Measurement and interpretation of maximal aerobic power in children

Bareket Falk[1,2], Raffy Dotan[2]

Department of Kinesiology, Centre for Bone and Muscle Health, Faculty of Applied Health Sciences, Brock University, St Catharines, ON, Canada, L2S 3A1

Faculty of Applied Health Sciences, Brock University, St Catharines, ON, Canada, L2S 3A1

Citation: Falk B, Dotan R. Measurement and Interpretation of Maximal Aerobic Power in Children. Pediatr Exerc Sci. 2019 May 1;31(2):144-151. doi: 10.1123/pes.2018-0191. Epub 2018 Dec 19. PMID: 30567470.

DOI: 10.1123/pes.2018-0191

Corresponding author:

Bareket Falk
Department of Kinesiology
Faculty of Applied Health Sciences
Brock University

St Catharines, Ontario, L2S 3A1, Canada
Tel.: 905-688-5550 ext.4979   Fax: 905-688-8364
E-mail: bfalk@brocku.ca

Abstract

The assessment of maximal aerobic power ($\dot{V}O_2$max), in both children and adults, is an invaluable tool for the evaluation of exercise performance capacity, and general physical fitness in clinical, athletic, public-health, and research applications. The complexity of means and considerations, as well as varying specific aims of $\dot{V}O_2$max-testing, has prevented the formulation of a universally-applicable, standard testing protocol, in general, and for children in particular. Numerous tester-controllable factors, such as exercise modality, metabolic measurement system, testing protocol, or data reduction strategies, can affect both the measurement and interpretation of $\dot{V}O_2$max data. Although the general guiding principles are similar, children differ from adults in several aspects. One notable difference is the absence of a discernible $\dot{V}O_2$ plateau in children. Thus, the proper choice of equipment and procedures may be different for children than for adults. It is, therefore, the aim of this article to highlight the general and pediatric-specific considerations that may affect $\dot{V}O_2$max measurement and interpretation of results.

Measurement of aerobic power – Why is it important?

Maximal aerobic power ($\dot{V}O_2$max) is one of the two main constituents of aerobic capacity – the other one being aerobic endurance (percentage of $\dot{V}O_2$max that can be maintained for given distances or durations). While aerobic endurance is difficult to quantify due to its duration dependency, maximal aerobic power is finite and better lends itself to unequivocal testing protocols. Moreover, performance in most habitual recreational, and athletic exertions is determined more by aerobic power than by aerobic endurance. For this reason, the measurement of $\dot{V}O_2$max has become a hallmark of clinical and athletic fitness assessment, as well as in research. In this review, we will focus on the measurement and interpretation of aerobic power.

The term ′$\dot{V}O_2$max’ is often referred to as the maximal, whole-body aerobic power, measured in exercise involving large muscle-mass (e.g., running, rowing, cycling), and this is how it will be used in this review. It is considered to reflect the upper-ceiling of the integrated ability of the cardiopulmonary and skeletal muscle systems to uptake, transport and utilize oxygen. Assessment of aerobic power is important for the understanding of physiological function (or dysfunction) and exercise capacity in both health and disease, or disability, as well as in athletic endeavours.

The assessment of maximal aerobic power is important clinically in providing insight into possible dysfunction or maladaptive responses to exercise, which may take place in numerous pulmonary, cardiovascular, or muscular pathological conditions; in determining disease progression; or in assessing efficacy of various therapeutic interventions (33, 37). Indeed, in a recent review, Pianosi et al. (37) argued that, in pediatrics, $\dot{V}O_2$peak can serve as an excellent biomarker of disease severity. In the athletic context it is paramount in assessing the status of, or training-induced changes in $\dot{V}O_2$max, one of the single-most important factors of athletic performance. In research, $\dot{V}O_2$max or $\dot{V}O_2$peak is determined to reflect aerobic fitness, to evaluate occupational work capacity, to determine relative exercise intensity in an exercise or training study, or to determine the effects of various interventions, as well as developmental or environmental factors.

There are numerous excellent resources delineating recommended protocols and procedures involved with exercise testing, including the assessment of maximal aerobic power, of pediatric populations in laboratory settings (1, 8, 34, 40). The purpose of this commentary is not to review and reiterate these protocols or procedures but rather, to highlight some of the pediatric-specific issues and considerations that may affect the measurement and interpretation of the main outcome of these measurements, namely, $\dot{V}O_2$max or $\dot{V}O_2$peak.

An inappropriate protocol, use of an improper ergometer or metabolic system, or inappropriate data reduction, could all lead to $\dot{V}O_2$max mis-determination with potentially direct consequences on the conclusions drawn and the actions taken (see also below under Data Interpretation). However, the importance of understanding the various issues involved in proper $\dot{V}O_2$max assessment, extends well beyond the mere quest for enhanced accuracy and reliability. The long list of factors affecting $\dot{V}O_2$max determination and interpretation that will be discussed in this review, and particularly factors that differentiate children from adults, provide the basis for understanding the shortcomings that have affected previously published pediatric data to various degrees. As will become clear by the issues raised in this review, more than a few of the available pediatric $\dot{V}O_2$max studies can be suspected as underestimating true aerobic power, more so than might be the case in adults. Such underestimation may affect our understanding of the magnitude and temporal changes of aerobic power during growth and maturation in the wider context of developmental physiology, as well as our understanding of the nature of more practical issues, such as that of aerobic trainability in children (14).

$\dot{V}O_2$max vs. $\dot{V}O_2$peak

Maximal aerobic power reflects the body's capacity to transport and utilize oxygen to generate muscular power. This is generally referred to as maximal oxygen uptake or $\dot{V}O_2$max. In the current exercise literature, maximal and peak $\dot{V}O_2$ are often used interchangeably and the distinction between the two terms is not always clear. Indeed, a recent set of commentaries (5, 18) highlight the diversity of perspectives on the terminology. $\dot{V}O_2$max is defined as the highest $\dot{V}O_2$ attainable under optimal conditions and is usually characterized by an inability to increase $\dot{V}O_2$ despite further increase in exercise intensity.

The highest measured oxygen consumption is often termed '$\dot{V}O_2$max', '$\dot{V}O_2$peak', or 'mode-specific $\dot{V}O_2$max'. However, these terms are not synonymous. The objective or subjective criteria often employed for ascertaining $\dot{V}O_2$max attainment (see below), are often not met, particularly in children (9, 28, 41, 42). Some view $\dot{V}O_2$max as a hypothetical value that, in practice, is approached to varying degrees, but often not quite attained (18). The term '$\dot{V}O_2$peak' has been introduced to include all cases in which $\dot{V}O_2$max has clearly not been attained or verified, or where it is highly likely that this is the case. It accounts, among other, for such cases as exercise employing too small a muscle mass (e.g., cycling, or arm ergometry, vs. treadmill running), or when VO_2 fails to demonstrate a plateauing effect (as is often the case with children) and the attainment of $\dot{V}O_2$max cannot be ascertained. Insufficient motivation or low tolerance for discomfort are occasionally also reasons for submaximal $\dot{V}O_2$ and then $\dot{V}O_2$peak rather than $\dot{V}O_2$max is the more appropriate term. In cases where the maximal $\dot{V}O_2$ response to a specific exercise mode is examined (e.g., upper-body testing the wheelchair-bound on the one hand, or high-level kayakers, on the other), the term "mode-specific $\dot{V}O_2$max" has been suggested (18).

Measurement issues

a) Exercise Mode and Modality

In both children and adults, treadmill running and cycle-ergometry are the two most commonly used testing modalities for determining $\dot{V}O_2$max and $\dot{V}O_2$peak. Treadmill-derived outcomes are typically 7−15% higher than those obtained in cycle-ergometry, likely due to the use of a greater muscle mass which, in turn, potentially increases oxygen demand, reduces local discomfort, and delays fatigue (2, 27, 28, 47). This difference is similar to that reported in adults (24). When the involved muscle mass is even smaller, e.g., as in upper-body ergometry, the noted discrepancies can be much greater. Each modality has its advantages and disadvantages (see below), but importantly, when comparing $\dot{V}O_2$max or $\dot{V}O_2$peak between studies, groups, or individuals, it is important to take note of the employed exercise modality.

While familiarization and habituation with treadmill walking and running may be required before the testee feels sufficiently comfortable to perform the test, an important treadmill advantage is that it involves a universally familiar movement skill. This is often not the case with cycling, particularly with an inability to pedal at sufficiently-high cadences. Treadmills typically are the ergometers of choice for $\dot{V}O_2$max testing also because running employs sufficiently large muscle mass which facilitates maximal $\dot{V}O_2$ values that are characteristically 10 or even 15% higher than corresponding cycle-ergometer values. Treadmills, however, have some important disadvantages, most notably lesser safety. Unlike cycle-, rowing-, or most other ergometers, one can fall off a treadmill, which may be a particular issue with young, inexperienced children, or individuals with disability. Indeed, young children often report feeling more comfortable on a cycle ergometer than on a treadmill (28, 46). A special safety harness, or a spotter, are often required to avert the risk of falling. Locomotor limitations, such as those encountered by overweight or obese individuals, may favour the use of an alternative, non-weight bearing ergometer. Secondly, while speed and incline can be controlled and recorded on a treadmill, power output (unlike most other ergometers) cannot be controlled nor quantified. From a practical perspective, a disadvantage of the treadmill is its sheer size and very limited portability. Treadmills can also be noisy and possibly intimidating to the uninitiated. When additional measurements other than $\dot{V}O_2$ are required (e.g., ECG, blood pressure, or lactate sampling), the jerky nature of running may be a hindrance.

The typical cycle-ergometer, while often preferred in clinical settings, is frequently geometrically inappropriate for young or small children (e.g., saddle height, crank length, pedal size, or arm reach). Although pediatric-specific cycle-ergometers and adjustable cranks that can accommodate even the smallest children are now commercially-available, important factors in maximizing cycling performance are proper setup of saddle height and arm reach and a firm foot-pedal connection via toe-clips or other means.

b) VO_2 Testing Equipment

Collection and Gas Analysis Systems. In terms of accuracy, the gold standard for gas collection and analysis is still the Douglas Bag (or equivalent) method. However, the advent of the online, automated ‘metabolic carts’, with their superior convenience and promptness of response, has largely rendered the Douglas Bag method as a historical anecdote, except for some specific special applications. In common practical use, only mixing-chamber-based and the breath-by-breath (BxB) metabolic measurement systems remain.

BxB systems, while becoming cheaper and more ubiquitous, ought to be systems of choice when monitoring fast-changing $\dot{V}O_2$ kinetics. Their fast-response, however, comes at the expense of accuracy (see Data Reduction section below). Thus, when accuracy of $\dot{V}O_2$max is important, mixing-chamber-type is the system of choice. A mixing chamber greatly minimizes the breathing-related fluctuations, thus increasing gas-analysis accuracy.

Mixing-Chamber and Dead-Space Size. Most mixing-chamber-type systems are geared for adult testing in terms of mixing-chamber and breathing-hose combined volume, which often exceeds 6 L. As children’s tidal volume can be £50% that of adults, it might take 2−3 times longer for the mixing-chamber-measured gas concentrations to reflect changes in expired air concentrations. This delay renders the mixing chamber a buffer that smoothens out $\dot{V}O_2$-response peaks and results in underestimation of the true maximal or peak values. Thus, the combined mixing-chamber/hose volume for prepubertal pediatric testing should, based on our experience, typically not exceed 3 L.

Breathing Interface. The mouthpiece/breathing-valve (+ nose-clip) combination has been the dominant setup. Its main advantage has been superior leak-proof operation, but this constrained breathing mode may be stressful to some. Breathing masks avoid that distress but cannot always guarantee proper seal. Current masks are better form-fitting and more leak-proof than earlier versions, even for small individuals. It should be noted that, depending on the particular metabolic measurement system, breathing masks are often not interchangeable with the traditional breathing valves. To assure best fit and minimal dead space for pediatric testing, the equipment chosen ought to be smaller or pediatric-specific (e.g., breathing hose, mask, mouthpiece).

c) Specificity

Generally, the treadmill would be the test modality of choice for $\dot{V}O_2$max determination, followed by the cycle-ergometer, if treadmill testing is inappropriate or impractical. However, when testing special populations, for example youth or adult athletes, it is advisable to employ the test modality most similar to their specialty, namely; runners on a treadmill, cyclists on a cycle-ergometer, rowers on a rowing ergometer, kayakers on a kayaking-specific-, or arm-cycling-ergometer, swimmers in a swimming flume or arm-cycling ergometer, etc. This would typically provide the most relevant data, even if those are $\dot{V}O_2$peak (or “mode-specific $\dot{V}O_2$max”) rather than true $\dot{V}O_2$max values.

d) Warm-Up

A warm-up prior to $\dot{V}O_2$max/peak testing can serve two important purposes. First and foremost, it prepares the body to better handle the rigors of the subsequent test. This effect can allow for a shorter subsequent test and, sometimes, even a higher final exercise intensity and measured $\dot{V}O_2$max/peak, at exhaustion. Inbar & Bar-Or (20) reported higher exhaustion $\dot{V}O_2$ among 7−9 year-old boy-non-athletes, following 15 min of intermittent warm-up at 60% $\dot{V}O_2$max, compared with no warm-up. Additionally, a warm-up can serve an excellent means of habituation to the test's conditions (e.g., incline, intensity, cadence) and to the testing equipment (see below). Most studies examining $\dot{V}O_2$max/peak in youth do not report warm-up procedures. Thus, by serving as an habituation and preparation, a proper warm-up reduces the chance of $\dot{V}O_2$max/peak underestimation.

There currently are no accepted warm-up protocols, but the following can serve as general warm-up guidelines: i) For warm-up specificity use the actual test's ergometer, or equivalent; ii) Increase intensity gradually, but do not approach exhaustion. Brief exposures to the maximal anticipated speed and incline, or power output can be introduced towards the end; iii) The warm-up should be ~5−8-minute-long; iv) Have the testee experience the applicable mouthpiece, mask, breathing valve, or headgear during the warm-up; v) The rest and setup interval from end of warm-up to start of test should preferably be 1−3 min.

e) **Test Protocol**

There is no one 'gold standard' protocol for $\dot{V}O_2$max/peak testing. Generally, the protocol ought to be progressive and start at a moderate intensity (see below). Given proper warm-up (see above), exercise intensity should be incremented in a manner that will induce exhaustion between 6 and 10 min (see below for exceptions). However, when there is no prior knowledge of the testee's general ability, and in many clinical settings, one may prefer to rely on a standard protocol. Many commonly-used, standard maximal aerobic-power protocols were developed for clinical use − mainly to examine ECG and blood pressure changes with exercise (e.g., Bruce or Balke protocols) − not for maximizing actual $\dot{V}O_2$max/peak values. Moreover, those protocols were developed for adults and their starting loads and step incrementation are usually inappropriate for children. Yet, they are still used today in clinical and research settings, also with children (17, 48). In the 1970's and 80's, several 'standardized' cycle-ergometry protocols were implemented in leading pediatric laboratories, which set the initial load and the progression according to body size. For example, by body mass (BM), height, or by estimated body surface area (BSA) (7, 8, 16, 21, 22)). Likewise, there are several modified treadmill protocols which have been used especially in pediatric clinical settings (see (43) for comprehensive details of cycle and treadmill protocols). Regardless of whether one follows a standardized protocol, or customizes one to the testee, the protocol ought to abide by the following general guidelines:

- <u>Starting load</u> should ideally be what elicits approx. 60% of the predicted $\dot{V}O_2$max, or ~50% of the expected final exercise intensity (following a proper warm-up). Unless the fitness level of the individual being tested is already known, determining the starting load may involve some guess work, which in turn might result in somewhat shorter or longer test durations.
- <u>Duration</u> should, ideally, be 6−10 min, assuming a proper warm-up was performed and allowing for the uncertainties surrounding starting load and load progression. Too short a test might be insufficient for $\dot{V}O_2$max/peak attainment, while an overly long one may induce fatigue and exhaustion before $\dot{V}O_2$max/peak has been attained. There are situations where it is preferred to have a somewhat longer test (e.g., when incorporating gas-exchange/ventilatory-threshold testing, which optimally

require more than 10-min). This can still be done without adversely affecting the $\dot{V}O_2$max/peak outcomes by adding the extra stages at the beginning, at lower intensities, rather than adding extra stages between 50 and 100% $\dot{V}O_2$max/peak. For example, a test starting at ~50% of the anticipated $\dot{V}O_2$max/peak and graded at 6% steps would last 9−10 min, but when started at 20% it would last 14−15 min using the same intensity incrementation. Importantly, the extra ~5 min, added by starting at such low intensity, eliminate the need for a preceding warm-up.

- Load incrementation could range between continuous ramping (e.g., 1 W every 2 s) to stepwise increases in exercise intensity (e.g., 1 km·hr^{-1}, 0.3 m·s^{-1}, or 20 W, every minute). In order to avoid a short test and based on our experience, we recommend that the load is incremented at a rate of ~6−12% of the anticipated final load. Incrementation rate is a direct function of the starting load and the expected test duration. Thus, for example, a cycle-ergometer test starting at 60W and reaching exhaustion at 120W would be 7-min long if incremented by 10W·min^{-1} (~8%·min^{-1}). If, however, a given incrementation rate (e.g., 7W or ~6%) proves, mid test, to be too small and leading to an overly long test, it can be upward-adjusted (e.g., to 10 or 15W / ~8−12%) so as to comply with the recommended upper limit of test duration.

- Controlling exercise variables. Eliciting one’s maximal performance partly depends on skill and, to a large extent, on a combination of force and velocity. On the treadmill, these are the treadmill incline and the running speed, respectively. On the cycle-ergometer these are determined by the applied resistance and the pedalling cadence. As there is an optimal force-velocity relationship for any physical activity, it is extremely important to function at or near that optimum to ensure that exhaustion occurs at the highest possible $\dot{V}O_2$ rather than prematurely encountered due to suboptimal mechanics. On the treadmill this would mean running at well under the individuals top speed or highest manageable incline. On the cycle-ergometer, due to the force–velocity interdependency, appropriate cadence plays a critical role (45). At a given exercise intensity, lower cadence means greater resistance, potentially resulting in premature exhaustion due to local muscular fatigue. Aside from trained cyclists, most individuals, adults and children alike, tend to pedal at sub-optimal cadence which, in turn, results in $\dot{V}O_2$max/peak underestimation. High cadence, on the other hand, is limited by skill and reduced efficiency which could also end a test prematurely. Based on our experience, a cadence of 80 rpm is typically well-tolerated by both children and adults. If comfortably doable, the testee may be advised to try and raise that cadence in the test's last few minutes as the effort approaches maximum. *Note*: raising the cadence is only possible on electromagnetically-braked cycle-ergometers (which maintain power output within a wide range of cadence), not on mechanically-braked ones.

- Measuring other variables in addition to VO_2. It is often desirable to measure additional variables in conjunction with VO_2. These may include heart rate and blood pressure, which may not require protocol modifications and could thus be regarded as fully compatible with $\dot{V}O_2$max/peak testing. However, when measured variables such as heart rate or lactate are meant to reflect the intensity of the stages in which they were taken (e.g., lactate response plot), stages of 3 min or longer would be required. This extends test duration 2–3-fold and may bring about pre-mature exhaustion, before $\dot{V}O_2$max/peak can be attained. While such protocols have been used successfully in some laboratories (e.g., (2)) they may not be ideal specifically for $\dot{V}O_2$max/peak testing.

f) Criteria for maximal effort

As indicated above, the classic criterion for the attainment of $\dot{V}O_2$max is an inability to increase $\dot{V}O_2$ despite an increment in work load, i.e., a $\dot{V}O_2$ plateau. There is no well-accepted definition of a plateau in pediatric testing, possibly because it is often absent (9, 28, 35, 41, 42). Criteria previously used in children are deviation of <2 ml·kg^{-1}·min^{-1} from the average value during the final 60 s of the test (41, 42), or >5% deviation from the projected $\dot{V}O_2$max in the last 60 s of the test (9). Since a plateau is often not observed in children, the challenge is not in its definition, but rather in the question of whether $\dot{V}O_2$max/peak has actually been attained. Adding to the $\dot{V}O_2$max determination challenge is the fact that in some individuals, mainly adults, $\dot{V}O_2$max may be reached prior to exhaustion, while in others, it may not be attained even at complete exhaustion (see Table 1). In pediatric studies, participants are 'simply' brought to volitional exhaustion, based on subjective perception by the participant or subjective evaluation by the researcher. The attainment of a true $\dot{V}O_2$max in this case may depend on the child's motivation and previous experience with the sensation of maximal exertion, as well as on the researcher's or clinician's encouragement and testing experience. Most children can be motivated to complete the exercise test to exhaustion, but this often requires an experienced clinician/researcher, who is familiar with their physical, as well as psychological or emotional responses. It has been argued that this reliance on experience, along with the common absence of the plateau criterion, may result in $\dot{V}O_2$max underestimation. Consequently, various secondary objective criteria, reflecting high physiological exertion, have been proposed and used, individually or in combination. These criteria include respiratory exchange ratio (RER) >1.0 or 1.1, HR >85% of predicted HRmax or >195 b/min, or blood lactate concentration >6 mM. It is unclear what the above 'acceptable' values are based upon, but it is evident that they too often result in substantial $\dot{V}O_2$max/peak underestimation (9, 38). Among adults, the use of such secondary criteria may result in up to 30–40% underestimation of true $\dot{V}O_2$max (30, 39). Similarly, Barker et al. (9) demonstrated that among 9–10 year-old recreationally-active children, the use of the 195 b/min HR, or 1.0 RER criteria, resulted in mean $\dot{V}O_2$max underestimations of 10 and 23%, respectively. It was also noted that, despite reaching $\dot{V}O_2$max in the incremental protocol (supramaximal-verified. See below), 30% of the participants did not display at least one of the ‘accepted’ secondary physiological criteria. Depending on the criterion, this would have resulted in rejection of valid $\dot{V}O_2$max values.

The use of the “Supramaximal Test” was first reported over 25 years ago by Rowland (41) and, more recently, by Barker et al. (9), to address this uncertainty surrounding true $\dot{V}O_2$max attainment. Administered 10–15 min following the $\dot{V}O_2$max/peak test, the ‘supramaximal’ test immediately applies 105% of the previous test’s final load (9), in children, and up to 110% in adults (38). It typically lasts 3–5 min. A resulting $\dot{V}O_2$ value equal to or lower than that achieved in the incremental $\dot{V}O_2$max/peak test validates the previous result, otherwise the new value is taken as $\dot{V}O_2$max, or the test is repeated. It is worthwhile noting, however, that a ‘supramaximal’ protocol cannot negate an improperly administered incremental test. It is also vulnerable to some of the ‘pitfalls’ of exercise testing (see Table 1), and may still provide an underestimate of $\dot{V}O_2$max.

In view of the difficulty of determining an objective endpoint to a progressive maximal test, particularly in children and especially in the clinical setting, some have suggested that submaximal indicators (e.g., ventilatory threshold, oxygen uptake efficiency slope) may be just as useful, if not more so, in diagnosing, managing and treating various diseases (12, 19). However, while submaximal protocols may be appropriate as diagnostic tools in many cases, submaximal indicators cannot be substituted for much of the information provided by maximal values. Moreover, the determination and interpretation criteria for these indicators are often just as elusive (52).

g) Data Reduction

The determination of $\dot{V}O_2$max/peak cannot rely on a fleeting spike in $\dot{V}O_2$, as could be encountered in BxB measurement, or even in short-term (e.g., 10–15 s) mixing-chamber averaging. Thus, a longer averaging period is needed to faithfully reflect actual metabolism. Stemming from the Douglas Bag era, where 1-min bags were typically collected, $\dot{V}O_2$max used to be defined as the highest $\dot{V}O_2$ at any given whole minute of the test. This may be a problem since a $\dot{V}O_2$max state could be more transient and shorter-lasting than 60 s. Moreover, that maximal state may straddle two successive collection periods. The advent of continuous, automated measurement systems has allowed for any duration of averaging periods (e.g., 10, 15, 20 s). Thus, averaging the highest consecutive periods over 30–45 s (e.g., 3x10 s, 2x15, or 2x20 s) can, at the same time, avoid $\dot{V}O_2$max underestimation and better pinpoint the time of its occurrence. Researchers often report the intervals at which $\dot{V}O_2$ is *recorded* by the metabolic system. They ought to also report the duration over which $\dot{V}O_2$max is calculated.

Data Interpretation

As summarized in Table 1, there are numerous 'pitfalls' that could cause over-, or more often, underestimation of $\dot{V}O_2$max. These pitfalls may be related to equipment set-up and calibration, choice of initial load or incrementation of exercise intensity, choice of ergometer, and data averaging. One of the main issues in conducting a maximal test is determining the endpoint of the test, specifically in children. That is, when trying to interpret results, it is imperative to determine whether $\dot{V}O_2$max (or $\dot{V}O_2$peak) was actually achieved. As discussed above and highlighted by Barker et al. (9), relying on secondary criteria can easily result in underestimation of true $\dot{V}O_2$max. Such underestimation can lead to underestimating an athlete's current fitness, or misdiagnosis of a patient's disease severity. In both cases, this underestimation could result in inappropriate evaluation and consequent actions. In research, when trying to characterize a sample, for example, underestimation may be variable. That is, assuming there is one true maximal value, but many potential 'underestimated' ones, $\dot{V}O_2$max/peak would be underestimated inconsistently, thus resulting in larger than true variability. If the means of two samples are compared (e.g., healthy vs. children with asthma), such large variability may lead to a false negative or type II error (undetectable group difference when it actually exists). If the efficacy of an intervention is examined (e.g., therapeutic intervention, athletic training), an underestimation of initial values (e.g., due to lack of familiarization, high initial loading, which are less likely to occur in the follow-up test) will lead to potential overestimation of the efficacy of the intervention or type I error (detecting improvement where none actually occurred). $\dot{V}O_2$max/peak overestimation, although less common, can occur due to the use of a breath-by-breath metabolic system, wrong calibration, or <30-s averaging periods (see above and in Table 1). This, of course, could also lead to inappropriate conclusions and recommendations, but in an opposite direction to that of underestimation errors.

An attempt is often made to compare an individual's $\dot{V}O_2$max/peak results with available normative values. This comparison may be specifically relevant in clinical settings, when diagnosing disease status, or in athletic settings such as in talent identification or in constructing fitness profiles. Various published values have been used as age- and sex-specific norms (7, 8, 10, 13, 29). However, the comparison is often hampered by the fact that the ‘norms’ do not take into account maturity status and related body-size differences, which would directly affect oxygen uptake. Moreover, ‘norms’ of healthy children generally represent a sample of volunteers rather than a random, unbiased population sample. Absolute $\dot{V}O_2$max/peak increases with age during the growing years, as was classically demonstrated by Astrand’s cross-sectional studies in 1952 (4), and more recently by others, using longitudinal designs (15, 25, 31). However, as discussed above, $\dot{V}O_2$max/peak values depend on many controlled factors (e.g., protocol/ procedure, exercise mode) and uncontrollable ones (e.g., climatic conditions). These are often ignored when comparing individual values to published ‘norms’. It makes little sense to compare treadmill $\dot{V}O_2$max results to ‘norms’ developed with cycle-ergometry. Even on the same ergometer, protocols and settings may differ (see above). Thus, all such comparisons should be made with caution.

The lack of well-accepted protocols for exercise testing and criteria for maximal aerobic power in children (see above), as well as the lack of well-accepted reporting practices (see below), render this seemingly simple task of comparison to norms quite daunting. Indeed, a concerted call for the harmonization of exercise testing and reporting, especially for clinical purposes, has recently been issued in order to make maximal and sub-maximal exercise-response comparisons more reliable and meaningful (3, 37). When possible, an individual’s $\dot{V}O_2$max/peak values could be compared with a database established at the testing laboratory, based on its particular equipment and procedures. Ideally, $\dot{V}O_2$max/peak values would also be compared with the individual’s own previous results, if they exist (for additional normalization and scaling considerations, see below under Data Reporting).

In research, $\dot{V}O_2$max/peak mean values of the examined sample are often compared with studies of a similar population. This comparison is vulnerable to similar pitfalls characteristic of comparing an individual's values to published 'norms', as discussed above. Numerous reviews of aerobic power in children grouped together studies which utilized treadmills, cycle-ergometers, and field tests to determine maximal aerobic power (e.g.,(6, 36). While pre–post training-intervention change in $\dot{V}O_2$max/peak may be less vulnerable to modality differences, grouping all modalities to deduce 'normative' or representative $\dot{V}O_2$max/peak values may be misleading. Thus, such comparisons should be made with caution.

Data Reporting

Oxygen uptake is a measure of volume over a given time (liters per minute) and is thus greatly dependent on body size. As children grow, absolute $\dot{V}O_2$max/peak (L/min) increases in accordance with increasing muscle and cardiac mass, as well as increasing blood and lung volumes. Thus, the absolute $\dot{V}O_2$max/peak of a 10-year-old child may be <50% that of an adult. When comparing children with adults, normal-weight with obese individuals, or when studying aerobic-power development with growth and maturation, a major question is whether or how to control (scale) for body size. Various approaches to scaling $\dot{V}O_2$max/peak have been adopted, depending on the pertinent question.

The most ubiquitous approach to $\dot{V}O_2$max/peak scaling involves ratio standards, whereby the absolute oxygen uptake is divided by some measure of body size. Although in mammals and homeotherms, in general, metabolic rate is most closely related to body surface area (BSA) (e.g., (26)), body mass or a derivative thereof is most commonly used. Since metabolic rate ($\dot{V}O_2$) and $\dot{V}O_2$max are known not to be linearly related to body mass (50, 51), and BSA best scales to 2/3 power of volume and mass, the body mass derivative often used to correct for size effects is body mass$^{0.67}$ (11, 44, 51). Other models propose that metabolic rate be scaled to body mass$^{0.75}$, based on dimensional analysis or nutrient transport models (11, 51). However, when comparing functional capacity in weight-bearing activities such as running, dividing by body mass is usually most informative. A difficulty associated with that is that body mass is largely determined by muscle mass and fat mass, and only the former directly contributes to oxygen uptake, while fat mass constitutes functional 'dead weight'. Thus, in some cases (e.g., when assessing overweight children), 'net' aerobic power is better expressed relative to fat-free mass. The latter, however, is a rough measure of muscle mass, since adiposity is typically assessed by indirect, approximate methods.

Empirically, none of the proposed exponents fully account for all observed $\dot{V}O_2$–body-mass relationships and, consequently, various sample-specific body mass exponents have been proposed (23, 32). Due to the large changes it undergoes during maturation, properly accounting for body mass is of pronounced importance in youth. This is especially important when interpreting longitudinal studies, particularly those which span the pubertal years during which there are large body-mass changes. It is also important in the clinical settings, where an individual's aerobic power value is used as an indicator of functional capacity, in conjunction with age-related 'norms' (see discussion on norms above).

A recent commentary by Welsman & Armstrong provide the background and methodical justification for the use of allometric scaling in children (49). The authors provide the scientific and statistical rationale and demonstrate that with the use of appropriate body mass exponents, the effect of body mass can effectively be removed. However, the derived body mass exponents are sample-specific, depending on sample characteristics and size, which may greatly vary from one sample to another. Indeed, in 20 different samples, the reported body mass exponents ranged from 0.37 to 0.94. Thus, there is no ‘ultimate’ universal exponent (11) and a practical recommendation is for users to carefully evaluate the suitability of any given body mass exponent to their particular sample and study objectives. Proper allometric scaling is particularly consequential when comparing children of widely different body sizes (which is most pronounced during pubertal maturation), or when comparing children and adults.

Ratio standards usually do not fully account for body size. Nevertheless, they are easily measured and practical. Therefore, from a practical perspective, they simplify the comparison of individuals with published ‘norms’ or the comparison between studies. Moreover, in specific cases where individuals of extreme body composition are assessed (very lean or obese), the use of fat-free mass, lean mass or muscle mass as the denominator may facilitate the comparison with normal-weight individuals.

Conclusion

While the general objectives and guidelines for $\dot{V}O_2$max/peak measurement and determination are not different with children than with adults, some important distinctions define the special considerations involved in pediatric testing. Foremost of those is the typical absence of $\dot{V}O_2$ plateau at test’s end that is not caused by deficient effort but leaves the tester with no unequivocal criterion for $\dot{V}O_2$max/peak attainment. The substitute criteria that have been extensively employed, even when used in combination, do not guarantee a correct determination of $\dot{V}O_2$max, either. These place a greater onus on the tester’s skill and experience and may compel the use of ‘supramaximal’ verification. The proper choice of pediatric-appropriate ergometers and metabolic measurement systems could also have great potential consequences in pediatric testing. Finally, the interpretation of the obtained data may require different accounting for body mass than with adults, particularly when comparing children of markedly varied sizes, or children vs. adults.

References

1. Armstrong N and van Mechelen W *Oxford textbook of children's sport and exercise medicine* 3rd ed. Oxford: Oxford University Press, 2017.
2. Armstrong N, Williams J, Balding J, Gentle P, and Kirby B. The peak oxygen uptake of British children with reference to age, sex and sexual maturity. *Eur J Appl Physiol Occup Physiol* 1991;62(5):369-375.
3. Ashish N, Bamman MM, Cerny FJ, Cooper DM, D'Hemecourt P, Eisenmann JC, Ericson D, Fahey J, Falk B, Gabriel D, Kahn MG, Kemper HC, Leu SY, Liem RI, McMurray R, Nixon PA, Olin JT, Pianosi PT, Purucker M, Radom-Aizik S, and Taylor A. The clinical translation gap in child health exercise research: a call for disruptive innovation. *Clinical and translational science* 2015;8(1):67-76.
4. Astrand PO *Experimental Studies of Physical Work Capacity in Relation to Sex and Age.* Copenhagen: Mundsgaard, 1952.
5. Azevedo P, Bhammar DM, Babb TG, Bowen TS, Witte KK, Rossiter HB, Brugniaux JV, Perry BD, Dantas de Lucas R, Turnes T, Sabino-Carvalho JL, Lopes TR, Zacca R, Fernandes RJ, McKie GL, Hazell TJ, Helal L, da Silveira AD, McNulty CR, Roberg RA, Nightingale TE, Alrashidi AA, Mashkovskiy E, Krassioukov A, Clos P, Laroche D, Pageaux B, Poole DC, Jones AM, Schaun GZ, de Souza DS, de Oliveira Barreto Lopes T, Vagula M, Zuo L, and Zhao T. Commentaries on Viewpoint: Vo2peak is an acceptable estimate of cardiorespiratory fitness but not Vo2max. *J Appl Physiol (1985)* 2018;125(1):233-240.
6. Baquet G, van Praagh E, and Berthoin S. Endurance training and aerobic fitness in young people. *Sports Med.* 2003;33(15):1127-1143.
7. Bar-Or O *Pediatric Sports Medicine for the Practitioner*. New York: Springer-Verlag, 1983.
8. Bar-Or O and Rowland T *Pediatric Exercise Medicine*. Champaign, IL: Human Kinetics, 2004.
9. Barker AR, Williams CA, Jones AM, and 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.
10. Blanchard J, Blais S, Chetaille P, Bisson M, Counil FP, Huard-Girard T, Berbari J, Boulay P, and Dallaire F. New Reference Values for Cardiopulmonary Exercise Testing in Children. *Med. Sci. Sports Exerc.* 2018;50(6):1125-1133.
11. Capellini I, Venditti C, and Barton RA. Phylogeny and metabolic scaling in mammals. *Ecology* 2010;91(9):2783-2793.
12. Cooper DM, Leu SY, Galassetti P, and Radom-Aizik S. Dynamic interactions of gas exchange, body mass, and progressive exercise in children. *Med. Sci. Sports Exerc.* 2014;46(5):877-886.

13. Cooper DM, Weiler-Ravell D, Whipp BJ, and Wasserman K. Growth-related changes in oxygen uptake and heart rate during progressive exercise in children. *Pediatr. Res.* 1984;18(9):845-851.
14. Dotan R. Children's Aerobic Trainability and Related Questions. *Res. Q. Exerc. Sport* 2017;88(4):377-383.
15. Geithner CA, Thomis MA, Vanden Eynde B, Maes HH, Loos RJ, Peeters M, Claessens AL, Vlietinck R, Malina RM, and Beunen GP. Growth in peak aerobic power during adolescence. *Med. Sci. Sports Exerc.* 2004;36(9):1616-1624.
16. Godfrey S, Davies CT, Wozniak E, and Barnes CA. Cardio-respiratory response to exercise in normal children. *Clin. Sci.* 1971;40(5):419-431.
17. Gomes EL, Carvalho CR, Peixoto-Souza FS, Teixeira-Carvalho EF, Mendonca JF, Stirbulov R, Sampaio LM, and Costa D. Active Video Game Exercise Training Improves the Clinical Control of Asthma in Children: Randomized Controlled Trial. *PloS one* 2015;10(8):e0135433.
18. Green S and Askew C. Vo2peak is an acceptable estimate of cardiorespiratory fitness but not Vo2max. *J Appl Physiol (1985)* 2018;125(1):229-232.
19. Hebestreit H, Staschen B, and Hebestreit A. Ventilatory threshold: a useful method to determine aerobic fitness in children? *Med. Sci. Sports Exerc.* 2000;32(11):1964-1969.
20. Inbar O and Bar-Or O. The effects of intermittent warm-up on 7-9 year-old boys. *Eur J Appl Physiol Occup Physiol* 1975;34(2):81-89.
21. James FW, Blomqvist CG, Freed MD, Miller WW, Moller JH, Nugent EW, Riopel DA, Strong WB, and Wessel HU. Standards for exercise testing in the pediatric age group. American Heart Association Council on Cardiovascular Disease in the Young. Ad hoc committee on exercise testing. *Circulation* 1982;66(6):1377A-1397A.
22. James FW, Kaplan S, Glueck CJ, Tsay JY, Knight MJ, and Sarwar CJ. Responses of normal children and young adults to controlled bicycle exercise. *Circulation* 1980;61(5):902-912.
23. Janz KF, Burns TL, Witt JD, and Mahoney LT. Longitudinal analysis of scaling VO2 for differences in body size during puberty: the Muscatine Study. *Med. Sci. Sports Exerc.* 1998;30(9):1436-1444.
24. Kamon E and Pandolf KB. Maximal aerobic power during laddermill climbing, uphill running, and cycling. *J. Appl. Physiol.* 1972;32(4):467-473.
25. Kemper HC, van Mechelen W, Post GB, Snel J, Twisk JW, van Lenthe FJ, and Welten DC. The Amsterdam Growth and Health Longitudinal Study. The past (1976-1996) and future (1997-?). *Int. J. Sports Med.* 1997;18 Suppl 3(S140-150.
26. Kleiber M. Body size and metabolic rate. *Physiol. Rev.* 1947;27(4):511-541.
27. Krahenbuhl GS, Skinner JS, and Kohrt WM. Developmental aspects of maximal aerobic power in children. *Exerc. Sport Sci. Rev.* 1985;13(503-538.
28. LeMura LM, von Duvillard SP, Cohen SL, Root CJ, Chelland SA, Andreacci J, Hoover J, and Weatherford J. Treadmill and cycle ergometry testing in 5- to 6-year-old children. *Eur. J. Appl. Physiol.* 2001;85(5):472-478.
29. McManus AM and Armstrong N. Maximal oxygen uptake, in *Cardiopulmonary Exercise Testing in Children and Adolescents*, Rowland TW, (Ed. Champaign, IL: Human Kinetics, 2018, pp. 79-94.

30. Midgley AW, Carroll S, Marchant D, McNaughton LR, and Siegler J. Evaluation of true maximal oxygen uptake based on a novel set of standardized criteria. *Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme* 2009;34(2):115-123.

31. Mirwald RL, Bailey DA, Cameron N, and Rasmussen RL. Longitudinal comparison of aerobic power in active and inactive boys aged 7.0 to 17.0 years. *Ann. Hum. Biol.* 1981;8(5):405-414.

32. Nevill AM and Holder RL. Scaling, normalizing, and per ratio standards: an allometric modeling approach. *J Appl Physiol (1985)* 1995;79(3):1027-1031.

33. Nixon PA, Orenstein DM, Kelsey SF, and Doershuk CF. The prognostic value of exercise testing in patients with cystic fibrosis. *N. Engl. J. Med.* 1992;327(25):1785-1788.

34. Paridon SM, Alpert BS, Boas SR, Cabrera ME, Caldarera LL, Daniels SR, Kimball TR, Knilans TK, Nixon PA, Rhodes J, Yetman AT, American Heart Association Council on Cardiovascular Disease in the Young CoAH, and Obesity in Y. Clinical stress testing in the pediatric age group: a statement from the American Heart Association Council on Cardiovascular Disease in the Young, Committee on Atherosclerosis, Hypertension, and Obesity in Youth. *Circulation* 2006;113(15):1905-1920.

35. Paterson DH, Cunningham DA, and Donner A. The effect of different treadmill speeds on the variability of VO2 max in children. *Eur J Appl Physiol Occup Physiol* 1981;47(2):113-122.

36. Pfeiffer KA, Lobelo F, Ward D, and Pate R. Endurance trainability of children and youth, in *The Young Athlete*, Heberstreit H and Bar-Or O, Eds.). Oxford: Blackwell Publishing Ltd., 2008, pp. 84-95.

37. Pianosi PT, Liem RI, McMurray RG, Cerny FJ, Falk B, and Kemper HC. Pediatric Exercise Testing: Value and Implications of Peak Oxygen Uptake. *Children (Basel)* 2017;4(1):

38. Poole DC and Jones AM. Measurement of the maximum oxygen uptake Vo2max: Vo2peak is no longer acceptable. *J Appl Physiol (1985)* 2017;122(4):997-1002.

39. Poole DC, Wilkerson DP, and Jones AM. Validity of criteria for establishing maximal O2 uptake during ramp exercise tests. *Eur. J. Appl. Physiol.* 2008;102(4):403-410.

40. Rowland TW *Cardiopulmonary Exercise Testing in Children and Adolescents*. Champaign, IL: Human Kinetics, 2018.

41. Rowland TW. Does peak VO2 reflect VO2max in children?: evidence from supramaximal testing. *Med. Sci. Sports Exerc.* 1993;25(6):689-693.

42. Rowland TW and Cunningham LN. Oxygen uptake plateau during maximal treadmill exercise in children. *Chest* 1992;101(2):485-489.

43. Sabath RJ, White DA, and Tenson KM. Exercise testing protocols, in *Cardiopulmonary Exercise Testing in Children and Adolescents*, Rowland TW, (Ed. Champaign, IL: Human Kinetics, 2018, pp. 23-40.

44. Schmidt-Nielsen K *Scaling: why is animal size so important*. Cambridge, UK: Cambridge University Press, 1984.

45. Swain DP and Wilcox JP. Effect of cadence on the economy of uphill cycling. *Med. Sci. Sports Exerc.* 1992;24(10):1123-1127.

46. Turley KR, Rogers DM, Harper KM, Kujawa KI, and Wilmore JH. Maximal treadmill versus cycle ergometry testing in children: differences, reliability and variability of responses. *Pediatric Exercise Science* 1995;7(49-60.

47. Turley KR and Wilmore JH. Cardiovascular responses to treadmill and cycle ergometer exercise in children and adults. *J. Appl. Physiol.* 1997;83(3):948-957.
48. van der Cammen-van Zijp MH, Ijsselstijn H, Takken T, Willemsen SP, Tibboel D, Stam HJ, and van den Berg-Emons RJ. Exercise testing of pre-school children using the Bruce treadmill protocol: new reference values. *Eur. J. Appl. Physiol.* 2010;108(2):393-399.
49. Welsman J and Armstrong N. Interpreting aerobic fitness in youth: the fallacy of ratio scaling. *Pediatric Exercise Science* 2019 31(2):xx-xx.
50. Welsman JR, Armstrong N, Nevill AM, Winter EM, and Kirby BJ. Scaling peak VO2 for differences in body size. *Med. Sci. Sports Exerc.* 1996;28(2):259-265.
51. West GB, Brown JH, and Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science* 1997;276(5309):122-126.
52. Yeh MP, Gardner RM, Adams TD, Yanowitz FG, and Crapo RO. "Anaerobic threshold": problems of determination and validation. *J Appl Physiol Respir Environ Exerc Physiol* 1983;55(4):1178-1186.

Table 1: Potential pitfalls in VO2 max measurement and their implications

Element	Pitfall	Consequence	V.O2 *V.O2* max Determination	Comments
Metabolic Measurement System	Breath-by-Breath System	Fast & large gas-concentration changes	Under- or over-estimation	Excellent for tracking fast-changing V.O2 *V.O2* kinetics, but compromises accuracy of V.O2 *V.O2* max determination
	Inappropriate Calibration	Unreliable gas concentrations & volumes	Under- or over-estimation	Follow instructions; Use certified calibration gases
	Large mixing-chamber & hose volume	Large dead space; lagging & buffered response	Underestimation	Pediatric setups are rare, but can be added
	Mask – inappropriate fit	Expiratory-gas & air leak	Underestimation	Modern masks are improved and generally provide good

				fit. Child-size mask is necessary.
Warm-Up	Absent or too short	Premature test termination	Underestimation	Aside from warm-up effect, habituation to test conditions; Gradual progression; 5−8 min
	Too intense or too long	Premature test termination	Underestimation	
Test Protocol	High starting load	Premature exhaustion	Underestimation	Incrementation can be adjusted during test to compensate for unsuitable starting load and initial increment size
	Too large increments	Premature exhaustion	Underestimation	
	Test too short	Lagging V.O$_2$ *V.02* response	Underestimation	
	Test too long	Exhaustion before V.O$_2$ *V.02* max	Underestimation	
Cycle-Ergometer	Inappropriate size or set-up	Premature exhaustion	Underestimation	Saddle too high; Cranks too long; Handlebar too far
	Smaller employed muscle-mass (relative to treadmill)	Lower O$_2$ demand	Underestimation	More so with arm-ergometry
	Low pedalling cadence	Premature exhaustion	Underestimation	Recommended: 80 RPM or higher
Treadmill	Inappropriate habituation	Premature termination	Underestimation	Extend habituation prior to, or as part of the warm-up
	Incline too steep	Premature exhaustion	Underestimation	Local muscular fatigue
	Speed too high	Premature exhaustion	Underestimation	Skill limitation & insecurity
Data Reduction	Short-period- or no averaging	Large peaks & fluctuations	Overestimation	Affected by non-physiologic fluctuations

	Too-long averaging period	Buffering effect; flattened response	Underestimation	Recommended: 30−45 s total. Average of 2−3 consecutive segments of 10−20 s each