Journals of Gerontology: BIOLOGICAL SCIENCES, 2015, 686–693
doi:10.1093/gerona/glu161
Original Article
Advance Access publication September 9, 2014

OXFORD

Original Article

Reliability of a Frailty Index Based on the Clinical Assessment of Health Deficits in Male C57BL/6J Mice

Hirad A. Feridooni,[1,*] Michael H. Sun,[1,*] Kenneth Rockwood,[2] and Susan E. Howlett[1,2,3]

[1]Department of Pharmacology and [2]Department of Medicine (Geriatric Medicine), Dalhousie University, Halifax, Nova Scotia, Canada. [3]Department of Cardiovascular Physiology, University of Manchester, UK.

*These authors contributed equally to this work.

Address correspondence to Susan E. Howlett, PhD, Department of Pharmacology, Dalhousie University, Sir Charles Tupper Medical Building, 5850 College Street, PO Box 15000, Halifax, NS B3H 4R2, Canada. Email: susan.howlett@dal.ca

Abstract

We investigated the reliability of a newly developed clinical frailty index (FI) that measures frailty based on deficit accumulation in aging mice. FI scores were measured by two different raters independently in a large cohort (n = 233) of 343–430 day-old male C57BL/6J mice. Inter-rater reliability was evaluated with correlation coefficients, the kappa statistic, and intra-class correlation coefficients (ICC) in three separate groups of mice (n = 45, 50, and 138 mice/group) sequentially over 3 months. After each group was evaluated, descriptions of techniques used to identify health deficits were amended. Mice had comparable overall FI scores regardless of rater (0.213±0.002 vs 0.212±0.002; p = .802), although discordant measures declined as techniques were refined. Correlation coefficients (r^2 values) between raters improved throughout the study and mean kappa values increased (mean ± SEM; 0.621±0.018, 0.764±0.017, and 0.836±0.009 for groups 1, 2, and 3; p < .05). Values for intra-class correlation coefficient also improved from .51 (95% confidence interval = 0.11–0.73) to .74 (0.54–0.85) and .77 (0.67–0.83). FI scores increased over 3 months (p < .05), but did not differ between raters. These results show a high overall inter-rater reliability when the clinical FI tool is used to assess frailty in a large cohort of mice.

Key Words: Frailty index—Deficit index—Deficit accumulation—Senescence—Ageing.

Decision Editor: Rafael de Cabo, PhD

Frailty can be defined as a state of increased vulnerability to adverse health outcomes for people of the same chronological age (1). It represents a major challenge in the clinical care of older adults, as frail individuals have longer hospitalizations, worse outcomes and higher mortality than do fit people (2). Despite the recognition that frailty is a major health care problem, the biology of frailty is not well understood. The limited success in linking the basic biology of aging with frailty arises, at least in part, because we lack scales to evaluate frailty in experimental models (3). The ability to quantify frailty in aging animal models is essential if we are to understand its biology and develop interventions that can attenuate frailty by targeting fundamental mechanisms of aging (3,4).

Several scales for the quantification of frailty have been used to measure frailty in people (5). One common approach is to quantify frailty with a "frailty index (FI)," in which an individual's potential health deficits (eg, clinical signs, diseases, laboratory abnormalities,

etc.) are counted and divided by the total number of items measured (6–8). We recently used a modification of this approach with a FI based on deficit accumulation in aging mice (9). We measured more than 30 health-related variables that provided information about activity levels, hemodynamic status, body composition, and metabolism in adult (12 month-old) and aged (30 month-old) mice of both sexes. We found that aged mice had significantly higher FI scores than did younger animals and that high FI scores predicted deficits in the structure and function of individual heart cells (9). Thus, a FI based on deficit accumulation can be used to quantify frailty and predict adverse outcomes in aging mice.

The techniques used to construct the FI in our original study were time-consuming, required access to specialized equipment and employed invasive methods (9). These requirements limit the utility of this method, especially in longitudinal studies of frailty in aging animals. To address this concern, we recently developed a simplified, non-invasive FI based on the clinical assessment of more than 30 potential deficits in a small cohort (n = 14) of aging mice (10). We used a checklist that combined readily apparent, published signs of clinical deterioration in mice. Our results showed that this simplified approach could be used to characterize frailty in aging mice (10) and that the FI scores achieved with this approach were similar to those measured in our original study (9). Importantly, we showed that the relationship between FI scores and age was virtually identical in mice and humans, when age was normalized to the maximal lifespan for each group (10).

The scales used to construct the mouse clinical FI require assessment across a range of domains, including the evaluation of integrative measures such grooming, strength, mobility, and measures of discomfort (10). Even though the article by Whitehead et al. (10) describes a detailed scoring system, it is possible that clinical impressions may vary from rater to rater. We have shown that the clinical FI scores exhibit very little test-to-test variability when administered by a single rater (10). Still, and especially if the clinical FI is to be used as an outcome measure to determine whether interventions can modify frailty in aging animal models, it must be reliable when used by different raters. The objectives of this study were: (a) to evaluate the mouse clinical FI in a large cohort of aging mice; (b) to determine the inter-rater reliability of this instrument and identify discordant measures; and (c) to refine the criteria used to construct the murine FI. The study used 343- to 430-day-old male C57BL/6J mice. Inter-rater reliability was measured in three groups of mice with standard correlation coefficients, the kappa statistic and with intra-class correlation coefficients (ICCs).

Methods

Experimental Animals

Three- to four-week-old male C57BL/6J mice (n = 233) were purchased from Charles River (St. Constant, Quebec). The mice were housed in groups in micro-isolator cages in the Carlton Animal Care Facility at Dalhousie University and aged to approximately one year before use in the present study. In a few experiments, young adult mice (≈6 months of age) were used. Mice were exposed to a 12-hour light/dark cycle and they had free access to food and water. The mice were fed a standard laboratory rodent diet (ProLab RMH 3500, Purina LabDiet, Aberfoyle, Ontario, Canada). Experiments followed the Canadian Council on Animal Care *Guide to the Care and Use of Experimental Animals* (CCAC, Ottawa, ON: Vol. 1, 2nd edition, 1993; Vol. 2, 1984); all protocols were approved by the Dalhousie University Committee on Laboratory Animals.

Measurement of Frailty With the Clinical Frailty Index

Two different raters independently calculated a unique FI score for each mouse based on the murine clinical FI tool we described previously and following the criteria outlined in that article (10). Assessments were performed between 10 am and 2 pm each day. Briefly, mice were placed in a fresh cage and moved to a dedicated small animal procedure room in the Carlton Animal Care Facility for evaluation. This procedure room was designed for behavioral testing, is located at the end of a quiet hall in the facility and we were its sole occupants during testing. Mice were weighed and their body surface temperature was measured at the abdomen with an infrared temperature probe (Infrascan; La Crosse Technology). An average of three temperature readings was used. The hearing test used a clicker of the type used to train dogs. The clinical FI score for each mouse was calculated using the checklist published previously (10). Clinical assessment included evaluation of the integument, musculoskeletal system, vestibulocochlear and auditory systems, ocular and nasal systems, digestive system, urogenital system, respiratory system, signs of discomfort, as well as the body weight (g) and body surface temperature (°C). A complete list of the clinical signs of deterioration and/or deficits evaluated in this study can be found in Supplementary Table 1.

Calculation of the FI Score

A simple deficit rating scale was used to compute the FI score for each animal. For each parameter, a score of 0 was given if there was no sign of a deficit, a score of 0.5 denoted a mild deficit and a score of 1 indicated a severe deficit. Deficits in body weight (g) and body surface temperature (°C) were scored based on their deviation from average reference values obtained from the entire cohort. Mean (±*SD*) reference values for weight were 48.6 ± 4.8 g and 48.7 ± 4.8 g for raters 1 and 2, respectively; average reference values for temperature were 30.6 ± 0.9°C for rater 1 and 30.2 ± 0.8°C for rater 2. Values that differed from reference values by less than 1 *SD* were scored as 0. Values that were ±1 *SD* with respect to the reference value were given a frailty value of 0.25; values that differed by ±2 *SD* scored 0.5, those that differed by ±3 *SD* scored 0.75 and values that were >3 *SD* above or below the mean received the maximal frailty value of 1. The frailty score for each of the 31 items on the checklist were added and the total was divided by the number of deficits measured (eg, 31 deficits) to yield a FI score between 0 and 1 for each animal. The possible frailty scores for each deficit are also illustrated in Supplementary Table 1.

Study Design

The mice were divided into three groups, an initial group with 45 mice (group 1), a second group with 50 mice (group 2), and a third group with 138 mice (group 3) for a total of 233 mice. After each group of mice had been evaluated by both raters, the scores were compared and areas of discrepancy were identified. After discussion between the two raters, techniques were refined and the descriptions of the criteria for clinical assessment of deficits were revised and clarified. Next, the second group of mice was evaluated and scores compared between raters as above. The refinement procedure was repeated and the final group of mice was evaluated.

Statistics

Data are presented as either the mean ± SEM or the mean ± *SD*, as indicated. Differences in FI scores between raters were calculated with a Student's *t*-test. Inter-rater reliability was measured in each of the three groups of mice in three ways: (a) Reliability was

compared with standard correlation coefficients. FI data obtained by raters 1 and 2 were fit with a simple linear regression and square of the correlation coefficient (r^2) was calculated to determine whether a linear relationship existed between scores measured by the two raters. (b) Inter-rater reliability was also calculated with the Cohen's kappa statistic, which takes into account agreement between raters that would occur by chance. An individual kappa value was calculated for each mouse and differences between the three groups of mice were evaluated with one-way analysis of variance. (c) The final test used to evaluate inter-rater reliability was the ICC with a two-way random model and consistency analysis; the 95% confidence interval (CI) was calculated for each ICC. In all cases, differences between groups were considered statistically significant when $p < .05$. Statistical analyses were performed either with SPSS (IBM SPSS Statistics, Version 21) or with Sigma Plot 11.0 (Systat Software, Inc., Point Richmond, CA). Graphs were created with Sigma Plot 11.0.

Results

Mean (±*SD*) physical characteristics of the three groups of mice as determined by each of the raters are shown in Table 1. As animals were rated on the same day by each rater, age was identical for both raters but increased significantly over the course of the study (Table 1). Mean values for weight did not differ between groups or raters (Table 1). Body surface temperature did vary between raters and in some cases between groups (Table 1). Even though temperature varied significantly, the variation was very small and is not likely to be biologically significant.

Figure 1A shows a scatterplot of the relationship between the FI and age for all the mice examined in this study by both raters. The figure shows that the FI scores generally increased with age, but individual scores at each age were highly variable (Figure 1A). Figure 1B shows that that there were no significant differences in the average (±SEM) FI scores obtained by raters 1 and 2 for any of the groups of mice examined in this study. Furthermore, the overall FI scores for all the mice used in the study were not significantly different between the two raters; values were 0.213 ± 0.002 for rater 1 and 0.212 ± 0.002 for rater 2 (mean ± SEM; $p = .802$; $n = 233$). On the other hand, Figure 1B shows that mean (±*SD*) scores for rater 1 increased with age (0.18 ± 0.03 for group 1; 0.21 ± 0.04 for group 2; 0.22 ± 0.03 for group 3); average scores for rater 2 increased between group 1 and group 2 (0.18 ± 0.03 to 0.22 ± 0.03) and then plateaued for group 3 (0.22 ± 0.03). Of note, FI scores were significantly higher in groups 2 and 3 when compared to group 1 as the mice increased in age (Figure 1B).

Table 1. Characteristics of Male C57BL/6J Mice Used in this Study

Characteristic*	Rater 1	Rater 2
Group 1		
Age (days)	349.6 ± 6.3	349.6 ± 6.3
Weight (g)	47.6 ± 5.6	47.5 ± 5.8
Body surface temperature (°C)	31.3 ± 0.8	30.9 ± 0.6†
Number of mice	45	45
Group 2		
Age (days)	374.8 ± 3.8‡	374.8 ± 3.8‡
Weight (g)	48.9 ± 4.0	49.0 ± 3.9
Body surface temperature (°C)	30.7 ± 0.9‡	30.1 ± 0.8†,‡
Number of mice	50	50
Group 3		
Age (days)	405.2 ± 11.8‡,§	405.2 ± 11.8‡,§
Weight (g)	48.9 ± 4.8	49.0 ± 4.7
Body surface temperature (°C)	30.4 ± 0.9‡,§	30.0 ± 0.8†,‡
Number of mice	138	138

Notes: *Values represent the mean ± *SD*. Weight and body surface temperature data were evaluated with two-way ANOVA with rater and group as main factors; differences between groups for age were assessed with a one-way ANOVA on ranks.

†Denotes significantly different from rater 1.

‡Denotes significantly different from group 1.

§Denotes significantly different from group 2.

Figure 2A shows the number of differences between raters for each individual item used to make up the FI score. The data are expressed as a percentage of the differences between raters in each of the three groups of mice examined. Items that differed by more than 25% were identified, as shown by the dashed line (Figure 2A). Figure 2A shows that the number of discrepancies between raters was highest for Group 1. Items that differed by more than 25% were: distended abdomen, gait, tremor, grip strength, body condition, head tilt, hearing loss, menace reflex, breathing rate/depth, and piloerection. Raters compared rating procedures, expanded the descriptions of techniques used for clinical assessment and evaluated mice in group 2. Figure 2A shows that the number of discrepancies declined for group 2, but still included hunched posture, tremor, hearing loss, menace reflex, and piloerection. The raters again refined and expanded the assessment criteria and evaluated group 3. The number of discrepancies again declined and only the hearing test and temperature varied by more than 25% between raters.

As shown in Figure 2A, the most disagreement between raters occurred with respect to body surface temperature and hearing loss. Importantly, these discrepancies were not resolved over the course of the study, so additional experiments were performed. The mice were originally tested in the experimental room in groups of 10. To determine whether the mice habituated to the sound of the clicker in the room, a separate group of young adult mice ($n = 11$) that could hear at baseline were repeatedly exposed to the clicker. Figure 2B shows that the percentage of mice responding to the clicker declined as the number of clicks increased. This demonstrates that the hearing test was not reliable unless the sound was novel. Differences between raters with respect to body temperature were also investigated further. Discrepancies were due to differences in the position of the probe relative to the mouse. We found that reliable and consistent recordings of body temperature could be made when the probe was positioned 2 cm directly above the centre of the abdomen. Based on the results of these investigations, the criteria and descriptions of the procedures used to construct the FI were modified. These modifications are shown as the entries in italics in Supplementary Table 2.

Reliability between raters was initially assessed with standard correlation coefficients, as shown in Figure 3. FI scores from rater 1 were plotted as a function of scores from rater 2 for each mouse and the data were fit with a simple linear regression (Figure 3A). For group 1, the square of the correlation coefficient (r^2) was .12 ($p = .02$). Figure 3B and C shows the values of r^2 increased from .34 ($p < .001$) for group 2 to .39 ($p < .001$) for group 3. We also used the kappa statistic to compare inter-rater reliability. Figure 4A shows that the mean kappa values improved over the course of the study (values increased from 0.61 ± 0.13 to 0.75 ± 0.11 and 0.82 ± 0.10 in groups 1, 2, and 3, respectively; $p < .05$). Figure 4B shows the average values for the ICC also increased from .51 (95% CI = 0.11–0.73) in group 1 to .74 (CI = 0.54–0.85), and .77 (CI = 0.67–0.83) in groups 2 and 3. This increase in ICC was statistically significant (Figure 4B; $p < .05$).

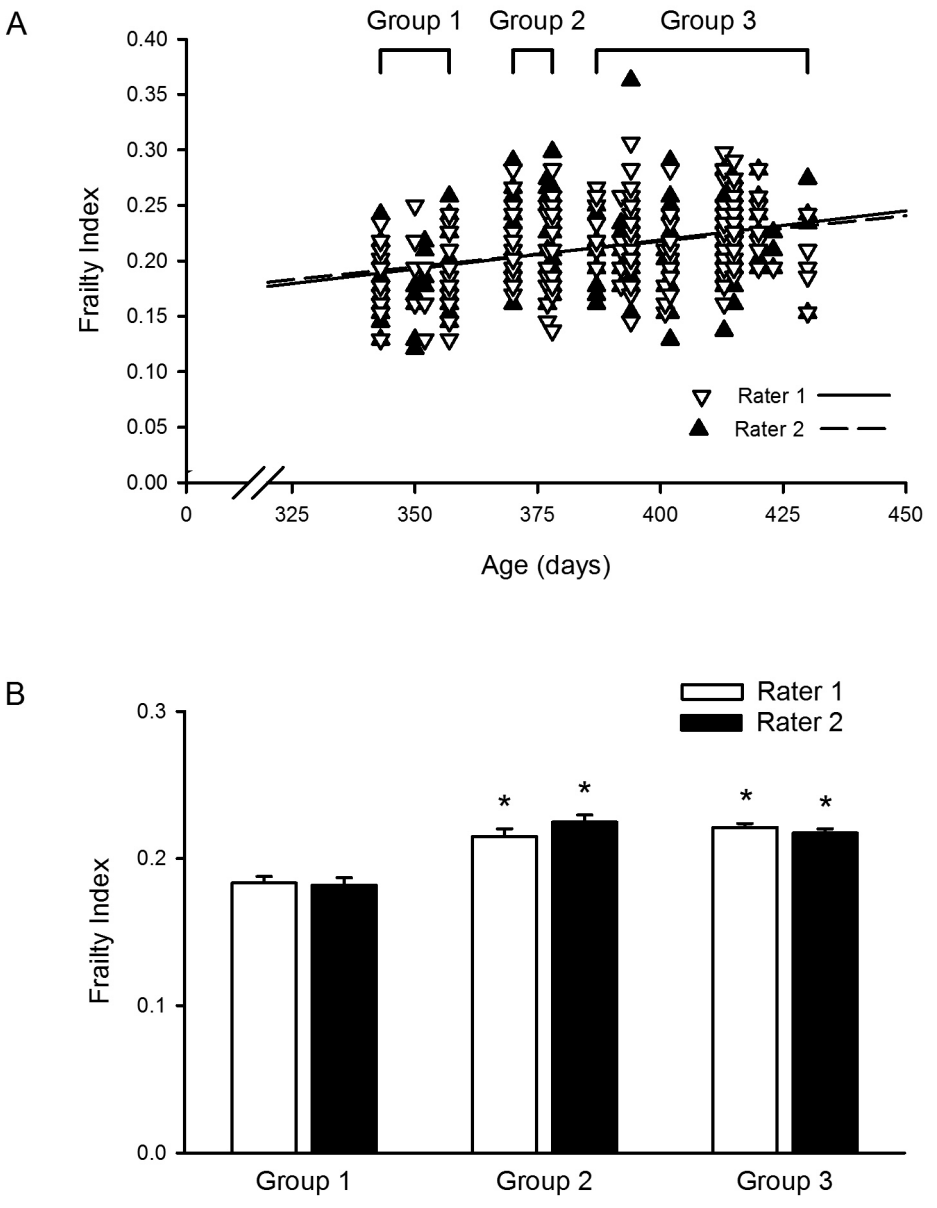

Figure 1. Frailty index scores in C57BL/6J mice. (**A**) Scatterplot of the relationship between the frailty index and age for scores from both raters shows that frailty index scores increased with age. Data from all mice evaluated in the study are illustrated in this panel (n = 233); each subgroup is identified at the top of this panel. (**B**) Values for the frailty index did not differ between raters for any of the groups used in this study, although there was an overall increase in frailty as mice increased in age. Frailty index scores for mice in group 1 were significantly lower than frailty index values in groups 2 and 3. Values represent the mean ± SEM. *Denotes significantly different from group 1 ($p < .05$; n = 45, 55, and 138 mice in groups 1, 2, and 3, respectively). Data were analyzed with a two-way ANOVA with rater and group as main factors.

Discussion

The overall goals of this study were to evaluate the newly described mouse clinical FI in a large cohort of 343- to 430-day-old mice, to determine the reliability of this instrument and to refine the techniques used to construct the index. When FI scores were compared across a range of ages in a large number of C57BL/6J mice, results showed that there were considerable differences in health status for mice of the same chronological age. This is consistent with the definition of frailty as variable vulnerability in animals of the same age. Interestingly, scores did not differ between raters for any of the three groups examined, although the number of discordant measures between raters declined as the techniques used to evaluate frailty were refined. This improvement in reliability was quantified as an increase in the correlation coefficients (r^2 values) between raters as the study progressed. Furthermore, both the average kappa values and the ICC values increased throughout the study. These data demonstrate that the relationship between health status, as assessed by the clinical FI, and chronological age is highly variable in older, 343- to 430-day-old C57BL/6J mice. Despite this variability, similar FI scores were obtained by two different raters and refinement of the techniques used to evaluate health deficits that make up the index led to a very high level of inter-rater reliability. These enhancements should improve the utility of this index as a tool to assess frailty in aging mice.

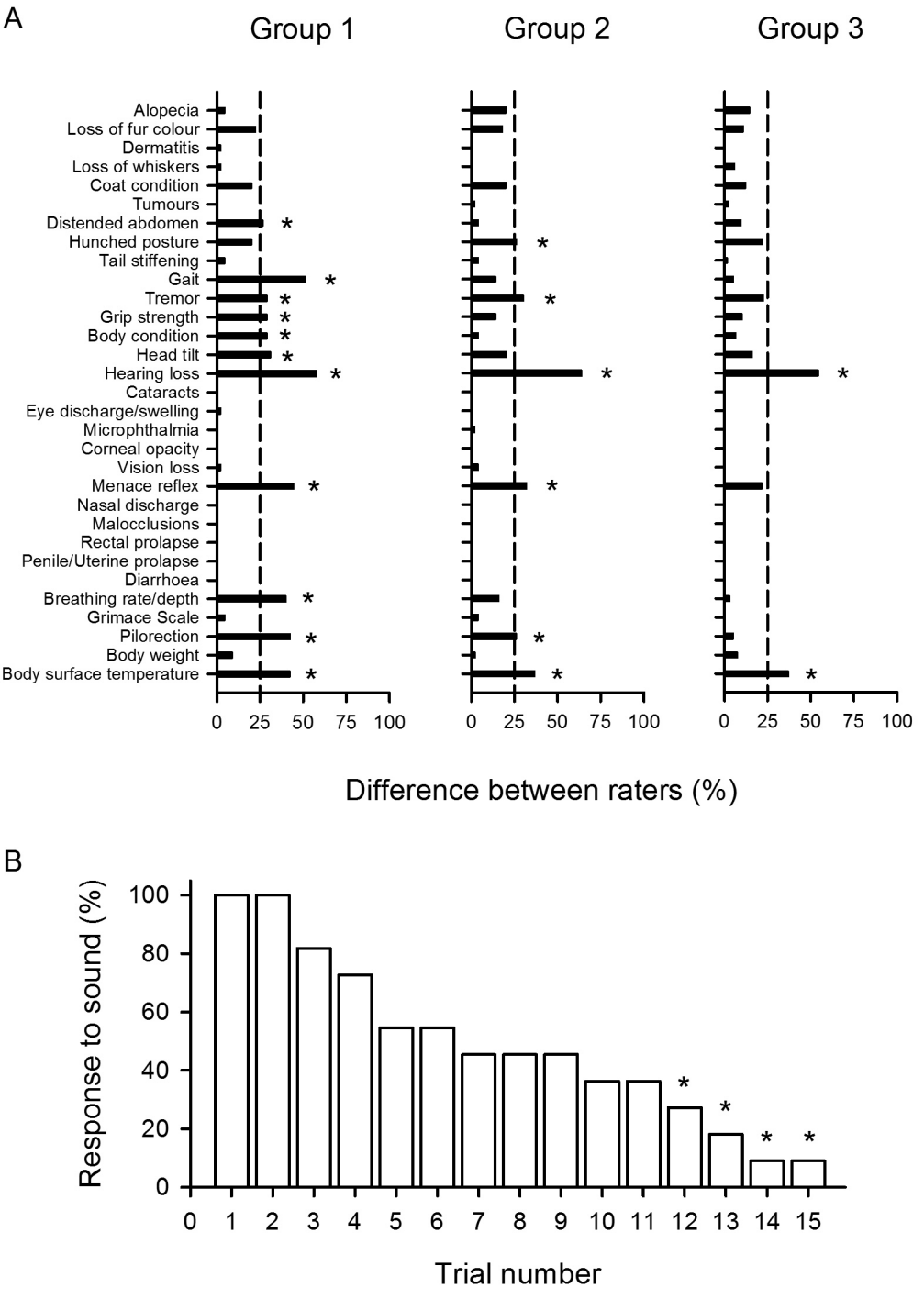

Figure 2. Refinement of techniques reduced differences between raters in the parameters used to construct the frailty index. (**A**) The number of differences between raters for each of the 31 deficits used to construct the frailty index was expressed as a percentage of the total number of mice in each group. For most items, differences between raters declined from group 1 to 3. The dashed line indicates a difference between raters of 25%; the asterisks indicate items that differed by more than 25% between raters (n = 45, 55, and 138 mice in groups 1, 2, and 3, respectively). (**B**) The test for hearing loss in the clinical frailty index was investigated in separate experiments. The percentage of young adult mice that responded to the clicker sound was plotted as a function of the number of clicks (trial number). The percentage of mice that responded to the clicker declined with repeated exposure. Data represent the number of mice that responded to the sound divided by the total number of young mice tested (n = 11 mice). *Denotes significantly different from trial 1 ($p < .05$; ANOVA on ranks).

The ability to quantify frailty in aging animal models has been identified as a key step in the effort to link the biology of aging with frailty (3). Indeed, several groups have recently developed different approaches to recognize and quantify frailty in aging animal models (9–12). These studies have generally adapted frailty scales that are commonly used to quantify frailty in people. For example, Liu et al. (12) developed a novel murine frailty scale based on the 5-point clinical "frailty phenotype" proposed by Fried et al. (13). In contrast, we have used a modification of the approach developed by Rockwood, Mitnitski et al. in humans (6,7), where frailty in mice is quantified in a FI measured as

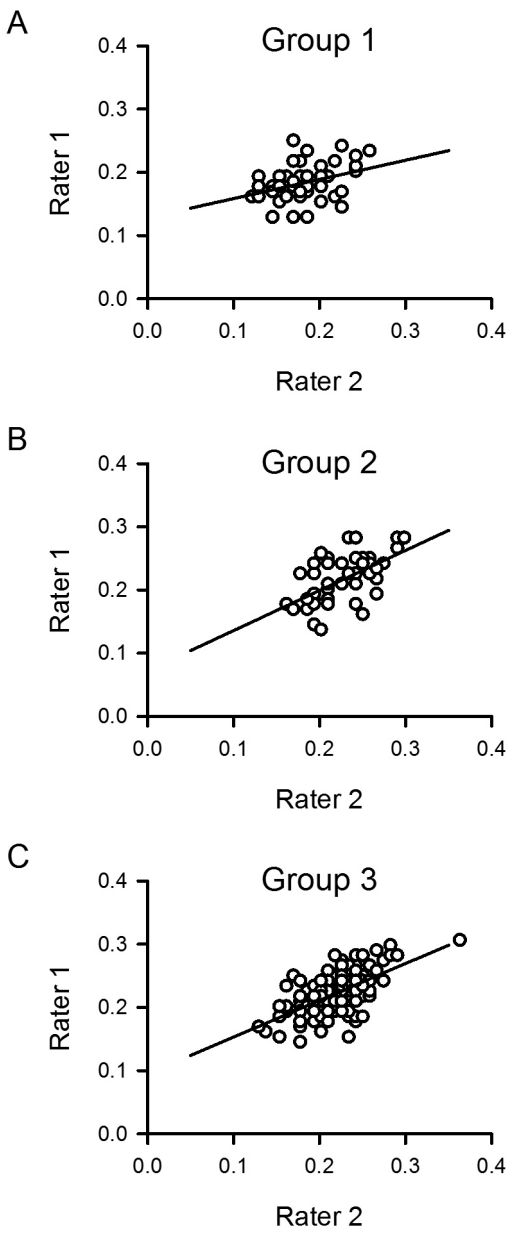

Figure 3. Reliability between raters compared with standard correlation coefficients. (**A**) Frailty index scores were obtained by each rater for mice in group 1. Data from rater 1 were plotted as a function of data from rater 2 and fit with a simple linear regression. The square of the correlation coefficient (r^2) was .12 (p = .02). (**B**, **C**) A similar approach was used to compare inter-rater reliability with correlation coefficients for groups 2 and 3. Values for r^2 increased from .34 (p < .001) for group 2 to .39 (p < .001) for group 3. (n = 45, 55, and 138 mice in groups 1, 2, and 3, respectively).

deficit accumulation (9,10). These novel assessment tools are an exciting development in the biology of frailty as they can potentially be used to quantify frailty and investigate the success of treatments to attenuate frailty in pre-clinical models. However, for clinically based frailty scales to be useful in different settings, they must be reliable (14–16). We previously showed that the murine clinical FI developed by our group showed little test-to-test variability when administered by a single rater (10). A major advance made in the present study is the demonstration that this clinical FI exhibits a high degree of inter-rater reliability when used in a large cohort of mice, so it is a reliable assessment tool.

In our initial study, we developed a standardized scoring system to measure health deficits in aging mice with a brief clinical exam (10). In the present study, when two independent raters used this scoring system to measure frailty, we found that there was some initial disagreement between raters on several health deficit measures in the first clinical evaluation. An important contribution made by the present study is that we have identified those items most likely to cause disagreement and we have more fully described the assessment procedure for each of these items. The expanded descriptions of the criteria used to define health deficits should help other laboratories operationalize this clinical FI.

When the FI is used to assess health status in humans, the relationship between health status and chronological age is highly variable (17,18), even though relative heterogeneity (coefficient of variation) declines with age (18). In our original description of the clinical FI, we found that the absolute variability of the index appeared to increase with age in a very small cohort (n = 14) of aging mice (10). In the present study, we have extended these observations to include data from a large number of C57BL/6J mice (n = 233) between the ages of 343–430 days of age. When we used the FI to assess the health status of these mice, we found that there was a great degree of variability in the health status of mice of the same age. These data demonstrate that the link between chronological age and health is highly variable, even in mice with similar genetic backgrounds, and suggest that population aging is diverse in these animals. Studies of interventions designed to influence frailty in animal models could select mice with different initial frailty levels to investigate the impact of potential treatments on mice with initial high or low frailty loads.

There is evidence that inflammation makes an important contribution to the development frailty in humans (19,20). Indeed, some studies that have investigated healthspan and frailty in animal models have focused on inflammation as a hallmark of frailty. For example, the interleukin-10 knockout mouse ($IL10^{tm/tm}$), which exhibits inflammation and an age-dependent reduction in skeletal muscle strength, also has been used to model frailty (21–24). As we used a non-invasive assessment tool to quantify frailty in this study, we did not directly evaluate the level of inflammation in the mice used in our study. However, in our previous work (10) we showed that dermatitis, which has been linked to inflammation (25), increased with age and frailty. This observation provides indirect evidence that inflammation is increased in frail older mice.

There is also evidence that sarcopenia contributes to the development frailty in humans (19,20). While sarcopenia was not investigated here, our clinical FI tool includes assessment of grip strength, gait disorders, and tremor, so it does reflect deficits in physical condition. Furthermore, in a previous study we compared clinical FI data with data from a FI based on performance measures in an open field (10). We found that higher clinical FI scores were associated with impaired performance as measured by activity levels (eg, total distance moved; average velocity of movement; rearing frequency). Therefore, high clinical FI scores are associated with functional impairment (10). We also previously used a dual energy X-ray absorptiometry (DEXA) scanner to demonstrate that changes in animal weight and body composition account for much of the FI variance when the FI is measured with a more invasive approach (9). Interestingly, Thompson and colleagues have proposed both a neuromuscular healthspan scoring system (11) and a FI based on physical signs of weakness (12) as tools to evaluate frailty in aging mice. A direct comparison of frailty levels obtained with our approach (10) and with the physical frailty methods described by others (11,12) in the same mice could be interesting.

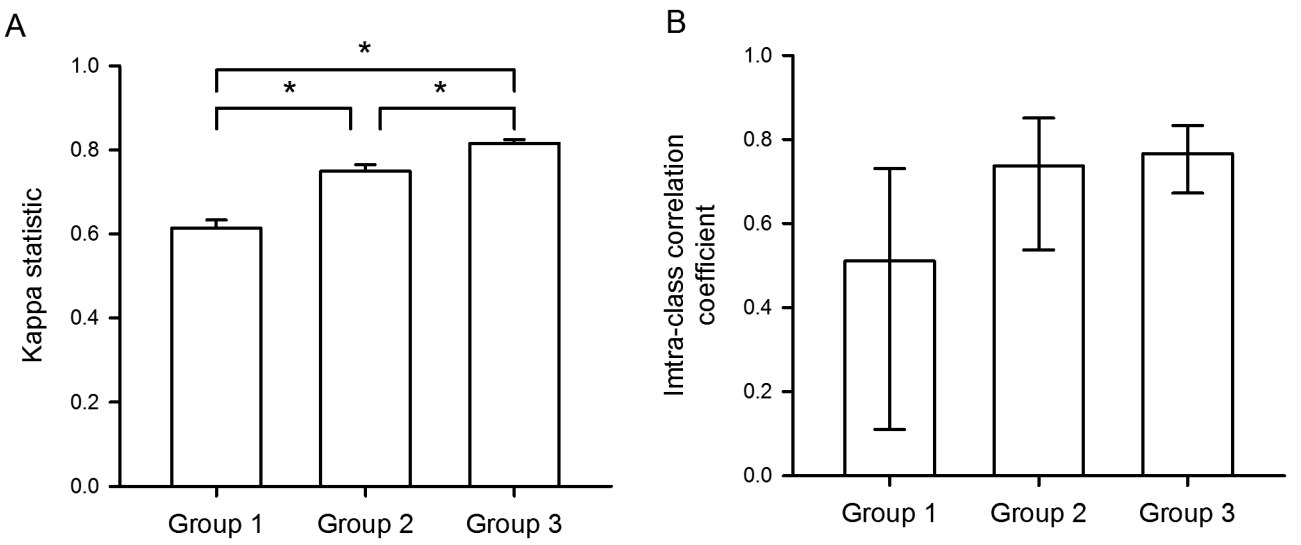

Figure 4. Inter-rater reliability as assessed by the kappa statistic and the ICC. (**A**) The kappa statistic was used to compare reliability between raters. Reliability increased progressively from group 1 to 3. Values represent the mean ± SEM and differences were evaluated with one-way analysis of variance. (**B**) The ICC was also used to evaluate inter-rater reliability. Values represent the correlation coefficients ± 95% CIs. ANOVA demonstrated that reliability between raters was statistically significant for the ICC. *Denotes significantly different from comparison group. In all cases, $p < .05$ and $n = 45$, 55, and 138 mice in groups 1, 2, and 3, respectively.

There are some limitations to the data presented here. We report FI data obtained from male C57BL/6J mice only, so results may not be directly applicable to female mice or to other strains of mice. It is possible that there are male–female differences in frailty in mice, especially since there is some evidence for sex differences in frailty in humans with most studies reporting that women have higher frailty levels than men (26). Still, whether there are sex differences in frailty in animal models is not yet clear. We did include a "head-to-head" comparison of male–female differences in our initial, small scale study of frailty in mice (9). Although we found that older males had higher FI scores than older females, this effect was not statistically significant (9). In a more recent study with the frailty assessment tool used in the present manuscript we found the opposite trend, with males somewhat less frail than females, although again this difference was not statistically significant (10). At present there is no evidence for a sex difference in frailty in mice and it may be that any sex difference is small and will only be detected in a larger sample.

Another potential limitation is the accuracy of the body surface temperature measurements made with an infrared temperature probe. To ensure the accuracy, we used an average of three temperature readings from each mouse. We found that the variance for temperature measurements was very low, which suggests that our technique is reproducible. An alternative approach would be to use a rectal probe to measure body temperature, although this would be a more invasive approach. Body temperature is an important variable to include in the FI as there is evidence that temperature declines between the ages of 2 and 30 months in male C57BL/6J mice (27). Importantly, studies have shown that a marked decline in body temperature occurs during the last 16 weeks of life in the mouse model (28), which suggests that a rapid decline in body temperature can be used as a marker imminent death.

The results of this study demonstrated that, even though FI scores increased with age, there was considerable variability in FI scores for mice of the same chronological age in this large cohort of C57BL/6J mice. This indicates that the link between chronological age and health is highly variable, even in mice with similar genetic backgrounds. This study also showed that the clinical FI tool exhibited high overall inter-rater reliability and that its reliability increased as the techniques used to evaluate clinical deficits were refined throughout the study. This novel assessment tool may be useful in evaluating the success of treatments designed to attenuate frailty and improve health in pre-clinical models, with the ultimate goal of translating findings to frail older adults.

Supplementary Material

Supplementary material can be found at: http://biomedgerontology.oxfordjournals.org/

Funding

This study was supported by grants from the Canadian Institutes for Health Research (MOP 126018) and the Fountain Innovation Fund of the Queen Elizabeth II Health Sciences Foundation. Kenneth Rockwood receives career support from the Dalhousie Medical Research Foundation as the Kathryn Allen Weldon Professor of Alzheimer Research.

Acknowledgments

The authors express their appreciation for excellent technical assistance provided by Peter Nicholl.

References

1. Rockwood K, Fox RA, Stolee P, Robertson D, Beattie BL. Frailty in elderly people: an evolving concept. *CMAJ*. 1994;150:489–495.
2. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. 2013;381:752–762. doi:10.1016/S0140-6736(12)62167-9
3. Kirkland JL. Translating advances from the basic biology of aging into clinical application. *Exp Gerontol*. 2013;48:1–5. doi:10.1016/j.exger.2012.11.014

4. Howlett SE, Rockwood K. New horizons in frailty: ageing and the deficit-scaling problem. *Age Ageing*. 2013;42:416–423. doi:10.1093/ageing/aft059
5. de Vries NM, Staal JB, van Ravensberg CD, Hobbelen JS, Olde Rikkert MG, Nijhuis-van der Sanden MW. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev*. 2011;10:104–114. doi:10.1016/j.arr.2010.09.001
6. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *ScientificWorldJournal*. 2001;1:323–336. doi:10.1100/tsw.2001.58
7. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8:24. doi:10.1186/1471-2318-8-24
8. Kulminski AM, Ukraintseva SV, Kulminskaya IV, Arbeev KG, Land K, Yashin AI. Cumulative deficits better characterize susceptibility to death in elderly people than phenotypic frailty: lessons from the Cardiovascular Health Study. *J Am Geriatr Soc*. 2008;56:898–903. doi:10.1111/j.1532-5415.2008.01656.x
9. 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
10. Whitehead JC, Hildebrand BA, Sun M, et al. A clinical frailty index in aging mice: comparisons with frailty index data in humans. *J Gerontol A Biol Sci Med Sci*. 2014;69:621–632. doi:10.1093/gerona/glt136
11. Graber TG, Ferguson-Stegall L, Kim JH, Thompson LV. C57BL/6 neuromuscular healthspan scoring system. *J Gerontol A Biol Sci Med Sci*. 2013;68:1326–1336. doi:10.1093/gerona/glt032
12. Liu H, Graber TG, Ferguson-Stegall L, Thompson LV. Clinically relevant frailty index for mice. *J Gerontol A Biol Sci Med Sci*. 2013. doi:10.1093/gerona/glt188
13. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56:M146–M156. doi:10.1093/gerona/56.3.M146
14. Hilmer SN, Perera V, Mitchell S, et al. The assessment of frailty in older people in acute care. *Australas J Ageing*. 2009;28:182–188. doi:10.1111/j.1741-6612.2009.00367.x
15. Jones DM, Song X, Rockwood K. Operationalizing a frailty index from a standardized comprehensive geriatric assessment. *J Am Geriatr Soc*. 2004;52:1929–1933. doi:10.1111/j.1532-5415.2004.52521.x
16. van Kempen JA, Schers HJ, Melis RJ, Olde Rikkert MG. Construct validity and reliability of a two-step tool for the identification of frail older people in primary care. *J Clin Epidemiol*. 2014;67:176–183. doi:10.1016/j.jclinepi.2013.08.008
17. Romero-Ortuno R. Frailty index in Europeans: association with determinants of health. *Geriatr Gerontol Int*. 2014;14:420–429. doi:10.1111/ggi.12122
18. Rockwood K, Mogilner A, Mitnitski A. Changes with age in the distribution of a frailty index. *Mech Ageing Dev*. 2004;125:517–519. doi:10.1016/j.mad.2004.05.003
19. Fulop T, Larbi A, Witkowski JM, et al. Aging, frailty and age-related diseases. *Biogerontology*. 2010;11:547–563. doi:10.1007/s10522-010-9287-2
20. Walston J, Hadley EC, Ferrucci L, et al. Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. *J Am Geriatr Soc*. 2006;54:991–1001. doi:10.1111/j.1532-5415.2006.00745.x
21. Akki A, Yang H, Gupta A, et al. Skeletal muscle ATP kinetics are impaired in frail mice. *Age (Dordr)*. 2014;36:21–30. doi:10.1007/s11357-013-9540-0
22. Ho YY, Matteini AM, Beamer B, et al. Exploring biologically relevant pathways in frailty. *J Gerontol A Biol Sci Med Sci*. 2011;66:975–979. doi:10.1093/gerona/glr061
23. Ko F, Yu Q, Xue QL, et al. Inflammation and mortality in a frail mouse model. *Age (Dordr)*. 2012;34:705–715. doi:10.1007/s11357-011-9269-6
24. Walston J, Fedarko N, Yang H, et al. The physical and biological characterization of a frail mouse model. *J Gerontol A Biol Sci Med Sci*. 2008;63:391–398.
25. Neuhaus B, Niessen CM, Mesaros A, Withers DJ, Krieg T, Partridge L. Experimental analysis of risk factors for ulcerative dermatitis in mice. *Exp Dermatol*. 2012;21:712–713. doi:10.1111/j.1600-0625.2012.01558.x
26. Hubbard RE, Rockwood K. Frailty in older women. *Maturitas*. 2011;69:203–207. doi:10.1016/j.maturitas.2011.04.006
27. Eleftheriou BE. Changes with age in protein-bound iodine (PBI) and body temperature in the mouse. *J Gerontol*. 1975;30:417–421. doi:10.1093/geronj/30.4.417
28. Trammell RA, Cox L, Toth LA. Markers for heightened monitoring, imminent death, and euthanasia in aged inbred mice. *Comp Med*. 2012;62:172–178.