



Research Protocol

Project Title

Peak cardiac output determined using inert gas rebreathing: A comparison of two exercise protocols

Principal Investigator

Martin Gibala, Ph.D.
Professor
Department of Kinesiology
McMaster University

Student Investigator

Billy Bostad, M.Sc.
Ph.D. Candidate
Department of Kinesiology
McMaster University

Background and rationale

The “gold standard” methods for determining cardiac output (Q) are the direct Fick and thermodilution techniques (1). Both are highly technical and very invasive, requiring arterial catheterization. Q can be estimated non-invasively using inert gas rebreathing (IGR). The IGR method involves taking 5-6 breaths from a closed-circuit rebreathing bag that contains a mixture of oxygen and inert gases. The Innocor (COSMED, Italy) is a popular commercially available device that uses oxygen (94%), an inert blood soluble gas (nitrous oxide, 5%), and an inert blood insoluble gas (sulfur hexafluoride, 1%) (2). Photoacoustic gas analyzers monitor the expired air and measure the disappearance rate of the blood soluble gas relative to the blood insoluble gas over the course of the rebreathing period to estimate Q.

IGR-derived Q using the Innocor correlates well with the direct Fick ($r=0.95$) and thermodilution ($r=0.94$) methods during maximal exercise (3, 4). A recent review comprising 10 clinical studies found that the correlation between IGR and the Fick method was higher than the correlation between the thermodilution and Fick method during submaximal steady state exercise (IGR vs Fick: $r=0.84$ [95% CI 0.74-0.90]; thermodilution vs Fick: $r=0.73$ [95% CI 0.61-0.82]), and that IGR demonstrated good Bland-Altman Limits of Agreement with the Fick method (1).

Moreover, IGR had a higher reliability than the thermodilution method (IGR typical error [TE]=7.2%; thermodilution TE=13.2%) (1). Further, the Fick and thermodilution methods yield greater inter-subject variability estimated by standard deviation during maximal exercise compared to IGR (Fick standard deviation [SD]=3.9 L/min; thermodilution SD=3.8 /min; IGR SD=2.2 L/min) (5). Therefore, IGR provides a noninvasive estimation of peak cardiac output (Q_{peak}) that is accurate and potentially more reliable than invasive measures. The latter may be

owing to the fact that the relative simplicity of the IGR method reduces assessor variability, and it is less susceptible to excessive movement during exercise.

A challenge inherent to the IGR technique for estimating Q_{peak} is the selection of an appropriate exercise protocol and deciding when to initiate the rebreathing procedure. The rebreathing requires a measurement period of ~ 10 s during “maximal” exercise or as close to this point as possible, while ensuring the participant can maintain the exercise load throughout the entire rebreathing period. There is no consensus regarding the optimal exercise protocol to measure IGR-derived Q_{peak} . One commonly-used protocol involves an incremental (step) exercise test with rebreathing initiated when the participant reaches a heart rate (HR) that is within 5 bpm of a previously determined peak value (the Q_{step} protocol) (6). Iterations of this protocol involve initiating rebreathing when the participant self-reports that they are ~ 30 s from exhaustion, or when the workload corresponds to the peak power output from a previously completed step test to exhaustion (7), as opposed to when a pre-determined HR is reached. The TE of step test determinations of Q_{peak} have been reported to be $\sim 7\%$ (1, 8).

We recently developed an exercise protocol for the IGR-derived estimation of Q_{peak} (the Q_{CL} protocol) (9) that was modelled after tests of $VO_{2\text{peak}}$ that involve a “verification phase” (10, 11). It involved a ramp exercise test to elicit $VO_{2\text{peak}}$, followed by a constant load exercise phase after a brief recovery interval, and the TE was 4.7% (9). The constant load phase was performed at $\sim 90\%$ of the workload that elicited $VO_{2\text{peak}}$, with rebreathing to estimate Q_{peak} initiated after 2 minutes. This protocol was developed based on previous research studies that found that constant load exercise performed at $>85\%$ W_{peak} can elicit comparable VO_2 values, and HR values that are within $\sim 5\%$ of those obtained at the end of a ramp $VO_{2\text{peak}}$ test (9–12). The use of a constant load phase to measure Q_{peak} allows for the measurement of this parameter during the same test as used



to determine $VO_{2\text{peak}}$, in contrast to the Q_{step} protocol which must be performed on a different day. This may be convenient for training studies (i.e., reducing the need for two tests). No study has compared Q_{peak} between the Q_{CL} and Q_{step} protocols, and no study has assessed the reliability of the Q_{CL} and Q_{step} protocols for determining Q_{peak} within the same group of participants.

Objectives

Primary objective

To test whether the Q_{peak} produced from the Q_{CL} protocol is not meaningfully different from the Q_{step} protocol using a non-inferiority design.

Secondary objectives

(1) to determine the reliability of the Q_{CL} and Q_{step} protocols, (2) to determine whether the VO_2 and HR obtained from the two experimental protocols are not different from the $VO_{2\text{peak}}$ and HR_{peak} obtained from the baseline $VO_{2\text{peak}}$ test, and (3) to determine whether the VO_2 measured by the Innocor is lower than the metabolic cart.

Hypothesis

We hypothesize that the Q_{CL} protocol will produce a Q_{peak} that is not different (by a margin of noninferiority of 0.5 L/min) compared to the Q_{step} protocol. A Q_{peak} of 0.5 L/min was the TE in a previous study in our laboratory (9), and is therefore deemed to be the minimum meaningful difference for determinations of Q_{peak} .



Methods

Participants

Participants will be recruited from the McMaster community and surrounding area through printed posters, web-based advertisements, and word of mouth.

Inclusion criteria:

- Adults aged 18-35 years.
- Meeting the Canadian 24-hour Movement Guidelines for Adults for aerobic physical activity, i.e. ≥ 150 minutes of moderate to vigorous activity per week (13). This will be assessed using the Canadian Society for Exercise Physiology Get Active Questionnaire (GAQ) (14).

Exclusion criteria:

- Diagnosis of any cardiovascular, respiratory, or metabolic disease that would preclude participation from a clinical standpoint, as determined by answering ‘yes’ to any of the questions on the first page of the GAQ.

Details of the experimental protocol, purpose and potential risks of participation will be explained to all participants and written informed consent will be obtained prior to participation in the study. This study has been approved by the Hamilton Integrated Research Ethics Board (Project # 13339).

Sample size determination

A calculation performed using G*Power (v 3.1.9.2) for a two-tailed dependent means (matched pairs) t-test estimated that a sample size of 34 was required to detect a medium effect size ($d_z=0.5$) with 80% power at an alpha level of 0.05 (Figure 1). A medium effect size was deemed

reasonable based on determinations made in G*Power using our hypothesized minimum meaningful difference (0.5 L/min) and typical means, standard deviations and correlations determined in our laboratory and reported in the literature for Q_{peak} (5, 9, 15).

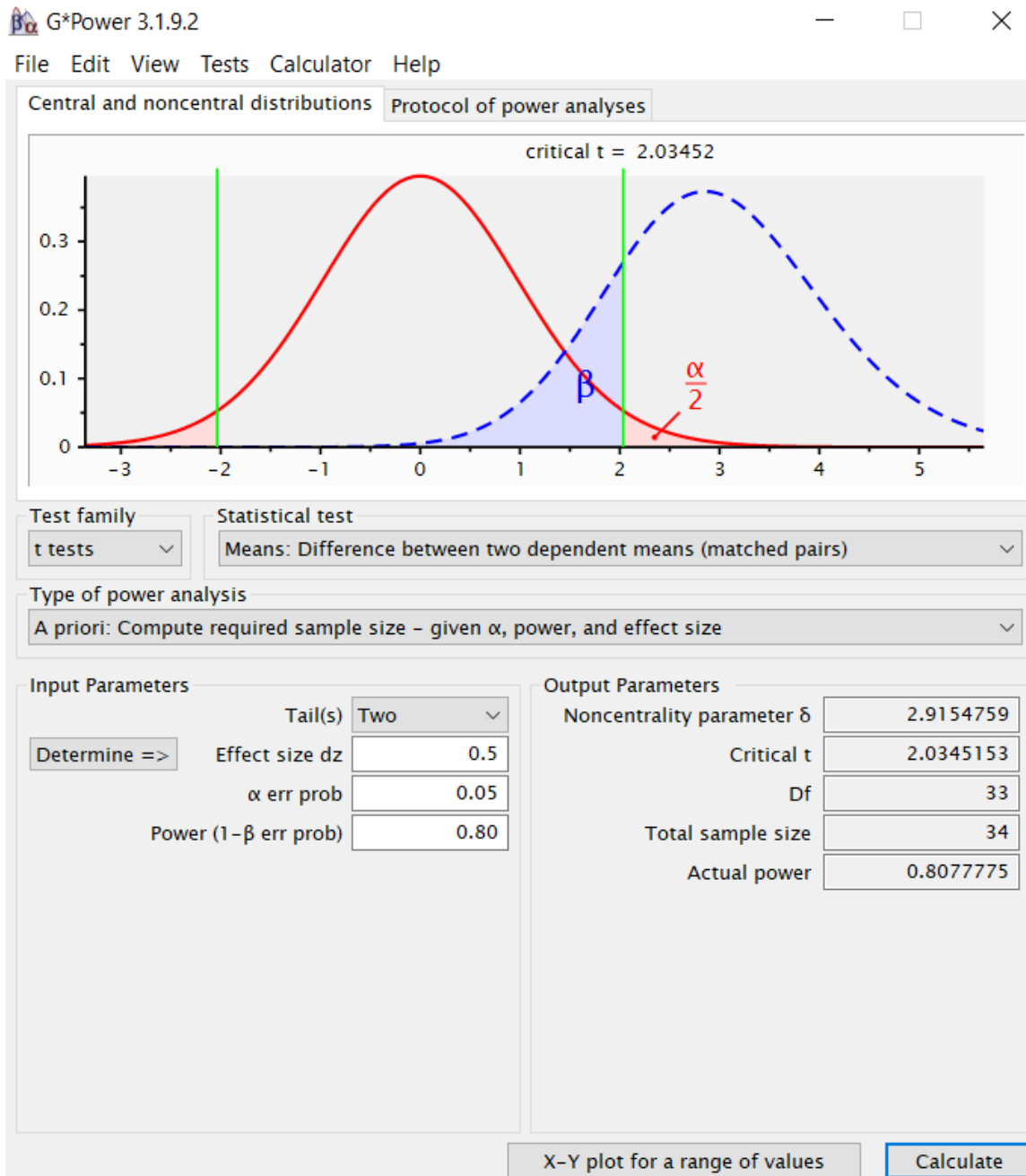


Figure 1: G*Power sample size output.

Experimental procedures

This study will use a randomized within-subject crossover design. All data collection will take place in the Human Performance Laboratory at McMaster University in Hamilton, Ontario, Canada. Following preliminary screening and recruitment into the study, participants will attend the laboratory on 9 separate occasions, each separated by ≥ 48 hours (Figure 2). The first visit will involve the determination of $VO_{2\text{peak}}$ (Quark CPET metabolic cart, COSMED, Italy), HR_{peak} (Polar A3, Finland), and W_{peak} (Lode Excalibur Sport V2.0, Groningen, The Netherlands). On the second visit, participants will be familiarized with the inert gas rebreathing procedure and measurement of Q will be performed at rest and during a brief exercise test consisting of a 3-minute warm up at 50 W followed by an immediate increase to 80% of peak workload (W_{peak}) with rebreathing initiated after 2 minutes at 80% W_{peak} . The third visit will involve a verification test with the same measures as the first test. The highest values for $VO_{2\text{peak}}$, HR_{peak} and W_{peak} will be used to determine appropriate parameters and for comparative purposes with the subsequent main experimental trials. There will then be a 1-week break before beginning the experimental trials to reduce the likelihood of a testing-induced training effect. The experimental trials will consist of 6 different visits to the laboratory using a randomized repeated measures design (Experimental trials, Figure 2). Participants will perform the Q_{CL} protocol on 2 separate occasions and the Q_{step} protocol on 2 separate occasions. Q_{peak} (Innocor, COSMED, Italy) and HR_{peak} (Polar) will be assessed during each of these visits. On an additional 2 visits, participants will perform the Q_{CL} (one visit) and Q_{step} protocols (one visit) with VO_2 measured in place of Q_{peak} using the same metabolic cart as the one used for the baseline $VO_{2\text{peak}}$ test (Quark CPET metabolic cart, COSMED, Italy). These tests are referred to as $VO_{2\text{CL}}$ and $VO_{2\text{step}}$ in Figure 2. All 6 experimental trials will be randomized for each participant. The first Q_{CL} and the first Q_{step}

test will be used to compare Q_{peak} *between* protocols, and the values obtained from the first vs the second Q_{CL} and Q_{step} tests will be used to determine the TE *within* each protocol. The VO_{2peak} from the VO_{2CL} and VO_{2step} protocols will be compared to the VO_{2peak} from the baseline VO_{2peak} test.

Participants will be instructed to maintain their habitual diet between testing sessions and abstain from strenuous exercise and alcohol consumption for a minimum of 24 hours before testing. A 24-hour dietary log will be completed before every experimental trial testing session.

Participants will be instructed to arrive to the laboratory hydrated. Testing sessions will take place at the same time of day ± 1 hour for each participant. All exercise protocols will use a cycle ergometer (Lode Excalibur Sport V2.0, Groningen, The Netherlands).

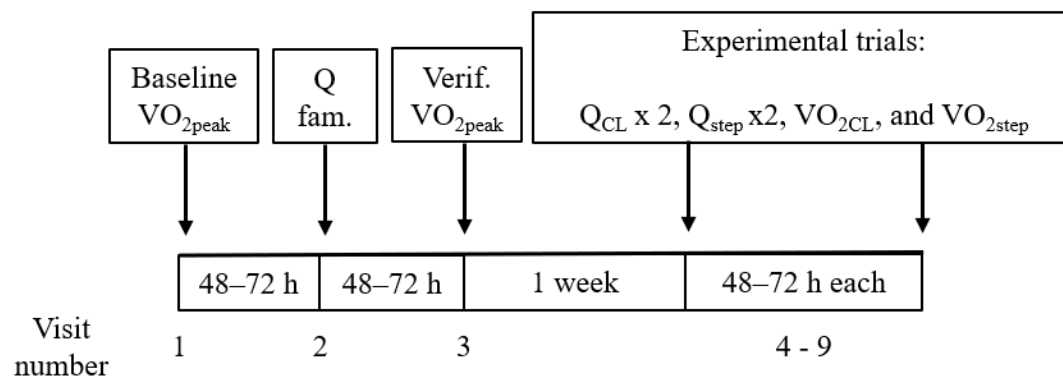


Figure 2: Schematic of study design. Q fam., cardiac output familiarization; Verif, Verification; h, hours. Each set of 2 experimental trials will be performed in random order.

Randomization

The order of the 6 experimental trials will be determined using simple randomization in Microsoft Excel. For each participant, a random, 15 decimal place number between 0 and 1 will be assigned to each experimental trial using the “=rand()” formula. The experimental trials will

be sorted using the “sort largest to smallest” function to determine the order in which the trials will be performed. The random sequence generation will be performed by a study investigator who is not involved in participant enrollment and will inform the other investigators of the order of experimental trials for each participant 24 hours before their first session. Due to the nature of the experiments, we are unable to blind the participants or researchers to the order of experimental tests; however, we will limit performance bias by blinding participants to the study hypotheses.

Primary outcome measure

Q_{peak} determined using the first Q_{CL} and Q_{step} protocols.

Secondary outcome measures

Reproducibility of Q_{peak} for each of the two IGR protocols based on the 1st and 2nd experimental trial for each test.

HR_{peak} determined during to the two IGR protocols.

$VO_{2\text{peak}}$, HR_{peak} , and W_{peak} determined during the baseline and verification $VO_{2\text{peak}}$ tests.

$VO_{2\text{peak}}$ between the $VO_{2\text{peak}}$ test, the $VO_{2\text{CL}}$ protocol and the $VO_{2\text{step}}$ protocol.

$VO_{2\text{peak}}$ between the Q_{CL} vs $VO_{2\text{CL}}$ tests and the Q_{step} vs $VO_{2\text{step}}$ tests.

$VO_{2\text{peak}}$ tests

For the determination of $VO_{2\text{peak}}$, participants will perform a progressive exercise test to maximal voluntary exertion using an electromagnetically braked cycle ergometer (Lode Excalibur Sport V2.0, Groningen, The Netherlands). Following a 3 min warm-up at a fixed workload of 50 W, a ramp protocol will be applied with a linear workload increase of 1 W every 2 s (30 W/min). The

precise protocol will be standardized for a given participant and intend to achieve a ramp duration of ~10 min (range: 5-15 min). Pedaling cadence will be chosen by the participant and will be required to be ≥ 60 rpm. A 3-min recovery phase will be performed at 50 W. Gas exchange and ventilatory variables will be continuously determined using a metabolic cart (Quark CPET metabolic cart, COSMED, Italy). Data will be averaged over 10-s intervals and $VO_{2\text{peak}}$ will be defined as the highest 30-s average over three consecutive intervals. Heart rate will be recorded continuously (Polar A3, Finland) and HR_{peak} will be defined as the highest 2-s average. Data-based cutoffs for age-stratified secondary exhaustion criteria based on peak respiratory exchange ratio and age-predicted maximal heart rate (16) will be used to verify that the test involved maximal effort. The highest $VO_{2\text{peak}}$, HR_{peak} , and W_{peak} between the baseline and verification tests will be regarded as the ‘true’ peak.

Q_{CL} protocol

This protocol involves a ramp exercise test to exhaustion, followed by a constant work rate test that is initiated after 10 minutes of active recovery (Figure 3). The ramp phase uses the same protocol as the $VO_{2\text{peak}}$ test (described above). The constant work rate phase will begin with a 1-minute warmup at 50 W. The warmup will be followed by an immediate increase to an intensity equivalent to 90% of the W_{peak} elicited during the baseline ramp $VO_{2\text{peak}}$ test (90% W_{peak}). HR (Polar A3, Finland) and VO_2 (Innocor, COSMED, Italy) will be monitored continuously throughout the test. The rebreathing procedure to measure Q_{peak} (Innocor, COSMED, Italy) will be initiated after 2 minutes of cycling at 90% W_{peak} .

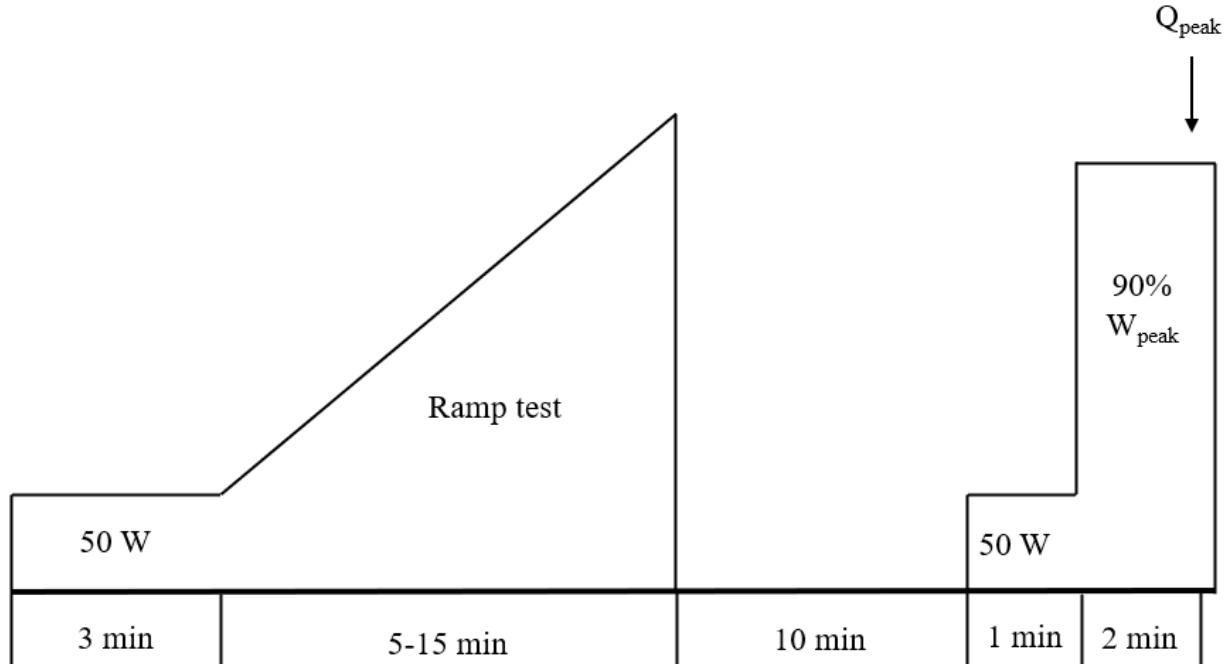


Figure 3: Q_{CL} protocol.

Q_{step} protocol

This test is modelled after an incremental exercise protocol that has commonly been used to assess Q_{peak} (6, 15, 17) including to assess potential changes in Q_{peak} after an exercise intervention (15). The test will begin with a 3-minute warm-up at 50 W and workload will subsequently be increased by 30 W every 1 minute (Figure 4). HR (Polar A3, Finland) and VO_2 (Innocor, COSMED, Italy) will be monitored continuously throughout the test. The rebreathing procedure to measure Q_{peak} (Innocor, COSMED, Italy) will be initiated when the participant reaches a HR that is within 5 bpm of their previously determined HR_{peak} obtained from baseline VO_{2peak} testing (visit #1 or 3 from Figure 2). W_{peak} will be calculated as $W_{peak} = W_{completed} + 30 \times (t/60)$ where $W_{completed}$ is the last completed workload and t is the time in s maintained during the final workload.

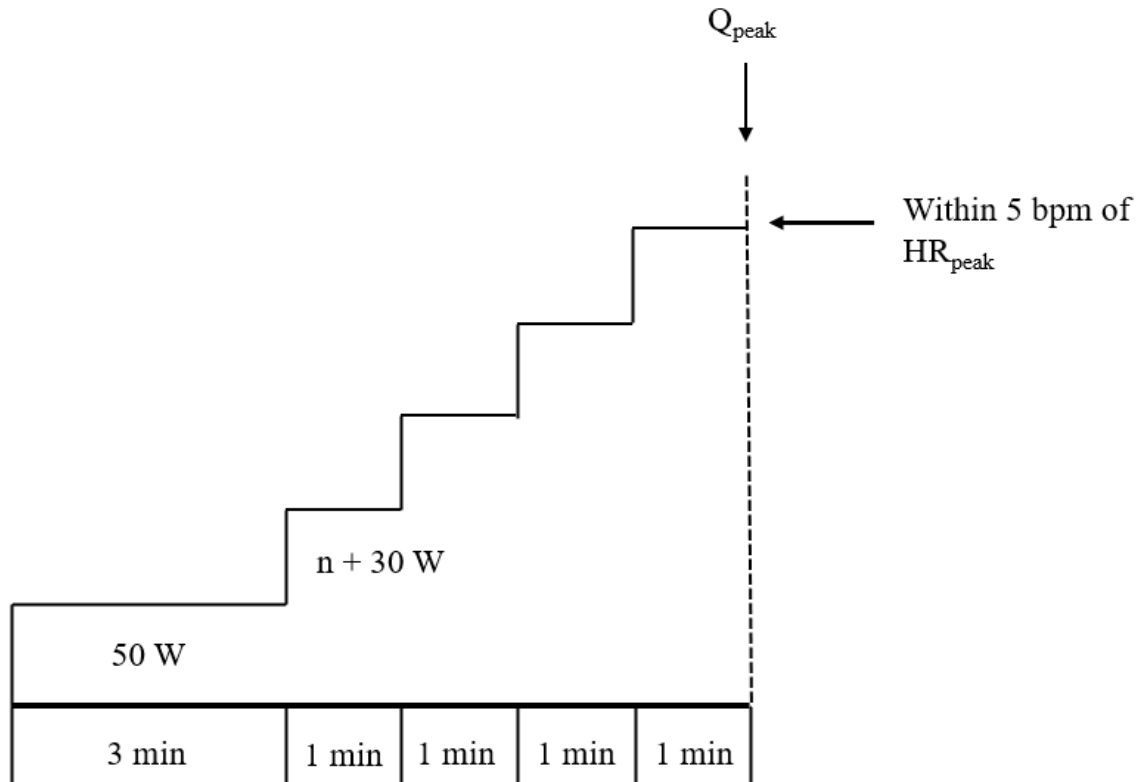


Figure 4: Q_{step} protocol.

VO_{2CL} protocol

The VO_{2CL} protocol will be identical to the Q_{CL} protocol (Figure 3), except VO_2 will be monitored continuously using the same metabolic cart as the $VO_{2\text{peak}}$ test (Quark CPET metabolic cart, COSMED, Italy), and Q_{peak} will not be assessed. The constant load $90\% W_{\text{peak}}$ phase will be 2 min and 15 s in duration; the final 15 s is added to account for the approximate time it would take for the IGR to occur.

VO_{2step} protocol

The VO_{2step} protocol will be identical to the Q_{step} protocol (Figure 4), except VO₂ will be monitored continuously using the same metabolic cart as the VO_{2peak} test (Quark CPET metabolic cart, COSMED, Italy), and Q_{peak} will not be assessed. Participants will continue cycling for an additional 15 s after reaching a HR that is within 5 bpm of HR_{peak} to account for the approximate time it would take for the IGR to occur.

Statistical analysis

A two-tailed paired samples t-test will be performed to compare each of Q_{peak}, HR_{peak}, and W_{peak} between the two experimental conditions (Q_{CL} vs Q_{step}). A two-tailed paired samples t-test will be performed to compare each of VO_{2peak}, HR_{peak}, and W_{peak} between the baseline and verification VO_{2peak} tests, to compare Q_{peak} between the first and second Q_{CL} and Q_{step} tests, and to compare VO_{2peak} between the VO_{2peak} test and each of VO_{2CL} and VO_{2step}. Significance will be set to p<0.05. Effect sizes will be reported as Cohen's d_z and 95% confidence intervals (CI) will also be determined. Q_{peak} estimated by the Q_{CL} protocol will be considered noninferior to the Q_{step} protocol if the 95% CIs for the change in Q_{peak} fall within the margin of noninferiority (i.e., 0.5 L/min) (18, 19).

Reliability of each experimental protocol and of the VO_{2peak} tests will be calculated as the TE using the following formula (20):

$$TE = (SD_{diff} \div \sqrt{2} \div grand\ mean) \times 100$$

Where SD_{diff} is the standard deviation of the difference scores between the 2 measurements, and the grand mean is the mean of all measurements included in the analysis. The TE will be expressed as a percentage.



Perspectives and significance

IGR provides a noninvasive method for determining Q_{peak} , and assessing potential changes elicited by acute and chronic exercise interventions. The proposed work will systematically compare two methods for determining IGR-derived Q_{peak} , to assess potential differences between protocols as well as reliability of each method. This research will provide novel comparative data and inform decisions by researchers regarding choice of Q_{peak} protocol in future studies.

Strategies to reduce bias

Selection bias

The order of experimental trials will be randomized for each participant using Microsoft Excel by an investigator not involved in participant enrollment. This investigator will inform the other investigators of the trial order after baseline testing, and 24 hours before the first experimental trial for each participant. The investigator performing the randomization will play a secondary role in data collection.

Performance and detection bias

Given the distinct nature of each experiment, it is not possible to blind investigators or participants to the order of experimental trials. However, we will limit performance bias by blinding participants to the study hypotheses.

Reporting bias

The current protocol, which clearly states the primary and secondary outcomes and hypotheses, will be published in an open access repository prior to participant recruitment.

Summary of analysis plan

| <u>Question</u> | <u>Hypothesis</u> | <u>Data collection plan</u> | <u>Analysis plan</u> | <u>Interpretation</u> |
|---|--|---|--|--|
| Is Q_{peak} determined by Q_{CL} noninferior to Q_{step} ? | Q_{CL} is noninferior to Q_{step} . | Measure Q_{peak} following Q_{CL} and Q_{step} using a randomized within-subject design. | Calculate 95% CIs for the mean difference in Q_{peak} between Q_{CL} and Q_{step} . | Q_{CL} is noninferior to Q_{step} if the lower bound of the CI is within 0.5 L/min. |
| What is the reproducibility of the Q_{CL} and Q_{step} protocols? | Reproducibility will be similar between Q_{CL} and Q_{step} . | Measure Q_{peak} during 2 Q_{CL} tests and 2 Q_{step} tests in a randomized order. | Calculate the TE between 1) the 2 Q_{CL} tests, and 2) the 2 Q_{step} tests. | Result will provide an indication of Q_{peak} reproducibility. |
| Is the HR_{peak} different between Q_{CL} and Q_{step} ? | HR_{peak} will not be significantly different between Q_{CL} and Q_{step} . | Measure HR_{peak} during Q_{CL} and Q_{step} using a randomized within-subject design. | Perform a paired t-test for HR_{peak} between Q_{CL} and Q_{step} . | If HR_{peak} is not significantly different between Q_{CL} and Q_{step} , it suggests participants are similarly close to “max” values during both protocols. |
| What is the reproducibility of the $VO_{2\text{peak}}$ test? | The reproducibility for the $VO_{2\text{peak}}$ tests will be approximately 5%. | Measure $VO_{2\text{peak}}$ during the baseline and verification $VO_{2\text{peak}}$ tests. | Calculate the TE of the baseline and verification $VO_{2\text{peak}}$ tests. | Result will provide an indication of $VO_{2\text{peak}}$ test reproducibility. |
| Is the $VO_{2\text{peak}}$ measured from the Innocor noninferior to the Quark metabolic cart? | $VO_{2\text{peak}}$ from the Innocor will be inferior to the Quark metabolic cart. | Measure $VO_{2\text{peak}}$ during identical tests using the Innocor and the Quark metabolic cart (i.e., Q_{CL} vs $VO_{2\text{CL}}$ and Q_{step} vs $VO_{2\text{step}}$). | Calculate 95% CIs for the mean difference in $VO_{2\text{peak}}$ between Q_{CL} vs $VO_{2\text{CL}}$ and Q_{step} vs $VO_{2\text{step}}$. | Q_{CL} is noninferior to $VO_{2\text{CL}}$, and Q_{step} is noninferior to $VO_{2\text{step}}$ if the lower bound of the CI is within a margin of |

| | | | | |
|---|--|--|--|--|
| | | | | noninferiority of 0.2 L/min. |
| Do the experimental tests elicit VO_{2peak} that is noninferior to the VO_{2peak} test? | VO_{2peak} during the experimental tests will be noninferior to the VO_{2peak} test. | Measure VO_{2peak} during the VO_{2CL} , VO_{2step} , and baseline VO_{2peak} tests. | Calculate 95% CIs for the mean difference in VO_{2peak} between VO_{2CL} vs the baseline VO_{2peak} test and VO_{2step} vs the baseline VO_{2peak} test. | VO_{2CL} and VO_{2step} are noninferior to the baseline VO_{2peak} test if the lower bound of their respective CIs are within a margin of noninferiority of 0.2 L/min. |

Acknowledgments

We thank our colleagues who have provided input on the development of this proposal including those whose substantive contributions will be recognized with authorship on the resulting manuscript.

References

1. Hardin EA, Stoller D, Lawley J, et al. Noninvasive Assessment of Cardiac Output: Accuracy and Precision of the Closed-Circuit Acetylene Rebreathing Technique for Cardiac Output Measurement. *J Am Heart Assoc.* 2020;9(17):e015794.
2. Ayotte B, Seymour J, McIlroy M. A new method for measurement of cardiac output with nitrous oxide. *J Appl Physiol.* 1970;28(6):863–6.
3. Christensen P, Clemensen P, Andersen PK, Henneberg SW. Thermodilution versus inert gas rebreathing for estimation of effective pulmonary blood flow. *Crit Care Med.* 2000;28(1):51–6.
4. Agostoni P, Cattadori G, Apostolo A, et al. Noninvasive measurement of cardiac output during exercise by inert gas rebreathing technique: a new tool for heart failure evaluation. *J Am Coll Cardiol.* 2005;46(9):1779–81.
5. Jarvis SS, Levine BD, Prisk GK, et al. Simultaneous determination of the accuracy and precision of closed-circuit cardiac output rebreathing techniques. *J Appl Physiol.* 2007;103(3):867–74.
6. Siebenmann C, Rasmussen P, Sørensen H, et al. Cardiac output during exercise: A comparison of four methods. *Scand J Med Sci Sport.* 2015;25(1):e20–7.
7. Schierbauer J, Hoffmeister T, Treff G, Wachsmuth NB, Schmidt WFJ. Effect of Exercise-Induced Reductions in Blood Volume on Cardiac Output and Oxygen Transport Capacity. *Front Physiol.* 2021;12:679232.
8. Fontana P, Boutellier U, Toigo M. Reliability of measurements with Innocor TM during exercise. *Int J Sports Med.* 2009;30:747–53.
9. Bostad W. Twelve weeks of sprint interval training increases peak cardiac output in previously untrained individuals - accepted for publication. *Eur J Appl Physiol.* 2021
10. Iannetta D, de Almeida Azevedo R, Ingram CP, Keir DA, Murias JM. Evaluating the suitability of supra-PO_{peak} verification trials after ramp-incremental exercise to confirm the attainment of maximum O₂ uptake. *Am J Physiol Regul Integr Comp Physiol.* 2020;319(3):R315–22.
11. Rossiter HB, Kowalchuk JM, Whipp BJ. A test to establish maximum O₂ uptake despite no plateau in the O₂ uptake response to ramp incremental exercise. *J Appl Physiol.* 2006;100(3):764–70.
12. Macpherson R, Hazell T, Olver T, Paterson D, Lemon R. Run sprint interval training improves aerobic performance but not maximal cardiac output. *Med Sci Sports Exerc.* 2011;43(1):115–22.

13. Ross R, Chaput JP, Giangregorio LM, et al. Canadian 24-Hour Movement Guidelines for Adults aged 18-64 years and Adults aged 65 years or older: an integration of physical activity, sedentary behaviour, and sleep. *Appl Physiol Nutr Metab.* 2020;45(10):S57–102.
14. CSEP. Get Active Questionnaire. Canadian Society for Exercise Physiology, Ottawa, ON, Canada. Available from <http://www.csep.ca/home>; 2017.
15. Montero D, Cathomen A, Jacobs RA, et al. Haematological rather than skeletal muscle adaptations contribute to the increase in peak oxygen uptake induced by moderate endurance training. *J Physiol.* 2015;593(20):4677–88.
16. Wagner J, Niemeyer MAX, Infanger D, et al. New Data-based Cutoffs for Maximal Exercise Criteria across the Lifespan. *Med Sci Sports Exerc.* 2020;52(9):1915–23.
17. Bonne TC, Doucende G, Flück D, et al. Phlebotomy eliminates the maximal cardiac output response to six weeks of exercise training. *Am J Physiol - Regul Integr Comp Physiol.* 2014;306(10):752–60.
18. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. Reporting of noninferiority and equivalence randomized trials: Extension of the CONSORT 2010 statement. *JAMA.* 2012;308(24):2594–604.
19. Schumi J, Wittes J. Through the looking glass: understanding non-inferiority. *Trials.* 2011;12(106).
20. Hopkins WG. Measures of reliability in sports medicine and science. *Sport Med.* 2000;30(1):1–15.