C57BL/6 Neuromuscular Healthspan Scoring System

Ted G. Graber, Lisa Ferguson-Stegall, Jong-Hee Kim, and LaDora V. Thompson

University of Minnesota Medical School, Department of Physical Medicine and Rehabilitation, Minneapolis, Minnesota.

Address correspondence to LaDora V. Thompson, PhD, University of Minnesota Medical School, 420 Delaware St. SE, Minneapolis, MN 55455. Email: thomp067@umn.edu

Developing a scoring system based on physiological and functional measurements is critical to test the efficacy of potential interventions for sarcopenia and frailty in aging animal models; therefore, the aim of this study was to develop a neuromuscular healthspan scoring system (NMHSS). We examined three ages of male C57BL/6 mice: adults (6–7 months old, 100% survival), old (24–26 months old, 75% survival), and elderly group (>28 months old, ≤50% survival)—as well as mice along this age continuum. Functional performance (as determined by the rotarod and inverted-cling grip test) and in vitro muscle contractility were the determinants. A raw score was derived for each determinant, and the NMHSS was then derived as the sum of the individual determinant scores. In comparison with individual determinants, the NMHSS reduced the effect of individual variability within age groups, thus potentially providing an enhanced ability to detect treatment effects in future studies.

THE rapidly aging population has caused a worldwide demographic shift that presents multiple challenges for the socioeconomic sustainability of the medical establishment (1,2). Sarcopenia, the age-related loss of skeletal muscle mass and strength, has become a more prevalent condition as more people live longer. Furthermore, quality of life during aging, or healthspan, becomes an important concern. Sarcopenia can ultimately cause the elderly to lose their independence and contributes to the onset of frailty—costing billions of dollars per year (direct cost of sarcopenia was estimated at $18.5 billion in 2002) (3,4). It is in the best interest of society, both from a financial standpoint and as a quality of life issue, to find successful interventions for this age-associated muscle-wasting condition.

Biogerontology research has traditionally focused on mechanisms to extend life span (5). Recently, however, there has been a paradigm shift toward improving healthspan being as or even more important than increasing life span (5–7). In order to address the translation of therapies from the bench to bedside, critical elements must be in place such as animal models, experimental protocols, and assessment tools, which evaluate interventions designed to improve healthspan (8). Moreover, to date, there is no assessment tool to evaluate neuromuscular function in the mouse (healthspan), which would address the component of sarcopenia.

A common challenge in interpreting aging research is the wide statistical variability within outcome measurement data. The standard deviation (SD), commonly used to measure data variability, is a component of the standard error (SE), which plays a major role in determining statistically significant differences between experimental groups. The coefficient of variation is defined as the SD divided by the mean of the measurement. The coefficient of variation thus describes the spread of the data in relation to the mean. However, mathematical constructs, such as multiple regressions, are used to reduce the effect of data spread by limiting the impact of covariates. Therefore, creating mathematical constructs to ameliorate statistical variation is one way to address wide data spread within groups in aging studies.

The primary purpose of this study was to develop a *NeuroMuscular Healthspan Scoring System* (NMHSS) that described healthspan in the C57BL/6 mice and reduced the effect of variability within the selected outcome measures. Specifically, we hypothesized that the large individual variability of an outcome measure within groups that is often a bane to detecting treatment effects in aging studies would be attenuated by this scoring system. Therefore, the first step was to evaluate the performance and muscle contractility changes that occur with age in the mouse model to select the determinants of the scoring system. Then, we constructed the scoring system based on the age-group means and scores predicted by equations developed from multiple linear regression analysis of each determinant. Using these multiple regression equations as a component within the scoring system allowed us to lower the coefficient of variation of the NMHSS in comparison with the individual determinants.

Methods

Animals

Male C57BL/6 mice, from the National Institute on Aging colony, of three specific age groups (adults, $n = 20$, 100% survival, 6–7 months old; old, $n = 12$, 75% survival,

24–26 months; elderly, $n = 23$, <50% survival, >28 months) were selected for initial analysis (9,10). These age groups are translatable to human ages of young adult, middle age, and elderly adults (9,11). Additional ages (from 2 to 32 months) were used to establish the multiple linear regression models for the NMHSS. In total, 70 animals were tested for contractile physiology and 99 for each of the performance tests (rotarod and grip tests). Animals were housed in a central specific pathogen-free facility, and experiments were conducted under Institutional Animal Care and Use Committee approved protocols ensuring optimal ethical and humane treatment. Because of animals lost to natural death and other variables, not all animals were tested in all parameters. Body and heart mass were determined on day of sacrifice.

Functional Measurements

Rotarod.—Rotarod testing is a well-established measure of overall motor function (12,13). The mice are placed on a rotating cylinder (Pan Lab Lsi Rota-Rod /RS Model 8200), and the time spent or duration before falling is recorded. The device is set to either a run (static revolutions per minute [rpm]) or an accelerating (rpm consistently increases over time—accelerates from 4 to 40 rpm in 30 seconds or 2, 5, or 10 minutes) protocol. The mice were acclimated to the rotarod before testing to familiarize the mice with the device and test protocols. Briefly, each mouse was acclimated on the rotarod device for three trials per day on three consecutive days. On Day 4, the performance of the mouse was tested using an accelerating protocol (5-minute slope to a maximum of 40 rpm), and the time the mouse remained on the rotarod was recorded. This accelerating protocol was performed three times and the average of the three trials was calculated.

Inverted-cling grip test.—The inverted-cling grip test is a measure of overall strength and muscular endurance of the mouse (14). This test consisted of placing the mouse on a cage-like wire grid and then inverting the grid (over a padded surface).

A custom-designed device was constructed for the grip test to ensure consistency between trials. The mice were placed on the grid that was located on the hinged lid of the device. The grid was held perpendicular to the device for 3 seconds, and then the lid was closed so that the mouse was inverted. The time the mouse held on before falling to the padded surface at the bottom of the device was recorded to a maximum of 180 seconds. The results were averaged between the two trials.

Whole muscle physiology.—An in vitro isolated muscle preparation was used to measure the contractile properties of the extensor digitorum longus (EDL) muscles. Immediately

after being weighed (massed), the EDL muscle was placed in a tissue bath filled with Krebs–Ringer buffer (115 mM NaCl, 5.9 mM KCl, 1.2 mM $MgCl_2$, 1.2 mM NaH_2PO_4, 1.2 mM Na_2SO_4, 2.5 mM $CaCl_2$, 25 mM $NaHCO_3$, and 10 mM glucose, pH 7.3), oxygenated (5% carbon dioxide and 95% oxygen), and maintained at 25°C via a circulating water system. The muscle, viewed under a 1.75× swing arm magnifying glass, was tied with number 4 suture line at both myotendinous junctions and then suspended from a force transducer and static clamp between two platinum electrodes in an oxygenated bath filled with Krebs-Ringer buffer. The muscle was then tested for contractile function.

Equipment and settings for measuring muscle contractile function.—A dual bath physiology system (Aurora Scientific, Aurora, Ontario, Canada) was used to measure isolated whole muscle contractility. This system consisted of two force transducers (0.5 N, Model 300B), two stimulators (Model 701B), one Dual Lever A/D Interface (Model 604B), one Dual System Signal Interface, customized Aurora software Dynamic Muscle Control (version 4.1.4.6), Dynamic Muscle Analysis (version 3.2), and a temperature control with water bath unit (Model 912 Polyscience, Niles, IL). The physiological system was controlled from a dedicated Windows compatible computer. The stimulators were set to biphase modality and the electrical output to 1,000 milliAmps at 30 V.

Determination of muscle optimal length.—Optimal muscle length (L_0) was determined using the peak twitch force–length curve. Briefly, the muscle length was increased until the peak twitch force (P_t) was achieved. Subsequently, the muscle length was measured between myotendinous junctions using calipers. This measurement is L_0 (optimal length). L_0 is the muscle length where the sarcomeres are at optimal length to produce maximal force (15).

Determination of maximum isometric force.—The force–frequency curve was used to obtain the maximum or peak isometric force. Briefly, isometric force, with the muscle set at L_0, was determined at various frequencies (hertz, Hz): a single pulse (twitch), 10, 40, 80, 120, 150, and 180 Hz and then again for a single twitch. The 300 ms pulse length was software controlled by Dynamic Muscle Control. Each contraction was preceded by a 1-minute rest period. The maximum isometric force (P_0) was defined as the peak force generated and was generally achieved at 150 Hz for the EDL.

Neuromuscular healthspan scoring system.—The NMHSS is an index composed of a composite of three scores from the individual outcome determinants (rotarod, grip test, and EDL maximum isometric force). Before determining the composite of scores for the NMHSS, it was necessary to determine two

important mathematical terms: (i) the statistical mean of each outcome determinant by age group, defined as the mean, and (ii) the predicted outcome measurement for each determinant, which was calculated using a multiple linear regression equation. Next, the score for each outcome determinant was calculated from two components. The first component was the ratio of the actual measurement/mean for the age group, and the second component was the ratio of the actual measurement/predicted measurement for the age group. Subsequently, the score for each determinant was defined as the mean of the two ratios. Finally, the overall healthspan score (NMHSS) was the sum of the three individual outcome scores (rotarod, grip test, and EDL maximum isometric force). Figure 1 shows an example of how to create a rotarod score using a representative mouse from the elderly age group.

The NMHSS identifies how different the animal's actual performance was from the group mean, identifies how far the animal's actual performance is from the predicted, and accounts for variability within each outcome measure. An animal whose performance score is at the mean for all three determinants (actual measurement = mean, and actual measurement = predicted measurement) will have a NMHSS of "3." NMHSS more than 3 would suggest a healthier animal compared with an animal with a NMHSS less than 3. Table 1 illustrates the use of the Healthspan Scoring System with two examples of "old" mice (one with a low grip test and one with a high grip test) from our actual population and a fictitious average old mouse. The data for the fictitious average old mouse were derived from our reference group for each parameter.

Statistical Analysis

Statistical differences between means were determined by one-way analysis of variance (ANOVA) and analysis of covariance. Data are presented as means ± standard error (*SEM*) when appropriate. Significance was set at α = 0.05 for ANOVA, analysis of covariance, logistic regressions, and simple/multiple linear regressions. The post hoc test for ANOVA was Tukey–Kramer Honestly Significant Difference, and a Bonferroni correction was used for analysis of covariance. Factor analysis, using a promax rotation to determine principal components contributing to the variability of each set of outcome measures, was performed. SPSS (IBM Corporation, Armonk, New York) was used for statistical analysis aside from the power analyses (for a 3 × 2 ANOVA to detect a 15% treatment effect at 80% power), which were conducted with the PASS software package (from NCSS LLC, Kaysville, Utah).

Results

Functional Measurements

Rotarod.—The effect of age on overall motor function is summarized in Figures 2–4. The mean rotarod performance, the time spent on the rod during an accelerating protocol, declined 40% across the age groups (adult versus elderly group, $p = .002$; old versus elderly group, $p = .037$; Figure 2). Moreover, there is a negative correlation between age and rotarod performance ($R = −0.373$, $p = .001$; Figure 3) explaining 14% of the variability ($r^2 = 0.139$). Figure 3 shows poor rotarod performance,

NMHSS = Sum of Raw Scores (Rota-Rod + Grip Test + EDL Force)

$$\text{Rota-Rod Raw Score} = \left(\frac{Actual}{Mean} + \frac{Actual}{Predicted}\right) \div 2 = \left(\frac{115}{60} + \frac{115}{127.9}\right) \div 2 = 1.41$$
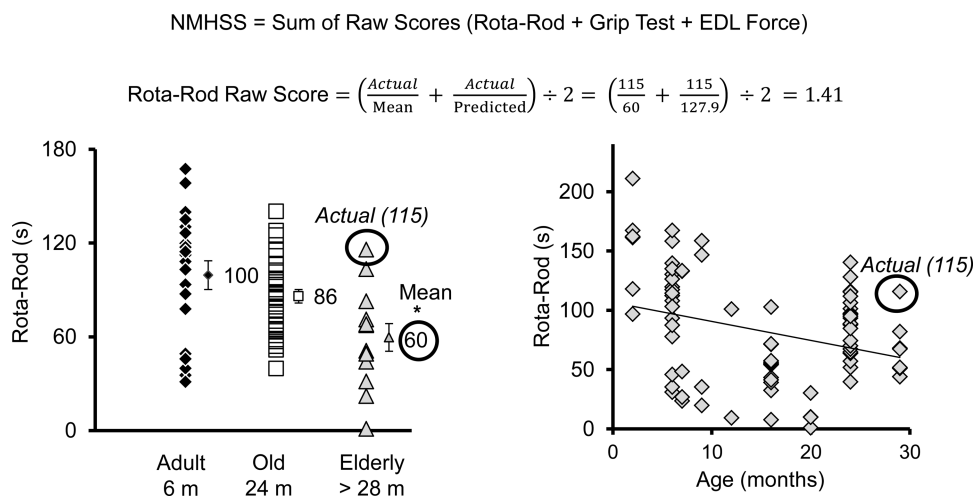

Figure 1. Production of rotarod raw score. The figure shows an example of how to create a rotarod raw score using a representative animal from the elderly age group (>28 months of age). Numbers in italics are the unique values associated with this animal. The first term of the equation is produced by taking the actual score of the animal in the test (*115.0* s) and then dividing by the mean score for the age group (60.0 s), which in this case equals 1.92. The second term of the equation consists of the actual score on the test (*115.0* s) divided by the predicted score (*127.9* s), which is produced by the model described by the multiple linear regression [$n = 35$, $R = 0.720$, $r^2 = 0.518$, $p < .001$; model: Rotarod (s) = 237.951 − 1.798 * body mass (*31.3* g) − 291.995 * heart mass (*0.184* mg) = *127.86*], which equals *0.899*. Both terms are then added (*1.92 + 0.899 = 2.819*) and divided by two to get the mean, *1.410*. This raw score for rotarod, 1.410, is then added to the raw score for grip test and EDL force to obtain the overall NMHSS score. The same process is performed for each outcome measure. In the case of this particular mouse, the grip test raw score was 0.709 and the EDL force raw score was 0.783. The total NMHSS score of this particular mouse was 2.902.

Table 1. Comparison of NMMHS Score Construction

A.	Average Fictional Old Mouse (AF)*		Low Grip Old Mouse (LG)		High Grip Old Mouse (HG)	
	Mean	Score	Measure	Score	Measure	Score
Grip mean	68.0	1	34.5	0.51	116.5	1.71
Grip predicted	72.97	0.93	69.98	0.51	76.34	1.53
Grip raw score		0.97		0.5		1.62
P_0 mean	319.00	1	349.31	1.1	338.23	1.06
P_0 predicted	279.40	1.14	276.34	1.26	284.18	1.19
P_0 raw score		1.07		1.18		1.13
Rotarod (R) mean	86.1	1	64.7	0.7	57.0	0.61
R. predicted	122.20	0.70	117.78	0.55	132.42	0.43
R. raw score		0.85		0.62		0.52
NMHSS score		2.89		2.30		3.27
B.	Covariates	Mean	LG	HG		
	L_0 (mm)	12.12	12.00	12.30		
	Body mass (g)	33.10	34.00	32.09		
	Heart mass (g)	0.19	0.20	0.16		

Notes. (**A**) The table shows a comparison of the scores that comprise the NMHSS in two actual examples of mice from the old cohort (both 24-mo old mice), one with a low performance in grip test (LG) and the other with a high performance (HG). The first two columns describe the values that would be obtained by putting the mean values for the age group into the NMHSS equation. The derivation for each component of the NMHSS index follows the formula given in Figure 1. (**B**) The values for the covariates used in multiple linear regression equations to produce the predicted terms are shown in this table. mm, millimeters; g, grams.

*It should be noted that the fictional old mouse has values different from the mean of the old mouse cohort in Figure 12 because the values given are from the means for the entire sample of the population, not just the cohort described in Figure 12.
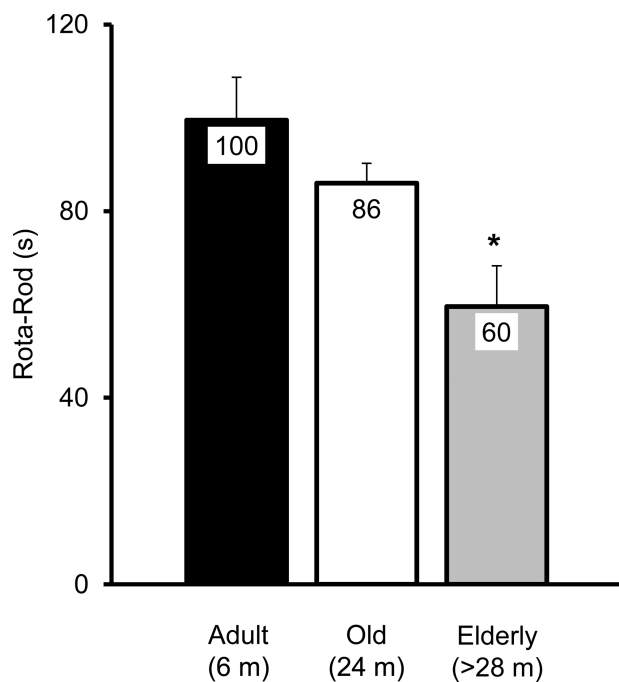

Figure 2. Rotarod performance declines with age. Rotarod performance, the time before falling, is defined as the average time in seconds (s) of three trials. The number of mice per group were 20, 32, and 13 for the adult (6 months, m), old (24 months, m), and elderly mice (>28 months, m), respectively. Values are means (100, 86, and 60) for adult, old, and elderly mice) ± *SEM*. Results are from a one-way ANOVA ($F = 6.3$, $p = .003$) with a Tukey–Kramer Honestly Significant Difference post hoc analysis. *elderly mice significantly different from adult ($p = .002$) and old mice ($p = .037$).
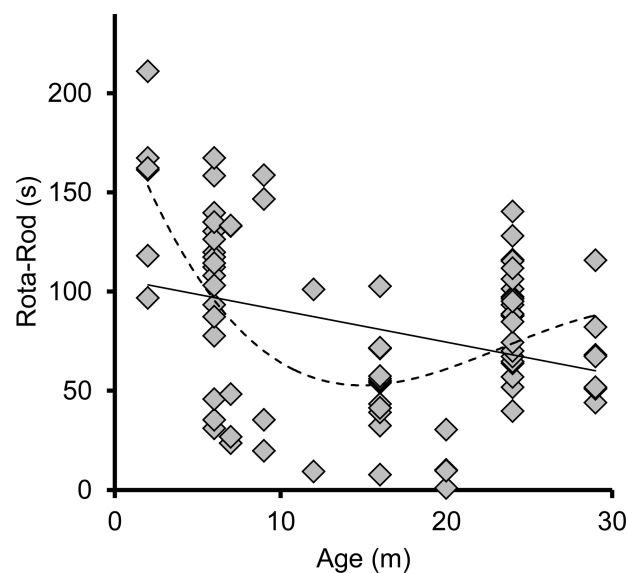

Figure 3. Rotarod performance declines with age. This graph shows both a simple linear regression ($r = -0.373$, $R^2 = 0.139$, $p = .001$, $n = 99$; equation: $y = -1.876 * age + 99.699$), which explains 14% of the variability, and a third-degree polynomial regression ($R = 0.588$, $r^2 = 0.3453$, $p < .001$, $n = 99$; equation: $y = -0.0157 * age^3 + 1.1005 * age^2 - 22.375 * age + 193.66$), in which 35% of the variation is explained. The third-degree polynomial regression demonstrates the survivor effect (animals at the oldest ages are healthier than some animals in middle age—ie, 16–20 months). Note the dip in performance of the 16–20-month animals.

represented as a "dip," in mice between the ages of 12 and 20 months. Although the simple linear regression showed a negative correlation between age and rotarod performance, the third-degree polynomial regression better explained the reduced performance at the middle-age groups and the increased performance at the youngest and oldest ages.

In order to determine if rotarod performance predicts the age of the mouse, a simple logistic regression ($r^2 = 0.19$,
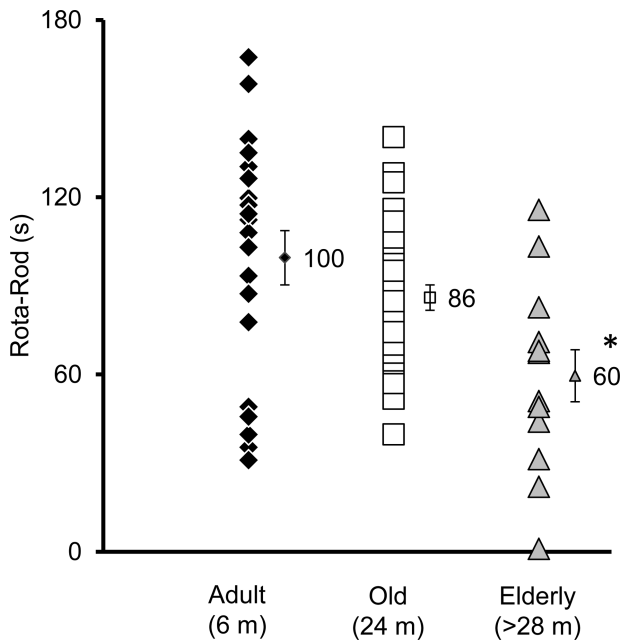
Figure 4. Large variability in rotarod performance within each age group. Rotarod performance ranges from 31 to 167 s in adult, from 40 to 140 s in old mice, and from 1 to 116 s in elderly mice. s: the average time in seconds (of three trials) spent on rotarod before falling. Values 100, 86, and 60 represent the mean and the bar represents ±*SEM*. m: months old. *elderly mice significantly different from adult (*p* = .002) and old mice (*p* = .037).
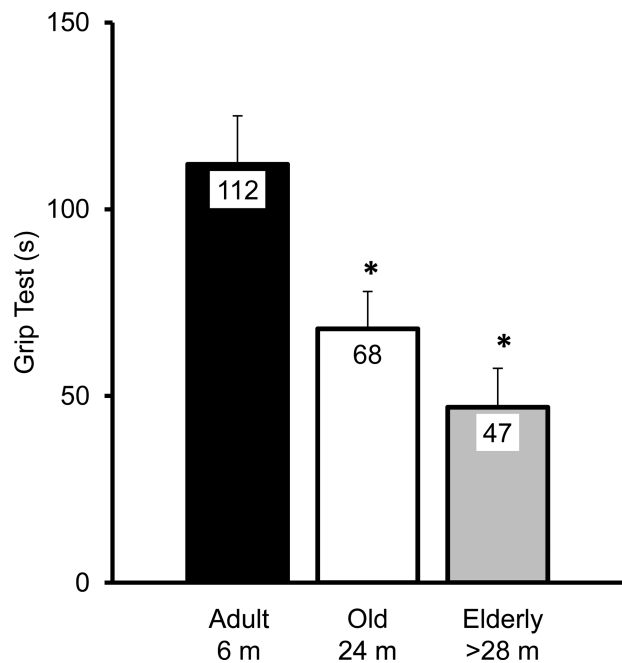

Figure 5. Performance on grip test declines with age. Grip test is defined as the average time (duration in seconds, s) before falling from the grid. The numbers of animals per group were as follows: 18, 20, and 8 for the adult (6 months old, m), old (24 months old, m), and elderly animals (>28 months old, m), respectively. Values are means (112, 68, 47 for adult, old, and elderly animals) ± *SEM*. Results from a one-way ANOVA (*F* = 6.8, *p* = .003) with a Tukey–Kramer Honestly Significant Difference post hoc analysis show that adults are different from old (*p* = .002) and from elderly animals (*p* = .006).

X^2 = 12.1, *p* = .002) was calculated. The regression model used was: Adult (1) = reference group, Old (2) = 1.734 − 0.014 * rotarod time, Elderly (3) = 2.903 − 0.042 * rotarod time. The logistic regression correctly classified the mice into their proper age groups 57% of the time.

Figure 4 highlights the variability between individual mice and shows the broad range of ability to stay on the rotarod within each of the age groups. Adult mice were able to remain on the accelerating rotarod from 31 to 167 seconds, whereas the elderly mice remain from 1 to 116 seconds. To determine the major contributors to the variability, factor analysis and hierarchical multiple regression analysis were performed among all the experimental parameters. The multiple linear regression analysis showed that 52% of the rotarod variability could be explained by body mass and heart mass ([*n* = 35, *R* = 0.720, r^2 = 0.518, *p* < .001; model equation: rotarod (s) = 237.951 − 1.798 * body mass (g) − 291.995 * heart mass (mg)]). There was a positive correlation between age and body mass (*R* = 0.429, *p* = .010) in the animals that were rotarod tested.

Using body mass as a covariant (analysis of covariance, *F* = 5.94 and *p* = .006), rotarod performance was found to be statistically different between adult and old mice (Bonferroni-adjusted post hoc test, *p* = .0332) and between adult and elderly mice (*p* = .011). There was no statistical difference between the old and elderly groups.

Grip test.—The effect of age on overall muscle strength, as measured by the inverted-cling grip test, is summarized in Figures 5–7. Using a one-way ANOVA to test the difference in mean performances, a significant age-related decline in grip strength was found between the adult and the elderly mice (61% reduction, *p* = .006) and between the adult and the old mice (39% reduction, *p* = .002); however, there was no significant difference in performance between the old and elderly groups (Figure 5).

Using simple linear regression, grip test performance was found to be negatively correlated with age in months (*R* = −0.419) and 18% of the variability could be accounted for (r^2 = 0.176, *p* = .001). The linear regression equation used was grip test (s) = −1.876 * age (m) + 99.699 (Figure 6).

Figure 7 highlights the wide individual variability within each age group. In the adult group, the lowest functioning mouse held onto the grid for 41.5 seconds, whereas the highest functioning mice held on for the maximum (180 seconds). In contrast to the adult group, the best performing mouse in the elderly group held on for only 94 seconds and the weakest mouse for 8.5 seconds. To determine whether the grip test performance can be used to classify the mice into their respective age groups even with this wide variation, without using any adjuster, a simple logistic regression was performed. The regression model used was: Adult (1) = reference group, Old (2) = 1.686 − 0.018 * age at
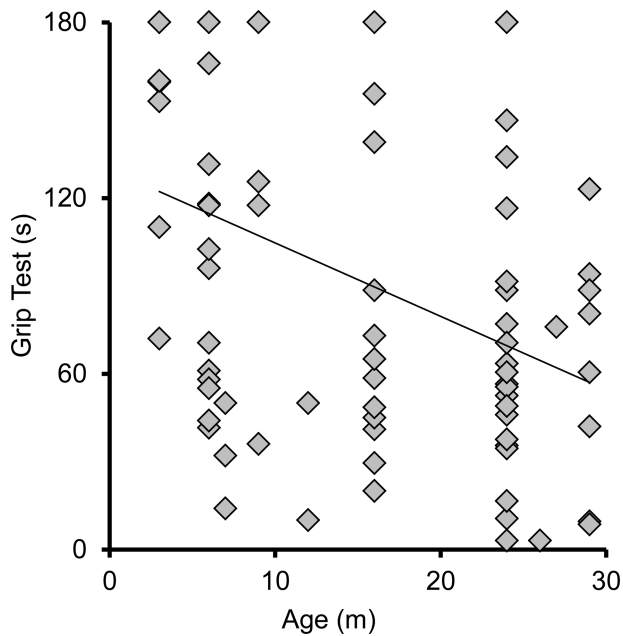
Figure 6. Grip test declines with age. This graph shows a simple linear regression ($R = -0.419$, $r^2 = 0.176$, $p = .001$, $n = 99$; equation: Grip (s) = -1.876 age (m) + 99.699), in which age explains 18% of the variation. s, seconds; m, months.
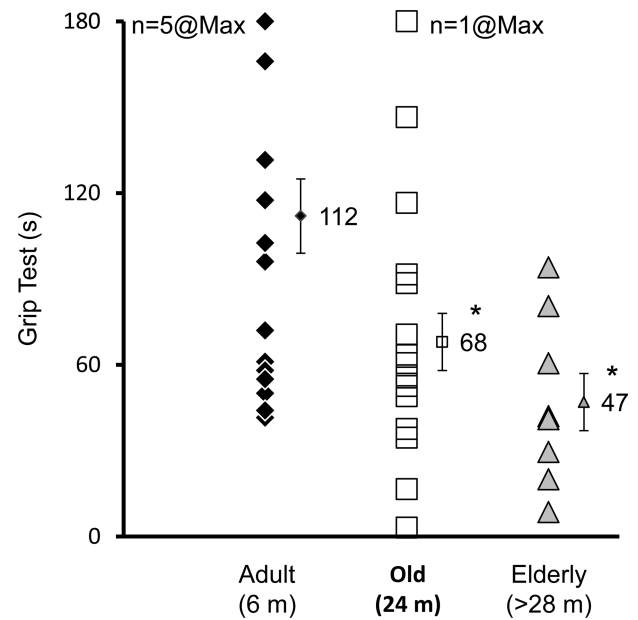

Figure 7. Grip test performance shows wide individual variation within each age group. Grip test performance ranges from 41.5 seconds (s) to the maximum 180 s (adult), from 3 s to the maximum 180 s (old group), and from 8.5 to 94 s in the elderly group. Values 112, 68, and 47 represent the mean and the bar represents ± *SEM*. m, months old. *significantly different (from adult group) mean ($p < .05$).

grip test, Elderly (3) = 1.760 – 0.036 * age at grip test. The results indicate that the age group of the mouse is predicted correctly 56.6% of the time ($r^2 = 0.24$, $X^2 = 12.5$, $p = .002$).

To determine the major contributors to the variability in grip test, factor analysis and hierarchical multiple regression analysis were performed among all the experimental parameters. Age and body mass were the major sources of the variability. When corrected for age and body mass, 22% of the variability is accounted for [multiple linear regression, $n = 29$, $R = 0.471$, $r^2 = 0.222$, $p = .038$; model equation: grip test (s) = 213.024 – 1.240 * age at grip test (m) −3.332 * body mass (g)].

Muscle contractile physiology.—Using in vitro methodology, we documented contractile properties for the EDL muscles, in the three age groups and along our healthspan continuum (2–32 months). The main contractile property investigated was peak tetanic force (P_0) of the EDL.

EDL peak tetanic force (P_0).—EDL P_0 declined significantly with age (28%, $p < .001$), from a mean of 388 mN in adult to a mean of 281 mN in the elderly group (Figure 8). The decline from adult to old (mean 319 mN) trended toward significance (−18%, $p = .09$), with no significant difference between the old and elderly groups (−12%, $p = .43$). There was a correlation between age and EDL P_0 ($R = -0.569$). A simple linear regression of peak tetanic force of the EDL with the age in months ($n = 53$, $R = -0.569$, $r^2 = 0.324$, $p < .001$; with the equation: $P_0 = -4.452$ * age (m) + 420.128) showed that age accounted for 32% of the variability (Figure 9).

Figure 10 highlights the wide variability of EDL P_0 within each specific age group. For instance, the P_0 of the EDL from the adult group ranged from a high of 482 mN to a low of 286 mN, whereas the P_0 of the EDL muscles from the elderly group ranged from a high of 427 mN to a low of 182 mN. Notably, 35% of the mice from the elderly group performed better than the average mouse in the old group. Furthermore, 27% of the mice from the adult group performed worse than the average mouse in the old group.

To test whether the EDL P_0 was predictive of the mouse age group, a simple logistic regression was performed ($r^2 = 0.39$, $X^2 = 17.25$, $p < .001$). P_0 classified the animal into the elderly age group correctly 90% of the time using the following model equation: Adult (1) = reference group, Old (2) = not significant, Elderly group (3) = 8.13 – 0.023 * P_0. In contrast, this regression model was not able to significantly classify the animal into the old age group. When using the old group as the reference, however, the regression model classifies the adult group 66% of the time.

In order to explain the individual variability within each age group, a factor analysis was performed followed by a multiple linear regression using muscle length and age as covariates. Forty-six percent of the variance (adjusted $r^2 = 0.46$) was explained by the P_0 multiple regression equation (P_0 [mN]= 117.484 – 4.976 age (m) + 26.094 L_0 [mm]).

Relationship between the outcome measurements.— A simple linear regression of the determinants is shown in Figure 11, demonstrating that there was no significant correlation between rotarod and EDL P_0 (or between grip
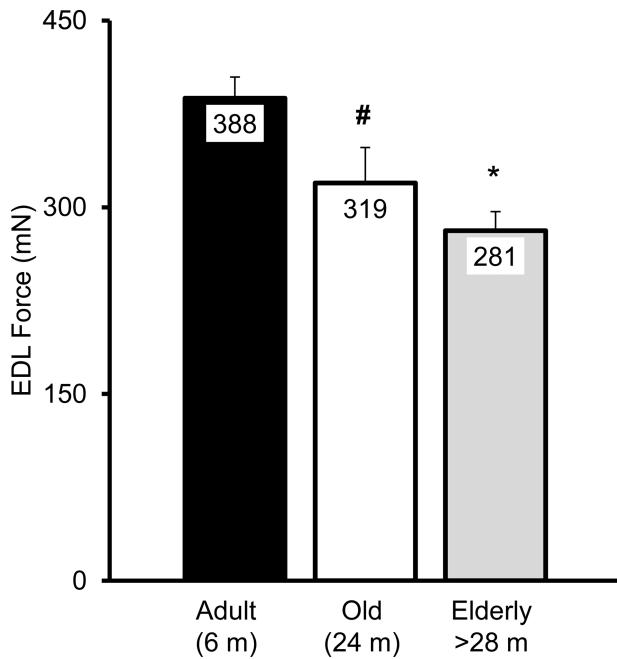
Figure 8. The peak force produced by the EDL muscle declines with age. The number of animals per group were 15, 7, and 20 for the adult, old, and elderly groups, respectively. Means (in milliNewtons, mN) were 388, 319, and 281 for the adult, old, and elderly groups ± *SEM*. Results are from a one-way ANOVA ($F = 10.16$, $p < .001$) with a Tukey–Kramer Honestly Significant Difference post hoc analysis: adult different from old (#$p = .09$) and from elderly (*$p < .001$). m, months.
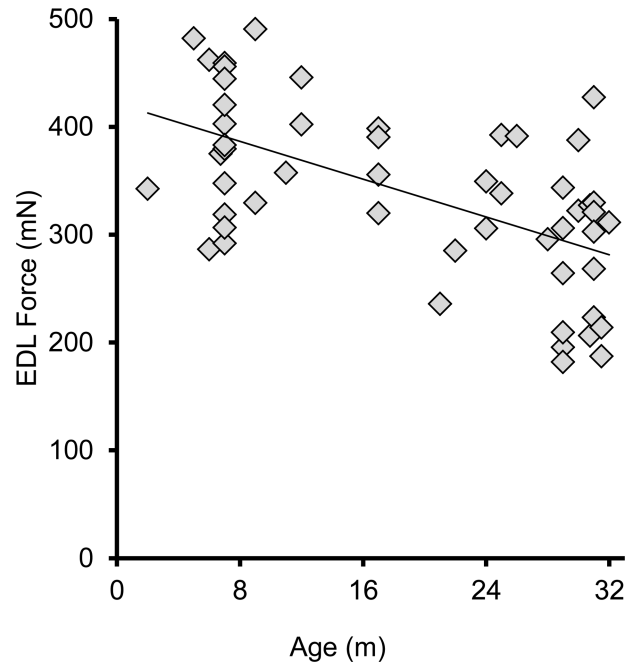

Figure 9. Force generation by the EDL muscle declines with age. This graph shows a simple linear regression ($n = 53$, $R = -0.569$, $r^2 = 0.324$, $p < .001$; equation: $P_0 = -4.452 * \text{age (m)} + 420.128$), in which age explains 32% of the variability of P_0. mN, milliNewtons; m, months.

test and EDL force). There is a weak correlation, however, between grip test and rotarod ($r^2 = 0.11$). The three outcome measures are relatively independent measurements of neuromuscular health, each representing different or unique aspects of the mouse performance ability.

Neuromuscular healthspan scoring system.—The NMHSS for a cohort of mice is demonstrated in Figure 12 ($n = 15$ adults, $n = 5$ old animals, and $n = 19$ elderly animals). The mean NMHSS score for the adult animals was 3.01 and ranged from 1.68 to 4.70 ($SEM = 0.196$). The mean NMHSS score for the elderly animals was 3.05 and ranged from 1.71 to 4.44 ($SEM = 0.154$). In contrast, however, the mean NMHSS score for the old animals was 2.67 and ranged from 2.14 to 3.24 ($SEM = 0.211$). Figure 1 shows how the NMHSS is determined, using a single rotarod raw score as an example.

In order to quantify and compare the amount of individual variability inherent within each age group, coefficients of variations (standard deviation/mean) were evaluated. When comparing the coefficient of variation of the NMHSS with the coefficient of variation of our outcome measures, there was a reduction in variation of 2-fold for rotarod and 3.7-fold for grip test; although there was a slight increase (0.23) from P_0.

To demonstrate the utility of the NMHSS ability to reduce variability, we compared power analyses (80% power to

detect a 15% difference in a 3 × 2 ANOVA—designed as three age groups each with two treatment groups) using the mean of the elderly age group for each outcome measure (rotarod, grip test, and P_0). Table 2 summarizes the results of the power analyses and the CVs. Notably, the number of animals needed for detection of the 15% difference is reduced by 77%, 87%, and 21% for rotarod, grip test, and P_0, respectively.

Discussion

The purpose of this study was to develop a neuromuscular healthspan scoring system that will be used to evaluate treatments for sarcopenia. We hypothesized the wide variability within groups, which is often a bane to detecting treatment effects in aging studies, will be attenuated by this scoring system. The scoring system consisted of a mathematical construct designed as an index of determinants, which both compared the means of relevant outcome measurements and used multiple regressions to alleviate the effect of covariates. The main findings included a significant reduction in coefficient of variation with the NMHSS compared with the coefficients of variation of two of the outcome measures (rotarod and grip test). This resulted in an increased ability to detect differences between groups, reflected by a reduction in the number of animals needed to detect a difference of 15% at 80% power in a 3 × 2 ANOVA in the NMHSS as compared with all three determinants alone. As expected, the main outcome measures (determinants: rotarod, grip
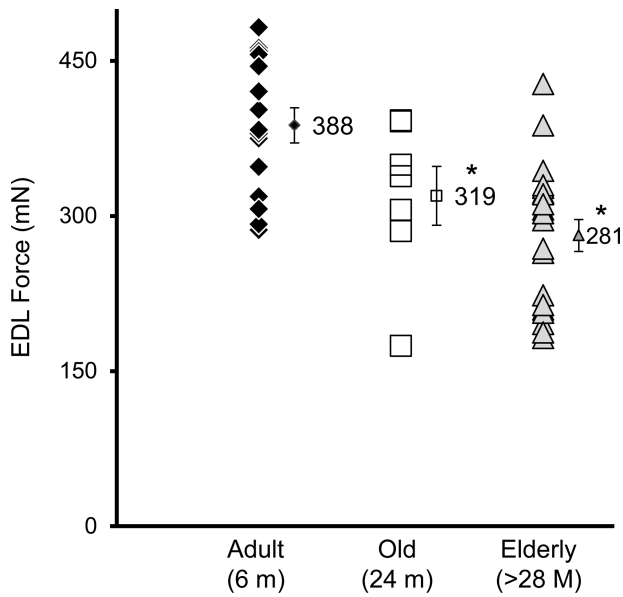
Figure 10. Force generation of the EDL shows wide variability within age groups. The P_0 ranges from 286 to 482, from 174 to 392, and from 182 to 427 in the adult, old and elderly groups, respectively. Numbers 388, 319, and 281 represent the mean, and the bar represents ± *SEM* for adult, old and elderly groups, respectively. m, months; mN, milliNewtons. Significantly different (*p < .05) from adult group.

test, and EDL P_0) declined with age (from adult to elderly animals—40%, −61%, and −28%, respectively).

Scoring indices or testing batteries have been developed to measure frailty and to predict life span in the mouse (12,16). In humans, there are multiple types of testing regimens designed to measure disability, frailty, mental health status, and predict biological age (17–20). However, a neuromuscular healthspan scoring system has not been developed. Neuromuscular health is defined as muscle force production combined with functional performance. This ability decreases, on average, in an age-dependent manner (21–23). In this study, our definition of neuromuscular healthspan is the ability to maintain an optimal level of performance (eg, running and jumping) and strength/power output over the life span adequate to perform activities of daily living. Hence, the scoring system we developed quantifies neuromuscular healthspan within age groups.

The components of the NMHSS combine to present an overall picture of the neuromuscular health of the animal—both in comparison with the peers within its age group and with respect to what level of ability would be predicted by the multiple linear regression equation at the animals' given age. Our mathematical construct (the mean of the ratios of the actual/mean and actual/predicted values) was a successful way to reduce variability of the individual determinants (rotarod, grip test, and P_0) by removing the effect of covariates using multiple linear regression (r^2 = 0.52, 0.22, and 0.46 for rotarod, grip test, and P_0, respectively).

One significant outcome of reducing variability is lowering the effective *SD* by reducing the spread of data, which
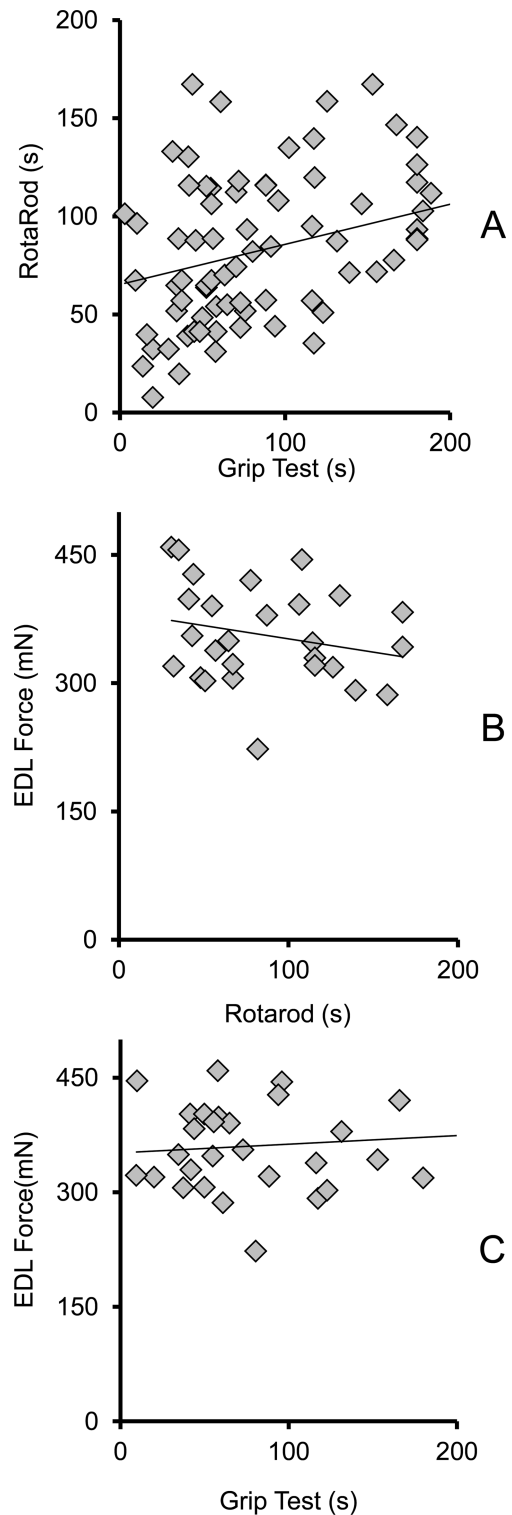

Figure 11. Little correlation between outcome measures of rotarod, grip test, and EDL force. mN, milliNewtons and s, seconds. **Panel A**: Grip test and rotarod are weakly correlated. Rotarod/grip test regression (n = 66): Rotarod (s) = 0.281 * grip (s) + 55.078, R = 0.334, r^2 = 0.111, p = .007. **Panel B**: Rotarod is not significantly correlated with EDL force. Rotarod/EDL force regression (n = 27): P_0(mN) = −0.3107 * rotarod (s) + 382.84, R = 0.232, r^2 = 0.054, p = .244. **Panel C**: Grip test is not significantly correlated with EDL force. Grip test/EDL force regression (n = 30): P_0 (mN) = 0.1124 * grip (s) + 351.64, R = 0.10, r^2 = 0.01, p = .606.
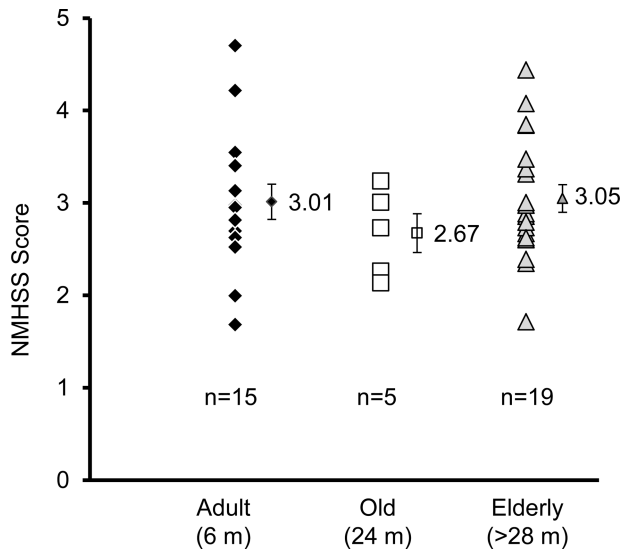
Figure 12. NMHSS scores for adult, old, and elderly mice. NMHSS ranges from 1.68 to 4.70 (adult), from 2.14 to 3.24 (old), and from 1.70 to 4.44 in the elderly group. Values 3.01, 2.67, and 3.05 represent the mean and the bar represents ± *SEM*. m, months old.

Table 2. NMHSS Reduces Effect of Variability

Elderly Mice	*SD*	Mean	CV	*n* (at 80% power)
NMHSS	0.67	3.05	0.22	11
Rotarod	30.8	68.3	0.45	48
Grip	35.3	59.8	0.59	82
P_0	62.3	285.6	0.17	14

Notes: The coefficient of variation, CV, of the NMHSS is lower (0.22) than the CV of either the rotarod (0.45) or the grip test (0.59). The end result is that the number of animals needed to achieve an 80% power (using means and *SD*s from the elderly cohort from Figure 12 in a 3 × 2 ANOVA at α = 0.05 with the desired detectable difference being 15%) is much lower using the NMHSS (11) than the other tests (48, 82, and 14 for the rotarod, grip, and EDL P_0, respectively).

then reduces the *SEM*. This makes it easier to detect differences between group means because the 95% confidence interval around the mean becomes smaller—thus it becomes easier to achieve statistical significance. Instead of reducing variability (or *SD*), often the number of animals used (*n*) is increased to help achieve the same effect. In this study, summarized in Table 2 and with data taken from the actual cohort of elderly animals (Figure 12), the marked reduction in animals needed (sample size) is evident. Specifically, Table 2 documents a power analysis by using NMHSS, for a 3 × 2 ANOVA, to detect a 15% treatment effect at 80% power in the actual elderly group.

As noted in the Methods section, the NMHSS identifies how far the animal's actual performance is from the age-group mean, identifies how far the animal's actual performance is from the predicted, and accounts for variability of each set of outcome measures. Assessing the relative neuromuscular health is another advantage of using the NMHSS. An animal whose performance score is at the mean for all three determinants (actual measurement = mean, and actual measurement = predicted measurement) will have a NMHSS of "3." NMHSS more than 3 would suggest a healthier animal compared with an animal with a NMHSS less than 3.

As highlighted in Figure 12, the mean NMHSS of the adult, old, and elderly animals from the test cohort were 3.01, 2.67, and 3.05, respectively. The adult and elderly means indicated that the groups themselves tended to perform, on average, very close to what would be predicted. The old animals, lower than 3 score can be interpreted that the old group of animals performed, on average, at a lower than expected level.

Collectively, in this study, our young and elderly group performed as expected, whereas the old group was less capable. A unique application of the NMHSS is as assessment tool to describe the collective neuromuscular health of the age group cohorts and to describe the relative neuromuscular health of an individual mouse. Thus, there is a potential to assess frailty within a group, by setting a cutoff value—below which an animal is considered frail or weak. For example, two standard deviations below the mean cohort NMHSS could be used to declare an animal a weak member within the group, whereas, two standard deviations below the average value of an age group (ie, the old average animal NMHSS is 2.89, from Table 1) could be used to declare an animal frail.

Determinant validity: Rotarod, grip test, and EDL peak tetanic force.—One necessary component of scoring systems is the use of outcome measurements that ensure validity (24–26). Validity is defined as accurately measuring the intended measurement (26). The determinants of the NMHSS were carefully selected and vetted to ensure maximum validity.

Rotarod and neuromuscular healthspan.—We chose the rotarod as our first determinant to measure overall motor function because it is one of the most common functional tests traditionally used for neuromuscular evaluation (12–14). The outcome measurement is how long the animal can stay on the device. The mode of the rotarod operation that we chose (acceleration) requires the animals to not only keep their balance and run on a spinning rod but also to continually produce more power to keep up with the acceleration of the device. Because the animals will also tire from the exertion of the effort required to keep up with spinning rod, the measurement also contains a component of endurance. Therefore, this device demonstrates both face and content validity for our desired measurements (overall motor function: balance, coordination, gait speed, and endurance) with only one measurement (time in seconds on device) because the time on the device directly correlates with the linear acceleration of the revolutions per minute. This outcome measure is similar to gait speed or timed-up-and-go testing

used in humans, which are both used to document disability and frailty in humans (17).

We used the two variables, body and heart mass, that contributed most to the variability in the multiple linear regression. These two adjustors are appropriate and valid contributors to variability. Heart mass may very well be an indicator of underlying cardiovascular conditions. For example, an enlarged heart may signify heart valve problems, cardiomyopathy, coronary artery disease, or hypertension—any of which has potential to negatively affect cardiovascular performance (one aspect of the rotarod test component that encompasses endurance) (27,28). Controlling for heart mass has the advantage of reducing the effect of these potential conditions on the outcome measurement. Bodyweight has obvious implications on performance because a large and/or obese mouse may not be as agile or as fatigue resistant in comparison with a mouse of normal weight.

Rotarod and the survivor effect.—There was a clear dip in performance evidenced on Figure 3. The survivor effect may play a role in this dip. The survivor effect theory suggests that the strongest and healthiest animals will live to the oldest ages (less healthy/robust animals will succumb to disease before reaching the oldest ages); therefore, in some cases the performance of the older individuals will exceed that of their younger counterparts (6).

Grip test and neuromuscular healthspan.—Another well-characterized measurement of neuromuscular ability is the grip test (12–14). There are different ways to measure grip strength, including the use of force transducers attached to trapeze arms (the animal grasps the bar and is pulled backwards by the tail, outcome being the force measured when the animal lets go) and suspending the mice from a grid and measuring how long they can hold on before they fall. Although the former measurement has less involvement of muscle stamina and more directly measures strength, the latter gives information about both strength and endurance.

We, therefore, chose the inverted-cling grip test to be our second functional test, and built a custom testing device to ensure that the reliability of the test was maximized by making the conditions of each test identical. The face and content validity of the test are evident in that the outcome measurement (eg, how long can the mouse suspend itself before falling) measures the ability of the mouse to support its body weight (strength) for a given amount of time (endurance). This outcome measure would be similar to a human pull-up test.

EDL contractility and neuromuscular healthspan.—One advantage of using the mouse model is that we can isolate individual muscles for whole muscle in vitro contractile physiology measurement. The EDL is primarily a fast fiber type muscle and thus is more sensitive to age-related muscle

dysfunction (29). P_0 represents raw force production—an absolute measurement of strength. This outcome measure is, therefore, somewhat comparable with a one-repetition maximum measurement of a weight lifting exercise, but may be more reliable and valid because the muscle receives maximum stimulation because there is no voluntary component.

Conclusions

Both functional ability and strength are impaired with age in the C57BL/6 mouse as evidenced by declines in grip test, rotarod, and EDL P_0. This was in agreement with other investigations (12,13,15,30). There is, however, wide variation in the ability of individual animals. NMHSS, as a mathematical construct, is a much more sensitive instrument than the outcome measurements alone—due to lower coefficients of variation. This leads to an increase in power that allows detection of differences in means with ANOVA using far fewer animals than would be needed to detect the same difference in the separate outcome measures. The NMHSS may well become a very valuable tool for researchers to assess interventions in future studies.

In summary, the NMHSS reduces variability, increases power, and serves as an assessment tool for neuromuscular ability. We postulate that in future investigations, the principles of the NMHSS may be adapted to producing other types of scoring systems in addition to providing researchers with a tool to assess sarcopenia interventions. By substituting other outcome measurements and by carefully considering validity, other types of scoring systems (eg, cardiovascular, immune response, and others) could be produced using the principles behind the NMHSS.

References

1. Herrell JH. Health care expenditures: the approaching crisis. *Mayo Clin Proc*. 1980;55:705–710.
2. Peterson PG. Gray dawn: the global aging crisis. *Foreign Affairs*. 1999;78:42–55.
3. Janssen I, Heymsfield SB, Ross R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. *J Am Geriatr Soc*. 2002;50:889–896.
4. Janssen I, Shepard DS, Katzmarzyk PT, Roubenoff R. The health-care costs of sarcopenia in the United States. *J Am Geriatr Soc*. 2004;52:80–85. doi:10.1111/j.1532-5415.2004.52014.x.
5. Sierra F. Biology of aging summit report. *J Gerontol A Biol Sci Med Sci*. 2009;64:155–156. doi:10.1093/gerona/gln069.
6. Murphy TE, Han L, Allore HG, Peduzzi PN, Gill TM, Lin H. Treatment of death in the analysis of longitudinal studies of

gerontological outcomes. *J Gerontol A Biol Sci Med Sci*. 2011;66:109–114. doi:10.1093/gerona/glq188.

7. Fries JF. Aging, natural death, and the compression of morbidity. *N Engl J Med*. 1980;303:130–135. doi:10.1056/NEJM198007173030304.

8. Kirkland JL, Peterson C. Healthspan, translation, and new outcomes for animal studies of aging. *J Gerontol A Biol Sci Med Sci*. 2009;64:209–212. doi:10.1093/gerona/gln063.

9. National Institute on Aging. *Aged Rodent Colonies Handbook*. Available at: http://www.nia.nih.gov/ResearchInformation/ScientificResources/AgedRodentColoniesHandbook/StrainSurvivalInformation.htm. Accessed April 15, 2012.

10. Miller RA, Nadon NL. Principles of animal use for gerontological research. *J Gerontol A Biol Sci Med Sci*. 2000;55:B117–B123. doi:10.1093/gerona/55.3.B117.

11. Arias E. United States Life Tables, 2007. *National Vital Statistics Reports*, vol. 59, no. 9. Hyattsville, MD: National Center for Health Statistics; 2011. Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_09.pdf. Accessed February 2, 2012.

12. Ingram DK, Reynolds MA. Assessing the predictive validity of psychomotor tests as measures of biological age in mice. *Exp Aging Res*. 1986;12:155–162. doi:0.1080/03610738608259454.

13. Ingram DK, Archer JR, Harrison DE, Reynolds MA. Physiological and behavioral correlates of lifespan in aged C57BL/6J mice. *Exp Gerontol*. 1982;17:295–303.

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

15. Brooks SV, Faulkner JA. Contractile properties of skeletal muscles from young, adult and aged mice. *J Physiol (Lond)*. 1988;404:71–82.

16. Parks RJ, Fares E, Macdonald JK, et al. A procedure for creating a frailty index based on deficit accumulation in aging mice. *J Gerontol A Biol Sci Med Sci*. 2012;67:217–227. doi:10.1093/gerona/glr193.

17. Whetstone LM, Fozard JL, Metter EJ, et al. The physical functioning inventory: a procedure for assessing physical function in adults. *J Aging Health*. 2001;13:467–493. doi:10.1177/089826430101300402.

18. Pialoux T, Goyard J, Lesourd B. Screening tools for frailty in primary health care: a systematic review. *Geriatr Gerontol Int*. 2012;12:189–197. doi:10.1111/j.1447-0594.2011.00797.x.

19. Cheung HN, Power MJ. The development of a new multidimensional depression assessment scale: preliminary results. *Clin Psychol Psychother*. 2012;19:170–178. doi:10.1002/cpp.1782.

20. Borkan GA, Norris AH. Assessment of biological age using a profile of physical parameters. *J Gerontol*. 1980;35:177–184.

21. Brooks SV, Faulkner JA. Maximum and sustained power of extensor digitorum longus muscles from young, adult, and old mice. *J Gerontol*. 1991;46:B28–B33.

22. Narici MV, Maffulli N. Sarcopenia: characteristics, mechanisms and functional significance. *Br Med Bull*. 2010;95:139–159. doi:10.1093/bmb/ldq008.

23. Wright VJ, Perricelli BC. Age-related rates of decline in performance among elite senior athletes. *Am J Sports Med*. 2008;36:443–450. doi:10.1177/0363546507309673.

24. Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil*. 1993;74:531–536. doi:10.1016/0003-9993(93)90119-U.

25. Sim J, Arnell P. Measurement validity in physical therapy research. *Phys Ther*. 1993;73:102–110.

26. Ingram DK. Toward the behavioral assessment of biological aging in the laboratory mouse: concepts, terminology, and objectives. *Exp Aging Res*. 1983;9:225–238. doi:10.1080/03610738308258457.

27. Machackova J, Barta J, Dhalla NS. Myofibrillar remodeling in cardiac hypertrophy, heart failure and cardiomyopathies. *Can J Cardiol*. 2006;22:953–968 doi:10.1016/S0828-282X(06)70315-4.

28. Widmaier EP, Raff H, Strang KT. *Vander's Human Physiology The Mechanisms of Body Function*. 11th ed. New York: McGraw-Hill Higher Education; 2006.

29. Brunner F, Schmid A, Sheikhzadeh A, Nordin M, Yoon J, Frankel V. Effects of aging on Type II muscle fibers: a systematic review of the literature. *J Aging Phys Act*. 2007;15:336–348.

30. Fahlstrom A, Zeberg H, Ulfhake B. Changes in behaviors of male C57BL/6J mice across adult life span and effects of dietary restriction. *Age: J Am Aging Assoc*. 2012;34:1435–1452. doi:10.1007/s11357-011-9320-7.