



Published in final edited form as:

Physiol Behav. 2009 October 19; 98(4): 433–440. doi:10.1016/j.physbeh.2009.07.006.

Repeatability of Exercise Behaviors in Mice

Amy M Knab¹, Robert S Bowen¹, Trudy Moore-Harrison¹, Alicia Trynor Hamilton¹, Michael J Turner¹, and J Timothy Lightfoot¹

¹Department of Kinesiology, University of North Carolina, Charlotte

Abstract

Purpose—Measurements of exercise behaviors in rodents such as maximal treadmill endurance and physical activity are often used in the literature; however, minimal data are available regarding the repeatability of measurements used for these exercise behaviors. This study assessed the repeatability of a commonly used maximal exercise endurance treadmill test as well as voluntary physical activity measured by wheel running in mice.

Methods—Repeatability of treadmill tests were analyzed for both inbred and outbred mice in addition to a 10 week repeatability analysis using Balb/cJ mice (n=20). Voluntary daily physical activity was assessed by distance, duration, and speed of wheel running (WR). Physical activity measurements on days 5 and 6 of WR in a large cohort (n=739) of both inbred and outbred mice were compared.

Results—No significant differences ($p>0.05$) in exercise endurance were found between different cohorts of Balb/cJ and DBA/2J mice indicating strains overall generally test the same; however, significant differences between tests were seen within BaD2F₂ animals ($p<0.001$). Bland-Altman analysis revealed lack of agreement between weekly endurance tests within mouse, and correlation analysis showed lack of consistent correlations between weekly endurance tests within mouse. No significant differences were found for WR measurements within mouse between days ($p=0.99$). High correlations between days within mouse for WR were found ($r=0.74-0.85$).

Conclusions—High intra-mouse variability between repeated endurance tests suggests that treadmill testing in an enclosed chamber with shock grid for motivation to run in mice is not repeatable. Conversely, high correlation and agreement between days of wheel running measurements suggest that voluntary activity (WR) is repeatable and stable within individual mice.

Keywords

running wheel; endurance; treadmill; physical activity

Introduction

Most measurements of exercise behavior in humans (e.g. exercise endurance, VO_2 max, activity level) have been shown to be repeatable within subject [8,28,41]. With this precedence, measurements of exercise endurance and daily physical activity in rodents are often used to

Corresponding Author: Amy Knab, knabam@appstate.edu Address: Appalachian State University, Human Performance Laboratory, Plants for Human Health Institute, North Carolina Research Campus, 600 Laureate Way, Kannapolis, NC 28081, Telephone: 704-250-5352, Fax: 704-250-5423.

¹**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

investigate regulating mechanisms associated with exercise that are difficult to measure in humans [22,23,26]. Given the high test-retest repeatability for human exercise behavior measurement, it is natural to assume that endurance tests in rodents would also be repeatable and stable. However, repeatability of exercise measurements in rodents must be established to ensure valid physiological conclusions from such studies. In the current study, we hypothesized that both treadmill running (using an enclosed chamber treadmill) and wheel running are repeatable measurements in mice.

Exercise behavior testing in rodents usually consists of either the determination of exercise endurance/capacity and/or voluntary daily activity. Forced exercise capacity tests in rodents generally use small treadmills encapsulated by a chamber to assess maximal exercise endurance and/or VO_2max [18,22,24,36,39]. These treadmill protocols typically use a variety of stimuli (e.g. shock grid, tail tapping, or high pressure bursts of air) to motivate the animal to run. Treadmill testing for assessment of endurance/aerobic capacity in rodents has been generally preferred to swimming tests since rodents do not display consistent swimming behaviors (e.g. animals will bob, float, and/or dive) and these behaviors skew any data investigating aerobic capacity [20]. Several variations of exercise treadmill protocols have been used with rodents [4,18,22,24,25,31,36,39]; however, in the current literature, limited studies report a measure of repeatability of forced treadmill testing within animal [6,17,31]. These studies report within animal repeatability of VO_2max measurements, using enclosed treadmill protocols ranging from $r=0.42$ to 0.97 [6,17,31]. In spite of the wide use of exercise endurance treadmill testing in rodents, no repeatability measures of maximal running time using enclosed chambers (as opposed to the repeatability of the measurement of VO_2max itself) have been reported. Koch and colleagues used a protocol consisting of five consecutive endurance tests on consecutive days [18] and have reported that “120 runs in 24 female rats were found not to be different from a normal distribution as assessed by the Kolmogorov-Smirnov test”. Unfortunately, it was not noted whether the five tests differed significantly from each other, and it is not clear whether this is a good indicator of repeatability. Thus, although some papers present some form of repeatability of VO_2max measurements in rodents, no studies have systematically analyzed the within subject repeatability of forced exercise treadmill tests (measured by run time) in rodents.

The other most common measurement of exercise behavior in rodents involves the determination of daily voluntary activity levels using wheel running [16,19,23,27,35,38,43]. Much like exercise endurance, day-to-day wheel running within strains of rodents has been assumed to be repeatable; however, little data is published regarding this assumption. Friedman and colleagues [17] evaluated several locomotor behaviors including wheel running in 35 random bred male ICR mice and reported a $r\text{-value}=0.852$ (with deletion of one outlier) between days 6 and 7 of wheel running. Additionally, Swallow et al. [36] tested 577 male and female mice selectively bred for high-wheel running activity and reported a $r\text{-value}=0.787$ for females, and a $r\text{-value}=0.868$ for males for repeatability of wheel running between days 5 and 6 of data collection.

Given the relative paucity of the data regarding the repeatability of rodent exercise behavioral measurements in the literature, the goal of this study was to examine the repeatability of commonly used forced exercise treadmill tests and daily voluntary physical activity measurements in several cohorts of inbred and outbred mice.

Methods

Overview

A variety of different mouse cohorts were used in the completion of this study. Archived, unpublished data from several previous studies [23,24,26,27] as well as data collected

specifically for this project are reported in this paper. All procedures were reviewed and approved by the University of North Carolina Charlotte Institutional Animal Care and Use Committee, conformed to the animal care policies of the U.S. Department of Agriculture (USDA), and conformed to the *Resource Book for the Design of Animal Exercise Protocols* [20]. All animals were housed in the University Vivarium with 12 hour light/dark cycles, were provided standard rodent chow (Harlan Teklad) and water *ad libitum*, and were weighed weekly. Mice used in maximal exercise treadmill tests were group housed four mice per cage and identified using ear punches. Mice used during wheel running experiments were single housed in rat size cages and identified using a unique mouse number as well all other identifying information on cage cards.

Animals Used

Exercise Endurance repeatability—Using data from previous studies, we first investigated whether exercise endurance was similar within inbred strain between different mouse cohorts separated in time. This question directly addressed whether exercise endurance within a particular strain of mouse was stable over time and was determined by comparing exercise endurance from two cohorts of Balb/cJ and DBA/2J inbred mice tested in the same manner in 1999 [24] and in 2005 (unpublished data). Both cohorts of mice were oriented to the treadmill twice, and we used an open treadmill, which allowed manual stimulation of the animal (tapping the tail) in conjunction with a shock grid to encourage running. Otherwise, the procedures used were the same as that addressed below. The strains tested in 1999 consisted of eight female Balb/cJ (weight = 19.0 ± 1.2 g) and seven female DBA/2J mice (weight = 16.9 ± 1.4 g), while the 2005 cohort consisted of 10 female Balb/cJ (weight = 20.6 ± 0.8 g) and 10 female DBA/2J mice (weight = 20.4 ± 1.6 g).

To determine repeatability of exercise endurance in outbred mice at two distinct time points, we compared repeated exercise endurance tests from 80 BaD2F₂ outbred mice that were tested using a sealed metabolic chamber that used a shock grid as the sole means to motivate exercise. Before testing began, mice were oriented to the treadmill twice by allowing the animal to sit in the chamber on the treadmill for 3 minutes, followed by 15 minutes of exercise at a speed of 16 m/min. These 80 mice were chosen from a cohort of 300 F₂ mice because they exhibited either high (n=40) or low (n=40) endurance during a maximal endurance test conducted using methods outlined below and previously published [24]. These mice were developed by reciprocally crossing high endurance Balb/cJ and low endurance DBA/2J inbred strains [24], and exercise endurance of the BaD2F₂ mice was measured at 86.3 ± 7.2 days (weight = 23.1 ± 3.1 g) and 140.1 ± 5.3 days of age (weight = 24.9 ± 2.7 g).

Finally, to investigate within mouse repeatability of exercise endurance across shorter time spans, but without intervening exercise training, 20 Balb/cJ mice (10 female, 10 male), were exercise endurance tested using the sealed metabolic treadmill approximately every seven days after one orientation to the treadmill (see below). Balb/cJ mice were chosen for this protocol because previous studies have shown this strain to perform well on forced treadmill tests [24]. The males were tested every seven days starting at age 41.5 ± 0.5 days. To eliminate possible sex hormone effects on exercise endurance, the female mice were tested during the diestrous phase of the estrous cycle which was determined by the presence of cornified epithelial cells in a vaginal smear [3]. This testing began when the females were 44.6 ± 0.5 days of age and given the normal length of the estrous cycle (≈ 4 –5 days, with diestrous lasting 2–2.5 days), endurance treadmill testing was accomplished approximately once every seven days.

Physical Activity Repeatability—We also determined if measurement of voluntary physical activity using a running wheel was repeatable. The data used to determine the repeatability of physical activity were taken from a large dataset using a base cohort of 739

mice from 22 inbred strains (n= 367; 129s1/SvImJ, A/J, AKR/J, Balb/cJ, C3H/HeJ, C3Heb/FeJ, C57BL/10J, C57BL/6J, C57BLKS/J, C57L/J, CAST/Ei, CBA/J, CE/J, DBA/2J, LP/J, MRL/MpJ, NZB/BinJ, PL/J, SM/J, SPRET/Ei, SWR/J, WSB/Ei) and from 2 outbred strains developed in our laboratory (n=372, C3C5F₁, C3C5F₂). Within this large cohort, there were 324 females and 415 males. Given that the highest activity levels for mice generally occur between 9 and 12 weeks of age [37], we attempted, where possible, to draw data for the day 5/day 6 repeatability comparison when the mice were 68–69 days of age (i.e. 9 weeks + 5 days). Thus, the average age of the mice for the day 5–6 comparison was 69.7±7.4 days. In 34 cases, data for the repeatability comparison was shifted from day 5–6 to day 4–5 or to day 6–7 because of equipment sensor failure on either day 5 or 6 of wheel running exposure.

Forced Maximal Endurance Testing

Similar methods were used to determine exercise endurance for all mice [24,26] with the exception of the use of an open treadmill or a sealed, metabolic treadmill (5.08 cm × 38 cm; Columbus Instruments, Columbus, OH). All mice, regardless of the treadmill used, had one or two orientation exposures to the treadmill, each separated by at least 48 hours from the other orientation exposure or an exercise endurance test. In all cases, the front eight cm of the treadmill chamber was covered to provide a dark area for the mice to run toward. The first orientation exposure consisted of placing the mouse on the treadmill and letting the mouse walk on the treadmill at 16 m/min for 15 minutes. A shock grid mounted at the back of the treadmill delivered a 3.0 mA current [22,36] to provide motivation for exercise. The treadmill endurance protocol consisted of a series of stages and has been described previously [24,26]. Briefly, each stage was three minutes long with the initial stage being a period of acclimatization. At the end of the first three minutes, the speed was increased to 16 m/min and then increased by four m/min every three minutes until a maximum speed of 40 m/min. If the mouse was still running at this stage the grade was increased every three minutes by five percent. The test was ended when the mouse sat on the shock grid at the back of the treadmill for five seconds, or if the protocol was maxed out at 36 minutes, 40 m/min, and 15% grade. As is evident from the data shown in this paper, a mouse very rarely reached the end of the protocol, thus the repeatability of the data is was not jeopardized by unclear endpoints.

To determine if exercise endurance measurement was repeatable over a longer period when tested weekly, each Balb/cJ mouse was endurance tested once a week for a period of ten weeks. As noted earlier, female mice were only tested during the diestrous phase of the estrous cycle when estrogen levels are lowest. In order to quantify any technician bias, five male and five female mice were randomly assigned to one of two technicians and these technicians conducted the endurance tests on the same ten mice each week throughout the study.

Voluntary Physical Activity Measurement

Daily running on the wheel was measured using methods described previously [21,23,27]. Briefly, mice were housed individually, with a running wheel (circumference 450mm; Ware Manufacturing, Phoenix, AZ) mounted in each cage. The wheels were equipped with a magnet mounted on the outside surface and the top of the cage was equipped with a magnetic sensor (BC500; Sigma Sport, Olney, IL). Each cage computer was calibrated for the wheel circumference allowing for accurate measurement of distance (km) and time the animals ran on the wheel (duration = mins). Speed of activity (m/min) on both days was calculated by dividing daily distance by daily duration of exercise. The data were collected every 24 hours for 7–21 days and data collected on days 5 and 6 were used for repeatability testing. The wheels were checked manually each day to assure sensor alignment and free turning of the wheel. “Coasting” by the mice, where the mice stopped running while the wheel continued to turn with the mouse still on the wheel, was not a concern due to three factors: 1) The running wheels used had a metal solid-surface and thus, they could not grip the wheel to coast unlike if the

wheel surface were mesh; 2) the wheels had a diameter that was too small for the mouse to run up one side and then coast as the wheel re-centered from the unequal weight on one side of the wheel; and 3) two cross axis bars attaching the wheel to the axle prevented the mice from jumping off the wheel while it was still turning, thus requiring that the mouse stop the wheel before getting off and removing any excess wheel spinning. In addition, anecdotally our research team has never observed the mice coasting the running wheels we use to measure daily activity.

Statistical Analysis

All analyses were conducted using JMP software (ver. 7.0, SAS Institute, Cary, NC) and the alpha value was set *a priori* at 0.05. Several analyses were used depending upon the questions being examined. A two-way ANOVA (factors = strain and year tested) was used to determine the overall *stability* (not repeatability) of exercise endurance between different mouse cohorts separated by time. A two-way ANOVA (factors = endurance classification and time of measure) with a repeated measure on one factor (time of measure) was used to determine differences in exercise endurance data within a cohort of F₂ mice that were classified on the basis of one exercise endurance test. Bland-Altman analysis [9] was used to determine 95% levels of agreement between tests (repeatability) within the F₂ mice. In short, the differences between repeated measures within an individual subject were assessed relative to the average between the two measures for the single individual. The average difference across all subjects was used to determine the levels of agreement in which 95% (± 1.96 standard deviation) of the subject were expected to lie. Determination of the repeatability of exercise endurance every week for 10 weeks within the same cohort of animals was also accomplished using Bland-Altman analysis. Additionally, pairwise correlations between all 10 weeks of endurance testing were conducted to determine the association of endurance test results across the 10 repeated endurance tests. Two-way ANOVA (time of measurement and sex) with time of measurement being a repeated factor, was used to assess differences in running data. Where appropriate, Tukey's *post-hoc* analysis were used to examine significant main effects.

A two-way ANOVA (day of measurement and sex) was used to initially determine if sex played a role in the repeatability of any of the physical activity measurements. If sex exerted a non-significant main effect, the analysis was repeated using paired t-tests with each running wheel index (i.e. distance run, duration of exercise, and speed of exercise) to determine differences in activity level between days 5 and 6 of exposure to a running wheel. Additionally, Bland-Altman analysis [9] was used to determine levels of agreement and repeatability between day 5–6 of wheel running distance for all 739 mice.

Results

Different groups of Balb/cJ and DBA/2J mice were endurance tested in 1999 and 2005. Results in Figure 1 show that endurance test performance was not different between these measurements, within strains of mice (Balb/cJ mice, $p=0.55$; DBA/2J mice, $p=0.51$) despite being separated by approximately six years. This result does not show repeatability; however it does imply stability of the endurance phenotype within strain over time.

A large cohort of F₂ outbred mice ($n=300$) were exercise endurance tested at 12 weeks of age and the top 40 performing animals were classified as “high endurance” and the lowest 40 performing animals were classified as “low endurance”. A second endurance test was conducted on these 80 mice within seven weeks of the original test. Figure 2 shows a Bland-Altman analysis showing lack of agreement between test 1 and test 2 within mouse, indicating low repeatability. In the second test, the high endurance mice exhibited significantly less endurance (test 1: 28.1 ± 2.0 min, test 2: 21.7 ± 7.0 min, $p<0.001$) than on their first test.

Conversely, the low endurance mice exhibited significantly higher endurance (test 1: 9.7 ± 4.2 min, test 2: 15.5 ± 8.3 min, $p < 0.001$) than on their first test (not displayed in graphical form).

In comparing the 10 weeks of endurance testing among male and female mice, no difference in association between max endurance tests were attributed to sex. Thus, all animals were combined, and pairwise correlations were completed for all 20 Balb/cJ mice for each week of endurance testing (Table 1). When compared using two-way ANOVA, starting at week four, significant differences were found between males and females in overall average run time with males running a significantly longer duration than females ($p = 0.035$). Two-way ANOVA also showed significant differences between exercise endurance tests across weeks in the female mice ($p = 0.041$). The coefficient of variation within each mouse between exercise endurance tests over the 10 weeks was very high for both males and females (average CV = 37.0, CV = 51.0 respectively). Fig. 3 shows a Bland-Altman analysis which indicated a very low agreement in endurance scores within mouse between weeks 1–2, 5–6, and 9–10. No technician bias was found to have been associated with the variation in endurance scores ($p > 0.05$, $t = 1.97$) and body weight was not correlated with endurance performance (males, $r = 0.26$; females, $r = -0.15$).

In regard to wheel running repeatability, female and male mice exhibited similar repeatability measures in distance, duration, and speed (data not shown). Thus, when all mice were pooled, there were no significant differences found between days 5 and 6 in distance, duration, or speed (Fig. 4). Additionally, high correlations between days 5 and 6 (distance, $r = 0.74$; duration, $r = 0.74$; speed, $r = 0.85$) indicate repeatability within mouse for physical activity measurements. The Bland-Altman analysis on the wheel-running data (Fig. 5) also highlights the high level of agreement between activity indices within mice.

Discussion

Over the past several years, studies examining both maximal endurance phenotypes and physical activity phenotypes in rodents have been reported in an effort to assess the genetic/biological factors involved in the regulation of these exercise behaviors [14,18,22,23,26,27,31,32,36,40]. Given the relative consistency of these measures of exercise behaviors in humans (e.g. $VO_2\text{max}$ tests) and in smaller reported cohorts of mice, all of which assessed repeatability of $VO_2\text{max}$ measurements [6,17,31], it has been natural to assume that these measures were repeatable in mice. In addition, given the fact that $VO_2\text{max}$ is a good predictor of exercise endurance in humans [5,11], and has been shown repeatable, the assumption could be made that maximal endurance tests used to assess endurance in rodents [4,18,24,40] would also be repeatable. Our finding of within strain consistency of overall endurance in different cohorts of mice over a six year period (Fig. 1) and the repeatability of voluntary wheel-running measurements (Fig. 4, Fig. 5) support this assumption. However, over the course of several years and a number of studies, a lack of consistency in repeat testing of mouse maximal endurance became apparent in our lab (Fig. 2). This evidence, led us to conduct the 10 week repeatability of max endurance outlined in Table 1 and Figure 3, and combined, these data raise questions regarding the repeatability of this method of maximal endurance measurement in mice.

Forced Maximal Endurance Tests

Conducting endurance treadmill tests in rodents can be difficult. It has been noted [6,20] that anywhere from 10–25% of rats will refuse to run on a treadmill, even with orientation exposures. Given the difficulty of having rodents perform forced endurance tests, it is surprising that relatively few studies have reported repeatability results of maximal exercise endurance or $VO_2\text{max}$ using a graded treadmill protocol in mice. Rezende and colleagues [31] measured $VO_2\text{max}$ during endurance treadmill tests in mice ($n = 48$) selectively bred for high wheel running and reported repeatability of $VO_2\text{max}$ during treadmill tests as $r = 0.42$.

Uniquely, Rezende and colleagues (29) also reported using a subjective scale to assess the quality of the treadmill tests. Any “poor trials” were not included in the analysis [31] suggesting that there was some acknowledgment that animals may not repeatedly run to exhaustion.

Other interpretations of rodent exercise capacity repeatability may be hampered by methodological limitations. Bedford and colleagues [6] tested the repeatability of a ten-stage graded treadmill test in rats (n=18) and reported a reliability coefficient of 0.97. However, Bedford and colleagues operationally defined $VO_2\text{max}$ as “one in which there is less than a 5% increase in VO_2 with increase in work intensity.” This operational definition was different than what is normally used in literature - allowing the rodent to run to exhaustion - and this operational definition difference may contribute to their observation of higher repeatability values compared to other studies. We have anecdotally observed that even in using four different rodent metabolic carts and three different forced exercise modalities, that oxygen consumption values in rodents often peak very early in a forced endurance test and then decline in spite of continued increases in workload. Speculatively, this type of response is most likely due to the common set-up of most commercially available sealed, rodent exercise metabolic chambers that allows the animal to remove their ventilatory stream from the gas sampling airstream when the mouse runs farther back on the treadmill. Support for this suggestion comes from earlier work by Friedman et al. [17] that tested the repeatability of several locomotor behaviors in random bred ICR mice (n=38) and reported a repeatability for $VO_2\text{max}$ of $r=0.809$. In this study, the authors used the peak VO_2 measurement during a test as the $VO_2\text{max}$ regardless of whether this peak measurement occurred at the end of the test when mice were exhausted and unwilling to run farther or if the peak was reached earlier in the test but the animal continued to run beyond this point. Thus, our observations, combined with both Friedman and colleagues’ [17] and Bedford et al.’s [6] studies suggest that repeatability of a forced exercise test in a rodent may depend upon the operational definition of the primary measure (e.g. $VO_2\text{max}$) used as well as the testing equipment used.

Since measurement of maximal aerobic capacity in rodents can be challenging, graded treadmill protocols have also been used to measure maximal endurance without measurement of $VO_2\text{max}$ [4,18,24,40]. To date, repeatability of exercise endurance measures using this type of protocol has not been reported. Koch et al. [18] initially implemented an endurance testing protocol which consisted of a week of increasing orientation bouts on a treadmill, followed by endurance max testing in the second week for five consecutive days to assess heritability of exercise endurance in rats. These authors reported that within sex, variation in the five consecutive max endurance tests “was found not to be different from a normal distribution”. However, in none of the publications where this endurance testing model has been used, has it been noted whether the five tests differed significantly from each other, nor whether possible physiological training effects of the five consecutive max tests occurred. Regardless of whether these items were considered, the exhibition of a normal distribution across repeated testing does not indicate repeatability. For example, in the current study, the repeated testing we did over a ten-week time period (Table 1, and Fig. 3), was still normally distributed (Shapiro-Wilk W test, $p=0.12$) in spite of exhibiting an approximately 400% difference in day to day results and virtually no test-retest significant association. Therefore, a set of repeated measures can have a normal distribution, yet be significantly different within-subject and thus, not be repeatable test-test. While it appears that measurement of $VO_2\text{max}$ in rodents may be repeatable under specific conditions, the data found in our study (Table 1, Fig. 2, and Fig. 3) indicate that measurement of endurance in mice (as assessed by time to exhaustion in a sealed treadmill chamber using a shock grid for motivation) may not be repeatable.

Indeed, one possible reason for the lack of repeatability within our studies was the use of a sealed metabolic chamber with shock grid. During our early use of an open treadmill which allowed manual encouragement of running (using tail tapping) we observed consistency within

strain, even across several years (Fig. 1). Musch, et al. have recently supported this suggestion of the need for manual encouragement when they reported repeatability of endurance testing in rats using both shock and physical encouragement [10]. However, the use of an enclosed treadmill, while necessary for metabolic measures, eliminates the possibility of using manual encouragement as a supplement for electric shock. Thus, the lack of an additional means of motivating continued running when using a sealed treadmill may be a large factor in whether measurements of endurance are repeatable or not.

Another possibility to explain the lack of repeatability with the sealed treadmill is that testing using electric shock as motivation may be more of a psychological stressor to the animal than other exercise measurements such as voluntary wheel running. This hypothesis is supported indirectly by several studies. First, the use of various means of motivation for running during forced treadmill tests (e.g. electric shock, puffs of air, tapping of the tail) may induce a negative response in the animal similar to that of chronic psychological stress, and this could mask true exercise behaviors [29]. One such negative response is the observation that brain derived neurotrophic factor (BDNF) decreases after forced exercise in rodent models similar to the effects seen during immobilization stress [1,2]. However, in humans, treadmill exercise has been shown to increase BDNF levels [15], as has wheel running in mice [13]. Thus, treadmill exercise in rodents may not be an appropriate model for the comparison of the response to treadmill exercise in humans due to the psychological stress to the rodent, which may in turn contribute to this measurement's lack of repeatability in mice.

The significant difference observed between male and female mice during the repeated 10 week endurance study was unexpected, but may be related to the time of measurement within the estrous cycle. Female mice were only endurance tested during the diestrous phase of their cycles, which corresponded to periods of low estrogen. There have been no studies investigating the effects of the estrous cycle on exercise endurance in rodents; however, numerous other studies have suggested that estrogen may play a role in the regulation of overall physical activity patterns [33]. Thus, while it cannot be definitively concluded that the low estrogen levels are responsible for the sex difference seen in average exercise endurance in this study, the wide test-to-test variation seen in both females and males across time (Fig. 3) and the lack of significant test-retest association (Table 1, Fig. 3) lends support to the finding that exercise endurance measured in a sealed treadmill is not repeatable. Furthermore, the observation of no significant differences in exercise endurance between the 10 repeated tests in the male mice may have occurred because the variation between tests were so large that statistical significance for repeated measures ANOVA may have been undetectable. This hypothesis is supported by the lack of agreement between repeated tests in both the male and female mice as shown by the large 95% limits in the Bland-Altman analysis (Fig. 3). Additionally, it is worth noting that none of the animals in this maximal treadmill protocol actually reached the end of the protocol before stopping; thus, variation in the endpoints of the protocol did not contribute to the overall variation observed. Therefore, although males and females were significantly different in average run time on the endurance tests, both sexes were similar in their lack of repeatability in this measure. The large variation in maximal exercise endurance we observed with repeated testing is relevant given the repeatable nature of maximal endurance testing in humans (8–10%) [28] and the growing number of studies that are using maximal endurance testing without repeatability monitoring to distinguish between exercise and pharmacological treatments in animals [30,42].

Voluntary Physical Activity (Running Wheel)

Our large cohort data, in addition to the available literature, suggest that physical activity as measured by running wheel activity in rodents is a repeatable phenotype (Fig. 4–Fig 5). Swallow and colleagues [35] reported high repeatability of running wheel activity on days 5

and 6 of measurement in selectively bred mice (n=287 females $r=0.787$; n=273 males $r=0.868$). In addition, Rezende and colleagues measured $VO_2\text{max}$ during wheel running in selectively bred female mice (n=48) and reported repeatability of $VO_2\text{max}$ measurements during running wheel activity as $r=0.844$ [31], indicating that both running wheel activity and $VO_2\text{max}$ achieved during wheel activity are repeatable. Similar to humans, levels of BDNF increase in the brain following voluntary physical exercise in mice [7], possibly helping to explain the repeatability of this phenotype in rodent models.

It is also warranted to speculate that the repeatability of wheel running in rodents is due to the voluntary and perhaps the innate nature of this activity. Rowland [34] described the idea of an intrinsic biological control of energy expenditure in animals. From an evolutionary standpoint, it would be beneficial for organisms to maintain energy balance, and he proposed this was done by an “activity-stat” mechanism. Rowland proposed several lines of evidence, including genetics, for this “activity-stat” mechanism which would theoretically work centrally to control amount of intrinsic physical activity, and thus, energy expenditure [34]. Supporting the hypothesis of an “activity-stat” is the observation that genetically different strains of mice differ in the level of voluntary wheel running [23]. Because this “activity-stat” would be regulated centrally and would be intrinsic to individual animals, this could explain why the measurement of voluntary physical activity has been shown to be repeatable in our data and the available rodent literature.

Conclusions

Although within strain variation appears stable over time when an open treadmill is used (Fig. 1), exercise endurance measurements using sealed treadmills repeated on the same mouse are not repeatable. Crabbe and colleagues [12] employed a well designed study to show that inbred strains of mice differ in behavioral phenotypes depending on the laboratory setting. Even though different technicians and slightly different laboratory settings were employed for the different cohorts of mice outline in Figure 1, these two strains, as groups, tested the same over time. The different reported values for repeatability of $VO_2\text{max}$ testing in rodents in the literature could be partially explained by the evidence presented by Crabbe and colleagues; however, in the current study, even when repeated maximal endurance testing in the same lab, under the same conditions, with the same technicians was employed (Table 1) the results suggest high variability in this behavioral test (Fig. 3). It may be possible to repeatedly endurance test rodents using other methods; however, the results of this study indicate using an enclosed treadmill with a shock grid for aversive stimuli that produces a negative stimulus to encourage mice to run to “exhaustion” is not a repeatable measure for assessing exercise endurance in mice. In contrast, daily physical activity as assessed by distance, duration, and speed on a running wheel appears highly repeatable in both inbred and outbred mice. The level of voluntary physical activity an animal performs appears to be both genetically and biologically regulated possibly influencing the high repeatability of this phenotype. The observations in this study are critical in considering results from current and future exercise behavior literature that investigates the role of various biological factors involved in the regulation of exercise behaviors in rodents.

Acknowledgments

The authors would like to thank all involved in the data collection of both endurance and wheel running data for this paper including Keeley Loiseau, Ellie Friesen, Jessica Moser, Paul Downey, Matt Yost, and Sarah Carter. We would also like to thank Sean Courtney for technical help and discussion involving repeatability of $VO_2\text{max}$ in rodents. We would like to thank Dr. Steve Kleeberger for his guidance and also Drs. Timothy Musch and Ed Howley for their reviews and comments on this manuscript. The project described was supported by grant numbers DK61635 and AR050085 from NIH NIDDK and NIAMS (J.T. Lightfoot, A.M. Knab, R.S. Bowen, T. Moore-Harrison, A.T. Hamilton, M.J. Turner).

Funding: The project described was supported by grant numbers DK61635 and AR050085 from NIH NIDDK and NIAMS (J.T. Lightfoot, A.M. Knab, R.S. Bowen, T. Moore-Harrison, A.T. Hamilton, M.J. Turner).

References

1. Adlard PA, Cotman CW. Voluntary exercise protects against stress-induced decreases in brain-derived neurotrophic factor protein expression. *Neuroscience* 2004;124:985–992. [PubMed: 15026138]
2. Aguiar AS Jr, Tuon T, Pinho CA, Silva LA, Andreazza AC, Kapczynski F, Quevedo J, Streck EL, Pinho RA. Mitochondrial IV complex and brain neurotrophic derived factor responses of mice brain cortex after downhill training. *Neurosci. Lett* 2007;426:171–174. [PubMed: 17904742]
3. Allen E. The oestrous cycle in the mouse. *The American Journal of Anatomy* 1922;30:297–371.
4. Barbato JC, Koch LG, Darvish A, Cicila GT, Metting PJ, Britton SL. Spectrum of aerobic endurance running performance in eleven inbred strains of rats. *J. Appl. Physiol* 1998;85:530–536. [PubMed: 9688730]
5. Bassett DR Jr, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med. Sci. Sports Exerc* 2000;32:70–84. [PubMed: 10647532]
6. Bedford TG, Tipton CM, Wilson NC, Oppliger RA, Gisolfi CV. Maximum oxygen consumption of rats and its changes with various experimental procedures. *J. Appl. Physiol* 1979;47:1278–1283. [PubMed: 536299]
7. Belke TW, Wagner JP. The reinforcing property and the rewarding aftereffect of wheel running in rats: a combination of two paradigms. *Behav. Processes* 2005;68:165–172. [PubMed: 15686826]
8. Bingisser R, Kaplan V, Scherer T, Russi EW, Bloch KE. Effect of training on repeatability of cardiopulmonary exercise performance in normal men and women. *Med. Sci. Sports Exerc* 1997;29:1499–1504. [PubMed: 9372488]
9. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat. Methods Med. Res* 1999;8:135–160. [PubMed: 10501650]
10. Copp SW, Davis RT, Poole DC, Musch TI. Reproducibility of endurance capacity and VO₂peak in male Sprague-Dawley rats. *J. Appl. Physiol* 2009;106:1072–1078. [PubMed: 19213934]
11. Coyle EF, Coggan AR, Hopper MK, Walters TJ. Determinants of endurance in well-trained cyclists. *J. Appl. Physiol* 1988;64:2622–2630. [PubMed: 3403447]
12. Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. *Science* 1999;284:1670–1672. [PubMed: 10356397]
13. Duman CH, Schlesinger L, Russell DS, Duman RS. Voluntary exercise produces antidepressant and anxiolytic behavioral effects in mice. *Brain Res* 2008;1199:148–158. [PubMed: 18267317]
14. Dumke CL, Rhodes JS, Garland T Jr, Maslowski E, Swallow JG, Wetter AC, Cartee GD. Genetic selection of mice for high voluntary wheel running: effect on skeletal muscle glucose uptake. *J. Appl. Physiol* 2001;91:1289–1297. [PubMed: 11509528]
15. Ferris LT, Williams JS, Shen CL. The effect of acute exercise on serum brain-derived neurotrophic factor levels and cognitive function. *Med. Sci. Sports Exerc* 2007;39:728–734. [PubMed: 17414812]
16. Festing MF. Wheel activity in 26 strains of mouse. *Lab. Anim* 1977;11:257–258. [PubMed: 926756]
17. Friedman WA, Garland T Jr, Dohm MR. Individual variation in locomotor behavior and maximal oxygen consumption in mice. *Physiol. Behav* 1992;52:97–104. [PubMed: 1529020]
18. Koch LG, Meredith TA, Fraker TD, Metting PJ, Britton SL. Heritability of treadmill running endurance in rats. *Am. J. Physiol* 1998;275:R1455–R1460. [PubMed: 9791061]
19. Koteja P, Garland T Jr, Sax JK, Swallow JG, Carter PA. Behaviour of house mice artificially selected for high levels of voluntary wheel running. *Anim. Behav* 1999;58:1307–1318. [PubMed: 10600154]
20. Kregel, KC.; Allen, DL.; Booth, FW.; Fleshner, MR.; Henrikson, EJ.; Musch, TI.; O'Leary, DS.; Parks, CM.; Poole, DC.; Ra'anan, AW.; Sheriff, DD.; Sturek, MS.; Toth, LA. Resource book for the design of animal exercise protocols. American Physiological Society; 2006. Exercise Protocols Using Rats and Mice.
21. Leamy L, Pomp D, Lightfoot JT. An epistatic genetic basis for physical activity traits in mice. *J. Hered.* 2008

22. Lerman I, Harrison BC, Freeman K, Hewett TE, Allen DL, Robbins J, Leinwand LA. Genetic variability in forced and voluntary endurance exercise performance in seven inbred mouse strains. *J. Appl. Physiol* 2002;92:2245–2255. [PubMed: 12015333]
23. Lightfoot JT, Turner MJ, Daves M, Vordermark A, Kleeberger SR. Genetic influence on daily wheel running activity level. *Physiol. Genomics* 2004;19:270–276. [PubMed: 15383638]
24. Lightfoot JT, Turner MJ, Debate KA, Kleeberger SR. Interstrain variation in murine aerobic capacity. *Med. Sci. Sports Exerc* 2001;33:2053–2057. [PubMed: 11740298]
25. Lightfoot JT, Turner MJ, Kleinfehn AM, Jedlicka AE, Oshimura T, Marzec JM, Gladwell W, Leamy L, Kleeberger S. Quantitative trait loci (QTL) associated with maximal exercise endurance in mice. *J. Appl. Physiol* 2007;103:105–110. [PubMed: 17412788]
26. Lightfoot JT, Turner MJ, Knab AK, Jedlicka AE, Oshimura T, Marzec J, Gladwell W, Leamy LJ, Kleeberger SR. Quantitative trait loci associated with maximal exercise endurance in mice. *J. Appl. Physiol* 2007;103:105–110. [PubMed: 17412788]
27. Lightfoot JT, Turner MJ, Pomp D, Kleeberger SR, Leamy LJ. Quantitative trait loci (QTL) for physical activity traits in mice. *Physiol. Genomics* 2008;32:401–408. [PubMed: 18171721]
28. McArdle WD, Katch FI, Pechar GS. Comparison of continuous and discontinuous treadmill and bicycle tests for max Vo₂. *Med. Sci. Sports* 1973;5:156–160. [PubMed: 4747636]
29. Moraska A, Deak T, Spencer RL, Roth D, Fleshner M. Treadmill running produces both positive and negative physiological adaptations in Sprague-Dawley rats. *Am J Physiol Regul Integr Comp Physiol* 2000;279:R1321–R1329. [PubMed: 11004000]
30. Narkar VA, Downes M, Yu RT, Emblar E, Wang YX, Banayo E, Mihaylova MM, Nelson MC, Zou Y, Juguilon H, Kang H, Shaw RJ, Evans RM. AMPK and PPARdelta agonists are exercise mimetics. *Cell* 2008;134:405–415. [PubMed: 18674809]
31. Rezende EL, Chappell MA, Gomes FR, Malisch JL, Garland T Jr. Maximal metabolic rates during voluntary exercise, forced exercise, and cold exposure in house mice selectively bred for high wheel-running. *J. Exp. Biol* 2005;208:2447–2458. [PubMed: 15939783]
32. Rhodes JS, Garland T Jr, Gammie SC. Patterns of brain activity associated with variation in voluntary wheel-running behavior. *Behav. Neurosci* 2003;117:1243–1256. [PubMed: 14674844]
33. Richter CP. Animal behavior and internal drives. *The Quarterly Review of Biology* 1927;2:307–343.
34. Rowland TW. The biological basis of physical activity. *Med. Sci. Sports Exerc* 1998;30:392–399. [PubMed: 9526885]
35. Swallow JG, Carter PA, Garland T Jr. Artificial selection for increased wheel-running behavior in house mice. *Behav. Genet* 1998;28:227–237. [PubMed: 9670598]
36. Swallow JG, Garland T Jr, Carter PA, Zhan WZ, Sieck GC. Effects of voluntary activity and genetic selection on aerobic capacity in house mice (*Mus domesticus*). *J. Appl. Physiol* 1998;84:69–76. [PubMed: 9451619]
37. Swallow JG, Koteja P, Carter PA, Garland T. Artificial selection for increased wheel-running activity in house mice results in decreased body mass at maturity. *J. Exp. Biol* 1999;202:2513–2520. [PubMed: 10460738]
38. Turner MJ, Kleeberger SR, Lightfoot JT. Influence of genetic background on daily running-wheel activity differs with aging. *Physiol. Genomics* 2005;22:76–85. [PubMed: 15855385]
39. Ways JA, Cicila GT, Garrett MR, Koch LG. A genome scan for Loci associated with aerobic running capacity in rats. *Genomics* 2002;80:13–20. [PubMed: 12079278]
40. Ways JA, Smith BM, Barbato JC, Ramdath RS, Pettee KM, DeRaedt SJ, Allison DC, Koch LG, Lee SJ, Cicila GT. Congenic strains confirm aerobic running capacity quantitative trait loci on rat chromosome 16 and identify possible intermediate phenotypes. *Physiol. Genomics* 2007;29:91–97. [PubMed: 17179209]
41. Wergel-Kolmert U, Wisen A, Wohlfart B. Repeatability of measurements of oxygen consumption, heart rate and Borg's scale in men during ergometer cycling. *Clin. Physiol. Funct. Imaging* 2002;22:261–265. [PubMed: 12402448]
42. Wisloff U, Najjar SM, Ellingsen O, Haram PM, Swoap S, Al-Share Q, Fernstrom M, Rezaei K, Lee SJ, Koch LG, Britton SL. Cardiovascular risk factors emerge after artificial selection for low aerobic capacity. *Science* 2005;307:418–420. [PubMed: 15662013]

43. Yashiro M, Kimura S. Effect of voluntary exercise on physiological function and feeding behavior of mice on a 2% casein diet or a 10% casein diet. *J. Nutr. Sci. Vitaminol. (Tokyo)* 1979;25:23–32. [PubMed: 448475]

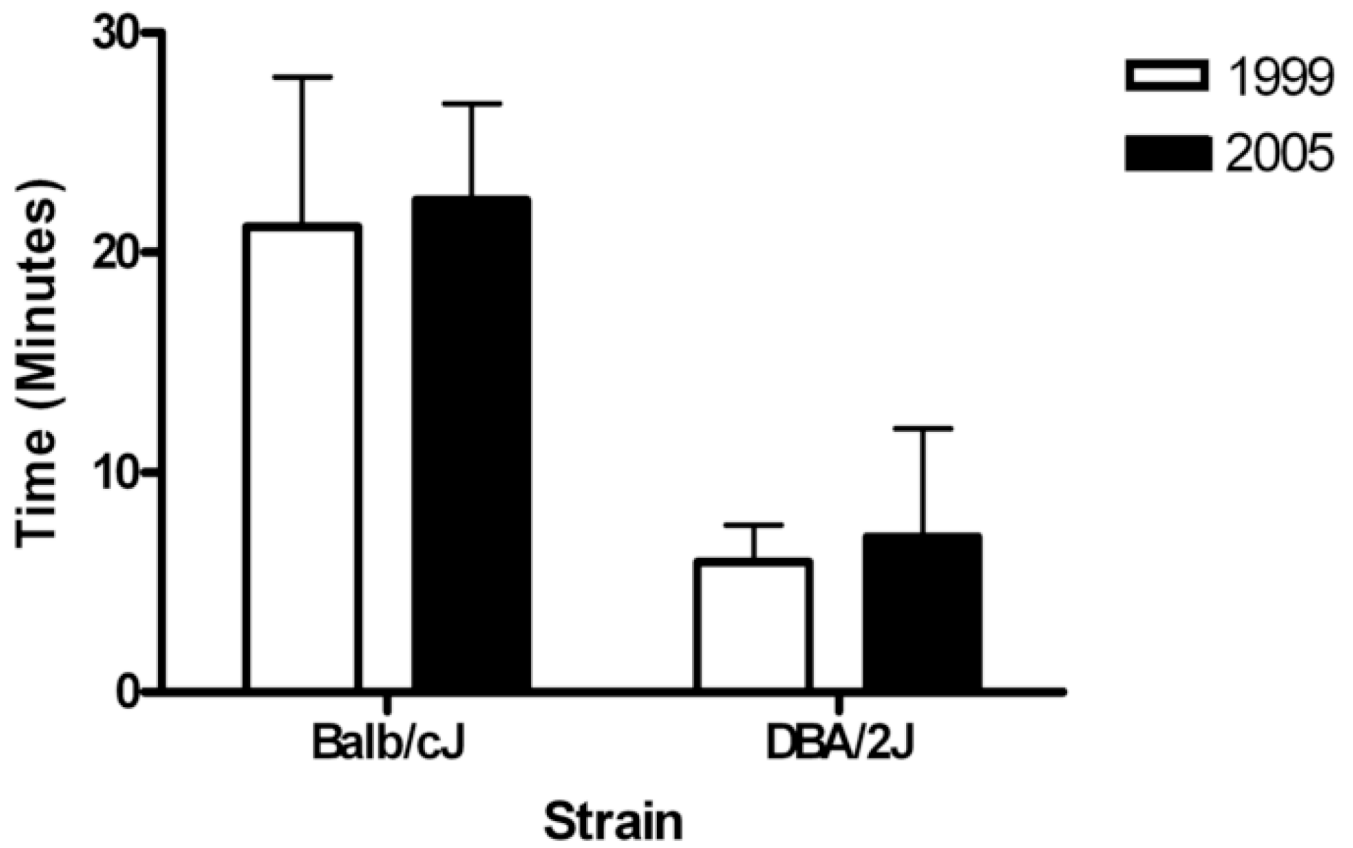
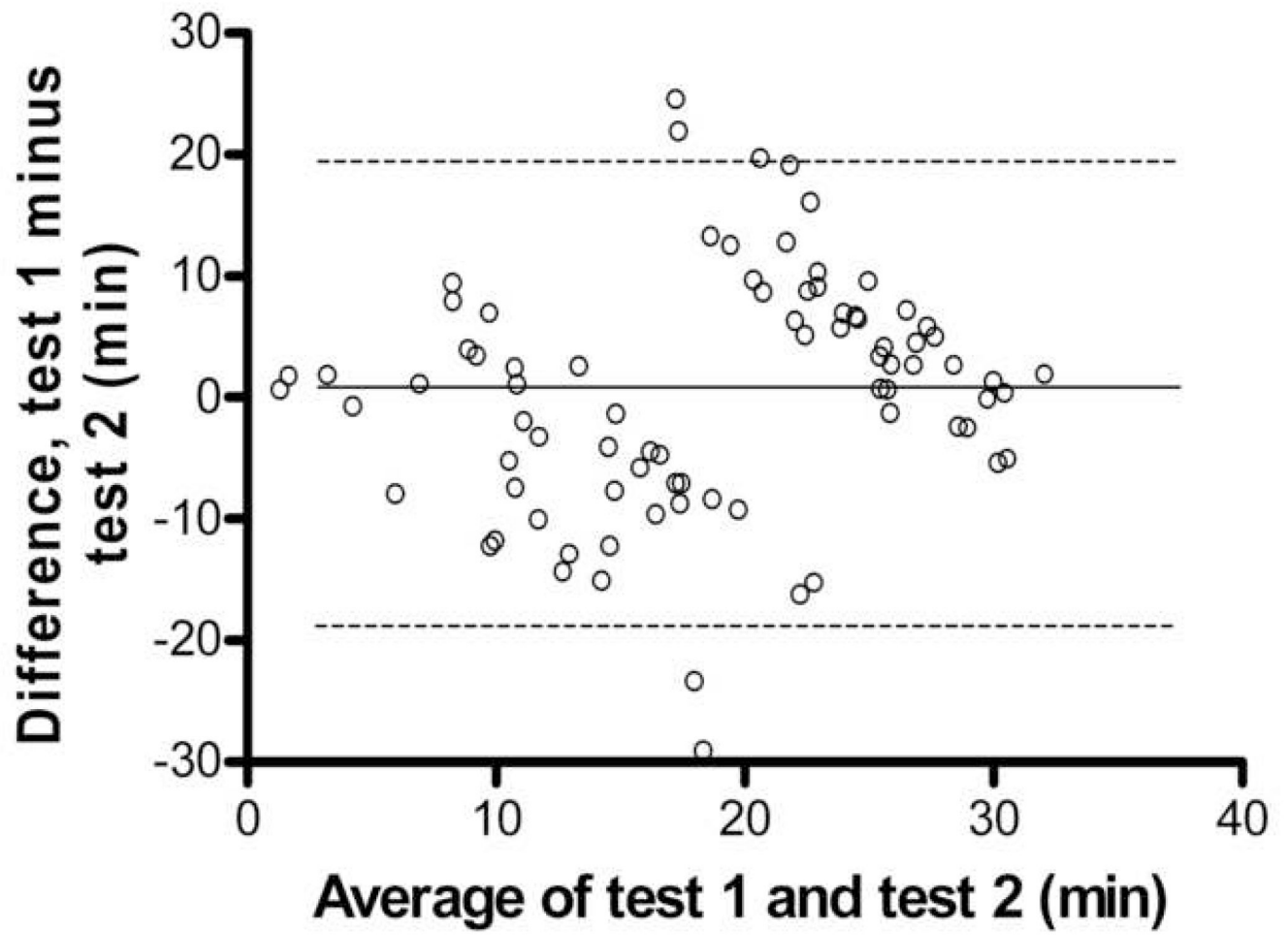


Figure 1. Average Time (and standard deviations) in minutes of two different cohorts of Balb/cJ mice and DBA/2J mice separated by 5 years. No significant differences were found between years within either strain (Balb/cJ mice, $p=0.55$; DBA/2J mice, $p=0.51$).



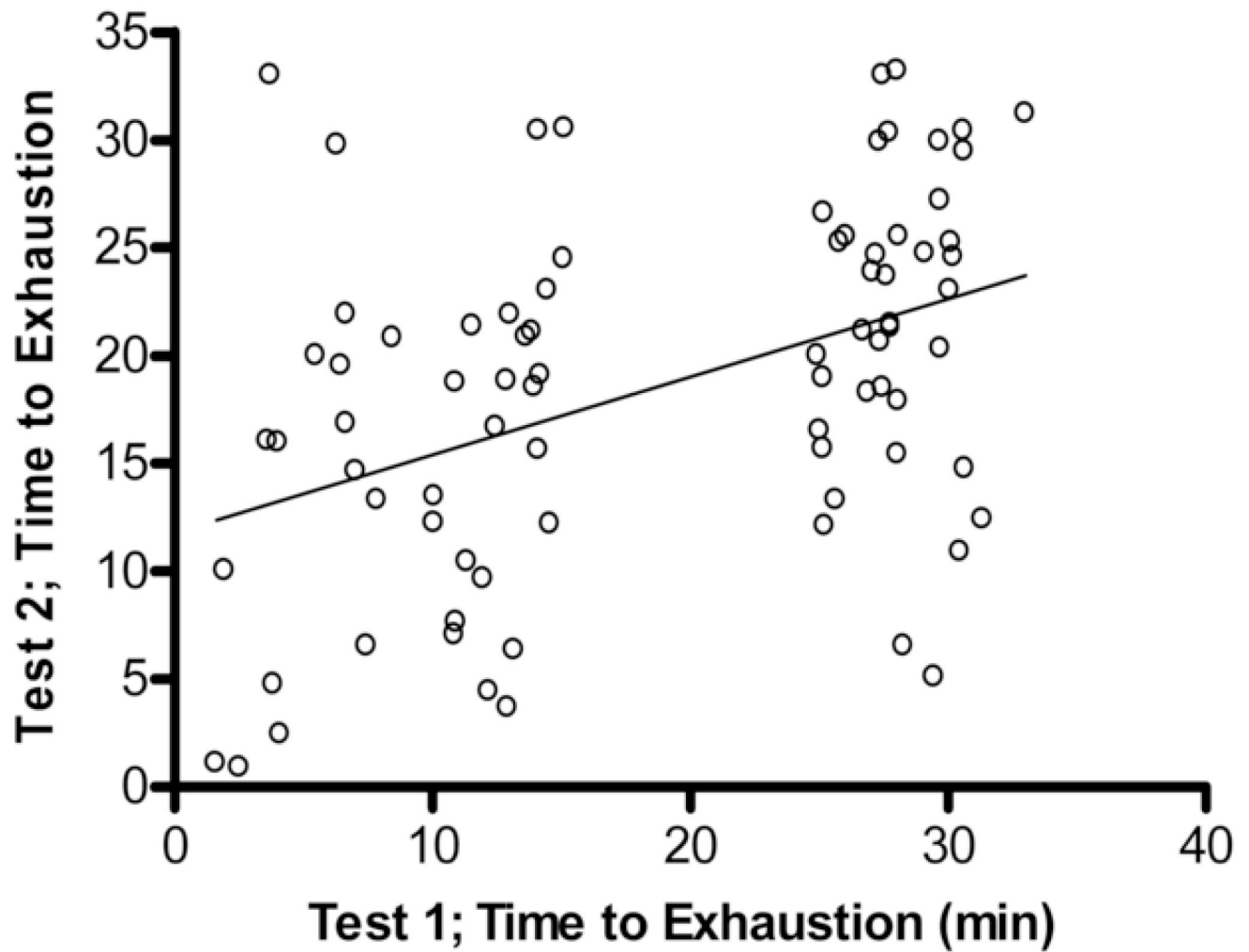
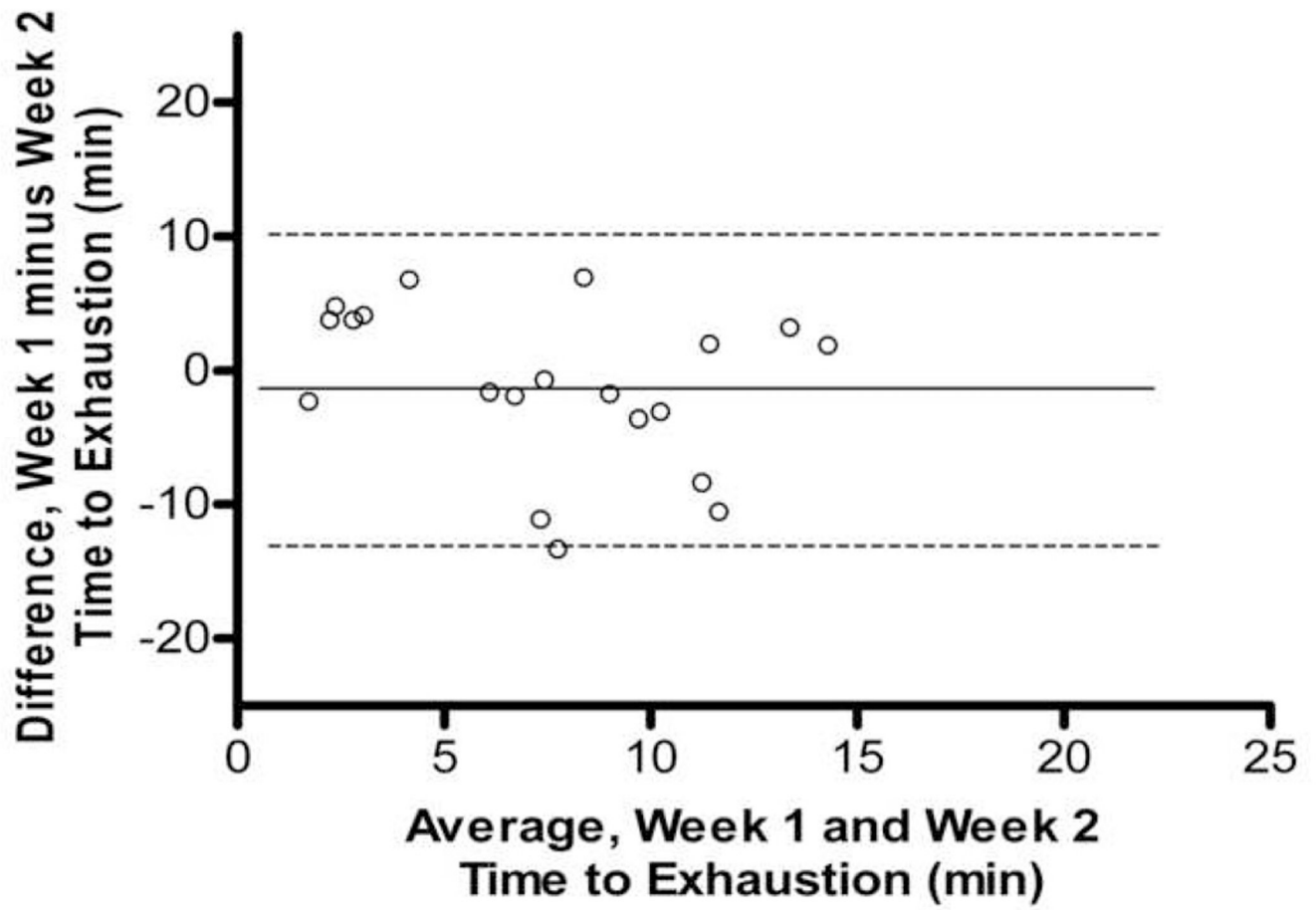
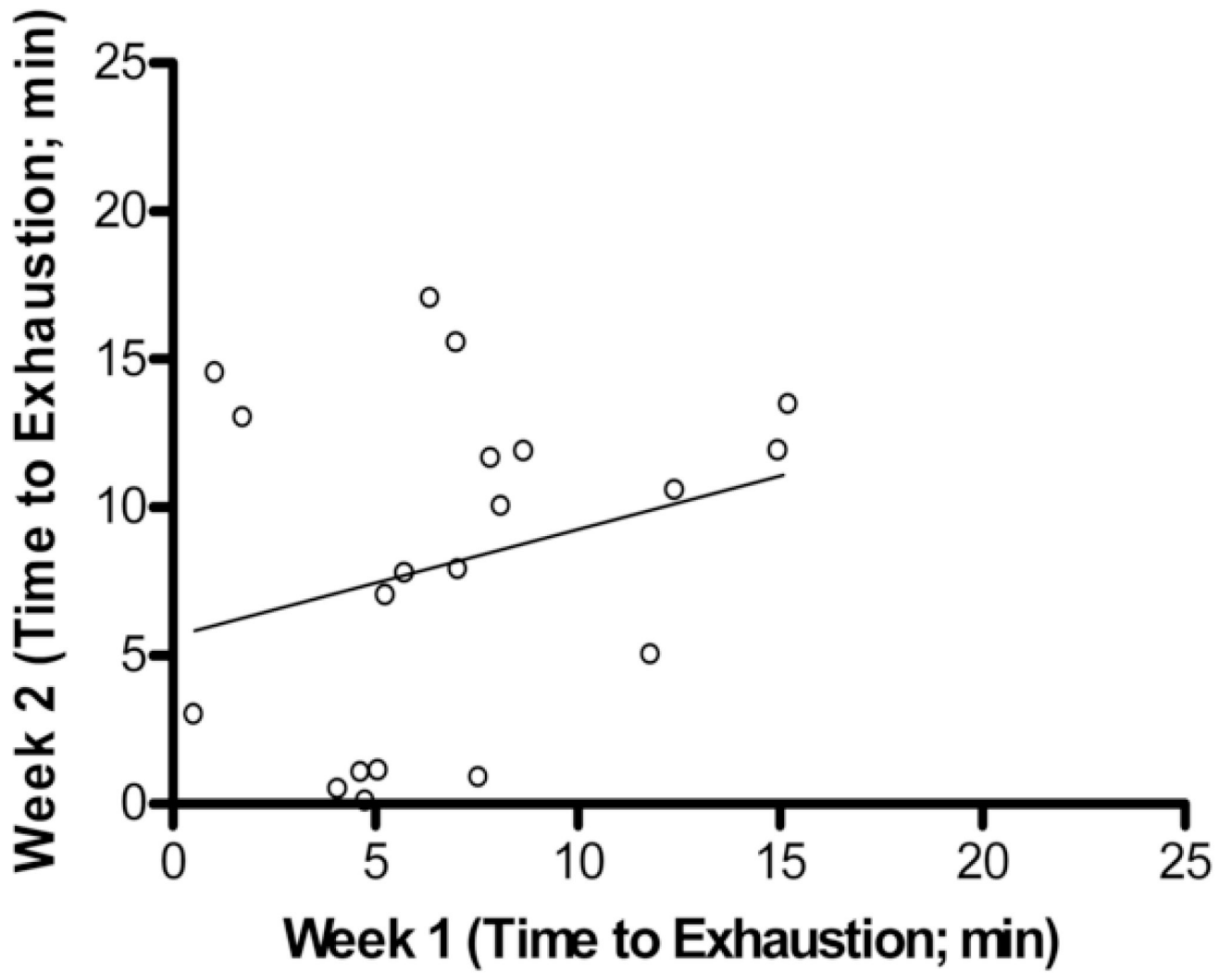
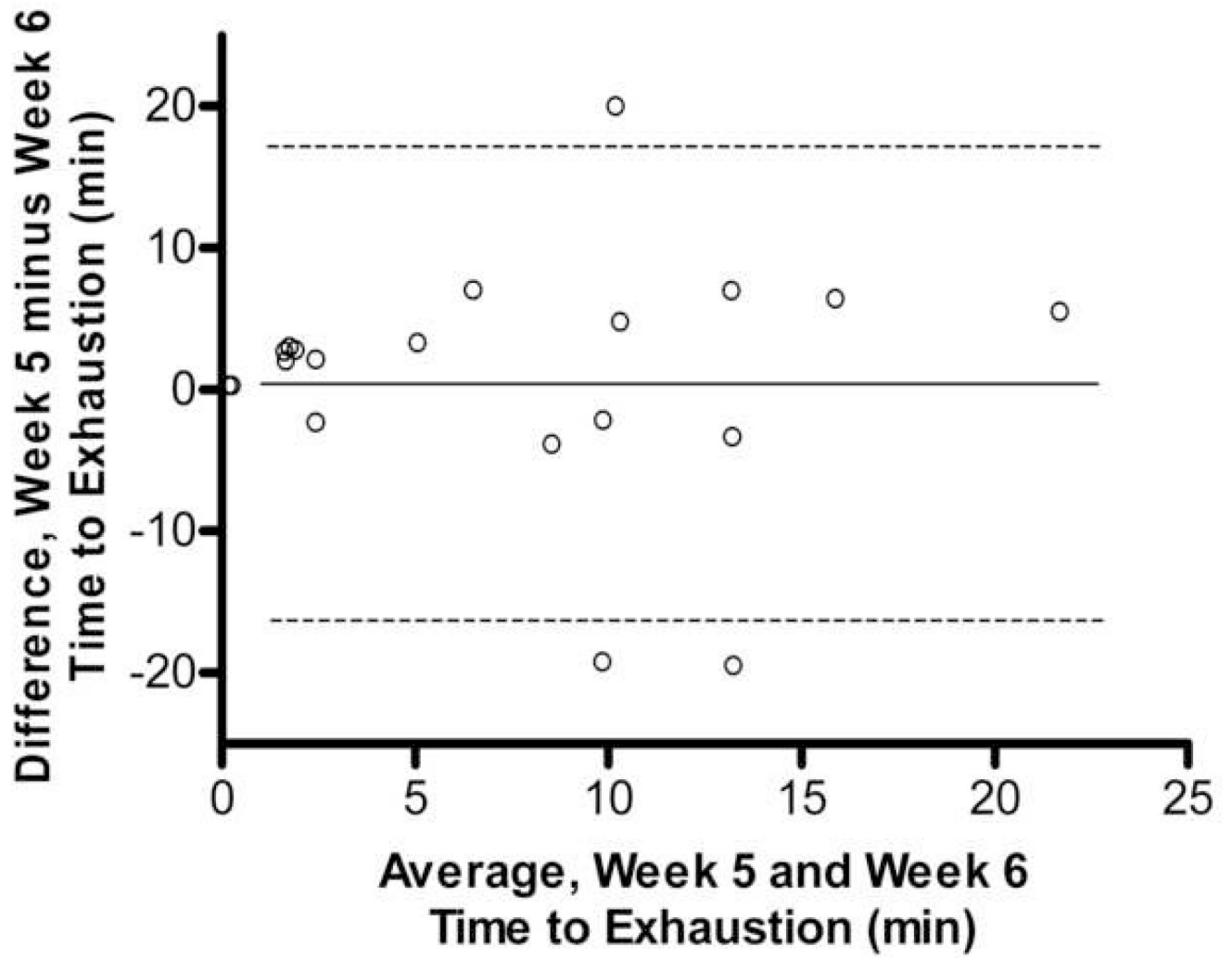
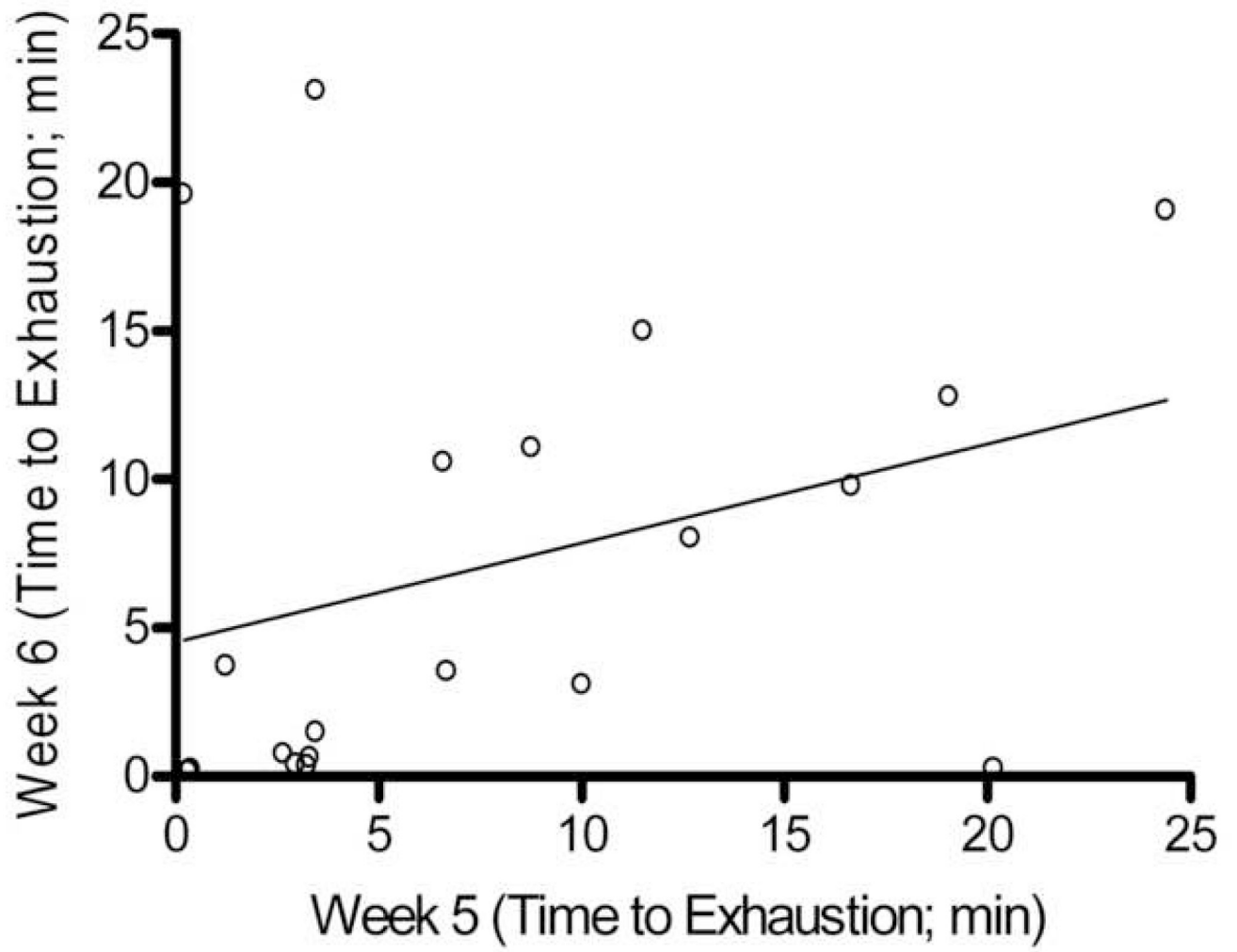


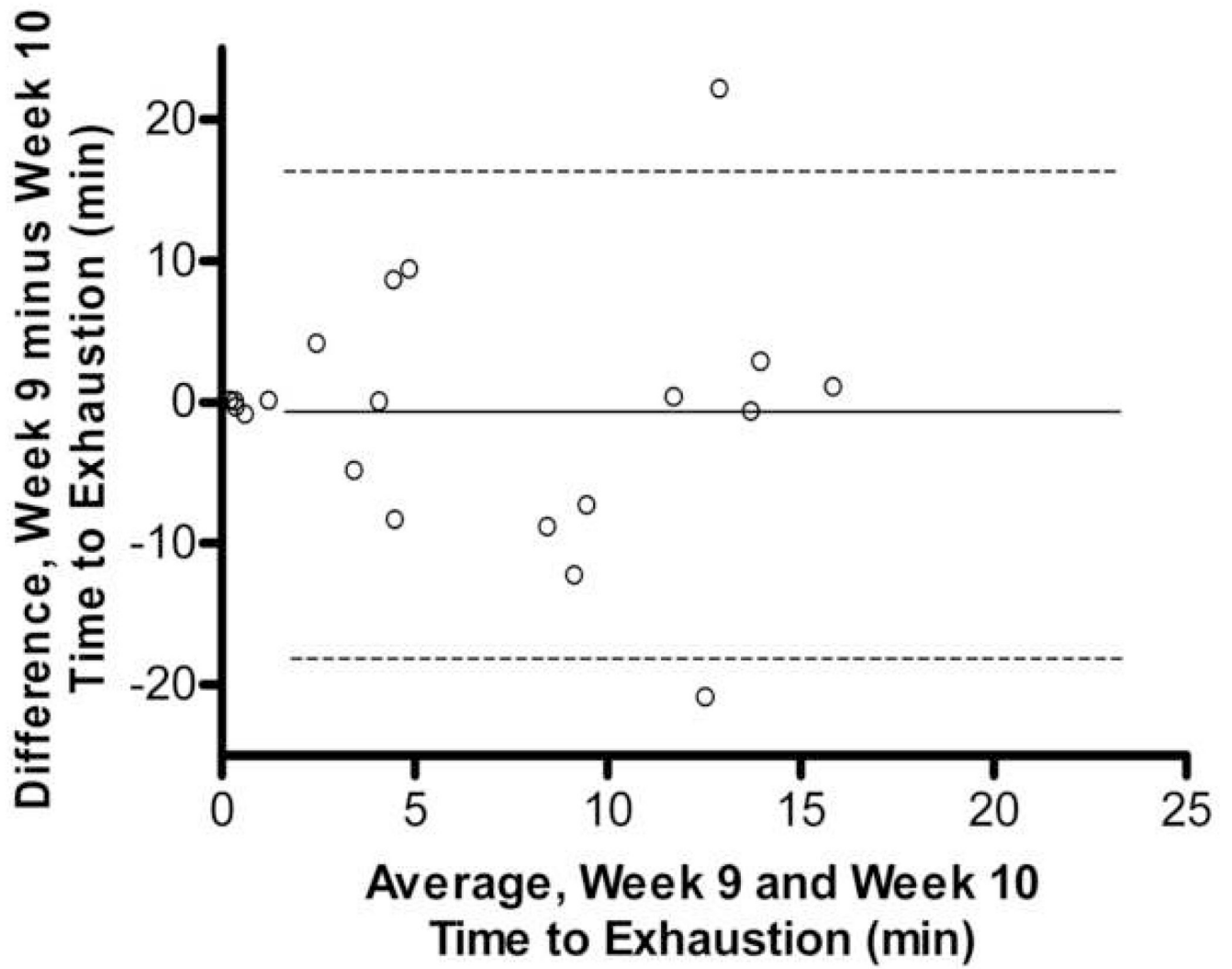
Figure 2. Bland-Altman Analysis for repeated exercise endurance testing. **A.** Difference between endurance test 1 and endurance test 2 versus average of endurance test 1 and endurance test 2 with the 95% limits of agreement (dashed lines) and mean difference (solid line) between tests test 1 and test 2. **B.** Scatterplot test 1 versus test 2, $r = 0.43$, $p > 0.05$.











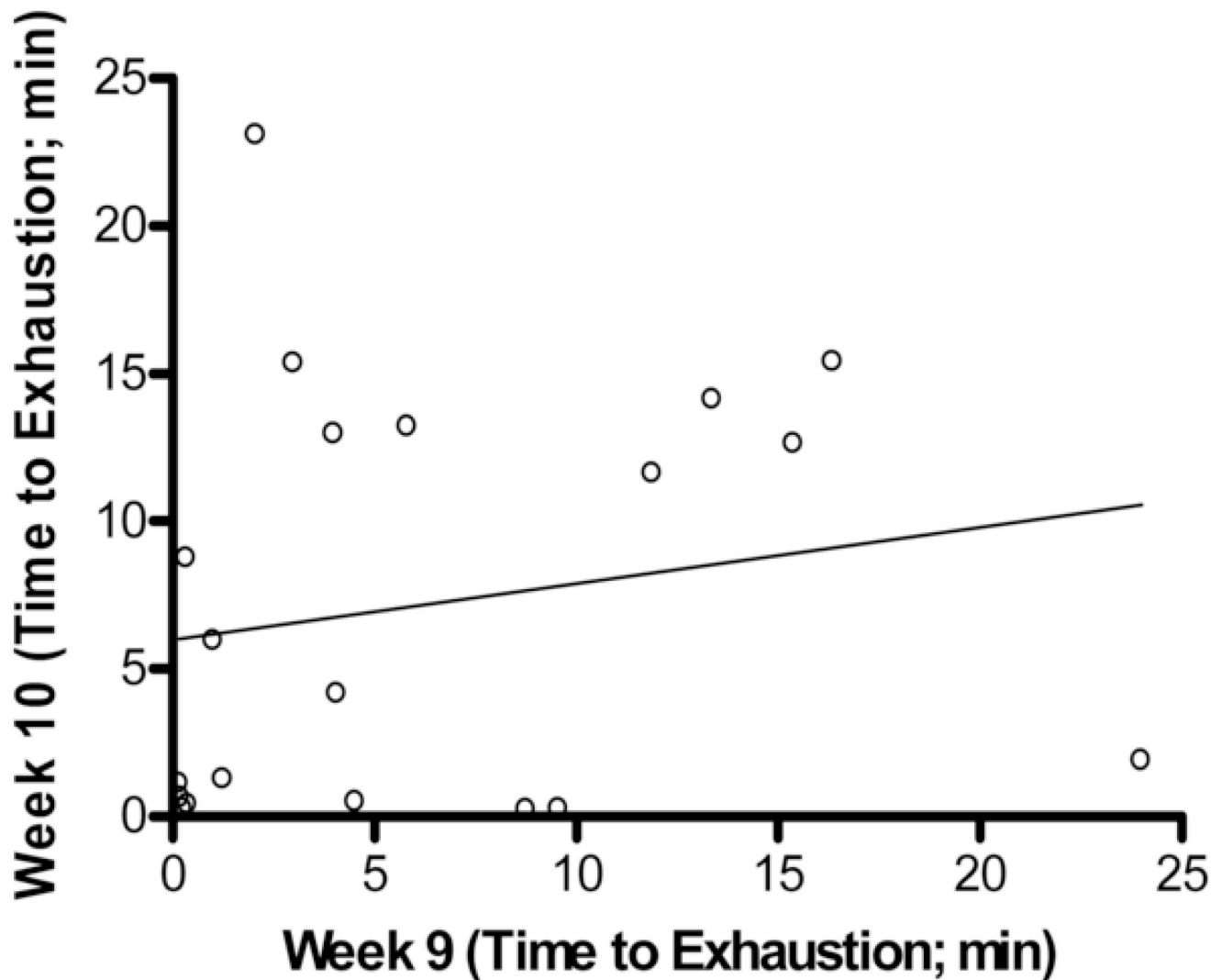


Figure 3.

Bland-Altman Analysis for repeated exercise endurance testing across 10 weeks. **A.** Difference between endurance test during week 1 and week 2 versus average of endurance test during week 1 and week 2 with the 95% limits of agreement (dashed lines) and mean difference (solid line) between endurance tests. **B.** Scatterplot of endurance test during week 1 versus endurance test during week 2, $r = 0.27$, $p > 0.05$. **C.** Difference between endurance test during week 5 and week 6 versus average of endurance test during week 5 and week 6. **D.** Scatterplot of endurance test during week 5 versus endurance test during week 6, $r = 0.33$, $p > 0.05$. **E.** Difference between endurance test during week 9 and week 10 versus average of endurance test during week 9 and week 10. **F.** Scatterplot of endurance test during week 9 versus endurance test during week 10, $r = 0.18$, $p > 0.05$.

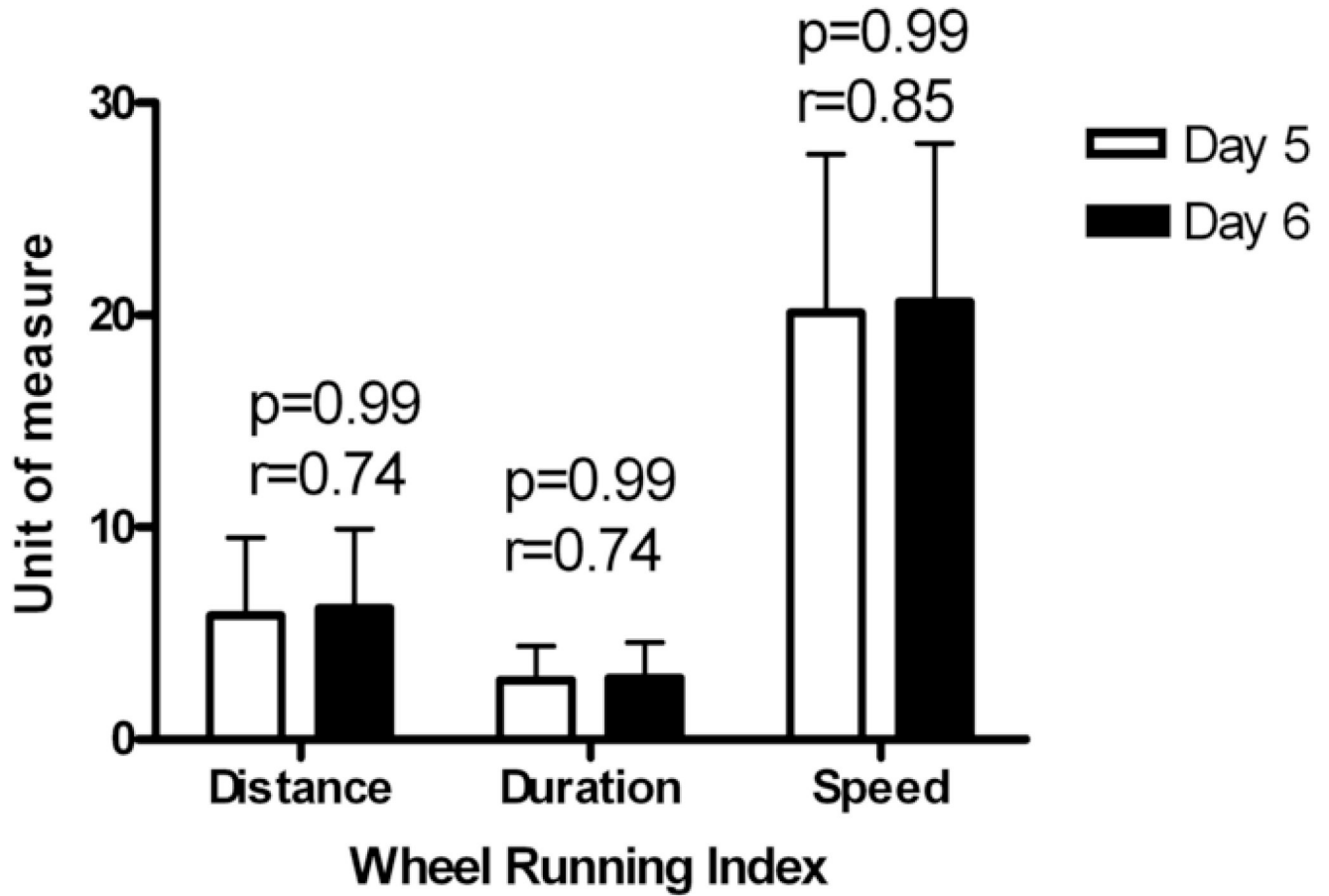


Figure 4.

Comparison of wheel running indices between day 5 and 6 of wheel running exposure in inbred and outbred mice (n=739). No significant differences were found between the two days of measurement for any index (Distance, Km/day; duration, min/day*100; and speed, m/min). Correlation values are also reported for each index.

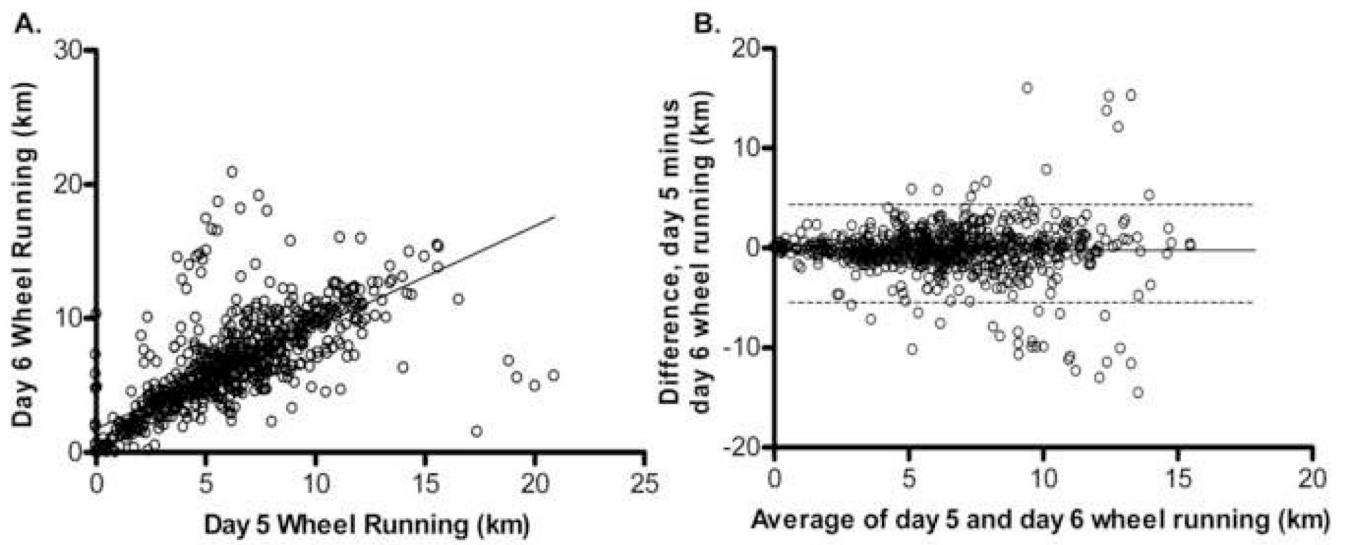


Figure 5. Bland-Altman Analysis for repeated wheel running distance measurements on days 5 and 6 of wheel running. **A.** Difference between distance run on day 5 and day 6 versus average distance run on day 5 and day 6 with the 95% limits of agreement (dashed lines) and mean difference (solid line) between test 1 and test 2. **B.** Scatterplot of distance run on day 5 versus day 6, $r = 0.74$, $p = 0.99$.

Table 1
Correlation Values of Endurance for Each Week of Testing in 20 Balb/c/J mice

Ten Balb/c/J male and ten Balb/c/J female mice were endurance tested once a week for ten weeks. No differences in association between max endurance tests were attributed to sex, thus matched pairs correlation values are shown in the table for all mice for all ten weeks. Lack of consistency is apparent by the lack of pattern of significant correlations. Two-way ANOVA showed no significant differences in male mice across weeks ($p=0.33$), but there were significant differences in female mice across weeks ($p=0.04$) (data not shown).

Week	1	2	3	4	5	6	7	8	9	10
1	---	.27	.34	-.15	.23	.02	-.08	-.08	-.04	.25
2		---	.35	.35	.21	.16	.47*	.03	.05	.41
3			---	.08	.44*	.12	.30	.10	-.08	.34
4				---	.16	.75***	.76***	.66***	.41	.38
5					---	.33	.30	.10	.22	.44
6						---	.51*	.52*	.62**	.33
7							---	.55*	.24	.51*
8								---	.42	.47*
9									---	.18
10										---

* significance at $p<0.05$

** significance at $p<0.0001$