RENWICK, J.R.M., PREOBRAZENSKI, N., GIUDICE, M.D., SWINTON, P.A. and GURD, B.J. 2023. Including supramaximal verification reduced uncertainty in VO_{2peak} response rate. *Applied physiology, nutrition and metabolism* [online], (accepted).

Including supramaximal verification reduced uncertainty in VO_{2peak} response rate.

RENWICK, J.R.M., PREOBRAZENSKI, N., GIUDICE, M.D., SWINTON, P.A. and GURD, B.J.

2023

© 2023 The Author(s). Permission for reuse can be obtained from copyright.com.

This document was downloaded from https://openair.rgu.ac.uk

SEE TERMS OF USE IN BOX ABOVE

DISTRIBUTED UNDER LICENCE

Including supramaximal verification reduced uncertainty in VO_{2peak} response rate

John R.M. Renwick,[1*] Nicholas Preobrazenski,[1,2*] Michael D. Giudice,[1] Paul A. Swinton,[3] and Brendon J. Gurd[1]

[1] School of Kinesiology and Health Studies, Queen's University, Kingston, Ontario, Canada

[2] Faculty of Medicine, University of Ottawa, Ottawa, ON K1N 6N5, Canada

[3] School of Health Sciences, Robert Gordon University, Aberdeen AB10 7QE, United Kingdom

* Contributed equally to manuscript.

Corresponding Author:

Brendon J. Gurd, PhD

Telephone: 613-533-6000 x79023

Fax: 613-533-6000

Email: gurdb@queensu.ca

ABSTRACT

Background: Many reports describe using a supramaximal verification phase - exercising at a power output higher than the highest power output recorded during an incremental cardiopulmonary test - to validate VO_{2max}. The impact of verification phases on estimating the proportion of individuals who increased VO_{2peak} in response to high-intensity interval training (HIIT) remains an underexplored area in the individual response literature.

Methods: This analysis investigated the influence of same-day and separate-day verification phases during repeated measurements (incremental tests – INCR1 and INCR2; incremental tests + supramaximal verification phases – INCR1+ and INCR2+) of VO_{2peak} on typical error (TE) and the proportion of individuals classified as responders (i.e. the response rate) following four weeks of HIIT (n=25) or a no-exercise control period (n=9).

Results: Incorporation of supramaximal verification consistently reduced the standard deviation of individual response, typical error, and confidence interval widths. However, variances were statistically similar across all groups ($p>0.05$). Response rates increased when incorporating either one (INCR1 to INCR1+; 24% to 48%, $p=0.07$) or two (INCR2 to INCR2+; 28% to 48%, $p=0.063$) supramaximal verification phase(s). However, response rates remained unchanged when either zero-based thresholds or smallest worthwhile difference response thresholds were used (50% and 90% confidence intervals, all $p>0.05$).

Conclusion: Supramaximal verification phases reduced random variability in VO_{2peak} response to HIIT. Compared with separate-day testing (INCR2 and INCR2+), the incorporation of a same-day verification (INCR1+) reduced CI widths the most. Researchers should consider using a same-day verification phase to reduce uncertainty and better estimate VO_{2peak} response rate to HIIT.

Keywords: supramaximal verification phase, response classification, maximal oxygen uptake, incremental testing, responder, typical error, individual response to exercise, response rate.

INTRODUCTION

Although research examining individual variability in exercise training-induced maximal oxygen uptake (VO_{2max}) response has become common in recent years (Bonafiglia et al. 2021b), the ideal method for determining changes in VO_{2max} in response to a given exercise intervention – herein referred to as 'VO_{2peak} response' – remains unclear. Concurrently, exercise researchers have debated the need to validate VO_{2max} using a supramaximal verification phase - exercise at a power output higher than the highest power output recorded during an incremental cardiopulmonary test (Rossiter et al. 2006; Astorino et al. 2009; Midgley and Carroll 2009; Bowen et al. 2012; Poole and Jones 2017). To date, seemingly few studies have incorporated supramaximal verification when classifying VO_{2peak} responses to exercise training. While some researchers argue supramaximal verification provides limited additional insight for the added financial cost and participant burden (Murias et al. 2018; Iannetta et al. 2020; Wagner et al. 2021), the impact of including supramaximal verification phases on estimating VO_{2peak} response rate remains an underexplored area of the individual response literature.

Estimating response rate can be achieved by modelling endurance or high-intensity interval training (HIIT) responses for each individual and calculating the proportion that exceed a threshold. This modelling depends on observed changes in outcomes before and after an intervention (Scharhag-Rosenberger et al. 2012; Astorino and Schubert 2014). Observed changes during an intervention incorporate variability attributable to measurement error (instrumentation error and day-to-day biological variability), within-subject variability (chronic changes attributable to behavioral/environmental factors external to the intervention), and variability attributable to exercise training (interindividual differences in trainability) (Hopkins 2000; Hecksteden et al. 2015; Swinton et al. 2018; Bonafiglia et al. 2019). Measurement error can be quantified by calculating typical error (TE) from the variability in a measure when an individual performs repeated tests in the absence of an intervention (Hopkins 2000; Swinton et al. 2018). Additionally, repeat testing and use of the mean of observed values reduces TE by reducing the influence of measurement error (Hopkins 2000; Monach 2012). Of note, using verification phases following an incremental cardiopulmonary test increases the number of repeated VO_{2peak} measurements pre- and post-intervention. It therefore seems reasonable that adding verification phases to VO_{2peak} tests will reduce the influence of measurement error on observed change scores and thereby reduce uncertainty in modelling individual responses and the overall response rate (Swinton et al. 2023).

The purpose of the current analysis was to investigate the influence of verification phases during repeat measurement of VO_{2peak} on TE and the response rate following exercise training. This research provides insight into whether clinicians and researchers should utilize supramaximal verification phases to improve classification of individual response following exercise training.

MATERIALS AND METHODS

Subjects

Thirty-four recreationally active (self-reported < 3 hours of physical activity per week), healthy young males (n=18; 13 of whom were from Del Giudice et al. 2020 and females (n=16) were included in the current study (age, 21.8±2.1 yrs; height, 172.6±9.9 cm; weight, 71.7±4.4 kg,

Table 1 and Figure 1). Inclusion criteria were as follows: between 18-30 years of age, < 3 hours of physical activity per week, no concurrent involvement in exercise training, body mass index < 30 kg/m^2. Exclusion criteria were as follows: cardiovascular or metabolic disease, current oral medication user, and current smoker. Participants were asked to maintain their habitual physical activity levels throughout the study. All participants provided written informed consent before participation, and all experimental procedures were approved by the Health Sciences Human Research Ethics Board at Queen's University (#6021938).

Experimental Design

The current study combined data from one previously published single-group, exercise training study (Del Giudice et al. 2020) and one unpublished randomized controlled trial. Data collection took place between June 20th, 2017 and November 19th, 2017 in the Queen's Muscle Physiology Lab in Kingston, Ontario. All participants completed a familiarization incremental ramp test (i.e. VO_{2peak} test) with a same-day supramaximal verification prior to the start of the experimental protocol to mitigate potential learning effects (Edgett et al. 2018). The term VO_{2peak} is used because attainment of VO_{2max} on an individual basis was not statistically confirmed (Midgley et al. 2008; Midgley and Carroll 2009; Poole and Jones 2017). Following familiarization, participants underwent two incremental ramp tests with supramaximal verification before and after the four-week training period (Figure 2). Participants consumed a standardized meal the night before each VO_{2peak} test (Stauffer's Sauté Sensations [520 kcal; 74 g carbohydrate, 10 g fat, 32 g protein]) and arrived at the laboratory in the morning following a 12-h overnight fast. Upon arrival, participants were fed a standardized breakfast (bagel [181 kcal] with 15 g of cream cheese [44 kcal]). Thirty minutes after consuming breakfast, participants completed a VO_{2peak} test on a motorized treadmill following an incremental test protocol with a supramaximal verification phase. Following baseline testing, participants were randomly allocated using random computer-generated numbers to a 3-day high-intensity interval training (HIIT) group (n=14) or a no-exercise control (n=11). Allocation was not concealed. We also included a non-randomized 4-day HIIT group (n=17; all males), individual VO_{2peak} data from these participants have been published previously (Del Giudice et al. 2020). Experimental testing procedures were the same for all three groups, but a skeletal muscle biopsy was added for all 4-day HIIT group (data not used in the present study), and it was performed 24 hours prior to their first incremental ramp test.

Gas exchange and heart rate were collected throughout the incremental and verification phase testing using the same metabolic cart (Moxus AEI Technologies, Pittsburgh, PA) and heart rate monitor (Polar Team2 Pro, Kempele, Finland), respectively. The highest 30-second average VO_2 was calculated for each test. The incremental test protocol consisted of three minutes of resting data collection (participants were asked to stand on the treadmill and breathe normally) followed by a five-minute warm-up with the treadmill set to 2.5 mph at an incline of 2 and subsequent increases of either incline or speed every two minutes until volitional fatigue (see Supplementary Table 1 for details – also published in Del Giudice et al. 2020). Following the incremental test protocol, participants were provided with a minimum 10 minutes of rest prior to commencing a supramaximal verification phase. The metabolic cart was not re-calibrated in between phases. During the supramaximal verification phase, participants ran until volitional fatigue at a speed that was 0.5 mph faster than the final stage attempted during the incremental test protocol. These protocols were used at pre- and post-testing. Time to fatigue (TTF) was recorded as the duration (seconds) of the incremental test. All exercise was supervised and was performed on the same motorized treadmill (SportsArt, City, USA). Participants were not taking any nutritional supplements during the study. They were also asked to refrain from exercising for 24 hours before, and from alcohol and caffeine for 12 hours before all experimental sessions.

Training Protocol

Participants trained on the same motorized treadmill either three or four times per week for four weeks by the same group of trainer(s). Each training session consisted of four, four-minute intervals at 90–95% HR_{max} with three minutes of active recovery at 70–75% HR_{max} between intervals. If the target HR was not attained two minutes into each four-minute interval, speed or incline were adjusted based on participant preference. Each session began with a 10-minute warm-up at 70–75% HR_{max} and ended with a five-minute cool down at 70–75% HR_{max} (40 minutes total). HR, speed and incline were recorded 30-s before the end of each interval during all training sessions. Speed and incline were adjusted by a trained volunteer during training sessions to ensure appropriate training intensity. Participants nor trainers were blinded.

Statistical Analysis

Modelled responses for all outcomes were calculated by subtracting post-intervention values from pre-intervention values. Final analysis included VO_2 data only from participants who completed a familiarization incremental test with a same-day supramaximal verification, two PRE and two POST incremental tests that each had a same-day supramaximal verification phase.

Two-way mixed ANOVAs (time x group) were used to examine group-level changes in relative VO_{2peak} (data for INCR2+ presented in Figure 3D) and time to fatigue (average of both incremental tests; see INCR2 in Figure 3C) following training. A two-way mixed ANOVA (group x method) was also used to compare change scores for VO_{2peak} and TTF between CTL and Exercise (3-day and 4-day HIIT groups) and across different methods used to determine VO_{2peak} (INCR1; INCR1+, etc.). Any significant interaction or main effects were subsequently analyzed using Bonferroni post-hoc analyses. Corresponding effect sizes were calculated and interpreted using partial eta squared (η_p^2) values (small <0.01; medium=0.059; large >0.14) (Cohen 1988). Within-group effect sizes were calculated using Cohen's d_{av} (small=0.2; medium=0.5; large=0.8) (Cohen 1988; Lakens 2013). Pooled (CTL, 3-day and 4-day HIIT) SDs of change scores were used for VO_{2peak} and TTF Cohen's d_{av} calculations.

Within-subject coefficients of variation (CV) were used to indicate reproducibility (Hopkins 2000). Two-way mixed effects models with absolute agreement were used to examine test-retest reliability (e.g. intraclass correlation coefficients [ICC] with 95% confidence intervals [CI]). ICCs with 95% CIs <0.5, between 0.5 and 0.75, between 0.75 and 0.9, and >0.9 indicated poor, moderate, good, and excellent reliability, respectively (Koo and Li 2016).

Individual response classification was calculated using typical errors (TE) calculated using the standard deviations (SD) of ΔVO_{2peak} from the no-exercise control group (n=11):

(1) $TE = \frac{SD_{CTL}}{\sqrt{2}}$

We used Swinton et al.'s (see supplemental file from Swinton et al. 2018) method to model VO_{2peak} responses using 50% and 95% CIs based on the typical error (TE) of averaged VO_{2peak} from individual change (POST-PRE) in the: 1) first incremental test ("INCR1"; 1.99 mL/kg/min), 2) first incremental test and associated verification phase ("INCR1+"; 1.41 mL/kg/min), 3) average of the two incremental tests ("INCR2"; 1.72 mL/kg/min), and 4) average of the two incremental tests and two verification phases ("INCR2+"; 1.37 mL/kg/min) (see

Figure 3). This approach was chosen as we believe it can help answer a question raised by exercise researchers and practitioners: Does the burden associated with additional tests (addition of second incremental and/or a supramaximal verification phase [SupraV]) improve ability to classify individual response? Responders were identified as participants with 50% or 95% CIs that lay above a zero-based (0 mL/kg/min) or clinically-based response threshold (1.75 mL/kg/min) (Bonafiglia et al. 2018). CIs were calculated using the following equations (Swinton et al. 2023):

(2) $50\%\ CI\ Limits = (\Delta VO_{2max}) \pm (0.67 \times \sqrt{2} \times TE)$

(3) $95\%\ CI\ Limits = (\Delta VO_{2max}) \pm (1.96 \times \sqrt{2} \times TE)$

Following previous work (Montero and Lundby 2017; Bonafiglia et al. 2018; Swinton et al. 2018; Pickering and Kiely 2019; Ross et al. 2019; Bonafiglia et al. 2021b), we have opted against labelling individuals as 'non-responders' when classifying individual response. Instead, we use the term 'uncertain' to reflect individuals who are less likely to have experienced benefit following intervention. A McNemar's test was used to determine whether each method (INCR1, INCR1+, INCR2, INCR2+) elicited similar response rates for group-level changes in VO_{2peak}.

The SD of individual response (SD_{IR}) and the standard error (SE) for each SD_{IR} value was calculated to construct 90% CI's in Microsoft Excel using the methods forwarded by Atkinson and Batterham (Atkinson and Batterham 2015) and Hecksteden et al. (Hecksteden et al. 2018) as we have done previously (Bonafiglia et al. 2019a, 2021a, 2021b). Because participants in the 4-day HIIT group were not randomized (Figure 1), analysis of the 4-day HIIT group violates the assumptions of independence required for the SD_{IR} analysis (Atkinson and Batterham 2015). Therefore, SD_{IR} analyses were performed for participants from the 3-day HIIT and no-exercise control group (see Figure 1). Levene's tests were used to compare interindividual variability (i.e., standard deviation of VO_{2peak} change scores) between groups.

ANOVAs, corresponding effect sizes, and ICCs were performed using SPSS version 25 (IBM Corp., Armonk, N.Y., USA). All other analyses were performed in GraphPad Prism Version 8.0. Outcome assessors were not blinded. Statistical significance was set at $p<0.05$, and all data are presented as mean±SD.

RESULTS

Of the 86 participants screened, 25 and 17 met inclusion criteria for the 3-day/CTL arm and the 4-day training arm, respectively (Figure 1). Eight participants were excluded from final analyses due to incomplete data, and 34 participants completed all physiological testing (CTL: n=12, 3-day HIIT n=9, 4-day HIIT: n=13) (Figure. 1). Each result represents data from these 34 participants. Table 1 presents baseline participant characteristics for all groups.

A significant effect of time ($p=0.0003$, $\eta_p^2=0.31$), group ($p<0.0001$, $\eta_p^2=0.56$) and interaction (group x time) ($p=0.0001$, $\eta_p^2=0.44$) for relative VO_{2peak} (mL/kg/min) was observed when using an average of all incremental and supramaximal verification (i.e. INCR2+) test data. Post-hoc analyses revealed VO_{2peak} increased significantly following 4-day HIIT (+4.12±2.65 mL/kg/min; $p<0.001$, $d_{av}=0.63$), but not following 3-day HIIT (+1.12±1.89; $p>0.05$, $d_{av}=0.12$) nor CTL (-0.78±2.37; $p>0.05$, $d_{av}=-0.12$). Significant ($p<0.001$) effects of time ($\eta_p^2=0.53$), group ($\eta_p^2=0.47$) and interaction ($\eta_p^2=0.46$) were observed for time to fatigue (TTF). Post-hoc analyses revealed that TTF increased following 3-day (+69.5±52.8 s; $p<0.001$, $d_{av}=0.35$) and 4-day (+72.7±33.3; $p<0.001$, $d_{av}=0.71$) but not CTL (-12.1±36.6; $p>0.05$, $d_{av}=-0.09$). Mean changes in VO_{2peak} are reported for CTL, 3-day and 4-day HIIT for each protocol method (INCR1, INCR1+, etc.) in Table 2. A significant (p=0.0003) main effect of group (η_p^2=0.26) was found for mean change in VO_{2peak}. However, no significant (p>0.05) effect of condition (i.e. INCR, INCR+, etc.) or interaction effect (method x group) was observed. Post-hoc analyses revealed that the 4-day HIIT group exhibited significantly greater improvements in VO_{2peak} compared to the 3-day HIIT group across all conditions except one (INCR1) and in all conditions when compared to the CTL group (Table 2).

CVs for incremental test and supramaximal verification VO_{2peak} values were 4.5% and 3.1%, respectively. As presented in Table 3, all ICCs demonstrated good or excellent reliability. Incorporation of supramaximal verification consistently reduced the SD of change and TE in the CTL group and shortened confidence interval widths (see Table 2, Figures 4 and 5). However, Levene's tests revealed variance across all groups was statistically similar (p>0.05). Figure 5 depicts how adding a supramaximal verification (i.e., groups INCR1+ and INCR2+) reduced the 95% confidence intervals around an observed change in VO_{2peak} for a representative subject.

The addition of the supramaximal verification either had no impact or increased the number of participants classified as responders using both ZBT and SWC response thresholds and 50% and 95% CIs (Table 2). Although response rates were increased when incorporating one (24% to 48%, *p*=0.07) or two (28% to 48%, *p*=0.063) supramaximal verification phases (see Table 2 [ZBT-95]; Figure 4). McNemar tests revealed that these changes failed to reach statistical significance for either ZBT or SWC response thresholds using 50% and 95% CIs (all *p*>0.05). Table 2 also presents SD_{IR} for each method. Interestingly, only INCR1+ had a positive SD_{IR}, indicating a lack of evidence for interindividual differences in trainability.

DISCUSSION

This study investigated the influence of supramaximal verification phases during repeat measurement of VO_{2peak} on TE, individual confidence interval widths, and the response rate following HIIT. We tested the hypotheses that incorporating supramaximal verification phases to VO_{2peak} testing would minimize the influence of measurement error on observed change scores, and thus, reduce uncertainty in modelling individual response. Although we failed to observe statistically significant impact of verification phases on SD of change or response rates – likely owing to sample size limitations – our results are generally consistent with our hypotheses. Specifically, our results suggest the addition of supramaximal verification phases narrow confidence interval widths, decrease uncertainty in modelling individual response, and increase the response rate.

Supramaximal verification phase reduces the influence of measurement error

Quantifying measurement error – comprised of instrumentation and biological 'noise' – helps contextualize data from interventions. If measurement error is random, the variability generated over repeated measurements results in observed values are normally distributed around an individual's true value (Hopkins 2000; Swinton et al. 2023). Thus, taking the mean of several measurements at a single time point minimizes measurement error and improves measurement accuracy (Hopkins 2000; Hecksteden et al. 2015; Swinton et al. 2023).

In this study, we incorporated supramaximal verification phases following an incremental test. The increased number of repeated VO_{2peak} measurements pre- and post-intervention reduced the SD of change in the non-exercise control group (Table 2). Although reductions in SD failed to yield statistically significant Levene's tests, our results suggest that verification phases can improve measurement accuracy of VO_{2peak} change scores by reducing measurement error.

Interestingly, the addition of verification phases reduced the SD of change in the non-exercise control group (INCR1=1.99; INCR1+=1.41; INCR2+=1.37) to a greater extent that a separate day incremental test (INCR2=1.72). This is likely due to greater variation in observed values across separate testing sessions (Swinton et al. 2023). Our data appear to suggest that adding same-day supramaximal verification improves measurement accuracy of changes in VO_{2peak} to a greater extent than separate-day testing. However, these results should be confirmed in additional studies utilizing different patient populations and larger samples.

Supramaximal verification reduces uncertainty in individual response classification

'Precision medicine' is a concept gaining popularity throughout various scientific disciplines (König et al. 2017). Precision exercise medicine involves personalizing exercise prescription – including initial prescription and subsequent modification - to maximize individual response (Ross et al. 2019). Although initial prescription should likely be based on protocols known to elicit the largest mean changes for the outcome(s) of interest (Atkinson et al. 2019; Bonafiglia et al. 2021c), subsequent modifications to exercise prescriptions will benefit from more accurate estimates of response. In the current data set, verification phase-associated reductions in the control group's SD reduced CI widths because our CIs were constructed by adding and subtracting a multiple of TE to each observed score (see equation 2 and 3) (Swinton et al. 2023). Smaller CIs reduced the magnitude of observed change required for an individual to be classified as a responder and thus reduced the likelihood of classifying an individual's response as "uncertain". This effect is illustrated for a representative participant in Figure 5. Reducing uncertainty in individual response classification would allow practitioners to make prescription modifications with increased confidence, especially when participants fail to demonstrate a meaningful response in their outcome(s) of interest. Because INCR1+ reduced CI width the most, practitioners monitoring an individual's VO_{2peak} response should consider incorporating a same-day supramaximal verification phase following an incremental test.

Impact of supramaximal verification phase and CI width on group response rate

Mean change and interindividual variability in observed response influence response rates to an exercise intervention (Bonafiglia et al. 2021c). When utilizing individual confidence intervals, the SD of change in the control group also contributes to response rates via its impact on CI width (Schulhauser et al. 2021). Thus, in the current study, response rates were determined by three factors: i) mean change in the exercise group, ii) interindividual variability (SD of change) in the exercise group, and/or iii) SD of change (and TE/CI width) in the control group. Although previous studies have primarily attributed increased response rates in CRF, body composition, exercise performance, and strength outcomes (Walsh et al. 2020; Islam et al. 2020; Bonafiglia et al. 2022b) to changes in mean response, we failed to observed a statistically significant change for mean changes in VO_{2peak} across conditions (Table 2). Despite this, and despite non-significant differences between conditions, incorporating verification phase data increased response rates in all but one condition (ZBT-50, INCR2 to INCR 2+). Response rates doubled (24 to 48%, p=0.07) and nearly doubled (28 to 48%, p=0.063) in the INCR1+ and INCR2+ conditions for ZBT-95, respectively (Figure 4). Because the SD of change in the exercise group was only reduced with the addition of a second day of testing (INCR1=2.92; INCR1+=2.85; INCR2=2.42; INCR2+=2.38) (Table 2) verification phases appear to improve response rate estimates by a combination factors ii and iii above. However, response rates across methods failed to reach statistical significance. The impact of verification phases on response rate certainty will be better understood with larger sample sizes.

Although mean changes in VO_{2peak} were not impacted by determination method (INCR1, INCR1+, etc.), the reduction in variability associated with additional measurements resulted in a progressively larger interaction effect size (INCR1, η^2_p=0.185; INCR1+, η^2_p=0.228; INCR2, η^2_p=0.257; INCR2+, η^2_p=0.300). This result highlights the ability of repeat tests in general – and verification phases specifically – to improve the sensitivity of studies to detect group level differences in VO_{2peak}.

Interestingly, despite reducing the SD of change in both exercise and control groups (albeit to a lesser degree than verification phases), a second incremental test had less robust effects on CI width and response rates. This suggests that researchers and practitioners interested in estimates of response rate would be better served by incorporating a same-day verification phase than a separate day incremental-test.

This study demonstrated the largest response rates (see Table 2) when using a ZBT with 50% CIs. This result corroborates recent findings that classification method heavily influences response rates (Schulhauser et al. 2021). While large response rates may seem desirable, thresholds failing to consider error will inflate response rates compared with more conservative thresholds considering both error and a smallest worthwhile change/minimal clinically important difference (Hecksteden et al. 2018; Schulhauser et al. 2021; Bonafiglia et al. 2021b). While the utility of using more conservative thresholds has been argued elsewhere (Swinton et al. 2018), there is currently no agreement in the literature on the best method(s) for response rate estimation.

In summary, our data demonstrate the ability of supramaximal verification phases to uncertainty and variation in both control (TE/CI width) and exercise comparator arms, suggesting they may be a valuable addition for future studies designed to examine VO_{2peak} response rates. Although interpretation of these results should be tempered by the lack of statistical significance for between-group response rates, we believe verification phases can improve precision of estimates of response rates and should be considered in future work.

LIMITATIONS

Because this is a secondary analysis, we did not appropriately power this study to detect differences between groups or methods. That being said, we did not observe statistically significant differences in SD of change in the control group, and response rates across groups. The studies contributing data to our analyses were also neither designed nor adequately powered to test for any sex-based differences in response to HIIT. Consequently, conducting sex-based analyses in the present study would not yield valid results. The potential influence of sex on training responsiveness to HIIT remains an important area for future research.

Given the relatively small sample size of the current study, future studies with larger sample sizes, a priori power calculations, and risk of bias mitigating practices should test whether incorporating supramaximal verification phases reduce uncertainty in individual response classification (Preobrazenski et al. 2020; Bonafiglia et al. 2022a). Although we used a group-based approach to classify responses, we acknowledge individualized approaches may have greater utility in studies using different populations and/or sample sizes (Swinton et al. 2018; Hecksteden et al. 2018; Bonafiglia et al. 2019b). We also acknowledge different incremental test protocols and populations can influence VO_{2peak} data (Gordon et al. 2012; Beltz et al. 2016). Although we incorporated robust outlier-detection protocols to improve data accuracy (Del Giudice et al. 2020), individual VO_{2peak} response classifications may differ based on testing protocol (e.g., intensity prescribed for verification phase, duration of recovery between phases), modality (ergometer vs. treadmill), and population. Thus, it remains unclear whether verification phases (including their prescribed mode and intensity) impact VO_{2peak} response classifications in athletic, older, unmotivated, or clinical populations following HIIT. Interrogating whether verification phase VO_{2peak} data can reduce uncertainty in response classification across a range of populations is currently unknown, but represents an important future direction.

The additional financial cost and participant burden are drawbacks to incorporating supramaximal testing and repeat testing. Both were not quantified in the current study. However, a supramaximal verification phase can be completed in at least 13 minutes (when including the break in between phases) (Scharhag-Rosenberger et al. 2011; Astorino 2020) and can reduce TE and CI width. These observations may persuade researchers and practitioners to justify adding at least a same-day supramaximal verification test (i.e., INCR1+). Additionally, we suspect researchers employ familiarization tests more frequently than practitioners. Participants in the current study underwent familiarization, which presumably reduced test-retest variability to an unmeasured degree.

CONCLUSION

This report showcases the impact of repeat incremental tests and supramaximal verification phases on measurement error, individual response classification and group response rates for VO_{2peak}. Including a same-day verification phase minimized the impact of random variability (attributable to measurement error and within-subject variability) in the control group, reducing TE and CI width. Although any repeated measurement of VO_{2peak} reduced CI width, adding a same-day verification (INCR1+) reduced CI width the most. We therefore recommend using a same-day verification phase to reduce uncertainty in individual VO_{2peak} response classifications.

Competing interests statement: The authors declare there are no competing interests.

Funding statement: This study was supported by funding provided to B.J.G. from the Natural Sciences and Engineering Research Council of Canada (NSERC; grant no. 402635).

Data availability statement: Data generated or analyzed during this study are available from the corresponding author upon reasonable request. As per APNM requirements, a post-recruitment registration of this project has been uploaded to Open Science Framework (https://osf.io/br5us).

REFERENCES

Astorino, T. A., DE LA Rosa, A. B., Clark, A., & DE Revere, J. L. (2020). Verification Testing to Confirm VO2max Attainment in Inactive Women with Obesity. International journal of exercise science, **13**(4), 1448–1458.

Astorino, T.A., and Schubert, M.M. 2014. Individual Responses to Completion of Short-Term and Chronic Interval Training: A Retrospective Study. PLoS ONE **9**(5): e97638. doi:10.1371/journal.pone.0097638.

Astorino, T.A., White, A.C., and Dalleck, L.C. 2009. Supramaximal testing to confirm attainment of VO2max in sedentary men and women. Int J Sports Med **30**(4): 279–284. doi:10.1055/s-0028-1104588.

Atkinson, G., and Batterham, A.M. 2015. True and false interindividual differences in the physiological response to an intervention: True and false interindividual differences. Exp Physiol **100**(6): 577–588. doi:10.1113/EP085070.

Atkinson, G., Williamson, P., and Batterham, A.M. 2019. Issues in the determination of "responders" and "non-responders" in physiological research. Exp Physiol **104**(8): 1215–1225. doi:10.1113/EP087712.

Beltz, N.M., Gibson, A.L., Janot, J.M., Kravitz, L., Mermier, C.M., and Dalleck, L.C. 2016. Graded Exercise Testing Protocols for the Determination of VO2 max: Historical Perspectives, Progress, and Future Considerations. Journal of Sports Medicine **2016**: 1–12. doi:10.1155/2016/3968393.

Bonafiglia, J.T., Brennan, A.M., Ross, R., and Gurd, B.J. 2019a. An appraisal of the SD IR as an estimate of true individual differences in training responsiveness in parallel-arm exercise randomized controlled trials. Physiological Reports **7**(14): e14163. doi:10.14814/phy2.14163.

Bonafiglia, J.T., Islam, H., Preobrazenski, N., and Gurd, B.J. 2022a. Risk of bias and reporting practices in studies comparing VO2max responses to sprint interval vs. continuous training: A systematic review and meta-analysis. Journal of Sport and Health Science **11**(5): 552–566. doi:10.1016/j.jshs.2021.03.005.

Bonafiglia, J.T., Islam, H., Preobrazenski, N., Ma, A., Deschenes, M., Erlich, A.T., Quadrilatero, J., Hood, D.A., and Gurd, B.J. 2021a. Examining interindividual differences in select muscle and whole-body adaptations to continuous endurance training. Experimental Physiology **106**(11): 2168–2176. doi:10.1113/EP089421.

Bonafiglia, J.T., Nelms, M.W., Preobrazenski, N., LeBlanc, C., Robins, L., Lu, S., Lithopoulos, A., Walsh, J.J., and Gurd, B.J. 2018. Moving beyond threshold-based dichotomous classification to improve the accuracy in classifying non-responders. Physiological Reports **6**(22). doi:10.14814/phy2.13928.

Bonafiglia, J.T., Preobrazenski, N., and Gurd, B.J. 2021b. A Systematic Review Examining the Approaches Used to Estimate Interindividual Differences in Trainability and Classify Individual Responses to Exercise Training. Front. Physiol. **12**: 665044. doi:10.3389/fphys.2021.665044.

Bonafiglia, J.T., Preobrazenski, N., Islam, H., Walsh, J.J., Ross, R., Johannsen, N.M., Martin, C.K., Church, T.S., Slentz, C.A., Ross, L.M., Kraus, W.E., Kenny, G.P., Goldfield, G.S., Prud'homme, D., Sigal, R.J., Earnest, C.P., and Gurd, B.J. 2021c. Exploring Differences in Cardiorespiratory Fitness Response Rates Across Varying Doses of Exercise Training:

A Retrospective Analysis of Eight Randomized Controlled Trials. Sports Med **51**(8): 1785–1797. doi:10.1007/s40279-021-01442-9.
Bonafiglia, J.T., Ross, R., and Gurd, B.J. 2019b. The application of repeated testing and monoexponential regressions to classify individual cardiorespiratory fitness responses to exercise training. Eur J Appl Physiol **119**(4): 889–900. doi:10.1007/s00421-019-04078-w.
Bonafiglia, J.T., Swinton, P.A., Ross, R., Johannsen, N.M., Martin, C.K., Church, T.S., Slentz, C.A., Ross, L.M., Kraus, W.E., Walsh, J.J., Kenny, G.P., Goldfield, G.S., Prud'homme, D., Sigal, R.J., Earnest, C.P., and Gurd, B.J. 2022b. Interindividual Differences in Trainability and Moderators of Cardiorespiratory Fitness, Waist Circumference, and Body Mass Responses: A Large-Scale Individual Participant Data Meta-analysis. Sports Med **52**(12): 2837–2851. doi:10.1007/s40279-022-01725-9.
Bowen, T.S., Cannon, D.T., Begg, G., Baliga, V., Witte, K.K., and Rossiter, H.B. 2012. A novel cardiopulmonary exercise test protocol and criterion to determine maximal oxygen uptake in chronic heart failure. J Appl Physiol (1985) **113**(3): 451–458. doi:10.1152/japplphysiol.01416.2011.
Cohen, J. 1988. Statistical power analysis for the behavioral sciences. *In* 2nd ed. L. Erlbaum Associates, Hillsdale, N.J.
Del Giudice, M., Bonafiglia, J.T., Islam, H., Preobrazenski, N., Amato, A., and Gurd, B.J. 2020a. Investigating the reproducibility of maximal oxygen uptake responses to high-intensity interval training. Journal of Science and Medicine in Sport **23**(1): 94–99. doi:10.1016/j.jsams.2019.09.007.
Edgett, B.A., Bonafiglia, J.T., Raleigh, J.P., Rotundo, M.P., Giles, M.D., Whittall, J.P., and Gurd, B.J. 2018. Reproducibility of peak oxygen consumption and the impact of test variability on classification of individual training responses in young recreationally active adults. Clinical Physiology and Functional Imaging. doi:10.1111/cpf.12459.
Gordon, D., Mehter, M., Gernigon, M., Caddy, O., Keiller, D., and Barnes, R. 2012. The effects of exercise modality on the incidence of plateau at VO2max. Clinical Physiology and Functional Imaging **32**(5): 394–399. doi:10.1111/j.1475-097X.2012.01142.x.
Hecksteden, A., Kraushaar, J., Scharhag-Rosenberger, F., Theisen, D., Senn, S., and Meyer, T. 2015. Individual response to exercise training - a statistical perspective. Journal of Applied Physiology **118**(12): 1450–1459. doi:10.1152/japplphysiol.00714.2014.
Hecksteden, A., Pitsch, W., Rosenberger, F., and Meyer, T. 2018. Repeated testing for the assessment of individual response to exercise training. Journal of Applied Physiology **124**(6): 1567–1579. doi:10.1152/japplphysiol.00896.2017.
Hopkins, W.G. 2000. Measures of reliability in sports medicine and science. doi:10.2165/00007256-200030010-00001.
Iannetta, D., de Almeida Azevedo, R., Ingram, C.P., Keir, D.A., and Murias, J.M. 2020. Evaluating the suitability of supra-PO $_{peak}$ verification trials after ramp-incremental exercise to confirm the attainment of maximum O $_2$ uptake. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **319**(3): R315–R322. doi:10.1152/ajpregu.00126.2020.
Islam, H., Siemens, T.L., Matusiak, J.B.L., Sawula, L., Bonafiglia, J.T., Preobrazenski, N., Jung, M.E., and Gurd, B.J. 2020. Cardiorespiratory fitness and muscular endurance responses immediately and 2 months after a whole-body Tabata or vigorous-intensity continuous

training intervention. Appl Physiol Nutr Metab **45**(6): 650–658. doi:10.1139/apnm-2019-0492.

König, I.R., Fuchs, O., Hansen, G., von Mutius, E., and Kopp, M.V. 2017. What is precision medicine? Eur Respir J **50**(4): 1700391. doi:10.1183/13993003.00391-2017.

Koo, T.K., and Li, M.Y. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine. doi:10.1016/j.jcm.2016.02.012.

Lakens, D. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. **4**. doi:10.3389/fpsyg.2013.00863.

Midgley, A.W., Bentley, D.J., Luttikholt, H., McNaughton, L.R., and Millet, G.P. 2008. Challenging a dogma of exercise physiology: does an incremental exercise test for valid VO 2 max determination really need to last between 8 and 12 minutes? Sports Med **38**(6): 441–447. doi:10.2165/00007256-200838060-00001.

Midgley, A.W., and Carroll, S. 2009. Emergence of the verification phase procedure for confirming 'true'$\dot{V}O_{2max}$. Scandinavian Journal of Medicine & Science in Sports **19**(3): 313–322. doi:10.1111/j.1600-0838.2009.00898.x.

Monach, P.A. 2012. Repeating tests: different roles in research studies and clinical medicine. Biomarkers in Medicine **6**(5): 691–703. doi:10.2217/bmm.12.57.

Montero, D., and Lundby, C. 2017. Refuting the myth of non-response to exercise training: 'non-responders' do respond to higher dose of training: Trainability and exercise dose. J Physiol **595**(11): 3377–3387. doi:10.1113/JP273480.

Murias, J.M., Pogliaghi, S., and Paterson, D.H. 2018. Measurement of a True V˙O2max during a Ramp Incremental Test Is Not Confirmed by a Verification Phase. Front. Physiol. **9**: 143. doi:10.3389/fphys.2018.00143.

Pickering, C., and Kiely, J. 2019. Do Non-Responders to Exercise Exist—and If So, What Should We Do About Them? Sports Med **49**(1): 1–7. doi:10.1007/s40279-018-01041-1.

Poole, D.C., and Jones, A.M. 2017. Measurement of the maximum oxygen uptake $\dot{V}O_{2max}$: $\dot{V}O_{2peak}$ is no longer acceptable. Journal of Applied Physiology **122**(4): 997–1002. doi:10.1152/japplphysiol.01063.2016.

Preobrazenski, N., Islam, H., Drouin, P.J., Bonafiglia, J.T., Tschakovsky, M.E., and Gurd, B.J. 2020. A novel gravity-induced blood flow restriction model augments ACC phosphorylation and *PGC-1α* mRNA in human skeletal muscle following aerobic exercise: a randomized crossover study. Appl. Physiol. Nutr. Metab. **45**(6): 641–649. doi:10.1139/apnm-2019-0641.

Ross, R., Goodpaster, B.H., Koch, L.G., Sarzynski, M.A., Kohrt, W.M., Johannsen, N.M., Skinner, J.S., Castro, A., Irving, B.A., Noland, R.C., Sparks, L.M., Spielmann, G., Day, A.G., Pitsch, W., Hopkins, W.G., and Bouchard, C. 2019. Precision exercise medicine: understanding exercise response variability. Br J Sports Med **53**(18): 1141–1153. doi:10.1136/bjsports-2018-100328.

Rossiter, H.B., Kowalchuk, J.M., and Whipp, B.J. 2006. A test to establish maximum O2 uptake despite no plateau in the O2 uptake response to ramp incremental exercise. J Appl Physiol (1985) **100**(3): 764–770. doi:10.1152/japplphysiol.00932.2005.

Scharhag-Rosenberger, F., Carlsohn, A., Cassel, M., Mayer, F., and Scharhag, J. 2011. How to test maximal oxygen uptake: a study on timing and testing procedure of a supramaximal verification test. Appl. Physiol. Nutr. Metab. **36**(1): 153–160. doi:10.1139/H10-099.

Scharhag-Rosenberger, F., Walitzek, S., Kindermann, W., and Meyer, T. 2012. Differences in adaptations to 1 year of aerobic endurance training: individual patterns of nonresponse: Variability of training responses. Scandinavian Journal of Medicine & Science in Sports **22**(1): 113–118. doi:10.1111/j.1600-0838.2010.01139.x.
Schulhauser, K.T., Bonafiglia, J.T., McKie, G.L., McCarthy, S.F., Islam, H., Townsend, L.K., Grisebach, D., Todd, C., Gurd, B.J., and Hazell, T. 2021. Individual patterns of response to traditional and modified sprint interval training. Journal of Sports Sciences **39**(10): 1077–1087. doi:10.1080/02640414.2020.1857507.
Swinton, P.A., Hemingway, B.S., Gallagher, I.J., and Dolan, E. 2023. Statistical Methods to Reduce the Effects of Measurement Error in Sport and Exercise: A Guide for Practitioners and Applied Researchers.
Swinton, P.A., Hemingway, B.S., Saunders, B., Gualano, B., and Dolan, E. 2018. A Statistical Framework to Interpret Individual Response to Intervention: Paving the way for personalised nutrition and exercise prescription. Frontiers in Nutrition **5**(May): 41. doi:10.3389/FNUT.2018.00041.
Wagner, J., Niemeyer, M., Infanger, D., Hinrichs, T., Guerra, C., Klenk, C., Königstein, K., Cajochen, C., Schmidt-Trucksäss, A., and Knaier, R. 2021. Verification-phase tests show low reliability and add little value in determining $\dot{V}O_2max$ in young trained adults. PLoS ONE **16**(1): e0245306. doi:10.1371/journal.pone.0245306.
Walsh, J.J., Bonafiglia, J.T., Goldfield, G.S., Sigal, R.J., Kenny, G.P., Doucette, S., Hadjiyannakis, S., Alberga, A.S., Prud'homme, D., and Gurd, B.J. 2020. Interindividual variability and individual responses to exercise training in adolescents with obesity. Appl. Physiol. Nutr. Metab. **45**(1): 45–54. doi:10.1139/apnm-2019-0088.

FIGURE CAPTIONS

1. Participant flow diagram. CTL, no-exercise control group; HIIT, high-intensity interval training.
2. Study protocols. Participants completed two incremental tests before (PRE) and after (POST) four weeks of HIIT or a no-exercise control period. See manuscript text for details. HIIT, high-intensity interval training.
3. Illustration of the four methods **(A-D)** used to calculate mean changes in peak oxygen uptake (VO_{2peak}) values from before (PRE) and after (POST) 4 weeks of HIIT in 34 participants. Solid black lines in a step-like formation represent incremental ramp tests, and shaded grey boxes represent supramaximal verifications. **(A)**: INCR1 - Calculated difference (Δ) between the 1^{st} incremental test at PRE and the 1^{st} incremental test at POST. **(B)**: INCR1+ - Calculated difference between the averaged VO_{2peak} from the 1st incremental test and its supramaximal verification at PRE and the averaged VO_{2peak} from the 1st incremental test and its supramaximal verification at POST. **(C)**: INCR2 - Calculated difference between the averaged VO_{2peak} of two incremental tests at PRE and two incremental tests at POST. **(D)**: INCR2+ - Calculated difference between the averaged VO_{2peak} of two incremental tests and their supramaximal verifications at PRE and the averaged VO_{2peak} of two incremental tests and their supramaximal verifications at POST. HIIT, high-intensity interval training.
4. Individual response classification following four methods used to calculate mean changes in peak oxygen uptake (VO_{2peak}). Green circles represent participants whose lower limit of their 95% confidence interval (CI) exceeds zero. Individual responses to VO_{2peak} from the 3-day and 4-day high-intensity interval training groups are ordered from smallest to

largest change according to VO_2 calculation method (A). Methods ‘INCR1+’ and ‘INCR2+’ contain supramaximal verification phases. Visualization of participants are ordered according to INCR1 (A) observed change score.

5. An example of inconsistent individual response classification across four VO_{2peak} calculation methods (data from participant 6). A green dot with 95% confidence intervals represents a positive response using a zero-based threshold (ZBT). CI width for each method is ± 4.59, 2.91, 3.55, 2.82, respectively. Methods ‘INCR1+’ and ‘INCR2+’ contain supramaximal verification phases. For interest, the smallest worthwhile change (SWC) threshold has been graphically displayed.

Table 1. Baseline participant characteristics (n=34).

Participants	All (n=34)	CTL (n=9)	3-day (n=12)	4-day (n=13)
Age (years)	21.8 ± 2.1	21.7 ± 2.5	22.2 ± 1.8	21.5 ± 2.3
Sex (M/F)	(18/16)	(2/7)	(3/9)	(13/0)
Height (cm)	172.6 ± 9.9	164.1 ± 10.1	171.1 ± 7.3	179.8 ± 6.4
Body weight (kg)	71.8 ± 12.7	63.3 ± 11.7	75.0 ± 14.3	74.6 ± 9.7
VO_{2peak} (mL/kg/min)	50.8 ± 9.8	45.0 ± 6.0	45.5 ± 8.9	59.7 ± 5.2 *
TTF (s)	1227 ± 186	1156 ± 126	1199 ± 187	1432 ± 96 *

CTL, no exercise control group; TTF = time to fatigue; values are presented as mean ± standard deviation.
* Significantly different from CTL and 3-day ($p<0.05$).

Table 2. Changes (Δ) in VO_{2peak}, proportion of response and SD_{IR} after various VO_{2peak} calculation methods (n=34).

	INCR1	INCR1+	INCR2	INCR2+
CTL (n=9)				
Mean ΔVO_{2peak} ± *SD*	-0.50 ± 2.82	-0.38 ± 2.00	-0.63 ± 2.43	-0.78 ± 1.94
TE	1.99	1.41	1.72	1.37
3-day HIIT (n=12)				
Mean ΔVO_{2peak} ± *SD*	1.51 ± 2.37	1.44 ± 2.17	1.35 ± 1.80	1.13 ± 1.86
4-day HIIT (n=13)				
Mean ΔVO_{2peak} ± *SD*	3.67 ± 3.19*	4.21 ± 2.94*†	3.77 ± 2.51*†	3.91 ± 2.15*†
CI width (±)				
95%	4.59	2.91	3.55	2.82
50%	1.36	0.97	1.18	0.94
Responders (%)				
ZBT-50	60	72	68	68
ZBT-95	24	48	28	48
SWC-50	48	52	40	48
SWC-95	16	24	12	20
SD_{IR} (90% CI) (n=21)	-1.67 (-3.13 - 2.05)	0.56 (-2.05 – 2.18)	-1.72 (-2.78 – 1.36)	-0.77 (-2.08 – 1.76)

CI, confidence interval; CTL, no exercise control group; HIIT, high-intensity interval training; SD_{IR}, standard deviation of individual response; SWC, smallest worthwhile change; ZBT, zero-based threshold. SD_{IR} was calculated using the 3-day exercise group (n=12) and CTL group (n=9). *Significantly different from CTL (p<0.01), †Significantly different from 3-day HIIT (p<0.05).

Table 3. Intraclass correlation coefficient (ICC) for each calculation method of VO_{2peak} (n=34).

VO_{2peak} calculation method		ICC with 95% CIs		
		Δ INCR1+	Δ INCR2	Δ INCR2+
(A)	Δ INCR1	**0.934** [0.873 to 0.966]	**0.878** [0.770 to 0.937]	**0.822** [0.672 to 0.907]
(B)	Δ INCR1+		**0.855** [0.731 to 0.925]	**0.903** [0.815 to 0.950]
(C)	Δ INCR2			**0.928** [0.861 to 0.963]
(D)	Δ INCR2+			

VO_{2peak}, peak oxygen uptake in mL/kg/min; CI, confidence interval; Δ, POST-PRE difference. Data are means ± SD.

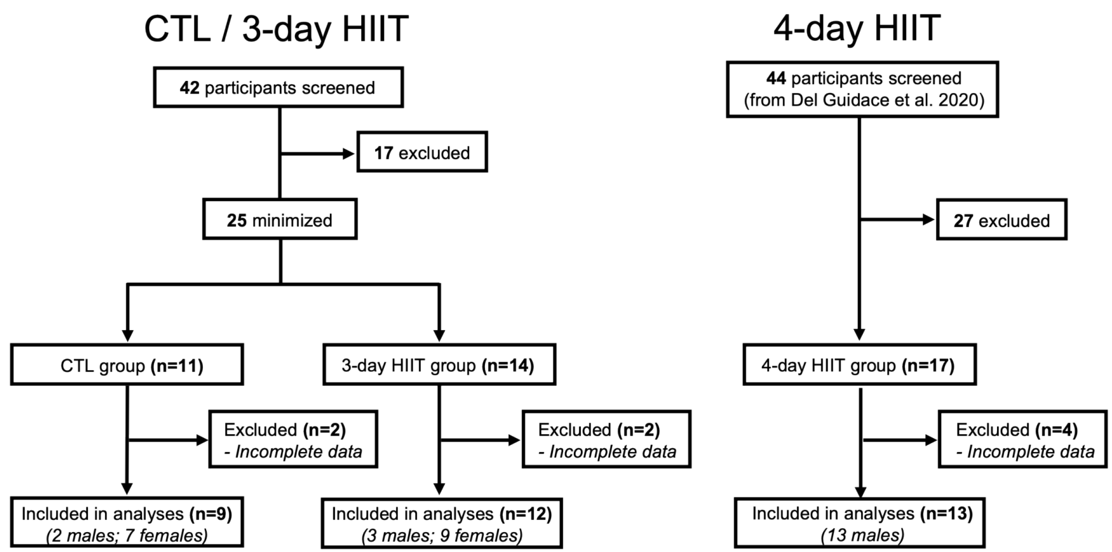


Figure 1. Participant flow diagram. CTL, no-exercise control group; HIIT, high-intensity interval training.

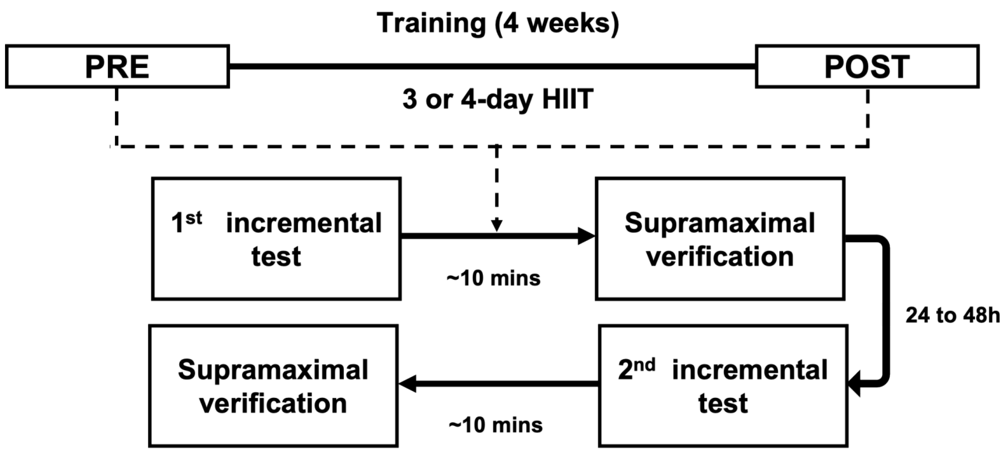
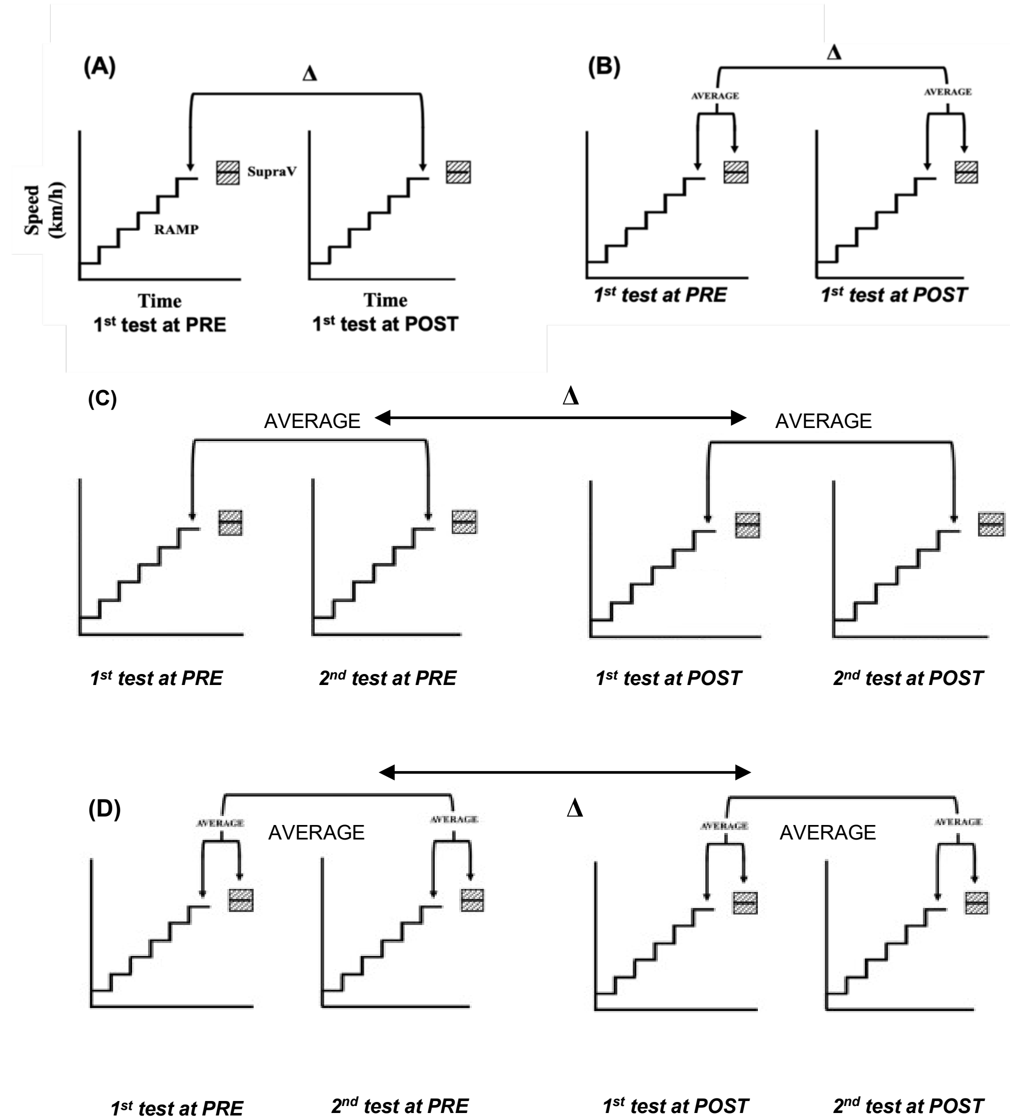


Figure 2. Study protocols. Participants completed two incremental tests before (PRE) and after (POST) four weeks of HIIT or a no-exercise control period. See manuscript text for details. HIIT, high-intensity interval training.

Figure 3. Illustration of the four methods **(A-D)** used to calculate mean changes in peak oxygen uptake (VO_{2peak}) values from before (PRE) and after (POST) 4 weeks of HIIT in 34 participants. Solid black lines in a step-like formation represent incremental ramp tests, and shaded grey boxes represent supramaximal verifications. HIIT, high-intensity interval training.

(A) : INCR1 - Calculated difference (Δ) between the 1st incremental test at PRE and the 1st incremental test at POST.
(B) : INCR1+ - Calculated difference between the averaged VO_{2peak} from the 1st incremental test and its supramaximal verification at PRE and the averaged VO_{2peak} from the 1st incremental test and its supramaximal verification at POST
(C) : INCR2 - Calculated difference between the averaged VO_{2peak} of two incremental tests at PRE and two incremental tests at POST.
(D) : INCR2+ - Calculated difference between the averaged VO_{2peak} of two incremental tests and their supramaximal verifications at PRE and the averaged VO_{2peak} of two incremental tests and their supramaximal verifications at POST.



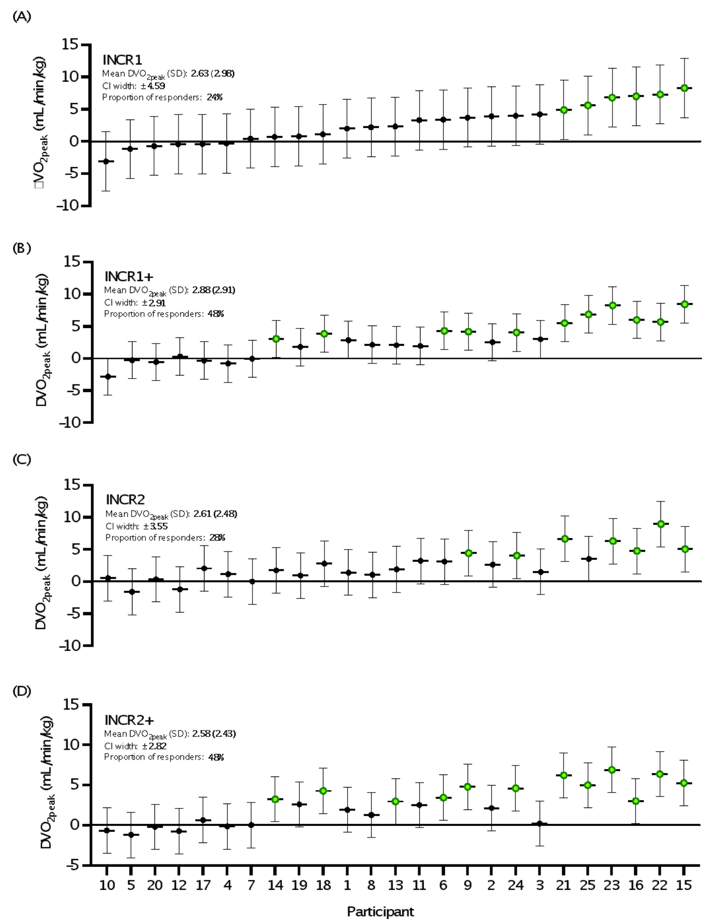


Figure 4. Individual response classification following four methods used to calculate mean changes in peak oxygen uptake (VO_{2peak}). Green circles represent participants whose lower limit of their 95% confidence interval (CI) exceeds zero. Individual responses to VO_{2peak} from the 3-day and 4-day high-intensity interval training groups are ordered from smallest to largest change according to VO_2 calculation method (A). Methods 'INCR1+' and 'INCR2+' contain supramaximal verification phases. Visualization of participants are ordered according to INCR1 (A) observed change score.

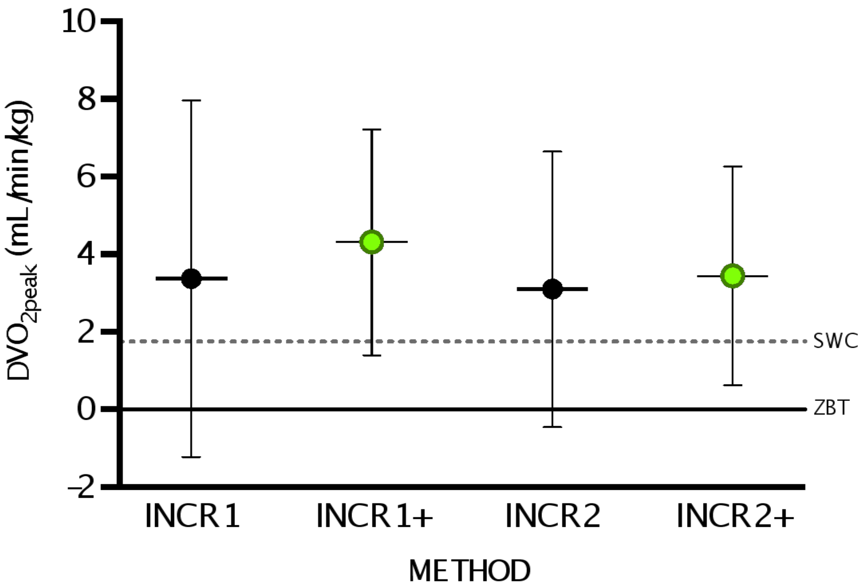


Figure 5. An example of inconsistent individual response classification across four VO $_{2peak}$ calculation methods (data from participant 6). A green dotwith 95% confidence intervals represents a positive response using a zero-based threshold (ZBT). CI width for each method is ± 4.59, 2.91, 3.55, 2.82, respectively. Methods 'INCR1+' and 'INCR2+' contain supramaximal verification phases. For interest, the smallest worthwhile change (SWC) threshold has been graphically displayed.