Evaluation of "true" $\dot{V}O_{2max}$ based on a novel set of standardised criteria

Adrian W. Midgley[1], Sean Carroll[2], David Marchant[3], Lars R. McNaughton[1], Jason Siegler[1]

[1]Department of Sport, Health and Exercise Science, University of Hull, Hull, UK

[2]Carnegie Research Institute, Leeds Metropolitan University, Leeds, UK

[3]Department of Sport and Physical Activity, Edge Hill University, Ormskirk, UK

Corresponding author:

Adrian W. Midgley, PhD

Department of Sport, Health and Exercise Science

University of Hull

Cottingham Road

Hull

HU6 7RX

ENGLAND

Phone: 44 (0)1482-466432

Fax: 44 (0)1482-465149

Email: A.W.Midgley@hull.ac.uk

Abstract

Background. Criteria are used to identify whether a subject has elicited a $\dot{V}O_{2max}$. **Objective.** Evaluate the validity of traditional $\dot{V}O_{2max}$ criteria and propose a novel set of criteria. **Research design and methods.** Twenty athletes completed a $\dot{V}O_{2max}$ test consisting of an incremental phase and a subsequent supramaximal phase to exhaustion (verification phase). Traditional and novel $\dot{V}O_{2max}$ criteria were evaluated. Novel criteria were: 1) $\dot{V}O_2$ plateau defined as a difference between modelled and actual $\dot{V}O_{2max}$ >50% of the regression slope of the individual $\dot{V}O_2$-workrate relationship; 2) same as criterion 1, but for maximal verification $\dot{V}O_2$; 3) difference ≤4 beats·min^{-1} between maximal heart rate values in the two phases. **Results.** Satisfying the traditional $\dot{V}O_2$ plateau criterion was largely an artefact of the between-subject variation in the $\dot{V}O_2$-workrate relationship. Secondary criteria, supposedly an indicator of maximal effort, were often satisfied long before volitional exhaustion, even as low an intensity as 61% $\dot{V}O_{2max}$. No significant mean differences were observed between the incremental and verification phases for $\dot{V}O_2$ ($t = 0.4$; $p = 0.7$) or heart rate ($t = 0.8$; $p = 0.5$). The novel $\dot{V}O_2$ plateau criterion, $\dot{V}O_{2max}$ verification criterion and maximal heart rate verification criterion were satisfied by 17, 18 and 18 subjects, respectively. The small individual absolute differences in $\dot{V}O_2$ between incremental and verification phases, observed in most subjects, provided additional confidence that $\dot{V}O_{2max}$ was elicited. **Conclusion.** Current $\dot{V}O_{2max}$ criteria were not valid and novel criteria should be further explored.

Key Words: Maximal oxygen uptake, Plateau, Primary, Secondary, Testing, Verification phase

INTRODUCTION

Determination of the maximal oxygen uptake ($\dot{V}O_{2max}$) is one of the most common procedures performed in the exercise physiology laboratory. Test procedures that increase the reliability and validity of $\dot{V}O_{2max}$ determination therefore have widespread applicability. One such test procedure is the application of criteria used to assess whether an individual has attained a 'true' $\dot{V}O_{2max}$.

The primary criterion traditionally used for establishing that a true $\dot{V}O_{2max}$ has been attained is a $\dot{V}O_2$ plateau, defined as a small or no increase in $\dot{V}O_2$ in response to an increase in workrate (Taylor et al. 1955). Where no $\dot{V}O_2$ plateau is evident, secondary criteria have been used to indicate whether an individual has given a maximum effort. In exercise tests that incorporate a large proportion of muscle mass, if a maximum effort has been given, it can been assumed that $\dot{V}O_{2max}$ has probably been attained, irrespective of the occurrence of a $\dot{V}O_2$ plateau (Duncan et al. 1997). Secondary criteria have included the attainment of threshold values for the respiratory exchange ratio and heart rate during the incremental exercise test, and post-exercise blood lactate concentration (Howley et al. 1995). The rationale for secondary criteria is that although there is a strong theoretical basis for the $\dot{V}O_2$ plateau concept, many individuals do not exhibit a clearly definable $\dot{V}O_2$ plateau despite giving an apparent maximal effort (Doherty et al. 2003; Froelicher et al. 1974; Rossiter et al. 2006). The application of currently used primary and secondary criteria in evaluating true $\dot{V}O_{2max}$ has been criticised (Misquita et al. 2001; Niemela et al. 1980). In particular, Poole et al. (2008) recently investigated the validity of currently used $\dot{V}O_{2max}$ criteria and concluded that the threshold values typically used for the secondary criteria can be satisfied at exercise intensities as low as 73% $\dot{V}O_{2max}$. If the criticisms directed at the currently used $\dot{V}O_{2max}$ criteria are correct, then there is a need for a new set of criteria.

The verification phase of a $\dot{V}O_{2max}$ test, first proposed by Thoden et al. (1982), involves a single square wave bout of exercise performed shortly after the incremental phase (Thoden et al. 1982). Several recent studies have investigated the utility of the verification phase for establishing true $\dot{V}O_{2max}$ (Foster et al. 2007; Midgley et al. 2006; Rossiter et al. 2006). However, two of these studies (Foster et al. 2007; Rossiter et al. 2006) did not apply a verification phase criterion threshold to individual subjects, but instead, validated the verification phase by comparing the mean maximal $\dot{V}O_2$ values obtained in the incremental and verification phases. This approach has been criticised since it does not provide support as to whether an individual test has produced a true $\dot{V}O_{2max}$

(Noakes 2008). The only study to-date that used a verification criterion threshold for each test concluded that further research needs to be conducted to establish appropriate verification criteria and test protocols (Midgley et al. 2006).

The present study had two aims. The first aim was to extend the work of Poole et al. (2008) and further evaluate the validity of traditional primary and secondary $\dot{V}O_{2max}$ criteria using a larger sample and threshold values currently used during experimental research (Midgley et al. 2007b). In addition to presenting new findings on this topic, we hypothesised that the present study would confirm the findings of Poole et al. (2008). The second aim was to propose a new set of standardised $\dot{V}O_{2max}$ criteria. These criteria were based on a regression approach to the $\dot{V}O_2$ plateau criterion (Poole et al. 2008; Rossiter et al. 2006) and new criteria derived from the verification phase procedure (Thoden et al. 1982).

MATERIALS AND METHODS

Subjects. Twenty male athletes (10 runners, 10 cyclists) recruited from local athletic clubs volunteered to participate in the study (subject characteristics shown in Table 1). All subjects regularly engaged in competitive races and were apparently healthy, free from injury, not taking any medications and were non-smokers. The experimental procedures had previously been reviewed and approved by the departmental ethics committee (University of Hull) for research on human subjects. After being informed of potential risks and discomforts of participation, subjects gave written informed consent.

Determination of $\dot{V}O_{2max}$. Runners completed a treadmill test and cyclists a cycle ergometer test for the determination of $\dot{V}O_{2max}$. The tests for runners and cyclists differed only by ergometer type. The $\dot{V}O_{2max}$ test protocol involved warm-up, incremental, recovery, and verification phases. The warm-up phase consisted of 5-min of exercise at the same workrate as the initial stage of the incremental phase. The incremental phase consisted of 1 $km \cdot h^{-1}$ or 30 W increments every minute and was continued to the subject's limit of tolerance. The initial workrate was selected so that subjects reached their limit of tolerance in approximately 10-12 min. The recovery phase involved 10 min of passive rest, with a fingertip blood sample (~75 μl) taken 3 min into the recovery period for measurement of blood lactate concentration. The subsequent verification phase consisted of exercise for 2 min at 50% WR_{max} (the workrate attained in the last completed stage of the incremental phase), 1 min at 70% WR_{max}, and then to the limit of tolerance at a workrate equivalent to one stage higher than WR_{max}. The total duration of the verification phase was expected to be around 4.5 min. A schematic of the different phases of the $\dot{V}O_{2max}$ test is shown in Figure 1. Participants were familiarised with the test equipment and procedures during a prior visit to the laboratory.

Equipment. Running tests were performed on a computer-controlled motorised treadmill (Ergo ELG 55, Woodway GmbH, Rhein am, Germany) set at a 1% gradient (Jones and Doust, 1996) and cycling tests on a computer-controlled electromagnetically braked cycle ergometer (SRM, Schoberer Rad Mebtechnik, Julich, Germany). Respired air was analysed breath-by-breath using an automated open-circuit gas analysis system (Quark b^2, Cosmed Srl, Rome, Italy). The gas analysers were calibrated immediately before the incremental and verification phases of each test using ambient air and certified standard gases containing 16.0 ± 0.02% oxygen and 5.0 ± 0.02% carbon dioxide (Cryoservice Ltd, Worcester, UK). The gas analysers were also calibrated immediately after the verification phase to confirm the analysers were still functioning properly. The turbine flow meter used for the determination of minute ventilation was calibrated with a 3-L calibration syringe (Cosmed Srl, Rome, Italy) immediately before each test. Heart rate was continuously measured using a telemetric heart rate monitor integrated into the Quark b^2 analysis system (model T41, Polar Electro Oy, Kempele, Finland). Age-predicted maximal heart rate was calculated as 220-age. Blood lactate concentration was determined after each test using an automated system (YSI 2300 STAT PLUS®, Yellow Springs Instruments, Yellow Springs, Ohio). The lactate analyser was calibrated immediately prior to each test according to the manufacturer's instructions using a standard of known lactate concentration.

Criteria for $\dot{V}O_{2max}$ test evaluation. Breath-by-breath $\dot{V}O_2$ data were filtered using the default settings of the filtering function in the Quark b^2 data management software to remove errant breaths caused by swallowing, coughs, sighs, etc, that tend to mask the underlying physiological response (Lamarra et al. 1987). The $\dot{V}O_2$ data were then 30-s stationary time-averaged and the highest 30-s average in the incremental and verification phases were regarded as $\dot{V}O_{2max}$ and $\dot{V}O_{2verif}$, respectively. The 30-s stationary time average provides a good compromise between removing noise while maintaining the underlying trend in relatively rapidly changing $\dot{V}O_2$ data, such as that observed in the verification phase. An ordinary least squares linear regression line was then fitted to the 4 min of $\dot{V}O_2$ data immediately preceding the last 2 min of incremental phase $\dot{V}O_2$ data (SPSS® for Windows software, release 15.0, SPSS Inc., Chicago, IL). This approach was used to 'capture' the linear portion of the $\dot{V}O_2$ response by avoiding the influence of any non-linear $\dot{V}O_2$ kinetics in the early incremental phase response and any deviation from linearity at the end of the incremental phase (a similar approach to that used by Rossiter et al. 2006). To ensure this goal was achieved, a scatterplot of $\dot{V}O_2$ versus time was visually inspected for evidence of deviation from linearity before the last 2 min of the incremental phase. Under such circumstances, the regression line was fitted to the 4 min of $\dot{V}O_2$ data immediately preceding the start of the deviation from linearity. The regression line was then extrapolated to the end of the incremental phase (last completed 30 s) to obtain a modelled $\dot{V}O_{2max}$ value. This regression line was also extrapolated to the supramaximal workrate used in the verification phase to obtain a modelled $\dot{V}O_{2verif}$ value. Heart rate data were 5-s stationary time-averaged and the highest 5-s average in the incremental and verification phases were regarded as HR_{max} and HR_{verif}, respectively. To investigate random error in heart rate measurement, the differences between six successive 5-s averages during the end of the warm-up period (where a relative steady state heart rate was evident) were calculated for each subject. The random error component was used to establish an appropriate heart rate verification criterion threshold. Whether each subject had attained a "true" $\dot{V}O_{2max}$ was then evaluated using a new set of standardized criteria:

Criterion 1. $\dot{V}O_2$ plateau. A $\dot{V}O_2$ plateau was considered to be evident if the difference between the modelled and actual $\dot{V}O_{2max}$ (mL·min^{-1}) was greater than 50% of the regression slope for the linear portion of the $\dot{V}O_2$-workrate relationship;

Criterion 2. $\dot{V}O_{2max}$ verification. A difference between the modelled and actual $\dot{V}O_{2verif}$ (mL·min^{-1}) that was greater than 50% of the regression slope for the linear portion of the $\dot{V}O_2$-workrate relationship.

Criterion 3. Maximal heart rate verification. A difference of $\leq$4 beats·min^{-1} between HR_{max} and HR_{verif}.

The rationale for these criteria thresholds is provided in the results and discussion sections. When none of the above criteria were satisfied, this was regarded as sufficient evidence to suggest that the maximal $\dot{V}O_2$ value in the incremental phase was not the subject's true $\dot{V}O_{2max}$. If either criterion 1 or 2 was satisfied, then this was accepted as sufficient evidence that a $\dot{V}O_2$ plateau had occurred and $\dot{V}O_{2max}$ had been elicited. If criterion 3 was satisfied, then this was accepted as sufficient evidence that the subject provided a maximal effort and that the $\dot{V}O_{2max}$ was 'probably' elicited. The validity of traditional $\dot{V}O_{2max}$ criteria using commonly used threshold values (Midgley et al. 2007b) also were evaluated.

Evaluation of readiness and willingness to give a maximal effort. Task-specific measures of motivational components (Tenenbaum et al. 2005) were administered immediately before the test to establish each subject's readiness and willingness to give a maximum effort. Confidence was addressed through the single item of "How confident are you in pushing yourself to the limit of your exercise tolerance without giving up at an earlier point?" measured on a scale of 0 (not at all confident) to 100 (very, very confident). Other task specific components were measured on a 5-point Likert-type scale ranging from 1 (not at all / poor) to 5 (very much / excellent): (a) commitment and determination: "How committed and determined are you to exercise as long as you can?", (b) perceived ability and competence: "How do you perceive your competence in exercising to achieve your $\dot{V}O_{2max}$?", (c) effort: "How much effort do you intend on investing in this task?". Low scores on more than two of the four task-specific variables was accepted as sufficient evidence to suggest a subject was not ready or willing to invest a maximal effort. A low score was regarded as <60 for Confidence and <3 for the other items.

Statistical analyses: Statistical analyses were completed using SPSS® for Windows software (release 15.0; SPSS Inc., Chicago, IL). Assumptions of each statistical procedure were checked and verified. The sample data were described using the mean (SD) where normality was plausible, otherwise the median (interquartile range) were used. The differences between maximal $\dot{V}O_2$ and heart rate values during incremental and verification phases were analysed using two-tailed paired samples t tests. Differences between runners and cyclists responses to the incremental phase were analysed using two-tailed independent samples t tests. The variation in steady-state heart rate was analysed using a one-way ANOVA for repeated measures. Statistical significance was accepted at $p < 0.05$.

RESULTS

Mean (SD) responses for the incremental and verification phases of the $\dot{V}O_{2max}$ test are shown in Table 2.

Traditional $\dot{V}O_{2max}$ criteria

The number of subjects who satisfied different threshold values of the traditional primary and secondary $\dot{V}O_{2max}$ criteria are shown in Table 3. The data show that the number of subjects who satisfied each criterion was highly dependent on the threshold value that was used. Figure 2 shows the cumulative frequency of the number of subjects who satisfied respiratory exchange ratio (panel A) and heart rate (panel B) criteria thresholds at different percentages of $\dot{V}O_{2max}$. The 1.05, 1.10 and 1.15 thresholds for the respiratory exchange ratio criterion were satisfied at median (interquartile range) % $\dot{V}O_{2max}$ values of 91 (9), 95 (7) and 100 (6), respectively. The 85%, 90%, 95% and 100% age-predicted maximal heart rate thresholds for the heart rate criterion were satisfied at median (interquartile range) % $\dot{V}O_{2max}$ values of 78 (8), 87 (6), 96 (8) and 100 (1) respectively. The most liberal threshold values for the respiratory exchange ratio and heart rate criteria could be satisfied at as little as 61% and 68% $\dot{V}O_{2max}$, respectively. The heart rate criterion appeared particularly sensitive to changes in the threshold value used to define it. At 90% $\dot{V}O_{2max}$, for example, 17 of the 20 subjects satisfied the 85% age-predicted maximal heart rate threshold, but none had satisfied the 100% threshold.

The slope of the $\dot{V}O_2$-workrate relationship in the incremental phase was 225 (51) $mL \cdot min^{-1}$ per 1 $km \cdot h^{-1}$ increment for running (at a constant 1% gradient) and 325 (SD 54) $mL \cdot min^{-1}$ per 30 W increment for cycling. The mean r^2 for the regression models was 0.963 (SD 0.030) for running and 0.986 (SD 0.013) for cycling. Cyclists had significantly higher $\dot{V}O_2$-workrate slopes than the runners (mean difference 100 $mL \cdot min^{-1}$; 95% CI 51, 150 $mL \cdot min^{-1}$; $t = 4.3$; $p = <0.001$), as well as significantly higher maximal respiratory exchange ratios (mean difference 0.07; 95% CI 0.02, 0.12; $t = 2.8$; $p = 0.01$) and post-exercise blood lactate concentrations (mean difference 3.2 mM; 95% CI 1.4, 5.0 mM; $t = 3.8$; $p = 0.002$). However, the differences between the age-predicted maximal heart rate and the observed maximal heart rate were not significant between groups (mean difference 2.6 $beats \cdot min^{-1}$; 95% CI -8.5, 13.7; $t = 0.5$; $p = 0.6$).

Novel standardized $\dot{V}O_{2max}$ criteria

Seventeen subjects satisfied the $\dot{V}O_2$ plateau criterion, where two of the three subjects who did not satisfy the criterion exhibited a marked accelerated $\dot{V}O_2$ response at the end of the incremental phase. Eighteen subjects satisfied the $\dot{V}O_{2max}$ verification criterion. The three subjects who did not satisfy the $\dot{V}O_2$ plateau criterion also exhibited the smallest differences between the actual and observed $\dot{V}O_{2verif}$. Since there were no significant differences between runners and cyclists for the $\dot{V}O_{2max}$ - $\dot{V}O_{2verif}$ and HR_{max}-HR_{verif} differences, the data were pooled for inferential statistical analyses. The mean (SD) $\dot{V}O_{2max}$ - $\dot{V}O_{2verif}$ difference of 21 (230) $mL \cdot min^{-1}$ was not statistically significant (95% CI -86, 129; $t = 0.4$; $p = 0.7$). Figure 3 shows the HR_{max}-HR_{verif} differences for the incremental and verification phases. The mean (SD) HR_{max}-HR_{verif} difference of 0.5 (2.7) $beats \cdot min^{-1}$ was not statistically significant ($t = 0.8$; $p = 0.5$; 95% CI -0.8, 1.7). Eighteen subjects satisfied the maximal heart rate verification criterion. Out of the two subjects that did not satisfy the maximal heart rate verification criterion, one did not satisfy either of the other two criteria and one satisfied both of the other two criteria. Because the $\dot{V}O_{2max}$ - $\dot{V}O_{2verif}$ and HR_{max}-HR_{verif} differences approximated normal distributions (as evidenced by Q-Q plots), the differences were more clustered around the mean difference (i.e. close to zero for both $\dot{V}O_2$ and heart rate). Only six subjects had $\dot{V}O_{2max}$ - $\dot{V}O_{2verif}$ differences greater than 200 mL·min^{-1}, whereas only five subjects had HR_{max}-HR_{verif} differences greater than 2 beats·min^{-1}. However, a scatterplot of the $\dot{V}O_{2max}$ - $\dot{V}O_{2verif}$ differences clearly showed heteroscedastic errors (i.e. the differences increased as the subjects' $\dot{V}O_{2max}$ increased).

The standard deviation of the absolute differences between six successive 5-s heart rate averages during the end of the warm-up period, for all subjects, was 1.7 beats·min^{-1}. A non-significant one-way analysis of variance for repeated measures ($F = 0.6$; $p = 0.6$) provided evidence that heart rate was at a steady state during this time. The 4 beats·min^{-1} heart rate criterion threshold was derived from multiplying the within-subject standard deviation of the differences in steady-state 5-s heart rate averages by 1.96 (rounded up to the nearest integer).

All subjects scored between moderate to high on three or all of the four task-specific items used to evaluate each subject's readiness and willingness to invest a maximal effort. The median (interquartile range) for each item was as follows: Confidence, 91 (12); Determination and Commitment, 5.0 (0); Competence, 4.5 (1.0); and Effort, 5.0 (0).

DISCUSSION

Traditional $\dot{V}O_{2max}$ criteria

One of the main findings of this study was that the currently used primary and secondary criteria used to assess whether a true $\dot{V}O_{2max}$ has been attained, are not valid for continuously incremented test protocols widely used for the determination of $\dot{V}O_{2max}$. Our findings provide experimental support for the concerns expressed by others (Misquita et al. 2001; Niemela et al. 1980; Poole et al. 2008).

The most commonly used criterion threshold to define a $\dot{V}O_2$ plateau (Midgley et al. 2007b) is the 150 mL·min^{-1} proposed by Taylor et al. (1955), which represented half of the mean increase in $\dot{V}O_2$ [299.3 (SD 86.5) mL·min^{-1}] in response to a 2.5% increase in treadmill grade. The mean (SD) $\dot{V}O_2$-workrate slope found in the present study [275 (SD 72) mL·min^{-1}] was similar to that reported by Taylor et al. (1955). This large between-subject variation around the mean response meant that satisfying the $\dot{V}O_2$ plateau criterion threshold was largely an artefact of differences in individual $\dot{V}O_2$-workrate slopes. The criterion threshold of 150 mL·min^{-1} was between 36% and 90% of the slope for each subject, thereby representing a large deviation from linearity for some subjects and for others, only a small deviation. Previous studies also have used arbitrary $\dot{V}O_2$ plateau criterion thresholds of 100, 200, and 280 mL·min^{-1} (Midgley et al. 2007b). The 280 mL·min^{-1} criterion threshold was higher than the $\dot{V}O_2$-workrate slope for ten subjects in the presents study, and therefore, this criterion would have been satisfied even if there was no downward deviation of the $\dot{V}O_2$-workrate relationship. In contrast, a 100 mL·min^{-1} threshold was less than 33% of the $\dot{V}O_2$-workrate slope for six of the 20 subjects and would have been comparatively difficult to achieve. In addition to being highly dependent on between-subject differences in the $\dot{V}O_2$-workrate slope, the traditional $\dot{V}O_2$ plateau criterion is dependent on the test protocol and the associated expected increase in $\dot{V}O_2$ per unit of time. The cyclists had significantly higher $\dot{V}O_2$-workrate slopes than the runners (probably related to higher workrate increments in the cycling protocol), indicating that, all other things being equal, the traditional $\dot{V}O_2$ plateau threshold would have been more easily satisfied by cyclists.

Figure 2A-B shows that many subjects satisfied threshold values for the respiratory exchange ratio and heart rate criteria at exercise intensities notably lower than those that elicited $\dot{V}O_{2max}$. Threshold values for the respiratory exchange ratio and age-predicted maximal heart rate criteria of 1.10 and 90%, respectively, have been used widely (Midgley et al. 2007b). At 90% $\dot{V}O_{2max}$, four of the 20 subjects had satisfied this respiratory exchange ratio criterion and 12 subjects had satisfied the heart rate criterion. The most liberal criterion thresholds for the respiratory exchange ratio and heart rate criteria, used previously in experimental research (Howley et al. 1995; Midgley et al. 2007b), were satisfied at exercise intensities as low as 61 and 68% $\dot{V}O_{2max}$,respectively. Poole et al. (2008) previously reported that a respiratory exchange ratio threshold of 1.10 and an age-predicted maximal heart rate threshold of 95% were satisfied at an exercise intensity as low as 73% $\dot{V}O_{2max}$ in eight apparently healthy men. Our results, conducted on a larger sample (n = 20), support the findings of Poole et al. (2008) that $\dot{V}O_{2max}$ can be 'confirmed' at values appreciably lower than the true $\dot{V}O_{2max}$.

Cyclists attained significantly higher respiratory exchange ratios and blood lactate concentrations than runners, indicating that the respiratory exchange ratio and blood lactate criteria are largely dependent on either the population undergoing $\dot{V}O_{2max}$ testing, the exercise modality, the test protocol, or a combination of these factors. Since the runners and cyclists were similar in competitive level and the incremental running and cycling test protocols were of similar duration, the differences in respiratory exchange ratio and blood lactate concentration are most likely to be explained by differences in the athletes' training, or differences in the physiological demands of the two modes of exercise. We did not record the athletes' typical training prior to testing so could not discern whether this explained some of the variance. A plausible explanation for the differences is that the cycling required a greater muscular force output than running, and therefore, elicited greater recruitment of fast twitch fibres. Fast twitch fibres are known to have a greater capacity for glycolytic metabolism and therefore can produce greater quantities of lactate than the more oxidative slow twitch fibres (Borges and Essén-Gustavsson, 1989). Greater plasma shifts, hemoconcentration (Senay, Jr., et al. 1980) and reduced blood flow in the legs (Matsui et al. 1978) during cycling, compared to running, also are possible explanations.

The test protocol dependency of the respiratory exchange ratio criterion is apparent when considering that longer incremental tests protocols used for $\dot{V}O_{2max}$ determination have been shown to elicit significantly lower respiratory exchange ratio values (Bentley and McNaughton 2003; Lukaski et al. 1989; Pollock et al. 1982). One study reported that despite no significant difference in $\dot{V}O_{2max}$, an incremental test with a mean duration of 9.1 (SD 0.8) min resulted in a mean respiratory exchange ratio of 1.21 (SD 0.05; range 1.18 to 1.26), compared to a significantly lower mean respiratory exchange ratio of 1.08 (SD 0.02; range 1.06 to 1.10) for a test with a mean duration of 24.4 (SD 2.6) min (Bentley and McNaughton 2003). Consequently, none of the subjects satisfied a respiratory exchange ratio criterion threshold of 1.15 in the long test, but the same subjects all satisfied the criterion in the short test. Failure to satisfy the respiratory exchange ratio criterion is therefore largely an artefact of test duration. Continuous test protocols have also been found to result in significantly lower respiratory exchange ratios and post-exercise blood lactate concentrations than discontinuous protocols (Duncan et al. 1997), further emphasising the test protocol dependency of these criteria.

A survey of currently used $\dot{V}O_{2max}$ criteria showed that from 79 of 207 studies that reported criteria for $\dot{V}O_{2max}$ tests, seven different threshold values were used for the $\dot{V}O_2$ plateau, two for blood lactate, eight for respiratory exchange ratio, and ten for the heart rate criterion (Midgley et al. 2007b). As would be expected, in the present study, the number of subjects who satisfied different criteria was dependent on the threshold values that were used (Table 3).

This sensitivity to different criterion thresholds was most evident for the $\dot{V}O_2$ plateau and heart rate criteria. There was an almost two-fold difference in the number of subjects who satisfied the criteria when changing from the most liberal to the most conservative threshold values. Except for lower blood lactate and respiratory exchange ratio thresholds for testing children (Armstrong and Welsman 1994) and occasional attempts to identify protocol specific $\dot{V}O_2$ plateaux (Mitchell et al. 1958), there has been no attempt to rationalise the use of these widely different criterion threshold values. The lack of standardisation in the use of $\dot{V}O_{2max}$ criteria thresholds exposes the procedure to misuse, by allowing researchers to choose criteria thresholds after the data has been collected that allow all the subjects to satisfy the criteria and be retained in the study (Midgley et al. 2007b). Such an approach would be counterproductive to the sole purpose of the $\dot{V}O_{2max}$ criteria.

Novel standardized $\dot{V}O_{2max}$ criteria

The non-significant mean $\dot{V}O_{2max}$ - $\dot{V}O_{2verif}$ difference of 21 mL·min^{-1} (0.5%) reported in the present study, is similar to the range of mean differences (12 to 47 mL·min^{-1}) reported in four studies using running (Foster et al. 2007; Midgley et al. 2006; Midgley et al. 2007a) and cycling (Foster et al. 2007; Rossiter et al. 2006) protocols. These non-significant mean $\dot{V}O_{2max}$ - $\dot{V}O_{2verif}$ differences indicate that the spread of the individual differences are random errors of $\dot{V}O_2$ determination around the mean difference, thereby validating the efficacy of the particular incremental test protocols used in these studies for eliciting a true $\dot{V}O_{2max}$. However, using this mean response approach could mask one or more individuals who have not given a maximal effort and therefore probably not elicited a true $\dot{V}O_{2max}$. The $\dot{V}O_{2max}$ verification criterion, in addition to other $\dot{V}O_{2max}$ criteria, should therefore always be applied on an individual basis.

The main advantage of applying linear regression to model the $\dot{V}O_2$-workrate relationship immediately prior to any potential plateau in the $\dot{V}O_2$ response of each individual subject, is that the derived $\dot{V}O_2$ plateau criterion threshold is specific to that subject and exercise test (including protocol, ergometer type and test occasion). This is in contrast to the traditional $\dot{V}O_2$ plateau criterion that is highly dependent on the test methodology and the mean response of the subject population from which the criterion was originally derived.

The rationale for the maximal heart rate verification criterion is that it is improbable that a subject would attain very similar peak heart rate values in two dissimilar bouts of exercise (incremental versus verification phase)(Midgley et al. 2007b). The utility of the verification phase is therefore largely dependent on minimising the HR_{max}-HR_{verif} differences. Midgley et al. (2006) previously reported a small but significantly lower mean maximal heart rate in the verification phase compared to the incremental phase. The authors suggested that the verification phase duration of 168 (SD 35) s may have provided insufficient time for the heart rate to reach its maximum. Since the maximal heart rate verification threshold in the present study was only 4 beats·min^{-1}, this bias towards a lower heart rate would have decreased the utility of the maximal heart rate verification criterion. The present study used a more prolonged multi-stage verification phase resulting in a mean duration of 276 (SD 20) s. The negligible mean HR_{max}-HR_{verif} difference suggests that the multi-stage verification phase was more efficacious than the single square wave verification protocol for maximal heart rate verification.

The original heart rate verification criterion threshold was 2 beats·min^{-1}, however, the authors concluded that this threshold may be too conservative because of random error in heart rate determination due to technical error of measurement and natural physiological causes (Maritz et al. 1961). These random errors have the potential to artificially elevate the maximal heart value attained in either the incremental or verification phase and increase the HR_{max}-HR_{verif} difference. The present study attempted to use a more objective maximal heart rate verification criterion threshold. A non-significant repeated measures one-way analysis of variance supported the view that heart rate was at a steady-state during the end of the warm-up period and that any variability in 5-s averaged heart rate values over time was random. The 4 beats·min^{-1} heart rate verification criterion threshold was established by multiplying the within-subject standard deviation of the differences in steady-state 5-s heart rate averages by 1.96 (rounded upwards to the nearest integer). The criterion of 1.96 times the within-subject standard deviation ensured that any HR_{max}-HR_{verif} differences greater that this value are unlikely to be due to random error and used as evidence that a submaximal effort may have been given on one of the two $\dot{V}O_{2max}$ test phases. One limitation of this criterion methodology is that the variability in steady state heart rate, used to establish the criterion threshold, would include normal physiological variation. However, the methodology used appeared to be the only available method for also ‘capturing’ random error due to technical error of measurement and other error from natural physiological causes.

Key aspects of valid $\dot{V}O_{2max}$ criteria are that they should be: 1) objective; 2) specific to the subject, exercise modality and protocol of each $\dot{V}O_{2max}$ test; and 3) not unduly affected by day-to-day variation in physiological responses. The novel set of $\dot{V}O_{2max}$ criteria proposed in the present study appear to satisfy the latter two of these properties, since the criteria are only dependent on the maximal physiological responses for the specific test that is being performed by a particular individual at that specific time and day. Objectivity also appears to be satisfied, except that the decision to use a 50% threshold value for the $\dot{V}O_2$ plateau and $\dot{V}O_{2max}$ verification criteria is largely arbitrary. The rationale for the 50% threshold was that if the $\dot{V}O_2$-workrate slope had decreased by at least half, this would indicate the subject was at or close to his or her $\dot{V}O_{2max}$. More stringent thresholds are problematic because they are close to measurement error values for $\dot{V}O_2$ determination (Howley et al. 1995).

Due to the requirement for the achievement and maintenance of high levels of exertion for accurate $\dot{V}O_{2max}$ testing, the present study measured task-specific motivational characteristics immediately prior to the $\dot{V}O_{2max}$ test. Midgley et al. (2007b) suggested that this 'auxiliary' information could be used to assist in establishing whether a subject is ready and willing to invest a maximal effort. These measures were based on recent social-cognitive psychological research showing that task-specific self-efficacy (confidence in adhering to the task) and motivational characteristics such as, readiness to invest effort, determination and commitment, and perceived competence significantly influence an individual's persistence at and tolerance of high levels of physical exertion (e.g. Tenenbaum et al. 2001; Tenenbaum et al. 2005). All subjects scored between moderate to high on three or all of the items used to measure task-specific motivational characteristics. Although these results do not confirm that subjects gave a maximal effort during the test, they do indicate that immediately prior to the test, all the participants in the study were at least ready and willing to invest a maximal effort. This finding improves the confidence one can place in the observed data regarding achievement of the $\dot{V}O_{2max}$ criteria proposed in the present study. However, further research is needed to evaluate these $\dot{V}O_2$ criteria using subjects who are heterogeneous in terms of their readiness and willingness to invest a maximal effort during the $\dot{V}O_{2max}$ test.

In conclusion, where invalid $\dot{V}O_{2max}$ values could alter the interpretation of research findings, the need for robust criteria to help identify subjects who may not have attained a true $\dot{V}O_{2max}$ is apparent. Traditional $\dot{V}O_{2max}$ criteria lack validity because they are considerably influenced by the incremental exercise test duration, exercise modality, and between-subject differences in maximal attainable values for each criterion. Satisfying particular criteria also is highly dependent on how criteria are defined. Upon considering these limitations, traditional $\dot{V}O_{2max}$ criteria should not be used and research should focus on developing a new set of criteria. Proposed criteria should be independent of the characteristics of the test protocol and subject being tested, so that the criteria can be universally and uniformly applied. The present study proposes a novel set of standardised $\dot{V}O_{2max}$ criteria that appear to satisfy these properties.

REFERENCES

Armstrong, N. and Welsman, J. R. 1994. Assessment and interpretation of aerobic fitness in children and adolescents. Exerc. Sport Sci. Rev. **22**: 435-476.

Bentley, D. J. and McNaughton, L. R. 2003. Comparison of W_{peak}, $\dot{V}O_{2peak}$ and the ventilation threshold from two different incremental exercise tests: relationship to endurance performance. J. Sci. Med. Sport. **6**: 422-435.

Borges, O. and Essén-Gustavsson, B. (1989). Enzyme activities in type I and II muscle fibres of human skeletal muscle in relation to age and torque development. Acta Physiol. Scand. **136**: 29-36.

Doherty, M., Nobbs, L. and Noakes, T. D. 2003. Low frequency of the "plateau phenomenon" during maximal exercise in elite British athletes. Eur. J. Appl. Physiol. **89**: 619-623.

Duncan, G. E., Howley, E. T. and Johnson, B. N. 1997. Applicability of $\dot{V}O_{2\,max}$ criteria: discontinuous versus continuous protocols. Med. Sci. Sports Exerc. **29**: 273-278.

Foster, C., Kuffel, E., Bradley, N., Battista, R. A., Wright, G., Porcari, J. P., Lucia, A. and Dekoning, J. J. 2007. $\dot{V}O_{2\,max}$ during successive maximal efforts. Eur. J. Appl. Physiol. **102**: 67-72.

Froelicher, V. F., Jr., Brammell, H., Davis, G., Noguera, I., Stewart, A. and Lancaster, M. C. 1974. A comparison of three maximal treadmill exercise protocols. J. Appl. Physiol. **36**: 720-725.

Howley, E. T., Bassett, D. R., Jr. and Welch, H. G. 1995. Criteria for maximal oxygen uptake: review and commentary. Med. Sci. Sports Exerc. **27**: 1292-1301.

Jones, A. M., Doust, J. H. 1996. A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. J. Sports Sci. **14**: 321-327.

Lamarra, N., Whipp, B. J., Ward, S. A. and Wasserman, K. 1987. Effect of interbreath fluctuations on characterizing exercise gas exchange kinetics. J. Appl. Physiol. **62**: 2003-2012.

Lukaski, H. C., Bolonchuk, W. W. and Klevay, L. M. 1989. Comparison of metabolic responses and oxygen cost during maximal exercise using three treadmill protocols. J. Sports Med. Phys. Fitness. **29**: 223-229.

Maritz, J. S., Morrison, J. F., Peter, J., Strydom, N. B. and Wyndham, C. H. 1961. A practical method of estimating an individual’s maximal oxygen intake. Ergonomics. **4**: 97-122.

Matsui, H., Kitamura, K. And Miyamura, M. 1978. Oxygen uptake and blood flow of the lower limb in maximal treadmill and bicycle exercise. Eur. J. Appl. Physiol. **40**: 57-62.

Midgley, A. W., McNaughton, L. R. and Carroll, S. 2006. Verification phase as a useful tool in the determination of the maximal oxygen uptake of runners. Appl. Physiol. Nutr. Metab. **31**: 541-548.

Midgley, A. W., McNaughton, L. R. and Carroll, S. 2007a. Time at $\dot{V}O_{2\,max}$ during intermittent treadmill running: test protocol dependent or methodological artefact? Int. J. Sports Med. **28**: 934-939.

Midgley, A. W., McNaughton, L. R., Polman, R. and Marchant, D. 2007b. Criteria for determination of the maximal oxygen uptake: A brief critique and recommendations for future research. Sports Med. **37**: 1019-1028.

Misquita, N. A., Davis, D. C., Dobrovolny, C. L., Ryan, A. S., Dennis, K. E. and Nicklas, B. J. 2001. Applicability of maximal oxygen consumption criteria in obese, postmenopausal women. J. Womens Health Gend. Based Med. **10**: 879-885.

Mitchell, J. H., Sproule, B. J. and Chapman, C. B. 1958. The physiological meaning of the maximal oxygen intake test. J. Clin. Invest. **37**: 538-547.

Niemela, K., Palatsi, I., Linnaluoto, M. and Takkunen, J. 1980. Criteria for maximum oxygen uptake in progressive bicycle tests. Eur. J. Appl. Physiol. Occup. Physiol. **44**: 51-59.

Noakes, T. D. 2008. Maximal oxygen uptake as a parametric measure of cardiorespiratory capacity: comment. Med. Sci. Sports Exerc. **40**: 585.

Pollock, M. L., Foster, C., Schmidt, D., Hellman, C., Linnerud, A. C. and Ward, A. 1982. Comparative analysis of physiologic responses to three different maximal graded exercise test protocols in healthy women. Am. Heart J. **103**: 363-373.

Poole, D. C., Wilkerson, D. P. and Jones, A. M. 2008. Validity of criteria for establishing maximal O_2 uptake during ramp exercise tests. Eur. J. Appl. Physiol. **102**: 403-410.

Rossiter, H. B., Kowalchuk, J. M. and Whipp, B. J. 2006. A test to establish maximum O_2 uptake despite no plateau in the O2 uptake response to ramp incremental exercise. J. Appl. Physiol. **100**: 764-770.

Senay, Jr., L.C., Rogers, G. and Jooste, P. 1980. Changes in blood plasma during progressive treadmill and cycle exercise. J. Appl. Physiol.: Respirat. Environ. Exercise Physiol. **49**: 59-65.

Taylor, H. L., Buskirk, E. and Henschel, A. 1955. Maximal oxygen intake as an objective measure of cardio-respiratory performance. J. Appl. Physiol. **8**: 73-80.

Tenenbaum, G., Hall, H. K., Calcagnini, N., Lange, R., Freeman, G. and Lloyd, M. 2001. Coping with physical exertion and negative feedback under competitive and self-standard conditions. J. Appl. Soc. Psychol. **31**: 1582-1626.

Tenenbaum, G., Lidor, R., Lavyan, N., Morrow, K., Tonnel, S. and Gershgoren, A. 2005. Dispositional and task-specific social-cognitive determinants of physical effort perseverance. J. Psychol. **139**: 139-157.

Thoden, J. S., MacDougall, J. D. and Wilson, B. A. 1982. Testing aerobic power. *In* Physiological testing of the elite athlete. *Edited by* J. D. MacDougall, H. A. Wenger and H. J. Green (eds). Mouvement Publications, Inc, Ithaca, NY, pp. 39-54.

Figure legends

FIGURE 1. Schematic of the different phases of the $\dot{V}O_{2max}$ (treadmill) test protocol. 'Calibrate' refers to the relative time points during the test when the gas analysers were calibrated.

FIGURE 2. Cumulative frequency of subjects (n = 20) who satisfied the respiratory exchange ratio (panel A) and heart rate (panel B) criteria at increasing percentages of $\dot{V}O_{2max}$. Threshold values of 1.05, 1.10 and 1.15 were used for the respiratory exchange ratio criterion and 85%, 90%, 95% and 100% age-predicted maximal heart rate (APMHR) for the heart rate criterion. Age-predicted maximal heart rate was calculated as 220 minus the subject's age. The criteria have all been used in previous experimental research (Midgley et al., 2007b). The lines do not all reach 20 on the Y axis because some subjects did not satisfy particular criterion thresholds. The figure shows that subjects often satisfied the different criteria thresholds at $\dot{V}O_2$ values well below $\dot{V}O_{2max}$. For example, eight subjects had satisfied the 90% APMHR criterion threshold (a common threshold criterion; Midgley et al. 2007b) at 85% $\dot{V}O_{2max}$.

FIGURE 3. Bland-Altman plot showing the incremental and verification phase differences for maximal heart rate. Most of the differences are close to the line of identity (at zero on the ordinate) and only three are greater than 3 beats·min^{-1}.

Table 1. Mean (SD) subject characteristics.

	Age (yr)	Height (m)	Body mass (kg)	Σ7 skinfolds (mm)†	% body fat
Runners (n = 10)	39.3 (6.9)	1.76 (0.07)	73.0 (9.9)	84.2 (39.9)	13.2 (6.4)
Cyclists (n = 10)	36.0 (4.1)	1.74 (0.06)	70.3 (6.9)	67.0 (25.1)	10.4 (4.0)
Total (n = 20)	37.6 (5.8)	1.75 (0.07)	71.6 (8.4)	75.6 (33.6)	11.8 (5.4)

† chest, midaxilla, abdomen, suprailliac, thigh, subscapula, triceps.

Table 2. Maximal responses to the $\dot{V}O_{2max}$ test. Mean (SD).

	Incremental phase t_{lim} (s)	Verification phase t_{lim} (s)	$\dot{V}O_{2max}$ (mL·min^{-1})	$\dot{V}O_{2verif}$ (mL·min^{-1})	HR_{max} (beats·min^{-1})	HR_{verif} (beats·min^{-1})	RER_{max}	[BLa] (mM)†
Running (n = 10)	684 (48)	270 (24)	3863 (394)	3915 (466)	177 (17)	178 (15)	1.16 (0.06)	8.3 (2.0)
Cycling (n = 10)	642 (78)	282 (18)	4054 (467)	3958 (381)	183 (8)	184 (8)	1.23 (0.03)‡	11.5 (1.2)‡
Total (n = 20)	660 (66)	276 (18)	3958 (432)	3937 (415)	180 (13)	181 (12)	1.19 (0.06)	10.2 (2.2)

t_{lim} = time to exhaustion. **$\dot{V}O_{2max}$** = maximal $\dot{V}O_2$ in the incremental phase. **$\dot{V}O_{2verif}$** = maximal $\dot{V}O_2$ in the verification phase. **HR_{max}** = maximal heart rate in the incremental phase. **HR_{verif}** = maximal heart rate in the verification phase**.** **RER_{max}** = maximal respiratory exchange ratio in the incremental phase. **[BLa]** = post-incremental phase blood lactate concentration. † Due to technical problems, data are only for 6 treadmill tests and 9 cycle ergometer tests. ‡ Significantly higher than the runners ($p < 0.05$).

Table 3. Number of subjects who satisfied traditional $\dot{V}O_{2\,max}$ criteria. Different threshold values have been used to highlight its effect on how many subjects satisfy each criterion. These threshold values were used in some of 207 experimental studies (that conducted $\dot{V}O_{2\,max}$ tests) published in four journals in 2005 and 2006 (Midgley 2007b).

Criterion	$\dot{V}O_2$ plateau			Heart rate				RER			Blood lactate †	
Threshold value	<200 $mL \cdot min^{-1}$	<150 $mL \cdot min^{-1}$	<100 $mL \cdot min^{-1}$	within 85% APMHR	within 90% APMHR	within 95% APMHR	within 100% APMHR	≥1.05	≥1.10	≥1.15	≥8 mM	≥10 mM
Runners (n = 10)	7	7	4	9	8	8	6	10	9	4	4	1
Cyclists (n = 10)	10	8	5	10	10	9	4	10	10	10	9	8
Total (n = 20)	17	15	9	19	18	17	10	20	19	14	13	9

APMHR = age-predicted maximal heart rate; **RER** = respiratory exchange ratio. † Due to technical problems, data are only for 6 treadmill tests and 9 cycle ergometer tests.

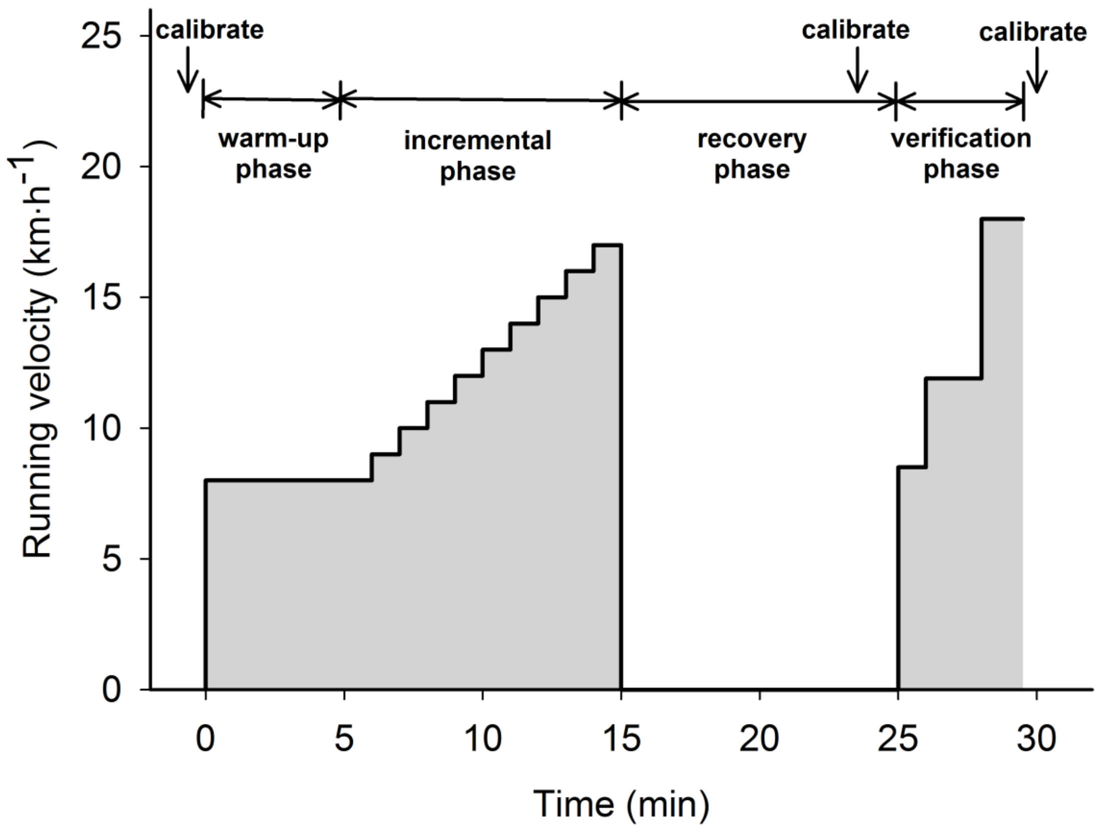
calibrate
calibrate
calibrate
warm-up phase
incremental phase
recovery phase
verification phase
Running velocity (km·h-1)
25
20
15
10
5
0
0
5
10
15
20
25
30
Time (min)

Figure 1.

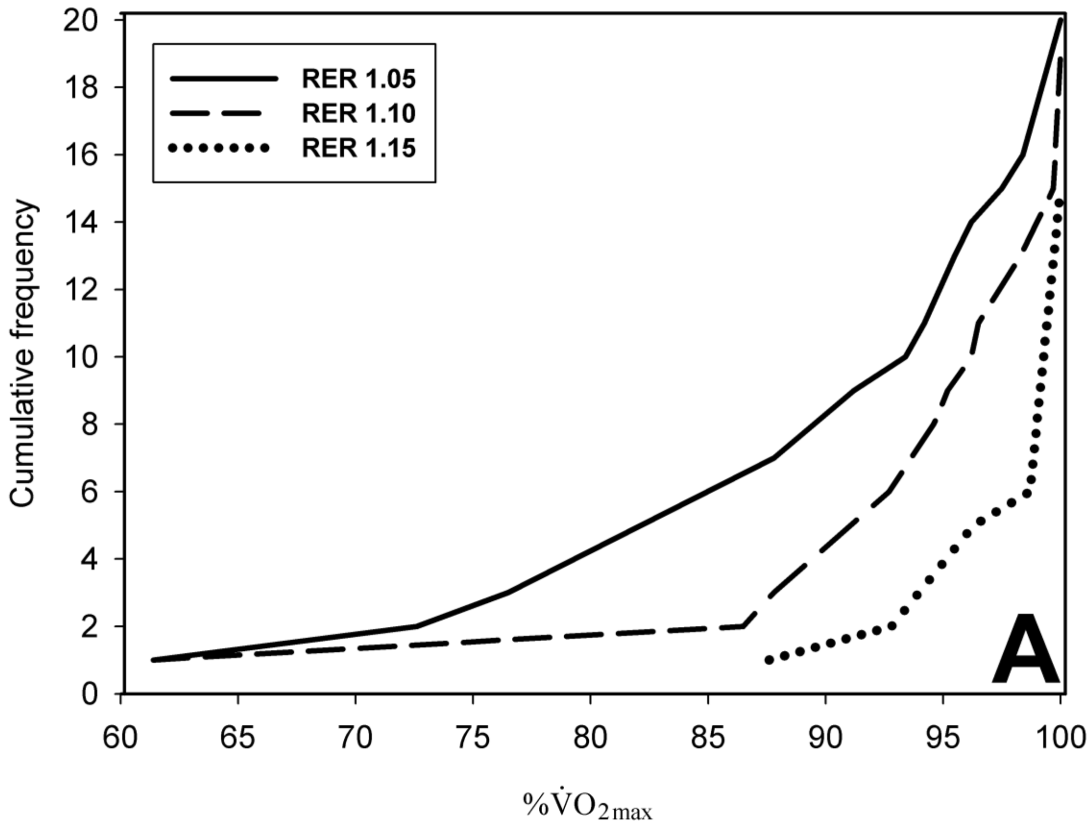

Figure 2A.

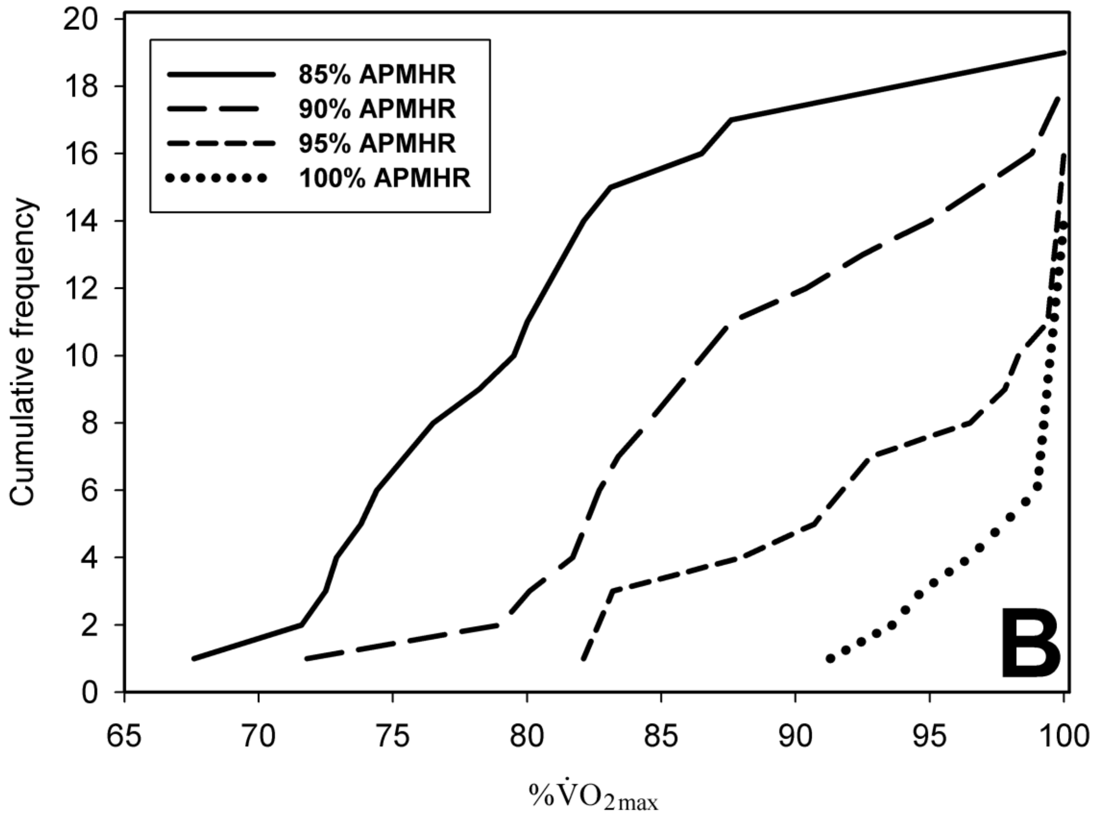

Figure 2B.

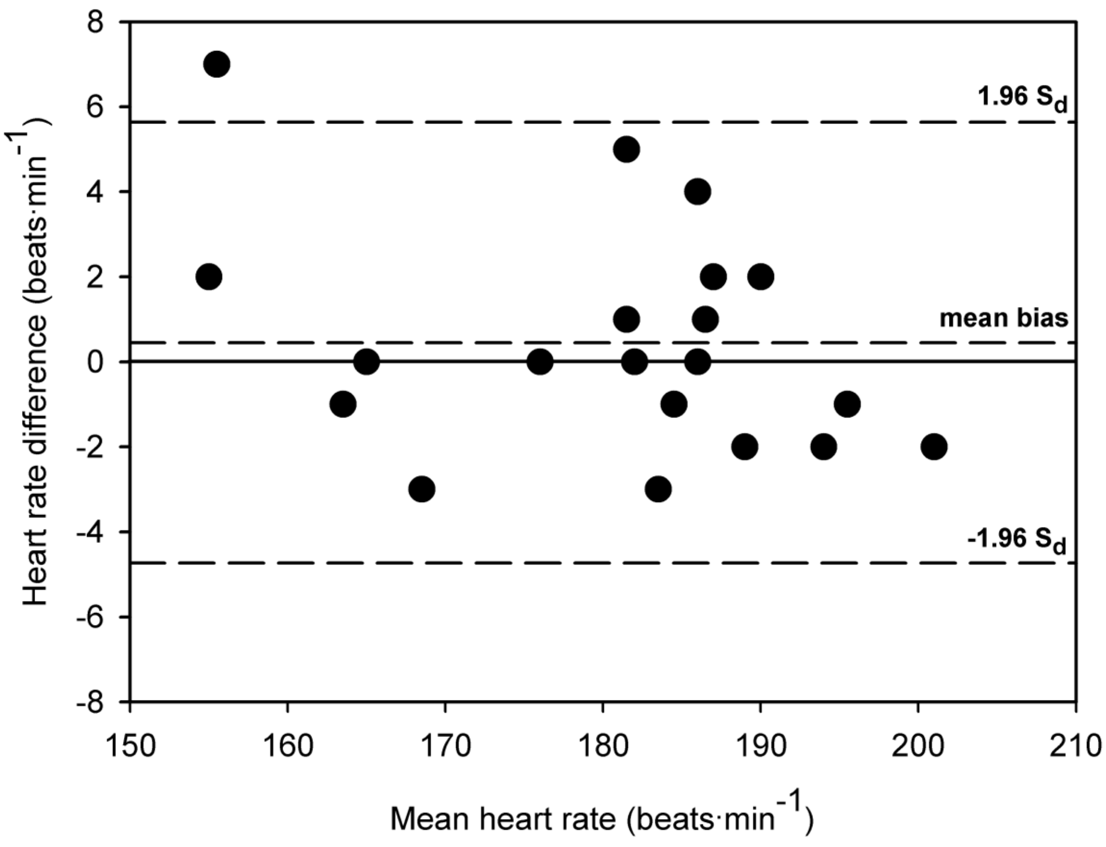
1.96 S_d
mean bias
-1.96 S_d
Heart rate difference (beats·min-1)
Mean heart rate (beats·min-1)
8
6
4
2
0
-2
-4
-6
-8
150
160
170
180
190
200
210

Figure 3.