



Published in final edited form as:

Med Sci Sports Exerc. 2011 December ; 43(12): 2381–2387. doi:10.1249/MSS.0b013e31822432d4.

Repeatability of aerobic capacity measurements in Parkinson Disease

Leslie I. Katzel, MD, PhD^{1,2}, John D. Sorkin, MD, PhD^{1,2}, Richard F. Macko, MD^{1,3,4}, Barbara Smith, RN, PhD⁵, Frederick M. Ivey, PhD^{1,3,4}, and Lisa M. Shulman, MD³

¹Baltimore Veterans Affairs Medical Center Geriatrics Research Education and Clinical Center, Baltimore, Maryland, 20201

²Department of Medicine, Division of Gerontology and Geriatric Medicine, Baltimore, Maryland, 20201

³Department of Neurology, University of Maryland School of Medicine, Baltimore, Maryland, 20201

⁴VA Rehabilitation Research & Development Maryland Exercise and Robotics Center of Excellence, Baltimore, Maryland, 20201

⁵University of Maryland School of Nursing, Baltimore, Maryland, 20201

Abstract

Purpose—Maximal or peak aerobic capacity (VO_2peak) during a maximal effort graded exercise test (GXT) is considered by many to be the “gold standard” outcome for assessing the impact of exercise training on cardiorespiratory fitness. The reliability of this measure in Parkinson Disease (PD) has not been established, where the degree of motor impairment can vary greatly and is influenced by medications. This study examined the reliability of VO_2peak during maximal effort GXT in subjects with PD.

Methods—Seventy healthy middle-aged and older subjects with PD, Hoehn and Yahr stage 1.5 to 3 underwent a screening/acclimatization maximal effort treadmill test followed by 2 additional maximal effort treadmill tests with repeated measurements of VO_2peak . A third VO_2peak test was performed in a subset of 21 subjects.

Results—The mean VO_2peak measurement was 2.4% higher in the second test compared to the first test (21.42 ± 4.3 ml/kg/min versus 21.93 ± 4.50 ml/kg/min, mean \pm SD, $p=0.03$). The intraclass correlation coefficients (ICC) for VO_2peak expressed either as ml/kg/min or L/min was highly reliable, with ICC of 0.90 and 0.94 respectively. The maximum heart rate (ICC of 0.91), and final speed achieved during the tests (ICC of 0.94) were also highly reliable, with the respiratory quotient (RQ) being the least reliable of the parameters measured (ICC of 0.65).

Conclusion—Our results demonstrate that measurement of VO_2peak is reliable and repeatable in subjects with mild to moderate PD, thereby validating use of this parameter for assessing the effects of exercise interventions on cardiorespiratory fitness.

Keywords

graded exercise test; reliability; cardiorespiratory fitness; maximal oxygen uptake; Parkinson Disease

Introduction

Resting tremor, bradykinesia, rigidity and postural instability with resulting impairment of gait and balance contribute to progressive impairment of mobility in Parkinson Disease (PD). The gait abnormalities in PD have been well described. Parkinson disease patients tend to walk slowly with shuffling and dragging steps, diminished arm swing and flexed forward posture gait with higher cadence and smaller stride length than non-PD patients. In more advanced disease, gait patterns include propulsive gait, steppage gait, waddling gait, and scissors gait. Despite significant advances in pharmacotherapy and deep brain stimulation, continuing development of new interventions to improve function remains a high priority for this patient population. We and others have proposed that treadmill based aerobic exercise training may improve cardiorespiratory fitness, ambulation and neurological performance in patients with PD (9,10,13,14,19,20). However, proper assessment of rehabilitation interventions including treadmill therapy depends on consistently reliable functional outcome measures. Maximal or peak aerobic capacity (VO_2peak) during a maximal effort graded exercise test (GXT) is often used as the “gold standard” outcome for assessing the impact of exercise training on cardiorespiratory fitness. For a measure to be a “gold standard”, the measure must be highly reproducible wherein repeated measurements obtained from a given subject under similar circumstances should show little within-subject variation. A measure which has this characteristic is referred to as having high reliability or precision. The reliability of this measure in PD has not been established, where the degree of motor impairment can vary greatly and is influenced by medications. No prior studies have examined the intraclass correlation (ICC) of repeated measurement of VO_2peak in PD. Hence, establishing repeatability of VO_2peak in PD is crucial, as a large variation in repeated measurements would confound the use of this parameter for assessing changes in cardiorespiratory fitness across rehabilitation interventions. This study examined the reliability of VO_2peak during maximal effort GXT in subjects with PD. As a secondary aim, we examined whether demographics or disease severity influenced the reproducibility of the measurement.

Methods

Subjects

Subjects were recruited from the University of Maryland Parkinson Disease and Movement Disorders Center, from the Baltimore VA Medical Center Parkinson’s Disease Clinic and via media advertisements from the community for participation in a randomized clinical trial of exercise training in Parkinson Disease. The University of Maryland Baltimore Institutional Review Board approved the protocol. After written informed consent was obtained at the first study visit, a history and physical examination including the Unified Parkinson’s Disease Rating Scale (UPDRS) (7) and Hoehn and Yahr staging (HY) (11) was performed by a neurologist with expertise in PD (LMS). Individuals underwent three timed 10 meter walks to evaluate their ability to ambulate and to assess their self-selected walking velocity. Self-selected velocity was defined as the average velocity during the three tests. To minimize the effects of pharmacological therapy, particularly with subjects on L-dopa, where there are well described fluctuations of wearing off or end-of-dose deterioration all study evaluations were performed while the subjects were “on” (i.e., experiencing benefit from their antiparkinsonian medications) at the same time of day with the same timing interval relative to their medication doses. Subjects used an additional dose of medication to maintain the “on” state for assessment when necessary.

The main study inclusion criteria were: 1) diagnosis of PD based on criteria of asymmetrical onset of at least two of the three cardinal features (resting tremor, bradykinesia, rigidity) with no atypical signs or exposure to dopamine blocking drugs 2) HY stage (5) 1.5 to 3

(while “on” for motor fluctuators), and presence of gait impairment or postural instability, as defined by a score of 1 or 2 on UPDRS (7) items #29 Gait or #30 Postural Stability (mild to moderate gait or balance impairment including slowing, dragging of the affected leg, freezing, or festination), 4) Age \leq 40, 5) Folstein mini mental status examination (8) score \geq 23, and 6) Unlikely to require PD medication adjustment for 4 months. The study exclusion criteria included: 1) unstable cardiac, pulmonary, liver or renal disease, 2) unstable hypertension or diabetes, 3) anemia, orthopedic or chronic pain condition restricting exercise, 4) unstable psychiatric illness, or 5) $>$ 20 minutes of aerobic exercise more than 3 times per week (to avoid a prior training effect). The individuals selected for participation did not have any of the 15 standard Brain Bank diagnostic criteria that exclude a PD diagnosis (such as rapidly progressive dementia and gaze palsy).

Procedures

Exercise Treadmill Testing

a) Screening exercise treadmill test: Following the initial study evaluation, a screening graded exercise treadmill test to voluntary exhaustion without VO_2 monitoring was performed using a customized manual protocol as previously described (19). All treadmill testing was performed in the early afternoon while the subjects were “on”. Stopping criteria were based on the American College of Sports Medicine criteria (1). This screening exercise treadmill test served to: 1) acclimate the subjects to walking on a treadmill; 2) evaluate for symptoms of overt coronary disease and to detect silent myocardial ischemia; 3) evaluate hemodynamic heart rate and blood pressure response to exercise; 4) observe gait patterns; and 5) determine whether there were any significant balance problems or other issues that would preclude their ability to safely exercise. All subjects wore a gait belt for safety, and a spotter stood behind subjects during the treadmill evaluations. Subjects were instructed to use the minimum level of handrail support for balance during the test.

Immediately preceding each test, baseline pre-exercise electrocardiogram and blood pressure were assessed with the subjects supine, seated and standing. After preliminary ECG and vital signs were obtained, the subjects started walking on the treadmill. The initial target speed for treadmill testing was the subject’s self-selected over ground walking velocity, with the incline set at 0%. In some subjects, the initial treadmill speed was adjusted slightly during the first 30 seconds by the exercise physiologist and clinician according to the subject’s tolerance with the pre-selected treadmill speed. The first stage was conducted for 2 minutes at 0% grade, the next stage was conducted for 2 minutes at 4% grade, and the grade was subsequently advanced by 2% every minute until voluntary exhaustion. In frailer subjects, the second stage was conducted at 2% instead of 4% to allow a more gradual increase in workload. Once the grade reached 10%, subjects were asked if the speed of the treadmill could be simultaneously advanced with grade (generally by 0.2 mph). The ECG was monitored continuously, and blood pressure measured during the first 3 stages of the tests and every 2 minutes during recovery. This protocol was well tolerated. Only one subject was excluded due to the development of chest pain during the test.

b) Graded exercise treadmill test with measurement of peak oxygen consumption: At the next study visit, generally one week later, subjects underwent a progressive GXT to voluntary exhaustion as described above with measurement of peak oxygen consumption ($\text{VO}_{2\text{peak}}$ using a Quark Cardio Pulmonary Exercise Testing metabolic analyzer (Cosmed, Rome, Italy). During the test, O_2 consumption, CO_2 production, and minute ventilation were measured breath-by-breath and values averaged for 20 second intervals. Subjects were instructed not to talk during the test as this is known to affect the depth of breathing and gas exchange. Based on our pilot study (19) we anticipated that many of these deconditioned subjects would not be able to obtain a true maximal aerobic capacity, defined as a plateau in

oxygen consumption during the final stage, maximal heart rate > 85% of age-adjusted predicted maximal heart rate, and respiratory quotient (RQ) or respiratory exchange ratio (RER) > 1.10. Therefore the values we obtained in many of the subjects represented peak effort values or VO_2peak as opposed to VO_2max . The VO_2 peak was based on the mean of the final two 20 second averages obtained during the final stage of the test.

One week later, subjects underwent a second evaluation of their VO_2 peak. The exercise physiologist set the initial speed of the test at the speed reached during the final stage of the prior test of peak VO_2 consumption, or based on comments noted during the prior test. The subjects exercised to voluntary exhaustion. Importantly, both the test supervision and the VO_2peak determinations were made by study staff who were blinded to the results of the prior treadmill tests. We report the baseline pre-exercise intervention data on 70 consecutive eligible subjects that performed two maximal effort treadmill tests with measurement of peak oxygen consumption.

To gain further insight into VO_2 peak test-retest reliability, a third maximal effort GXT with measurement of oxygen consumption was performed a week later in 21 subjects for whom the VO_2 peak measurements for the two treadmill tests varied by >5%. The exercise physiologist set the initial speed of the test at the speed reached during the final stage of the prior test of peak VO_2 consumption, or based on comments noted during the prior test. The subjects exercised to voluntary exhaustion. As in the prior tests, the supervising clinician was blinded to the results of the prior treadmill tests including the determination of the VO_2 peak.

Statistical analyses

Analyses were conducted using SAS software, version 9.2 (SAS Institute, Inc, Cary, North Carolina). Outcome measures obtained during GXT included peak oxygen consumption (expressed in L/min and mg/kg/min), maximal heart rate, maximal ventilation, RQ, maximum treadmill speed and maximum treadmill grade. All statistical tests were two-sided and performed at a significance level of 0.05.

Pearson product-moment correlations were computed from data obtained from the first and second VO_2peak tests. Unbiased intraclass correlations (r_k^2) (23) were computed using all available data, i.e. data obtained from the first two tests, and data for three tests in the 21 subjects on whom a third test was performed. The correlations (Pearson correlations) and intraclass correlations (ICC), along with their associated 95% confidence intervals were computed from 10,000 bootstrap samples of the original 70 subjects. Each bootstrap sample containing 70 rows of data was obtained by selecting with replacement from the original 70 subjects. The mean of the 10,000 bootstraps is reported as the mean for the statistics, the range of values including the 2.5 centile to the 97.5 centile of the distribution of the 10,000 values for each statistics is reported as the statistics 95% confidence interval. High reliability was defined as an ICC greater than 0.85.

We used random effects ANOVA to determine if there was a training effect across the time points. Three random effects models were examined; (a) random intercept, (b) random time, and (c) random intercept and time. We chose the best covariance structure using Corrected Akaike Information criteria. Bland-Altman plots were constructed (11), defined as the difference in values from test 2 and test 1 plotted versus the average of the values from tests 1 and 2. We computed the limits of agreement during Bland-Altman analysis. This was specified as the bias \pm 1.96 Standard deviation of the difference (average difference \pm 1.96 standard deviation of the difference). Linear multiple regression analyses were used to determine whether baseline characteristics and severity of PD affected the repeatability of the measurements.

Results

Physical characteristics and PD severity for the 70 subjects (50 men and 20 women), with a mean age of 65.3 ± 10.7 years (range 42 to 86 years) are summarized in Tables 1 and 2. Overall, based on the UDPRS and HY ratings, the subjects had moderately severe PD. However, there was a wide range of disease severity. Nine subjects (13%) had been treated with deep brain stimulation for their PD. Beyond their PD diagnosis, the population was healthy, with only four individuals (6%) having prior history of stable coronary artery disease, seven (10%) on medications for diabetes, and only one was a current smoker (1%). Only seven (10%) were on beta-blockers. As expected, a history of depression was common in this population with twenty-two subjects (31%) currently on antidepressant medication.

The treadmill testing was well tolerated. There were no falls or serious adverse events during the treadmill testing. There were initial concerns based on preliminary pilot studies by our group (1) that some subjects might have symptomatic orthostatic declines in their blood pressure going from supine to standing or during the recovery phase of the test. This did not prove to be a problem as only three subjects had asymptomatic orthostatic declines in BP that did not interfere with the assessment or warrant premature termination of the test.

Comparison of the 2 peak treadmill tests (VO_2peak) performed one week apart allowed for the determination of the ICC for a number of cardiopulmonary parameters (Tables 3,4). The overall mean VO_2peak measurements for tests one and two were 21.42 ± 4.30 ml/kg/min and 21.93 ± 4.50 ml/kg/min respectively ($p=0.03$), or 2.4% (0.56 ml/kg/min) higher on average for the second test. Bland-Altman plots of the within subject change for VO_2peak expressed in ml/kg/min versus the mean of tests 1 and 2 (figure 1) show the small but statistically significant increase of 0.56 (95% CI of -3.5 to 4.6) ml/kg/min between the first and second tests. The second test had higher VO_2peak values in 38 (54%) of the tests, consistent with a mild learning effect of the subjects and in the use of the final speed obtained in the first test to set the initial speed for the second test. During these tests, of the 63 subjects not on beta blockers, only 7 (11%) subjects achieved a true VO_2max based on a RER greater than 1.1 and achievement of $> 85\%$ age-predicted maximal heart rate.

The intraclass correlation coefficients (ICC) for the peak exercise tests (test 1 versus test 2) are summarized in Table 4. VO_2peak expressed both as ml/kg/min and L/min was highly reliable, ICC of 0.90 and 0.94 respectively. The maximum heart rate, and final speed achieved during the tests were also highly reliable in these patients. Final grade achieved during the tests was less reliable (ICC of 0.74) and fell below the 0.85 cutoff. The RQ proved the least reliable of the parameters measured during the GXT (ICC of 0.65).

We examined time (i.e. test number), age, sex, race, medical comorbidities, UPDRS total, UPDRS motor and HY stage to see if they predicted the change in VO_2 peak from one test to the next. None of these variables demonstrated significant bivariate correlations (both Spearman and Pearson) with VO_2peak expressed either in ml/kg/min (smallest p value 0.15) or L/min (smallest p value 0.25). Collectively the variables did not predict VO_2peak expressed either as ml/kg/min or L/min when they were included in a multiple regression.

To gain further insight into the reliability of the testing, and factors associated the variability, a third treadmill test was performed in 21 subjects in whom the first two tests differed by $> 5\%$. Fitness as assessed by VO_2peak , measured either in ml/kg/min or L/min improved with repeat testing (Table 5). The increases were small, but significant. VO_2 peak increased 0.56 ml/kg/min/test (a total of 1.2 ml/kg/min from first to third test) or 0.43 L/min/test (a total of 0.087 L/min from first to third test). Maximum heart rate, RQ, and maximum treadmill grade showed no significant increase across the three tests, however ventilation and maximum treadmill speed increased. Individual subject data revealed two

main testing patterns across the three tests. In 8 subjects, the third test had higher values than in either of the first two tests with a progressive increase in VO_2peak across the 3 tests in 5 subjects. This pattern is consistent with a learning effect. In 12 subjects, the third test was intermediate in value between the first and second test. This pattern is consistent with regression to the mean. In only one out of 21 subjects did we observe a progressive decline in VO_2peak across the three tests.

Discussion

Our results are the first to demonstrate that measurement of VO_2peak is reliable and repeatable in subjects with moderate PD, thereby validating use of this parameter for assessing the functional adaptation across rehabilitation interventions. The ICCs were 0.90 or higher for several peak treadmill testing measures including VO_2peak (expressed either in ml/kg/min or L/min), maximum heart rate, and final speed achieved. We observed a small, 2.4% mean increase in VO_2peak and other cardiovascular parameters such as maximal heart rate during the second test, consistent with subjects pushing themselves slightly harder during the second test. In almost all instances of disparate results (>5%) between the first and second test, the difference in measured VO_2peak was readily attributable at the time of the testing to differences in final grade and or speed achieved during the test, and subjective comments from the patients as to whether they were having a “good” day. Importantly, there were no falls or serious adverse events during these tests. None of the subjects had symptomatic orthostatic declines in blood pressure during the pre-test phase going from supine to standing or during recovery.

We performed a pubmed literature search using a variety of keywords, but were unable to find any other papers validating the reliability of treadmill testing with assessment of VO_2peak in subjects with PD. Importantly the reliability and repeatability parameters observed in PD subjects compare favorably to healthy subjects without PD. Generally, healthy controls have ICCs in the range of 0.85 to 0.98 for VO_2 peak and maximal heart rate (4,20,21). Turner et al. (22) reported high correlations for peak heart rate ($r=0.94$) and peak oxygen uptake ($r=0.98$) in subjects with peripheral arterial disease and claudication. Riebe (16) also reported high intraclass correlations for peak oxygen consumption ($r = 0.97$) in patients with peripheral arterial disease. We previously reported reliability coefficients of 0.87 for heart rate and 0.92 for VO_2peak in hemiparetic stroke patients studied using a GXT protocol similar to the one employed in the current PD study (6,12). In a large study of patients with congestive heart failure (HF Action), the mean VO_2peak was not different in tests 1 and 2 (3). However, the HF Action investigators reported a large within subject variability of VO_2peak between the two tests of 6.6%, with small but statistically significant increases in time on treadmill and RQ during the second test compared to the first test. The authors attributed the variation to intrinsic subject factors such as daily hemodynamic and volume status fluctuations. In comparing our results to other studies, particularly those done in healthy subjects, it is important to recognize that some of the studies employed treadmill tests set at the same speed and grade with the stated goal of attempting to reproduce values. This is in contrast to our approach which relied on staff discretion and participant feedback to dictate speed and grade progression to volitional exhaustion. Interpretation and comparison with other studies also must take into consideration differences in the ways data is expressed and analyzed.

Several studies have examined aerobic capacity in subjects with PD. A number of studies conducted testing on cycle ergometers making a direct comparison to the present treadmill study difficult. Overall, study results for VO_2peak in PD have been mixed with some reporting values similar to those of age-matched, healthy non-PD controls (5,15,18). Interestingly, Canning (5) reported no correlation between disease severity and VO_2peak in

individuals with mild to moderate PD. Although our subjects had VO_2peak values that were generally 20% lower than age-matched controls without PD studied in our laboratory (17), there was substantial heterogeneity in the VO_2peak and walking speeds.

There were several factors and testing methodology limitations that may contribute to increased variability in the measurement of VO_2peak . First, for safety reasons due to gait problems and increased risk of falling, subjects were allowed to hold on to the side or front railing and in many instances used the front railing of the treadmill for balance support. It is well established that hand support decreases workload and measured VO_2 . Subjects were repeatedly told to minimize hand rail support. Nevertheless, their inability to walk without rail support may have contributed to variability in the measurements and the modest upward trend in VO_2peak . Second, inherent biological variability in PD symptomatology is likely to affect reproducibility. Since PD is unique in its robust motor response to dopaminergic medications, it is particularly vulnerable to increased variability of motor performance. We attempted to minimize this by performing the tests at the same time of day with same time interval between medication dose and testing. Subjects often commented on the nature of their PD symptoms during the testing visits, but we did not objectively rate this source of variation. Third, there was inherent variability in the effort that these deconditioned subjects were willing to exert from day to day. Based on heart rate and respiratory quotient criteria, few of the subjects were able to achieve a true VO_2max . Fourth, technical limitations related to reproducibility of the metabolic equipment calibration prohibited exact replication of the measurement accuracy between testing visits.

The strengths and weaknesses of our study warrant comment. One strength of the study is that the investigative team has extensive experience in treadmill testing subjects with a variety of chronic medical conditions such as stroke, peripheral arterial disease, congestive heart failure, and HIV. Importantly, our group had performed preliminary pilot studies in PD before embarking on the current larger study. Lessons learned from these other populations with chronic disease and pilot PD participants were employed in this study. Another strength of the study was the relatively large number of subjects studied ($n=70$) and the range of disease severity. All participants had well characterized PD and all were receiving optimal medical management. Among the weaknesses was the potential for “volunteer bias”. Because volunteers were carefully screened for medical comorbidities that could impact on their ability to safely exercise, results from our population sub-group may not be generalizable to the general population of subjects with PD. We excluded patients with HY stages 4 and 5 (when “off”), limiting the applicability of our results to those with less severe disease. Those with more severe PD would likely have less reproducible measurements during treadmill testing based on wider motor fluctuations. Another limitation is that we studied well-controlled medicated PD patients who are “on” and that these results may not translate to those PD patients not on medications, or in whom the symptoms are not well-controlled.

In summary, our results demonstrate that GXT with measurement of aerobic exercise capacity is reproducible and can be safely performed in subjects with mild to moderate PD. Despite an observed learning effect across all three tests, the magnitude of the effect was small, averaging approximately 2.6% from the first to the second test. Our current recommendation based on these data is that two measurements of VO_2peak are warranted for intervention studies. Although a single test will probably suffice for characterizing fitness levels in cross-sectional studies, accounting for learning effects is important in longitudinal studies for filtering out changes that are not associated with the effects of rehabilitation interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Michael J. Fox Foundation for Parkinson's Research, The National Institute on Aging (NIA) Claude D. Pepper Older Americans Independence Center NIH grant P30-AG02874, VA Rehabilitation Research & Development Maryland Exercise and Robotics Center of Excellence, and the Baltimore VA Medical Center GRECC. We also wish to acknowledge the hard work and efforts of Terra Hill, Jessica Hammers, Kate Fisk, Brad Hennessie and other members of the study team.

The results of the present study do not constitute endorsement by ACSM

References

1. Thompson, WR.; Gordon, NF.; Pescatello, IS, editors. ACSM[®] guidelines for exercise testing and prescription. eighth edition. Philadelphia, Pennsylvania: Lipincott Williams and Wilkins; 2010. Chapter 5. Clinical exercise testing; p. 105-134.
2. Bland JM, Altman DG. Lancet. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1(8476):307–310. [PubMed: 2868172]
3. Bensimhon DR, Leifer ES, Ellis SJ, Fleg JL, Keteyian SJ, Piña IL, Kitzman DW, McKelvie RS, Kraus WE, Forman DE, Kao AJ, Whellan DJ, O'Connor CM, Russell SD. HF-ACTION Trial Investigators. 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). Am J Cardiol. 2008; 102(6):712–717. Epub 2008 Jul 9. [PubMed: 18773994]
4. Blessinger J, Sawyer B, Davis C, Irving BA, Weltman A, Gaesser G. Reliability of the VmaxST portable metabolic measurement system. Int J Sports Med. 2009; 30(1):22–26. Epub 2008 Jul 23. [PubMed: 18651368]
5. Canning CG, Alison JA, Allen NE, Groeller H. Parkinson's disease: an investigation of exercise capacity, respiratory function, and gait. Arch Phys Med Rehabil. 1997; 78(2):199–207. [PubMed: 9041903]
6. Dobrovolsky CL, Ivey FM, Rogers MA, Sorkin JD, Macko RF. Reliability of treadmill exercise testing in older patients with chronic hemiparetic stroke. Arch Phys Med Rehabil. 2003; 84(9): 1308–1312. [PubMed: 13680566]
7. Fahn, S.; Elton, RL. Members of the UPDRS Development Committee. Unified Parkinson's Disease Rating Scale. In: Fahn, S.; Marsden, CD.; Calne, DB.; Goldstein, M., editors. Recent developments in Parkinson's disease. Vol. vol 2. Macmillan Health Care Information; 1987. p. 153-164.
8. Folstein MF, Folstein SE, Mchugh PR. Mini-mental state: a practical method for grading the cognitive state off patients for the clinician. J Psychiatr Res. 1975; 12(3):189–198. [PubMed: 1202204]
9. Herman T, Giladi N, Gruendlinger L, Hausdorff JM. Six weeks of intensive treadmill training improves gait and quality of life in patients with Parkinson's disease: a pilot study. Arch Phys Med Rehabil. 2007 Sep; 88(9):1154–1158. [PubMed: 17826461]
10. Herman T, Giladi N, Hausdorff JM. Treadmill training for the treatment of gait disturbances in people with Parkinson's disease: a mini-review. J Neural Transm. 2009 Mar; 116(3):307–318. Epub 2008 Nov 4. [PubMed: 18982238]
11. Hoehn MM, Yahr MD. Parkinsonism: Onset, progression and mortality. Neurology. 1967; 17(5): 427–442. [PubMed: 6067254]
12. Macko RF, Katzell LI, Yataco A, Tretter LD, DeSouza CA, Dengel DR, Smith GV, Silver KH. Low-velocity graded treadmill stress testing in hemiparetic stroke patients. Stroke. 1997; 28(5): 988–992. [PubMed: 9158639]
13. Miyai I, Fujimoto Y, Yamamoto H, et al. Long-term effect of body weight-supported treadmill training in Parkinson's disease: A randomized controlled trial. Arch Phys Med Rehabil. 2002; 83(10):1370–1373. [PubMed: 12370870]

14. Protas EJ, Mitchell K, Williams A, Qureshy H, Caroline K, Lai EC. Gait and step training to reduce falls in Parkinson's disease. *Neurorehabilitation*. 2005; 20(3):183–190. [PubMed: 16340099]
15. Protas EJ, Stanley RK, Janovic J, Mac Neill B. Cardiovascular and metabolic responses to upper- and lower extremity exercise in men with idiopathic Parkinson's disease. *Phys Ther*. 1996; 76(1): 34–40. [PubMed: 8545491]
16. Riebe D, Patterson RB, Braun CM. Comparison of two progressive treadmill tests in patients with peripheral arterial disease. *Vasc Med*. 2001; 6(4):215–221. [PubMed: 11958386]
17. Rosen MJ, Sorkin JD, Goldberg AP, Hagberg JM, Katzell LI. Predictors of age-associated decline in maximal aerobic capacity: a comparison of four statistical models. *J Appl Physiol*. 1998; 84(6): 2163–2170. [PubMed: 9609813]
18. Saltin B, Landin S. Work capacity, muscle strength and SDH activity in both legs of hemiparetic patients and patients with Parkinson's disease. *Scand J Clin Lab Invest*. 1975; 35(6):531–538. [PubMed: 1198056]
19. Skidmore FM, Patterson SL, Shulman LM, Sorkin JD, Macko RF. Pilot safety and feasibility study of treadmill aerobic exercise in Parkinson disease with gait impairment. *J Rehabil Res Dev*. 2008; 45(1):117–124. [PubMed: 18566930]
20. Skinner JS, Wilmore KM, Jaskolska A, Jaskolski A, Daw EW, Rice T, Gagnon J, Leon AS, Wilmore JH, Rao DC, Bouchard C. Med Sci Sports Exerc. Reproducibility of maximal exercise test data in the HERITAGE family study. *Med Sci Sports Exerc*. 1999; 31(11):1623–1628. [PubMed: 10589867]
21. Snell PG, Stray-Gundersen J, Levine BD, Hawkins MN, Raven PB. Maximal oxygen uptake as a parametric measure of cardiorespiratory capacity. *Med Sci Sports Exerc*. 2007; 39(1):103–107. [PubMed: 17218891]
22. Tuner SL, Easton C, Wilson J, Byrne DS, Rogers P, Kilduff LP, Kingsmore DB, Pitsiladis YP. Cardiopulmonary responses to treadmill and cycle ergometry exercise in patients with peripheral vascular disease. *J Vasc Surg*. 2008; 47(1):123–130. [PubMed: 18178463]
23. Weiner, BJ. *Statistical Principles in Experimental Design*. 2nd Edition. New York: McGraw-Hill Book Company; 1971. p. 283-296.

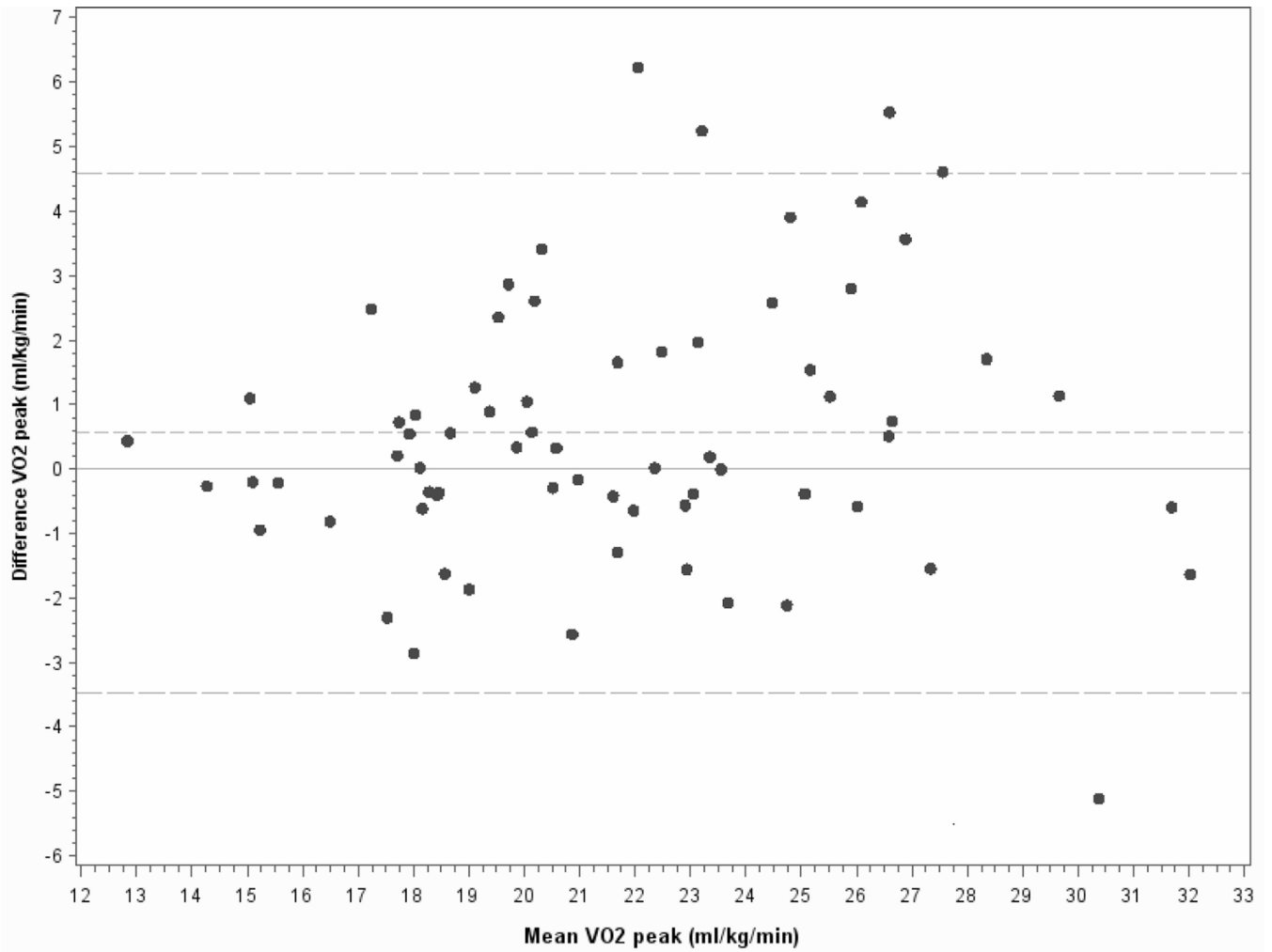


Figure 1. Bland-Altman plots for VO₂peak in ml/kg/min. Horizontal dotted lines show the small, but statistical significant mean difference between the tests and 5% and 95% confidence limit intervals (± 1.96 SD).

Table 1

Physical characteristics and severity of PD

	Mean \pm SD	Range
Age (years)	65.3 \pm 10.7	42 to 86
Self selected walking speed (mph)	2.43 \pm 0.52	1.12 to 3.9
Duration PD symptoms (months)	83 \pm 52	10 to 249
UPDRS Total score	47.44 \pm 12.67	15 to 89
UPDRS I: Mentation, behavior, and mood	2.20 \pm 1.35	0 to 8
UPDRS II: Activities of daily living	12.74 \pm 5.0	3 to 26
UPDRS III: Motor examination	32.5 \pm 9.53	11 to 59
UPDRS IV: Complications of therapy	2.81 \pm 1.91	0 to 8
Hoehn and Yahr Stage	2.17 \pm 0.37	1.5 to 3
Hoehn and Yahr Stage 1.5	N=1(1%)	
Hoehn and Yahr Stage 2.0	N=54 (77%)	
Hoehn and Yahr Stage 2.5	N=5 (7%)	
Hoehn and Yahr Stage 3.0	N=10 (14%)	

Table 2

PD motor and dyskinesia symptoms (N=70)

	N (%)
% with dyskinesia (proportion of waking day) UPDRS question #32	
0 (none)	50 (71%)
1 (1–25% of waking day)	17 (24%)
2 (26–50% of waking day)	2 (3%)
3 (51–75% of waking day)	1 (1%)
% with motor fluctuations (proportion of waking day) UPDRS question #39	
0 (none)	28 (43%)
1 (1–25% of waking day)	37 (53%)
2 (26–50% of waking day)	4 (6%)
3 (51–75% of waking day)	1 (1%)

Table 3

Cardiovascular Parameters and Their Change from First to Second Test

Measure	First test mean \pm SD	Second test mean \pm SD	With Person Difference mean \pm SEM	P value
VO ₂ peak (ml/kg/min)	21.4 \pm 4.3	22.0 \pm 4.5	0.56 \pm 0.25	0.03
VO ₂ peak (L/min)	1.85 \pm 0.55	1.90 \pm 0.56	0.04 \pm 0.03	0.10
Maximal heart rate (bpm)	131 \pm 20	133 \pm 22	2.2 \pm 1.0	0.04
Maximum ventilation (L/min)	61.0 \pm 9.1	63.4 \pm 19.9	2.4 \pm 1.0	0.02
RQ	1.02 \pm 0.10	1.04 \pm 0.09	0.01 \pm 0.01	0.24
Final speed during test (mph)	2.4 \pm 0.6	2.5 \pm 0.7	0.12 \pm 0.03	0.001
Final grade during test (%)	12.2 \pm 3.0	21.1 \pm 2.6	-0.11 \pm 0.29	0.70

Table 4

Intra-class correlations test 1 versus test 2 (N=70)

Variable	ICC	95% confidence interval
VO ₂ peak (ml/kg/min)	0.90	0.86 to 0.94
VO ₂ peak (L/min)	0.94	0.91 to 0.97
Maximal heart rate (bpm)	0.91	0.86 to 0.95
RQ	0.65	0.51 to 0.77
Final speed during test (mph)	0.94	0.90 to 0.97
Final grade during test (%)	0.74	0.62 to 0.83

Table 5

Subset of subjects with increased variability in whom 3 tests were performed

	First GXT (N=21)	Second GXT (N=21)	Third GXT (N=21)
VO ₂ peak (ml/kg/min)	21.7 ± 3.5 *	22.4 ± 3.9	22.8 ± 3.7
VO ₂ peak (L/min)	1.92 ± 0.59 *	1.97 ± 0.56	2.01 ± 0.54
Maximal heart rate	131 ± 23	136 ± 24	137 ± 27
Maximal ventilation	61.0 ± 19.4 *	64.2 ± 20.1	66.1 ± 18.6
RQ	1.01 ± 0.09	1.03 ± 0.09	1.03 ± 0.07
Final speed during test	2.3 ± 0.5 *	2.5 ± 0.6	2.5 ± 0.6
Final grade during test	12.1 ± 3.2	12.5 ± 2.7	12.6 ± 2.2

Note :

* values for first test differed significantly ($p < 0.05$) than values obtained during the second and third tests.