Reproducibility of endurance capacity and $\dot{V}O_2$peak in male Sprague-Dawley rats

Steven W. Copp, Robert T. Davis, David C. Poole, and Timothy I. Musch

Departments of Kinesiology, Anatomy, and Physiology, Kansas State University, Manhattan, Kansas

Submitted 4 December 2008; accepted in final form 6 February 2009

Copp SW, Davis RT, Poole DC, Musch TI. Reproducibility of endurance capacity and $\dot{V}O_2$peak in male Sprague-Dawley rats. J Appl Physiol 106: 1072–1078, 2009. First published February 12, 2009; doi:10.1152/japplphysiol.91566.2008.—The rat model of treadmill running is an invaluable tool for the investigation of experimentally and pathologically induced alterations in exercise performance. Interpretation of such data often presumes knowledge of the within-rat reproducibility of performance measures; however, the literature is bereft of this information. We tested the hypothesis that within-rat exercise endurance capacity and peak $O_2$ uptake ($\dot{V}O_2$peak) are highly reproducible across five measurements spanning ~5 wk when assessed with treadmill performance protocols. Male Sprague-Dawley rats ($n = 13$) performed five graded exercise tolerance tests to fatigue and five maximal exercise tests on a motor-driven treadmill for determination of endurance capacity and $\dot{V}O_2$peak, respectively. There were no differences ($P = 0.47$) in average time to fatigue among any of the five exercise tolerance tests (average range 45.9–52.1 min), and the average within-rat coefficient of variation (CV) over the five runs was 0.13. There were no differences ($P > 0.05$) among the average CVs from any consecutive weekly exercise tolerance tests (range of 4 CVs 0.06–0.10). As expected with the increase in body mass, relative $\dot{V}O_2$peak decreased (average range from 80.1 to 75.7 ml·kg$^{-1}$·min$^{-1}$, $P < 0.05$) throughout the five maximal exercise tests. However, there were no differences ($P = 0.63$) in the average within-rat CVs among any consecutive $\dot{V}O_2$peak tests (range of 4 CVs 0.03–0.04), and the average within-rat CV for all five tests was 0.06. The present data obtained from the protocols described herein demonstrate that within-rat measurements of endurance capacity and $\dot{V}O_2$peak are highly reproducible. These results have significant implications for improving and refining exercise testing and experimental designs.

Exercise performance; reproducibility; maximal oxygen uptake; treadmill running

EXERCISE ENDURANCE CAPACITY and the maximal capacity to transport and utilize $O_2$ [i.e., peak $O_2$ uptake ($\dot{V}O_2$peak)] during large muscle mass, dynamic-type exercise are fundamental parameters of integrated organ system function and health. Indeed, reductions in these parameters are hallmarks of many pathological conditions (e.g., heart failure and aging) (11, 23). In humans, measurements of exercise performance have been demonstrated to be highly reproducible within a subject (4); thus such reproducibility provides confidence in the ability to detect experimentally or pathologically induced alterations.

Ethical considerations limit the invasiveness of exercise testing procedures in human populations; therefore, rodent models of exercise performance are commonly employed to investigate the metabolic, cardiovascular, and pulmonary regulatory mechanisms of physiological function during exercise.

Accurate interpretation of many of these studies requires the tacit assumption that tests of endurance capacity and $\dot{V}O_2$peak are highly reproducible across subsequent and/or multiple repeated exercise bouts. In the rat, there is a broad range of various treadmill-running protocols for exercise tolerance and maximal exercise tests. It is surprising, therefore, that no rigorous analysis of exercise performance reproducibility has been performed in the rat, particularly because of the diversity of protocols, equipment, and running techniques employed in various laboratories (1, 2, 7–10, 12, 15, 17, 21).

The absence of such data may compromise the design and interpretation of investigations that assess changes in endurance capacity and/or $\dot{V}O_2$peak. For instance, many scientific questions are best addressed using a repeated-measures randomized crossover design, which allows each animal to serve as its own control. Unfortunately, if there is concern that confounding variables, such as changes in body mass, endurance capacity, or $\dot{V}O_2$peak that would hinder interpretation might occur within the experimental time frame, the investigator would likely utilize different groups of rats for control and experimental conditions. This consideration, combined with the requirement to overcome the population variability, necessitates the use of far more animals to achieve the same statistical power achieved with the randomized repeated-measures crossover design. Thus, development of protocols that permit highly reproducible endurance capacity and $\dot{V}O_2$peak measurements across time would be of great scientific utility.

The purpose of the present investigation was to determine the within-rat reproducibility across five tests of exercise endurance capacity and $\dot{V}O_2$peak assessed via treadmill testing protocols routinely used in our laboratory. We tested the hypothesis that endurance capacity and $\dot{V}O_2$peak are highly reproducible across time (several weeks).

METHODS

Animals. Male Sprague-Dawley rats ($n = 13$, initial age 2–4 mo) were obtained from Charles River Laboratories (Boston, MA). On arrival at Kansas State University, all rats were housed two per cage in facilities approved by the Association for the Assessment and Accreditation of Laboratory Animal Care and maintained on a 12:12-h light-dark cycle, with the light cycle beginning at 7 AM and ending at 7 PM. Food and water were provided ad libitum up to ~1 h before each test. All experimental procedures were approved by the Institutional Animal Care and Use Committee at Kansas State University and performed according to American Physiological Society guidelines for rodent exercise testing (20).

Experimental protocol. All rats were familiarized with running on a motor-driven treadmill for 5–10 min/day for 5 days at 25 m/min up to 10.220.33.3 on October 22, 2017 http://jap.physiology.org/ Downloaded from

Address for reprint requests and other correspondence: T. I. Musch, Dept. of Anatomy and Physiology, College of Veterinary Medicine, Kansas State Univ., Manhattan, KS 66506-5802 (e-mail: musch@vet.ksu.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
EXERCISE PERFORMANCE REPRODUCIBILITY IN RATS

a 10% grade. On familiarization and within a maximum of 3 days of the final acclimatization run, the experimental protocol was initiated: each rat performed five exercise tolerance tests to fatigue for the determination of endurance capacity and five maximal exercise tests for the determination of VO$_2$peak. The two different testing protocols were performed in alternating order, with no tests performed on consecutive days. The amount of time separating similar testing protocols (i.e., the average number of days between consecutive exercise tolerance or maximal exercise tests) was 9.8 ± 0.8 days. Our data demonstrate that this experimental design likely did not result in a training effect, as indicated by the lack of increase in endurance capacity and no change or a slight decrease in relative VO$_2$peak. All tests were performed between 8 AM and 4 PM, with the rat order randomly assigned to account for possible time-of-day bias in the results. The handler for each run was also randomized. Throughout the ~5-wk protocol, the room temperature was maintained between 20.5 and 22°C, and no additional fans or cooling devices were used. It was previously determined that exercise tolerance and maximal exercise tests using similar protocols result in rectal temperatures of 40.5 ± 0.1°C, which is unlikely to induce fatigue per se (T. I. Musch, unpublished observations).

**Determination of endurance capacity.** Each rat was weighed before the initiation of the exercise tolerance protocol for determination of endurance capacity. The protocol consisted of a progressive exercise test in which each rat initially ran at a speed of 25 m/min up a 10% grade for 15 min. Thereafter, the treadmill grade was held constant while the speed was increased by 5 m/min every 15 min until the rat was unable/unwilling to maintain pace with the treadmill belt (i.e., unable to come off the back of the treadmill lane for >3 s), despite encouragement to do so by application of manual bursts of high-pressure air aimed at the hindlegs. In nearly all cases, fatigue was immediately preceded by a lowering of the hindquarters and a raised snout, resulting in a significantly altered gait. Immediately after the end of each exercise tolerance test, the rat was removed from the treadmill and fatigue was confirmed by a significant attenuation of the righting reflex when the animal was placed on its back. Specifically, fatigued rats take several seconds or longer to right themselves. In contrast, the righting reflex occurs instantly in nonfatigued rats when they are placed in the supine position. In ~10% of tests, no gait change or obvious attenuation of the righting reflex was observed, and the exercise tolerance test was repeated after ≥24 h of recovery. In addition to the unwillingness/inability to continue running, the presence of either criterion (noticeable gait change or attenuation of the righting reflex) alone was accepted as a valid indicator of fatigue. The time to fatigue was recorded to the nearest second, and each rat was reweighed on completion of the test. The results reflect the rat’s postexercise test weight.

**Statistical analysis.** Values are means ± SE. Repeated-measures ANOVA were used to compare exercise tolerance tests, maximal exercise tests, and coefficients of variation (CVs, calculated as standard deviation/mean) across repeated tests. The within-rat CV was used in conjunction with group means as a measure of reproducibility, which is a better indicator of measurement reproducibility than the commonly reported coefficient of determination ($r^2$) and correlation significance (see DISCUSSION). Where a significant difference was detected, Tukey’s post hoc test was used to determine where the specific differences existed between tests or CVs. Pearson’s correlations were used to identify relationships between variables. The significance level was set at $P < 0.05$.

**RESULTS**

**Body weights.** The average initial weight of the rats was 352 ± 13 (range 278–396) g and increased ($P < 0.05$) to 473 ± 10 (range 396–513) g by the end of the ~5-wk experimental protocol.

**Endurance capacity.** The average time and distance run to fatigue for each of the five exercise tolerance tests are presented in Fig. 1. The analysis of the endurance capacity was not different whether it was expressed as time or distance run to fatigue. There were no significant differences in average time ($P = 0.47$) or distance run ($P = 0.42$) to fatigue among any of the five exercise tolerance tests. When analyzed for reproducibility between consecutive tests, there were no significant differences ($P = 0.64$) in the average within-rat CV among any of the consecutive runs (i.e., tests 1–2, 2–3, 3–4, and 4–5; Fig. 2, top). The CV for tests 1 and 2 and for tests 1–3 combined (0.06 and 0.09, respectively) were not different ($P > 0.05$); however, the CV for tests 1–4, as well as across all five tests (0.13 for both), were greater ($P < 0.05$) than the CV for tests 1 and 2 (Fig. 2). There were no significant correlations ($P > 0.05$) between the CV and time between tests or the CV and change in body weight between consecutive exercise tolerance tests.

**VO$_2$peak.** The average relative (ml·min$^{-1}·kg^{-1}$) and absolute (l/min) VO$_2$peak for the five maximal exercise tests is presented in Fig. 3 and a representative test for one animal is presented in Fig. 4. Average RER values obtained at peak exercise for each of the maximal exercise tests are shown in Table 1. In
previous studies, RER > 1.0 was used as an indicator of a successful maximal exercise test (28, 29). Although tests were repeated if RER was < 1.0, RER > 1.0 per se did not, and should not, constitute a successful test (i.e., proof of maximal VO₂) (26). There were no differences (P > 0.05) in relative VO₂peak among any of the consecutive tests. However, relative VO₂peak values were lower for tests 4 and 5 (P < 0.05) than for test 1, and VO₂peak was lower for test 5 (P < 0.05) than for tests 2 and 3. When analyzed for reproducibility, there were no differences (P > 0.05) among the average within-rat CV for any of the consecutive maximal exercise tests (Fig. 2, bottom). Contrary to the exercise tolerance tests, the within-rat CV for the maximal exercise tests did not change (P > 0.05) as an increasing number of maximal exercise tests were performed, and the average within-rat CV across all five tests was 0.06. Consistent with previous reports (2), VO₂peak was significantly negatively correlated (r = −0.34, P < 0.01) with body mass.

**DISCUSSION**

To our knowledge, this is the first investigation to establish a measure of reproducibility for treadmill-running tests of aerobic/anaerobic endurance capacity and VO₂peak in rats. The results of this study demonstrate the use of practical, reproducible progressive exercise tolerance and maximal exercise test protocols for the determination of endurance capacity and VO₂peak, respectively (Figs. 2 and 5). Using these protocols, we have identified several major considerations pertaining to experimental and test protocol design within the rat model of treadmill exercise performance. Specifically, interventions using endurance capacity as the outcome measure should con-

![Fig. 1. Top: average run time to fatigue among rats (n = 13) during 5 separate exercise tolerance tests for determination of endurance capacity. Bottom: average distance run to fatigue. P > 0.05 for both.](image)

![Fig. 2. Top: average (n = 13) within-rat coefficient of variation (CV) among an increasing number of exercise tolerance (endurance capacity) and maximal exercise [peak O₂ uptake (VO₂peak)] tests. *Significantly different (P < 0.05) from tests 1–2. Bottom: average within-rat CV between specific pairs of maximal exercise tests (P > 0.05).](image)

![Fig. 3. Average VO₂peak across 5 maximal exercise tests plotted as absolute and relative values. *P < 0.05 vs. test 1. *P < 0.05 vs. test 2. †P < 0.05 vs. test 3.](image)
sider the excellent within-rat reproducibility for any two consecutive exercise tolerance tests (Figs. 2 and 5). However, variability may increase as more runs and elapsed time are introduced into the experimental design. Investigators employing experimental perturbations that potentially alter \( \dot{V}O_2 \text{peak} \) should also be cognizant of high within-rat reproducibility during consecutive maximal exercise tests and note that relative \( \dot{V}O_2 \text{peak} \) \((\text{ml} \cdot \text{min}^{-1} \cdot \text{kg}^{-1})\) decreases over time, likely because of increases in body mass.

Comparison with previous literature. There is a paucity of available literature specifically addressing the reproducibility of treadmill performance in rats. Notwithstanding this consideration, \( \dot{V}O_2 \text{peak} \) data from the present investigation agree closely with values reported previously from our laboratory (10, 21) and others (2, 28, 29). Specifically, \( \dot{V}O_2 \text{peak} \) reproducibility over multiple tests is supported (similar mean values) by the control animals from studies by Woodman and colleagues (28, 29), even though within-rat reproducibility was not specifically reported in these investigations. In a classic study of treadmill-running performance in the rat, Bedford et al. (2) used a 10-stage fatiguing protocol to make repeated within-rat measurements of \( \dot{V}O_2 \text{peak} \) and reported a within-rat test/retest reliability coefficient of 0.83–0.97 between tests (although this is not a direct assessment of reproducibility, see Methodological and analytic considerations). Bedford et al. also noted that consideration of differences and/or changes in mass are vital in cross-sectional or longitudinal investigations measuring \( \dot{V}O_2 \text{peak} \). The results of the present investigation demonstrate that significant changes in body mass of the rat may occur within a very few weeks. Therefore, as suggested by Bedford et al. and supported by the present results, body mass is an obligatory consideration, and experimental protocols that compare \( \dot{V}O_2 \text{peak} \) values among different groups of rats or within the same rats over time must control for body mass.

Contrary to maximal exercise tests, there are no data that assess the reproducibility of tests of exercise tolerance for the determination of endurance capacity in rats. Koch and colleagues (16) reported a normal distribution for 120 endurance runs performed by 24 female Sprague-Dawley rats; however, this analysis does not address the within-rat reproducibility. Many investigations have utilized the mouse model for measurement of endurance capacity (18, 19, 27). Although endurance running reproducibility has not been reported in the mouse model and, thus, comparisons are limited, the rat is more suitable for invasive blood sampling, drug administration, and hemodynamic measurements via intravenous and

<table>
<thead>
<tr>
<th>Table 1. RER at peak exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak O(_2) Uptake Test</td>
</tr>
<tr>
<td>Test1</td>
</tr>
<tr>
<td>Test2</td>
</tr>
<tr>
<td>Test3</td>
</tr>
<tr>
<td>Test4</td>
</tr>
<tr>
<td>Test5</td>
</tr>
</tbody>
</table>

Values are means ± SE. RER, respiratory exchange ratio. \( P > 0.05 \) for all values.
intra-arterial catheters, and this fact should be considered in selection of the most appropriate and feasible experimental model.

Implications for exercise performance testing in rats. Because of the lack of knowledge of exercise performance measurement reproducibility, it is appropriate to compare test protocol designs and evaluate the physiological responses to varying protocols to facilitate the universal adoption of practical, reproducible measures of exercise performance.

The maximal exercise test protocol utilized here is a ramp-style protocol that increases the treadmill speed progressively until the rat can no longer maintain pace with the treadmill belt. By increasing the speed ~5–10 m/min, the present protocol ensures that an inordinate amount of time is not spent in any one stage, particularly at high workloads, thereby resulting in fatigue and inhibiting the attainment of true peak physiological variables (3). In addition, individual rats may differ in their behavioral response to changes in treadmill speed. Therefore, investigators should be trained to carefully control the rate of increases (i.e., slow, gradual increases vs. rapid increases) in treadmill speed, such that the individual personalities of each rat do not affect the test results (i.e., rapid changes in treadmill speed may cause some rats to stop running prematurely).

The endurance exercise protocol utilized in the present investigation was a graded exercise tolerance test with an initial speed of 25 m/min [i.e., ~65% \( \dot{V}O_2\text{peak} \) (23)]. With use of this graded protocol, fatigue was induced in ~50 min with an average final speed of 40 m/min. In contrast, many investigations have utilized protocols (either graded or constant) that limit speeds to ~30 m/min. At these submaximal levels, time to fatigue for untrained rats may vary from ~35 to ~250 min among laboratories (7, 8, 12), even when reported within the same rat population, while exercise-trained rats may run continuously for >9 h (12, 17). It is difficult to physiologically account for the vast differences in time to fatigue among studies utilizing similar rat populations and relatively slow treadmill speeds. This substantial variability suggests that endurance tests are often terminated as a result of other confounding variables (e.g., unwillingness to run), rather than the specific physiological consequences of the selected running velocities.

In this regard, the relation between time to fatigue and running velocity at high intensities is described by a rectangular hyperbola, where a finite amount of work (\( W' \)) can be performed above the asymptote of velocity (or power, i.e., critical velocity; Fig. 6). \( W' \) represents an anaerobic energy store and has been considered analogous to the “anaerobic work capacity,” consisting primarily of energy derived from creatine phosphate and anaerobic glycolysis. Fatigue occurs when \( W' \) is expended, and the critical velocity represents the highest velocity attainable without sustained use of the \( W' \) component. Above critical velocity, exercise tolerance is highly predictable from \( W' \) and, up to the point of fatigue, creatine phosphate and intramuscular pH fall and diprotonated

Fig. 6. Critical velocity (dashed line) is the asymptote of velocity-time to fatigue relationship for running at high intensities. Curve is constructed by having individuals perform separate constant-speed running bouts at various velocities (e.g., solid circles 1–4) that lead to fatigue in ~2–30 min. As reviewed in Jones et al. (14), the curvature parameter (\( W' \)) represents a finite energy store component (note the same hatched area for different velocities). Data points demonstrate vast variability of times to fatigue with constant-speed exercise tolerance protocols below 30 m/min (8, 12; data collected as part of Ref. 22). Inset: conditions such as exercise training (13, 25) and aging (24) would shift position of the curve and asymptote upward or downward in accordance with change in direction of \( \dot{V}O_2\text{peak} \), thus altering the speed at which the exercise tolerance protocol would be terminated. CHF, chronic heart failure.

Fig. 7. Individual rat \( \dot{V}O_2\text{peak} \) values from test 2 plotted against test 1. Similar to the example given by Bland and Altman (5), coefficient of determination (\( r^2 \)) = 0.19 and \( P = 0.13 \) defined here would lead to the erroneous conclusion that these consecutive \( \dot{V}O_2\text{peak} \) tests were not reproducible. However, similar mean values (dashed lines) from tests 1 and 2, small range of \( \dot{V}O_2\text{peak} \) values, and average within-rat CV between tests 1 and 2 of 0.04 ± 0.01 suggest that these two measurements were highly reproducible. Indeed, this CV falls within the expected measurement error of \( \dot{V}O_2\text{peak} \).
phosphate concentration rises inexorably to high, possibly limiting, levels (14). In marked contrast, at or below critical velocity, exercise tolerance becomes extremely prolonged, and the end point is poorly predictable. At these velocities, creatine phosphate is not depleted, and fatigue only occurs, at least theoretically, because of substrate (i.e., muscle glycogen) depletion, unless, of course, other confounding constraints (i.e., unwillingness to run) exist. Critical velocity occurs at ~70–90% of VO₂peak, and it is thus likely that healthy, and especially exercise-trained, rats running at ≤30 m/min and 10% grade (i.e., <70% VO₂peak) are below their critical velocity. For example, observations from our laboratory indicate that consecutive constant-speed endurance tests performed at 27 m/min up to 15% grade result in an average within-rat CV almost three to four times greater (0.29 ± 0.05; T. I. Musch, unpublished observations) than that reported in the present investigation. The design of our present exercise tolerance protocol ensures that fatigue occurs within a reasonable time in healthy, untrained rats (i.e., <1 h vs. up to ~250 min) and, by utilization of a progressive incremental test design, suggests that this protocol would be well suited in rat populations, where a change in VO₂peak and, therefore, critical velocity would be anticipated [i.e., exercise trained (13, 25) and aged (24) rats]. Importantly, validation of the critical velocity relationship in rats begs further evaluation.

Methodological and analytic considerations. A major consideration of the present protocols is the requirement for investigators to analyze changes in gait as a criterion for the validity of performance tests. During an exercise tolerance or a maximal exercise test, obvious gait changes, such as the lowering of the hindquarters, dropping of the tail, and elevation of the snout leading up to the cessation of running constitute perhaps the best criterion for the assessment of valid measurements (second to a direct observation of no increase in VO₂, despite increases in treadmill speed during measurement of VO₂peak). Once the test has ended, the inability or unwillingness of the rat, when placed in the supine position, to right itself provides additional less subjective assurance that fatigue/exhaustion is present.

For all tests, investigators must consider the running behavior of the rat on initiation of the test. It is our experience that <10% of rats will continue to turn and fight the treadmill belt for extended periods of time before settling into a normal running gait at the beginning of a test. In these instances, the rat may fatigue prematurely because of energy expended at the initiation of the test, and these animals should be excluded from data analysis. Thorough and carefully conducted acclimatization to the treadmill may help reduce these occurrences. The observation of an altered gait requires highly subjective analysis to assess true physiological fatigue and experienced handling of the treadmill to ensure the safety of the rat. Individuals performing exercise performance tests on rodents should be able to adequately demonstrate both of these skills, which will enhance the validity and reproducibility of performance measurements on rats.

The present statistical analysis uses the group mean in combination with the within-rat CVs between and among tests to examine reproducibility. Reproducibility analyses commonly consist of the calculation of correlation coefficients, r², and assessment of the significance of that correlation between two independent measurements. In the present investigation, correlation analysis would be invalid for determining reproducibility for the following reasons: 1) small, but statistically significant, mean differences in VO₂peak between various tests would conflict with high r² and significant Pearson’s correlations that would otherwise signify a reproducible test, and 2) the presence of a significant correlation depends crucially on the range of the measured quantity. In the present investigation, the small range of values present for endurance capacity and VO₂peak among animals often led to low r² and nonsignificant P values, which would mistakenly support the conclusion that reproducibility was poor or absent (Fig. 7) (5). The use of the within-rat CV provides an objective measure of reproducibility that may be compared with other species or populations or within the same species, even when their mean endurance capacity and VO₂peak values are significantly different from those of normal controls (i.e., exercise trained and/or heart failure rats).

Conclusions. This is the first investigation to specifically analyze and report measures of reproducibility during repeated exercise tolerance and maximal exercise tests spanning ~5 wk in rats. We have demonstrated that, over this time period, when using the protocols described here and employing rigorous and practiced rat-handling techniques, aerobic/anaerobic endurance capacity and VO₂peak are highly reproducible over multiple tests. However, important considerations for experimental design include the introduction of increasing variability as more exercise tolerance tests are performed and the potential confounding influence of increases in body mass on VO₂peak. The results of the present study are pertinent to the accurate interpretation of measures of endurance capacity and VO₂peak in the rat model of treadmill exercise performance and have significant implications for improving and refining exercise protocols and experimental designs. Specifically, notwithstanding differences in laboratory equipment and rat-handling techniques among individual investigators, use of the protocols described here establishes the scientific basis for utilization of the scientific “gold standard” randomized crossover experimental design when endurance capacity and VO₂peak are the prominent outcome measures sought.

ACKNOWLEDGMENTS

The authors thank K. Sue Hageman, Kyle F. Herspring, Lauren E. Hammel, and Daniel M. Hirai for contributions to the study and Dr. Timothy Lightfoot for helpful comments.

GRANTS

This study was funded, in part, by American Heart Association-Heartland Affiliate Grant-in-Aid 750009Z.

REFERENCES


