REVIEW | Cores of Reproducibility in Physiology

Measurement of the maximum oxygen uptake \( \dot{V}O_{2\text{max}} \); \( \dot{V}O_{2\text{peak}} \) is no longer acceptable

David C. Poole\(^1\) and Andrew M. Jones\(^2\)

\(^1\)Departments of Kinesiology, Anatomy, and Physiology, Kansas State University, Manhattan, Kansas; and \(^2\)Sport and Health Sciences, St. Lake’s Campus, University of Exeter, Exeter, United Kingdom

Submitted 2 December 2016; accepted in final form 25 January 2017

Poole DC, Jones AM. Measurement of the maximum oxygen uptake \( \dot{V}O_{2\text{max}} \); \( \dot{V}O_{2\text{peak}} \) is no longer acceptable. J Appl Physiol 122: 997–1002, 2017. First published February 2, 2017; doi:10.1152/japplphysiol.01063.2016.—The maximum rate of \( O_2 \) uptake (i.e., \( \dot{V}O_{2\text{max}} \)), as measured during large muscle mass exercise such as cycling or running, is widely considered to be the gold standard measurement of integrated cardiopulmonary-muscle oxidative function. The development of rapid-response gas analyzers, enabling measurement of breath-by-breath pulmonary gas exchange, has facilitated replacement of the discontinuous progressive maximal exercise test (that produced an unambiguous \( \dot{V}O_{2\text{work rate plateau definite for \( \dot{V}O_{2\text{max}} \}) \) with the rapidly incremented or ramp testing protocol. Although this is more suitable for clinical and experimental investigations and enables measurement of the gas exchange threshold, exercise efficiency, and \( \dot{V}O_2 \) kinetics, a \( \dot{V}O_{2\text{work rate plateau is not an obligatory outcome. This shortcoming has led to investigators resorting to so-called secondary criteria such as respiratory exchange ratio, maximal heart rate, and/or maximal blood lactate concentration, the acceptable values of which may be selected arbitrarily and result in grossly inaccurate \( \dot{V}O_{2\text{max}} \) estimation. Whereas this may not be an overriding concern in young, healthy subjects with experience of performing exercise to volitional exhaustion, exercise test naïve subjects, patient populations, and less motivated subjects may stop exercising before their \( \dot{V}O_{2\text{max}} \) is reached. When \( \dot{V}O_{2\text{max}} \) is a or the criterion outcome of the investigation, this represents a major experimental design issue. This CORP presents the rationale for incorporation of a second, constant work rate test performed at \( \sim 110\% \) of the work rate achieved on the initial ramp test to resolve the classic \( \dot{V}O_2, \text{work rate plateau that is the unambiguous validation of } \dot{V}O_{2\text{max}} \). The broad utility of this procedure has been established for children, adults of varying fitness, obese individuals, and patient populations.

cardiopulmonary exercise testing; incremental exercise; constant-load exercise; oxygen transport; cardiorespiratory disease

A variety of exhausting exercise tests has been used to predict all-cause mortality in healthy and patient populations (15; see Ref. 53 for review and critique). However, it is only the maximum oxygen uptake (\( \dot{V}O_{2\text{max}} \)) achieved during severe-intensity large muscle mass exercise such as running, cycling, or swimming that actually measures the upper ceiling of the \( O_2 \) transport/utilization system. Thus \( \dot{V}O_{2\text{max}} \) assesses the integrated functioning of the pulmonary, cardiovascular, and muscle systems to uptake (diffusive \( O_2 \) transport at the lung and muscle microvasculature), transport (conductive \( O_2 \) transport), and utilize \( O_2 \) predominantly in the contracting muscle mitochondria. The foundational premise of \( \dot{V}O_{2\text{max}}, \) namely that there exists a speed of locomotion or rate of work above which \( \dot{V}O_2 \) fails to increase further, dates back at least to 1923 and the ideas of Hill and Lupton (21).

A PubMed search for \( \dot{V}O_{2\text{max}} \) yields in excess of 9,000 citations and underscores the foundational importance of this concept for understanding physiological function and exercise performance in health. Assessment of \( \dot{V}O_{2\text{max}} \) also has substantial clinical utility for measuring and understanding dysfunction in aging and a host of pathological conditions that impact pulmonary, cardiovascular, and muscle systems from chronic heart failure and diabetes to HIV-AIDS. Moreover, the power of \( \dot{V}O_{2\text{max}} \) to noninvasively determine the efficacy of exercise training programs and other ergogenic strategies in health as well as therapeutic interventions in disease conditions is tremendous.

Before the advent of rapidly responding gas analyzers and breath-by-breath pulmonary gas exchange measurements, \( \dot{V}O_{2\text{max}} \) was measured most commonly using a discontinuous series of progressively higher constant-speed or constant work
classical criterion of a plateau in $\dot{V}O_2$ definitive for $\dot{V}O_2_{max}$ was also of several minutes, between adjacent exercise steps, the facie evidence of $\dot{V}O_2_{max}$), remain linear (same slope), or, more yielded the classical $\dot{V}O_2$-work rate plateau emblematic of $\dot{V}O_2_{max}$ in the vast required time. This protocol, which may take up to an hour to conduct, to 5-min rest periods, and continued until the subject could no longer sustain a sequence of discrete work bouts of 4- to 5-min duration, interspersed with 4- the speed or work rate progressively, ideally over the course of usound technologies, the maximal incremental or ramp test- exchange threshold detection rather than achievement of $\dot{V}O_2_{max}$ per se, for 8 –12 min although this recommendation was based on resolution of gas exchange threshold detection rather than achievement of $\dot{V}O_2_{max}$ per se, for which durations between 5 and 26 min, depending in part on whether cycling or running is used, are effective (30). Subject A portrays the $\dot{V}O_2$-work rate plateau definitive for $\dot{V}O_2_{max}$. However, responses seen in subjects B and C are other outcomes of this testing protocol and may demand a subsequent validation test to provide assurance that $\dot{V}O_2_{max}$ was achieved.

rate steps, each being continued for several minutes until a quasi-steady state (if achievable) or exhaustion intervened (Fig. 1, top). If researchers allow rest and recovery periods, also of several minutes, between adjacent exercise steps, the classical criterion of a plateau in $\dot{V}O_2$ definitive for $\dot{V}O_2_{max}$ was achieved in most instances (4, 18, 48). However, in part because these procedures were laborious, time consuming, and, as such, ill-suited to clinical assessments, there was interest in developing shorter continuous-type tests (28). In the 1970s, concomitant with the advent of rapidly responding $O_2$ and CO$_2$ analyzers coupled with instantaneous respiratory gas flow measurements using pneumotachographs or later turbines and ultrasound technologies, the maximal incremental or ramp testing protocol became popular. This test constitutes increasing the speed or work rate progressively, ideally over the course of 8–12 min, to the limit of the subject’s tolerance (or willingness to continue) (8). Unless the test is protracted beyond this duration, which allows development of the $\dot{V}O_2$ slow component expressed as an upward curvature of the $\dot{V}O_2$-work rate relationship, $\dot{V}O_2$ typically increases as a close-to-linear function of work rate (cycle) or treadmill speed/incline until close proximity to exhaustion, where three response profiles are possible (Fig. 1, bottom). Specifically, the trajectory of the $\dot{V}O_2$ response near maximum may lessen or plateau (prima facie evidence of $\dot{V}O_2_{max}$), remain linear (same slope), or, more rarely, accelerate (upward curvature; see Ref. 55). Economy of time renders this test more suitable for clinical evaluation and subject assessment than the previous discontinuous test but often (i.e., ð50% in healthy subjects and likely more in patient populations) (13, 14, 25) fails to produce a discernible $\dot{V}O_2$ plateau or leveling off even when the actual $\dot{V}O_2_{max}$ is, in fact, achieved (13, 41, 56). This situation is accentuated in children (12), the elderly (46), and extremely unfit subjects (19, 25), as well as patients suffering from cardiovascular and/or pulmonary disease (50).

Despite this substantial shortcoming, the so-called incremental or ramp test coupled with breath-by-breath ventilation and gas exchange measurements has become one of the most powerful paradigms for experimental and clinical cardiopulmonary assessment. Used correctly, the incremental/ramp test can appraise four sentinel parameters of aerobic function: the gas exchange threshold, exercise efficiency (i.e., $V\dot{O}_2$-work rate slope in ml·min$^{-1}$·W$^{-1}$ for cycling), $V\dot{O}_2$ kinetics (i.e., time constant, $\tau$), and potentially $V\dot{O}_2_{max}$ (16, 35, 54). However, when used incorrectly, for instance with patient populations or naïve subjects who are inexperienced or unwilling to push themselves to exhaustion, this testing paradigm can result in substantial underestimation of $V\dot{O}_2_{max}$, which would need to be acknowledged and carefully considered during data interpretation. Whereas the validation test protocol (see A Viable Strategy for $V\dot{O}_2_{max}$ Validation below) may conceivably yield a plateau at a submaximal $V\dot{O}_2$, this would have to occur coincidentally and, although not studied specifically, would be expected to be a relatively rare occurrence.

**The $V\dot{O}_2peak$ vs. $V\dot{O}_2_{max}$ Expedient**

For young, healthy subjects that are used to driving themselves to exhaustion while cycling or running, the incremental/ ramp cardiopulmonary exercise test yields a highly reproducible $V\dot{O}_2_{max}$ irrespective of exercise test protocol, work rate forcing function, or pacing strategy (10). However, this may not be assured with exercise-naïve, unmotivated, and/or clinical populations. A serious concern here is that, when a training or therapeutic paradigm is being tested, the efficacy of the intervention may be overestimated in repeated testing as the subject gains experience and possibly enhanced confidence. Thus, if an accurate $V\dot{O}_2_{max}$ was not measured initially, whether or not the limit of the integrated $O_2$ transport/utilization system has actually been increased cannot be determined. In an attempt to mitigate this substantial source of experimental error, investigators have resorted to the term $V\dot{O}_2peak$ (~5,200 PubMed references to date) as simply the highest $V\dot{O}_2$ reached on a given test. Unfortunately, this procedure cannot discriminate among subjects who cease exercise because of lack of motivation, perceived discomfort, or a plethora of other reasons, none of which are related necessarily to their maximal rate of $O_2$ transport/utilization (38).

**So-Called Secondary Criteria Used to Validate $V\dot{O}_2_{max}$**

In those instances where a $V\dot{O}_2$ plateau is not attained as definitive evidence of $V\dot{O}_2_{max}$, investigators commonly elect to substantiate that $V\dot{O}_2_{max}$ was actually achieved by utilization of so-called secondary criteria assumed to validate $V\dot{O}_2_{max}$. These criteria most commonly include one or more of the following: heart rate (HR) $\leq$10 beats/min or $\leq$5% of the age-predicted (220-age) maximum, blood lactate concentration ≥8 mM, or
respiratory exchange ratio (RER) > 1.00, 1.10, or 1.15 (31, 38). The obvious problem with this “one size fits all” approach is that there is a broad range of maximal values for each of these variables in healthy and especially patient populations. Thus perusal of the literature reveals that HRmax has a 95% confidence interval of ±35 beats/min (5, 11, 26), maximal RER varies from 1.0 to 1.44 (46, 50), and maximal blood lactate concentration varies from 4 to 17 mM (5, 16, 50). Moreover, in the same subjects, changing the slope of the ramp (i.e., speed of the work rate increment) can alter these criteria in opposite directions. Specifically, faster ramps yield higher RERs but lower HRs at the same V\(_{\text{O2}}\)max (7, 8, 50, 52, 54). Also, equally pernicious with regard to so-called V\(_{\text{O2}}\)max criteria is that distinct populations may have very different HRmax values. For instance, extremely sedentary adults and also some populations of children may have lower HRmax values than predicted (11, 19). For North American children/young adults aged 8–18 yr, HRmax averages 187 beats/min (11) compared with 205 beats/min in Scandinavian children (5). An additional consideration here is that patient populations taking β-blockers will experience bradycardia during exercise and thus evince very low HRmax values.

Utilization of these criteria can allow for a 30–40% underestimation of the true V\(_{\text{O2max}}\) and/or an errant rejection of tests in which subjects had actually achieved their V\(_{\text{O2max}}\) (31, 38). At its least destructive, employment of secondary criteria increases variability and weakens experimental design, thereby increasing the likelihood of finding a false negative (type II error, incorrectly retaining a false null hypothesis); at its worst, it may create a false positive (type I error, incorrectly rejecting a true null hypothesis).

A Viable Strategy for V\(_{\text{O2max}}\) Validation

In contrast to the notion that a discrete work rate yields a predictable V\(_{\text{O2}}\), recognition of the existence of the V\(_{\text{O2}}\) slow component (17, 35), especially during constant work rate exercise in the severe-intensity domain (i.e., above critical power, CP) (39), means that, provided the exercise is of sufficient duration, V\(_{\text{O2max}}\) will be achieved for any supra-CP work rate (Fig. 2) (35–37, 42). However, there is a smaller range of constant work rates, but also see above the maximum obtained on the ramp test, which are sustainable for several minutes, each of which will yield V\(_{\text{O2max}}\) before exhaustion (Fig. 3). Accordingly, and consistent with the earliest recognition that “an obligatory plateau of V\(_{\text{O2}}\) provides unequivocal identification of V\(_{\text{O2max}}\)” (21, 50), a plateau of V\(_{\text{O2}}\) vs. work rate (or speed) can be simply constructed from just two exercise bouts (i.e., a maximal ramp/incremental and a subsequent exhausting constant work rate test at a higher work rate than that achieved at the end of the ramp test, Fig. 4). The key is that, although all work rates above CP but below that at which the highest V\(_{\text{O2}}\) was measured on the ramp test can bring V\(_{\text{O2}}\) to V\(_{\text{O2max}}\), the subsequent validation test must be at a higher work rate than attained in the ramp test. There is not a straightforward answer to the question: How much higher? The investigators must select a work rate that is sufficiently higher than that attained on the ramp test to give the V\(_{\text{O2}}\) signal for the higher work rate the opportunity to emerge from the extant
noise. In the event that the subsequent test produces a plateau signifying \( V_{O2\max} \), this signal would be lower than expected for the work rate based on the previous \( V_{O2} \)-work rate slope, thereby allowing construction of the relationship shown in Fig. 4. On the other hand, the work rate must be sustainable for sufficient duration that the kinetics allow achievement of the same or a greater \( V_{O2} \) if such is possible. To date, a work rate ~10% higher than that completed on the previous ramp test has proven effective (3, 29, 31, 40) but may not be ideal for all populations or circumstances. Resolution of this specific issue would be valuable. Imposition of work rates below the maximum achieved on the ramp test may allow construction of a plateau of \( V_{O2} \) against work rate but will not extend the relationship to the higher work rate critical to define \( V_{O2\max} \) (45; see also the RISE-95 test validated in chronic heart failure patients, Ref. 7).

For exercise tests run on the motor-driven treadmill, the same principles apply. However, the potential flexibility of increasing treadmill speed and/or incline is available. For experienced runners, Midgley and colleagues (29, 31) have advocated an increased running speed of 0.5 km/h for the verification bout and, in addition to the levelling-off criteria for \( V_{O2} \), a difference of less than 4 beats/min in HR between the maximum reached on the ramp/incremental and the verification bout (30). As was true for the cycle ergometry paradigm, identifying the correct work rate (i.e., speed/incline) for the treadmill verification bout within the specific pertinent populations tested is a valuable goal. Even when, or if, this goal is achieved, it is important that the investigator build into the experimental design the latitude to select sufficient verification bouts to provide definitive evidence for \( V_{O2\max} \).

The specific rest period intervening between the ramp and subsequent validation test does not seem to be critical. Specifically, 20 min has been recommended (33), but, whereas longer may be advisable for patient populations, healthy subjects may want or require far less (i.e., 5 or 10 min). Procedurally, asking subjects or patients to perform two exhausting bouts of exercise within the same laboratory or clinical testing session is feasible, and typically subjects prefer the constant work rate test to the ramp test. The limit of tolerance during this constant work rate test is usually 3–6 min. Depending on the subject’s \( V_{O2} \) kinetics, work rates that induce exhaustion in much less than 3 min might be in the so-called “extreme” work rate domain and therefore fail to achieve \( V_{O2\max} \) (22, 35, 56).

To date, this validation protocol has been performed successfully with children (6), healthy sedentary (2), athletic (51), and obese (43, 57) adults, and patient populations (e.g., cystic fibrosis, 44). It is also notable that, if, in a particular investigation, CP is being measured using repeated exhaustive severe-intensity constant work rate tests, one or more of these can be used to validate \( V_{O2\max} \) provided that they are above the maximum power output achieved on the ramp test. Selecting a particular work rate that is, for example, 10% above that which achieved the highest \( V_{O2} \) on the ramp test means that the absolute level to which \( V_{O2} \) rises is essentially independent of the contribution of the \( V_{O2} \) slow component, \( V_{O2} \) kinetics, or the tolerable duration of the exercise (22, 23, 34, 47, 56).

One alternative protocol to the exhausting constant work rate test recommended above is the 3-min all-out effort (49). This entails the subject cycling as fast as possible against a fixed resistance causing the power output to peak within a few seconds before falling precipitously to asymptote at CP after ~2 min. On this test, \( V_{O2\max} \) is typically achieved within ~1 min and sustained for the remainder of the test despite developed power falling far below that achieved at \( V_{O2\max} \) during ramp incremental exercise. Although the full potential of this test across populations remains to be determined, its extremely strenuous nature may be contraindicated in at least some patient populations and even in healthy individuals who are unused to maximal exhausting exercise. Murgatroyd et al. (32) have reported that an exhausting ramp incremental test followed immediately by a maximal-effort 3-min variable-power sprint test effectively validates \( V_{O2\max} \), suggesting that key parameters of aerobic fitness can be appraised within a single exercise test. Unfortunately, the variable power condition for both of these options confounds construction of the relationship demonstrated in Fig. 4 that defines \( V_{O2\max} \). Recently, there has been interest in the use of self-paced incremental exercise...
tests (using 5 × 2 min stages at fixed increments of rating of perceived exertion) for the determination of VO$_{2\text{max}}$ (27). Such a procedure, which has been shown to produce similar VO$_{2\text{max}}$ values to traditional ramp incremental tests (9, 20, 24), might be considered to have greater ecological validity than conventional protocols in which work rate is externally controlled and incremented in a strictly linear fashion. Moreover, the “closed-loop” nature of the self-paced protocol obviates the necessity for the researcher or clinician to estimate starting work rates and ramp rates. The utility of self-paced maximal exercise tests in the laboratory or clinical setting requires further evaluation but does not negate the value of the verification phase advocated herein, which has the singular advantage of allowing construction of the graphical solution to VO$_{2\text{max}}$ shown in the top panel of Fig. 4.

What Constitutes a “Plateau” for VO$_{2\text{max}}$ Identification

During incremental exercise, a plateau in VO$_2$ is considered to represent the attainment of the definitive VO$_{2\text{max}}$. However, exactly how the existence of this plateau is defined is rarely considered. The criterion proposed by Taylor et al. (48) of a change in VO$_2$ of <2.1 ml·kg$^{-1}$·min$^{-1}$ between consecutive stages as the subject approaches exhaustion is often applied indiscriminately and without consideration of the expected increment in VO$_2$ for the ramp or incremental rate being utilized. Moreover, the window over which VO$_2$ is averaged (i.e., 15, 30, or 60 s) can significantly influence both the absolute VO$_2$ and the proportion of plateaus identified (1, 23). The inclusion of a verification phase in the test battery helps to circumvent these issues. For example, consider a subject who completes a ramp incremental test and achieves a VO$_{2\text{peak}}$ of 2,500 ml/min at a peak work rate of 200 W (i.e., baseline of 500 ml/min during “unloaded” cycling and a VO$_2$ gain of 10 ml·min$^{-1}$·W$^{-1}$ throughout the test). If the verification phase is completed at 10% above the peak work rate (i.e., 220 W), then the expected VO$_2$ would be 2,700 ml/min. Given a possible error in VO$_2$ measurement of ~3% (~80 ml/min), a plateau could be accepted if VO$_2$ was ≤ 2,620 ml/min.

Midgley and colleagues have advocated prediction of the “expected” VO$_2$ for the verification test using a least-squares linear regression line fitted to the ramp test VO$_2$ data for minute −6 to −2 before exhaustion (30, 31). The slope of this line is then extended to the absolute work rate (e.g., 100% of maximum work rate on the ramp test). A VO$_2$ plateau and therefore VO$_{2\text{max}}$ verification has been defined as a difference between the modeled and actual VO$_2$ of >50% of the regression slope as determined above. This approach is advocated in preference to choosing some arbitrary VO$_2$ plateau “threshold” of 100, 200, or even 280 ml/min (31).

Conclusions

The VO$_{2\text{max}}$ is an essential paradigm in the evaluation of both the integrated capability of an individual’s pulmonary-cardiovascular-metabolic systems to transport and utilize O$_2$ to support muscle contraction and the efficacy of interventions designed to enhance one or more components of the O$_2$ transport/utilization pathway. Measurement of VO$_{2\text{max}}$ will therefore remain a vital technique in the arsenal of researchers in exercise physiology, sport scientists, and clinical exercise specialists. In this article, we have highlighted some of the methodological issues that can thwart accurate assessment of VO$_{2\text{max}}$ and cautioned against the acceptance of VO$_{2\text{peak}}$ measured during ramp incremental exercise as a maximum value in all but those who are familiar with maximal exercise testing and are highly motivated. We advocate the inclusion of a short constant work rate verification phase, completed at a higher work rate than that achieved during the ramp test, in the exercise test battery to enable verification of the VO$_{2\text{max}}$-Fundamental tenets of this approach are that it is objective, specific to the particular subject and exercise testing format or modality, and sufficiently robust that it is not impacted substantially by day-to-day differences in physiological responses (29, 30, 31).

ACKNOWLEDGMENTS

D. C. Poole first discussed the practical and graphical solution to the ramp test VO$_{2\text{max}}$ dilemma featured herein with the incomparable respiratory physiologist Professor Brian J. Whipp at the American College of Sports Medicine national meeting in Salt Lake City, Utah in 1990. We thank Dr. Daniel M. Hirai for expert advice.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

REFERENCES


J Appl Physiol • doi:10.1152/japplphysiol.01063.2016 • www.jappl.org


