DunedinPACE: A DNA methylation biomarker of the Pace of Aging

DW Belsky[1,2]
A Caspi[3-6]
DL Corcoran[6]
K Sugden[4]
R Poulton[7]
L Arseneault[3]
A Baccarelli[8]
K Chamarti[4]
X Gao[9]
E Hannon[10]
HL Harrington[4]
R Houts[4]
M Kothari[2]
D Kwon[2]
J Mill[10]
J Schwartz[11]
P Vokonas[12]
C Wang[11]
B Williams[4]
TE Moffitt[3-6]

1. Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA
2. Robert N Butler Columbia Aging Center, Columbia University Mailman School of Public Health, New York, NY, USA
3. Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AF, United Kingdom
4. Department of Psychology and Neuroscience, Duke University, Durham, NC, USA
5. Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA
6. Center for Genomic and Computational Biology, Duke University, Durham, NC, USA
7. Department of Psychology and Dunedin Multidisciplinary Health and Development Research Unit, University of Otago, Otago, NZ
8. Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY, USA
9. Department of Occupational and Environmental Health Sciences, Peking University, Bejing, China
10. Complex Disease Epigenetics Group, University of Exeter, Exeter, United Kingdom
11. Department of Environmental Health Sciences, Harvard TH Chan School of Public Health, MA, USA
12. Veterans Affairs Normative Aging Study, Veterans Affairs Boston Healthcare System, Department of Medicine, Boston University School of Medicine, MA, USA

Key Words. Biological Aging, Methylation, Geroscience, Healthspan, Biomarker, Epigenetics

Acknowledgement

This research was supported by US-National Institute on Aging grants AG032282, AG061378, AG066887, and UK Medical Research Council grant MR/P005918/1.

The Dunedin Multidisciplinary Health and Development Research Unit is supported by the New Zealand Health Research Council Programme Grant (16-604), and the New Zealand Ministry of Business, Innovation and Employment (MBIE). We thank the Dunedin Study members, Unit research staff, and Study founder Phil Silva.

Understanding Society data come from The UK Household Longitudinal Study, which is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council (ES/M008592/1). Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/. Data governance was provided by the METADAC data access committee, funded by ESRC, Wellcome, and MRC (2015-2018: MR/N01104X/1; 2018-2020: ES/S008349/1)

The Normative Aging Study is supported by the National Institute of Environmental Health Sciences (grants P30ES009089, R01ES021733, R01ES025225, and R01ES027747). The VA Normative Aging Study is supported by the Cooperative Studies Program/Epidemiology Research and Information Center of the U.S. Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center, Boston, Massachusetts.

The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195, HHSN268201500001I and 75N92019D00031). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.

The E-Risk Study is supported by the UK Medical Research Council (grant G1002190), the US National Institute of Child Health and Development (grant HD077482), and the Jacobs Foundation. The generation of DNA methylation data was supported by the American Asthma Foundation.

This work used a high-performance computing facility partially supported by grant 2016-IDG-1013 (HARDAC+: Reproducible HPC for Next-generation Genomics") from the North Carolina Biotechnology Center.

We would like to acknowledge the assistance of the Duke Molecular Physiology Institute Molecular Genomics Core for the generation of data for the manuscript.

AC, DC, DWB, KS, RP, and TEM are listed as inventors on a Duke University and University of Otago invention that was licensed to a commercial entity.

ABSTRACT

Measures to quantify changes in the pace of biological aging in response to intervention are needed to evaluate geroprotective interventions for humans. We used data from the Dunedin Study 1972-3 birth cohort tracking within-individual decline in 19 indicators of organ-system integrity across four time points spanning two decades to model Pace of Aging. We distilled this two-decade Pace of Aging into a single-time-point DNA-methylation blood-test using elastic-net regression and a DNA-methylation dataset restricted to exclude probes with low test-retest reliability. We evaluated the resulting measure, named DunedinPACE, in five additional datasets. DunedinPACE showed high test-retest reliability, was associated with morbidity, disability, and mortality, and indicated faster aging in young adults with childhood adversity. DunedinPACE effect-sizes were similar to GrimAge Clock effect-sizes. In analysis of morbidity, disability, and mortality, DunedinPACE and added incremental prediction beyond GrimAge. DunedinPACE is a novel blood biomarker of the pace of aging for gerontology and geroscience.

INTRODUCTION

Guided by research on the cellular hallmarks of aging, model-organism studies are yielding treatments that slow the accumulation of deterioration in cells and organs (Barzilai et al., 2018; Bellantuono, 2018; Campisi et al., 2019). However, human geroprotector trials face a barrier when it comes to outcome measures: humans' long lifespan makes it prohibitively time-consuming to test whether a treatment extends healthspan (Justice et al., 2018). What is needed is a way to measure each clinical trial participant's personal pace of biological aging, before, during, after, and at long-term follow-up, to test whether a gero-protective therapy slows that pace, and whether benefits fade out (Belsky et al., 2017; Moffitt et al., 2017; Sierra et al., 2021).

An ideal research design for isolating biological patterns that differentiate faster and slower aging is longitudinal follow-up of a cohort of individuals who are all the same chronological age. We originally applied this design to test the geroscience-derived hypothesis that the process of biological aging was ongoing in healthy young adults and, critically, that it was already variable across individuals decades before most chronic diseases onset (Belsky et al., 2015; Elliott et al., 2021). We studied a single-year birth cohort of all 1,037 babies born in the city of Dunedin, New Zealand, during 1972-73 and followed-up through midlife. Because aging can be understood as gradual, progressive deterioration simultaneously affecting different organ systems, we tracked declines in the cardiovascular, metabolic, renal, hepatic, immune, periodontal, and pulmonary systems of each participant using multiple biomarker measurements of each organ system, first at ages 26, 32, and 38 years, and subsequently including a fourth assessment at age 45 years. Modelling this one-of-a-kind dataset yielded a metric which quantified how slowly or rapidly each participant had been aging. We called this metric the Pace of Aging.

People with faster Pace of Aging tend to experience more rapid aging-related decline in physical and cognitive functions and develop early signs of brain aging (Belsky et al., 2015; Elliott et al., 2019, 2021; Rasmussen et al., 2019). This evidence suggests that intervention to slow Pace of Aging could preserve functions lost with aging and extend healthspan. However, Pace of Aging takes many years to measure and requires data from several types of clinical and biological assays, limiting its utility in clinical trials. To address this limitation, we applied machine-learning tools to genome-wide DNA methylation data from Dunedin cohort members'

blood samples to distill the multi-timepoint, multi-assay Pace of Aging measure into a single-timepoint, single-assay blood biomarker previously described in this journal: DunedinPoAm (for Dunedin Study Pace of Aging from methylation) (Belsky et al., 2020). DunedinPoAm was designed to measure Pace of Aging biological change over time from a single blood sample. Like the original Pace of Aging, people with faster DunedinPoAm scores more often experienced declines in cognitive and physical functioning by midlife and showed more rapid facial aging (Belsky et al., 2020); in older adults, faster DunedinPoAm predicted increased risk of disease and death (Belsky et al., 2020; Graf et al., In Press); in young people, experiences of early-life adversity were linked to faster DunedinPoAm (Belsky et al., 2020; Raffington et al., 2021).

DunedinPoAm provided proof of concept that Pace of Aging could be measured from a single blood test. But DunedinPoAm was based on biological change observed over just twelve years spanning the transition from young adulthood to midlife, limiting the scope of biological change that could be observed. And this biological change was measured from just three assessment waves, limiting precision. In addition, similar to other DNA methylation composites measured from microarray data, the measure had only moderate test-retest reliability (Higgins-Chen et al., 2021), limiting its value to clinical trials seeking to test within-participant changes from pre-treatment baseline to post-treatment follow-up. To address these limitations, we incorporated new data from the Dunedin cohort extending Pace of Aging follow-up to include a fourth measurement occasion in the fifth decade of life (Elliott et al., 2021) and restricted DNA-methylation data to exclude probes with poor reliability (Sugden et al., 2020) to refine our DNA-methylation biomarker of Pace of Aging. The longer follow-up period, which allowed us to observe more aging-related change in system integrity, and the additional measurement occasion included in analysis improve the power and precision of our modeling of longitudinal change in system integrity (Brandmaier et al., 2018; Rast and Hofer, 2014). The refined DNA-methylation dataset used for machine-learning improves reliability of measurement. We name the novel algorithm DunedinPACE, for Dunedin (P)ace of (A)ging (C)alculated from the (E)pigenome.

As a new DNA methylation measure of biological aging, DunedinPACE joins a well-established battery of DNA methylation measures known as “clocks”, so-named for their accurate prediction of chronological age (Horvath and Raj, 2018). However, DunedinPACE is distinct from the DNA methylation clocks in the method of its design and in its interpretation. DunedinPACE is derived from analysis of longitudinal data collected from a cohort of

individuals who are all the same chronological age. It reflects differences between those individuals in the rate of deterioration in system integrity occurring over a fixed time interval, age 26 to age 45 years. It therefore has three distinguishing features that isolate aging from other confounds: (1) analysis of a single-year birth cohort distinguishes aging from confounding effects of exposures to factors that alter the methylome and that differ across generations, referred to hereafter as "cohort effects"; (2) initiation of follow-up in young adults distinguishes aging from the effects of disease and excludes contamination from survival bias; and (3) focus on changes in multi-organ-system integrity during adult life distinguishes ongoing aging processes from deficits established early in development. The clocks, in contrast, were developed using a single cross-section of data from samples of mixed-age and older adults. They do not distinguish aging from cohort effects, are vulnerable to survival bias and contamination of aging signal by disease processes, and do not distinguish aging-related changes from prenatal/childhood deficits. DunedinPACE, therefore, represents a complementary tool with useful advantages for DNA methylation quantification of biological aging.

RESULTS

We developed the DunedinPACE measure reported here by analyzing the pace of biological aging in a 1972-73 birth cohort (N=1037), the Dunedin Study. This analysis consisted of two parts. The first part of analysis followed our original method (Belsky et al., 2015) to quantify Pace of Aging from two decades of longitudinal organ-system integrity data (Elliott et al., 2021). First, we measured longitudinal changes in 19 biomarkers assessing cardiovascular, metabolic, renal, hepatic, immune, dental, and pulmonary systems, measured at ages 26, 32, 38 and, most recently in 2019, 45 years (**Supplementary Table S1** provides measurement details about the tracked biomarkers). Second, linear mixed-effects modelling was used to fit growth models to estimate each study member's personal rate of change for each of the 19 biomarkers. Third, these 19 personal rates of change were combined across biomarkers to calculate each Study member's personal Pace of Aging. In line with the geroscience hypothesis (Kaeberlein, 2013; Kennedy et al., 2014), which states that aging represents progressive decline across organ systems, we calculated each Study member's Pace of Aging as the sum of age-

dependent annual changes across all biomarkers: Pace of Aging$_i$ = $\sum_{B=1}^{19} \mu_{1iB}$. The resulting Pace of Aging was then scaled to a mean of 1, so that it could be interpreted with reference to an average rate of 1 year of biological aging per year of chronological aging. Study members showed wide variation in their Pace of Aging (Mean=1 biological year per chronological year, SD=0.29). Over the two decades that we measured biological aging, the Study member with the slowest Pace of Aging aged by just 0.40 biological years per chronological year, while the Study member with the fastest Pace of Aging accrued 2.44 biological years per chronological year (Elliott et al., 2021).

The second part of the analysis followed our original method to distill Pace of Aging into a blood test (Belsky et al., 2020). Using Illumina EPIC array DNA methylation data from blood collected at age 45, we developed DunedinPACE to provide a surrogate for the 20-year Pace of Aging. Briefly, we used elastic-net-regression (Zou and Hastie, 2005) to develop a DNA methylation algorithm to predict the 20-year Pace of Aging. Analysis included the subset of probes included on both the Illumina 450k and EPIC arrays which we previously determined to have acceptable test-retest reliability (ICC>0.4, n=81,2392) (Sugden et al., 2020). Following the method established by Horvath (Horvath, 2013) and used in our prior analysis, we fixed the alpha parameter of the elastic net at 0.5. The final lamda value selected by the analysis was 0.042. The resulting algorithm included 173 CpG sites. The CpGs included in DunedinPACE are listed in **Supplementary Table S2**.

DunedinPACE distills the 20-year Pace of Aging into a single-timepoint DNA methylation blood test. DunedinPACE demonstrated a high in-sample correlation with Pace of Aging to age 45 (r=0.78, **Figure 1a**). The value indicates improved fit of the DunedinPACE model to the 20-year Pace of Aging as compared to the original DunedinPoAm version's fit to the 12-year Pace of Aging (in-sample correlation r=0.6). Correlations among the 12- and 20-year Pace of Aging measures and the DunedinPoAm and DunedinPACE DNA methylation measures are shown in **Supplementary Figure 1a**.

Consistent with the high correlation of DunedinPACE with 20-year Pace of Aging, effect-sizes for associations of DunedinPACE with measures of physical and cognitive functioning and subjective signs of aging measured at age 45 in the Dunedin Study were similar to effect-sizes for the 20-year Pace of Aging. Effect-sizes for DunedinPACE are graphed in

Figure 1b. Comparison with effect-sizes for the 20-year Pace of Aging reported previously (Elliott et al., 2021; Rasmussen et al., 2019) are shown in **Supplementary Figure 1b**.

We also repeated analysis of decline in physical and cognitive functions included in our original 2020 *eLife* report of DunedinPoAm. This analysis tested associations between DunedinPACE and worsening in physical functions and subjective signs of aging between the age-38 and age-45 assessments and change in cognitive functioning from adolescent baseline to the age-45 assessment. Effect-sizes are graphed in **Figure 1c**.

Comparison of DunedinPACE effect-sizes for associations with function tests and subjective signs of aging to effect-sizes for the original DunedinPoAm and the DNA methylation clocks proposed by Horvath, Hannum et al., Levine et al. (PhenoAge Clock), and Lu et al. (GrimAge Clock) are shown in **Supplementary Figure 1c**. DunedinPACE effect-sizes are similar to or larger than those for DunedinPoAm and the GrimAge clock and larger than those for the other DNA methylation clocks. DunedinPACE and DunedinPoAm were developed from analysis of Pace of Aging in the Dunedin cohort; effect-sizes for these measures may be over-estimated in analyses of Dunedin data. We therefore conducted further validation analysis in four additional cohorts.

DunedinPACE shows exceptional test-retest reliability. A Pace of Aging measure should have high test-retest reliability in order to be able to quantify change from pre-treatment baseline to post-treatment follow-up in intervention studies or exposure-related changes in longitudinal observational studies. We first evaluated test-retest reliability of DunedinPACE using a publicly available database of 36 replicate adult-blood-sample Illumina 450k datasets (Lehne et al., 2015) (GEO Accession GSE55763). We estimated intraclass correlation coefficients (ICCs) for these replicates using mixed-effects regression. For the original DunedinPoAm, reliability was good (ICC=0.89 95% CI [0.79-0.94]). For DunedinPACE, reliability was excellent (ICC=0.96 [0.93-0.98]; **Figure 2**).We repeated this analysis in two additional databases of replicates reported by Sugden et al. (Sugden et al., 2020), a database of n=28 replicate Illumina EPIC-array datasets and a database of n=350 replicates in which one dataset was generated from the Illumina 450k array and the other dataset was generated from the Illumina EPIC array. DunedinPACE ICCs were 0.97 [0.94-0.98] in the EPIC-EPIC database and 0.87 [0.82-0.90] in the 450k-EPIC database.

Comparison of DunedinPACE reliability to reliabilities for the original DunedinPoAm and the DNA methylation clocks are shown in **Supplementary Figure 2**. DunedinPACE is as reliable or more so as compared to the GrimAge clock and more reliable in comparison to the other clocks, although methods have been proposed that may improve reliabilities for the other clocks (Higgins-Chen et al., 2021).

We conducted analysis to evaluate the extent to which improved fit to longitudinal Pace of Aging and better test-retest reliability for DunedinPACE as compared to the original DunedinPoAm might reflect (a) the four-time-point follow-up design of 20-year Pace of Aging and (b) the restriction of elastic net regression to develop DunedinPACE to probes that demonstrated higher test-retest reliabilities in blood in an independent sample. Results suggest that both four-time-points to measure Pace of Aging and restricting machine-learning analysis to CpG sites with higher test-retest reliability contributed to the technical improvement in DunedinPACE over DunedinPoAm (see **Supplementary Results**).

DunedinPACE is faster in chronologically older individuals. In demography, the pace of biological aging in a population can be estimated from the rate of increase in mortality risk from younger to older chronological ages (Hägg et al., 2019). In humans and many other species, the increase in mortality risk with advancing chronological age is nonlinear, suggesting that the pace of biological aging accelerates as we grow older (Gompertz, 1825; Kirkwood, 2015; Olshansky and Carnes, 1997). We tested if, consistent with this hypothesis, DunedinPACE was faster in chronologically older as compared to younger individuals in data from the Understanding Society Study (n=1175, age range 28-95). Chronologically older Understanding Society participants had faster DunedinPACE as compared to younger ones (r=0.32, **Figure 3 Panel A**). This correlation was threefold larger as compared to the correlation for the original DunedinPoAm (r=0.11).

Between-individual correlation of older chronological age with faster DunedinPACE could be confounded by differences in exposure histories across birth cohorts, i.e. not an effect of within-individual aging (Moffitt, 2020). To rule out such confounding, we conducted analysis of within-individual change in DunedinPACE in the US Veterans Administration Normative Aging Study (NAS). Among NAS participants with repeated-measures DNA methylation data (n=536),

DunedinPACE increased by b=0.021 (SE=0.003) per 5 years of follow-up, or approximately 0.2 standard deviation units, similar to what we previously reported for the original DunedinPoAm (Belsky et al., 2020).

DunedinPACE is faster in individuals measured to be biologically older using prior measures derived from DNA methylation and physiological data, and who report poorer self-rated health. We compared DunedinPACE with the epigenetic clocks proposed by Horvath, Hannum et al., Levine et al., and Lu et al. (Hannum et al., 2013; Horvath, 2013; Levine et al., 2018; Lu et al., 2019), with measurements of biological age derived from physiological data and with participants' subjective perceptions of their own health status in data from the UK-based Understanding Society Study (n=1175, mean age=58, SD=15, 42% male).

Epigenetic Clock Analysis. For each epigenetic clock, we calculated the measure referred to as epigenetic "age acceleration" by regressing participants' clock-estimated ages on their chronological ages and computing residual values. This measure is often used as an estimate of the aging rate. It quantifies the difference between the amount of aging expected based on a person's chronological age and the amount of aging observed based on DNA methylation. DunedinPACE was weakly correlated with age acceleration computed from the Horvath clock (r=0.13) and somewhat more strongly correlated with age acceleration computed from the Hannum and Levine clocks (Hannum r=0.26; PhenoAge r=0.36). These correlations were similar to those reported on our 2020 *eLife* article for the original DunedinPoAm measure. For the GrimAge clock proposed by Lu et al., the correlation with the age-acceleration residuals was stronger (r=0.58). Correlations are plotted in **Figure 3 Panel B**.

Physiology-based Measurements of Biological Age. We used published methods to measure participants' biological ages from blood-chemistry, blood-pressure, and lung function data collected at the same time as the DNA methylation data. We computed indices of biological age using three different methods: the Klemera and Doubal method (KDM) Biological Age estimates the age at which a person's physiology would appear normal in a reference population (Klemera and Doubal, 2006); the Phenotypic Age method estimates the age at which a person's physiology-predicted mortality risk would be approximately normal in a reference population (Levine et al., 2018); the homeostatic dysregulation method estimates the extent to which a person's physiology deviates from a state of health (Cohen et al., 2013). These algorithms

generate correlated but distinct estimates conceptually related to biological age (Hastings et al., 2019; Parker et al., 2019). Algorithms were parameterized using a set of biomarkers available in Understanding Society and included in Levine's original implementation of the Klemera-Doubal Biological Age (Levine, 2013) (albumin, alkaline phosphatase, blood urea nitrogen, creatinine, C-reactive protein, HbA1C, systolic blood pressure, forced expiratory volume in one second) and training data from the US National Health and Nutrition Examination Surveys (NHANES) (Kwon and Belsky, 2021). Consistent with analysis of epigenetic age acceleration, participants with more advanced biological aging measured from physiological data also showed faster DunedinPACE (KDM Biological Age Advancement r=0.30; Phenotypic Age Advancement r=0.32; Homeostatic Dysregulation r=0.09; **Figure 3 Panel C**).

Self-rated Health. Participants who rated themselves as being in worse health also showed faster DunedinPACE. The difference in DunedinPACE between those rating their health as excellent and poor was large (Cohen's d=0.74 [0.46-1.03]; **Figure 3 Panel C**).

Comparison of effect-sizes for associations with physiology-based measures of biological age and self-rated health between DunedinPACE and the original DunedinPoAm and the four DNA methylation clocks is shown in **Supplementary Figure 3**. DunedinPACE effect-sizes are similar to or larger than effect-sizes for DunedinPoAm and the GrimAge clock and larger in comparison to the other clocks.

Faster DunedinPACE is associated with morbidity, disability, and mortality. We conducted analysis of morbidity and mortality in older men in the VA Normative Aging Study (NAS) and in the Framingham Heart Study Offspring cohort.

NAS analysis was conducted using the same methods reported in our original article (Belsky et al., 2020). Briefly, analysis included n=771 older men with average age of 77 years (SD=7) at DNA methylation measurement. Over follow-up from DNA collection in 1999-2013, 46% died and 23% were diagnosed with a new chronic disease, including any of hypertension, type-2 diabetes, cardiovascular disease, chronic obstructive pulmonary disease, chronic kidney disease, and cancer. Participants with faster DunedinPACE were at increased risk for morbidity and mortality (for incident chronic disease morbidity, DunedinPACE HR=1.23 [1.07-1.42]; for prevalent chronic disease morbidity, DunedinPACE RR=1.16 [1.12-1.20]; for mortality, DunedinPACE HR=1.26 95% CI [1.14-1.40], **Figure 4A**).

Framingham analysis included n=2,471 members of the Offspring cohort (54% women) with average age of 66 years (SD=9) at DNA methylation measurement during 2005-2008. Over follow-up through 2018, **23%** died, **13%** were newly diagnosed with cardiovascular disease (CVD), and **6%** had a first stroke or transient ischemic attack (TIA). Disability follow-up was conducted through 2015 based on participant reports of limitations to activities of daily living (ADLs) on the Nagi, Katz, and Rosow-Brelsau scales. Participants with faster DunedinPACE at baseline were at increased risk for CVD, stroke/TIA, and mortality (CVD HR=1.39 [1.26-1.54]; stroke/TIA HR=1.37 [1.19-1.58]; mortality HR=1.65 95% CI [1.51-1.79], **Figure 4B**). They were also more likely to develop disability (Nagi ADL IRR=1.40 [1.19-1.65]; Katz ADL IRR=1.33 [1.16-1.53]; Rosow-Breslau ADL IRR=1.39 [1.24-1.56]).

Comparison of effect-sizes for associations with morbidity, disability, and mortality between DunedinPACE and the original DunedinPoAm and the DNA methylation clocks are shown in **Supplementary Figure 4**. DunedinPACE effect-sizes were similar to those for DunedinPoAm and the GrimAge clock and larger than effect-sizes for the other measures.

Childhood exposure to poverty and victimization is associated with faster DunedinPACE. A Pace of Aging measure should be sensitive to histories of exposure associated with earlier onset of morbidity and shorter lifespan and should provide indications of faster aging before clinical signs of aging-related disease develop, in time for prevention. To test if DunedinPACE indicated faster aging in young people with histories of exposure thought to shorten healthy lifespan, we analyzed data from n=1658 members of the E-Risk Longitudinal Study. The E-Risk Study follows a 1994–95 birth cohort of same-sex twins. Blood DNA methylation data were collected when participants were aged 18 years. We analyzed two exposures associated with shorter healthy lifespan, childhood low socioeconomic status and childhood victimization. Socioeconomic status was measured from data on parents' education, occupation, and income (Trzesniewski et al., 2006). Victimization was measured from exposure dossiers compiled from interviews with the children's mothers and home-visit assessments conducted when the children were aged 5, 7, 10, and 12 (Fisher et al., 2015). The dossiers recorded children's exposure to domestic violence, peer bullying, physical and sexual harm by an adult, and neglect. 72% of the analysis sample had no victimization exposure, 21% had one type of victimization exposure, 4% had two types of exposure, and 2% had three or more types of exposure.

E-Risk participants with exposure to childhood poverty and victimization showed faster DunedinPACE (For low as compared to high childhood socioeconomic status, d=0.38 95% CI [0.25-0.51]; for polyvictimization as compared to no victimization, d=0.47 [0.17-0.77]; **Figure 5**).

Comparison of effect-sizes for associations with childhood poverty and victimization between DunedinPACE and the original DunedinPoAm and the DNA methylation clocks are shown in **Supplementary Figure 5**. DunedinPACE effect-sizes were similar to those for DunedinPoAm and the GrimAge clock and larger than effect-sizes for the other clocks.

Sensitivity Analysis. Results for analysis of all studies including covariate adjustment for cell counts and smoking are reported in **Supplementary Table 4**. Covariate adjustment for estimated cell counts and participant reports of smoking history resulted in modest attenuation of some effect-sizes, although DunedinPACE effect-sizes remained statistically different from zero in nearly all cases (associations with declines in balance and cognitive function were attenuated below the alpha=0.05 threshold in Dunedin Study analysis including smoking as a covariate; differences between un-victimized and those with one type of victimization were attenuated below the alpha=0.05 threshold in E-Risk Study analysis including smoking as a covariate; differences between high- and middle-SES groups were attenuated below the alpha=0.05 threshold in E-Risk Study analysis excluding smokers). Compared to the original DunedinPoAm, covariate adjustment for smoking tended to have a smaller impact on effect-size estimates for DunedinPACE. In follow-up analysis of the Understanding Society Study, we found that participant-reported smoking history explained 30% of variance in DunedinPoAm, but only 13% of variance in DunedinPACE. Results were similar in analysis of the Framingham Heart Study Offspring Cohort (R^2=23% for DunedinPoAm vs. 12% for DunedinPACE).

Incremental predictive value of DunedinPACE in an analysis of morbidity, disability, and mortality. DunedinPACE is distinct from DNA methylation clocks in the method of its development and in its interpretation. However, DunedinPACE is moderately correlated with some DNA methylation clocks, in particular GrimAge (see **Figure 3 Panel B**). We therefore conducted analysis to test if DunedinPACE contributed new information about health-span and lifespan over and above existing DNA methylation clocks. We tested if DunedinPACE

associations with morbidity, disability, and mortality were statistically independent of each of the DNA methylation clocks within the Framingham Heart Study. In the case of the GrimAge clock, which was developed to predict mortality using this Framingham dataset, this analysis provides an especially rigorous test.

DunedinPACE associations with health-span endpoints were little-changed by covariate adjustment for the Horvath, Hannum, or PhenoAge Clocks. In models adjusted for GrimAge, DunedinPACE associations with mortality, CVD, and disability were attenuated, but remained statistically different from zero (mortality HR=1.24 [1.49-1.74], CVD HR=1.18 [1.05-1.34], Nagi ADL IRR=1.27 [1.02-1.58], Katz ADL IRR=1.26 [1.02-1.54], Rosow-Breslau ADL IRR=1.27 [1.08-1.50]); associations with stroke were similar to unadjusted models (HR=1.33 [1.05-1.69]). Results for all models are reported in **Supplementary Table 5**. Thus, Dunedin PACE adds incremental prediction over and above all clocks studied here.

DISCUSSION

We developed a novel DNA methylation measure of the pace of biological aging from analysis of two decades of longitudinal data from members of the Dunedin Study 1972-3 birth cohort. The new measure is named DunedinPACE. As in our original report (Belsky et al., 2020), we utilized a two-step approach that first modeled the Pace of Aging from within-individual changes in organ-system integrity markers and then distilled this longitudinal change measure into a single-time-point DNA-methylation blood test. DunedinPACE is therefore distinct from the DNA methylation clocks in both theory and method. The clocks estimate the progress of aging up to a cross-section in time. DunedinPACE is a DNA methylation estimate of the Pace of Aging, the ongoing rate of decline in system integrity.

The DunedinPACE measure advances DNA methylation measurement of the Pace of Aging for clinical trials of geroprotective therapies beyond our original DunedinPoAm measure in four ways: (1) Pace of Aging analysis included 20 years of follow-up (instead of 12); (2) Pace of Aging analysis included four time points of measurement (instead of three); (3) DNA methylation modeling of Pace of Aging excluded CpG sites for which probes were determined to have low test-retest reliability in blood datasets; (4) the algorithm to implement DunedinPACE

includes a normalization step that allows DunedinPACE values for individual samples to be compared to the Dunedin Study cohort in which the measure was developed.

As a result of these advances, DunedinPACE provides a more precise measurement of the Pace of Aging as compared to the original DunedinPoAm and shows higher test-retest reliability. Interestingly, in contrast to reports for some epigenetic clocks (Higgins-Chen et al., 2021), the restriction of machine learning analysis to a more-reliably-measured subset of CpG sites on the Illumina EPIC array did not harm criterion validity of DunedinPACE; effect-sizes for prediction of morbidity and mortality were as large or larger in comparison to DunedinPoAm. Analysis of this more precise and reliable measurement revealed stronger associations with signs of aging. DunedinPACE showed a stronger correlation with chronological age as compared to the original DunedinPoAm, consistent with demographic evidence that biological processes of aging accelerate in later life (Finch and Crimmins, 2016). In analysis of construct and criterion validity, DunedinPACE was associated with functional decline in midlife adults; it was correlated with measures of biological age derived from blood chemistry and DNA methylation data, and with research participants' subjective perceptions of their own health; it was predictive of morbidity and mortality in midlife and older adults; and it indicated faster Pace of Aging in young adults with histories of exposure to poverty and victimization. DunedinPACE effect-sizes were similar to or larger than those for the original DunedinPoAm measure even as DunedinPACE was less influenced by participant smoking history. In analysis that took smoking history into account, DunedinPACE effect-sizes were more robust than those for DunedinPoAm, most notably in analysis of mortality in the Framingham Heart Study and of early-life adversity in the E-Risk cohort.

DunedinPACE effect-sizes in analysis of construct and criterion validity were also larger than those for benchmark DNA methylation clocks proposed by Horvath, Hannum et al., and Levine et al. (PhenoAge), and similar to those for the clock proposed by Lu et al. (GrimAge), and were robust to covariate adjustment for smoking and DNA methylation estimates of cell counts. Moreover, in Framingham Heart Study Offspring Cohort analysis of incremental predictive validity, DunedinPACE associations with morbidity, disability, and mortality were robust to covariate adjustment for the DNA methylation clocks, including GrimAge, although some associations were attenuated in GrimAge-adjusted models. Framingham analysis represents

a decisive test of value added by DunedinPACE over and above GrimAge because GrimAge was developed within the Framingham cohort.

In sum, DunedinPACE represents a novel measure of aging that is conceptually and empirically distinct from the DNA methylation clocks and that constitutes an advance on the original DunedinPoAm measure. DunedinPACE can complement existing DNA methylation measures of aging in analysis of observational studies and randomized trials to help advance the frontiers of geroscience.

We acknowledge limitations. Foremost, the Dunedin Study sample we analyzed to develop DunedinPACE is a relatively modestly sized cohort and is drawn from a single country. As additional cohorts develop the multi-time-point organ-system-function datasets to model Pace of Aging, integrating these data into DNA methylation algorithms to measure Pace of Aging may enhance precision. Nevertheless, DunedinPACE does show consistent evidence of criterion and construct validity in several additional datasets, boosting confidence in its external validity. In addition, our work on DunedinPACE thus far has not addressed population diversity in biological aging. The Dunedin cohort and the Understanding Society, NAS, Framingham, and E-Risk cohorts were mostly of white European descent. Follow-up of DunedinPACE in more diverse samples is needed to establish cross-population validity. Importantly, our original DunedinPoAm measure has been followed-up in diverse cohorts and has showed evidence of consistent validity across race/ethnic subgroups (Crimmins et al., 2021; Graf et al., In Press; Raffington et al., 2021; Schmitz et al., 2021). Criterion validity analyses conducted in this article do not consider specific diseases or causes of death because the cohorts used for follow-up are relatively small. Geroscience theory predicts aging should increase risk of chronic disease and death overall. Analysis of DunedinPACE in larger datasets with more deaths will allow for tests of whether particular incident diseases or causes of death drive associations. Comparative analyses in this article include only other DNA methylation measures of aging. Recent reports suggest valuable information about aging and healthspan may also be captured in metabolomic and proteomic datasets (Deelen et al., 2019; Eiriksdottir et al., 2021; Jansen et al., 2021; Lehallier et al., 2019; Tanaka et al., 2020). Future studies should compare DunedinPACE to measures derived from these biological levels of analysis. As with all proposed measures of biological aging, ultimately establishing DunedinPACE as a surrogate endpoint for healthspan

will require evidence that it is modifiable by intervention and that intervention-induced changes result in changes in healthy-lifespan phenotypes (Justice et al., 2018; Prentice, 1989).

Within the bounds of these limitations, our analysis establishes DunedinPACE as a novel single-time-point measure that quantifies Pace of Aging from a blood test. It can be implemented in Illumina 450k or in EPIC array data, making it immediately available for testing in a wide range of existing datasets as a complement to existing methylation measures of aging. Code to compute DunedinPACE is publicly available on GitHub. Through collaboration with ongoing cohort studies, DunedinPACE will soon be available from the US Health and Retirement Study and other studies. Because of its high test-retest reliability, DunedinPACE offers a unique measurement for intervention trials and natural experiment studies investigating how the rate of aging may be changed by behavioral or drug therapy, or by environmental modification.

MATERIALS AND METHODS

Data Sources

Data were used from five studies: Dunedin Study, Understanding Society Study, the Normative Aging Study (NAS), the Framingham Heart Study, and the Environmental Risk (E-Risk) Longitudinal Twin Study. In addition, we accessed the Gene Expression Omnibus dataset GSE55763. The datasets and measures analyzed within each of them are described below.

The Dunedin Study is a longitudinal investigation of health and behavior in a complete birth cohort. Study members (N=1,037; 91% of eligible births; 52% male) were all individuals born between April 1972 and March 1973 in Dunedin, New Zealand (NZ), who were eligible based on residence in the province and who participated in the first assessment at age 3. The cohort represents the full range of socioeconomic status on NZ's South Island and matches the NZ National Health and Nutrition Survey on key health indicators (e.g., BMI, smoking, physical activity, GP visits) (Poulton et al., 2015). The cohort is primarily white (93%) (Poulton et al., 2015). Assessments were carried out at birth and ages 3, 5, 7, 9, 11, 13, 15, 18, 21, 26, 32, 38 and, most recently, 45 years, when 94% of the 997 study members still alive took part. At each assessment, each study member is brought to the research unit for a full day of interviews and examinations. Study data may be accessed through agreement with the Study investigators

(https://moffittcaspi.trinity.duke.edu/research-topics/dunedin). Dunedin Study measures of physical and cognitive functioning and subjective signs of aging are described in detail in **Supplementary Table 3**.

GSE55763 is a publicly available dataset including technical replicate Illumina 450k DNA methylation array data for 36 adult human samples (Lehne et al., 2015). Data are available from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55763).

Understanding Society is an ongoing panel study of the United Kingdom population (https://www.understandingsociety.ac.uk/). During 2010-12, participants were invited to take part in a nurse's exam involving a blood draw. Of the roughly 20,000 participants who provided clinical data in this exam, methylation data have been generated for just under 1,200. We analyzed data from 1,175 participants with available methylation and blood chemistry data. Documentation of the methylation (University of Essex) and blood chemistry (University of Essex) data resource is available online (https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/health/user-guides/7251-UnderstandingSociety-Biomarker-UserGuide-2014.pdf).

Physiology-based Biological Age Measures. We measured biological aging from blood chemistry, systolic blood pressure, and lung-function data using the algorithms proposed by Klemera and Doubal, Levine, and Cohen (Cohen et al., 2013; Klemera and Doubal, 2006; Levine, 2013; Levine et al., 2018), trained in data from US National Health and Nutrition Examination Surveys (NHANES; https://wwwn.cdc.gov/nchs/nhanes/Default.aspx) following the methods described by Hastings (Hastings et al., 2019) and using the software developed by Kwon (https://github.com/dayoonkwon/BioAge). Following the procedure in our original eLife article (Belsky et al., 2020), we included 8 of 10 biomarkers included in Levine's original implementation of the Klemera-Doubal method Biological Age (Levine, 2013): albumin, alkaline phosphatase (log), blood urea nitrogen, creatinine (log), C-reactive protein (log), HbA1C, systolic blood pressure, and forced expiratory volume in 1 second (FEV_1). We omitted total cholesterol because of evidence this biomarker shows different directions of association

with aging in younger and older adults (Arbeev et al., 2016). Cytomegalovirus optical density was not available in the Understanding Society database.

Self-rated Health. Understanding Society participants rated their health as excellent, very-good, good, fair, or poor. We standardized this measure to have Mean=0, Standard Deviation=1 for analysis.

The Normative Aging Study (NAS) is an ongoing longitudinal study on aging established by the US Department of Veterans Affairs in 1963. Details of the study have been published previously (Bell et al., 1972). Briefly, the NAS is a closed cohort of 2,280 male veterans from the Greater Boston area enrolled after an initial health screening to determine that they were free of known chronic medical conditions. Participants have been re-evaluated every 3–5 years on a continuous rolling basis using detailed on-site physical examinations and questionnaires. DNA from blood samples was collected from 771 participants beginning in 1999. We analyzed blood DNA methylation data from up to four repeated assessments conducted through 2013 (Gao et al., 2019; Panni et al., 2016). Of the 771 participants with DNA methylation data, n=536 (70%) had data from 2 repeated assessments and n=178 (23%) had data from three or four repeated assessments. We restricted the current analysis to participants with at least one DNA methylation data point. The NAS was approved by the Department of Veterans Affairs Boston Healthcare System and written informed consent was obtained from each subject before participation.

Mortality. Regular mailings to study participants have been used to acquire vital-status information and official death certificates were obtained from the appropriate state health department to be reviewed by a physician. Participant deaths are routinely updated by the research team and the last available update was on 31 December 2013. During follow-up, n=354 (46%) of the 771 NAS participants died.

Chronic Disease Morbidity. We measured chronic disease morbidity from participants medical histories and prior diagnoses (Gao et al., 2021; Lepeule et al., 2018). We counted the number of chronic diseases to compose an ordinal index with categories of 0, 1, 2, 3, or 4+ of the following comorbidities: hypertension, type-2 diabetes, cardiovascular disease, chronic obstructive pulmonary disease, chronic kidney disease, and cancer.

The Framingham Heart Study tracks the development of cardiovascular disease in three generations of families recruited in Framingham Massachusetts beginning in 1948 (Tsao and Vasan, 2015). We analyzed data from the Offspring Cohort, the second generation of study participants, who were recruited beginning in 1971. Blood samples used for DNA methylation analysis were collected from Offspring Cohort members at their 8th follow-up visit during 2005-2008. We analyzed data for n=2,471 cohort members with available DNA methylation and mortality data. This sample was 54% women and had an average age of 66 years (SD=9) at the time blood draws for DNA methylation analysis were conducted. Data for the Framingham Study were obtained from dbGaP (phs000007.v32.p13).

Mortality. Mortality ascertainment in our dataset extended through 2018 (dbGaP accession pht003317.v9.p13, dataset vr_survdth_2018_a_1268s). Participants contributed a maximum of 14 years of follow-up time. A total of 575 deaths were recorded at an average age of 81 years (SD=9).

Cardiovascular Disease (CVD). CVD ascertainment in our dataset extended through 2018 (dbGaP accession pht003316.v9.p13, dataset vr_survcvd_2018_a_1267s). N=2,075 participants free of cardiovascular disease at baseline contributed a maximum of 14 years of follow-up time during which 327 incident cases of CVD were recorded at an average age of 77 years (SD=6).

Stroke/ Transient Ischemic Attack (TIA). Stroke/TIA ascertainment in our dataset extended through 2018 (dbGaP accession pht006024.v4.p13, dataset vr_svstktia_2018_a_1270s). N=2,364 participants with no history of stroke/TIA at baseline contributed a maximum of 14 years of follow-up time during which 149 events were recorded at an average age of 78 years (SD=8).

Disability. Disability was measured from participant reports about limitations to activities of daily living (ADLs) during interviews conducted at the 8th and 9th assessment waves. ADL information was collected using the Nagi, Katz, and Rosow-Breslau scales (Katz et al., 1970; Nagi, 1976; Rosow and Breslau, 1966). Scales are described in **Supplementary Table 6**. Incident disability was analyzed for participants free of disability at baseline (Nagi n=1,830; Katz n=1,826; Rosow-Breslau n=1,670). Over follow-up, n=123 participants developed new limitations in Nagi ADLs; n=181 developed new limitations in Katz ADLs; and n=224 developed new limitations in Rosow-Breslau ADLs.

The Environmental Risk Longitudinal Twin Study tracks the development of a birth cohort of 2,232 British participants. The sample was drawn from a larger birth register of twins born in England and Wales in 1994-1995. Full details about the sample are reported elsewhere (Moffitt and E-risk Team, 2002). Briefly, the E-Risk sample was constructed in 1999-2000, when 1,116 families (93% of those eligible) with same-sex 5-year-old twins participated in home-visit assessments. This sample comprised 56% monozygotic (MZ) and 44% dizygotic (DZ) twin pairs; sex was evenly distributed within zygosity (49% male). Families were recruited to represent the UK population of families with newborns in the 1990s, on the basis of residential location throughout England and Wales and mother's age. Teenaged mothers with twins were over-selected to replace high-risk families who were selectively lost to the register through non-response. Older mothers having twins via assisted reproduction were under-selected to avoid an excess of well-educated older mothers. The study sample represents the full range of socioeconomic conditions in the UK, as reflected in the families' distribution on a neighborhood-level socioeconomic index (ACORN [A Classification of Residential Neighborhoods], developed by CACI Inc. for commercial use): 25.6% of E-Risk families lived in "wealthy achiever" neighborhoods compared to 25.3% nationwide; 5.3% vs. 11.6% lived in "urban prosperity" neighborhoods; 29.6% vs. 26.9% lived in "comfortably off" neighborhoods; 13.4% vs. 13.9% lived in "moderate means" neighborhoods, and 26.1% vs. 20.7% lived in "hard-pressed" neighborhoods. E-Risk underrepresents "urban prosperity" neighborhoods because such households are likely to be childless.

Home-visits assessments took place when participants were aged 5, 7, 10, 12, and 18 years, when 93% of the participants took part. At ages 5, 7, 10, and 12 years, assessments were carried out with participants as well as their mothers (or primary caretakers); the home visit at age 18 included interviews only with participants. Each twin was assessed by a different interviewer. These data are supplemented by searches of official records and by questionnaires that are mailed, as developmentally appropriate, to teachers, and co-informants nominated by participants themselves. The Joint South London and Maudsley and the Institute of Psychiatry Research Ethics Committee approved each phase of the study. Parents gave informed consent and twins gave assent between 5-12 years and then informed consent at age 18. Study data may be accessed through the Study investigators (https://moffittcaspi.trinity.duke.edu/research-topics/erisk).

Childhood Socioeconomic Status (SES) was defined through a standardized composite of parental income, education, and occupation (Trzesniewski et al., 2006). The three SES indicators were highly correlated (r=0.57–0.67) and loaded significantly onto one factor. The population-wide distribution of the resulting factor was divided in tertiles for analyses.

Childhood Victimization. As previously described (Danese et al., 2016), we assessed exposure to six types of childhood victimization between birth to age 12: exposure to domestic violence between the mother and her partner, frequent bullying by peers, physical and sexual harm by an adult, and neglect.

DNA Methylation Data

DNA methylation was measured from Illumina 450k Arrays in GSE55763, NAS, Framingham, and E-Risk and from Illumina EPIC 850k Arrays in the Dunedin Study and Understanding Society. DNA was derived from whole blood samples in all studies. Dunedin Study blood draws were conducted at the cohort's age-38 assessment during 2010-12 for the prior DunedinPoAm measure, and at the cohort's age-45 assessment during 2017-2019 for the new DunedinPACE measure. Understanding Society blood draws were conducted in 2012. NAS blood draws were conducted during 1999-2013. Framingham blood draws were conducted during 2005-2008. E-Risk blood draws were conducted at the cohort's age-18 assessment during 2012-13. Dunedin methylation assays were run by the Molecular Genomics Shared Resource at Duke Molecular Physiology Institute, Duke University (USA). Understanding Society and E-Risk assays were run by the Complex Disease Epigenetics Group at the University of Exeter Medical School (UK) (www.epigenomicslab.com). NAS methylation assays were run by the Genome Research Core of the University of Illinois at Chicago. Framingham methylation assays were run by the University of Minnesota and John's Hopkins University (dbGaP phs000724.v9.p13). Processing protocols for the methylation data from all studies have been described previously (Dai et al., 2017; Hannon et al., 2018; Marzi et al., 2018; Panni et al., 2016).

Methylation Clocks. We computed the methylation clocks proposed by Levine et al. and Lu et al. (Levine et al., 2018; Lu et al., 2019) using the methylation data provided by the individual studies and the Horvath Lab's webtool (http://dnamage.genetics.ucla.edu/new).

The new Dunedin Pace of Aging methylation algorithm (DunedinPACE) was developed using elastic-net regression analysis carried out in the Dunedin Study, as described in detail in the Results. The criterion variable was 20-year Pace of Aging. Development of the 20-year Pace of Aging is described in detail elsewhere (Elliott et al., 2021). Briefly, we conducted mixed-effects growth modeling of longitudinal change in 19 biomarkers measuring integrity of the cardiovascular, metabolic, renal, hepatic, pulmonary, periodontal, and immune systems. Biomarkers were measured at the age 26, 32, 38, and 45 assessments. Biomarkers are listed in **Supplementary Table 1**. For each biomarker, we estimated random slopes quantifying each participant's own rate of change in that biomarker. We then composited, the 19 slopes across the biomarkers to calculate a participant's Pace of Aging. Pace of Aging was scaled in units representing the mean trend in the cohort, i.e. the average physiological change occurring during one calendar year (N=955, M=1, SD=0.29). Of the N=818 Dunedin Study members with methylation data at age 45, N=817 had measured Pace of Aging (M=0.99, SD=0.30). This group formed the analysis sample to develop DunedinPACE.

To compute DunedinPACE in the GSE55763, Understanding Society, NAS, Framingham, and E-Risk datasets, we applied the scoring algorithm derived from elastic net regression in the Dunedin Study. CpG weights for the scoring algorithm are provided in **Supplementary Table 2**. To apply the scoring algorithm, a panel of 20,000 probes that represent the underlying distribution of all probes included in the analysis to develop DunedinPACE are drawn from the target dataset. This new dataset is then quantile-normalized to match the reference distribution from the Dunedin Study. The set of 20,000 probes is composed of the 173 probes that make up the DunedinPACE algorithm and 19,827 additional probes selected at random from the full set of probes included in analysis to develop DunedinPACE (i.e. probes passing quality control checks, which are included on Illumina 450k and EPIC arrays, and which were reliable at an ICC threshold of 0.4 in the analysis by Sugden et al. (Sugden et al., 2020)).

Statistical Analysis

We conducted analysis of Dunedin, GSE55763, Understanding Society, NAS, Framingham, and E-Risk, data using regression models. We analyzed continuous outcome data using linear regression. We analyzed count outcome data using Poisson and negative binomial

regression. We analyzed time-to-event outcome data using Cox proportional hazard regression. To estimate intraclass correlation coefficients for technical replicates in GSE55763, we used mixed-effects regression models. For analysis of repeated-measures longitudinal DNA methylation data in the NAS, we used generalized estimating equations to account for non-independence of repeated observations of individuals (Ballinger, 2004), following the method in previous analysis of those data (Gao et al., 2018), and econometric fixed-effects regression (Wooldridge, 2012) to test within-person change over time. For analysis in E-Risk, which include data on twin siblings, we clustered standard errors at the family level to account for non-independence of data. For regression analysis, methylation measures were adjusted for batch effects by regressing the measure on batch controls and predicting residual values. Dunedin Study, Understanding Society, Framingham, and E-Risk analyses included covariate adjustment for sex (the NAS included only men). Understanding Society, NAS, and Framingham analyses included covariate adjustment for chronological age. (Dunedin and E-Risk are birth-cohort studies and participants are all the same chronological age.) Sensitivity analyses testing covariate adjustment for leukocyte distributions and smoking are reported in **Supplementary Table 4.** Leukocyte distributions were measured from complete blood count data in the Dunedin Cohort and were estimated from DNA methylation data for other cohorts.

Code for analysis. Code used for analysis and to prepare figures will be accessible via GitHub upon acceptance of the article for publication.

Code to calculate DunedinPACE from Illumina 450k or Epic Array Data. R code to implement DunedinPACE will be available via GitHub upon publication of the article.

REFERENCES

Arbeev, K.G., Ukraintseva, S.V., and Yashin, A.I. (2016). Dynamics of biomarkers in relation to aging and mortality. Mech. Ageing Dev. *156*, 42–54.

Ballinger, G.A. (2004). Using generalized estimating equations for longitudinal data analysis. Organ. Res. Methods *7*, 127–150.

Barzilai, N., Cuervo, A.M., and Austad, S. (2018). Aging as a Biological Target for Prevention and Therapy. JAMA *320*, 1321–1322.

Bell, B., Rose, C.L., and Damon, A. (1972). The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. Aging Hum. Dev. *3*, 5–17.

Bellantuono, I. (2018). Find drugs that delay many diseases of old age. Nature *554*, 293–295.

Belsky, D.W., Caspi, A., Houts, R., Cohen, H.J., Corcoran, D.L., Danese, A., Harrington, H., Israel, S., Levine, M.E., Schaefer, J.D., et al. (2015). Quantification of biological aging in young adults. Proc. Natl. Acad. Sci. U. S. A. *112*, E4104-4110.

Belsky, D.W., Huffman, K.M., Pieper, C.F., Shalev, I., and Kraus, W.E. (2017). Change in the Rate of Biological Aging in Response to Caloric Restriction: CALERIE Biobank Analysis. J. Gerontol. Ser. A *73*, 4–10.

Belsky, D.W., Caspi, A., Arseneault, L., Baccarelli, A., Corcoran, D.L., Gao, X., Hannon, E., Harrington, H.L., Rasmussen, L.J., Houts, R., et al. (2020). Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. ELife *9*, e54870.

Brandmaier, A.M., von Oertzen, T., Ghisletta, P., Lindenberger, U., and Hertzog, C. (2018). Precision, Reliability, and Effect Size of Slope Variance in Latent Growth Curve Models: Implications for Statistical Power Analysis. Front. Psychol. *9*.

Campisi, J., Kapahi, P., Lithgow, G.J., Melov, S., Newman, J.C., and Verdin, E. (2019). From discoveries in ageing research to therapeutics for healthy ageing. Nature *571*, 183.

Cohen, A.A., Milot, E., Yong, J., Seplaki, C.L., Fülöp, T., Bandeen-Roche, K., and Fried, L.P. (2013). A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. Mech. Ageing Dev. *134*, 110–117.

Crimmins, E.M., Thyagarajan, B., Levine, M.E., Weir, D.R., and Faul, J. (2021). Associations of Age, Sex, Race/Ethnicity, and Education With 13 Epigenetic Clocks in a Nationally Representative U.S. Sample: The Health and Retirement Study. J. Gerontol. Ser. A *76*, 1117–1123.

Dai, L., Mehta, A., Mordukhovich, I., Just, A.C., Shen, J., Hou, L., Koutrakis, P., Sparrow, D., Vokonas, P.S., Baccarelli, A.A., et al. (2017). Differential DNA methylation and PM2.5 species in a 450K epigenome-wide association study. Epigenetics *12*, 139–148.

Danese, A., Moffitt, T.E., Arseneault, L., Bleiberg, B.A., Dinardo, P.B., Gandelman, S.B., Houts, R., Ambler, A., Fisher, H.L., and Poulton, R. (2016). The origins of cognitive deficits in victimized children: implications for neuroscientists and clinicians. Am. J. Psychiatry *174*, 349–361.

Deelen, J., Kettunen, J., Fischer, K., van der Spek, A., Trompet, S., Kastenmüller, G., Boyd, A., Zierer, J., van den Akker, E.B., Ala-Korpela, M., et al. (2019). A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nat. Commun. *10*, 3346.

Eiriksdottir, T., Ardal, S., Jonsson, B.A., Lund, S.H., Ivarsdottir, E.V., Norland, K., Ferkingstad, E., Stefansson, H., Jonsdottir, I., Holm, H., et al. (2021). Predicting the probability of death using proteomics. Commun. Biol. *4*, 1–11.

Elliott, M.L., Belsky, D.W., Knodt, A.R., Ireland, D., Melzer, T.R., Poulton, R., Ramrakha, S., Caspi, A., Moffitt, T.E., and Hariri, A.R. (2019). Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. Mol. Psychiatry 1–10.

Elliott, M.L., Caspi, A., Houts, R.M., Ambler, A., Broadbent, J.M., Hancox, R.J., Harrington, H., Hogan, S., Keenan, R., Knodt, A., et al. (2021). Disparities in the pace of biological aging among midlife adults of the same chronological age have implications for future frailty risk and policy. Nat. Aging *1*, 295–308.

Finch, C.E., and Crimmins, E.M. (2016). Constant molecular aging rates vs. the exponential acceleration of mortality. Proc. Natl. Acad. Sci. *113*, 1121–1123.

Fisher, H.L., Caspi, A., Moffitt, T.E., Wertz, J., Gray, R., Newbury, J., Ambler, A., Zavos, H., Danese, A., Mill, J., et al. (2015). Measuring adolescents' exposure to victimization: The Environmental Risk (E-Risk) Longitudinal Twin Study. Dev. Psychopathol. *27*, 1399–1416.

Gao, X., Colicino, E., Shen, J., Just, A.C., Nwanaji-Enwerem, J.C., Coull, B., Lin, X., Vokonas, P., Zheng, Y., and Hou, L. (2018). Accelerated DNA methylation age and the use of antihypertensive medication among older adults. Aging *10*, 3210.

Gao, X., Colicino, E., Shen, J., Just, A.C., Nwanaji-Enwerem, J.C., Wang, C., Coull, B., Lin, X., Vokonas, P., Zheng, Y., et al. (2019). Comparative validation of an epigenetic mortality risk score with three aging biomarkers for predicting mortality risks among older adult males. Int. J. Epidemiol.

Gao, X., Coull, B., Lin, X., Vokonas, P., Spiro, A., Hou, L., Schwartz, J., and Baccarelli, A.A. (2021). Short-term air pollution, cognitive performance and nonsteroidal anti-inflammatory drug use in the Veterans Affairs Normative Aging Study. Nat. Aging 1–8.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. Philos. Trans. R. Soc. Lond. 513–583.

Graf, G.H., Crowe, C.L., Kothari, M., Kwon, D., Manly, J.J., Turney, I.C., Valeri, L., and Belsky, D.W. (In Press). Testing Black-White disparities in biological aging in older adults in the United States: Analysis of DNA methylation and blood chemistry methods. Am. J. Epidemiol. 2021.03.02.21252685.

Hägg, S., Belsky, D.W., and Cohen, A.A. (2019). Developments in molecular epidemiology of aging. Emerg. Top. Life Sci. ETLS20180173.

Hannon, E., Gorrie-Stone, T.J., Smart, M.C., Burrage, J., Hughes, A., Bao, Y., Kumari, M., Schalkwyk, L.C., and Mill, J. (2018). Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. Am. J. Hum. Genet. *103*, 654–665.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., et al. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell *49*, 359–367.

Hastings, W.J., Shalev, I., and Belsky, D.W. (2019). Comparability of biological aging measures in the National Health and Nutrition Examination Study, 1999–2002. Psychoneuroendocrinology *106*, 171–178.

Higgins-Chen, A.T., Thrush, K.L., Wang, Y., Kuo, P.-L., Wang, M., Minteer, C.J., Moore, A.Z., Bandinelli, S., Vinkers, C.H., Vermetten, E., et al. (2021). A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. BioRxiv 2021.04.16.440205.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol. *14*, R115.

Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. *19*, 371–384.

Jansen, R., Han, L.K., Verhoeven, J.E., Aberg, K.A., van den Oord, E.C., Milaneschi, Y., and Penninx, B.W. (2021). An integrative study of five biological clocks in somatic and mental health. ELife *10*, e59479.

Justice, J.N., Ferrucci, L., Newman, A.B., Aroda, V.R., Bahnson, J.L., Divers, J., Espeland, M.A., Marcovina, S., Pollak, M.N., Kritchevsky, S.B., et al. (2018). A framework for selection of blood-based biomarkers for geroscience-guided clinical trials: report from the TAME Biomarkers Workgroup. GeroScience *40*, 419–436.

Kaeberlein, M. (2013). Longevity and aging. F1000Prime Rep. *5*.

Katz, S., Downs, T.D., Cash, H.R., and Grotz, R.C. (1970). Progress in development of the index of ADL. The Gerontologist *10*, 20–30.

Kennedy, B.K., Berger, S.L., Brunet, A., Campisi, J., Cuervo, A.M., Epel, E.S., Franceschi, C., Lithgow, G.J., Morimoto, R.I., Pessin, J.E., et al. (2014). Geroscience: linking aging to chronic disease. Cell *159*, 709–713.

Kirkwood, T.B.L. (2015). Deciphering death: a commentary on Gompertz (1825) 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies.' Philos. Trans. R. Soc. B Biol. Sci. *370*.

Klemera, P., and Doubal, S. (2006). A new approach to the concept and computation of biological age. Mech. Ageing Dev. *127*, 240–248.

Kwon, D., and Belsky, D.W. (2021). A toolkit for quantification of biological age from blood chemistry and organ function test data: BioAge. GeroScience *43*, 2795–2808.

Lehallier, B., Gate, D., Schaum, N., Nanasi, T., Lee, S.E., Yousef, H., Moran Losada, P., Berdnik, D., Keller, A., Verghese, J., et al. (2019). Undulating changes in human plasma proteome profiles across the lifespan. Nat. Med. *25*, 1843–1850.

Lehne, B., Drong, A.W., Loh, M., Zhang, W., Scott, W.R., Tan, S.-T., Afzal, U., Scott, J., Jarvelin, M.-R., Elliott, P., et al. (2015). A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol. *16*, 37.

Lepeule, J., Litonjua, A.A., Gasparrini, A., Koutrakis, P., Sparrow, D., Vokonas, P.S., and Schwartz, J. (2018). Lung function association with outdoor temperature and relative humidity and its interaction with air pollution in the elderly. Environ. Res. *165*, 110–117.

Levine, M.E. (2013). Modeling the rate of senescence: Can estimated biological age predict mortality more accurately than chronological age? J. Gerontol. A. Biol. Sci. Med. Sci. *68*, 667–674.

Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y., et al. (2018). An epigenetic biomarker of aging for lifespan and healthspan. Aging *10*, 573–591.

Lu, A.T., Quach, A., Wilson, J.G., Reiner, A.P., Aviv, A., Raj, K., Hou, L., Baccarelli, A.A., Li, Y., Stewart, J.D., et al. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging *11*, 303–327.

Marzi, S.J., Sugden, K., Arseneault, L., Belsky, D.W., Burrage, J., Corcoran, D.L., Danese, A., Fisher, H.L., Hannon, E., Moffitt, T.E., et al. (2018). Analysis of DNA Methylation in Young People: Limited Evidence for an Association Between Victimization Stress and Epigenetic Variation in Blood. Am. J. Psychiatry appi.ajp.2017.17060693.

Moffitt, T.E. (2020). Behavioral and Social Research to Accelerate the Geroscience Translation Agenda. Ageing Res. Rev. *63*, 101146.

Moffitt, T.E. and E-risk Team (2002). Teen-aged mothers in contemporary Britain. J. Child Psychol. Psychiatry *43*, 727–742.

Moffitt, T.E., Belsky, D.W., Danese, A., Poulton, R., and Caspi, A. (2017). The Longitudinal Study of Aging in Human Young Adults: Knowledge Gaps and Research Agenda. J. Gerontol. Ser. A *72*, 210–215.

Nagi, S.Z. (1976). An epidemiology of disability among adults in the United States. Milbank Mem. Fund Q. Health Soc. 439–467.

Olshansky, S.J., and Carnes, B.A. (1997). Ever since gompertz. Demography *34*, 1–15.

Panni, T., Mehta Amar J., Schwartz Joel D., Baccarelli Andrea A., Just Allan C., Wolf Kathrin, Wahl Simone, Cyrys Josef, Kunze Sonja, Strauch Konstantin, et al. (2016). Genome-Wide Analysis of DNA Methylation and Fine Particulate Matter Air Pollution in Three Study Populations: KORA F3, KORA F4, and the Normative Aging Study. Environ. Health Perspect. *124*, 983–990.

Parker, D.C., Bartlett, B.N., Cohen, H.J., Fillenbaum, G., Huebner, J.L., Kraus, V.B., Pieper, C., and Belsky, D.W. (2019). Association of Blood Chemistry Quantifications of Biological Aging With Disability and Mortality in Older Adults. J. Gerontol. Ser. A *glz219*.

Poulton, R., Moffitt, T.E., and Silva, P.A. (2015). The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. Soc. Psychiatry Psychiatr. Epidemiol. *50*, 679–693.

Prentice, R.L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. Stat. Med. *8*, 431–440.

Raffington, L., Belsky, D.W., Kothari, M., Malanchini, M., Tucker-Drob, E.M., and Harden, K.P. (2021). Socioeconomic Disadvantage and the Pace of Biological Aging in Children. Pediatrics.

Rasmussen, L.J.H., Caspi, A., Ambler, A., Broadbent, J.M., Cohen, H.J., d'Arbeloff, T., Elliott, M., Hancox, R.J., Harrington, H., Hogan, S., et al. (2019). Association of Neurocognitive and Physical Function With Gait Speed in Midlife. JAMA Netw. Open *2*, e1913123–e1913123.

Rast, P., and Hofer, S.M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. Psychol. Methods *19*, 133–154.

Rosow, I., and Breslau, N. (1966). A Guttman health scale for the aged. J. Gerontol.

Schmitz, L.L., Zhao, W., Ratliff, S.M., Goodwin, J., Miao, J., Lu, Q., Guo, X., Taylor, K.D., Ding, J., Liu, Y., et al. (2021). The Socioeconomic Gradient in Epigenetic Aging Clocks: Evidence from the Multi-Ethnic Study of Atherosclerosis and the Health and Retirement Study. MedRxiv 2021.03.01.21252660.

Sierra, F., Caspi, A., Fortinsky, R.H., Haynes, L., Lithgow, G.J., Moffitt, T.E., Olshansky, S.J., Perry, D., Verdin, E., and Kuchel, G.A. (2021). Moving geroscience from the bench to clinical care and health policy. J. Am. Geriatr. Soc. *n/a*.

Sugden, K., Hannon, E.J., Arseneault, L., Belsky, D.W., Corcoran, D.L., Fisher, H.L., Houts, R.M., Kandaswamy, R., Moffitt, T.E., Poulton, R., et al. (2020). Patterns of Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement. Patterns *1*, 100014.

Tanaka, T., Basisty, N., Fantoni, G., Candia, J., Moore, A.Z., Biancotto, A., Schilling, B., Bandinelli, S., and Ferrucci, L. (2020). Plasma proteomic biomarker signature of age predicts health and life span. ELife *9*, e61073.

Trzesniewski, K.H., Moffitt, T.E., Caspi, A., Taylor, A., and Maughan, B. (2006). Revisiting the association between reading achievement and antisocial behavior: New evidence of an environmental explanation from a twin study. Child Dev. *77*, 72–88.

Tsao, C.W., and Vasan, R.S. (2015). Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int. J. Epidemiol. *44*, 1800–1813.

University of Essex Institute for Social and Economic Research, NatCen Social Research, and University of Exeter Medical School. (2017). Understanding Society: DNA methylation data based on Illumina methylation EPIC array.

University of Essex Institute for Social and Economic Research and National Centre for Social Research, Understanding Society: Waves 2 and 3 Nurse Health Assessment, 2010- 2012 (UK Data Service).

Wooldridge, J.M. (2012). Introductory Econometrics: A Modern Approach, 5th ed. (Cengage Learning).

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. *67*, 301–320.

Figure 1. DunedinPACE correlation with 20-year Pace of Aging and association with decline in physical and cognitive functions and subjective signs of aging. Panel A shows the association of DunedinPACE with the 20-year Pace of Aging in the Dunedin Study cohort (r=0.78). Panel B shows effect-sizes for associations of DunedinPACE with measures of physical and cognitive functioning and subjective signs of aging in Dunedin Study members at age 45 years. Colors indicate groupings of measures (physical functions in shades of orange, cognitive functions in shades of yellow, subjective signs of aging in shades of green). Stars next to labels for Physical Limitations and Facial Aging indicate that these meaures were reverse-coded for this analysis. Panel C shows binned scatterplots of associations of DunedinPACE with declines in physical and cognitive function and subjective signs of aging. Plotted points show average values for "bins" of approximately 20 Study members. Regression slopes are estimated from the raw, un-binned data. In Panel C, y-axis scales are denominated in units of standard deviations computed from the baseline measurement for all outcomes except self-rated health, for which the y-axis shows probability of incident fair/poor self-rated health. Changes were calculated over the interval between age-38 and age-45 assessments for all measures except cognition, for which change was calculated over the interval between the age-13 and age-45 assessments. DunedinPACE and DunedinPoAm were developed from analysis of Pace of Aging in the Dunedin cohort; effect-sizes for these measures may be over-estimated.

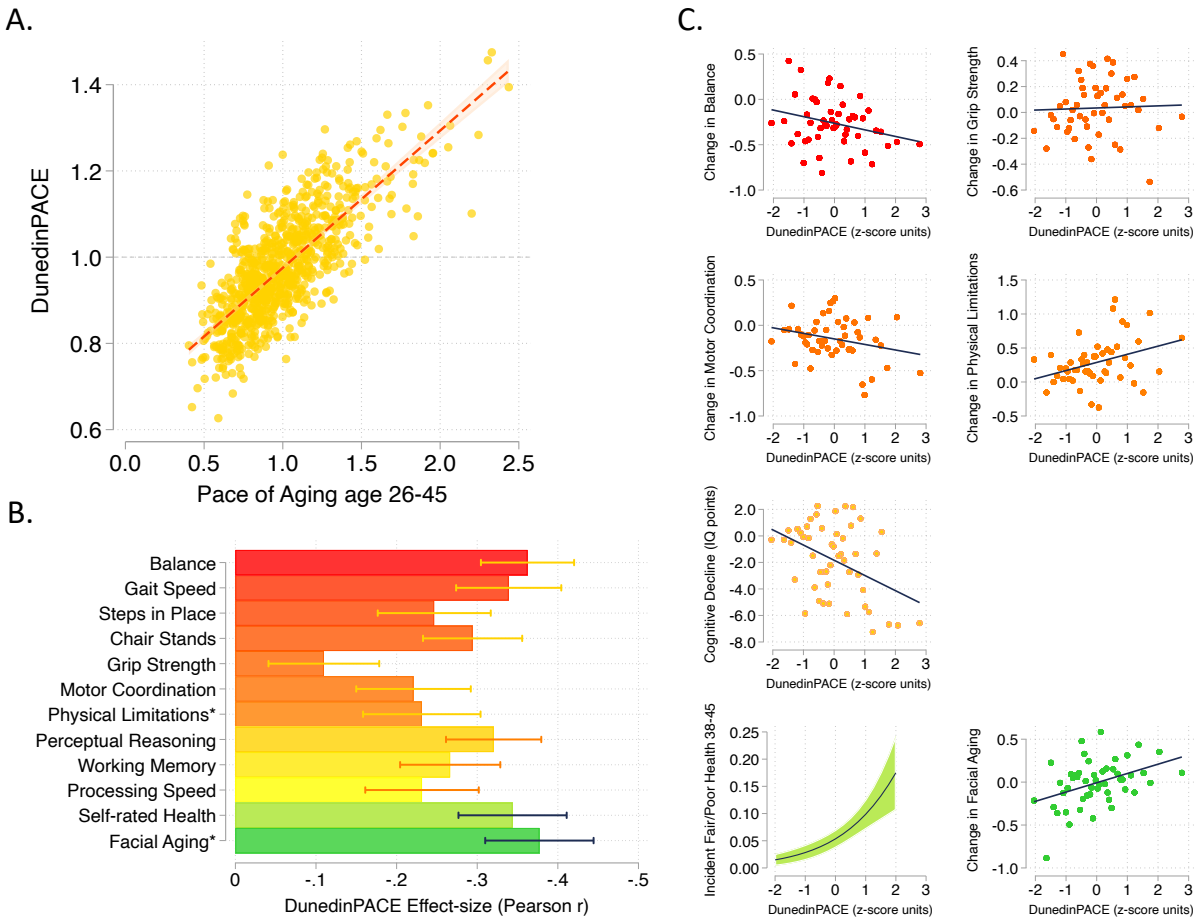


Figure 2. Test-retest reliability of DunedinPACE. The figure graphs DunedinPACE values for replicate Illumnina450k datasets for n=36 individuals in the dataset published by Lehne and colleagues (Lehne et al., 2015) (GEO Accession GSE55763). The ICC for DunedinPACE in the Lehne dataset is 0.96 95% CI [0.92-0.98].

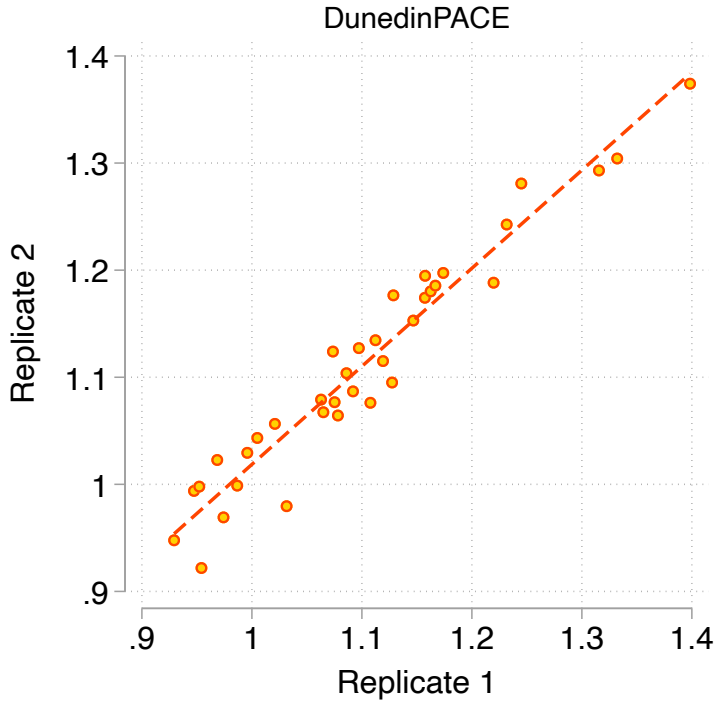


Figure 3. Associations of DunedinPACE with chronological age, epigenetic clocks, Physiology-based measures of biological age, and self-rated health in Understanding Society. Panel A shows a scatterplot and fitted slopes illustrating the association between chronological age (x-axis) and DunedinPACE (y-axis) in women and men in the Understanding Society sample. Data for women are plotted with yellow dots (orange slope) and for men with blue crosses (navy slope). The figure illustrates a positive association between chronological age and DunedinPACE (Pearson r = 0.32). Panel B shows a matrix of correlations and association plots among DunedinPACE and age-acceleration residuals of Horvath, Hannum, Levine-PhenoAge and Lu-GrimAge epigenetic clocks. The diagonal cells of the matrix list the DNA methylation measures. The half of the matrix below the diagonal shows scatter plots of associations. For each scatter-plot cell, the y-axis corresponds to the variable named along the matrix diagonal to the right of the plot and the x-axis corresponds to the variable named along the matrix diagonal above the plot. The half of the matrix above the diagonal lists Pearson correlations between the DNA methylation measures. For each correlation cell, the value reflects the correlation of the variables named along the matrix diagonal to the left of the cell and below the cell. Panel C graphs scatterplots of associations of DunedinPACE with three physiology-based measures of biological age (KDM Biological Age Advancement, r=0.30 [0.24-0.36]; Phenotypic Age Advancement, r=0.32 95% CI [0.26-0.38]; and Homeostatic Dysregulation r=0.09 [0.03-0.16]) and a plot of DunedinPACE means and 95% confidence intervals by self-rated health category (r=0.20 [0.15-0.26]).

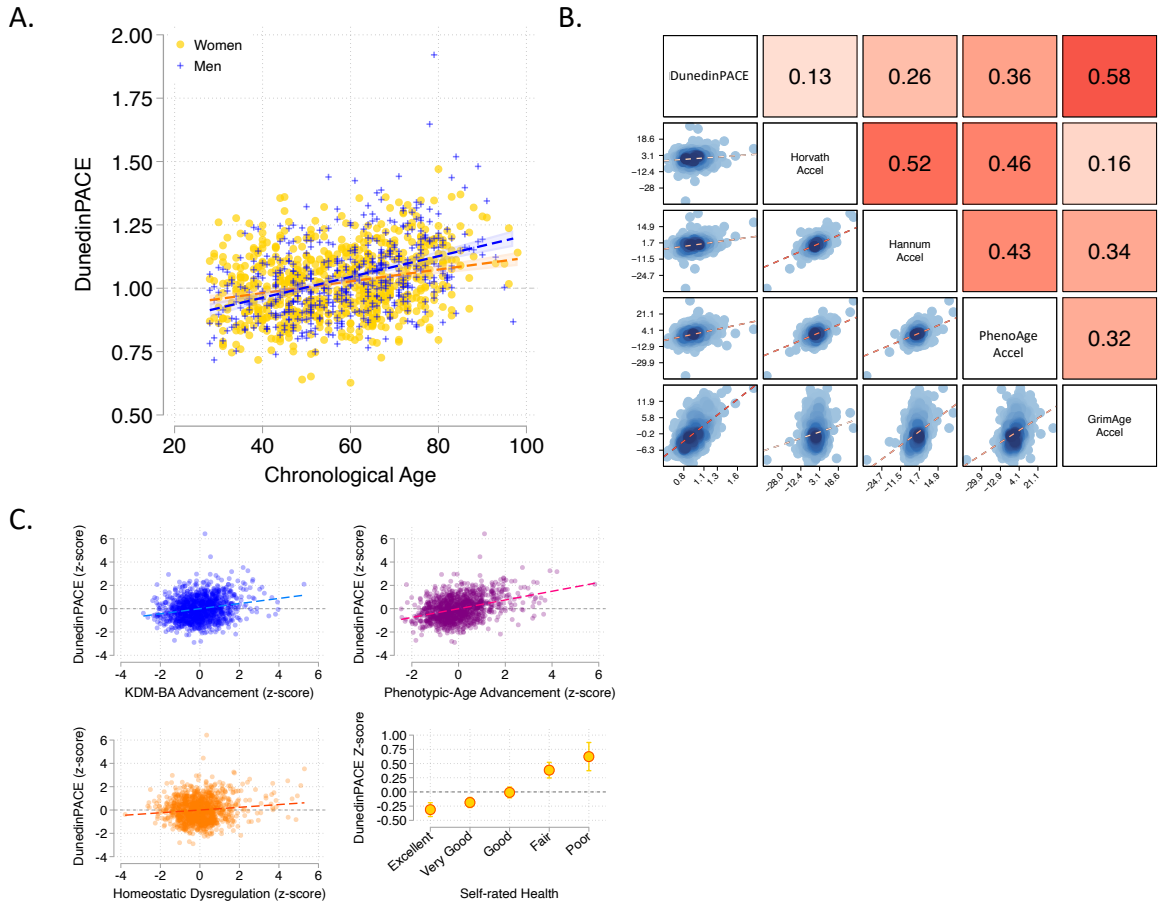
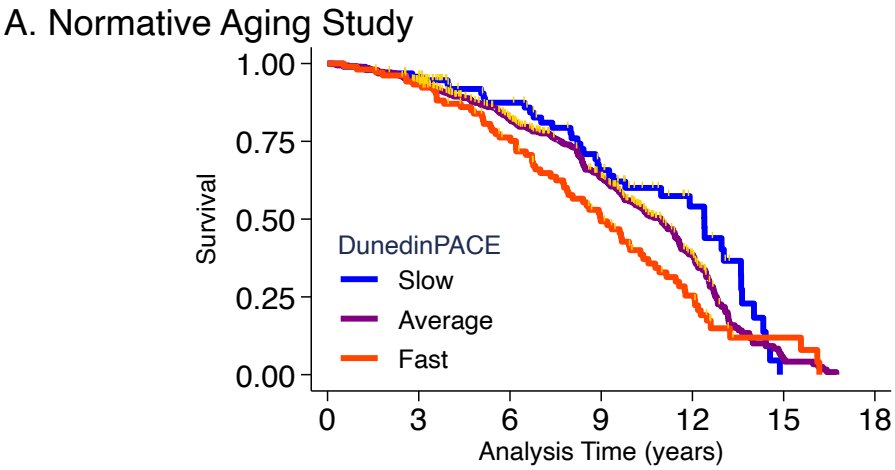
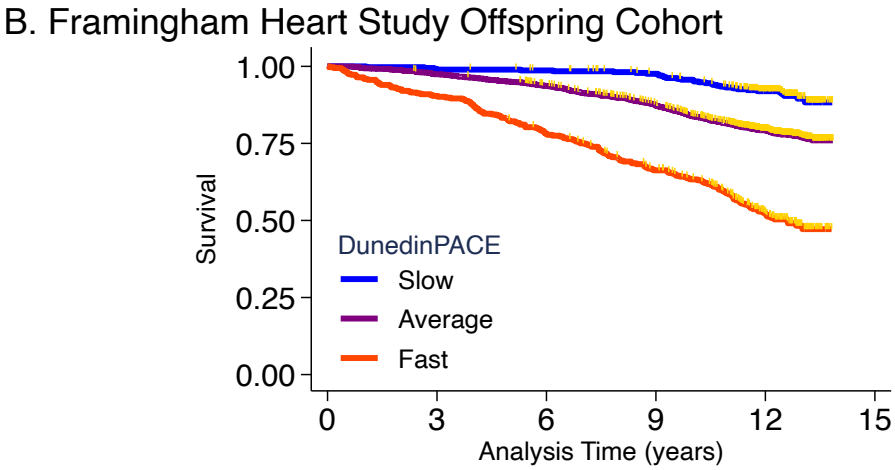


Figure 4. Association of DunedinPACE with mortality. Panel A shows mortality in the Normative Aging Study (NAS). Panel B shows mortality in the Framingham Heart Study Offspring Cohort. The figure plots Kaplan-Meier curves for three groups of participants in each of the two cohorts: those with DunedinPACE 1 SD or more below the mean ('slow' DunedinPACE, blue line); those with DunedinPACE within 1 SD of the mean ('average' DunedinPACE, purple line); and those with DunedinPACE 1 SD or more above the mean ('fast' DunedinPACE, red line). Censoring of participants prior to death is indicated with a gold hash marks. The table below the figure details the number of participants at risk per 3-year interval and, in parentheses, the number who died during the interval.



Number At Risk (Deaths)

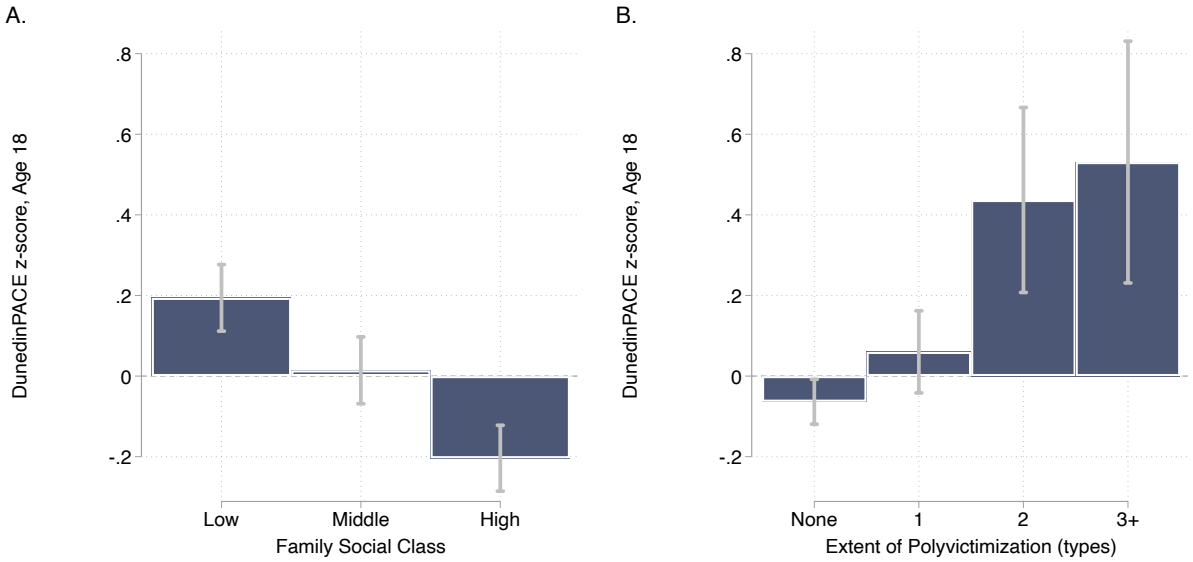
Slow DunedinPACE	95	(4)	87	(6)	59	(13)	38	(5)	16	(13)	0	(0)	0
Average DunedinPACE	496	(31)	440	(45)	295	(60)	187	(55)	66	(43)	6	(6)	0
Fast DunedinPACE	104	(7)	95	(16)	68	(23)	40	(17)	17	(7)	3	(3)	0



Number At Risk (Deaths)

Slow DunedinPACE	378	(4)	374	(1)	369	(4)	355	(20)	251	(5)	0
Average DunedinPACE	1724	(44)	1678	(65)	1597	(106)	1452	(129)	946	(23)	0
Fast DunedinPACE	365	(35)	330	(45)	283	(42)	231	(44)	127	(8)	0

Figure 5. DunedinPACE levels by strata of childhood socioeconomic status (SES) and victimization in the E-Risk Study. Panel A (left side) plots means and 95% CIs for DunedinPACE measured at age 18 among E-Risk participants who grew up low, middle, and high socioeconomic status households. Panel B (right side) plots means and 95% CIs for DunedinPACE measured at age 18 among E-Risk participants who experienced 0, 1, 2, or three or more types of victimization through age 12 years.



Supplementary Results

Correlation with criterion Pace of Aging and test-retest reliability for DunedinPACE, original DunedinPoAm, and alternative versions of these algorithms developed using different sets of CpG sites. DunedinPACE and DunedinPoAm were both developed from machine-learning analysis fitting DNA methylation data to Pace of Aging measures derived from longitudinal analysis of organ-function and blood chemistry data within the Dunedin Study.

There are two key differences between the measures. First, the Pace of Aging measure modeled to develop DunedinPoAm included data from only three repeated measures spanning 12 years of follow-up. In contrast, the Pace of Aging measure modeled to develop DunedinPACE included data from four repeated measurements collected over 20 years of follow-up. Second, DunedinPoAm was developed from a DNA methylation dataset including all probes included on both the Illumina450k and EPIC arrays. In contrast, the DNA methylation dataset used to develop DunedinPACE was restricted to a subset of CpG sites meeting a test-retest reliability threshold.

These differences should have two consequences for DunedinPACE. First, the inclusion of a 4th timepoint in Pace of Aging analysis should increase precision of measurement and result in improved fit of DNA methylation data to the 4-timepoint Pace of Aging criterion as compared to the 3-timepoint version. (A less precise Pace of Aging will be harder for the elastic net machine learning analysis to model since more of the variation will be random noise.) Second, the restriction of machine learning to more reliable CpG sites should result in improved test-retest reliability.

To further test these improvements, we conducted a sensitivity analysis. We repeated the machine-learning analysis to construct DunedinPACE using the unrestricted set of CpG sites included in the original DunedinPoAm analysis. We refer to the resulting measure as "All-probes DunedinPACE". In parallel, we repeated the machine-learning analysis to construct DunedinPoAm using the subset of more-reliable CpG sites included in the original DunedinPACE analysis. We refer to the resulting measure as "Reliable Probes DunedinPoAm". We then repeated our analyses testing correlation of the DNA methylation measures with the criterion Pace of Aging and our analysis of test-retest reliability.

Consistent with the expectation that a 4th timepoint would increase precision of Pace of Aging measurement and result in improved criterion fit, both versions of DunedinPACE showed stronger correlation with 4-timepoint Pace of Aging than did the versions of DunedinPoAm with 3-timepoint Pace of Aging. For algorithms trained on all probes, r=0.87 DunedinPACE vs. r=0.58 for original DunedinPoAm. For algorithms trained on reliable probes, r=0.77 for DunedinPACE vs. r=0.65 for DunedinPoAm.

Consistent with the expectation that training the algorithm using more reliable probes would improve test-retest reliability, both versions of the reliable-probes-trained scores showed superior test-retest reliability to the all-probes trained scores. For DunedinPACE, reliable-probes ICC=0.96 vs. 0.72 for all-probes. For DunedinPoAm, reliable-probes ICC=0.90 vs. 0.89 for all probes.

Supplementary Table 1. Measurements included in modeling Pace of Aging in the Dunedin Study Birth Cohort. We measured Pace of Aging from repeated assessments of a panel of 19 biomarkers: Body mass index (BMI), Waist-hip ratio, Glycated hemoglobin, Leptin, Blood pressure (mean arterial pressure), Cardiorespiratory fitness (VO_2Max), Forced vital capacity ratio (FEV_1/FVC), Forced expiratory volume in one second (FEV_1), Total cholesterol, Triglycerides, High density lipoprotein (HDL), Lipoprotein(a), Apolipoprotein B100/A1 ratio, estimated Glomerular Filtration Rate (eGFR), Blood Urea Nitrogen (BUN), High Sensitivity C-reactive Protein (hs-CRP), White blood cell count, mean periodontal attachment loss (AL), and the number of dental-caries-affected tooth surfaces (tooth decay). This list includes two new biomarkers, leptin and caries, not included in the original Pace of Aging (Belsky et al., 2015), both of which have now been assessed at multiple waves, allowing growth curve modeling. Telomere length was dropped because of an emerging and yet-unresolved field-wide debate about its measurement (Nettle et al., 2021). All other biomarkers were the same. The 20-year Pace of Aging measure is described in detail elsewhere (Elliott et al., 2021).

Body mass index	Height was measured to the nearest millimeter using a Seca 264 Wireless Stadiometer. Weight was measured to the nearest 0.1 kg using calibrated scales. Individuals were weighed in light clothing. Body mass index (BMI) was calculated using the standard formula weight in kilograms divided by height in meters squared.
Waist-hip ratio	Waist girth was the perimeter at the level of the noticeable waist narrowing located between the costal border and the iliac crest. Hip girth was taken as the perimeter at the level of the greatest protuberance and at about the symphysion pubic level anteriorly. Measurements were repeated and the average used to calculate waist-hip ratio.
Glycated hemoglobin (HbA1C)	Whole blood glycated hemoglobin concentration (expressed as a percentage of total hemoglobin) was measured by ion exchange high performance liquid chromatography (BioRad D-100, Hercules, Calif.), a method certified by the US National Glycohemoglobin Standardization Program (http://www.ngsp.org/).
Leptin	Serum leptin (μg/L) was measured using Human Leptin RIA kit (Cat# HL-81K, Linco Research, Missouri, USA) (Ages 32 & 38) and the Quantikine ELISA Human Leptin Immunoassay (Cat# SLP00, R&D Systems Inc, Minneapolis, MN) (Age 45) according to the manufacturer's instructions.
Blood pressure (mean arterial pressure)	Systolic and diastolic blood pressure were assessed according to standard protocols with a Hawksley random-zero sphygmomanometer with a constant deflation valve (age 32 & 38) and with a BpTRU™ Vital Signs Monitor BPM 200 (age 45). Mean arterial pressure (MAP) was calculated using the formula Diastolic Pressure+1/3(Systolic Pressure - Diastolic Pressure).
Cardiorespiratory fitness (VO_2Max)	Cardiorespiratory fitness was assessed by measuring heart rate in response to a submaximal exercise test on a friction-braked cycle

	ergometer. Dependent on the extent to which heart rate increased during a 2-min 50 W warm-up, the workload was adjusted to elicit a steady heart-rate in the range 130–170 beats per minute. After a further 6-min constant power output stage, the maximum heart rate was recorded and used to calculate predicted maximum oxygen uptake adjusted for body weight in milliliters per minute per kilogram (VO_2max) according to standard protocols.
Lung function (FEV_1 and FEV_1/FVC)	We calculated post-albuterol forced expiratory volume in one second (FEV_1) and the ratio of FEV_1 to forced vital capacity (FVC; FEV_1/FVC) using measurements from spirometry conducted with a Sensormedics body plethysmograph (Sensormedics Corporation, Yorba Linda, CA, USA).
Total cholesterol, triglycerides, and high-density lipoprotein (HDL) cholesterol	Serum non-fasting total cholesterol, triglycerides, and high-density lipoprotein (HDL) cholesterol levels (mmol/L) were measured by colorimetric assay on a Hitachi 917 analyzer (ages 26-32), a Modular P analyzer (age 38), and a Cobas c702 analyzer (age 45).
Lipoprotein(a)	Serum lipoprotein(a) (nmol/L) was measured by a particle-enhanced immunoturbidimetric assay on a Hitachi 917 analyzer (ages 26-32), a Modular P analyzer (age 38), and a Cobas c502 analyzer (age 45).
Apolipoprotein B100/A1 ratio	Serum apolipoprotein A1 (g/L) and apolipoprotein B100 (g/L) were measured by immunoturbidimetric assay on a Hitachi 917 analyzer (ages 26-32), a Modular P analyzer (age 38), and a Cobas c502 (age 45), and the ratio between the two was calculated.
eGFR (estimated glomerular filtration rate)	Serum creatinine (umol/L) was measured by kinetic colorimetric assay on a Hitachi 917 analyzer (age 32), Modular P analyzer (age 38), and Cobas c702 (age 45) (Roche Diagnostics, Mannheim, Germany). eGFR was estimated utilizing the CKD-Epi formula calculated from serum creatinine.
Blood urea nitrogen (BUN)	Blood urea nitrogen (mmol/L) was measured by kinetic UV assay at ages 26 (Hitachi 917 analyzer) and 45 (Cobas c702 analyzer), and by kinetic colorimetric assay at ages 32 (Hitachi 917 analyzer) and 38 (Modular P analyzer), , and Cobas c702 analyzer at age 45.
High sensitivity C-reactive protein (hs-CRP)	Serum C-reactive protein (mg/L) was measured by high sensitivity immunoturbidimetric assay on a Hitachi 917 analyzer (age 32), a Modular P analyzer (age 38), and a Cobas c702 (age 45). Values were log-transformed for analysis.
White blood cell count	Whole blood white blood cell counts ($x10^9$/L) were measured by flow cytometry with a Coulter STKS (Coulter Corporation, Miami, FL) (age 26), a Sysmex XE2100 (Sysmex Corporation, Japan) (age 32), and a Sysmex XE5000 (Sysmex Corporation, Japan) (ages 38 and 45). Counts were log-transformed for analysis.
Mean periodontal attachment loss (AL)	Calibrated dentists used a PCP-2 periodontal probe (Hu-Friedy; Chicago) to measure gingival recession (the distance from the cementoenamel junction to the gingival margin) and probing depth (the distance from the probe tip to the gingival margin) in millimeters

	at three sites (mesiobuccal, buccal, and distolingual) per tooth, excluding third molar teeth. Periodontal attachment loss for each site was computed by summing gingival recession and probing depth. and then averaged across all periodontally examined teeth. Periodontal examinations were conducted with half-mouth examinations at age 26 and full-mouth examinations at ages 32, 38, and 45 years.
Caries-affected tooth surfaces	Calibrated dentists examined the teeth for caries and restorations following the World Health Organization Oral Health Surveys methodology. Four surfaces were considered for anterior teeth (canines and incisors): buccal, lingual, distal, and mesial; a fifth surface, occlusal, was considered for premolar and molar teeth. Tooth surfaces were classified as having untreated caries (DS) if a cavitated carious lesion was present, as filled (FS) if a dental restoration was present (including crowns), and missing due to caries (MS) if the participant indicated that a given tooth had been removed due to decay or failed dental restorative work. DS, MS, and FS counts were summed to obtain a DMFS score (ranging from 0 to 148 surfaces). Caries experience was expressed as the % of tooth surfaces that had been caries-affected, excluding surfaces of teeth that were unerupted, lost due to trauma, extracted for reasons other than caries (impaction, orthodontic treatment, or periodontal disease), or could not be visualized by the examiner.

Supplementary Table 2. CpG sites included in the DunedinPACE DNA methylation algorithm.

We will post Supplemental Table 2 once the article is accepted for publication.

Supplementary Table 3. Physical and cognitive functioning and subjective signs of aging measures in the Dunedin Study

Physical Functioning (N=817 with DNAm data)	
Balance	Balance was measured using the Unipedal Stance Test as the maximum time achieved across three trials of the test with eyes closed (Bohannon et al., 1984; Springer et al., 2007; Vereeck et al., 2008).
Gait Speed	Gait speed (meters per second) was assessed with the 6-m-long GAITRite Electronic Walkway (CIR Systems, Inc) with 2-m acceleration and 2-m deceleration before and after the walkway, respectively. Gait speed was assessed under 3 walking conditions: usual gait speed (walk at normal pace from a standing start, measured as a mean of 2 walks) and 2 challenge paradigms, dual-task gait speed (walk at normal pace while reciting alternate letters of the alphabet out loud, starting with the letter "A," measured as a mean of 2 walks) and maximum gait speed (walk as fast as safely possible, measured as a mean of 3 walks). We calculated the mean of the 3 individual walk conditions to generate our primary measure of composite gait speed (Rasmussen et al., 2019).
Steps in Place	The 2-min step test was measured as the number of times a participant lifted their right knee to mid-thigh height (measured as the height half-way between the knee cap and the iliac crest) in 2 minutes at a self-directed pace (Jones and Rikli, 2002; Rikli and Jones, 1999).
Chair Stands	Chair rises were measured as the number of stands a participant completed in 30 seconds from a seated position (Jones et al., 1999; Jones and Rikli, 2002).
Grip Strength	Handgrip strength was measured for the dominant hand (elbow held at 90°, upper arm held tight against the trunk) as the maximum value achieved across three trials using a Jamar digital dynamometer (Mathiowetz et al., 1985; Rantanen T et al., 1999).
Motor Coordination	At ages 38 and 45, we measured motor functioning as the time to completion of the Grooved Pegboard Test with the dominant hand.
Physical Limitations	Physical limitations were measured with the 10-item RAND 36-Item Health Survey 1.0 physical functioning scale (Ware and Sherbourne, 1992). Participant responses ("limited a lot", "limited a little", "not limited at all") assessed their difficulty with completing various activities, e.g., climbing several flights of stairs, walking more than 1 km, participating in strenuous sports, etc. Scores were reversed to reflect physical limitations so that a high score indicates more limitations.
Decline in Physical Functioning	Tests of balance and grip strength and interviews about physical limitations were completed at both the age-38 and age-45 Dunedin Study assessments. We measured decline across the 7-year measurement interval by subtracting the age-38 test score from the age-45 test score.
Cognitive Functioning (N=814 with DNAm data)	
Cognitive Functioning	The Wechsler Adult Intelligence Scale-IV (WAIS-IV) (Wechsler, 2008) was administered to the participants at age 45 years, yielding the IQ. In addition to full scale IQ, the WAIS-IV measures four specific domains of cognitive function: Processing Speed, Working Memory, Perceptual Reasoning, and Verbal Comprehension.

Cognitive Decline	IQ is a highly reliable measure of general intellectual functioning that captures overall ability across differentiable cognitive functions. We measured IQ from the individually administered Wechsler Intelligence Scale for Children-Revised (WISC-R; averaged across ages 7, 9, 11, and 13) (Wechsler, 2003) and the Wechsler Adult Intelligence Scale-IV (WAIS-IV; age 45) (Wechsler, 2008). We measured IQ decline by comparing scores from the WISC-R and the WAIS-IV.
Subjective Signs of Aging (N=817 with DNAm data)	
Self-rated Health	Study members rated their health on a scale of 1-5 (poor, fair, good, very good, or excellent).
Facial Aging	Facial Aging is the subjective perception of aged appearance based on a facial photograph and is proposed as a clinically-useful marker of mortality risk (Christensen et al., 2009). Facial Aging measurement in the Dunedin Study was based on ratings by an independent panel of 8 raters of each participant's facial photograph (Belsky et al., 2015; Shalev et al., 2014). Facial Aging was based on two measurements of perceived age. First, Age Range was assessed by an independent panel of 4 raters, who were presented with standardized (non-smiling) facial photographs of participants and were kept blind to their actual age. Raters used a Likert scale to categorize each participant into a 5-year age range (i.e., from 20-24 years old up to 70+ years old) (interrater reliability = .77). Scores for each participant were averaged across all raters. Second, Relative Age was assessed by a different panel of 4 raters, who were told that all photos were of people aged 45 years old. Raters then used a 7-item Likert scale to assign a "relative age" to each participant (1="young looking", 7="old looking") (interrater reliability = .79). The measure of perceived age at 45 years, Facial Age, was derived by standardizing and averaging Age Range and Relative Age scores.
Subjective Decline	Self-rated Health and Facial Aging were measured at both the age-38 and age-45 assessments. We measured decline in self-rated health as incident fair/poor health reported at the age-45 assessment. We measured acceleration in Facial Aging by computing the difference in Facial Aging Z-scores between the age-45 and age-38 assessments.

Supplementary Table 4. Effect-sizes for analysis of DunedinPACE, DunedinPoAm, and DNA methylation clocks in models with covariate adjustment for age and sex only and for models with additional covariate adjustment for DNA methylation estimated cell counts and for smoking. Effect-sizes are reported for a one standard deviation increment of DunedinPACE. Outcome variables in Dunedin Study, Understanding Society Study, and E-Risk Study (Panels A, B, C, and E) were z-transformed for analysis; effect-sizes are interpretable as Pearson r or Cohen's d. For Normative Aging Study and Framingham Analysis (Panel C), effect-sizes for time-to event outcomes are reported as hazard ratios (HR), effect-sizes for prevalent chronic disease are reported as Risk Ratios (RR), and effect-sizes for incident disability are reported as incidence rate ratios (IRR).

Panel A. Effect-sizes for associations of DunedinPACE, DunedinPoAm, and DNA methylation clocks with functional assessments at age 45 in the Dunedin Study. Effect-sizes are standardized coefficients from linear regressions of the outcomes on the aging measures. The models adjusted for cell counts (middle column of results) were adjusted for cell counts measured from complete blood count data (white blood cell count and percentages of lymphocytes, monocytes, neutrophils, eosinophils, and basophils). The models adjusted for smoking (left-most column of results) included a covariate for number of pack-years smoked.

Panel B. Effect-sizes for associations of DunedinPACE, DunedinPoAm, and DNA methylation clocks with functional decline in the Dunedin Study. Decline was measured as change between the age-38 and age-45 assessments for all measures except cognition, for which change was computed over the interval beginning with an adolescent baseline test at age 13 and ending with the most recent cognitive assessment at age 45. Effect-sizes are coefficients from linear regressions of change scores on aging measures for all outcomes except incident fair/poor health, for which effect-sizes are risk ratios (RR) estimated from Poisson regression of incidence on the aging measures. Change scores were computed by standardizing baseline and follow-up measurements using the mean and standard deviation (SD) of the baseline assessment and computing the difference in standardized scores. A single unit of change therefore corresponds to 1 SD of variation at baseline. Effect-sizes are reported for a one SD difference in the aging measures. The models adjusted for smoking (left-most column of results) included a covariate for number of pack-years smoked.

Panel C. Effect-sizes for associations of DunedinPACE, DunedinPoAm, and DNA methylation clocks with clinical-biomarker measures of biological age and self-rated health in the Understanding Society Study. Effect-sizes are standardized coefficients from linear regressions of the outcomes on the DNA methylation measures of aging. The models adjusted for smoking included covariates for smoking status and current smoking quantity (cigarettes per day).

Panel D. Effect-sizes for associations of DunedinPACE, DunedinPoAm, and DNA methylation clocks with mortality, morbidity, and disability in the Normative Aging Study and the Framingham Heart Study Offspring Cohort. Effect-sizes are hazard ratios (HR) for time-to-event analyses and relative risks (RR) and incidence rate ratios (IRR) for repeated-measures analyses. Analysis of mortality was conducted in both cohorts. Analysis of incident and prevalent chronic disease morbidity was conducted only in the Normative Aging Study.

Time-to-event analysis of cardiovascular disease (CVD) and stroke/transient ischemic attack (TIA) and repeated measures analysis of incident disability were conducted only in the Framingham Heart Study Offspring Cohort. The models adjusted for smoking included covariates for smoking history (pack-years smoked) in the Normative Aging Study and for smoking status and current smoking quantity (cigarettes per day) in the Framingham Heart Study.

Panel E. Effect-sizes for associations of childhood socioeconomic status (SES) and childhood victimization with aging measures in the E-Risk Study. Effect-sizes are standardized coefficients (Cohen's d) from linear regressions of the aging measures on exposure groups (low and middle SES vs. high SES; one, two, and three or more types of victimization vs. no victimization). Models included covariate adjustment for sex. Standard errors were clustered at the family level to account for non-independence of sibling data.

Supplementary Table 5. Independent associations of DunedinPACE with mortality, morbidity, and disability in models including covariate adjustment for DNA methylation clocks. The table reports effect-sizes for DunedinPACE and DNA methylation clocks from time-to-event analysis of mortality, cardiovascular disease (CVD), and stroke or transient ischemic attack (TIA) and repeated-measures analysis of incident limitations to activities of daily living (ADLs) in the Framingham Heart Study Offspring cohort. Each model includes DunedinPACE, one of the DNA methylation clocks, and covariates for age and sex. Time-to-event model effect-sizes are reported as hazard ratios (HR). Repeated-measures model effect-sizes are reported as incidence-rate ratios (IRR).

	Time-to-Event Analysis of Mortality, Cardiovascular Disease (CVD) Diagnosis, and Stroke or Transient Ischemic Attack (TIA)								
	Mortality			CVD			Stroke/TIA		
	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
DunedinPACE	1.64	[1.48-1.82]	6.46E-22	1.39	[1.25-1.54]	8.08E-10	1.42	[1.18-1.70]	1.57E-04
Horvath Clock	1.02	[0.95-1.10]	0.584	1.04	[0.95-1.14]	0.429	1.00	[0.86-1.17]	0.998
DunedinPACE	1.60	[1.43-1.78]	8.59E-17	1.34	[1.21-1.49]	1.73E-08	1.35	[1.13-1.61]	0.001
Hannum Clock	1.09	[1.01-1.17]	0.019	1.12	[1.02-1.23]	0.020	1.17	[1.00-1.36]	0.052
DunedinPACE	1.57	[1.40-1.75]	2.11E-15	1.35	[1.21-1.50]	9.25E-08	1.33	[1.07-1.65]	0.011
PhenoAge Clock	1.14	[1.03-1.25]	0.009	1.10	[1.00-1.20]	0.053	1.20	[1.02-1.40]	0.025
DunedinPACE	1.24	[1.12-1.38]	6.89E-05	1.18	[1.05-1.34]	0.007	1.33	[1.05-1.69]	0.019
GrimAge Clock	1.61	[1.49-1.74]	1.30E-33	1.33	[1.19-1.49]	5.74E-07	1.11	[0.91-1.35]	0.295

	Repeated Measures Analysis of Incident Limitation to Activities of Daily Living (ADLs)								
	Nagi ADLs			Katz ADLs			Rosow-Breslau ADLs		
	IRR	95% CI	p-value	IRR	95% CI	p-value	IRR	95% CI	p-value
DunedinPACE	1.39	[1.17-1.65]	1.49E-04	1.31	[1.14-1.50]	1.02E-04	1.40	[1.24-1.57]	2.36E-08
Horvath Clock	1.05	[0.88-1.26]	0.565	1.11	[0.98-1.26]	0.091	0.96	[0.89-1.05]	0.385
DunedinPACE	1.37	[1.14-1.64]	6.63E-04	1.30	[1.13-1.51]	3.07E-04	1.37	[1.20-1.57]	2.84E-06
Hannum Clock	1.08	[0.91-1.28]	0.381	1.06	[0.94-1.19]	0.367	1.04	[0.91-1.19]	0.562
DunedinPACE	1.40	[1.14-1.72]	0.001	1.26	[1.06-1.50]	0.007	1.43	[1.25-1.64]	3.54E-07
PhenoAge Clock	1.00	[0.77-1.30]	0.973	1.13	[0.95-1.35]	0.161	0.93	[0.81-1.07]	0.298
DunedinPACE	1.27	[1.02-1.58]	0.032	1.26	[1.02-1.54]	0.029	1.27	[1.08-1.50]	0.005
GrimAge Clock	1.18	[0.94-1.48]	0.158	1.10	[0.90-1.34]	0.357	1.15	[0.97-1.37]	0.098

Supplementary Table 6. Items included in Nagi, Katz, and Rosow-Breslau scales of limitations to Activities of Daily Living (ADLs).

Nagi ADL Scale. Count of activities for which participants reported a lot of difficulty or inability to perform.
Pulling or pushing large objects Stooping, crouching, or kneeling Reaching or extending arms below shoulder level Reaching or extending arms above shoulder level Writing, handling, or fingering small objects Standing in one place for long periods (15 minutes) Sitting for long periods (one hour) Lifting or carrying weights under 10 lbs Lifting or carrying weights over 10 lbs
Katz ADL Scale. Count of activities for which participants required assistance or could not do themselves.
Dressing Bathing Eating Transferring (getting in and out of a chair) Toileting
Rosow-Breslau ADL Scale. Count of activities participants were not able to do.
Heavy work around the house Walk half a mile without assistance Walk up and down one flight of stairs

SUPPLEMENT REFERENCES

Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, Harrington H, Israel S, Levine ME, Schaefer JD, Sugden K, Williams B, Yashin AI, Poulton R, Moffitt TE. 2015. Quantification of biological aging in young adults. *Proc Natl Acad Sci USA* **112**:E4104-4110. doi:10.1073/pnas.1506264112

Bohannon RW, Larkin PA, Cook AC, Gear J, Singer J. 1984. Decrease in timed balance test scores with aging. *Phys Ther* **64**:1067–1070.

Christensen K, Thinggaard M, McGue M, Rexbye H, Hjelmborg J v B, Aviv A, Gunn D, van der Ouderaa F, Vaupel JW. 2009. Perceived age as clinically useful biomarker of ageing: cohort study. *BMJ* **339**. doi:10.1136/bmj.b5262

Elliott ML, Caspi A, Houts RM, Ambler A, Broadbent JM, Hancox RJ, Harrington H, Hogan S, Keenan R, Knodt A, Leung JH, Melzer TR, Purdy SC, Ramrakha S, Richmond-Rakerd LS, Righarts A, Sugden K, Thomson WM, Thorne PR, Williams BS, Wilson G, Hariri AR, Poulton R, Moffitt TE. 2021. Disparities in the pace of biological aging among midlife adults of the same chronological age have implications for future frailty risk and policy. *Nat Aging* **1**:295–308. doi:10.1038/s43587-021-00044-4

Jones CJ, Rikli RE. 2002. Measuring functional fitness in older adults. *The Journal on active aging* **1**:24–30.

Jones CJ, Rikli RE, Beam WC. 1999. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Research quarterly for exercise and sport* **70**:113–119.

Mathiowetz V, Kashman N, Volland G, Weber K, Dowe M, Rogers S. 1985. Grip and pinch strength: normative data for adults. *Arch Phys Med Rehabil* **66**:69–74.

Nettle D, Gadalla SM, Lai T-P, Susser E, Bateson M, Aviv A. 2021. Telomere length measurement for longitudinal analysis: implications of assay precision. *Am J Epidemiol*. doi:10.1093/aje/kwab025

Rantanen T, Guralnik JM, Foley D, et al. 1999. MIdlife hand grip strength as a predictor of old age disability. *JAMA* **281**:558–560. doi:10.1001/jama.281.6.558

Rasmussen LJH, Caspi A, Ambler A, Broadbent JM, Cohen HJ, d'Arbeloff T, Elliott M, Hancox RJ, Harrington H, Hogan S, Houts R, Ireland D, Knodt AR, Meredith-Jones K, Morey MC, Morrison L, Poulton R, Ramrakha S, Richmond-Rakerd L, Sison ML, Sneddon K, Thomson WM, Hariri AR, Moffitt TE. 2019. Association of Neurocognitive and Physical Function With Gait Speed in Midlife. *JAMA Netw Open* **2**:e1913123–e1913123. doi:10.1001/jamanetworkopen.2019.13123

Rikli RE, Jones CJ. 1999. Functional fitness normative scores for community-residing older adults, ages 60-94. *Journal of aging and physical activity* **7**:162–181.

Shalev I, Caspi A, Ambler A, Belsky DW, Chapple S, Cohen HJ, Israel S, Poulton R, Ramrakha S, Rivera CD, Sugden K, Williams B, Wolke D, Moffitt TE. 2014. Perinatal complications and aging indicators by midlife. *Pediatrics* **134**:e1315-1323. doi:10.1542/peds.2014-1669

Springer BA, Marin R, Cyhan T, Roberts H, Gill NW. 2007. Normative values for the unipedal stance test with eyes open and closed. *J Geriatr Phys Ther* **30**:8–15.

Thomas S. 2017. Telomeres as Sentinels for Environmental Exposures, Psychosocial Stress, and Disease Susceptibility: Workshop Summary. *NIEHS-NIA Workshop on Telomeres*.

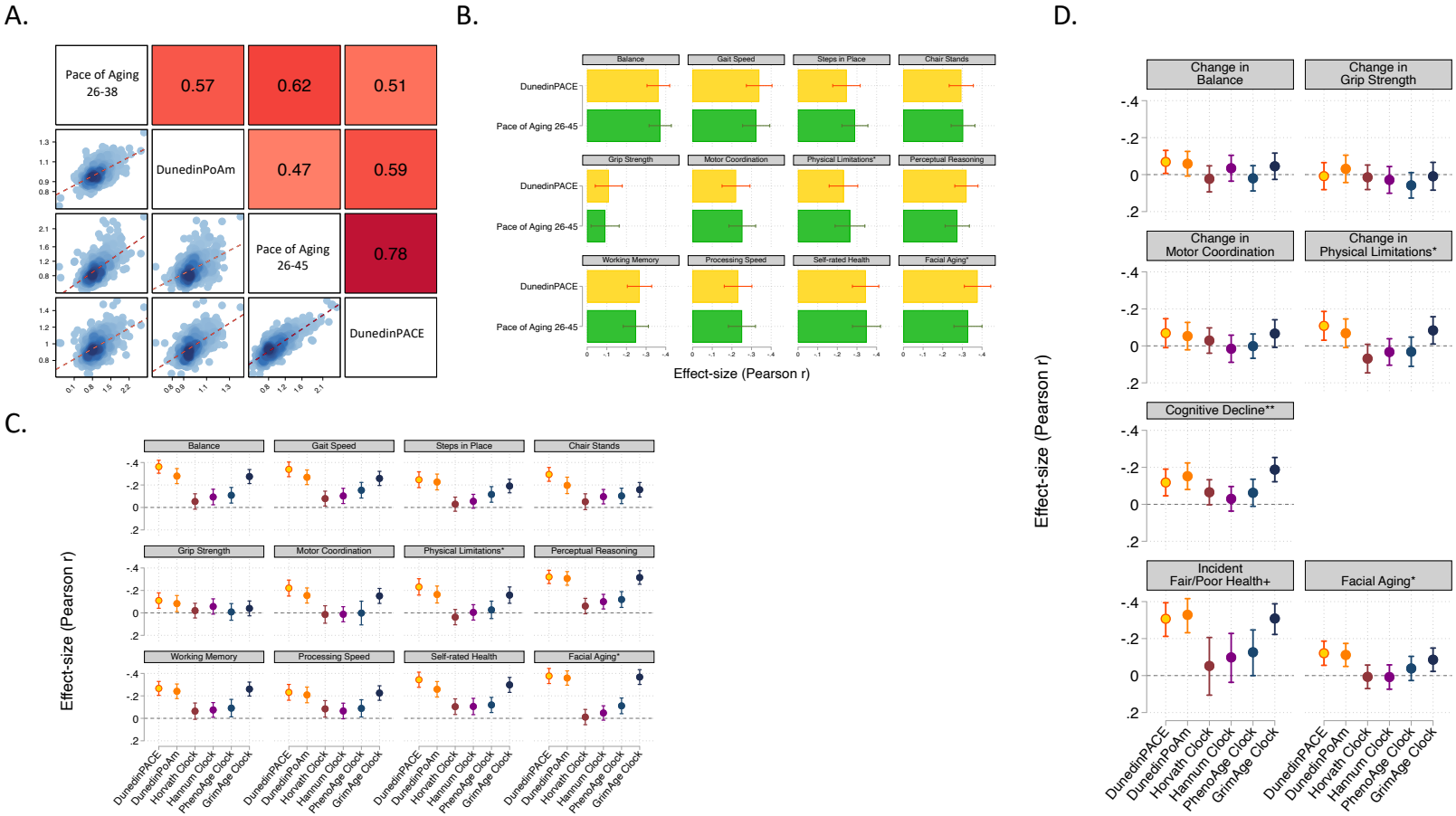
Vereeck L, Wuyts F, Truijen S, Van de Heyning P. 2008. Clinical assessment of balance: Normative data, and gender and age effects. *International Journal of Audiology* **47**:67–75. doi:10.1080/14992020701689688

Ware JE Jr, Sherbourne CD. 1992. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Medical Care* **30**:473–483.

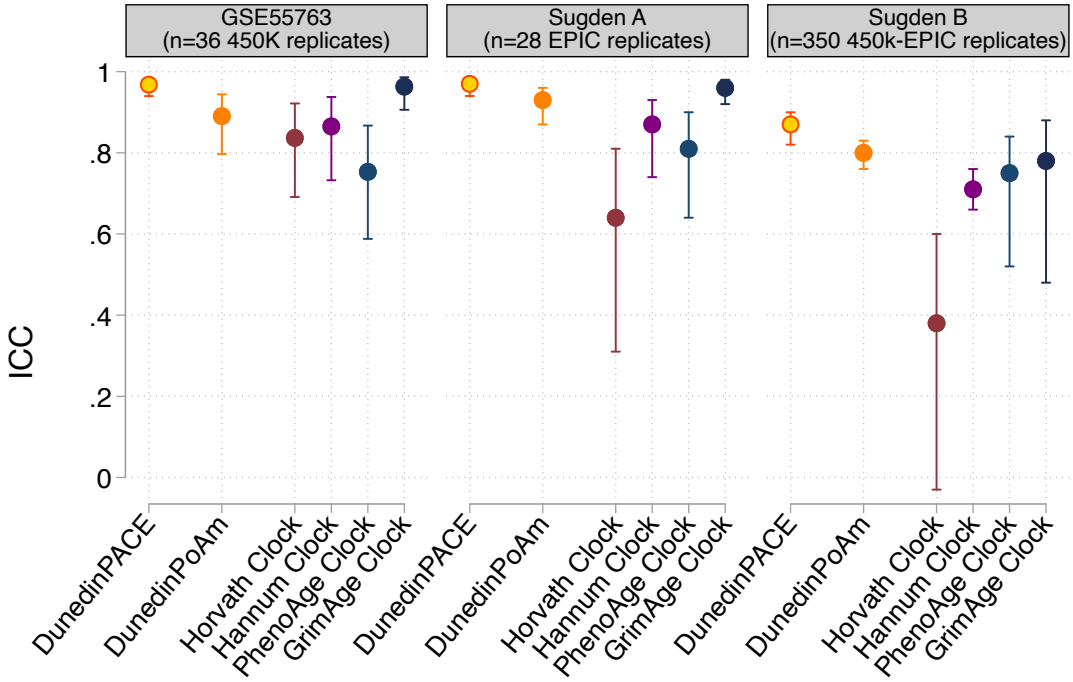
Wechsler D. 2008. Wechsler Adult Intelligence Scale, 4th ed. San Antonio, TX: Pearson Assessment.

Wechsler D. 2003. Wechsler Intelligence Scale for Children, 4th (UK Version). ed. San Antonio, TX: Harcourt Assessment.

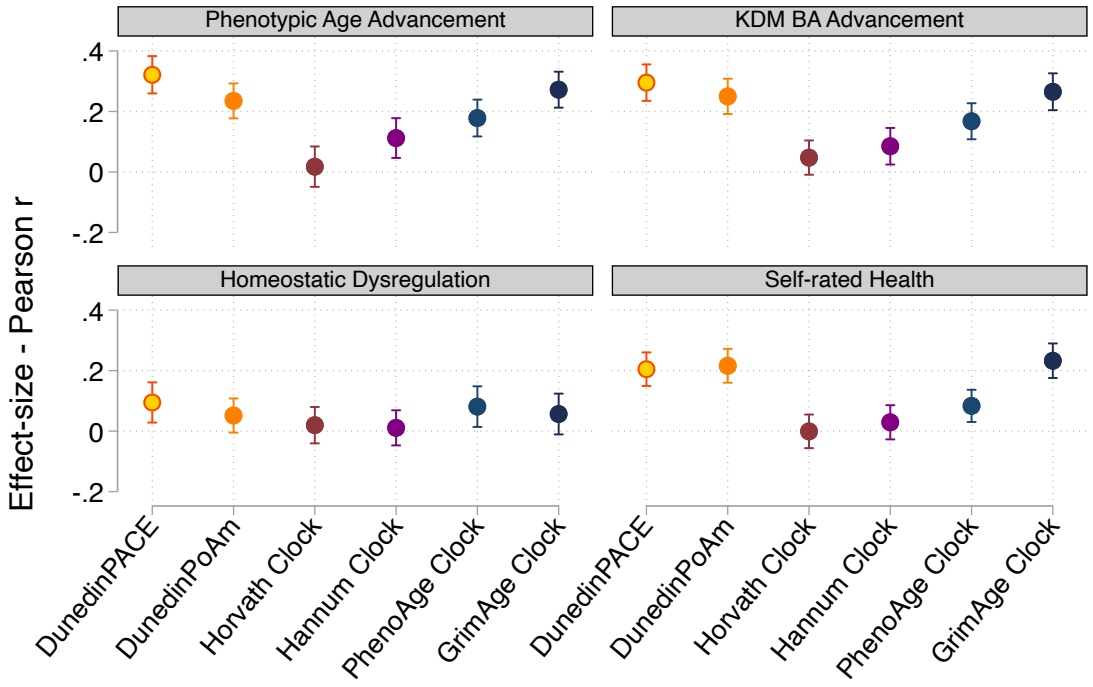
Supplementary Figure 1. DunedinPACE provides a DNA methylation surrogate measure of the 20-year Pace of Aging. Panel A shows correlations among the 12-year and 20-year Pace of Aging measures and the DunedinPoAm and DunedinPACE DNA methylation measures. Panel B graphs effect-sizes for associations of DunedinPoAm and the 20-year Pace of Aging with measures of physical and cognitive functioning and subjective signs of aging measured when Dunedin Study participants were aged 45 years. Panel C graphs effect-sizes for associations of DunedinPACE, DunedinPoAm, and the DNA methylation clocks proposed by Horvath, Hannum et al., Levine et al. (PhenoAge), and Lu et al. (GrimAge) with measures of physical and cognitive functioning and subjective signs of aging measured when Dunedin Study participants were aged 45 years. Panel D graphs effect-sizes for associations of aging measures with measures of change in physical functioning and subjective signs of aging over ages 38-45 years and with cognitive functioning from adolescent baseline to age-45 follow-up. In Panels C and D, the * indicates measures reverse-coded for analysis. DunedinPACE and DunedinPoAm were developed from analysis of Pace of Aging in the Dunedin cohort; effect-sizes for these measures may be over-estimated.



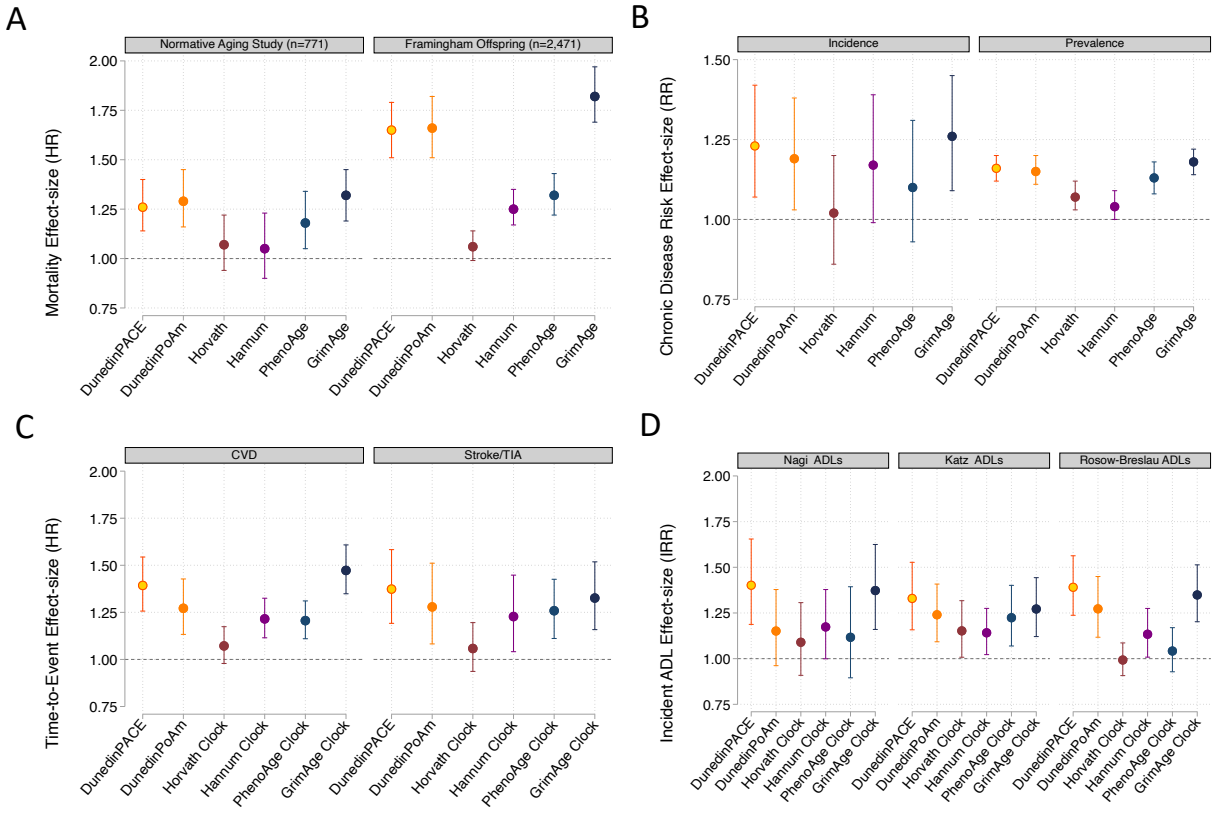
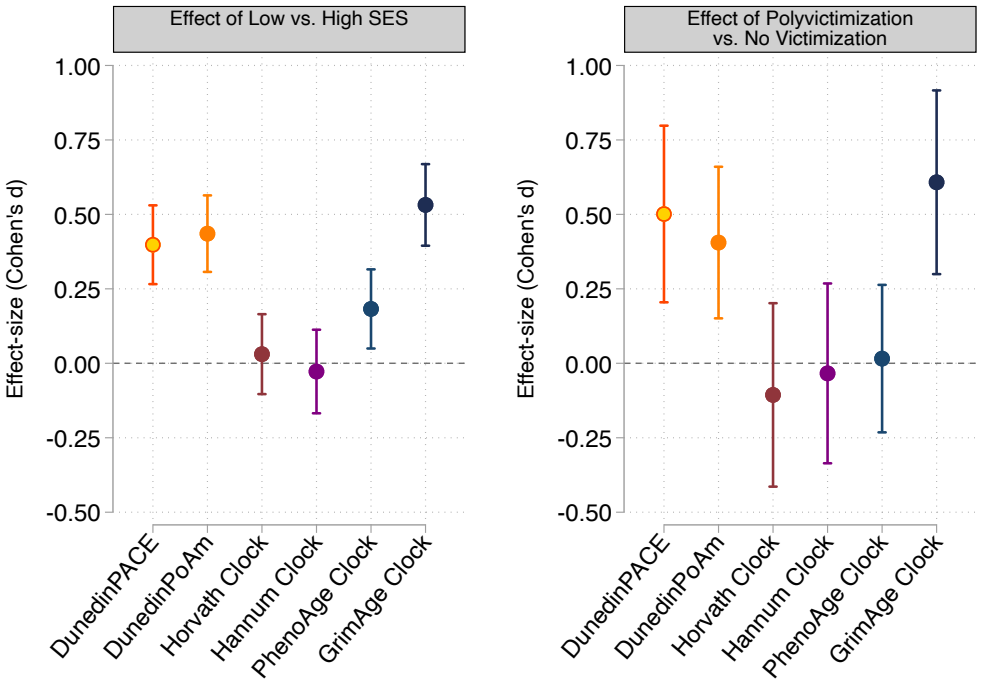
Supplementary Figure 2. Reliability of DunedinPACE, DunedinPoAm, and DNA methylation clocks. The figure plots intraclass correlation coefficients (ICCs) estimated from replicate DNA methylation datasets for DunedinPACE, original DunedinPoAm, and the DNA methylation clocks proposed by Horvath, Hannum et al., Levine et al. (PhenoAge), and Lu et al. (GrimAge). Error bars show 95% confidence intervals. Data are graphed from three datasets, 36 Illumina 450k technical replicates from GSE55763 (Lehne et al., 2015), and 28 Illumina EPIC array technical replicates and 350 Illumina 450k-EPIC replicates from Sugden et al. (Sugden et al., 2020). The Sugden et al. EPIC technical replicate dataset (Sugden A) included arrays analyzed as part of the Dunedin Study dataset. The Sugden et al. 450k-EPIC dataset (Sugden B) was used to identify probes for inclusion in the machine learning analysis from which DunedinPACE was derived.



Supplementary Figure 3. Effect-sizes for associations of DunedinPACE, DunedinPoAm, and DNA methylation clocks with physiology-based measures of biological age and self-rated health. Figure shows effect-sizes estimated from the Understanding Society data (n=1,175). Effect-sizes are reported as standardized regression coefficients interpretable as Pearson r values. Error bars show 95% confidence intervals. DNA methylation clocks were residualized for chronological age prior to analysis. Models included covariates for chronological age and sex. Physiology-based measures of biological age were computed from Understanding Society biomarker data (albumin, alkaline phosphatase, creatinine, C-reactive protein, blood urea nitrogen, glycated hemoglobin, systolic blood pressure, and forced expiratory volume in 1 second) based on algorithms derived in data from the US NHANES according to the methods developed by Levine et al. (Levine et al., 2018), Klemera and Doubal (Klemera and Doubal, 2006), and Cohen et al. (Cohen et al., 2013).



Supplementary Figure 4. Effect-sizes for associations of DunedinPACE, DunedinPoAm, and DNA methylation clocks with mortality, morbidity, and disability. Panel A plots effect-sizes for analysis of mortality in the Normative Aging Study and the Framingham Heart Study Offspring cohort. Effect-sizes are reported as hazard ratios (HR) per standard deviation increment of the aging measures estimated from Cox proportional hazard regression. Panel B plots effect-sizes for analysis of incident and prevalent chronic disease morbidity in the Normative Aging Study. Effect-sizes are reported as relative risks (RR) per standard deviation increment of the aging measures estimated from Poisson regression. Panel C plots effect-sizes for analysis of time-to-diagnosis with cardiovascular disease (CVD) and time to stroke or transient ischemic attack (TIA) in the Framingham Heart Study Offspring Cohort. Effect-sizes are reported as hazard ratios (HR) per standard deviation increment of the aging measures estimated from Cox proportional hazard regression. Panel D plots effect-sizes for analysis of incident disability from repeated-measures of limitations to activities of daily living (ADLs). For all panels, error bars show 95% confidence intervals. DNA methylation clocks were residualized for chronological age prior to analysis. All models included covariates for chronological age and sex.



Supplementary Figure 5. Effect-sizes for associations of childhood socioeconomic conditions and childhood victimization history with DunedinPACE, DunedinPoAm, and DNA methylation clocks at age 18. Panel A plots effect-sizes for low socioeconomic status (SES) in childhood (vs. high SES in childhood). Panel B plots effect-sizes for childhood polyvictimization (vs. no childhood victimization). Effect-sizes are reported as Cohen's d. For both panels, error bars show 95% confidence intervals. DNA methylation clocks were residualized for chronological age prior to analysis. All models included sex as a covariate. (All E-Risk participants are the same chronological age; no age covariate is needed.)