Published in final edited form as:
Am J Cardiol. 2008 September 15; 102(6): 712–717. doi:10.1016/j.amjcard.2008.04.047.

Reproducibility of Peak Oxygen Uptake and Other Cardiopulmonary Exercise Testing Parameters in Patients with Heart Failure (From the Heart Failure and A Controlled Trial Investigating Outcomes of exercise traiNing)

Daniel R. Bensimhon, MD[a], **Eric S. Leifer, PhD**[b], **Stephen J. Ellis, PhD**[c], **Jerome L. Fleg, MD**[b], **Steven J. Keteyian, PhD**[d], **Ileana L. Piña, MD**[e], **Dalane W. Kitzman, MD**[f], **Robert S. McKelvie, MD, PhD**[g], **William E. Kraus, MD**[h], **Daniel E. Forman, MD**[i], **Andrew J. Kao, MD**[j], **David J. Whellan, MD**[k], **Christopher M. O'Connor, MD**[c], **Stuart D. Russell, MD**[l], and **the HF-ACTION Trial Investigators**

a*Lebauer Cardiovascular Research Foundation, Greensboro, NC*

b*National Institutes of Health, Bethesda, MD*

c*Duke Clinical Research Institute, Durham, NC*

d*Division of Cardiovascular Medicine, Henry Ford Hospital, Detroit, MI*

e*Division of Cardiology, Case Western Reserve VA Medical Center, Cleveland, OH*

f*Section of Cardiology, Department of Medicine, Wake Forest University School of Medicine, Winston-Salem, NC*

g*Division of Cardiology, Hamilton Health Sciences*

h*Division of Cardiovascular Medicine, Duke University Medical Center, Durham, NC*

i*Cardiovascular Division, Brigham and Women's Hospital and Boston VA Medical Center, Boston MA*

j*Cardiovascular Consultants, Mid-America Heart Institute, Kansas City, MO*

k*Division of Cardiology, Jefferson Medical College, Philadelphia, PA*

l*Division of Cardiology, Johns Hopkins Hospital, Baltimore, MD*

Abstract

Peak oxygen uptake (pVO_2) is an important parameter in assessing the functional capacity and prognosis of patients with heart failure. In heart failure trials, the change in pVO_2 is often used to assess the effectiveness of an intervention. However, the within-patient variability of pVO_2 on serial testing may limit its usefulness. This study was designed to evaluate the within-patient variability of pVO_2 over two baseline cardiopulmonary exercise tests. As a sub study of the HF-ACTION (Heart Failure and A Controlled Trial Investigating Outcomes of exercise traiNing) trial, 398 subjects (73%

Correspondence to, Stuart D. Russell, MD, 600 North Wolfe Street, Carnegie 568, Baltimore, MD 21287, Phone: 410-955-5708, Fax: 410-955-3478, E-mail: srusse14@jhmi.edu.

male, 27% female, mean age 59 years) with HF and left ventricular ejection fraction ≤ 35% underwent two baseline cardiopulmonary exercise tests within 14 days. Mean pVO_2 was unchanged from test 1 to test 2 (15.16 ± 4.97 vs. 15.18 ± 4.97 mL/kg/min; p=0.78). However, the mean within-subject absolute change was 1.3 mL/kg/min (10th, 90^{th} percentiles = 0.1, 3.0 mL/kg/min), with 46% of subjects increasing and 48% decreasing on the second test. Other parameters, including the ventilation-to-carbon dioxide production slope and VO_2 at ventilatory threshold, also demonstrated significant within-subject variation with minimal mean differences between tests. In conclusion, peak oxygen consumption demonstrates substantial within-subject variability in heart failure subjects and should be taken into account in clinical applications. However, on repeat baseline cardiopulmonary exercise tests, there appears to be no familiarization effect for pVO_2 in heart failure subjects and, hence, in multicenter trials there is no need to perform more than one baseline cardiopulmonary exercise test.

Keywords

heart failure; cardiopulmonary exercise testing; test reproducibility

HF-ACTION (Heart Failure and A Controlled Trial Investigating Outcomes of exercise traiNing) is a National Heart, Lung, and Blood Institute-funded multi-center randomized controlled trial designed to study the effects of exercise training on morbidity and mortality in 2,331 subjects with left ventricular dysfunction (ejection fraction ≤ 35%) and NYHA Class II-IV symptoms.[1] As part of the trial, each subject underwent cardiopulmonary exercise testing at baseline, 3, 12 and 24 months to assess changes in peak oxygen consumption (pVO_2) and other physiologic parameters with exercise training. The protocol also pre-specified that the first five subjects at each clinical site (as well as the first 100 subjects overall) would perform two baseline tests within one week of each other to assess the reproducibility of testing at the sites and the within-subject variability of repeat testing. This paper reports the results of this reproducibility analysis.

Methods

The enrollment criteria and study design for the HF-ACTION trial have been previously published.[1] Subjects had a left ventricular ejection fraction ≤ 35% (due to ischemic or nonischemic cardiomyopathy) and were on stable doses (i.e., the same dose for at least six weeks prior to enrollment) of optimal drug therapy including an angiotensin converting enzyme inhibitor or angiotensin receptor blocker and a beta-blocker unless a contraindication was present. Of the subjects enrolled in HF-ACTION, 94% were on an angiotensin converting enzyme inhibitor or angiotensin receptor blocker and 95% were taking beta-blockers.

Of 401 subjects performing 2 baseline tests, 350 subjects (87%) used a modified Naughton treadmill protocol and 48 subjects (12%) used a 10 Watt/minute incremental cycle ergometry protocol at a total of 83 different clinical sites. Three (1%) subjects performed one of their baseline tests on the treadmill and the other on the cycle ergometer and were therefore excluded from the analysis. Ninety two percent of the subjects performed 2 tests within 7 days of each other; all subjects had both tests within 14 days of each other. Sites were instructed to perform the testing at the same time of day and at a constant interval from the last dose of beta-blocker. Subjects were instructed not to exercise between tests and there were no changes in medications between the tests. All tests were interpreted by the HF-ACTION Cardiopulmonary Exercise Core Laboratory. All sites performed routine calibrations of gas concentrations and flow prior to each exercise test. Peak VO_2 was defined as the highest VO_2 value for a given 15- or 20-second interval within the last 90 seconds of exercise or the first 30 seconds of recovery. Peak respiratory exchange ratio (RER) was defined as the highest recorded value for a 15- or 20-

second interval that occurred during the last 90 seconds of exercise. The selected peak RER had to correspond to appropriate VO_2 values and progress in a physiologic fashion from preceding ratios. The VO_2 at ventilatory threshold (VT) was independently determined by 2 readers using the V-slope method.[2] If the independently selected VO_2 at VT values were within 150 mL/min of each other, the values were averaged; if the chosen values differed by > 150 mL/min, then a third reader was employed, and the VO_2 at VT was chosen by an average of the 2 values that were in closest agreement. If the 2 closest readings still differed by > 150 mL/min or the VT could not be determined by the V slope method, the VO_2 at VT was considered "indeterminate." The Ve/VCO_2 slope was determined by measuring the slope across the entire course of exercise.[3] Prior to being approved for subject enrollment, each site had to have its cardiopulmonary exercise testing laboratory validated by the HF-ACTION Cardiopulmonary Exercise Core Laboratory. To be validated, all sites underwent baseline education in the proper conduct of an exercise test. This included instruction on encouraging the participants to achieve a maximal effort and how to perform a proper metabolic exercise test. Additionally, each site exercised 2 normal controls on the protocol used in the HF-ACTION trial. Comparisons of VO_2 at baseline and at the first 3 workloads were made to a group of normal controls. The site controls' VO_2 had to show a physiologic increase over exercise and be within 2 standard deviations of the Core Lab controls' VO_2 at each workload in order to be validated. All sites passed this validation step prior to enrolling any subjects and underwent repeat validation testing every 3 to 6 months during the trial.

Means are expressed using either ± standard deviation or with 10^{th} and 90^{th} percentiles. For all of the variables examined, within-subject variability from test 1 to test 2 was quantified by the within-subject absolute change (i.e., either increase or decrease from test 1 to test 2). The paired t-test was used to test for a statistical difference between the test 1 and test 2 values. Bland-Altman plots were used to visually assess whether the magnitude of within-subject variability in key variables varied with the magnitude of the measurements.[4] The unpaired t-test was used to test whether within-subject pVO_2 variability differed with respect to baseline dichotomous variables. Linear regression and Pearson's correlation coefficient were used to correlate baseline continuous variables and within-subject pVO_2 variability. In multivariate linear regression modeling, partial and multiple Pearson's correlation coefficients were computed to quantify the explained within-subject pVO_2 variability. In the multivariate model, a variable was included if its p-value in the univariate linear model was less than 0.2; no variable selection algorithm was performed after this initial selection. For each of the variables, the coefficient of variation (CV) was defined as the (within-person standard deviation/within-person mean) × 100%.

To examine whether within-subject variability depended on clinical site experience with cardiopulmonary exercise testing, sites were prespecified as either "more experienced" or "less experienced." An experienced site was defined as one for which: 1) a physician or exercise physiologist conducted the tests; 2) testing personnel had formal training in cardiopulmonary exercise testing; 3) the site performed at least 50 tests in the prior year; 4) the site passed all validation tests on their first attempt; and 5) the site did not have frequent HF-ACTION Core Laboratory queries regarding its testing.

All statistical hypothesis tests were 2-sided and were performed at a significance level of 0.05. Analyses were conducted with SAS software, version 8.2 (SAS Institute Inc., Cary, NC) and Splus software, version 7.0 (Insightful Corp., Seattle, WA).

Results

Table 1 gives the baseline demographics and peak VO_2 for the 398 subjects in the study. On the first cardiopulmonary exercise test, the mean pVO_2 was 15.2±5.0 mL/kg/min.

As shown in Table 2, mean exercise time increased significantly from test 1 to test 2. In contrast, mean pVO_2 was virtually identical on the two tests. Peak VO_2 was nearly as likely to increase from test 1 to test 2 (46% of the subjects) as it was to decrease (48% of the subjects). Moreover, pVO_2 had substantial within-subject variability, averaging 1.3 mL/kg/min over the two tests with a 90^{th} percentile change of 3.0 mL/kg/min. As shown in Figure 1A, within-subject pVO_2 variability was quite similar for subjects with lower versus higher pVO_2.

The VO_2 at VT also showed nearly identical mean values on test 1 and 2, but also a substantial amount of within-subject variability. Sixty-seven subjects (17%) had an indeterminate VT on at least one test. Figure 1B shows the within-subject variability of VO_2 at VT to be similar for subjects with lower versus higher values. In contrast, for the Ve/VCO_2 slope, there appears to be more variability for subjects with higher slopes, although there are fewer of these subjects (Figure 1C).

There was a very small, but statistically significant, difference in the peak RER between the two tests. As shown in Figure 1D and similar to pVO_2, the range of within-subject variability was quite similar for subjects with lower versus higher peak values. Although the change in the peak RER correlated positively with the change in pVO_2 between the two tests (Figure 2), changes in peak RER and pVO_2 between the two tests trended in opposite directions in 37% of subjects.

Within-patient variability in cardiopulmonary exercise parameters was similar across demographic subgroups (Table 3). Subjects at the more experienced sites had higher mean pVO_2 on the first test than subjects at the less experienced sites. This was largely because the Class II subjects at the more experienced sites had a higher mean pVO_2 on test 1 than those at the less experienced sites (17.1 ± 4.7 vs. 15.5 ± 5.0, $p=0.016$). Subject demographics did not differ significantly between the less- and more-experienced sites.

Univariate and multivariate analyses were performed to study the relationship between demographic and test 1 variables (since in clinical practice, a patient will typically undergo only one cardiopulmonary exercise test) with the change in pVO_2 from test 1 to test 2. As shown in Table 4a for categorical variables and Table 4b for continuous variables, pVO_2 on test 1 was the most significant predictor of the change in pVO_2 from test 1 to test 2, consistent with a regression to the mean effect. Additionally, VO_2 at VT on test 1 was statistically significant in univariate analyses due to its high correlation with pVO_2 on test 1 ($r = 0.83$). Other clinical and test variables including gender, NYHA class, race, and the Ve/VCO_2 slope on test 1 did not correlate with pVO_2 change from test 1 to test 2. Heart failure etiology was marginally statistically significant, with nonischemics more likely to increase their pVO_2 on test 2 and ischemics more likely to decrease pVO_2 on test 2. In the multivariate model (Table 5) for predictors of change in pVO_2 from test 1 to test 2, pVO_2 on test 1 was the most significant predictor. VO_2 at VT was not included in the multivariate model due to its high correlation with pVO_2.

Discussion

As a prespecified sub study of the HF-ACTION trial, we evaluated the reproducibility of pVO_2 and other important cardiopulmonary exercise parameters in 398 HF subjects who underwent two tests within 14 days of each other to assess the need for repeat baseline testing in all subjects. The major findings are as follows: 1) there was significant within-subject variability in pVO_2 between the two tests (average CV 6.6%) but the mean is the same between the two tests; 2) similar variability was seen in other important cardiopulmonary exercise variables, including parameters that are considered to be effort-independent such as Ve/VCO_2 slope (average CV 5.0%) and the VO_2 at VT (average CV 7.8%); 3) pVO_2 was as likely

to decline on the second of two baseline tests as it was to increase, suggesting there was no familiarity effect as is commonly seen with exercise time; and 4) the within-subject variability in pVO_2 was not well-explained by baseline ejection fraction, subject demographics, heart failure class or site experience with cardiopulmonary exercise testing. Thus, the within-subject variability appeared to be due primarily to intrinsic biological factors including, but not limited to, daily fluctuations in the subject's hemodynamic and heart failure status.

Numerous studies have shown that pVO_2 is a key factor in assessing the prognosis of heart failure subjects and gauging their suitability for advanced therapies such as heart transplantation.[5–8] Peak VO_2 has also been commonly used as a clinical endpoint in heart failure therapeutic trials. However, prior work has shown that pVO_2 can vary significantly over serial tests[5, 9–11] and some have suggested that more than one test should be performed at baseline for clinical trial purposes.[11] Previously, Elborn et al. performed three consecutive treadmill tests separated by two weeks on 30 subjects with systolic heart failure due to ischemic cardiomyopathy.[11] The mean pVO_2 improved significantly from the first to the second test (14.1 vs. 14.9 mL/kg/min, $p<0.005$) with no difference in pVO_2 between the second and the third test (14.9 vs. 14.8 mL/kg/min, p=NS). The average within-subject CV for the three tests was 6%. The authors concluded that a single baseline test was not sufficient for measuring the cardiopulmonary response to an intervention and suggested the performance of at least two tests for clinical research applications. In contrast, Russell et al. performed a series of five baseline maximal treadmill tests (three with gas exchange; 1^{st}, 3^{rd}, 4^{th}) on 81 men and women with symptomatic HF and found that while there was a significant improvement in exercise time between test 1 and test 3 (419±140 vs. 470±131 seconds; $p<0.05$), there was no significant change in pVO_2 (1119±376 vs. 1105±346 vs. 1123±400 mL/min; p = NS).[12] The authors concluded that a single baseline test was sufficient to measure pVO_2 for the evaluation of therapy or assessment of prognosis. Similarly, Marburger et al. in an older population of 9 patients found a change in pVO2 of 63 ml/min over serial tests and concluded that only one test was necessary.[13] Finally, Meyer et al. exercised 11 patients with severe heart failure and found a difference of 57 ml/min between the two studies.[14]

In the present study, the overall group means for pVO_2 on test 1 and test 2 were essentially identical as were the percentage of subjects whose pVO_2 decreased or increased on test 2. As seen previously, the mean exercise time increased from test 1 to test 2 with 63% of subjects increasing on test 2 and only 33% decreasing.[10–12, 15] However, despite the similar pVO_2 means, there was significant within-subject pVO_2 variability from test 1 to test 2. The average coefficient of variation was 6.6% and this within-subject variability was largely effort-independent. Although the change in the peak RER, an objective measure of subject effort, correlated positively with the change in pVO_2 between the two tests, in 37% of the subjects, the test 1-to-test 2 change in peak RER and pVO_2, respectively, trended in opposite directions. Somewhat surprisingly, effort-independent variables including Ve/VCO_2 slope (average coefficient of variability = 5.0%) and VO_2 at VT (average coefficient of variability = 7.8%) also showed substantial within-subject variability. This suggests that the majority of the variation is related to intrinsic subject factors such as daily hemodynamic and volume status fluctuations.

Our study had several limitations. First, although sites were instructed to perform both baseline tests at the same time of day, we did not exclude subjects from analysis who had different testing times. Likewise, although sites were instructed to conduct the test between three and ten hours after subjects took their dose of beta blocker, we did not control for the timing between the medications and the test. Third, most studies were done on a treadmill and these findings may not apply to cycle ergometry testing. Finally, although sites were instructed on how to encourage subjects during the test, we were unable to assure that subjects received the same level of encouragement on both baseline tests. However, the similar RER and Borg rates of

perceived exertion scores on tests 1 and 2 suggest that encouragement did not change between tests.

Acknowledgments

This trial is supported by the National Institutes of Health grant # 5U01HL063747

References

1. Whellan DJ, O'Connor CM, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, Leifer ES, Kraus WE, Kitzman DW, Blumenthal JA, Rendall DS, Houston-Miller N, Fleg JL, Schulman KA, Piña IL. HF-ACTION Trial Investigators. Heart failure and a controlled trial investigating outcomes of exercise training (HF-ACTION): design and rationale. Am Heart J 2007;153:201–211. [PubMed: 17239677]
2. Beaver WL, Wasserman K, Whipp WJ. A new method for detecting anaerobic threshold by gas exchange. J Appl Physiol 1986;60:2020–2027. [PubMed: 3087938]
3. Arena R, Myers J, Aslam SS, Varughese EB, Peberdy MA. Technical considerations related to the minute ventilation/carbon dioxide output slope in patients with heart failure. Chest 2003;124:720–727. [PubMed: 12907564]
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310. [PubMed: 2868172]
5. ATS/ACCP. Statement on cardiopulmonary exercise testing. Am J Respir Crit Care Med 2003;167(2): 211–277. [PubMed: 12524257]
6. Corra U, Mezzani A, Bosimini E, Giannuzzi P. Cardiopulmonary exercise testing and prognosis in chronic heart failure. Chest 2004;126:942–950. [PubMed: 15364777]
7. Mancini DM, Eisen H, Kussmaul W, Mull R, Edmunds LH, Wilson JR. Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. Circulation 1991;83:778–786. [PubMed: 1999029]
8. Mehra MR, Kobashigawa J, Starling R, Russell S, Uber PA, Parameshwar J, Mohacsi P, Augustine S, Aaronson K, Barr M. Listing criteria for heart transplantation: International Society for Heart and Lung Transplantation guidelines for the care of cardiac transplant candidates—2006. J Heart Lung Transplant 2006;25:1024–1042. [PubMed: 16962464]
9. Skinner JS, Wilmore KM, Jaskolska A, Jaskolski A, Daw EW, Rice T, Gagnon J, Leon AS, Wilmore JH, Rao DC, Bouchard C. Reproducibility of maximal exercise test data in the HERITAGE family study. Med Sci Sports Exerc 1999;31:1623–1628. [PubMed: 10589867]
10. Sullivan M, Genter F, Savvides M, Roberts M, Myers J, Froelicher V. The reproducibility of hemodynamic, electrocardiographic, and gas exchange data during treadmill exercise in patients with stable angina pectoris. Chest 1984;86:375–382. [PubMed: 6467998]
11. Elborn JS, Stanford CF, Nicholls DP. Reproducibility of cardiopulmonary parameters during exercise in patients with chronic cardiac failure. The need for a preliminary test. Eur Heart J 1990;11:75–81. [PubMed: 2106439]
12. Russell SD, McNeer FR, Beere P, Logan LJ, Higginbotham MB. Improvement in the mechanical efficiency of walking: an explanation for the “placebo effect” seen during repeated exercise testing of patients with heart failure. Am Heart J 1998;135:107–114. [PubMed: 9453529]
13. Marburger CT, Brubaker PH, Pollock WE, Morgan TM, Kitzman DW. Reproducibility of cardiopulmonary exercise testing in elderly patients with congestive heart failure. Am J Cardiol 1998;82:905–909. [PubMed: 9781977]
14. Meyer K, Westbrook S, Schwaibold M, Hajric R, Peters K, Roskamm H. Short-term reproducibility of cardiopulmonary measurements during exercise testing in patients with severe chronic heart failure. Am Heart J 1997;134:20–26. [PubMed: 9266779]
15. Froelicher VF, Brammell H, Davis G, Noguera I, Stewart A, Lancaster MC. A comparison of the reproducibility and physiologic response to three maximal treadmill exercise protocols. Chest 1974;65:512–517. [PubMed: 4826023]

Figure 1A

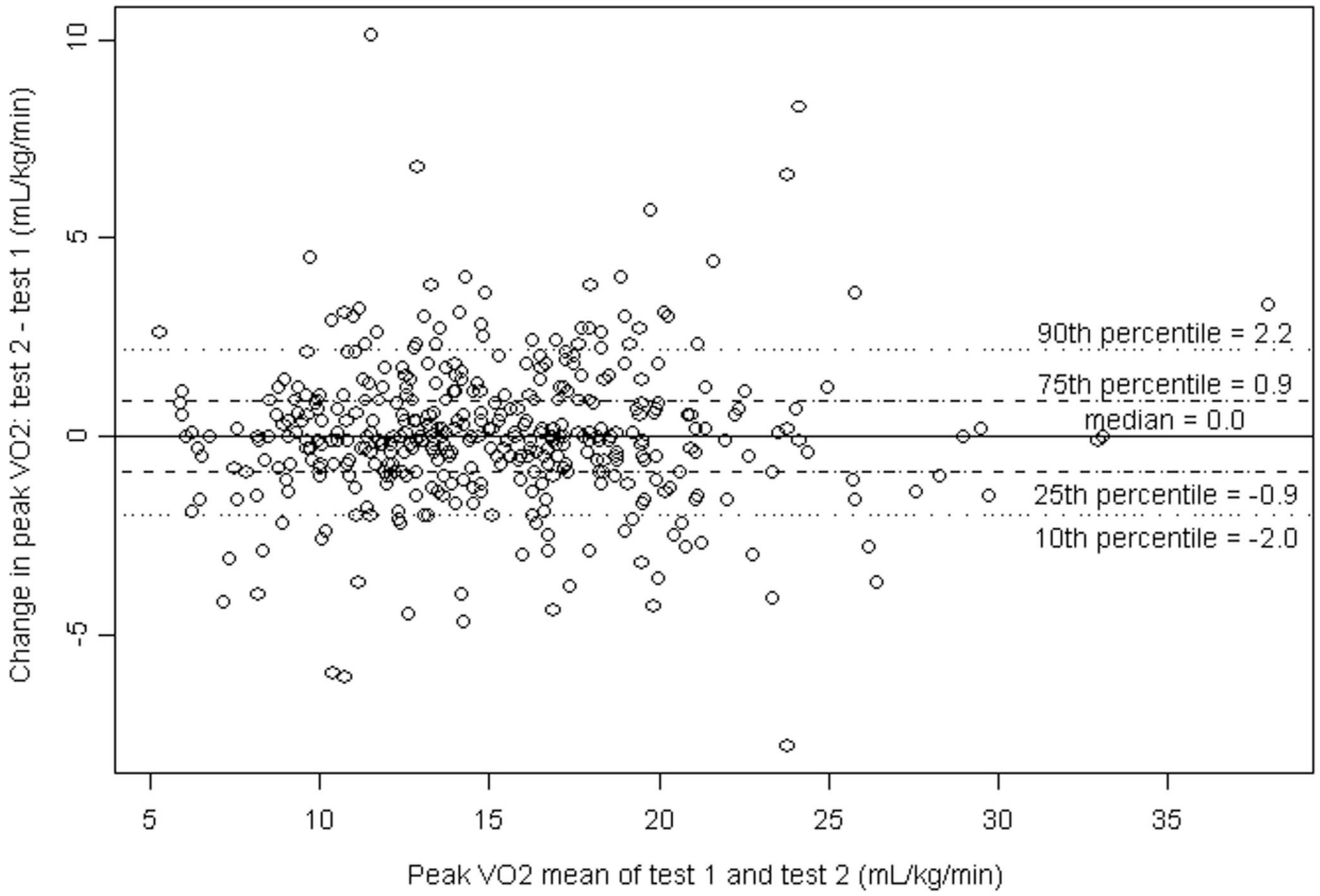


Figure 1B

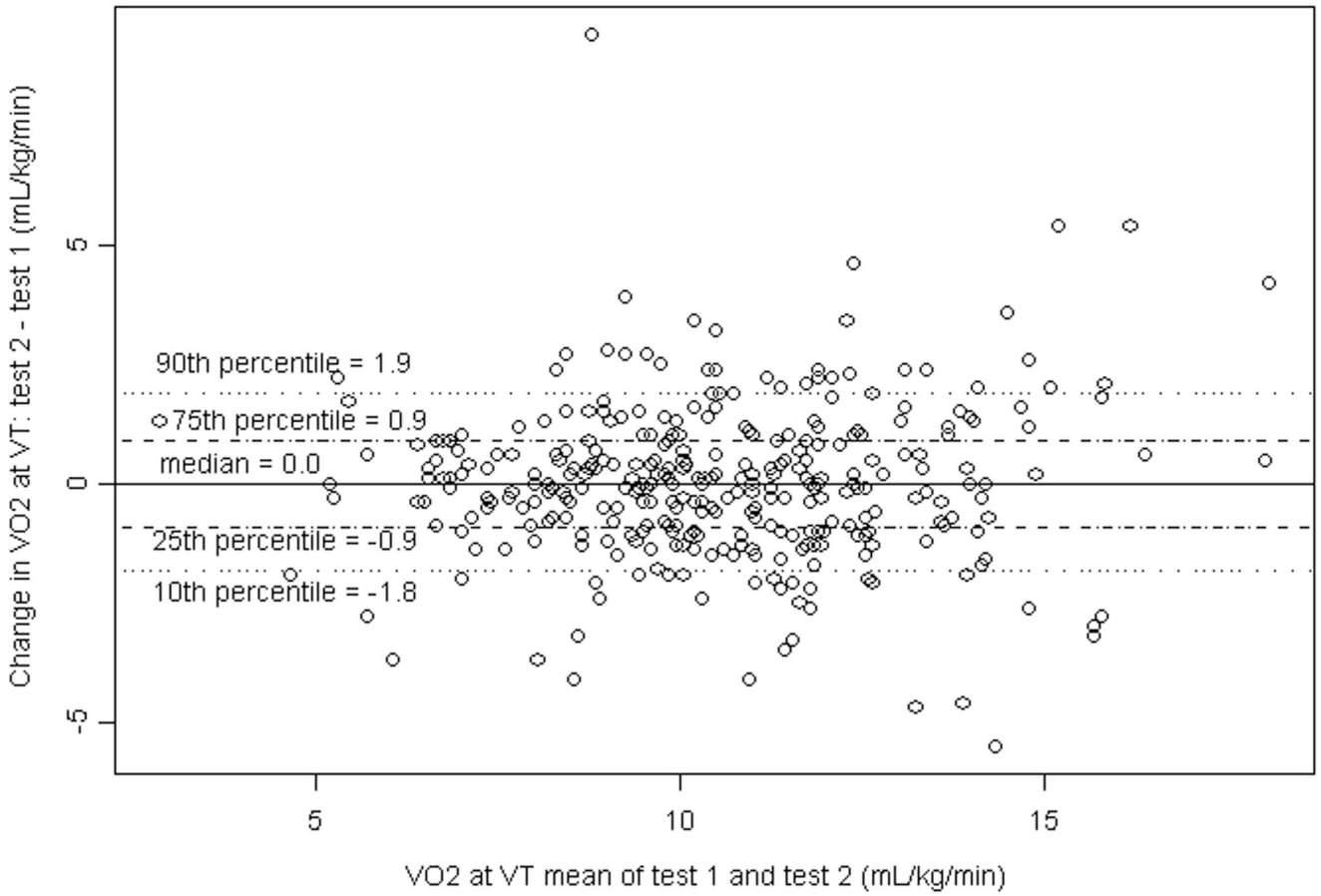
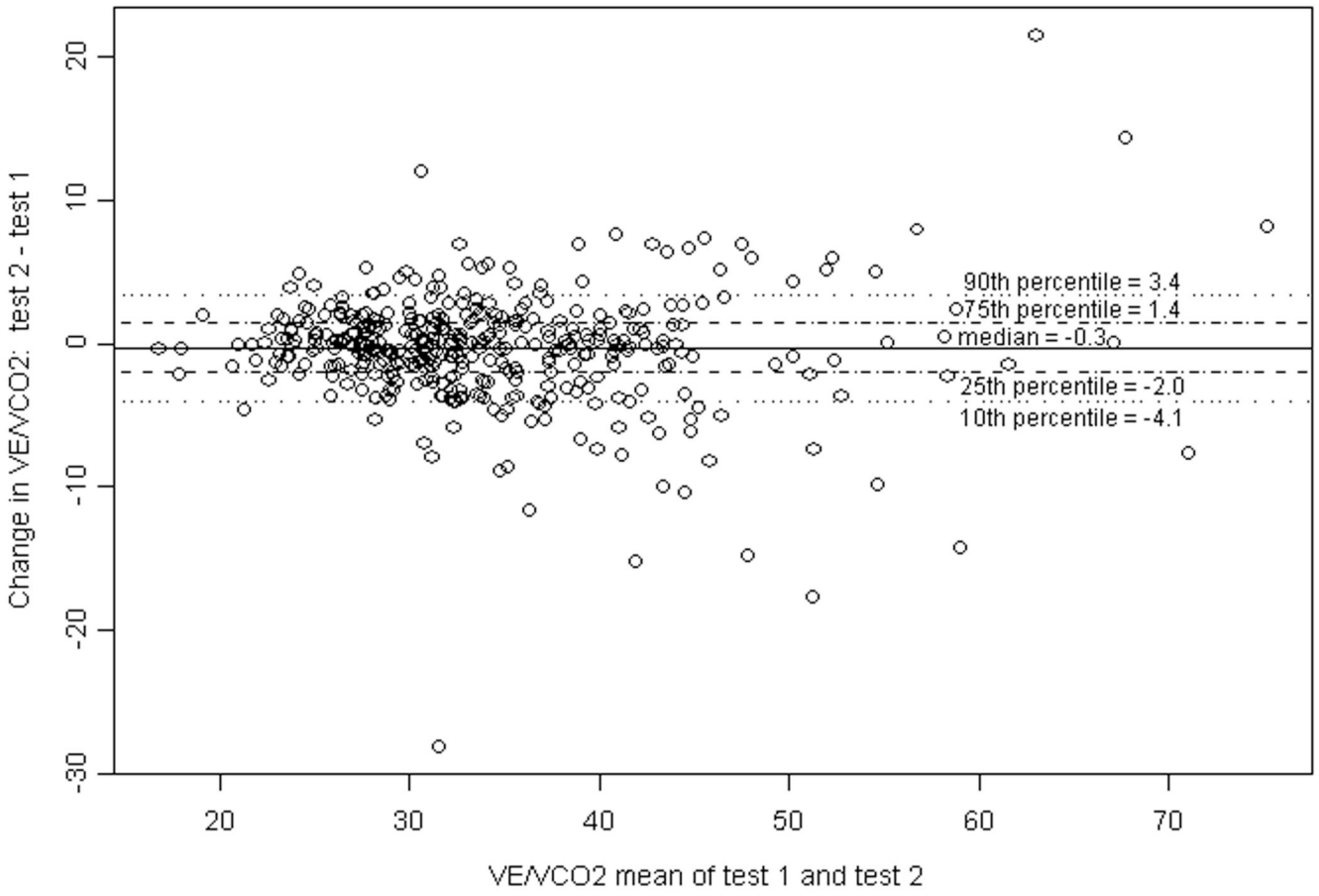


Figure 1C



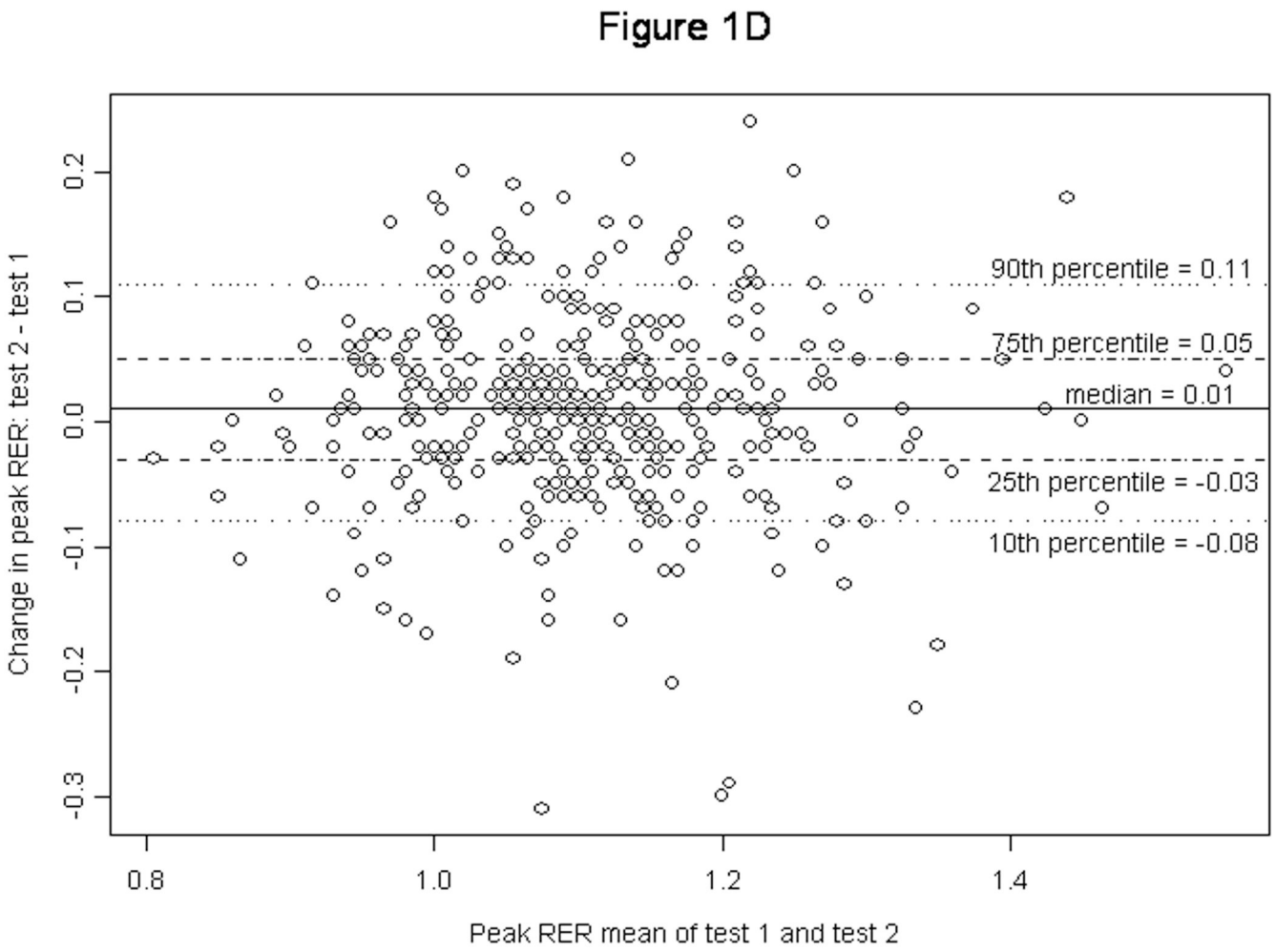


Figure 1.

Figure 1A. Bland-Altman Plot: Change in pVO_2 vs. mean pVO_2 for Test 1 and Test 2

Figure 1B. Bland-Altman Plot: Change in VO2 at ventilatory threshold vs. mean VO2 at ventilatory threshold for Test 1 and Test 2

Figure 1C. Bland-Altman Plot: Change in Ve/VCO_2 slope vs. mean Ve/VCO_2 slope for Test 1 and Test 2

Figure 1D. Bland-Altman Plot: Change in peak respiratory exchange ratio vs. mean peak respiratory exchange ratio for Test 1 and Test 2

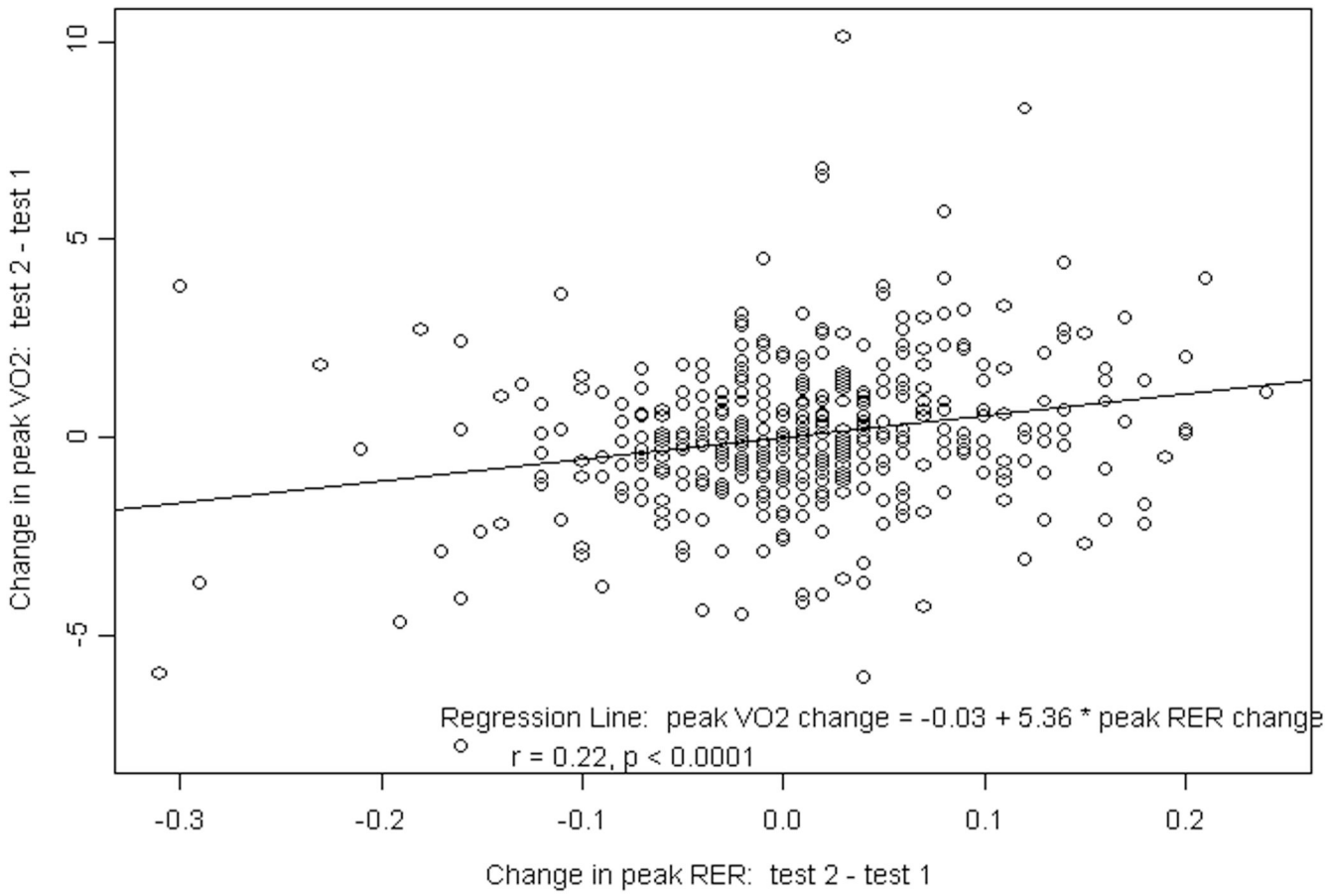


Figure 2.
Change in peak VO_2 as a function of change in peak respiratory exchange ratio

Table 1

Baseline demographics and peak oxygen consumption (n = 398)

Age (years)	59 (44, 74)*
Men/Women	73%/27%
White/AA/Other	68%/27%/5%
Weight (kg)	93 (67, 119)
Body mass index (kg/m^2)	31 (23, 39)
Etiology: Ischemic/Nonischemic	53%/47%
Ejection fraction (%)	26 (16, 36)
New York Heart Association II/III/IV	65%/34%/1%
Peak oxygen consumption, test 1 (mL/kg/min):	15.2 (9.3, 21.1)

*Continuous variables presented as mean (10th percentile, 90th percentile)

Table 2

Test-retest variability of cardiopulmonary exercise variables: mean (10^{th} percentile, 90^{th} percentile)

Variable	Test 1 mean (10th, 90th)	Test 2 mean (10th, 90th)	p-value*	Absolute change from test 1 to test 2 mean (10th, 90th)	% of tests for which T2 > T1	% of tests for which T1 > T2
Peak VO_2 (mL/kg/min) (n=398)	15.2 (9.3, 21.1)	15.2 (9.4, 21.1)	0.78	1.3 (0.1, 3.0)	46%	48%
Exercise time (s) (n**=393)	611 (320, 934)	637 (330, 942)	< 0.0001	60 (8, 140)	63%	33%
Ve/VCO_2 slope (n=390)	34.5 (25.3, 45.0)	34.1 (25.8, 44.2)	0.033	2.5 (0.2, 5.4)	44%	53%
VO_2 at VT (mL/kg/min) (n=331)	10.5 (7.4, 13.8)	10.6 (7.3, 13.7)	0.61	1.2 (0.1, 2.4)	50%	47%
Peak Borg RPE (n=397)	16.8 (13.0, 19.0)	16.9 (13.0, 20.0)	0.045	1.2 (0.0, 3.0)	36%	27%
HR at Peak Exercise (beats/min) (n=398)	121 (91, 148)	122 (92, 152)	0.042	8 (1, 17)	53%	41%
Peak RER (n=392)	1.11 (0.96, 1.25)	1.12 (0.98, 1.26)	0.004	0.06 (0.01, 0.13)	55%	39%

VO_2 = oxygen consumption, Ve/VCO_2 = ventilation per carbon dioxide, VT = ventilatory threshold, RPE = rating of perceived exertion, HR = heart rate, RER = respiratory exchange ratio.

* paired t-test of the null hypothesis that the mean of test 2 minus the mean of test 1 = 0

** number of subjects with non-missing values on both test 1 and test 2

Table 3

Test-retest variability of peak oxygen consumption (mL/kg/min) by demographic subgroups: mean (10^{th} percentile, 90^{th} percentile)

Variable	Test 1 mean (10^{th}, 90^{th})	Test 2 mean (10^{th}, 90^{th})	p-value*	Absolute change from test 1 to test 2 mean (10^{th}, 90^{th})	% of tests for which T2 > T1	% of tests for which T1 > T2
Site Experience						
Lower (n=272, 68%)	14.8 (9.3, 21.0)	14.8 (9.4, 20.9)	0.69	1.4 (0.1, 3.1)	44%	51%
Higher (n=126, 32%)	16.0 (9.1, 21.3)	16.0 (9.7, 21.2)	0.91	1.1 (0.1, 3.0)	50%	43%
Gender						
Female (n= 106, 27%)	14.2 (9.1, 19.3)	14.2 (8.8, 19.3)	0.96	1.2 (0.1, 2.3)	47%	47%
Male (n = 292, 73%)	15.5 (9.5, 21.8)	15.5 (9.6, 21.3)	0.73	1.4 (0.1, 3.1)	46%	49%
NYHA Class†						
II (n = 260, 65%)	16.1 (10.1, 22.0)	16.1 (10.5, 21.4)	0.57	1.4 (0.1, 3.2)	49%	47%
III (n = 134, 34%)	13.5 (8.5, 18.7)	13.4 (8.2, 19.1)	0.72	1.1 (0.1, 2.7)	42%	50%
Etiology						
Ischemic (n=211, 53%)	14.5 (8.8, 20.7)	14.4 (9.0, 19.9)	0.30	1.2 (0.1, 2.9)	45%	50%
Non-ischemic (n=187, 47%)	15.9 (10.4, 22.0)	16.1 (10.1, 22.6)	0.18	1.4 (0.1, 3.2)	48%	47%
Race						
African-American (n = 105, 27%)	14.1 (8.9, 19.4)	14.1 (8.5, 19.4)	0.84	1.4 (0.1, 3.1)	43%	52%
White (n=268, 68%)	15.6 (9.5, 21.9)	15.6 (9.7, 21.2)	0.95	1.3 (0.1, 3.0)	46%	49%

*paired t-test of the null hypothesis that the mean of test 2 minus the mean of test 1 = 0

†n = 4 (1%) NYHA Class IV subjects are excluded from the analysis.

Table 4

Table 4a. Univariate relationship of categorical variables to change in peak oxygen consumption (mL/kg/min) from Test 1 to Test 2

Variable	Mean change in Peak VO_2 from test 1 to test 2 (10^{th} percentile, 90^{th} percentile)	Within-variable difference	p-value*
Site experience			
Lower	0.05 (−2.1, 2.3)	0.07	0.74
Higher	−0.02 (−1.6, 1.5)		
Gender			
Female	−0.01 (−2.0, 1.8)	−0.05	0.82
Male	0.04 (−2.1, 2.3)		
NYHA Class			
Class II	0.07 (−2.2, 2.3)	0.12	0.51
Class III	−0.05 (−1.9, 1.7)		
Etiology			
Ischemic	−0.13 (−2.1, 1.8)	−0.33	0.09
Non-ischemic	0.20 (−2.0, 2.7)		
Race			
African-American	−0.04 (−2.2, 2.2)	−0.03	0.88
White	−0.01 (−2.1, 2.3)		

Table 4b. Univariate relationship of continuous variables to change in peak VO_2 from Test 1 to Test 2

Variable	mean (10^{th}, 90^{th})	Slope Estimate (95% conf. interval)	Correlation coefficient with change in peak VO_2	p-value
Age at baseline	59 (44, 74)	−0.008* (−0.09, 0.07)	−0.01	0.84
BMI at baseline	31 (23, 39)	0.01 (−0.02, 0.04)	0.04	0.43
Weight (kg) at baseline	93 (67, 119)	0.01† (−0.03, 0.06)	0.03	0.55
Ejection Fraction (echo core lab)	26 (16, 36)	0.02 (−0.01, 0.04)	0.07	0.19
Peak VO_2 (mL/kg/min) on first test	15.2 (9.3, 21.1)	−0.07 (−0.11, −0.04)	−0.19	0.0002
Peak RER on first test	1.11 (0.96, 1.25)	−1.62 (−3.27, 0.03)	−0.10	0.054
Ve/VCO_2 slope on first test	34.5 (25.3, 45.0)	0.01 (−0.01, 0.03)	0.04	0.38
VO_2 at VT (mL/kg/min) on first test	10.5 (7.4, 13.8)	−0.12 (−0.20, −0.04)	−0.15	0.004

* unpaired t-test of the null hypothesis that the mean change of group 1 – mean change of group 2 is 0

BMI = body mass index, VO_2 = oxygen consumption, RER = respiratory exchange ratio, $VeVCO_2$ = ventilation to carbon dioxide produced, VT = ventilatory threshold.

* slope estimate corresponds to 5 year increments

† slope estimate corresponds to 5 kg increments

Table 5

Multivariate predictors of change in peak VO_2 from Test 1 to Test 2

Variable	Slope Estimate (95% conf. interval)	Partial R^2	p-value
Peak VO_2 on first test	−0.08 (−0.12, −0.04)	0.04	< 0.0001
Peak RER on first test	−1.00 (−2.70, 0.69)	0.004	0.24
Ejection Fraction	0.025 (0.001, 0.049)	0.01	0.04
Etiology	−0.41* (−0.80, −0.03)	0.01	0.04

RER = respiratory exchange ratio

*Nonischemics have more of an increase in peak VO_2, on average, than ischemics do from test 1 to test 2.