




Comprehensive longitudinal non-invasive quantification of healthspan and frailty in a large cohort ($n = 546$) of geriatric C57BL/6 J mice

Serena Marcozzi · Giorgia Bigossi · Maria Elisa Giuliani · Robertina Giacconi · Maurizio Cardelli · Francesco Piacenza · Fiorenza Orlando · Agnese Segala · Alessandra Valerio · Enzo Nisoli · Dario Brunetti · Annibale Puca · Federico Boschi · Carlo Gaetano · Alessia Mongelli · Fabrizia Lattanzio · Mauro Provinciali · Marco Malavolta 

Received: 15 November 2022 / Accepted: 17 January 2023

© The Author(s), under exclusive licence to American Aging Association 2023

Abstract Frailty is an age-related condition characterized by a multisystem functional decline, increased vulnerability to stressors, and adverse health outcomes. Quantifying the degree of frailty in humans and animals is a health measure useful for translational geroscience research. Two frailty measurements, namely the frailty phenotype (FP) and the clinical frailty index (CFI), have been validated in mice and are frequently applied in preclinical research. However, these two tools are based on different concepts and do not necessarily identify the same mice

as frail. In particular, the FP is based on a dichotomous classification that suffers from high sample size requirements and misclassification problems. Based on the monthly longitudinal non-invasive assessment of frailty in a large cohort of mice, here we develop an alternative scoring method, which we called physical function score (PFS), proposed as a continuous variable that resumes into a unique function, the five criteria included in the FP. This score would not only reduce misclassification of frailty but it also makes the two tools, PFS and CFI, integrable to provide an overall measurement of health, named vitality score (VS) in aging mice. VS displays a higher association with mortality than PFS or CFI and correlates with

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11357-023-00737-1>.

S. Marcozzi · G. Bigossi · M. E. Giuliani · R. Giacconi · M. Cardelli · F. Piacenza · M. Provinciali · M. Malavolta (✉)
Advanced Technology Center for Aging Research, IRCCS INRCA, 60121 Ancona, Italy
e-mail: m.malavolta@inrca.it

S. Marcozzi · F. Lattanzio
Scientific Direction, IRCCS INRCA, 60124 Ancona, Italy

F. Orlando
Experimental Animal Models for Aging Unit,
Scientific Technological Area, IRCCS INRCA,
60015 Falconara Marittima (AN), Italy

A. Segala · A. Valerio
Department of Molecular and Translational Medicine,
University of Brescia, Viale Europa, 11, 25123 Brescia,
Italy

E. Nisoli
Center for Study and Research On Obesity, Department
of Medical Biotechnology and Translational Medicine,
University of Milan, Via Vanvitelli, 32, 20129 Milan, Italy

D. Brunetti
Medical Genetics and Neurogenetics Unit, Fondazione
IRCCS Istituto Neurologico Carlo Besta, 20126 Milan,
Italy

D. Brunetti
Department of Medical Biotechnology and Translational
Medicine, University of Milan, 20129 Milan, Italy

A. Puca
Department of Medicine, Surgery and Dentistry “Scuola
Medica Salernitana”, University of Salerno, Via Salvatore
Allende, 84081 Baronissi, Salerno, Italy

biomarkers related to the accumulation of senescent cells and the epigenetic clock. This longitudinal non-invasive assessment strategy and the VS may help to overcome the different sensitivity in frailty identification, reduce the sample size in longitudinal experiments, and establish the effectiveness of therapeutic/preventive interventions for frailty or other age-related diseases in geriatric animals.

Keywords Biomarkers of aging · Biogerontology · Cellular senescence · Epigenetic clock · Physical performance

Introduction

It is widely recognized that people age at different rates, with the health status of the elderly varying from healthy to frail. Frailty is generally defined as a clinical geriatric syndrome characterized by a state of significant vulnerability resulting from the age-associated decline in multiple physiological systems [1, 2]. Clinical studies demonstrated that frailty is associated with adverse outcomes (such as an increased risk of hospitalization, disability, and ultimately death), prompting the need to identify the frail individuals correctly. The most widely used tools in humans are the “Fried’s frailty phenotype” and the “Rockwood’s clinical frailty index.” Fried’s frailty phenotype (FP) defines frailty as a clinical syndrome characterized by three or more altered criteria, including unintentional weight loss, self-reported exhaustion, weakness, slow walking speed, and low physical activity [3]. The presence of one or two of these criteria identifies subjects as prefrail, thus allowing subject categorization into three distinct subgroups of risks (i.e., frail, prefrail, and robust). On the other hand, Rockwood’s clinical frailty index (CFI) defines frailty as a state

of accumulating clinical deficits, and the estimation of frailty is performed by counting the actual number of health deficits divided by the total number of items measured [4]. This way, CFI is rated from 0 to 1 (where a higher FI indicates greater frailty).

However, little is still known about the mechanisms underlying the development of frailty and potential interventions for its treatment or prevention. In this scenario, murine models represent a valuable option for understanding the biology of this complex phenomenon and for evaluating the effect of new therapies to counteract, slow down, or prevent the frailty phenotype. For this reason, FP and CFI have been recently reverse-translated from humans into assessment tools for mice [5].

In 2014, Liu and colleagues [6] translated Fried’s method to mice, by emulating the measurement of four of the five criteria. This original mouse FP was recently modified with the inclusion of body weight and the improvement of endurance analysis with a more established test [7, 8]. As in humans, it is used to categorize mice into robust (none of the criteria), pre-frail (one or two criteria), and frail (three or more criteria).

Similarly, Parks and colleagues [9] published a tool for the preclinical assessment of a continuum CFI in aging mice. However, this method was time-consuming and required specific equipment as well as invasive procedures, thus making it unsuitable for longitudinal studies. Building on this, Whitehead et al. [10] continued this work by developing a non-invasive CFI based on a checklist of 31 easily observed deficits related to biological systems (integument, musculoskeletal system, auditory system, urogenital system, ocular/nasal system, respiratory system) and signs of discomfort. Although CFI has sometimes been converted into a categorical variable by identifying cutoff points, its main advantage over FP lies in its continuous nature. Indeed, a continuous variable is subject to less risk of misclassification as it does not need to define risk thresholds (cutoff points). Moreover, when the two assessment instruments were compared in the same population, they did not necessarily identify the same subject as frail [11–13]. However, both FP and CFI predicted adverse outcomes and death [14, 15].

Here, we present, in a cohort of 546 naturally aged C57BL/6 mice, an alternative scoring method, which we called physical function score (PFS), proposed as a continuous variable that resumes into a unique function, the five criteria identified by Fried. This score

A. Puca
Cardiovascular Research Unit, IRCCS MultiMedica,
20138 Milan, Italy

F. Boschi
Department of Computer Science, University of Verona,
Strada Le Grazie 15, 37134 Verona, Italy

C. Gaetano · A. Mongelli
Laboratory of Epigenetics, Istituti Clinici Scientifici
Maugeri IRCCS, Via Maugeri 10, 27100 Pavia, Italy

would reduce misclassification of the frail subjects and make the two tools (PFS and CFI) integrable with each other. In line with this, we tested an overall score, the vitality score (VS), that would consider both the clinical and physical parameters, allowing unambiguous health assessment of mice and a more precise identification of the frail population. We also computed the ethical benefits, in terms of reduction of sample size, that can be obtained from introducing these scores in preclinical intervention studies aimed to improve health span. Furthermore, we assessed the correlation between the newly created frailty assessment scores and two biological aging indices: epigenetic age and senescent cell accumulation.

Materials and methods

Animals and experimental design

All experiments were performed according to the European Community Council Directives of 2010/63/UE. The protocol was approved according to current Italian law (D.Lgs. n. 26/2014) by the Organismo Preposto al Benessere Animale (OPBA, animal care and health committee) of IRCCS INRCA and by the General Direction of Animal Health and Veterinary Drugs of the Italian Ministry of Health with the authorization no. 392/2019-PR ($n=446$ mice explicitly recruited for this study, of which $n=53$ carrying the p16-3MR transgene in the same background), authorization no. 130/2018-PR ($n=50$, control mice used in a longitudinal study aimed to assess the effects of gene therapy with the Longevity-Associated Variant-LAV of BPIFB4), and approval no. 1074/2021-PR ($n=50$, control mice used in a longitudinal study aimed to assess the effects of natural bioactive compounds with a nutritional supplement).

Based on the date of enrolment in the study, we subdivided the population into two cohorts of $n=152$ (2018–2019 cohort) and $n=394$ mice (2020–2021 cohort).

Mice functional phenotype

Measurement of clinical frailty index and clinical health score

We measured both the clinical frailty index (CFI) and physical frailty in mice as previously described [10, 16–18]. All frailty measurements were performed

within the SPF animal facility of INRCA in a dedicated area. The CFI score for each mouse was calculated using the previously published checklist and method [10]. The clinical health score (CHS) was computed as $1 - \text{CFI}$. This value ranges from 1 (optimal health) to 0. However, there are very few cases below 0.4, and those are usually mice that require immediate humane suppression.

Measurement of physical function score and frailty phenotype

The measurement of the PFS and FP in mice was performed following the same procedure described to translate the FP screening performed in humans [3] to mice [6, 8, 15]. To ensure testing reliability, we performed multiple measurements for each of the five criteria of the frailty assessment (shrinking, weakness, exhaustion, slowness, and sedentarily). All measurements performed to define FP and PFS are schematically described in Supplementary Table S1.

- 1) Body Size Score: This criterion was assessed through a composite score reflecting the body condition of the mice that included current body weight and body length. The weight was corrected to the previous measurement in the case we detected an increase in weight due to a tumor or distended abdomen. The body length (nose to base of tail) was measured during the locomotor activity test when the mouse was walking straight. The locomotor activity area was calibrated with a ruler, and the mean of 3 length measurements was performed using the video analysis software Tracker (v. 5.1.5, <https://physlits.org/tracker/>).
- 2) Strength Score: This criterion was assessed through a composite score reflecting the forelimb grip strength of the mice, including the measurements from 4 different tests.
 - a) Grip strength meter test (Ugo Basile, Varese, Italy) with a plastic grid [16, 19].
 - b) Grip strength meter test (Ugo Basile, Varese, Italy) with an iron bar.
 - c) Home cage lift test. The mouse gently held by the base of the tail to the top of an empty cage put above a scale with rapid response

and the mean of the two most negative peaks from about ten attempts is collected.

- d) Gripping weights lift test [20] with the modification described elsewhere [16, 21, 22]. To maximally motivate the mice, we used two different sets of weights. A first set constituted (as previously reported) of balls of tangled fine gauge stainless steel wire used to prevent limestone scale formation in domestic kettles, and a second set constituted of balls of tangled gauge stainless steel wire of the same diameter as the cage of the mice (2 mm). We observed that once the mice terminated the trial with the first set of weights, they were usually able to lift one or more heaviest weights of the second set. So, we recorded both scores and used the mean as the final score.
- 3) Endurance Score: This criterion was assessed through a composite score reflecting the endurance capacity of the mice that included treadmill distance (program: starting at 5 rpm for 2 min and increasing speed from 5 to 50 m/s in 2700 s), mean time to fall at rotarod test (program: starting at 5 m/s for 2 min and increasing speed from 5 to 40 rpm in 300 s), and the score of the gripping weights lift test normalized to body weight. As previously observed [16], this last measurement was indeed correlated with the other endurance measurements (as the test includes an endurance component due to the repeated lifting of weights) [20]. When a mouse was unable to perform a test due to severe illness and required humane suppression, it was assigned the minimum value for the test.
 - 4) Speed Score: This criterion was assessed through a composite score reflecting four different measurements related to the speed of the mice during their normal locomotion.
 - a) We analyzed the distribution of the time spent by the mouse in different speed intervals in an open field test (whole test duration 5 min). The speed intervals considered were as follows: I1 (0–1 cm/s), I2 (1–5 cm/s), I3 (5–10 cm/s), I4 (10–15 cm/s), I5 (15–20 cm/s), I6 (20–25 cm/s), I7 (25–30 cm/s), I8 (30–35 cm/s), I9 (35–40 cm/s), I10 (40–90 cm/s). We recorded the highest speed interval that the mouse ran for at least 3 s and assigned the mean speed of the interval as the value of the test (e.g., 12.5 for I4 and 37.5 cm/s for I9) as previously described [21, 22]. Locomotor activity was conducted by a 5-min open field test on a white wood chamber (72×72×30 cm) surmounted by a Logitech Brio Ultra HD Webcam 4 K 1080 P 60FPS (Logitech Lausanne Switzerland). Tracking and analysis were performed with Biobserve Viewer3 (Biobserve GmbH, Germany) [21, 22].
 - b) An additional measurement for speed was obtained by recording the maximum speed of the rotarod test.
 - c) As previously reported, there is a strong rationale in support of the relationship between walking speed and stride length, especially in older individuals [23]. Therefore, we also included in this score the mean of the measurements of the stride length of the mice obtained from the footprint test [24] and video tracking. This last measurement was obtained by calculating the distance between two consecutive hindlimbs paws during a straight walk in the open field arena (manual tracking of the paws was performed with Tracker v. 5.1.5, <https://physl.ets.org/tracker/>).
 - 5) Activity Score: This criterion was assessed by recording two measurements related to the dynamic behavior of the mice during a locomotor activity test:
 - a) total track length (total distance in cm) and
 - b) the time the mice were not resting (speed above 1 cm/s) during the locomotor activity test. Both measurements were recorded automatically by Biobserve Viewer3 (Biobserve GmbH, Germany).

Development of the composite functional scores and detection of the frailty phenotype

The results from the multiple measurements related to the same criterion were combined in a unique score following, separately for each sex, this procedure: (1)

the variables were normalized with the MIN–MAX method: $Z = (X_i - \text{Min}) / (\text{MAX} - \text{MIN})$ (where X_i is the measurement, Z the normalized data, MIN and MAX are the minimum and maximum values for X recorded in the population of mice); (2) the variables assigned to the same criterion were averaged to create a composite score for each criterium. This approach provided five quantitative composite scores: body size, strength, endurance, speed, and activity score. An overall score representative of physical decline, named PFS, was computed as the mean of the composite scores of the five criteria.

To avoid bias due to potential outliers, we used as MAX value the 95th percentile for all measurements, excluding weight. In this last case, we used as MAX the reference weights for mice aged 10–20 months in our colony (28.4 and 33.8 for females and males, respectively). Indeed, the weight of C57BL/6/J mice increases until 10 months, remains relatively stable from 10 to 20 months, and only later starts to decline. However, pathological conditions, such as distended abdomen and tumors, may induce a paradoxical increase in weight in some old and very old mice. In the case we detected an increase in weight associated with the above-mentioned pathological conditions, we used the weight value recorded in the previous observation (before the pathological problem appeared). Hence, each measurement ranges from 1 (or slightly above) to 0. Following the percentiles used by Fried et al. in humans [3] and by others in mice [6, 8, 15–17], mice that fell in the bottom 20% of our old cohort for the composite score computed for each criterion were considered positive for frailty for that given criterion. Mice with three or more positive frailty criteria were identified as frail.

Vitality score

The VS was obtained by calculating the arithmetic mean of the individual values of CHS and PFS. This process was used for each mouse studied at any given month of age. This newly created health assessment score ranges from 1 (optimal health) to 0.

Bioluminescence assay

Bioluminescence imaging (BLI) of experimental animals was performed with an IVIS Spectrum

imaging system (PerkinElmer). Briefly, mice were intraperitoneally injected with 100 μl of Xenolight RediJect Coelenterazine h (PerkinElmer #760,506) and were immediately anesthetized in an oxygen-rich induction chamber with 2% isoflurane. Images were captured after 20 min of incubation to allow substrate distribution. Bioluminescent images were obtained with mice in the dorsal position. Anesthesia was maintained during the entire imaging process by using a nose cone isoflurane-oxygen delivery device in the light-tight chamber. Bioluminescence, acquired by the CCD camera, was quantified by the Living Image® software drawing a region of interest (ROI) around the interested area and measuring the radiance (photons/seconds/cm²/steradian) emitted by the surface within the ROI (Fig. S1a).

Epigenetic age analysis

DNA methylation levels were analyzed at three age-associated CG dinucleotides (CpGs) as described previously [25]. Briefly, blood samples of C57BL/6 J mice of the chronologically aged C57BL/6 J mice ($n=44$) were obtained by ocular bleeding ($\sim 200 \mu\text{l}$) of living mice from the ocular vena and stored at $-80 \text{ }^\circ\text{C}$ until epigenetic analysis.

Genomic DNA from 200 μl murine whole blood was isolated by the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA was quantified by Nanodrop Spectrophotometers (ZYMO Research). Five hundred nanograms of extracted genomic DNA was subsequently bisulfite-converted with the epiTect Fast DNA Bisulfite kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Converted DNA was quantified using Nanodrop spectrophotometer (ZYMO RESEARCH). Thirty nanograms of bisulfite-converted DNA was initially subjected to PCR amplification. Target sequences for *Prima1*, *Hsf4*, and *Kcns1* were amplified by PyroMark PCR kit (Qiagen) using forward and reverse primers containing handle sequences for the subsequent barcoding step. PCR was run under the following conditions: 95 $^\circ\text{C}$ for 15 min; 45 cycles of 95 $^\circ\text{C}$ for 30 s, 56 $^\circ\text{C}$ for 30 s, 72 $^\circ\text{C}$ for 30 s; final elongation at 72 $^\circ\text{C}$ for 10 min. Primers were purchased at Metabion and the sequences are provided in Supplementary Table S2.

Five microliters of PCR products was subsequently immobilized to 2 μ l Streptavidin Sepharose High Performance Bead (Cytiva) and finally annealed to 20 μ l sequencing primer (0.375 μ M) for 5 min at 80 °C. Amplicons were sequenced using PyroMark Q24 Advanced Reagents (Qiagen) on PyroMark Q24 System (Qiagen, Hilden, Germany) and the methylation levels calculated by PyroMark Q CpG software (Qiagen).

Predicted age was calculated applying the coefficients obtained from a regression model developed in test samples from the same colony of mice (Fig. S1b).

Statistical analysis

Generalized linear mixed model analysis (SPSS 26.0) was used to consider the longitudinal design of the study in mice. The identifier of each mouse, gender, and age were included in the model. The linear models were developed assuming linear distribution with the identity link function. The Satterthwaite approximation with a robust estimator was used for unbalanced data and violation of the assumptions. The correlation of frailty parameters with biomarkers was studied by Spearman and Linear Mixed models correlation. Association with mortality was investigated by Linear Mixed models. Differential survival patterns were estimated by Cox-regression, taking into account possible confounder variables (age at inclusion and gender).

Sample size estimation for future studies with a longitudinal design was performed with the online power computation for linear models GLIMMPSE [26]. Sample size estimation for studies with a cross-sectional design was estimated with GPower 3.1 (Kiel University, Germany).

Results

The characteristics of the study population

Two cohorts of C57BL/6 J mice (Table 1) housed in the Geriatric Mouse Clinic of INRCA under SPF condition from 2018 to 2021 were used in this study. Five-hundred forty-six mice (46% females), mean age at the inclusion of 23.02 ± 1.86 months (males: 23.04 ± 1.85 months; females: 22.99 ± 1.88 months), were included in the study. All mice underwent monthly phenotyping and most of them (483 mice; 252 males and 231 females) were followed longitudinally until natural death or euthanized for aging-associated diseases. The remaining mice (censored) were followed longitudinally until used for organ explants that served as controls in other histochemical or molecular studies. Precisely, the follow-up number (functional and frailty assessment, one each month) ranged from 1 to 19 and the mean age of the animals at death was 29.11 ± 3.84 . Figure S2 shows the survival curve of the study population. The overall number of recorded phenotypes at each month of age is listed in Supplementary Table S3.

Table 1 Description of study population

	Total	2018–2019	2020–2021
<i>n</i>	546	152	394
Male	293	102	191
Female	253	50	203
<i>n</i> recorded deaths (<i>n</i> censored)	483 (63)	141 (11)	342 (52)
Male	252 (41)	91 (11)	161 (30)
Female	231 (22)	50 (0)	181 (22)
Age at enrollment (Mean \pm SD)	23.02 ± 1.86	22.01 ± 2.51	23.41 ± 1.36
Male	23.04 ± 1.85	22.35 ± 2.36	23.41 ± 1.38
Female	22.99 ± 1.88	21.32 ± 2.68	23.40 ± 1.35
Age at death (Mean \pm SD)	29.11 ± 3.84	30.11 ± 3.90	28.69 ± 3.75
Male	29.03 ± 3.93	30.09 ± 4.05	28.43 ± 3.75
Female	29.17 ± 3.74	30.16 ± 3.63	28.92 ± 3.74
<i>n</i> follow-up (Mean \pm SD)	7.43 ± 3.85	8.94 ± 4.62	6.85 ± 3.34
Male	7.17 ± 3.98	8.49 ± 4.67	6.46 ± 3.36
Female	7.73 ± 3.69	9.86 ± 4.43	7.21 ± 3.29

Frailty phenotype

The longitudinal design of this study allowed us to assess the age-related variations in the health status of the same mice throughout the whole lifespan. The five criteria were analyzed to identify mice with a frail phenotype according to Fried's frailty phenotype (FP). The prevalence of physical frailty was estimated at each month of age. As shown in Fig. S3a, the percentage of mice classified as frail after 20 months increased exponentially with age, with about 55% of the population being frail at 35 months. Moreover, using this approach, it is possible to predict mortality within 1, 2, 3, and 6 months (Supplementary Table S4). As a side note, categorizing mice as pre-frail also predicts 6-month mortality (Supplementary Table S4).

To improve the frailty classification system in mice, we converted the classic categorical FP into a quantitative measurement of each physical criterion (Physical Scores). In our modified method, each estimated parameter was assigned a value ranging from 1 (highest functionality) to 0 (lowest functionality).

All five "Physical Scores" showed a significant age-related decline (Fig. 1a–e). Expressly, the Body Size Score (Fig. 1a) declined linearly with advancing age, with similar trends in both males and females. Similarly, the Strength Score (Fig. 1b), which represents the average of all measurements taken to detect grip strength, exhibited an age-related decline slightly more marked in females than in males. On the contrary, the Endurance Score (Fig. 1c) showed two different trends: it seems fairly preserved until the 25th month in both sexes and then accelerates the decay, particularly in female mice. Similar to Body Size and Strength Scores, the Speed Score (Fig. 1d) linearly declines with age in both males and females. The Activity Score (Fig. 1e) displays the highest variability among the scores measured. Notwithstanding, the activity level declines with advancing age in females and, more markedly, in male mice.

We also computed an overall quantitative index of physical functionality that we named physical function score (PFS, Fig. 2a,b and Fig. S4a,b). Interestingly, this variable was not influenced by sex in its temporal decline; at the same time, it seems to follow two different slopes: a very gradual decay up to 26 months followed by a faster acceleration (Fig. 2a and Fig. S4a). Moreover, PFS is significantly

associated with mortality: graphs in Fig. 2b and Fig. S4b show that PFS rapidly declines when the death outcome approaches. This association was independent of sex and remained significant even when chronological age was included in the analysis.

To test which frailty phenotype classification method best predicts mortality, we first sought to identify an optimal PFS cutoff value to define the onset of frailty based on the FP classification. Specifically, through the decision tree model and the ROC analysis, we identified a PFS cutoff point of 0.55. ROC curve analysis indicated that the AUC was 0.967 (95% CI, 0.960–0.974) and that a PFS cutoff point of 0.55 showed a sensitivity of 81.8% and specificity of 94.10% for the prediction of frailty (data not shown). The analysis reported in Fig. S5 describes the results obtained with the decision tree model: the prevalence of frailty is 86.6% in animals with a quantitative value of PFS below 0.51, and the prevalence of frailty is still about 50% for a PFS value between 0.51 and 0.55. Overall, this PFS cutoff identifies 72.7% of mice classified as frail by FP. The bar graph depicted in Fig. S3b showed the exponential increase with aging in the percentage of mice classified as frail.

Interestingly, when the original and the modified frailty classifications methods were compared in their ability to predict death, we found that both classification methods powerfully predict mortality within 1, 2, 3, and 6 months (Supplementary Table S5); however, when a restricted population of younger old mice (24–25 months of age) was selected, PFS cutoff point of 0.55 was more effective in predict mortality than FP and age (Table 2). Moreover, when a 0.63 PFS cutoff point (Fig. S5) was used to identify pre-frail status in younger mice, PFS was more effective than FP in predicting 6-month death (Table 2 and Supplementary Table S5).

Clinical frailty index analysis

Alongside the frailty phenotype, Rockwood's clinical frailty index (CFI) measures the proportion of accumulated deficits in an individual among a list of all potential ones, i.e., symptoms, signs, disabilities, diseases, and laboratory test results. As expected, the CFI in our study population increases with aging (Fig. S6a) following a linear trend (Fig. S6b), especially in female individuals. Moreover, the CFI was closely associated with mortality (Fig. S6c,d).

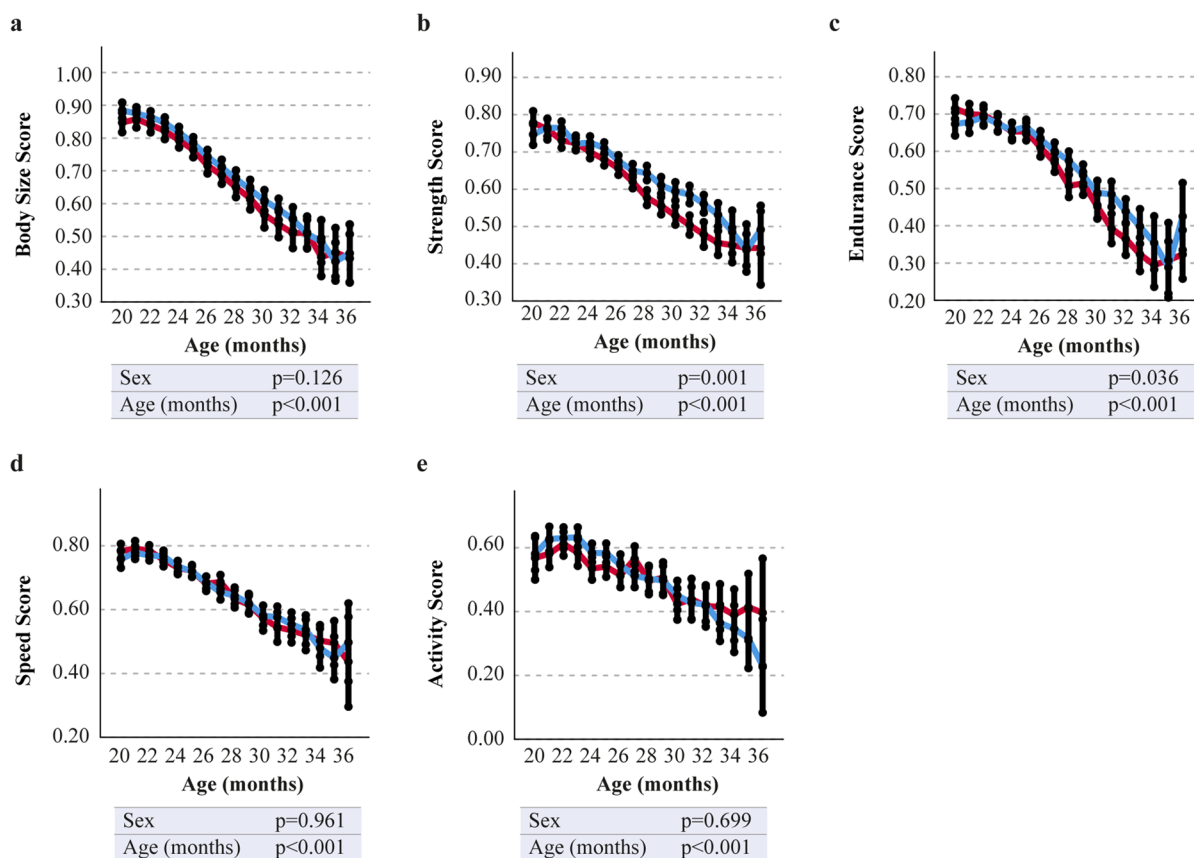


Fig. 1 Quantitative estimation of the five Physical Scores. C57BL/6 J mice ($n=546$) were monitored each month from the inclusion (20 months) up to death. Graphs show the trend of Body Size Score (a), Strength Score (b), Endurance Score (c), Speed Score (d), and Activity Score (e) with advancing age, both in male (blue line) and female (red line) mice. Time

is expressed in months. Values are reported as the mean estimates obtained by linear mixed model analysis for longitudinal data using sex, cohort, and age (months) as fixed factors. Test of fixed effects parameters (sex and age) are reported inside the figure

Based on previous studies, mice were categorized as frail using a CFI's cutoff of 0.25 [27]. Indeed, using this threshold we found that, at 21 months, only a small percentage of mice were classified as frail (Fig. S3c). However, this percentage increased exponentially with aging, and about 65% of the population was found to be frail at 35 months.

Vitality score

We found that PFS and CFI do not identify the same mice as frail (Fig. S7). The quantitative parameter, PFS, provides an additional advantage in the analysis of health in a longitudinal study of mice: creating a unique index that summarizes the frailty phenotype and frailty index. To this aim, the

CFI was converted to its complementary (1-CFI) renamed clinical health score (CHS, Fig. 2c,d and Fig. S4c,d), and used to compute the vitality score (VS, Fig. 2e,f and Fig. S4e,f) as reported in the methods section. VS displays a gradual decline with age up to 28 months, followed by a faster acceleration (Fig. 2e and Fig. S4e). As already observed for PFS and CFI, VS also maintains its association with mortality, declining rapidly when the death outcome approaches (Fig. 2f and Fig. S4f). This association is independent of sex and remained significant even when chronological age was included in the analysis.

ROC curve analysis reported in Fig. S4g–i showed that VS was consistently more effective than the other quantitative scores (PFS and CFI) in predicting

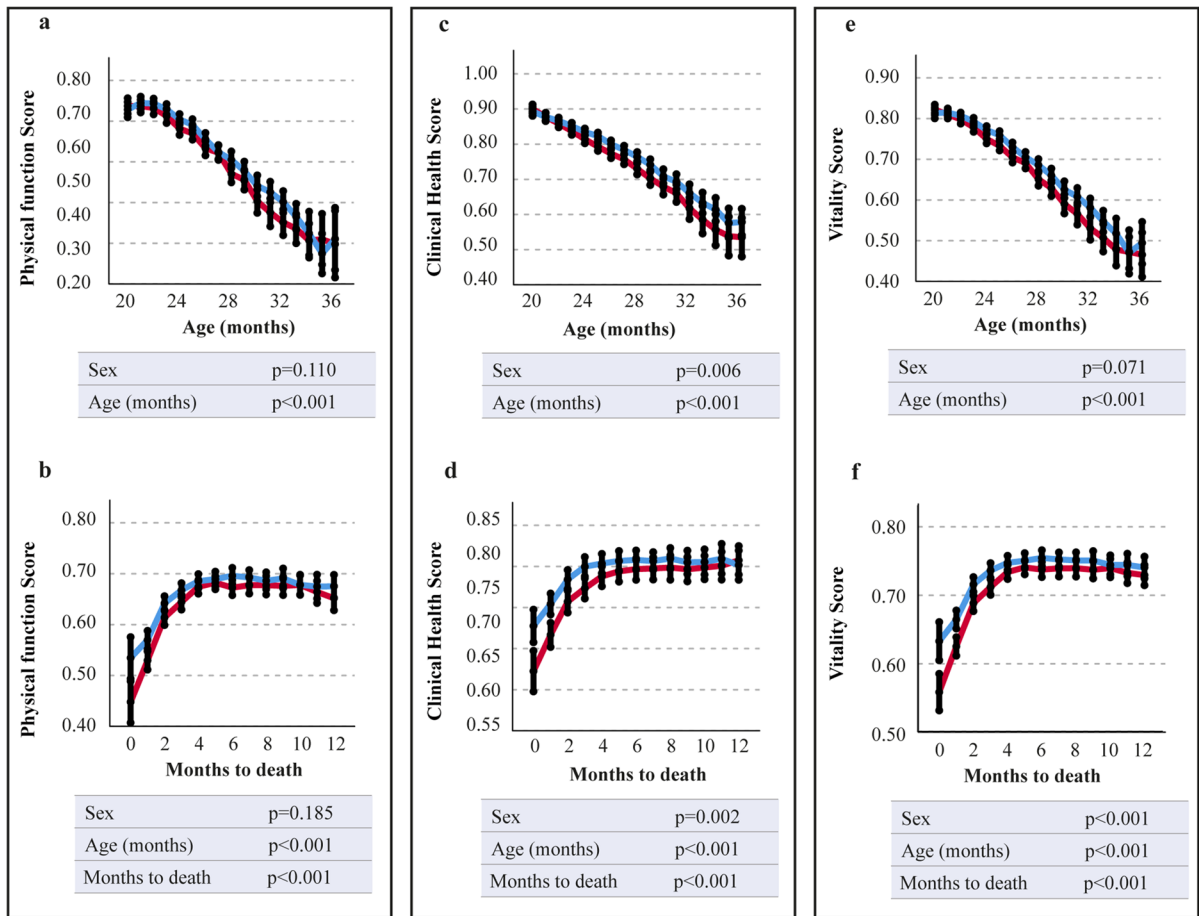


Fig. 2 Association of the physical function score (PFS), clinical health score (CHS), and vitality score (VS) with aging and mortality. C57BL/6 J ($n=546$) mice were monitored each month from the inclusion (20 months) up to death. (a–f) PFS, CHS, and VS decline with advancing age. Data of PFS (a, b), CHS (c, d), and VS (e, f) are reported as a function of age (a,

c, e) or months to death (b, d, f). Values are reported as the mean estimates obtained by the linear mixed model analysis for longitudinal data. Tests of fixed effects parameters (sex, age, or months to death) are reported inside the figure. Data from male (blue line) and female (red line) mice are presented

mortality at 1, 2, and 3 months, highlighting the validity of using this parameter.

Intending to identify a cutoff point for frailty classification, we related the animals identified as frail either by FP or by the CFI (Combined Frailty Parameter, CFP) to their quantitative value of VS (Fig. S8a). The analysis in Fig. S8b described the results obtained with the decision tree model: the prevalence of frailty is 67.0% in animals with a quantitative value of VS below 0.68, and the prevalence of frailty is still about 40% for a VS value between 0.68 and 0.71. Overall, this PFS cutoff identifies 88.45% of mice classified as frail by the CFP. By ROC curve analysis (AUC=0.963; 95% CI=0.957–0.970), a PFS value

of 0.71 shows a sensitivity of 89.70% and specificity of 89.60% for the prediction of frailty (data not shown).

The bar graph depicted in Fig. S3d and Fig. S3e shows the exponential increase with aging in the percentage of mice classified as frail by CFP and VS.

Analysis of mortality prediction (Supplementary Table S6) shows that both CFP and VS classifications methods powerfully predict 1-, 2-, 3-, and 6-month mortality. However, when a time-restricted subpopulation of mice (24–25 months of age, Table 3) was selected, CFP lost mortality prediction effectiveness; in contrast, VS cutoff point of 0.71 not only predicted

Table 2 The association between prefrail and frail status, defined according to Fried's method (FP) or physical function score (PFS), and mortality in 24- to 25-month-old mice

			Coefficient*	Sig	95% Confidence Interval		Exp (coefficient)	95% confidence interval for exp(coefficient)		
					Lower	Upper		Lower	Upper	
24–25 months	Frail	1 month	<i>Age (months)</i>	−0.062	0.717	−0.398	0.274	0.940	0.671	1.315
			<i>FP</i>	1.032	0.049	0.004	2.061	2.808	1.004	7.852
			<i>PFS</i>	2.036	<i>p</i> < 0.001	1.085	2.987	7.658	2.958	19.821
		2 months	<i>Age (months)</i>	−0.134	0.318	−0.396	0.129	0.875	0.673	1.138
			<i>FP</i>	1.224	0.004	0.384	2.064	3.400	1.468	7.875
			<i>PFS</i>	1.339	<i>p</i> = 0.001	0.526	2.152	3.816	1.692	8.604
		3 months	<i>Age (months)</i>	0.245	0.037	0.015	0.476	1.278	1.015	1.609
			<i>FP</i>	1.107	0.009	0.283	1.931	3.026	1.328	6.896
			<i>PFS</i>	0.948	0.019	0.158	1.739	2.582	1.171	5.691
		6 months	<i>Age (months)</i>	0.012	0.913	−0.199	0.223	1.012	0.819	1.249
			<i>FP</i>	1.901	0.009	0.476	3.326	6.693	1.609	27.836
			<i>PFS</i>	1.587	0.006	0.466	2.708	4.887	1.593	14.996
	Pre-frail	6 months	<i>Age (months)</i>	−0.033	0.768	−0.252	0.186	0.968	0.777	1.205
			<i>FP</i>	0.357	0.081	−0.044	0.758	1.429	0.957	2.134
			<i>PFS</i>	1.308	<i>p</i> < 0.001	0.756	1.860	3.698	2.129	6.421

*Coefficient table of fixed effects parameters obtained by fitting the linear mixed model. Significant coefficients are highlighted in bold

1-, 2-, 3-, and 6-month mortality but was also more effective than age.

Sample size estimation

The sample size of preclinical studies in geriatric mice is one of the critical limit and ethical issues as regards the “Reduction” from the 3Rs principles (Replacement, Reduction, and Refinement). In order to understand if the introduction of the new scores can provide ethical benefits, we used the data (means, variance and co-correlations) originating from the 2020–2021 cohort to compute the sample size required to estimate the efficacy of an intervention expected to improve 10%, 20%, or 30% of the three health measurements described in this study (CFI, PFS, and VS) compared to a control group. We used GLMMSI [26] to estimate the sample size in the case of a prospective longitudinal study (control vs treatment) with a 3- or 6-month follow-up starting with mice aged 24 months, and GPower for a cross-sectional design aimed at detecting differences between the treated and control groups at 27 or 30 months. Interestingly, we found that the sample

size is relatively large when using only CFI as the outcome, but it is reduced using PFS and even more for VS (Table 4). As expected, a longitudinal design can further help reduce the study's sample size.

Biomarkers

The association between PFS and VS with some biomarkers of aging such as senescent cell accumulation (analyzed by total body luciferase activity in p16-3MR mice) and epigenetic age was investigated.

The accumulation of senescent cells is considered the primary cause of chronic inflammation underpinning aging and age-related diseases. In fact, the cells expressing p16(Ink4a), the principal gene regulating cellular senescence, gradually increases during the mice's lifetime. To evaluate the correlation between our frailty scores and the total accumulation of senescent cells in the body, the dorsal luminescence of p16-3MR mice was analyzed. Indeed, in this murine model, p16(Ink4a) drives the expression of the renilla luciferase, allowing us to quickly identify the presence of senescent cells through in vivo BLI. Data reported

Table 3 The association between frail status, defined according to the Combined Frailty Parameter (CFP) or vitality score (VS), and mortality in 24- to 25-month-old mice

			Coefficient*	Sig	95% Confidence Interval		Exp (coefficient)	95% confidence interval for exp(coefficient)	
					Lower	Upper		Lower	Upper
24–25 months	1 month	<i>Age (months)</i>	0.034	0.837	−0.292	0.360	1.035	0.747	1.434
		<i>CFP</i>	0.576	0.131	−0.172	1.325	1.779	0.842	3.761
		<i>VS</i>	1.637	<i>p</i> < 0.000	0.932	2.341	5.138	2.541	10.388
	2 months	<i>Age (months)</i>	−0.067	0.620	−0.331	0.197	0.935	0.719	1.218
		<i>CFP</i>	0.549	0.103	−0.112	1.211	1.732	0.894	3.357
		<i>VS</i>	1.191	<i>p</i> < 0.000	0.604	1.779	3.292	1.829	5.924
	3 months	<i>Age (months)</i>	0.282	0.017	0.050	0.515	1.326	1.051	1.673
		<i>CFP</i>	0.441	0.170	−0.190	1.072	1.555	0.827	2.923
		<i>VS</i>	0.846	0.005	0.262	1.430	2.331	1.300	4.180
	6 months	<i>Age (months)</i>	0.071	0.513	−0.142	0.285	1.074	0.867	1.330
		<i>CFP</i>	0.416	0.228	−0.261	1.093	1.516	0.770	2.982
		<i>VS</i>	1.451	<i>p</i> < 0.000	0.795	2.107	4.267	2.213	8.226

*Coefficient table of fixed effects parameters obtained by fitting the linear mixed model. Significant coefficients are highlighted in bold

Table 4 Estimated total sample size for two groups of mice (control vs treatment) starting at age 24 months on the basis of different outcomes and study designs (longitudinal or cross sectional) with 3- or 6-month follow-up

Outcome	Expected improvement*	27 months (3-month follow-up)		30 months (6-month follow-up)	
		Longitudinal	Cross sectional	longitudinal	Cross sectional
		<i>N</i> (<i>N</i> corr.) ^a	<i>N</i> (<i>N</i> corr.) ^a	<i>N</i> (<i>N</i> corr.) ^a	<i>N</i> (<i>N</i> corr.) ^a
<i>Clinical frailty index (CFI)</i>	10%	186 (238)	404 (518)	294 (554)	412 (778)
	20%	48 (62)	104 (134)	76 (144)	106 (200)
	30%	24 (30)	48 (62)	36 (68)	48 (90)
<i>Physical function score (PFS)</i>	10%	52 (66)	80 (102)	140 (264)	164 (310)
	20%	16 (20)	22 (28)	38 (72)	44 (84)
	30%	8 (10)	12 (15)	18 (33)	20 (38)
<i>Vitality score (VS)</i>	10%	16 (20)	40 (52)	46 (86)	68 (128)
	20%	8 (10)	12 (16)	14 (26)	20 (38)
	30%	6 (8)	8 (10)	8 (15)	10 (18)

*Expected improvement is considered with respect to the data of normal aging mice computed at age 27 or 30 months. *N* corr., sample size corrected for mortality attrition, estimating the loss from the respective survival curves

in Fig. 3a–c showed that dorsal bioluminescence increases as physical function and vitality score decrease and, consistently, as CFI increases.

The associations between epigenetic age and frailty scores (Fig. 3d–f) showed a statistically significant negative correlation between epigenetic age and mouse PFS or VS and a positive

correlation with CFI, confirming that epigenetic age decreases significantly with increasing frailty status. Interestingly, the highest R^2 was obtained from the correlation between the epigenetic age and VS, thus confirming that this parameter may better represent the health status in aging than any single measurement of frailty.

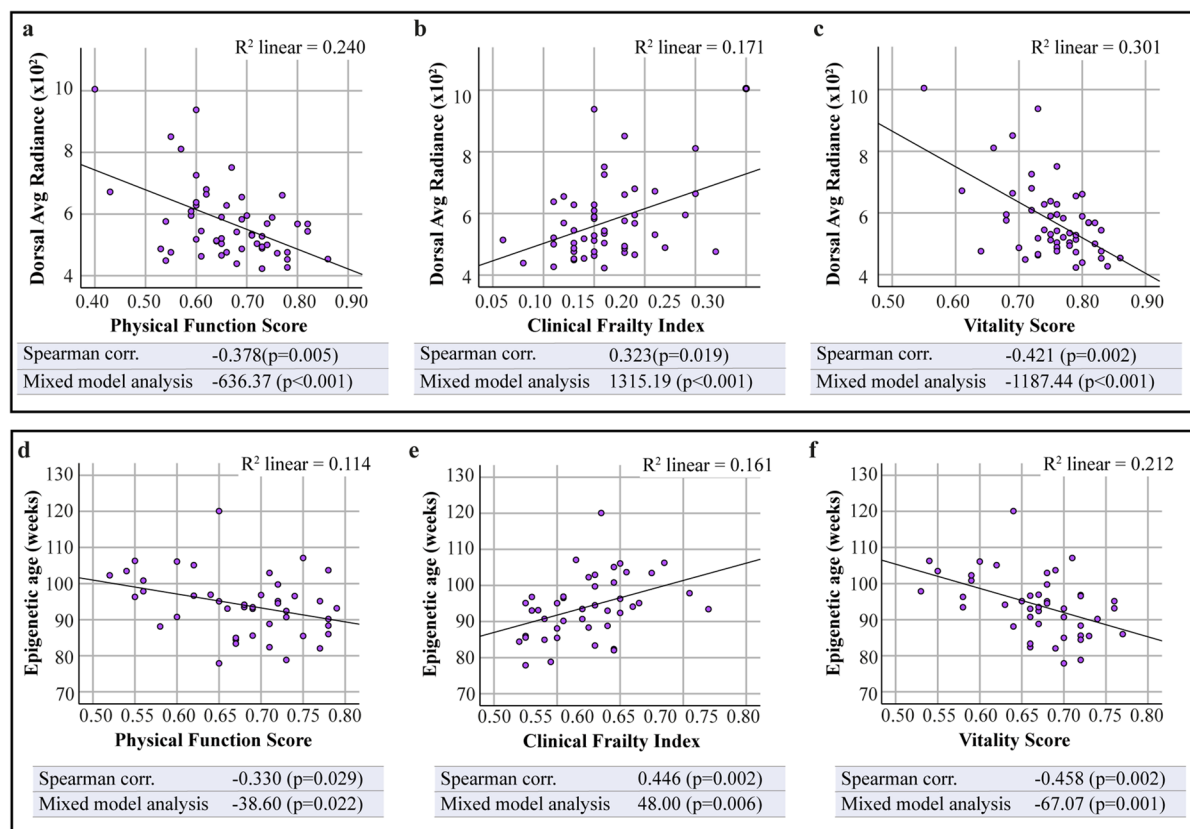


Fig. 3 Correlation between frailty scores with some biomarkers of aging. (**a–c**) Bioluminescent average radiance (photons/seconds/cm²/steradian) emitted by a cohort ($n=53$) of chronologically aged p16-3MR mice significantly correlates with PFS (**a**), CFI (**b**), and VS (**c**). The average radiance for individual mice are depicted. (**d–f**) Epigenetic age evaluated in a cohort

($n=44$) of chronologically aged C57BL/6 J mice significantly correlates with PFS (**g**), CFI (**h**), and VS (**i**). The R^2 coefficient is shown on each of the graphs. Correlation coefficient and p values analyzed by Spearman correlation and mixed model analysis are reported

Validation model

To assess the efficiency of our modified frailty scores to discriminate between different treatments/conditions, we subdivided our original cohort of mice into two groups with different age at death: short-lived (mice that died within 27 months of age) and long-lived (mice that died after 33 months of age). The mean age at death was 25.09 ± 1.54 and 34.91 ± 1.75 months for short-lived and long-lived, respectively (Supplementary Table S7). Cox regression survival curve confirmed the reduced lifespan in short-lived group compared to the long-lived (Fig. 4a). To test whether differences in lifespan correlated with a significantly reduced physical

functionality, PFS was analyzed longitudinally for each animal in the two life-classes. Interestingly, the function depicted in Fig. 4b, showed that the two groups under analysis differed significantly in PFS trend (short-lived vs long-lived: $p<0.001$): short-lived mice showed an accelerated decline in PFS as early as 23 months of age, unlike long-lived mice in which it was observed later in life, after 31 months. A deeper analysis also revealed that these significant differences between the two classes under analysis are related to a significant reduction in Body Size, Endurance, Speed, and Activity Scores, while the Strength Score is maintained between the two groups, consistently with data already reported in literature [28] (Fig. S9a–e).

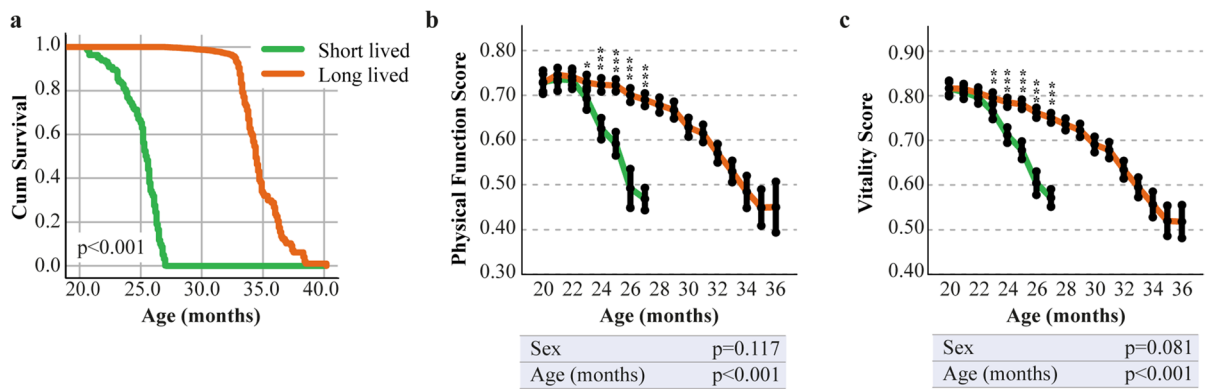


Fig. 4 PFS and VS validation. (a–c) Survival function of short-lived (green line, $n=164$) vs long-lived (orange line, $n=90$) mice (e). PFS (e) and VS (f) were monitored from 20 months up to death in short-lived and long-lived groups. Time is expressed in months. Values are reported as

mean \pm SD. Statistics to compare data between groups was performed by mixed model analysis for longitudinal data, using group, sex, and age (months) as fixed factors. Test of fixed effects parameters (sex and age) are reported inside the figure. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Furthermore, similarly to PFS, a significant difference is also observed for VS (short-lived vs long-lived: $p < 0.001$), with short-lived animals showing an earlier decline than long-lives (Fig. 4c).

Discussion

The employment of aging mice in the study of frailty provides an indispensable model to understand the biological mechanisms underlying this complex phenomenon and to develop new strategies to improve human healthspan and lifespan. Indeed, preclinical studies limited to the sole lifespan extension are inadequate to establish and validate interventions with the potential to improve human health throughout aging [29]. To this end, the translation of the screening tools currently used in humans, namely Rockwood's CFI and Fried's FP, into tools for studying naturally aging rodent models represents an important step forward [6–10].

The mouse CFI measures frailty as a continuum between 0 (very fit) and 1 (severely frail) and has already emerged as a promising new preclinical model for frailty assessment. In fact, from 2012 to date, it has been applied in a wide range of research areas such as the study of the impact of strain, dietary, pharmaceutical interventions, and pathologies on aging [16, 30–38].

Conversely, the translation of FP to the murine model was first introduced by Liu and colleagues in 2014 [6] and subsequently expanded and optimized by the same authors and others [7, 8, 39]. This murine FP adapts to mice the same five functional criteria used in Fried's human model (i.e., shrinkness, weakness, poor endurance/exhaustion, slowness, and low activity); the identification of a specific cutoff point for each criterion [6–8, 28, 39] allows to categorize an animal as frail if at least three criteria are below the established cutoff point, pre-frail if one or two frailty criteria are altered, and robust if no markers of frailty have been identified. This method has already been used in murine models to analyze the effects of exercise interventions, senolytics, and chronic medications on frailty [7, 38–40]. However, this categorical approach usually requires a very high sample size and it is subject to misclassification due to the need for specific threshold values. In addition, relatively few mice have been used in previous studies, and only one functional test is usually chosen to evaluate a single frailty criterion, thus reducing the assessment's reproducibility. Over time (longitudinally) assessment of the FP in mice, as well as sex-related frailty differences, have been poorly studied [28, 41].

The primary purpose of this study was to propose a new non-categorical method for assessing the physical phenotype of frailty in aging mice as an alternative to Fried's method. Specifically, the five functional criteria of frailty were analyzed longitudinally, every

month until death, in a cohort of naturally aging mice ($n=546$). Unlike current assessment tools, which categorize mice as fulfilling or not a specific frailty criterion (categorical method), we rescaled each functional criterion into an intuitive continuous value (Physical Score), ranging from 1 (highest functionality) to 0 (lowest functionality). Our results showed an age-related decline in the five Physical Scores, with significant sex-related differences in the Strength and Endurance Scores. A recent investigation in humans highlighted the relevance of studying different subtypes of prefrailty which, in turn, are suspected to belong to different biological syndromes [42]. As the present method quantifies each individual frailty criterion with an intuitive score, it can be perfectly suited to help the investigation in this field of research. However, the main advantage of this analysis lies in the possibility of combining the five Physical Scores into a unique quantitative index that we have called physical function score (PFS): the result is a continuous variable enabling the overall physical performance to be traced longitudinally, that declines significantly with age and that strongly associates with the overall mortality.

Notwithstanding the advantage of having a continuous variable, we additionally identified cutoff points for the PFS to create categories (frail, pre-frail, healthy) that could be compared with the FP results. Similar to evidence in the literature [8, 28, 41, 43], the FP predicts mortality effectively in our mouse cohort; however, our data highlighted that PFS, using a frailty cutoff point of 0.55, associated with mortality better than the classical PF method when tested in a subpopulation of time restricted mice (24–25 months). Furthermore, introducing a cutoff of 0.63 for categorizing the pre-frail phenotype makes it possible with our new score to predict 6-month mortality, whereas PF fails in this aim. These features make PFS particularly interesting in the study of frailty biological processes and in the identification of innovative treatments for frailty.

Another critical point in frail status evaluation is the absence of a unique, comprehensive assessment method. In fact, several studies have shown that the two dominant approaches for measuring frailty either in humans or mice (FP and CFI) identify different subjects as frail [11–13]. Using a very large cohort of mice, we confirmed the concept that some animals are categorized as frail only by CFI and others only by PF. Clinical and physical parameters should be considered to standardize the classification method and to uniquely

and more accurately identify frail subjects. To this end, we created a unique frailty assessment tool, the vitality score (VS), derived from the arithmetic mean of both parameters analyzed in this work, PFS and CFI. As for PFS, VS inversely correlates with advancing age, reflecting the age-related accumulation of deficits, and declines rapidly as the death outcome approaches. In addition, it is remarkable that this combined score predicts 1-, 2-, and 3-month mortality better than the two variables from which it is derived. One reason the VS may perform better than clinical or physical assessments could be attributed to the ability of VS to provide more information for the overall health of the mice, thus enabling better identification of frail individuals. To further confirm the advantage of using our newly created score, a frailty cutoff point of 0.71 was identified to create ordinal values that could be compared with the results obtained by combining subjects classified as frail with the FP assessment or CFI (Combined Frailty Parameter, CFP). We found that both CFP and VS powerfully predict 1-, 2-, 3-, and 6-month mortality. However, when a time-restricted subpopulation of mice (24–25 months) was considered, only VS was associated with mortality outcomes, whereas both CFP and age lost their mortality prediction effectiveness. These data suggested a more significant potential of our newly created function to identify frail individuals with a higher risk of adverse events.

Therefore, the two newly developed scores presented in this paper display a remarkable ability to predict death compared to the currently available tools. The high requirement of sample size may be a limit for preclinical studies using geriatric mice [44]. This is the natural consequence of the significant phenotypic variations observed in aged laboratory mice [45]. When CFI is used as the primary outcome, it should be taken into account that this parameter displays an additional source of variation due to sex differences and to the subjective scoring of the deficits. However, as reported in Table 4, the assessment of PFS and VS may greatly help reduce the study's sample size. To further validate the use of these assessment tools in preclinical studies, the two scores were (1) correlated with biomarkers of aging (i.e., senescent cells accumulation and epigenetic age) and (2) tested on their ability to discriminate between different experimental conditions.

Using the p16-3MR transgenic mouse model, which allow senescent cell detection in living animals [46], we observed that both PFS and CFI

significantly correlated with increasing dorsal bioluminescence. Interestingly, this correlation was much more potent when VS was considered, thus suggesting that this parameter could be a helpful tool in studies aimed at reducing senescent cell burden with senolytics [47]. Similar results were also obtained in the correlation between the three parameters and the epigenetic age. Also in this case, the best correlation is observed with VS. Importantly, a recent study reported that the CFI outperforms DNA methylation age in predicting mortality [48]. Hence, the finding that the newly developed VS is a better predictor of survival than the CFI suggests that VS may be considered a reliable measurement of the biological age in mice. Moreover, one of the most relevant advantages of the VS is that it can be obtained using non-invasive tools, thus matching scientific and ethical requirements in the best way. The validation model (long-lived *vs* short-lived mice) demonstrated that both PFS and VS frankly discriminate between different experimental conditions, confirming the high potential of using these two tools in studying new strategies to improve healthspan and lifespan.

The main limitation of our study relates to the development of the VS: in fact, it was created by considering the arithmetic mean between CFI and PFS. Albeit, its computation could be improved, for example, by assigning different weights to the individual components of the score, we deem that a simple procedure to develop the score can display practical advantages. Moreover, further studies are needed to determine whether one of the two areas of frailty, clinical and physical, has greater relevance and with what magnitude.

An important aspect that has not been thoroughly examined in our study and others is the connection between underlying pathology and frailty status in mice. A recent study on aging female mice found that none of the physical performance parameters was related to the presence of tumors or organ dysfunction, indicating that mice can maintain their physical function despite underlying diseases [49].

However, since mice primarily die from malignant diseases, it is possible to assume that their presence may affect the evaluation of the five criteria included in the physical performance assessment [29]. One area of concern is the estimation of weight loss, as tumors or distended abdomen can cause a

paradoxical increase in weight in some old and very old mice. For this reason, in case we detected an increase in weight associated to distended abdomen or tumor, we corrected the weight to the previous measurement to compute the “Body Size Score.”

Anyway, it still remains to be investigated the sensitivity of the other individual frailty assessment criteria to the consequences of malignancies.

Regardless of the limitation, this study dramatically demonstrates the potential to apply a non-categorical, non-invasive physical performance assessment tool in the study of aging, allowing longitudinal assessment of the five functional criteria of frailty and the creation of a unique physical function variable, the so-called PFS. This model also sets the groundwork for integrating physical parameters with clinical assessment data. Indeed, both frailty phenotype and clinical frailty index have validity in frailty research, but they identify different mice as frail. The vitality score, herein presented, leads to the creation of a comprehensive function that overcomes the different sensitivity in frailty identification of the in-use assessment tools, allowing for a more thorough and precise analysis of, for example, the prevalence of frailty or the effectiveness of therapeutic/preventive interventions. Our future work aims to develop additional non-invasive cognitive, respiratory, and cardiac indexes that may provide extensive mouse health phenotyping associated with the physical performance score and the clinical frailty index.

Acknowledgements The authors would like to acknowledge the Centro Piattaforme Tecnologiche (CPT) of Verona University for the IVIS Spectrum usage. The authors wish also to acknowledge Andrea Amoroso, Daniele Pierelli and Huilca Quispe Doctor Ambrosio from Charles River Laboratory Italia, as well as Beatrice Bartozzi and Gianni Bernardini for their precious technical support.

Author contribution Conceptualization: MM EN, AV. Data curation: MM, SM. Formal analysis: SM. Funding acquisition: AP, MM, AV, MP, EN, FL. Investigation: MM, GB, AS, MEG, RG. Methodology: MM, FB. Project administration: FL, MP. Supervision: AP, MP, FL. Visualization: SM. Writing—original draft: SM, MM. Writing—review and editing: MP, FP, RG, EN, AV.

Funding This work was supported by The Network IRCCS AGING “Rete Nazionale di Ricerca sull’invecchiamento e la longevità attiva – Implementazione della RoadMap nella ricerca sull’Aging (IRMA)” to

FL; by Fondazione Cariplo (grant 2016–1006) “Multicomponent Analysis of Physical Frailty Biomarkers” to EN, MP, and AV; and by Ricerca Corrente funding from Italian Ministry of Health to IRCCS-INRCA (MM and MP) and to IRCCS MultiMedica (AP).

Data availability The data that support the findings of this study are available from the corresponding author (MM), upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

- Dent E, et al. Management of frailty: opportunities, challenges, and future directions. *Lancet*. 2019;394(10206):1376–86. [https://doi.org/10.1016/S0140-6736\(19\)31785-4](https://doi.org/10.1016/S0140-6736(19)31785-4).
- Clegg A, et al. Frailty in elderly people. *Lancet*. 2013;381(9868):752–62. [https://doi.org/10.1016/S0140-6736\(12\)62167-9](https://doi.org/10.1016/S0140-6736(12)62167-9).
- Fried LP, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146–56. <https://doi.org/10.1093/gerona/56.3.m146>.
- Rockwood K, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ*. 2005;173(5):489–95. <https://doi.org/10.1503/cmaj.050051>.
- Howlett SE, Rutenberg AD, Rockwood K. The degree of frailty as a translational measure of health in aging. *Nature Aging*. 2021;1(8):651–65. <https://doi.org/10.1038/s43587-021-00099-3>.
- Liu H, et al. Clinically relevant frailty index for mice. *J Gerontol A Biol Sci Med Sci*. 2014;69(12):1485–91. <https://doi.org/10.1093/gerona/glt188>.
- Gomez-Cabrera MC, et al. A new frailty score for experimental animals based on the clinical phenotype: inactivity as a model of frailty. *J Gerontol A Biol Sci Med Sci*. 2017;72(7):885–91. <https://doi.org/10.1093/gerona/glw337>.
- Baumann CW, Kwak D, Thompson LV. Assessing onset, prevalence and survival in mice using a frailty phenotype. *Aging*. 2018;10(12):4042–53. <https://doi.org/10.18632/aging.101692>.
- Parks RJ, 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(3):217–27. <https://doi.org/10.1093/gerona/glr193>.
- Whitehead JC, 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(6):621–32. <https://doi.org/10.1093/gerona/glt136>.
- Malmstrom TK, Miller DK, Morley JE. A comparison of four frailty models. *J Am Geriatr Soc*. 2014;62(4):721–6. <https://doi.org/10.1111/jgs.12735>.
- Hogan DB, et al. Comparing frailty measures in their ability to predict adverse outcome among older residents of assisted living. *BMC Geriatr*. 2012;12:56. <https://doi.org/10.1186/1471-2318-12-56>.
- Todorovic S, et al. Frailty index and phenotype frailty score: sex- and age-related differences in 5XFAD transgenic mouse model of Alzheimer’s disease. *Mech Ageing Dev*. 2020;185:111195. <https://doi.org/10.1016/j.mad.2019.111195>.
- Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. *J Gerontol A Biol Sci Med Sci*. 2007;62(7):738–43. <https://doi.org/10.1093/gerona/62.7.738>.
- Kane AE, et al. A comparison of two mouse frailty assessment tools. *J Gerontol A Biol Sci Med Sci*. 2017;72(7):904–9. <https://doi.org/10.1093/gerona/glx009>.
- Malavolta M, et al. LAV-BPIFB4 associates with reduced frailty in humans and its transfer prevents frailty progression in old mice. *Aging*. 2019;11(16):6555–68. <https://doi.org/10.18632/aging.102209>.
- Kane AE, et al. Implementation of the mouse frailty index. *Can J Physiol Pharmacol*. 2017;95(10):1149–55. <https://doi.org/10.1139/cjpp-2017-0025>.
- Feridooni HA, et al. Reliability of a frailty index based on the clinical assessment of health deficits in male C57BL/6J mice. *J Gerontol A Biol Sci Med Sci*. 2015;70(6):686–93. <https://doi.org/10.1093/gerona/glu161>.
- Castro B, Kuang S. Evaluation of muscle performance in mice by treadmill exhaustion test and whole-limb grip strength assay. *Bio Protoc*. 2017;7(8):e2237. <https://doi.org/10.21769/BioProtoc.2237>.
- Deacon RM. Measuring the strength of mice. *J Vis Exp*. 2013;(76):2610. <https://doi.org/10.3791/2610>.
- Malavolta M, et al. Recovery from mild Escherichia coli O157:H7 infection in young and aged C57BL/6 mice with intact flora estimated by fecal shedding, locomotor activity and grip strength. *Comp Immunol Microbiol Infect Dis*. 2019;63:1–9. <https://doi.org/10.1016/j.cimid.2018.12.003>.
- Grohn KJ, et al. C60 in olive oil causes light-dependent toxicity and does not extend lifespan in mice. *GeroScience*. 2021;43(2):579–91. <https://doi.org/10.1007/s11357-020-00292-z>.
- Morio Y, Izawa KP, Omori Y, Katata H, Ishiyama D, Koyama S, et al. The relationship between walking speed and step length in older aged patients. *Diseases*. 2019;7(1):17. <https://doi.org/10.3390/diseases7010017>.
- Fernagut PO, et al. A simple method to measure stride length as an index of nigrostriatal dysfunction in mice. *J Neurosci Methods*. 2002;113(2):123–30. [https://doi.org/10.1016/S0165-0270\(01\)00485-x](https://doi.org/10.1016/S0165-0270(01)00485-x).
- Han Y, Eipel M, Franzen J, Sakk V, Dethmers-Ausema B, Yndriago L, et al. Epigenetic age-predictor for mice based on three CpG sites. *Elife*. 2018;7:e37462. <https://doi.org/10.7554/eLife.37462>.
- Kreidler SM, Muller KE, Grunwald GK, Ringham BM, Coker-Dukowitz ZT, Sakhadeo UR, et al. GLIMPSE: online power computation for linear models with and without a baseline covariate. *J Stat Softw*. 2013;54(10):i10. <https://doi.org/10.18637/jss.v054.i10>.
- Romero-Ortuno R. An alternative method for frailty index cut-off points to define frailty categories. *Eur Geriatr Med*. 2013;4(5):299–303. <https://doi.org/10.1016/j.eurger.2013.06.005>.
- Kwak D, Baumann CW, Thompson LV. Identifying characteristics of frailty in female mice using a

- phenotype assessment tool. *J Gerontol A Biol Sci Med Sci*. 2020;75(4):640–6. <https://doi.org/10.1093/gerona/glz092>.
29. von Zglinicki T, et al. Frailty in mouse ageing: a conceptual approach. *Mech Ageing Dev*. 2016;160:34–40. <https://doi.org/10.1016/j.mad.2016.07.004>.
 30. Kane AE, et al. Impact of longevity interventions on a validated mouse clinical frailty index. *J Gerontol A Biol Sci Med Sci*. 2016;71(3):333–9. <https://doi.org/10.1093/gerona/glu315>.
 31. Kane AE, et al. Sex differences in healthspan predict lifespan in the 3xTg-AD mouse model of Alzheimer's disease. *Front Aging Neurosci*. 2018;10:172. <https://doi.org/10.3389/fnagi.2018.00172>.
 32. Kane AE, et al. Acetaminophen hepatotoxicity in mice: effect of age, frailty and exposure type. *Exp Gerontol*. 2016;73:95–106. <https://doi.org/10.1016/j.exger.2015.11.013>.
 33. Jansen HJ, et al. Atrial structure, function and arrhythmogenesis in aged and frail mice. *Sci Rep*. 2017;7:44336. <https://doi.org/10.1038/srep44336>.
 34. Antoch MP, et al. Physiological frailty index (PFI): quantitative in-life estimate of individual biological age in mice. *Aging*. 2017;9(3):615–26. <https://doi.org/10.18632/aging.101206>.
 35. Keller K, et al. Chronic treatment with the ACE inhibitor enalapril attenuates the development of frailty and differentially modifies pro- and anti-inflammatory cytokines in aging male and female C57BL/6 mice. *J Gerontol A Biol Sci Med Sci*. 2019;74(8):1149–57. <https://doi.org/10.1093/gerona/gly219>.
 36. Huizer-Pajkos A, et al. Adverse geriatric outcomes secondary to polypharmacy in a mouse model: the influence of aging. *J Gerontol A Biol Sci Med Sci*. 2016;71(5):571–7. <https://doi.org/10.1093/gerona/glv046>.
 37. Tang Y, et al. Pre-existing weakness is critical for the occurrence of postoperative cognitive dysfunction in mice of the same age. *PLoS One*. 2017;12(8):e0182471. <https://doi.org/10.1371/journal.pone.0182471>.
 38. Mach J, et al. Preclinical frailty assessments: phenotype and frailty index identify frailty in different mice and are variably affected by chronic medications. *Exp Gerontol*. 2022;161:111700. <https://doi.org/10.1016/j.exger.2022.111700>.
 39. Seldeen KL, et al. High intensity interval training improves physical performance and frailty in aged mice. *J Gerontol A Biol Sci Med Sci*. 2018;73(4):429–37. <https://doi.org/10.1093/gerona/glx120>.
 40. Xu M, et al. Senolytics improve physical function and increase lifespan in old age. *Nat Med*. 2018;24(8):1246–56. <https://doi.org/10.1038/s41591-018-0092-9>.
 41. Baumann CW, Kwak D, Thompson LV. Sex-specific components of frailty in C57BL/6 mice. *Aging*. 2019;11(14):5206–14. <https://doi.org/10.18632/aging.102114>.
 42. Romero-Ortuno R, et al. Is phenotypical prefrailty all the same? A longitudinal investigation of two prefrailty subtypes in TILDA. *Age Ageing*. 2019;49(1):39–45. <https://doi.org/10.1093/ageing/afz129>.
 43. Baumann CW, Kwak D, Thompson LV. Phenotypic frailty assessment in mice: development, discoveries, and experimental considerations. *Physiology (Bethesda)*. 2020;35(6):405–14. <https://doi.org/10.1152/physiol.00016.2020>.
 44. Ackert-Bicknell CL, et al. Aging research using mouse models. *Curr Protoc Mouse Biol*. 2015;5(2):95–133. <https://doi.org/10.1002/9780470942390.mo140195>.
 45. Alfaro I, et al. Empirical versus theoretical power and type I error (false-positive) rates estimated from real murine aging research data. *Cell Rep*. 2021;36(7):109560. <https://doi.org/10.1016/j.celrep.2021.109560>.
 46. Demaria M, et al. An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA. *Dev Cell*. 2014;31(6):722–33. <https://doi.org/10.1016/j.devcel.2014.11.012>.
 47. Dungan CM, Figueiredo VC, Wen Y, VonLehmden GL, Zdunek CJ, Thomas NT, et al. Senolytic treatment rescues blunted muscle hypertrophy in old mice. *Geroscience*. 2022;44(4):1925–40. <https://doi.org/10.1007/s11357-022-00542-2>.
 48. Kim S, et al. The frailty index outperforms DNA methylation age and its derivatives as an indicator of biological age. *Geroscience*. 2017;39(1):83–92. <https://doi.org/10.1007/s11357-017-9960-3>.
 49. Seldeen KL, et al. High intensity interval training improves physical performance in aged female mice: a comparison of mouse frailty assessment tools. *Mech Ageing Dev*. 2019;180:49–62. <https://doi.org/10.1016/j.mad.2019.04.001>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.