Original Research

Psychometrics of the Balance Beam Functional Test in C57BL/6 Mice

Melissa C Orenduff,[1,2,*] Erika T Rezeli,[1] Stephen D Hursting,[1,3] and Carl F Pieper[4]

Aging is associated with a progressive decline in physical function characterized by decreased mobility, which is an important risk factor for loss of independence and reduced quality of life. Functional testing conducted in animals has advanced our understanding of age-related changes in physical ability and contributed to the development of physiologic measurements that can be used to assess functional changes during aging. The balance beam test is one assessment tool used to measure age-related changes in balance and coordination. The goal of this study is to provide analytical examples and psychometric support of a protocol that has been analyzed to show how the number of successive test runs, foot slips, pauses, and hesitations affect the reliability of the primary outcome measure, which is the time to cross the beam. Our results suggest that conducting more than 1 training session, consisting of greater than or equal to 3 successful training runs, followed by at least one test session with no less than 2 successful runs (that is, runs without pauses or hesitations) provides a psychometrically sound outcome. The data presented here indicate that a psychometric approach can improve protocol design and reliability of balance beam measures in mice.

Abbreviations: ICC, interclass correlation coefficient

The world's population is aging. For the first time in history, the number of people who are 65 y of age or older has surpassed the number of children under 5 y of age.[26] Current population projections estimate that by 2050, one in 6 people in the world will be 65 y of age or older.[26] As this aging of the population intensifies, so does the number of older adults with physical limitations.[19] These limitations in physical function impair the ability to perform activities of daily living, reduce the quality of life, and increase the risk of falling.[9,18] One of the most important causes of reduced physical function is age-related loss of skeletal muscle mass.[27,29,30] Emerging evidence suggests the loss of skeletal muscle mass is not the only contributing factor to the decline in physical performance in older adults.[23] For instance, loss of muscle strength, power, and function, distinct features of aging muscle, are strongly associated with reduced mobility and physical ability in older adults.[23]

Our current understanding of physical function across the continuum of age-related changes has been largely guided by studies conducted in animal models of aging.[13,17] Rodents (primarily rats and mice) are the most commonly used experimental animals in aging research due to their relatively short developmental period and life span as compared with humans.[13] Many phenotypes and functional metrics associated with biologic aging are conserved across species.[1] Laboratory-based motor measurements have provided nume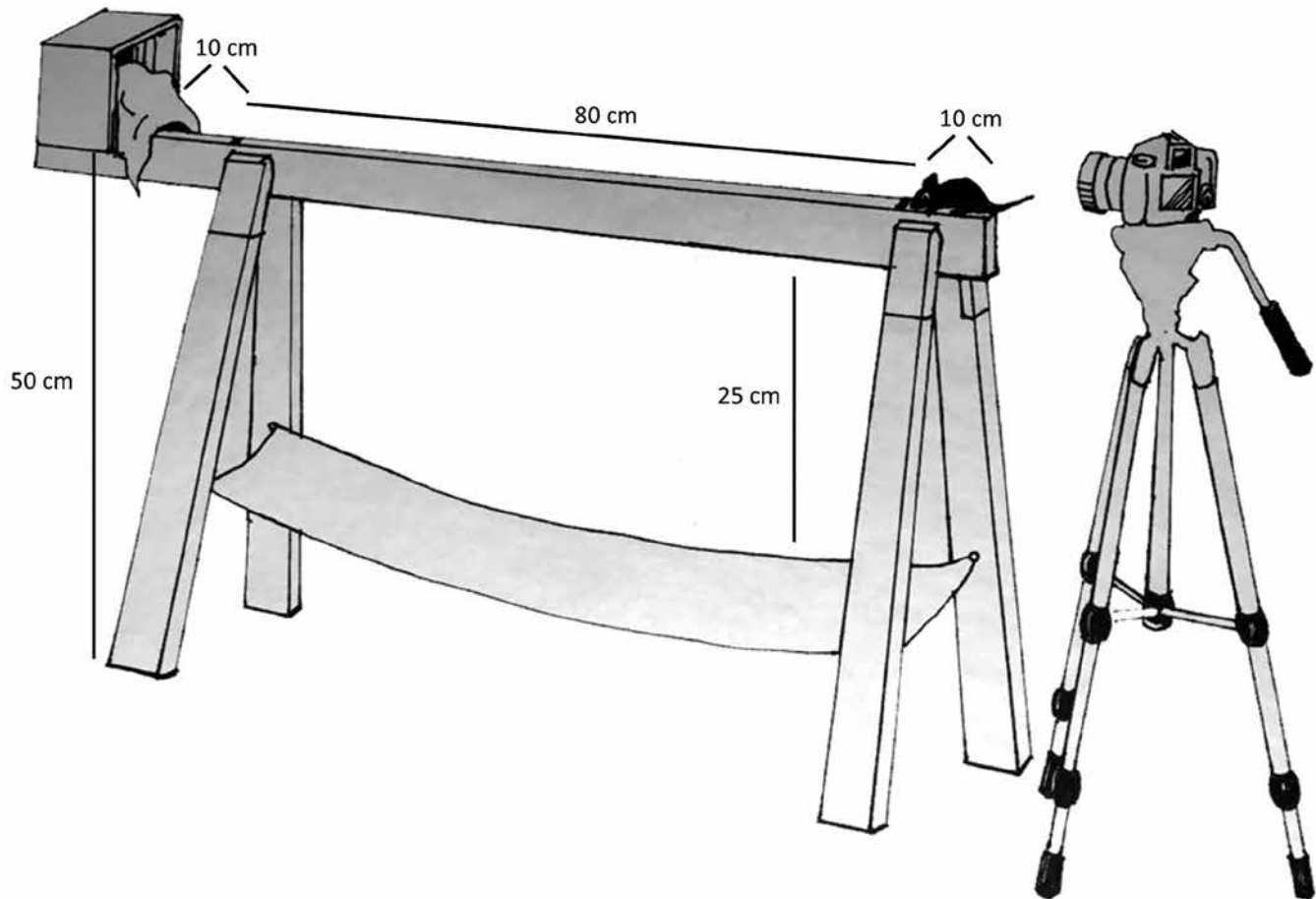rous assessments of age-related changes in physical function in rodents.[2,5,10,11] Often referred to as behavioral testing because of cognitive influences on physical ability, these tests measure functional status in rodents by assessing a spectrum of physical performance metrics including balance, coordination, agility, and strength.[12] Of these functional assays, the balance beam test is a well-established method used to assess several parameters of motor capabilities and gait-related activity in mice.[4,15] The balance beam apparatus is a simple construct, requires minimal equipment and is easy to assemble (Figure 1). Test measures typically consist of time to cross the beam and the number of foot slips. Performance on the beam allows an assessment of balance-related motor skills, provides a functional marker of locomotor differences between young and aged animals, and permits comparison of the effects of age, disease, and therapeutic interventions between cohorts with regard to motor function, coordination, and balance.[4,6,24,25,28]

Over time, the mechanics and experimental protocol of the balance beam test have undergone several modifications. One of the first applications of measuring balance and coordination was performed to characterize the locomotor skills of aged animals.[28] Early balance beam experiments tested the ability of the animal to remain on a narrow rod during a single 3-min interval. The test was conducted over 3 consecutive days, with latency to fall as the outcome variable.[4] More recent versions of the balance beam test employ flat boards or round rods suspended horizontally or inclined. The test requires mice to be trained to traverse the beam without being over-trained, as this can desensitize the innate motivation to escape.[15] Testing is conducted within 1 or 2 d after training or within a period of time during which the mouse is likely to retain memory of the learned task.[6] Measures obtained from the balance beam include (1) the amount of time required to travel a constant

Figure 1. Balance beam apparatus. A wooden beam (12 mm and 6 mm) is suspended 50 cm above the ground. To encourage the mouse to cross the beam to enter a secure location, a safe house with bedding is placed at the exit end of the beam. A net provides a soft landing in case the mouse falls from the beam. The camera at the end of the beam captures any foot slips as the mouse crosses the beam.

distance on the beam (typically 80 cm), (2) the number of foot slips encountered (captured on video recordings), and, in some cases, (3) the number of pauses or nudges required for the mouse to complete the task. In theory, the quantitative components of the balance beam test are relatively straightforward to measure when the animal completes a successful test run by crossing the beam without pausing midway or hesitating at the starting line or before entering the safe house at the end. However, due to unavoidable complications resulting from pauses and hesitations, which are considered intrinsic features of animal testing using the balance beam, the reliability of the primary outcome measure (that is, time to cross) may be questionable. Although thorough and well-crafted balance beam protocols provide useful instructional models,[4,6,8] a standardized analytical strategy is not available to account for potential performance inconsistencies that inherently arise in animal functional testing.

Our goal in this study is to provide a pragmatic, psychometrically based analysis to assess the reliability of measures between one successful as run compared with 2 other successful runs and to demonstrate the impact of pauses, hesitations, and foot slips on time to cross the beam. This analysis may provide a more consistent and standardized use of the balance beam test as a tool to assess balance and coordination in mice.

Materials and Methods

Test subjects. Ethics and procedures concerning the use and care of animals were approved and monitored by the Institu-tional Animal Care and Use Committee at University of North Carolina at Chapel Hill.

Eight-mo-old male (n = 45) and female (n = 45) C57BL/6J mice were purchased from the Jackson Laboratory (Sacramento, CA). Mice were housed and allowed to acclimate for an extended period of time (approximately 8 wk). This acclimation period allowed the animals to become fully adjusted to the environment and diet and to maintain stable body mass. During the acclimation period, 2 female mice were euthanized due to neurologic illness. The remaining mice (n = 43 female, n = 45 male) were included in the study. When testing began, mice were 10 mo of age, a developmental phase that is operationally defined in mice as middle-aged.[27] We selected this age to identify the transitional physiologic changes in balance and coordination that occur between young and old mice.

Mice were singly housed in standard Techniplast Green Line cages (floor area = 501 cm^2/77.6 in^2) (Tecniplast USA, West Chester, PA) and provided with Crink-l'Nest α-dry bedding, crinkle paper enrichment, (The Andersons, Goldsboro, NC), and a polycarbonate mouse housing tent (Datesand Limited, United Kingdom). Mice were maintained on a 12:12 light:dark cycle with lights on at 0700 and lights out at 1900 and had free access to water and a purified control diet (D12450J) formulated by Research Diets, New Brunswick, NJ. Body composition was conducted 7 d prior to the first training session by using quantitative magnetic resonance imaging (qMRI) (UNC Nutrition
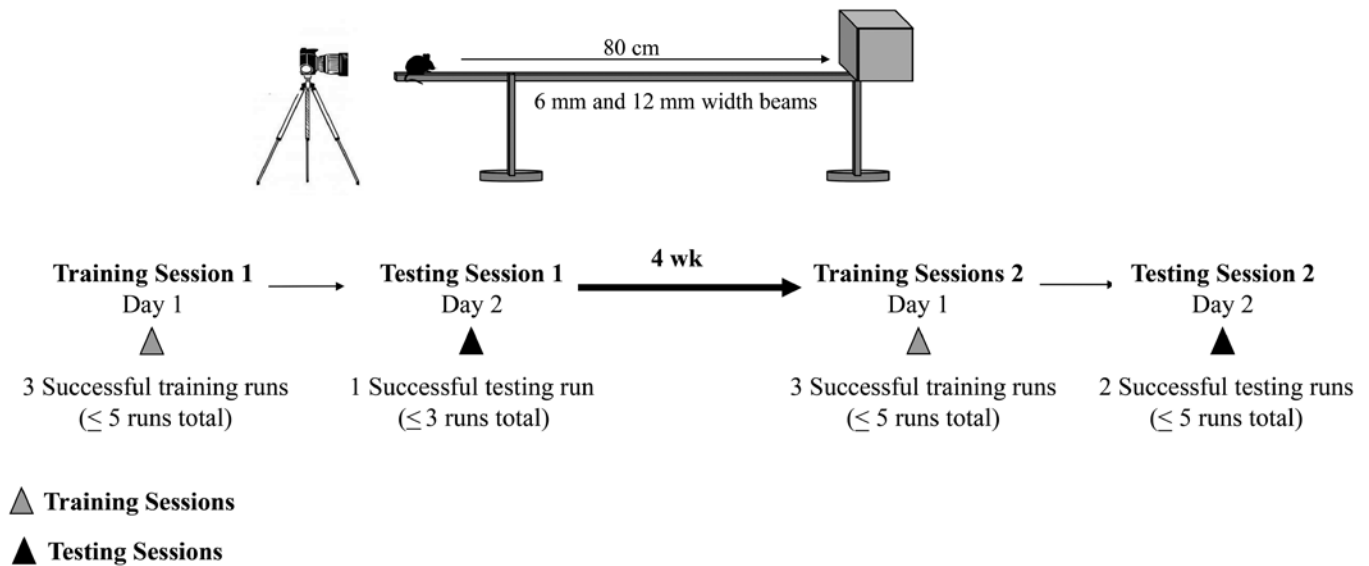
Figure 2. Schedule for training and test sessions. Training sessions are indicated with gray triangles (fx1) and Test sessions are indicated with black triangles (fx2).

Obesity Research Center: Animal Metabolism Phenotyping Core, Chapel Hill, NC).

Procedure. The procedures used in this study were adapted from protocols developed previously.[4,6] Briefly, the balance beam apparatus (Figure 1) is composed of 2 smooth, wooden beams 1 m in length and 6 mm or 12 mm in width. The beams are securely suspended 50 cm above the floor using sturdy trestles. Enclosed safe houses are placed at the escape ends of each of the 2 beams and bedding is added to encourage the mouse to enter. To prevent injury to mice should they fall off the beam, a net is tethered to the trestles 25 cm below the beam. A distance of 80 cm was marked on the beam, indicating the target distance for which the mouse is tested, with an extra distance of 10 cm behind the starting line to allow a space for placement of the mouse on the beam and 10 cm after the finish line in case the mouse pauses or hesitates right before entering the safe house. Mice that pause or hesitate are retested up to 3 times in test session 1 and up to 5 times in test session 2 on each beam (Figure 2). A stopwatch is used to time the mice from the start line to the finish line, and a video camera records foot slips. The video camera is placed on a tripod 50 cm from the starting end of the beam and positioned to view the mouse from behind. As illustrated in Figure 1, this position allows the cameras' point of view to capture the entire length of the beam and enable the investigator to observe right and left foot slips.

Training and testing sessions were conducted between 1000 and 1400 (the normal resting phase of nocturnal animals). However, in consideration of this interruption of their innate sleep/active cycle, mice were brought into the testing room 30 min prior to training and testing sessions to allow them time to fully awaken and acclimate to the test environment. Training and testing schedules are illustrated in Figure 2.

Training sessions. Training sessions were conducted the day before test sessions 1 and 2. Mice were trained on the 12 mm beam followed by the 6 mm beam. For the first training run, mice were placed on the beam (10 cm behind the starting line) and were guided as needed by light nudges from the investigator's index finger. This guidance was only provided if the mouse tried to reverse its direction or paused before the finish line. For subsequent training runs, the investigator used guidance only if the mouse required additional training; otherwise, the investigator

did not interfere with the mouse while completing the distance on the beam. Training sessions concluded after the mouse had completed 3 successful training runs (that is, free of pauses and hesitations) on both the 12 mm and 6 mm beams. A subset of mice that appeared to be unmotivated or otherwise reluctant to traverse the beam required additional training; however, no training sessions consisted of more than 5 training runs. Mice were given 30 s to rest in the safe house between training runs. After completing the training runs, mice were placed back in their home cages and returned to the mouse housing room.

Test session 1. The objective of test session 1 was to obtain one successful test run (no pauses or hesitations) on both the 12 mm and 6 mm beams with no more than 3 runs total per beam. During the test sessions, mice did not receive any motivational guidance and were allowed to rest in the safe house for 30 s in between runs. Mice were first tested on the 12 mm followed by the 6 mm beam. A stopwatch was used to measure the amount of time required to traverse the beam. Test runs were video recorded for later assessments of foot slips by the investigators.

Test session 2. The objective of test session 2 was to obtain 2 successful test runs (no pauses or hesitations) on both the 12 mm and 6 mm beams with no more than 5 runs total per beam. Four weeks after Test Session 1, mice were trained again as previously described. On the following day, mice completed test session 2. A stopwatch was used to measure the amount of time required to traverse the beam, and video recordings captured foot slips that were later assessed by investigators.

Statistical analysis. Means and standard deviations (SDs) were calculated separately for 12 mm and 6 mm beam widths based on test session(s) and run(s), unless noted otherwise. Statistical differences between and within each of the test sessions and runs were tested by mixed models to control for the inherent correlation of repeat measures, and all tests were performed at a Type-I error rate of 0.05. If the omnibus test showed that at least one variable was significant between groups (test session or sex), follow-up pairwise group comparisons were performed using contrasts.

The Spearman correlations between the 3 run measures (1 run in test session 1, 2 runs in test session 2); were calculated initially. Using these correlations, the reliability of a single random measure of time was calculated by classic test theory formulas for the Interclass correlation coefficient (ICC).[20] Because of differences

in the average time to cross between runs, the classic ICC formula (ICC = σ^2(mouse)/(σ^2(mouse) + σ^2(error))) was modified to incorporate the run effect (ICC = σ^2(mouse)/(σ^2(mouse) + σ^2(run) + σ^2(error))).[22]

The ICCs were calculated in 2 formats —(a) 2 test sessions, 3 runs (one in test session 1, 2 in test session 2); and (b) 1 test session, 2 runs. The latter was performed because the measured times in test session 1, run 1 were very different, both in level and correlation, than the 2 runs in session 2. Under both formats, the reliability was calculated by the Spearman–Brown prophesy formula:[16] $R_k = KR/(1 + (K − 1)R)$, where K = the amount of increase (for example, if K = 2, the number of measurements are doubled), R is the reliability of the baseline measure, and R_K is the resulting predicted reliability, to derive estimates of the reliability for the average of 2 compared with 4 runs.

Finally, using only the time to cross measures from successful runs, mixed model regression was employed to assess the impact of mouse and experimental protocol on the time measurement. In addition to time to cross and test session 1 or 2, number of foot slips, body mass, sex, and percent fat mass were measured and included to assess the impact of these variables. Separate regression analyses were performed on data collected from the 12 mm and 6 mm beams. Finally, to show the impact of pauses and hesitations on the time to cross measure, these regressions were performed again including all measures for both successful runs and runs with pauses and hesitations.

The data analysis for this paper was performed using SAS software, Version 9.4. Copyright 2021 SAS Institute, Cary, NC, USA. Figure 3 A and B were generated using GraphPad Prism 9, Version 9.0.2.

Results

Ten-month-old male and female C57BL/6J mice (*n* = 88) were used in this study (see Table 1 for characteristics of the mice). Measurements include time to cross, pauses, hesitations, and foot slips in 2 test sessions. In the first test session, a single

Table 1. Descriptive characteristics of the mice and Spearman correlation coefficients of time to cross: 12 mm and 6 mm Beams (*n* = 88)

Characteristics	Male	Female
Sex (n)	45	43
Body mass, g mean (SD)	32.5 (2.6)	24.3 (2.3)
Fat mass, g mean (SD)	4.3 (2.2)	3.3 (1.7)

successful run was required on both the 6 mm and 12 mm beams, with up to 3 runs total per beam. In the second test session, 2 successful runs were required for each of the beams, with a limit of 5 runs total per beam.

Test session 1. For the 12 mm beam, 76% of the mice (*n* = 67) completed successful runs on the first trial. Twenty of 21 (95%) of the mice that required a second trial were successful, while the remaining mouse was successful on the third trial. For the 6 mm beam, 56% (*n* = 49) were successful on the first trial. Among those (*n* = 39) that required a second run, 92% (*n* = 36) were successful. The remaining 3 mice were successful on the third run. Overall, 27% and 39% of runs were excluded due to pauses or hesitations on the 12 mm and 6 mm beams, respectively.

Test session 2. For the 12 mm beam, 56% (*n* = 50) completed 2 successful runs in the first 2 trials. An additional 36% (*n* = 32) required 3 runs, while the remaining 6 animals required 4 (*n* = 5) or 5 (*n* = 1) trials to complete 2 successful runs. For the 6 mm test, 61% (*n* = 54) performed 2 successful runs in 2 trials, whereas 27% (*n* = 24) required 3 trials. Of the remaining 10 animals, 7 required 4 trials, and 3 required 5 trials. At the conclusion of test session 2, 21% and 41% of trials were excluded due to pauses or hesitations on the 12 mm and 6 mm beams, respectively.

Differences between test sessions and successful test runs. Table 2 shows test session-dependent differences between successful testing runs for male and female mice. The average time (s) to cross for a successful run was substantially longer in test session 1 for both the 12 mm (3.90 s) and 6 mm (4.85 s) beams than either of the successful runs in test session 2 for the 12 mm (3.63 and 3.28 s) and 6 mm (4.59 and 4.66 s) beams for runs 1 and 2, respectively. In addition, the correlation coefficients of time
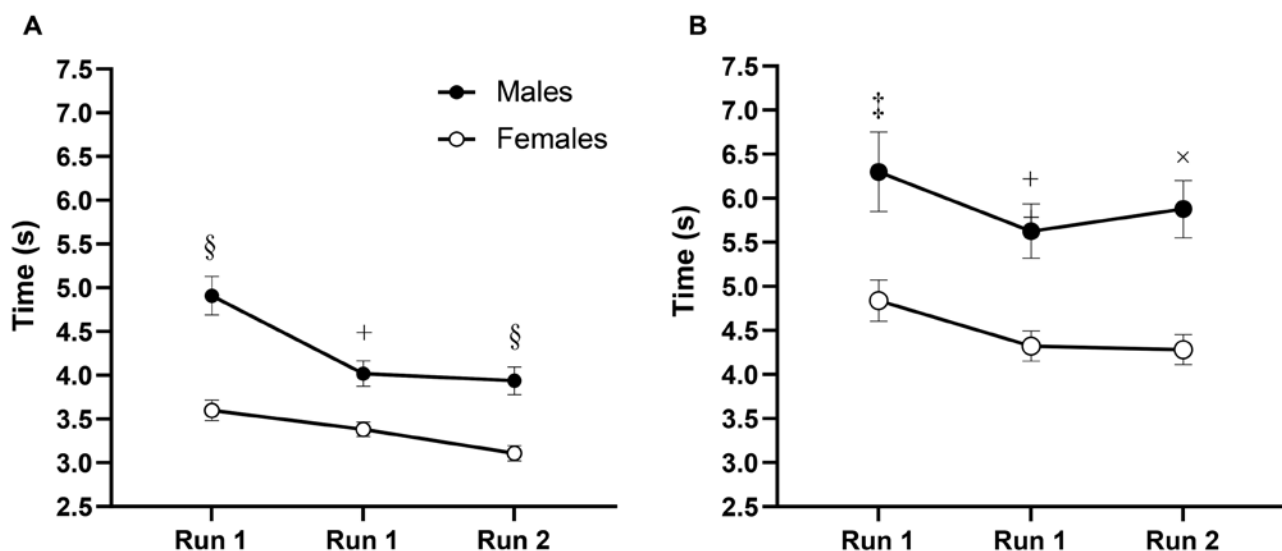
Figure 3. Time to Cross 12mm and 6mm Beams by Sex. (A) A time by sex interaction was observed for the 12mm beam (*P* = 0.017) and main effect for sex at each run. Females were significantly faster than males for all runs (session 1, run 1 (*P* < 0.0001), session 2, run 1 (*P* = 0.004), and session 2 run 2 (*P* = 0.0001). (B) Time to cross the 6mm beam. The main effect for sex was significant at each run. Female mice were significantly faster for all runs (session 1, run 1 (*P* = 0.001), session 2, run 1 (*P* = 0.005), and session 2 run 2 (*P* = 0.0004). Male mice are indicated by black filled circle (fx3) and female mice by white filled circle (fx4). Results shown are the means and ± SE ‡, *P* ≤ 0.001; §, *P* ≤ 0.0001; +, *P* ≤ 0.005; ×, *P* ≤ 0.0005

Table 2. Spearman correlation coefficients of time to cross: 12 mm and 6 mm beams (*n* = 88)

| | Test sessions and successful runs | | |
| | Session 1 | Session 2 | |
Variable	Run 1	Run 1	Run 2
12 mm beam			
Median (IQR)	3.90 (3.28, 4.97)	3.63 (3.11, 3.91)	3.28 (2.89, 3.94)
Spearman correlations			
Session 1 run 1	1.00		
Session 2 run 1 (*P* value)	0.394 (0.0001)	1.00	
Session 2 run 2 (*P* value)	0.433 (<0.0001)	0.712 (<0.0001)	1.00
6 mm beam			
Median (IQR)	4.85 (4.05, 6.14)	4.59 (3.78, 5.94)	4.66 (3.73, 5.76)
Spearman correlations			
Session 1 run 1	1.00		
Session 2 run 1 (*P* value)	0.470 (<0.0001)	1.00	
Session 2 run 2 (*P* value)	0.472 (<0.0001)	0.780 (<0.0001)	1.00

to cross between test sessions 1 and 2 for the 12 mm beam (run 1, 0.39; run 2, 0.43) and the 6 mm beam (run 1, 0.47; run 2, 0.47) were substantially lower than the correlation of time to cross for runs 1 and 2 in test session 2 for both the 12 mm (0.71) and 6 mm (0.78) beams. That is, runs 1 and 2 in test session 2 had a higher correlation than did either of the runs in test session 2 with test session 1. Male and female mice showed a significant time by sex interaction for differences in time to cross (*P* = 0.017). Latency to cross for female mice was significantly lower for each session and run that for male mice on both the 12mm (Figure 3 A) and 6mm (Figure 3 B) beams.

With these results in mind, we performed an analysis to assess the reliability of the summary measures from the 1 successful test run from test session 1 as compared with the 2 successful runs in test session 2. Shown in Table 3, we computed the Inter-Class Correlation using all successful runs (*n* = 3) in test session 2, and for test session 1 with 2 successful runs. The results from test session 1 did not meet the assumption of parallel measures (equal means and variances across all measures) (Table 3). Thus, we extended the model and added a random effect for 'run', irrespective of test session. However, the reduced correlation between test sessions 1 and 2 were substantively lowered in the ICC analysis when test session 1 data were included, and even the ICCs were lowered. Both ICCs were above the 0.80 threshold level if the results of test session 1 were not incorporated.[21] However, to achieve an estimated ICC above 0.8 with test session 1 data included, 2 test sessions and 2 successful runs would be required for the 6 mm beam, while 3 successful runs would be required for the 12 mm beam.

Regression analysis was conducted to support reliability estimates presented in this paper. Using MRI data, the results in Table 4 show predictors of time to cross for the 12 mm and 6 mm beams. Of the protocol factors measured (including session, run, pauses, hesitations, foot slips, sex, and fat mass), the test session was again related to time. Test session 1 times differed from test session 2 times (avg. of 2 runs) by nearly 0.7 s (*P* < 0.0001) for the 12mm beam, and approximately 0.5 s (*P* = 0.003) for the 6mm beam. Among individual factors for the 6 mm beam, sex was an important predictor in initial analyses including only sessions, runs, and sex. In addition, when all variables were added, body mass and foot-slips were related to time to cross, with higher body mass and more foot-slips associated with increased time.

Differences between test sessions and successful test runs with pauses and hesitations included. Using a repeated measures design to analyze all runs from test sessions 1 and 2, including both those completed successfully and those with pauses and hesitations, we found pauses and hesitations significantly lengthened time to cross. For test session 1, a pause/hesitation led to an increase of 1.2 s (*P* = 0.0002) for the 12mm beam and an increase of 1.6 s (*P* = 0.002) for 6mm beam. For test session 2, which required 2 successful runs, pauses and hesitations were associated with a 2.24 s increase in time (*P* = 0.0001) for the 12 mm beam test and a 0.88 s increase (*P* = 0.0144) for the 6 mm beam.

Table 3. Estimated InterClass Correlation Coefficients (ICCs) under various designs: 12 mm and 6 mm beams (*n* = 88)

Effect	β	*P* value	Overall *P* value
12 mm Beam			
Sex			0.464
Male	Ref.	Ref.	
Female	−0.369	0.464	
Test session			<0.0001
Session 1 run 1	0.696	<0.0001	
Session 2 run 1	0.126	0.282	
Session 2 run 2	Ref.	Ref.	
Body mass	0.064	0.259	0.259
Fat mass	−0.00427	0.863	0.863
Foot slips	0.161	0.125	0.125
6 mm beam			
Sex			0.208
Male	Ref.	Ref.	
Female	1.439	0.208	
Test session			0.003
Session 1 run 1	0.477	0.011	
Session 2 run 1	−0.142	0.450	
Session 2 run 2	Ref.	Ref.	
Body mass	0.342	0.008	0.009
Fat mass	−0.095	0.094	0.094
Foot slips	0.034	0.023	0.023

Ref, the reference group.

Table 4. Regression estimates: 12 mm and 6 mm beams ($n = 88$)

| Input protocol | Results for 1 test session: 2 runs (test session 2) | | Results for 2 test sessions: 1 run (test session 1) 2 runs (test session 2) | | |
	ICC 1 run	ICC 1 session /2 runs	ICC 1 session /1 run	ICC 2 sessions 1 run/ session	ICC 2 sessions 2 runs/ session
12MM					
Expected ICC	0.811	0.896	0.428	0.600	0.749
6 MM					
Expected ICC	0.786	0.880	0.632	0.775	0.872

Discussion

The balance beam is an important functional test used to detect changes in motor coordination and balance in animals. Here, we set out to evaluate current protocols and method development by testing the reliability of the time to cross as related to the number of test sessions and runs and the impact of foot slips, hesitations, and pauses. We found that the measures from test session 1 differed substantively from those obtained in test session 2. Test session 1 differed from test session 2 in average time to cross and in the correlation analysis, and these differences, in turn, negatively affected the estimate of reliability. This lower time to cross in test session 2 may indicate a practice effect, with decreasing time to cross as the number of runs increase from session 1 to session 2. However, the lower correlation coefficient suggests that the practice effect is not systematic across the animals and may or may not be important. From a methodological standpoint, this indicates that either practice should occur until the times are stabilized or, if the practice effects persist only during the period of time in which the mouse remembers the tasks, then measurements should contain both early and later trials. Our data suggest that reliability will improve by reducing the effect of pauses and hesitations on time to cross if the study used greater than or equal to 1 training session(s) consisting of greater than or equal to 3 successful training runs, followed by greater than or equal to 1 test session with no fewer than 2 successful test runs that are void of pauses and/or hesitations. Further, the association between foot slips and increased time to cross suggests that controlling for number of foot slips in the statistical analysis or including foot slips in the experimental design by repeating test trials with foot slips. In comparison to previous studies that used the balance beam to detect changes in balance and coordination in aging mice,[3,25] our findings suggest that the precision of those results may have been affected by failure to incorporate the confounding effects of pauses and hesitations in the analytical approach.

With respect to the effect of sex on latency to cross the beam, we observed a sex by time to cross interaction, suggesting that females completed the task faster than males. One possibility we considered for these differences is the impact of estrous cycles on female mice. At 10 mo of age, estrous cycles may potentially influence motor and behavioral performance and increase variability among female mice.[7,14] However, the error bars were not different by sex; thus, we conclude that estrous cycles did not substantially influence time to cross for female mice. A second possibility for the differences in time to cross the beam between male and female mice is body mass, and in fact, when the sex differences were controlled for body mass, the sex effect was no longer statistically significant. This result suggests that body mass is an important mediator of the sex/time to cross effect and may be an important covariate in the analytical approach.

This study has some limitations. Based on a previous study that that demonstrated over-training on the balance beam may encourage increased exploratory-like behaviors including increased and longer pauses and hesitations,[15] we used 2 training sessions and 2 test sessions. However, by not attempting additional training or test sessions, we were unable to comment on the potential effects of over-training. In fact, if addition test sessions had been conducted, the mice may have continued to improve their performance (decreased time to cross) or further stabilized their run times, thus increasing measures of reliability.

Functional performance testing in mice is a developing field and the measures have direct application to functional assessments in human aging research. The development of a standardized analysis to account for potential performance inconsistencies and confounding factors has yet to be established. By evaluating these measurement issues in functional testing in mice, our goal was to provide a methodological and analytical approach to improve the reliability and precision of outcome measures when using the balance beam test to assess coordination and balance in mice.

The results of this study speak to the design of balance beam protocol and the importance of using a robust psychometric approach in the methodological development of outcome measures, including number of successful runs, and considering potential confounding variables in the analysis approach. Integrating the results from this study or employing a psychometric component to the protocol development may assist in deriving efficient and optimal measures when using the balance beam test in animal models. Researchers designing a protocol for balance and coordination in mice may want to incorporate pauses and hesitations and other potential confounding variables such as sex and body mass into the design and analysis.

Acknowledgments

References

1. **Ackert-Bicknell CL, Anderson LC, Sheehan S, Hill WG, Chang B, Churchill GA, Chesler EJ, Korstanje R, Peters LL.** 2015. Aging research using mouse models. Curr Protoc Mouse Biol **5:**95–133. https://doi.org/10.1002/9780470942390.mo140195.

2. **Altun M, Bergman E, Edström E, Johnson H, Ulfhake B.** 2007. Behavioral impairments of the aging rat. Physiol Behav **92:**911–923. https://doi.org/10.1016/j.physbeh.2007.06.017.

3. **Battilana F, Steurer S, Rizzi G, Delgado AC, Tan KR, Handschin C.** 2020. Exercise-linked improvement in age-associated loss of balance is associated with increased vestibular input to motor neurons. Aging Cell **19:**1–13. https://doi.org/10.1111/acel.13274.

4. **Brooks SP, Dunnett SB.** 2009. Tests to assess motor phenotype in mice: a user's guide. Nat Rev Neurosci **10:**519–529. https://doi.org/10.1038/nrn2652.

5. **Carter CS, Sonntag WE, Onder G, Pahor M.** 2002. Physical performance and longevity in aged rats. J Gerontol A Biol Sci Med Sci **57:**B193–B197. https://doi.org/10.1093/gerona/57.5.B193.

6. **Carter RJ, Morton J, Dunnett SB.** 2001. Motor coordination and balance in rodents. Curr Protoc Neurosci **Chapter 8:**Unit 8.12. https://doi.org/10.1002/0471142301.ns0812s15.

7. **Chari T, Griswold S, Andrews NA, Fagiolini M.** 2020. The stage of the estrus cycle is critical for interpretation of female mouse social interaction behavior. Front Behav Neurosci **14:**1–9. https://doi.org/10.3389/fnbeh.2020.00113.

8. **Deacon RMJ.** 2013. Measuring motor coordination in mice. J Vis Exp **e2609:** doi: 10.3791/2609

9. **Dunsky A.** 2019. The effect of balance and coordination exercises on quality of life in older adults: a mini-review. Front Aging Neurosci **11:**1–10. https://doi.org/10.3389/fnagi.2019.00318.

10. **Forster MJ, Lal H.** 1999. Estimating age-related changes in psychomotor function: influence of practice and of level of caloric intake in different genotypes. Neurobiol Aging **20:**167–176. https://doi.org/10.1016/S0197-4580(99)00041-X.

11. **Ingram DK, Reynolds MA.** 1986. Assessing the predictive validity of psychomotor tests as measures of biological age in mice. Exp Aging Res **12:**155–162. https://doi.org/10.1080/03610738608259454.

12. **Justice JN, Carter CS, Beck HJ, Gioscia-Ryan RA, McQueen M, Enoka RM, Seals DR.** 2014. Battery of behavioral tests in mice that models age-associated changes in human motor function. Age (Dordr) **36:**583–592. https://doi.org/10.1007/s11357-013-9589-9.

13. **Justice JN, Cesari M, Seals DR, Shively CA, Carter CS.** 2015. Comparative approaches to understanding the relation between aging and physical function. J Gerontol A Biol Sci Med Sci **71:**1243–1253. https://doi.org/10.1093/gerona/glv035.

14. **Kopp C, Ressel V, Wigger E, Tobler I.** 2006. Influence of estrus cycle and ageing on activity patterns in two inbred mouse strains. Behav Brain Res **167:**165–174. https://doi.org/10.1016/j.bbr.2005.09.001.

15. **Luong TN, Carlisle HJ, Southwell A, Patterson PH.** 2011. Assessment of motor balance and coordination in mice using the balance beam. J Vis Exp 1–3. https://doi.org/10.3791/2376.

16. **Lewis-Beck MS, Bryman A, Liao TF.** 2004. The SAGE encyclopedia of social science research methods. Thousand Oaks (CA): SAGE Publications.

17. **Mitchell SJ, Scheibye-Knudsen M, Longo DL, de Cabo R.** 2015. Animal models of aging research: implications for human aging and age-related diseases. Annu Rev Anim Biosci **3:**283–303. https://doi.org/10.1146/annurev-animal-022114-110829.

18. **Mullen SP, McAuley E, Satariano WA, Kealey M, Prohaska TR.** 2012. Physical activity and functional limitations in older adults: the influence of self-efficacy and functional performance. J Gerontol B Psychol Sci Soc Sci **67B:**354–361. https://doi.org/10.1093/geronb/gbs036.

19. **National Research Council, Division on Engineering and Physical Sciences; Division of Behavioral and Social Sciences and Education; Board on Mathematical Sciences and Their Applications; Committee on Population; Committee on the Long-Run Macroeconomic Effects of the Aging US Population.** 2012. Aging and the Macroeconomy: Long-Term Implications of an Older Population. Washington (DC): National Academies Press.

20. **Nunnally JC, Bernstein IH.** 1994. Psychometric theory, 3rd ed. (McGraw-Hill Series in Psychology). New York (NY): McGraw-Hill Humanities/Social Sciences/Langua.

21. **Shrout PE.** 1998. Measurement reliability and agreement in psychiatry. Stat Methods Med Res **7:**301–317. https://doi.org/10.1177/096228029800700306.

22. **Shrout PE, Fleiss JL.** 1979. Intraclass correlations: uses in assessing rater reliability. Psychol Bull **86:**420–428. https://doi.org/10.1037/0033-2909.86.2.420.

23. **Tieland M, Trouwborst I, Clark BC.** 2017. Skeletal muscle performance and ageing. J Cachexia Sarcopenia Muscle **9:**3–19. https://doi.org/10.1002/jcsm.12238.

24. **Tung VW, Burton TJ, Dababneh E, Quail SL, Camp AJ.** 2014. Behavioral assessment of the aging mouse vestibular system. J Vis Exp 1–11. https://doi.org/10.3791/51605.

25. **Tung VW, Burton TJ, Quail SL, Mathews MA, Camp AJ.** 2016. Motor Performance is Impaired Following Vestibular Stimulation in Ageing Mice. Front Aging Neurosci **8:**1–10. https://doi.org/10.3389/fnagi.2016.00012.

26. **United Nations Department of Economic and Social Affairs Population Division.** 2019. World Population Ageing 2019: Highlights. New York (NY): United Nations.

27. **Volpi E, Nazemi R, Fujita S.** 2004. Muscle tissue changes with aging. Curr Opin Clin Nutr Metab Care **7:**405–410. https://doi.org/10.1097/01.mco.0000134362.76653.b2.

28. **Wallace JE, Krauter EE, Campbell BA.** 1980. Motor and reflexive behavior in the aging rat. J Gerontol **35:**364–370. https://doi.org/10.1093/geronj/35.3.364.

29. **Walston JD.** 2012. Sarcopenia in older adults. Curr Opin Rheumatol **24:**623–627. https://doi.org/10.1097/BOR.0b013e328358d59b.

30. **Wilkinson DJ, Piasecki M, Atherton PJ.** 2018. The age-related loss of skeletal muscle mass and function: Measurement and physiology of muscle fibre atrophy and muscle fibre loss in humans. Ageing Res Rev **47:**123–132. https://doi.org/10.1016/j.arr.2018.07.005.