindness and assistance with the medical examinations in study IV.

Friends from GIH, thank you all for your company through good and bad times.

Coffee, Pepsi Max, and chewing gum, "for making life a little more worth living, for lifting my spirits in times of need and simply for keeping me awake."

The GIH gym, a place for training, as well as a life style. All the time at the gym have resulted in a strong and robust physique. The hours at the gym have been an important ventilator in times of setbacks in research. All the strength training has also been a necessary preparation for my future career in the Swedish Armed Forces. A greeting to every single training buddy out there: you rock!

Norrland, with regard to the special and beloved people in Örnsköldsvik. I have always been welcomed with open arms when the stressful PhD-life in Stockholm has been too hard on me. I have lived, loved and laughed in the High Coast, and I want to send extra big thanks to the Lindnords and Göransson-Olsson.

My family, you are the reason I'm here. My mother, and my sisters Elvira and Lovisa – there are no words to describe what you mean to me. Dear father in heaven, I'm so grateful for all wisdom and perseverance you taught me while still around. I know that you would have been proud of me, for what I am, for what I have done, and for what I will do in the future.

Lastly, I want to thank all the **subjects**. Hundreds of people have volunteered for the studies, and they have contributed to science by cycling and running to complete exhaustion. Thank you so much!

The studies in this thesis have received economical support from the Swedish Armed Forces' Research Authority – research grant no. AF9220914.

9 REFERENCES

1. Bassett DR, Jr., Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. Med Sci Sports Exerc. 2000;32(1):70-84.
2. Bergh U, Kanstrup IL, Ekblom B. Maximal oxygen uptake during exercise with various combinations of arm and leg work. J Appl Physiol. 1976;41(2):191-6.
3. Rowell LB. Human circulation: regulation during physical stress. New York: Oxford University Press. 1986.
4. Hill AV, Long CNH, Lupton H. Muscular Exercise, Lactic Acid, and the Supply and Utilisation of Oxygen. Proceedings of the Royal Society of London Series B, Containing Papers of a Biological Character. 1924;97(681):84-138.
5. Hill AV, Lupton H. Muscular Exercise, Lactic Acid, and the Supply and Utilization of Oxygen. QJM: An International Journal of Medicine. 1923;os-16(62):135-71.
6. Liljestrand G, J. Lindhard. Respirationsversuche beim Gehen, Laufen, Ski- und Schlittschuhlaufen. Skand Arch Physiol. 1920(39):167-206.
7. Schutte NM, Nederend I, Hudziak JJ, Bartels M, de Geus EJ. Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. Physiol Genomics. 2016;48(3):210-9.
8. Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, et al. Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. J Appl Physiol (1985). 1999;87(3):1003-8.
9. Purge P, Jurimae J, Jurimae T. Hormonal and psychological adaptation in elite male rowers during prolonged training. J Sports Sci. 2006;24(10):1075-82.
10. Larsen HB. Kenyan dominance in distance running. Comp Biochem Physiol A Mol Integr Physiol. 2003;136(1):161-70.
11. Coyle EF. Improved muscular efficiency displayed as Tour de France champion matures. J Appl Physiol (1985). 2005;98(6):2191-6.
12. Saltin B, Astrand PO. Maximal oxygen uptake in athletes. J Appl Physiol. 1967;23(3):353-8.
13. Helgerud J. Maximal oxygen uptake, anaerobic threshold and running economy in women and men with similar performances level in marathons. European journal of applied physiology and occupational physiology. 1994;68(2):155-61.
14. Eriksen L, Gronbaek M, Helge JW, Tolstrup JS. Cardiorespiratory fitness in 16 025 adults aged 18-91 years and associations with physical activity and sitting time. Scand J Med Sci Sports. 2016;26(12):1435-43.
15. Ekblom B, Engstrom LM, Ekblom O. Secular trends of physical fitness in Swedish adults. Scand J Med Sci Sports. 2007;17(3):267-73.
16. Loe H, Rognmo O, Saltin B, Wisloff U. Aerobic capacity reference data in 3816 healthy men and women 20-90 years. PloS one. 2013;8(5):e64319.

17. Helgerud J, Hoydal K, Wang E, Karlsen T, Berg P, Bjerkaas M, et al. Aerobic high-intensity intervals improve VO2max more than moderate training. Med Sci Sports Exerc. 2007;39(4):665-71.
18. Psilander N, Wang L, Westergren J, Tonkonogi M, Sahlin K. Mitochondrial gene expression in elite cyclists: effects of high-intensity interval exercise. Eur J Appl Physiol. 2010;110(3):597-606.
19. Billat LV. Interval training for performance: a scientific and empirical practice. Special recommendations for middle- and long-distance running. Part II: anaerobic interval training. Sports Med. 2001;31(2):75-90.
20. Billat LV. Interval training for performance: a scientific and empirical practice. Special recommendations for middle- and long-distance running. Part I: aerobic interval training. Sports Med. 2001;31(1):13-31.
21. Blair SN, Kohl HW, 3rd, Paffenbarger RS, Jr., Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. JAMA. 1989;262(17):2395-401.
22. Blair SN, Kampert JB, Kohl HW, 3rd, Barlow CE, Macera CA, Paffenbarger RS, Jr., et al. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. JAMA. 1996;276(3):205-10.
23. Barry VW, Baruth M, Beets MW, Durstine JL, Liu J, Blair SN. Fitness vs. fatness on all-cause mortality: a meta-analysis. Prog Cardiovasc Dis. 2014;56(4):382-90.
24. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA. 2009;301(19):2024-35.
25. Morris CK, Ueshima K, Kawaguchi T, Hideg A, Froelicher VF. The prognostic value of exercise capacity: a review of the literature. American heart journal. 1991;122(5):1423-31.
26. Simmons RK, Griffin SJ, Steele R, Wareham NJ, Ekelund U, ProActive Research T. Increasing overall physical activity and aerobic fitness is associated with improvements in metabolic risk: cohort analysis of the ProActive trial. Diabetologia. 2008;51(5):787-94.
27. Zaccardi F, O'Donovan G, Webb DR, Yates T, Kurl S, Khunti K, et al. Cardiorespiratory fitness and risk of type 2 diabetes mellitus: A 23-year cohort study and a meta-analysis of prospective studies. Atherosclerosis. 2015;243(1):131-7.
28. Fedewa MV, Gist NH, Evans EM, Dishman RK. Exercise and insulin resistance in youth: a meta-analysis. Pediatrics. 2014;133(1):e163-74.
29. Saavedra JM, Escalante Y, Garcia-Hermoso A. Improvement of aerobic fitness in obese children: a meta-analysis. International journal of pediatric obesity : IJPO : an official journal of the International Association for the Study of Obesity. 2011;6(3-4):169-77.
30. McMurray RG, Bangdiwala SI, Harrell JS, Amorim LD. Adolescents with metabolic syndrome have a history of low aerobic fitness and physical activity levels. Dynamic medicine : DM. 2008;7:5.

31. Janz KF, Dawson JD, Mahoney LT. Increases in physical fitness during childhood improve cardiovascular health during adolescence: the Muscatine Study. Int J Sports Med. 2002;23 Suppl 1:S15-21.
32. Brown HE, Pearson N, Braithwaite RE, Brown WJ, Biddle SJ. Physical activity interventions and depression in children and adolescents : a systematic review and meta-analysis. Sports Med. 2013;43(3):195-206.
33. Hillman CH, Buck SM, Themanson JR, Pontifex MB, Castelli DM. Aerobic fitness and cognitive development: Event-related brain potential and task performance indices of executive control in preadolescent children. Dev Psychol. 2009;45(1):114-29.
34. Castelli DM, Hillman CH, Buck SM, Erwin HE. Physical fitness and academic achievement in third- and fifth-grade students. J Sport Exerc Psychol. 2007;29(2):239-52.
35. Fedewa AL, Ahn S. The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: a meta-analysis. Research quarterly for exercise and sport. 2011;82(3):521-35.
36. Stroth S, Kubesch S, Dieterle K, Ruchsow M, Heim R, Kiefer M. Physical fitness, but not acute exercise modulates event-related potential indices for executive control in healthy adolescents. Brain research. 2009;1269:114-24.
37. Prout W. On the quantity of carbonic acid gas emitted from the lungs during respiration at different times and under different circumstances. Thomson's Ann Philosoph. 1813(2):328-42.
38. Tipton CM. Historical perspective: origin to recognition. In: Tipton CM, editor. ACSM's advanced exercise physiology. Baltimore: Lippincott Williams & Wilkins; 2006. p. 30.
39. Douglas CG. A method for determining the total respiratory exchange in man. J Physiol. 1911;42:17-8.
40. Shephard RJ. Open-circuit respirometry: a brief historical review of the use of Douglas bags and chemical analyzers. Eur J Appl Physiol. 2017;117(3):381-7.
41. Macfarlane DJ. Automated metabolic gas analysis systems: a review. Sports Med. 2001;31(12):841-61.
42. Carter J, Jeukendrup AE. Validity and reliability of three commercially available breath-by-breath respiratory systems. Eur J Appl Physiol. 2002;86(5):435-41.
43. Hopker JG, Jobson SA, Gregson HC, Coleman D, Passfield L. Reliability of cycling gross efficiency using the Douglas bag method. Med Sci Sports Exerc. 2012;44(2):290-6.
44. Whipp BJ, Ward SA. Physiological determinants of pulmonary gas exchange kinetics during exercise. Med Sci Sports Exerc. 1990;22(1):62-71.
45. Myers J, Walsh D, Sullivan M, Froelicher V. Effect of sampling on variability and plateau in oxygen uptake. J Appl Physiol (1985). 1990;68(1):404-10.
46. Foss O, Hallen J. Validity and stability of a computerized metabolic system with mixing chamber. Int J Sports Med. 2005;26(7):569-75.

47. Rietjens GJ, Kuipers H, Kester AD, Keizer HA. Validation of a computerized metabolic measurement system (Oxycon-Pro) during low and high intensity exercise. Int J Sports Med. 2001;22(4):291-4.
48. Astrand PO, Saltin B. Maximal oxygen uptake and heart rate in various types of muscular activity. J Appl Physiol. 1961;16:977-81.
49. Åstrand P-O. Experimental studies of physical working capacity in relation to sex and age. Copenhagen, Denmark: Kungliga Gymnastiska Centralinstitutet; 1952.
50. Taylor HL, Buskirk E, Henschel A. Maximal oxygen intake as an objective measure of cardio-respiratory performance. J Appl Physiol. 1955;8(1):73-80.
51. Kamon E, Metz KF, Pandolf KB. Climbing and cycling with additional weights on the extremities. J Appl Physiol. 1973;35(3):367-70.
52. Hermansen L, Saltin B. Oxygen uptake during maximal treadmill and bicycle exercise. J Appl Physiol. 1969;26(1):31-7.
53. Fairshter RD, Walters J, Salness K, Fox M, Minh VD, Wilson AF. A comparison of incremental exercise tests during cycle and treadmill ergometry. Med Sci Sports Exerc. 1983;15(6):549-54.
54. Buchfuhrer MJ, Hansen JE, Robinson TE, Sue DY, Wasserman K, Whipp BJ. Optimizing the exercise protocol for cardiopulmonary assessment. Journal of applied physiology: respiratory, environmental and exercise physiology. 1983;55(5):1558-64.
55. Shephard RJ, Allen C, Benade AJ, Davies CT, Di Prampero PE, Hedman R, et al. The maximum oxygen intake. An international reference standard of cardiorespiratory fitness. Bull World Health Organ. 1968;38(5):757-64.
56. Dickinson S. The Dynamics of Bicycle Pedalliug. Proceedings of the Royal Society of London Series B, Containing Papers of a Biological Character. 1928;103(724):225-33.
57. Garry RC, Wishart GM. On the existence of a most efficient speed in bicycle pedalling, and the problem of determining human muscular efficiency. J Physiol. 1931;72(4):425-37.
58. Eckermann P, Millahn H. P. Der Einfluß der Drehzahl auf die Herzfrequenz und die Sauerstoffaufnahme bei konstanter Leistung am Fahrradergometer. Int Z Angew Physiol [einschl Arbeitsphysiol]. 1967;23(4):340–44.
59. Astrand I, Astrand PO, Hallback I, Kilbom A. Reduction in maximal oxygen uptake with age. J Appl Physiol. 1973;35(5):649-54.
60. Marsh AP, Martin PE. The association between cycling experience and preferred and most economical cadences. Med Sci Sports Exerc. 1993;25(11):1269-74.
61. Marsh AP, Martin PE, Foley KO. Effect of cadence, cycling experience, and aerobic power on delta efficiency during cycling. Med Sci Sports Exerc. 2000;32(9):1630-4.
62. Seabury JJ, Adams WC, Ramey MR. Influence of pedalling rate and power output on energy expenditure during bicycle ergometry. Ergonomics. 1977;20(5):491-8.

63. McKay GA, Banister EW. A comparison of maximum oxygen uptake determination by bicycle ergometry at various pedaling frequencies and by treadmill running at various speeds. European journal of applied physiology and occupational physiology. 1976;35(3):191-200.
64. Patterson RP, Moreno MI. Bicycle pedalling forces as a function of pedalling rate and power output. Med Sci Sports Exerc. 1990;22(4):512-6.
65. Whitty AG, Murphy AJ, Coutts AJ, Watsford ML. Factors associated with the selection of the freely chosen cadence in non-cyclists. Eur J Appl Physiol. 2009;106(5):705-12.
66. Sacchetti M, Lenti M, Di Palumbo AS, De Vito G. Different effect of cadence on cycling efficiency between young and older cyclists. Med Sci Sports Exerc. 2010;42(11):2128-33.
67. Bouchard C, Godbout P, Mondor JC, Leblanc C. Specificity of maximal aerobic power. European journal of applied physiology and occupational physiology. 1979;40(2):85-93.
68. Holmer I, Lundin A, Eriksson BO. Maximum oxygen uptake during swimming and running by elite swimmers. J Appl Physiol. 1974;36(6):711-4.
69. Howley ET, Bassett DR, Jr., Welch HG. Criteria for maximal oxygen uptake: review and commentary. Med Sci Sports Exerc. 1995;27(9):1292-301.
70. Nielsen M, Hansen O. Maximale körperliche Arbeit bei Atmung O2-reicher Luft1. Skandinavisches Archiv Für Physiologie. 1937;76(1-2):37-59.
71. Åstrand I. Aerobic work capacity in men and women with special reference to age. Stockholm, Sweden: Kungliga Gymnastiska Centralinstitutet; 1960.
72. Whipp BJ, Ward SA, Lamarra N, Davis JA, Wasserman K. Parameters of ventilatory and gas exchange dynamics during exercise. Journal of applied physiology: respiratory, environmental and exercise physiology. 1982;52(6):1506-13.
73. Barstow TJ, Mole PA. Linear and nonlinear characteristics of oxygen uptake kinetics during heavy exercise. J Appl Physiol (1985). 1991;71(6):2099-106.
74. Roston WL, Whipp BJ, Davis JA, Cunningham DA, Effros RM, Wasserman K. Oxygen uptake kinetics and lactate concentration during exercise in humans. The American review of respiratory disease. 1987;135(5):1080-4.
75. Whipp BJ. The slow component of O2 uptake kinetics during heavy exercise. Med Sci Sports Exerc. 1994;26(11):1319-26.
76. Billat VL, Richard R, Binsse VM, Koralsztein JP, Haouzi P. The V(O2) slow component for severe exercise depends on type of exercise and is not correlated with time to fatigue. J Appl Physiol (1985). 1998;85(6):2118-24.
77. Jones AM, McConnell AM. Effect of exercise modality on oxygen uptake kinetics during heavy exercise. European journal of applied physiology and occupational physiology. 1999;80(3):213-9.
78. Jones AM, DiMenna F, Lothian F, Taylor E, Garland SW, Hayes PR, et al. 'Priming' exercise and O2 uptake kinetics during treadmill running. Respiratory physiology & neurobiology. 2008;161(2):182-8.

79. Whipp BJ, Davis JA, Torres F, Wasserman K. A test to determine parameters of aerobic function during exercise. Journal of applied physiology: respiratory, environmental and exercise physiology. 1981;50(1):217-21.
80. Davies B, Daggett A, Jakeman P, Mulhall J. Maximum oxygen uptake utilising different treadmill protocols. Br J Sports Med. 1984;18(2):74-9.
81. Lawler J, Powers SK, Dodd S. A time-saving incremental cycle ergometer protocol to determine peak oxygen consumption. Br J Sports Med. 1987;21(4):171-3.
82. Faulkner J, Mauger AR, Woolley B, Lambrick D. The efficacy of a self-paced VO2max test during motorized treadmill exercise. International journal of sports physiology and performance. 2015;10(1):99-105.
83. Astorino TA, White, A.C., Dalleck, L.C. Supramax Testing to Confirm Attainment of VO2max in Sedentary Men and Women. Int J Sports Med. 2009;30(4):279-84.
84. Astorino TA, White, A.C. Assessment of anaerobic power to verify VO2max attainment. Clin Physiol Funct Imaging. 2010;30(4):294-300.
85. Scharhag-Rosenberger F, Carlsohn, A., Cassel, M., Mayer, F.,Scharhag, J. How to test maximal oxygen uptake: a study on timing and testing procedure of a supramaximal verification test. Appl Physiol Nutr Metab. 2011;36(1):153-60.
86. Midgley AW, McNaughton, L.R., Carroll, S. Verification phase as a useful tool in the determination of the maximal oxygen uptake of distance runners. Appl Physiol Nutr Metab. 2006;31(5):541-8.
87. Rowland T, Cunningham, L. Oxygen uptake plateau during maximal treadmill exercise in children. Chest. 1992;101:485-9.
88. Rowland TW. Does peak VO2 reflect VO2max in children?: evidence from supramaximal testing. Med Sci Sports Exerc. 1993;25(6):689-93.
89. Barker AR, Williams, C.A., Jones, A.M., Armstrong, N. Establishing maximal oxygen uptake in young people during a ramp cycle test to exhaustion. Br J Sports Med. 2011;45(6):498-503.
90. de Groot JF, Takken T, de Graaff S, Gooskens RH, Helders PJ, Vanhees L. Treadmill testing of children who have spina bifida and are ambulatory: does peak oxygen uptake reflect maximum oxygen uptake? Phys Ther. 2009;89(7):679-87.
91. Astorino TA, Willey J, Kinnahan J, Larsson SM, Welch H, Dalleck LC. Elucidating determinants of the plateau in oxygen consumption at VO2max. Br J Sports Med. 2005;39(9):655-60; discussion 60.
92. Beltz NM, Gibson AL, Janot JM, Kravitz L, Mermier CM, Dalleck LC. Graded Exercise Testing Protocols for the Determination of VO2max: Historical Perspectives, Progress, and Future Considerations. Journal of sports medicine (Hindawi Publishing Corporation). 2016;2016:3968393.
93. Duncan GE, Howley ET, Johnson BN. Applicability of VO2max criteria: discontinuous versus continuous protocols. Med Sci Sports Exerc. 1997;29(2):273-8.
94. Day JR, Rossiter HB, Coats EM, Skasick A, Whipp BJ. The maximally attainable VO2 during exercise in humans: the peak vs. maximum issue. J Appl Physiol (1985). 2003;95(5):1901-7.

95. Astorino T, Robergs RA., Ghiasvand, F., Marks, D., Burns, S. Incidence Of The Oxygen Plateau at VO2max During Exercise Testing To Volitional Fatigue. Journal of Exercise Physiology Online. 2000;3(4):1-12.
96. Armstrong N, Kirby BJ, McManus AM, Welsman JR. Aerobic fitness of prepubescent children. Annals of human biology. 1995;22(5):427-41.
97. Edvardsen E, Hem E, Anderssen SA. End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. PloS one. 2014;9(1):e85276.
98. Bangsbo J, Iaia FM, Krustrup P. The Yo-Yo intermittent recovery test : a useful tool for evaluation of physical performance in intermittent sports. Sports Med. 2008;38(1):37-51.
99. Bandyopadhyay A. Prediction of maximum oxygen uptake by using the heart rate ratio method in Indian university students. J Hum Ergol (Tokyo). 2013;42(1-2):55-63.
100. Weiglein L, Herrick J, Kirk S, Kirk EP. The 1-mile walk test is a valid predictor of VO(2max) and is a reliable alternative fitness test to the 1.5-mile run in U.S. Air Force males. Military medicine. 2011;176(6):669-73.
101. Krustrup P, Mohr M, Amstrup T, Rysgaard T, Johansen J, Steensberg A, et al. The yo-yo intermittent recovery test: physiological response, reliability, and validity. Med Sci Sports Exerc. 2003;35(4):697-705.
102. Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. United States Armed Forces medical journal. 1959;10(6):675-88.
103. Pollock ML, Bohannon RL, Cooper KH, Ayres JJ, Ward A, White SR, et al. A comparative analysis of four protocols for maximal treadmill stress testing. American heart journal. 1976;92(1):39-46.
104. Pollock ML, Foster C, Schmidt D, Hellman C, Linnerud AC, Ward A. Comparative analysis of physiologic responses to three different maximal graded exercise test protocols in healthy women. American heart journal. 1982;103(3):363-73.
105. Bruce RA, Blackmon JR, Jones JW, Strait G. Exercising testing in adult normal subjects and cardiac patients. Pediatrics. 1963;32:Suppl 742-56.
106. Margaria R, Aghemo P, Rovelli E. Indirect determination of maximal O2 consumption in man. J Appl Physiol. 1965;20(5):1070-3.
107. Maritz JS, Morrison JF, Peter J, Strydom NB, Wyndham CH. A practical method of estimating an individual's maximal oxygen intake. Ergonomics. 1961;4(2):97-122.
108. Astrand PO, Ryhming I. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during sub-maximal work. J Appl Physiol. 1954;7(2):218-21.
109. Astrand I. Aerobic work capacity in men and women with special reference to age. Acta Physiol Scand Suppl. 1960;49(169):1-92.
110. Washburn RA, Montoye HJ. The validity of predicting VO2max in males age 10-39. J Sports Med Phys Fitness. 1984;24(1):41-8.
111. Higginbotham MB, Morris, K.G., Williams, R.S., McHale, P.A., Coleman, R.E., Cobb F.R. Regulation of Stroke Volume during Submaximal and Maximal Upright Exercise in Normal Man. Circ Res. 1986;58:281-91.

112. Ekblom B. Effect of physical training on oxygen transport system in man. Acta Physiol Scand Suppl. 1968;328:1-45.
113. Åstrand P-O, Rodahl K. Textbook of Work Physiology. Physiological Bases of Exercise.: McGraw-Hill, Inc.; 1977.
114. Kenney WL, Wilmore JH, Costill DL. Physiology of Sport and Exercise: Human Kinetics; 2015.
115. Ekblom B, Hermansen L. Cardiac output in athletes. J Appl Physiol. 1968;25(5):619-25.
116. Keith A, Flack M. The Form and Nature of the Muscular Connections between the Primary Divisions of the Vertebrate Heart. Journal of anatomy and physiology. 1907;41(Pt 3):172-89.
117. Ekblom B, Goldbarg AN, Kilbom A, Astrand PO. Effects of atropine and propranolol on the oxygen transport system during exercise in man. Scandinavian journal of clinical and laboratory investigation. 1972;30(1):35-42.
118. Faria IE, Drummond BJ. Circadian changes in resting heart rate and body temperature, maximal oxygen consumption and perceived exertion. Ergonomics. 1982;25(5):381-6.
119. Wahlberg I, Astrand I. Physical work capacity during the day and at night. Work-Environment-Health. 1973(10):65-8.
120. Davies CT, Sargeant AJ. Circadian variation in physiological responses to exercise on a stationary bicycle ergometer. British journal of industrial medicine. 1975;32(2):110-4.
121. Noordhof DA, de Koning JJ, van Erp T, van Keimpema B, de Ridder D, Otter R, et al. The between and within day variation in gross efficiency. Eur J Appl Physiol. 2010;109(6):1209-18.
122. Ebbeling CB, Ward A, Puleo EM, Widrick J, Rippe JM. Development of a single-stage submaximal treadmill walking test. Med Sci Sports Exerc. 1991;23(8):966-73.
123. Brouha L, Health CW, Graybiel A. Step test simple method of measuring physical fitness for hard muscular work in adult men. Rev Canadian Biol. 1943(2):86-92.
124. Sykes K. Chester step test; resource pack (Version 3). Cheshire, UK: Chester College of Higher Education; 1998.
125. Ryhming I. A modified Harvard step test for the evaluation of physical fitness. Arbeitsphysiologie; internationale Zeitschrift fur angewandte Physiologie. 1953;15(3):235-50.
126. Bennett H, Parfitt G, Davison K, Eston R. Validity of Submaximal Step Tests to Estimate Maximal Oxygen Uptake in Healthy Adults. Sports Med. 2016;46(5):737-50.
127. Wahlund H. Determination of the Physical Working Capacity. A Physiological and Clinical Study with special reference to Standardization of Cardio-Pulmonary Functional Tests. Stockholm: Karolinska Institutet; 1948.
128. Ekblom B, Gjessing E. Maximal oxygen uptake of the Easter Island population. J Appl Physiol. 1968;25(2):124-9.
129. Sjöstrand T. Changes in the Respiratory organs of workmen at one oresmelding work. Acta Med Scand. 1947(196):687-99.

130. McMurray RG, Guion WK, Ainsworth BE, Harrell JS. Predicting aerobic power in children. A comparison of two methods. J Sports Med Phys Fitness. 1998;38(3):227-33.
131. Bland J, Pfeiffer K, Eisenmann JC. The PWC170: comparison of different stage lengths in 11-16 year olds. Eur J Appl Physiol. 2012;112(5):1955-61.
132. Strandell T. Circulatory studies on helathy old men. With special reference to the limitation of the maximal physical working capacity. Acta medica Scandinavica Supplementum. 1964;414:Suppl 414:1-44.
133. Kasch FW. The Validity of the Astrand and Sjostrand Submaximal Tests. Physician and Sports medicine.12(8):47-54.
134. Golding LA, Myers, C. R., & Sinning, W. E. Y's way to physical fitness (3rd ed.). (Champaign, IL: Human Kinetics) 1989.
135. Beekley MD, Brechue WF, deHoyos DV, Garzarella L, Werber-Zion G, Pollock ML. Cross-validation of the YMCA submaximal cycle ergometer test to predict VO2max. Research quarterly for exercise and sport. 2004;75(3):337-42.
136. Jamnick NA, By S, Pettitt CD, Pettitt RW. Comparison of the YMCA and a Custom Submaximal Exercise Test for Determining VO2max. Med Sci Sports Exerc. 2016;48(2):254-9.
137. von Dobeln W, Astrand I, Bergstrom A. An analysis of age and other factors related to maximal oxygen uptake. J Appl Physiol. 1967;22(5):934-8.
138. Teraslinna P, Ismail AH, MacLeod DF. Nomogram by Astrand and Ryhming as a predictor of maximum oxygen intake. J Appl Physiol. 1966;21(2):513-5.
139. Jette M. A comparison between predicted VO2 max from the Astrand procedure and the Canadian Home Fitness Test. Canadian journal of applied sport sciences Journal canadien des sciences appliquees au sport. 1979;4(3):214-8.
140. Cink RE, Thomas TR. Validity of the Astrand-Ryhming nomogram for predicting maximal oxygen intake. Br J Sports Med. 1981;15(3):182-5.
141. Siconolfi SF, Cullinane EM, Carleton RA, Thompson PD. Assessing VO2max in epidemiologic studies: modification of the Astrand-Rhyming test. Med Sci Sports Exerc. 1982;14(5):335-8.
142. Hartung GH, Blancq RJ, Lally DA, Krock LP. Estimation of aerobic capacity from submaximal cycle ergometry in women. Med Sci Sports Exerc. 1995;27(3):452-7.
143. Jessup GT, Terry JW, Landiss CW. Prediction of workload for the Astrand-Rhyming test using stepwise multiple linear regression. J Sports Med Phys Fitness. 1975;15(1):37-42.
144. Terry JW, Tolson H, Johnson DJ, Jessup GT. A workload selection procedure for the Astrand-Rhyming test. J Sports Med Phys Fitness. 1977;17(4):361-6.
145. Ekblom-Bak E, Bjorkman F, Hellenius ML, Ekblom B. A new submaximal cycle ergometer test for prediction of VO2max. Scand J Med Sci Sports. 2014;24(2):319-26.
146. Reilly T, Robinson G, Minors DS. Some circulatory responses to exercise at different times of day. Med Sci Sports Exerc. 1984;16(5):477-82.

147. Halicka-Ambroziak HD, Eberhardt A, Romanowski W, Klammer M. Evaluation of the method of indirect determination of the maximum oxygen uptake after Astrand and Ryhming in smokers and subjects in a state of emotional excitation. J Sports Med Phys Fitness. 1975;15(1):33-6.
148. Neovius K, Rasmussen F, Sundstrom J, Neovius M. Forecast of future premature mortality as a result of trends in obesity and smoking: nationwide cohort simulation study. European journal of epidemiology. 2010;25(10):703-9.
149. Legge BJ, Banister EW. The Astrand-Ryhming nomogram revisited. J Appl Physiol (1985). 1986;61(3):1203-9.
150. Borresen J, Lambert MI. Autonomic control of heart rate during and after exercise : measurements and implications for monitoring training status. Sports Med. 2008;38(8):633-46.
151. Buchheit M, Chivot A, Parouty J, Mercier D, Al Haddad H, Laursen PB, et al. Monitoring endurance running performance using cardiac parasympathetic function. Eur J Appl Physiol. 2010;108(6):1153-67.
152. Scharhag-Rosenberger F, Meyer T, Walitzek S, Kindermann W. Time course of changes in endurance capacity: a 1-yr training study. Med Sci Sports Exerc. 2009;41(5):1130-7.
153. Davies CT. Limitations to the prediction of maximum oxygen intake from cardiac frequency measurements. J Appl Physiol. 1968;24(5):700-6.
154. Peel AB, Thomas SM, Dittus K, Jones LW, Lakoski SG. Cardiorespiratory fitness in breast cancer patients: a call for normative values. J Am Heart Assoc. 2014;3(1):e000432.
155. Mijwel S, Cardinale D, Ekblom-Bak E, Sundberg CJ, Wengstrom Y, Rundqvist H. Validation of 2 Submaximal Cardiorespiratory Fitness Tests in Patients With Breast Cancer Undergoing Chemotherapy. Rehabil Oncol. 2016;34(4):137-43.
156. May AM, van Weert E, Korstjens I, Hoekstra-Weebers JE, van der Schans CP, Zonderland ML, et al. Monitoring training progress during exercise training in cancer survivors: a submaximal exercise test as an alternative for a maximal exercise test? Archives of physical medicine and rehabilitation. 2010;91(3):351-7.
157. De Backer IC, Schep G, Hoogeveen A, Vreugdenhil G, Kester AD, van Breda E. Exercise testing and training in a cancer rehabilitation program: the advantage of the steep ramp test. Archives of physical medicine and rehabilitation. 2007;88(5):610-6.
158. Andersen LB, Sardinha LB, Froberg K, Riddoch CJ, Page AS, Anderssen SA. Fitness, fatness and clustering of cardiovascular risk factors in children from Denmark, Estonia and Portugal: the European Youth Heart Study. International journal of pediatric obesity : IJPO : an official journal of the International Association for the Study of Obesity. 2008;3 Suppl 1:58-66.
159. Hurtig-Wennlof A, Ruiz JR, Harro M, Sjostrom M. Cardiorespiratory fitness relates more strongly than physical activity to cardiovascular disease risk factors in healthy children and adolescents: the European Youth Heart Study. European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on

Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology. 2007;14(4):575-81.
160. Rizzo NS, Ruiz JR, Hurtig-Wennlof A, Ortega FB, Sjostrom M. Relationship of physical activity, fitness, and fatness with clustered metabolic risk in children and adolescents: the European youth heart study. The Journal of pediatrics. 2007;150(4):388-94.
161. Esmaeilzadeh S. The association between depressive symptoms and physical status including physical activity, aerobic and muscular fitness tests in children. Environmental health and preventive medicine. 2015;20(6):434-40.
162. Turley KR, Wilmore JH. Cardiovascular responses to treadmill and cycle ergometer exercise in children and adults. J Appl Physiol (1985). 1997;83(3):948-57.
163. Vinet A, Nottin S, Lecoq AM, Obert P. Cardiovascular responses to progressive cycle exercise in healthy children and adults. Int J Sports Med. 2002;23(4):242-6.
164. Unnithan V, Rowland TW. Use of Oxygen Pulse in Predicting Doppler-Derived Maximal Stroke Volume in Adolescents. Pediatr Exerc Sci. 2015;27(3):412-8.
165. Nottin S, Vinet A, Stecken F, N'Guyen LD, Ounissi F, Lecoq AM, et al. Central and peripheral cardiovascular adaptations to exercise in endurance-trained children. Acta physiologica Scandinavica. 2002;175(2):85-92.
166. de Prado DM, Dias RG, Trombetta IC. Cardiovascular, ventilatory, and metabolic parameters during exercise: differences between children and adults. Arq Bras Cardiol. 2006;87(4):e149-55.
167. Leclair E, Berthoin S, Borel B, Thevenet D, Carter H, Baquet G, et al. Faster pulmonary oxygen uptake kinetics in children vs adults due to enhancements in oxygen delivery and extraction. Scand J Med Sci Sports. 2013;23(6):705-12.
168. Rowland T, Potts, J., Potts, T., Sandor, G., Goff, D., Ferrone, L. Cardiac responses to progressive exercise in normal children: a synthesis. Med Sci Sports Exerc. 2000;32(2):253-9.
169. Turley KR, Wilmore, J.H. Cardiovascular responses to submaximal exercise in 7- to 9-yr-old boys and girls. Med Sci Sports Exerc. 1997;29(6):824-32.
170. Eriksson BO, Grimby G, Saltin B. Cardiac output and arterial blood gases during exercise in pubertal boys. J Appl Physiol. 1971;31(3):348-52.
171. Melo X, Fernhall B, Santos DA, Pinto R, Pimenta NM, Sardinha LB, et al. The acute effect of maximal exercise on central and peripheral arterial stiffness indices and hemodynamics in children and adults. Appl Physiol Nutr Metab. 2016;41(3):266-76.
172. Turley KR. Cardiovascular responses to exercise in children. Sports Med. 1997;24(4):241-57.
173. Wirth A, Trager E, Scheele K, Mayer D, Diehm K, Reischle K, et al. Cardiopulmonary adjustment and metabolic response to maximal and submaximal physical exercise of boys and girls at different stages of maturity. European journal of applied physiology and occupational physiology. 1978;39(4):229-40.

174. Williams JR, Armstrong N, Kirby BJ. The 4 mM blood lactate level as an index of exercise performance in 11-13 year old children. J Sports Sci. 1990;8(2):139-47.
175. Bar-Or O, Shephard RJ, Allen CL. Cardiac output of 10- to 13-year-old boys and girls during submaximal exercise. J Appl Physiol. 1971;30(2):219-23.
176. Marta CC, Marinho DA, Izquierdo M, Marques MC. Differentiating maturational influence on training-induced strength and endurance adaptations in prepubescent children. Am J Hum Biol. 2014;26(4):469-75.
177. Armstrong N. Aerobic fitness and physical activity in children. Pediatr Exerc Sci. 2013;25(4):548-60.
178. Kolle E, Steene-Johannessen J, Andersen LB, Anderssen SA. Objectively assessed physical activity and aerobic fitness in a population-based sample of Norwegian 9- and 15-year-olds. Scand J Med Sci Sports. 2010;20(1):e41-7.
179. Janz KF, Burns TL, Witt JD, Mahoney LT. Longitudinal analysis of scaling VO2 for differences in body size during puberty: the Muscatine Study. Med Sci Sports Exerc. 1998;30(9):1436-44.
180. Rowland TW. Developmental aspects of physiological function relating to aerobic exercise in children. Sports Med. 1990;10(4):255-66.
181. Graves LE, Batterham AM, Foweather L, McWhannell N, Hopkins ND, Boddy LM, et al. Scaling of peak oxygen uptake in children: a comparison of three body size index models. Med Sci Sports Exerc. 2013;45(12):2341-5.
182. Tanner JM. Fallacy of per-weight and per-surface area standards, and their relation to spurious correlation. J Appl Physiol. 1949;2(1):1-15.
183. Cooper DM, Weiler-Ravell D, Whipp BJ, Wasserman K. Aerobic parameters of exercise as a function of body size during growth in children. Journal of applied physiology: respiratory, environmental and exercise physiology. 1984;56(3):628-34.
184. Paterson DH, McLellan TM, Stella RS, Cunningham DA. Longitudinal study of ventilation threshold and maximal O2 uptake in athletic boys. J Appl Physiol (1985). 1987;62(5):2051-7.
185. Cooper DM, Weiler-Ravell D, Whipp BJ, Wasserman K. Growth-related changes in oxygen uptake and heart rate during progressive exercise in children. Pediatric research. 1984;18(9):845-51.
186. Bouchard C, Dionne FT, Simoneau JA, Boulay MR. Genetics of aerobic and anaerobic performances. Exerc Sport Sci Rev. 1992;20:27-58.
187. Stewart KJ, Gutin B. The prediction of maximal oxygen uptake before and after physical training in children. J Hum Ergol (Tokyo). 1975;4(2):153-62.
188. Woynarowska B. The validity of indirect estimations of maximal oxygen uptake in children 11-12 years of age. European journal of applied physiology and occupational physiology. 1980;43(1):19-23.
189. Binyildiz PO. Prediction of maximal oxygen uptake in boys 11--13 years of age. European journal of applied physiology and occupational physiology. 1980;43(3):213-9.
190. Ekblom O. Testing the validity of three submaximal ergometer tests for estimating maximal aerobic capacity in children. Acta paediatrica. 2014;103(5):559-63.

191. Mahoney C. 20-MST and PWC170 validity in non-Caucasian children in the UK. Br J Sports Med. 1992;26(1):45-7.
192. Rowland TW, Rambusch JM, Staab JS, Unnithan VB, Siconolfi SF. Accuracy of physical working capacity (PWC170) in estimating aerobic fitness in children. J Sports Med Phys Fitness. 1993;33(2):184-8.
193. Wells CL, Scrutton EW, Archibald LD, Cooke WP, De la Mothe JW. Physical working capacity and maximal oxygen uptake of teenaged athletes. Med Sci Sports. 1973;5(4):232-41.
194. Bouny E. Mesure du travail dépénse dans Iémploi da la bicyclette. Paris Ac Sci O R. 1896(122):1891-5.
195. Von Dobeln W. A simple bicycle ergometer. J Appl Physiol. 1954;7(2):222-4.
196. Åstrand P-O. Work tests with the bicycle ergometer. 1964.
197. Faria I, Dix C, Frazer C. Effect of body position during cycling on heart rate, pulmonary ventilation, oxygen uptake and work output. J Sports Med Phys Fitness. 1978;18(1):49-56.
198. Welbergen E, Clijsen LP. The influence of body position on maximal performance in cycling. European journal of applied physiology and occupational physiology. 1990;61(1-2):138-42.
199. Franke WD, Betz CB, Humphrey RH. Effects of rider position on continuous wave Doppler responses to maximal cycle ergometry. Br J Sports Med. 1994;28(1):38-42.
200. Ryschon TW, Stray-Gundersen J. The effect of body position on the energy cost of cycling. Med Sci Sports Exerc. 1991;23(8):949-53.
201. Grappe F, Candau R, Busso T, Rouillon JD. Effect of cycling position on ventilatory and metabolic variables. Int J Sports Med. 1998;19(5):336-41.
202. Heil DP, Wilcox AR, Quinn CM. Cardiorespiratory responses to seat-tube angle variation during steady-state cycling. Med Sci Sports Exerc. 1995;27(5):730-5.
203. Heil DP, Derrick TR, Whittlesey S. The relationship between preferred and optimal positioning during submaximal cycle ergometry. European journal of applied physiology and occupational physiology. 1997;75(2):160-5.
204. Sutton JR, Reeves JT, Wagner PD, Groves BM, Cymerman A, Malconian MK, et al. Operation Everest II: oxygen transport during exercise at extreme simulated altitude. J Appl Physiol (1985). 1988;64(4):1309-21.
205. Ostler LM, Betts JA, Gore CJ. Gross cycling efficiency is not altered with and without toe-clips. J Sports Sci. 2008;26(1):47-55.
206. Swain DP. Energy cost calculations for exercise prescription: an update. Sports Med. 2000;30(1):17-22.
207. Von Dobeln W. Human standard and maximal metabolic rate in relation to fat-free body mass. Acta Physiol Scand Suppl. 1956;37(126):1-79.
208. Cotes JE. Relationships of oxygen consumption, ventilation and cardiac frequency to body weight during standardized submaximal exercise in normal subjects. Ergonomics. 1969;12(3):415-27.
209. Berry MJ, Storsteen JA, Woodard CM. Effects of body mass on exercise efficiency and VO2 during steady-state cycling. Med Sci Sports Exerc. 1993;25(9):1031-7.

210. Lafortuna CL, Proietti M, Agosti F, Sartorio A. The energy cost of cycling in young obese women. Eur J Appl Physiol. 2006;97(1):16-25.
211. Astrand I, Astrand PO, Stunkard A. Oxygen intake of obese individuals during work on a bicycle ergometer. Acta physiologica Scandinavica. 1960;50:294-9.
212. Zoladz JA, Duda K, Majerczak J. VO2/power output relationship and the slow component of oxygen uptake kinetics during cycling at different pedaling rates: relationship to venous lactate accumulation and blood acid-base balance. Physiological research. 1998;47(6):427-38.
213. Ettema G, Loras HW. Efficiency in cycling: a review. Eur J Appl Physiol. 2009;106(1):1-14.
214. Gaesser GA, Brooks GA. Muscular efficiency during steady-rate exercise: effects of speed and work rate. J Appl Physiol. 1975;38(6):1132-9.
215. Stebbins CL, Moore JL, Casazza GA. Effects of cadence on aerobic capacity following a prolonged, varied intensity cycling trial. Journal of sports science & medicine. 2014;13(1):114-9.
216. Rokowski RJ, Spodick DH. Prandial effects on cardiac function and responses. American heart journal. 1989;118(5 Pt 1):1078-82.
217. Fagan TC, Sawyer PR, Gourley LA, Lee JT, Gaffney TE. Postprandial alterations in hemodynamics and blood pressure in normal subjects. American Journal of Cardiology.58(7):636-41.
218. Myers J, Forman DE, Balady GJ, Franklin BA, Nelson-Worel J, Martin BJ, et al. Supervision of exercise testing by nonphysicians: a scientific statement from the American Heart Association. Circulation. 2014;130(12):1014-27.
219. Noonan V, Dean E. Submaximal Exercise Testing: Clinical Application and Interpretation. Physical Therapy. 2000;80(8):782-807.
220. Hirsch GL, Sue DY, Wasserman K, Robinson TE, Hansen JE. Immediate effects of cigarette smoking on cardiorespiratory responses to exercise. J Appl Physiol (1985). 1985;58(6):1975-81.
221. Van Duser BL, Raven PB. The effects of oral smokeless tobacco on the cardiorespiratory response to exercise. Med Sci Sports Exerc. 1992;24(3):389-95.
222. Jessup GT, Tolson H, Terry JW. Prediction of maximal oxygen intake from Astrand-Rhyming test, 12-minute run, and anthropometric variables using stepwise multiple regression. Am J Phys Med. 1974;53(4):200-7.
223. Jessup GT, Riggs CE, Lambert J, Miller WD. The effect of pedalling speed on the validity of the Astrand-Rhyming aerobic work capacity test. J Sports Med Phys Fitness. 1977;17(4):367-71.
224. Achten J, Jeukendrup AE. Heart rate monitoring: applications and limitations. Sports Med. 2003;33(7):517-38.
225. Holter NJ. New method for heart studies. Science (New York, NY). 1961;134(3486):1214-20.
226. Laukkanen RM, Virtanen PK. Heart rate monitors: state of the art. J Sports Sci. 1998;16 Suppl:S3-7.

227. Mier CM, Domenick MA, Wilmore JH. Changes in stroke volume with beta-blockade before and after 10 days of exercise training in men and women. J Appl Physiol (1985). 1997;83(5):1660-5.
228. Joyner MJ, Freund BJ, Jilka SM, Hetrick GA, Martinez E, Ewy GA, et al. Effects of beta-blockade on exercise capacity of trained and untrained men: a hemodynamic comparison. J Appl Physiol (1985). 1986;60(4):1429-34.
229. Gibbons RJ, Balady GJ, Bricker JT, Chaitman BR, Fletcher GF, Froelicher VF, et al. ACC/AHA 2002 guideline update for exercise testing: summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). Circulation. 2002;106(14):1883-92.
230. Whipp BJ, Wasserman K. Oxygen uptake kinetics for various intensities of constant-load work. J Appl Physiol. 1972;33(3):351-6.
231. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med. 1970;2(2):92-8.
232. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc. 1982;14(5):377-81.
233. Tanner JM. Growth at Adolescence, 2nd Edition. Oxford: Blackwell scientific publications. 1962.
234. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child. 1969;44(235):291-303.
235. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child. 1970;45(239):13-23.
236. Midgley AW, Carroll, S. Emergence of the verification phase procedure for confirming ‘true’ VO2max. Scand J Med Sci Sports. 2009;19(3):313-22.
237. Bak EE. Physical activity, cardiorespiratory fitness, and abdominal obesity in relation to cardiovascular disease risk – epidemiological studies. Stockholm, Sweden: Karolinska Institutet, Stockholm, Sweden; 2013.
238. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. The American journal of clinical nutrition. 1990;51(2):241-7.
239. Williams L. Reliability of predicting maximal oxygen intake using the Astrand-Ryhming nomogram. Research quarterly. 1975;46(1):12-6.
240. Larsen FJ, Schiffer TA, Sahlin K, Ekblom B, Weitzberg E, Lundberg JO. Mitochondrial oxygen affinity predicts basal metabolic rate in humans. FASEB J. 2011;25(8):2843-52.
241. Froberg A, Larsson C, Berg C, Boldemann C, Raustorp A. Accelerometer-measured physical activity among adolescents in a multicultural area characterized by low socioeconomic status. International journal of adolescent medicine and health. 2016.