nature aging

Letter

https://doi.org/10.1038/s43587-022-00357-y

# Effect of long-term caloric restriction on DNA methylation measures of biological aging in healthy adults from the CALERIE trial

Received: 8 September 2021

Accepted: 22 December 2022

Check for updates

R. Waziry[1,16], C. P. Ryan[1,16], D. L. Corcoran[2], K. M. Huffman[3], M. S. Kobor[4], M. Kothari[1], G. H. Graf[1,5], V. B. Kraus[3], W. E. Kraus[3], D. T. S. Lin[4], C. F. Pieper[6], M. E. Ramaker[3], M. Bhapkar[7], S. K. Das[8], L. Ferrucci[9], W. J. Hastings[10], M. Kebbe[11], D. C. Parker[12], S. B. Racette[13,14], I. Shalev[10], B. Schilling[15] & D. W. Belsky[2,5]

The geroscience hypothesis proposes that therapy to slow or reverse molecular changes that occur with aging can delay or prevent multiple chronic diseases and extend healthy lifespan[1–3]. Caloric restriction (CR), defined as lessening caloric intake without depriving essential nutrients[4], results in changes in molecular processes that have been associated with aging, including DNA methylation (DNAm)[5–7], and is established to increase healthy lifespan in multiple species[8,9]. Here we report the results of a post hoc analysis of the influence of CR on DNAm measures of aging in blood samples from the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE) trial, a randomized controlled trial in which $n = 220$ adults without obesity were randomized to 25% CR or ad libitum control diet for 2 yr (ref. [10]). We found that CALERIE intervention slowed the pace of aging, as measured by the DunedinPACE DNAm algorithm, but did not lead to significant changes in biological age estimates measured by various DNAm clocks including PhenoAge and GrimAge. Treatment effect sizes were small. Nevertheless, modest slowing of the pace of aging can have profound effects on population health[11–13]. The finding that CR modified DunedinPACE in a randomized controlled trial supports the geroscience hypothesis, building on evidence from small and uncontrolled studies[14–16] and contrasting with reports that biological aging may not be modifiable[17]. Ultimately, a conclusive test of the geroscience hypothesis will require trials with long-term follow-up to establish effects of intervention on primary healthy-aging endpoints, including incidence of chronic disease and mortality[18–20].

Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE) Phase 2 was a multi-center, randomized controlled trial conducted at three clinical centers in the United States[10]. It aimed to evaluate the time-course effects of 25% CR (that is, intake 25% below the individual's baseline level) over a 2-yr period in healthy adults (men aged 21–50 yr, premenopausal women aged 21–47 yr) with body mass

A full list of affiliations appears at the end of the paper. e-mail: Daniel.Belsky@Columbia.edu

https://doi.org/10.1038/s43587-022-00357-y

index (BMI) in the normal weight or slightly overweight range (BMI 22.0–27.9 kg m$^{-2}$). Participants were randomly assigned at a ratio of 2:1 to a CR behavioral intervention or to an ad libitum (AL) control group stratified by site, sex and BMI. Of 238 eligible individuals, CALERIE randomized $N = 220$ participants (145 CR intervention and 75 AL control; Fig. 1). Participants in the CR group were prescribed a 25% restriction in calorie intake based on energy requirements estimated from two 2-week doubly labeled water (DLW) measurement periods at baseline. The precise level of CR achieved was quantified by comparing energy intake (determined periodically throughout the trial by the DLW method[21]) during the CR intervention with baseline energy intake. The CALERIE Trial is described in more detail in Methods.

Blood DNAm data were generated at baseline and at least one follow-up timepoint for $n = 197$ participants (128 CR and 69 AL). Of this analysis sample, $n = 105$ (82%) CR participants and $n = 59$ (86%) AL participants had DNAm data available from all three timepoints (baseline, 12 months and 24 months). DNAm analysis is described in more detail in Methods. Participants had a mean age of 38 yr (s.d. = 7), 70% were women and 77% were white; there were no differences in age, sex or race/ethnicity between AL and CR at baseline (Table 1).

The goal of our analysis was to test the effect of CALERIE intervention on biological aging. We measured biological aging from blood DNAm using published algorithms. These algorithms aim to capture the accumulation of molecular changes that underlie the progressive loss of system integrity that occurs with advancing chronological age. Primary analysis focused on the PhenoAge[22] and GrimAge[23] second-generation DNAm clocks and the DunedinPACE[24] measure of pace of aging, all of which show strong associations with aging-related morbidity and mortality. We analyzed versions of the PhenoAge and GrimAge clocks constructed from DNAm principal components (PCs) (hereafter 'PC clocks'), which have superior technical reliability as compared with the original versions of these measures[25]; DunedinPACE was originally designed to have high technical reliability. Measures are described in detail in Table 2 and Methods. Associations of DNAm measures of aging with chronological age at preintervention baseline are shown in Supplementary Fig. 2. Mean values of the DNAm measures of aging in the CR and AL groups at baseline and each follow-up are reported in Supplementary Table 1. Intraclass correlation coefficients for tests of technical reliability and within-individual stability are reported in Supplementary Table 2.

We computed change scores for the DNAm measures of aging as the differences of 12-month and 24-month follow-up values from baseline values. For analysis, change scores were scaled so that effect sizes can be interpreted as standardized differences between means (Cohen's $d$). For PhenoAge and GrimAge clocks, change score values were scaled by the standard deviation of the difference between clock age and chronological age at pretreatment baseline. For DunedinPACE, which measures pace of aging (that is, change in biological age per chronological year), change score values were scaled by the standard deviation at pretreatment baseline. Scaled change scores are reported in Supplementary Table 3. Change scores are graphed in Fig. 2 and Supplementary Fig. 3.

To test the hypothesis that CR slowed biological aging, we conducted intent-to-treat (ITT) analysis which compared change scores between participants randomized to CR intervention and AL control group using repeated-measures analysis of covariance (ANCOVA) implemented under mixed models, following the approach used in past CALERIE analysis[26]. Model details are reported in Methods. We use $P < 0.005$ as a conservative threshold for statistical significance following guidance from leaders in the field[27]. As expected, participants' PhenoAge and GrimAge values tended to increase over time. However, change in PhenoAge and GrimAge values did not differ between CR and AL groups (for PhenoAge, 12-month $d = -0.03$ (95% confidence interval (95% CI) −0.19, 0.12), 24-month $d = 0.05$ (95% CI −0.11, 0.20), $P > 0.50$ for both; for GrimAge, 12-month $d = -0.04$ (95% CI −0.16, 0.07),

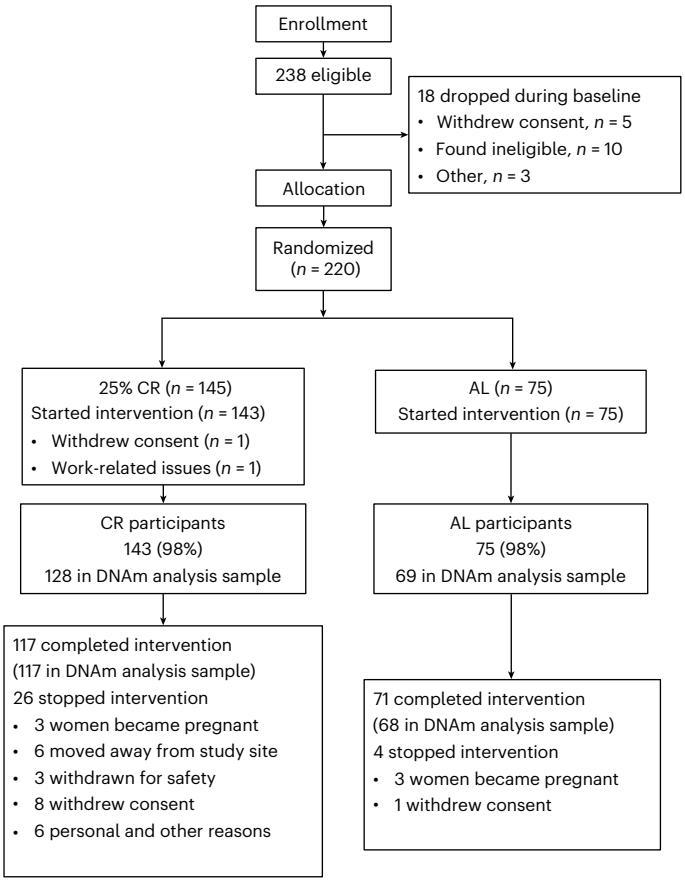



**Fig. 1 | Consort diagram for the CALERIE Trial.** DNAm was assayed from blood samples collected at baseline, 12 months and 24 months. Of the $n = 197$ participants for whom DNAm data were available from baseline and at least one follow-up assessment, baseline to 12-month change was measured for $n = 125$ CR and $n = 66$ AL participants and baseline to 24-month change was measured for $n = 117$ CR and $n = 68$ AL participants.

24-month $d = 0.05$ (95% CI −0.07, 0.17), $P > 0.40$ for both). CR treatment reduced participants' DunedinPACE by the 12-month follow-up and this reduction was maintained through follow-up at 24 months (12-month $d = -0.29$ (95% CI −0.45, −0.13), 24-month $d = -0.25$ (95% CI −0.41, −0.09), $P < 0.003$ for both). Standardized treatment effects on DunedinPACE correspond to a reduction in the pace of aging of 2–3%. These average treatment effects summarize diverse responses to intervention; for some treatment group participants, reductions in DunedinPACE were much larger, whereas, for others, DunedinPACE increased from baseline to follow-up. ITT results are reported in Supplementary Table 4.

In the CALERIE Trial, the %CR achieved by participants in the treatment group varied, with most participants achieving doses below the prescribed 25% (mean = 11.9, s.e.m. = 0.7%)[10]. We therefore conducted analyses (1) to test if those who achieved higher CR doses experienced larger treatment effects (dose–response); and (2) to quantify the treatment effect that would be expected among individuals achieving a high dose (we selected a dose of 20%, the 75th percentile of the CR distribution in the treatment group at 12 months, hereafter 'effect-of-treatment-on-the-treated' or TOT). To test dose–response, we stratified CR treatment group participants according to whether they achieved at least 10% CR and repeated ITT analysis. For DunedinPACE, the treatment effect in the >10% CR group was $d = -0.33$ at 12 months and $d = -0.33$ at 24 months, as compared with $d = -0.19$ at 12 months and $d = -0.14$ at 24 months in the <10% CR group. There was no evidence of a dose–response effect for PhenoAge or GrimAge. Full results are

https://doi.org/10.1038/s43587-022-00357-y

**Table 1 | Characteristics of CALERIE Trial participants at baseline**

| | AL | | CR | |
|---|---|---|---|---|
| | *N*/mean | (%/s.d.) | *N*/mean | (%/s.d.) |
| **CALERIE participant characteristics at baseline** | | | | |
| Age at baseline (yr) | 38 | (7) | 38 | (7) |
| Energy intake (kcal d$^{-1}$) | 2,045 | (481) | 2,124 | (558) |
| Sex | | | | |
| Women | 53 | (71%) | 100 | (69%) |
| Men | 22 | (29%) | 45 | (31%) |
| Race/ethnic identity | | | | |
| White | 57 | (76%) | 111 | (77%) |
| Black | 11 | (15%) | 16 | (11%) |
| Other | 7 | (9%) | 18 | (12%) |
| Study site | | | | |
| A | 25 | (33%) | 47 | (32%) |
| B | 27 | (36%) | 53 | (37%) |
| C | 23 | (31%) | 45 | (31%) |
| BMI stratum | | | | |
| 22.0–24.9 kg m$^{-2}$ | 37 | (54%) | 70 | (48%) |
| 25.0–27.9 kg m$^{-2}$ | 38 | (55%) | 75 | (52%) |
| **Analysis sample characteristics at baseline** | | | | |
| Age at baseline | 38 | (7) | 38 | (7) |
| Energy intake (kcal d$^{-1}$) | 2,023 | (469) | 2,130 | (540) |
| Sex | | | | |
| Women | 48.00 | (70%) | 89 | (70%) |
| Men | 21.00 | (30%) | 39 | (30%) |
| Race/ethnic identity | | | | |
| White | 52.00 | (75%) | 100 | (78%) |
| Black | 10 | (14%) | 13 | (10%) |
| Other | 7 | (10%) | 15 | (12%) |
| Study site | | | | |
| A | 24 | (35%) | 42 | (33%) |
| B | 25 | (36%) | 45 | (35%) |
| C | 20 | (29%) | 41 | (32%) |
| BMI stratum | | | | |
| 22.0–24.9 kg m$^{-2}$ | 33 | (48%) | 60 | (47%) |
| 25.0–27.9 kg m$^{-2}$ | 36 | (52%) | 68 | (53%) |

The table shows data for participants randomized to the AL control group and the CR treatment group. CALERIE included a total of *N*=220 participants (AL *n*=75, of whom *n*=71 completed the study; CR *n*=145, of whom *n*=118 completed the study). The analysis sample was composed of CALERIE participants for whom DNAm data were available at baseline and at least one follow-up assessment ('analysis sample'; *N*=197; AL *n*=69, CR *n*=128).

**Table 2 | DNAm clock and pace-of-aging measures included in CALERIE analysis**

**DNAm clocks**. DNAm clock measures of aging are algorithms that estimate biological age, the state of an organism's biology represented as the age at which that state would be typical in a reference population. The clocks we analyzed were developed to predict mortality risk. The age values computed by the clock algorithms correspond to the age at which predicted mortality risk would be approximately normal in the reference population used to develop the clock. We computed clock values based on versions of the clock algorithms developed from DNAm PCs (sometimes referred to as 'PC clocks')[18,21].

| | |
|---|---|
| PhenoAge clock | Based on analysis of nine blood chemistry markers, age and mortality data from the US National Health and Nutrition Examination Surveys (*n*=9,926 participants aged 18 yr and older; 23 yr of mortality follow-up); DNAm and blood chemistry data from the InCHIANTI Study (*n*=912 participants aged 21–100 yr); and the US Health and Retirement Study (*n*=3,593 participants aged 51–100 yr)[19]. |
| GrimAge clock | Based on analysis of eight plasma protein markers, smoking pack years, age, sex and mortality data from the Framingham Heart Study Offspring Cohort (*n*=2,356 participants aged 24–92 yr)[47–49]. |

**Pace of aging**. Pace-of-aging measures estimate the rate of biological aging, defined as the rate of decline in overall system integrity. Pace-of-aging values correspond to the years of biological aging experienced during a single calendar year. A value of 1 represents the typical pace of aging in a reference population; values above 1 indicate faster pace of aging; values below 1 indicate slower pace of aging.

| | |
|---|---|
| DunedinPACE | Based on analysis of pace of aging in the Dunedin Study (*n*=817 participants examined at ages 26, 32, 38 and 45 yr)[24]. Pace of aging was measured from within-person change over time in 19 blood chemistry and organ function test metrics of system integrity[24]. DNAm was measured at age 45 yr. |

reported in Supplementary Table 5. To test the TOT, we conducted instrumental variables (IV) analysis. IV analysis assumes that CALERIE intervention affected participants' biological aging only through its effect on their caloric intake. Our IV analysis estimated the % reduction in caloric intake each participant achieved because of the intervention and applied these estimates to quantify the effect of %CR on biological aging. In IV analysis, the effect of 20% CR on DunedinPACE was $d = -0.43$ (95% CI −0.67, −0.19) at 12 months and $d = -0.40$ (95% CI −0.67, −0.12) at 24 months ($P < 0.005$ for both). IV effect-size estimates for PhenoAge and GrimAge were small ($d = -0.13$–$0.01$; $P > 0.15$). TOT results are reported in Supplementary Table 4.

We tested sensitivity of results to changes in white blood cell populations in response to CALERIE intervention by including covariates in our models for DNAm estimates of cell counts[28]; these results were similar to unadjusted analyses (Supplementary Table 6).

We tested sex differences in treatment effects. We repeated ITT and TOT analyses with the addition of a product term testing interaction between the treatment variable and participant sex. Sex differences in treatment effects were not statistically different from zero in any of the models. Means of DNAm measures of aging are reported separately for men and women in Supplementary Tables 7 and 8. Sex-stratified treatment effects and tests of sex differences in treatment effects are reported in Supplementary Tables 9 and 10.

Previous studies have considered a broader set of DNAm measures of aging. In the interest of comparability across studies, we report results for so-called 'first-generation' clocks developed to predict chronological age and the original versions of the PhenoAge and GrimAge clocks in the Supplementary Information.

CR effects varied across the DNAm measures of aging we analyzed. CALERIE intervention slowed pace of aging as measured by DunedinPACE, whereas the CR intervention did not affect the PhenoAge and GrimAge DNAm clocks. All three measures have evidence for validity as biomarkers of aging, in particular, evidence of association with aging-related morbidity and mortality and with exposures associated with shortened healthy lifespan[24,29,30]. However, these DNAm measures were developed using different methods and reflect different models of aging. The PhenoAge and GrimAge clocks were developed to predict mortality risk at a single timepoint in mixed-age and older adults. This approach quantifies aging as a static construct of risk accumulated across the lifetime. In contrast, DunedinPACE was developed to predict multi-system physiological decline over two decades of follow-up from early adulthood to midlife. This approach quantifies aging as a dynamic construct reflecting change in risk accumulation. DunedinPACE may

https://doi.org/10.1038/s43587-022-00357-y

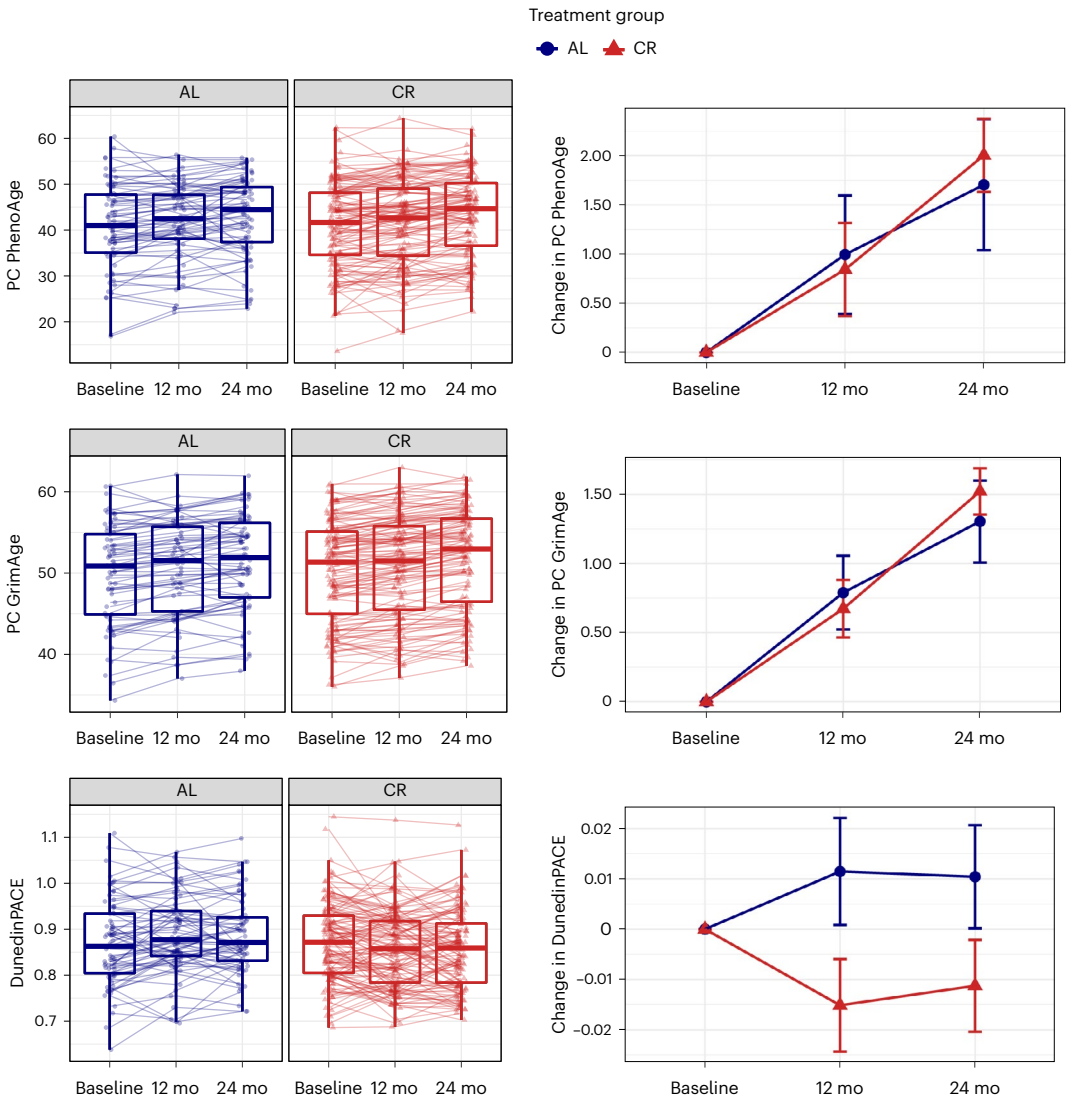



**Fig. 2 | Change from baseline to 12- and 24-month follow-up in DNAm measures of aging in the AL and CR groups in the CALERIE Trial.** The figure shows CALERIE Trial treatment effects on three DNAm measures of aging, the PC PhenoAge clock, the PC GrimAge clock and DunedinPACE. Values for the AL control group ($n = 69$ participants) are graphed in blue. Values for the CR treatment group ($n = 128$ participants) are graphed in red. For the PhenoAge and GrimAge DNAm clocks, values are denominated in 'years' of DNAm age. For the clocks, expected change under the null hypothesis is 1 yr at 12-month follow-up and 2 yr at 24-month follow-up. For the DunedinPACE measure, values are denominated in pace-of-aging units scaled to be interpretable as percentage difference in the rate of aging relative to the reference norm of 1 yr of biological decline per calendar year. For DunedinPACE, expected change under the null hypothesis is zero. The left column of the figure shows box plots of the observed values of the measures at baseline and 12- and 24-month follow-ups. The boxes show the interquartile range; the whiskers show 1.5× the interquartile range; the center line shows the median; individual participant data are plotted as dots and connected with lines. For the PC PhenoAge and PC GrimAge DNAm clocks, the box plots show similar patterns of increase in both AL and CR groups. For DunedinPACE, the box plot shows stability in the AL group and decrease in the CR group. The right column of the figure shows mean values of change from baseline and 95% CIs estimated from mixed models at the 12- and 24-month follow-ups for the AL and CR groups. There is no confidence interval estimated for baseline because change from baseline is exactly zero at this timepoint. For the PC PhenoAge and PC GrimAge DNAm clocks, mean change is similar in the AL and CR groups. For DunedinPACE, mean change is positive in the AL group (although the confidence interval overlaps zero at 24 months) and negative in the CR group. mo, months.

therefore be more sensitive than PhenoAge and GrimAge to changes induced by 2 yr of CALERIE intervention.

Our previous reports on CALERIE establish that CR intervention improved participants' cardiometabolic health and slowed aging-related changes in physiological system integrity[26,31,32]. In some cases, these effects are larger than the effects we observed for DunedinPACE (for example, $d = 0.2–0.3$ for DunedinPACE as compared with $d = 0.2–0.4$ for blood chemistry measures of biological age[32]). Changes in DunedinPACE in response to CR intervention mediated only small fractions of CR-induced changes in clinical measures (Supplementary Fig. 4). The purpose of DNAm analysis in CALERIE was to evaluate intervention effects at the molecular level, where aging processes are posited to originate[33]. Studies in subsets of CALERIE participants suggest effects of CR on molecular mechanisms of immune and metabolic regulation[34,35]. DunedinPACE findings broaden evidence of molecular changes in response to CR to a DNAm biomarker of aging established to predict morbidity and mortality.

Follow-up in the CALERIE Trial did not extend beyond the intervention. It is therefore unclear if the changes in DunedinPACE observed during the 2-yr intervention will translate into reduced morbidity and

mortality over the long term. In observational studies with long-term follow-up, individuals with slower DunedinPACE are better-off on a range of healthspan metrics, including showing reduced incidence of morbidity and increased survival[24,29]. These previous studies suggest that the CALERIE treatment effect of 2–3% slower pace of aging corresponds to a reduction in mortality risk of as much as 10–15%, similar in magnitude to the effect of smoking cessation intervention[36]. Additional follow-up of trial participants is required to determine whether CR-induced reductions to DunedinPACE in CALERIE will translate into disease prevention and increased healthy lifespan. Moreover, changes in DunedinPACE over follow-up showed substantial overlap between the CR treatment group and the AL control group; effect-size estimates imply close to 90% overlap of DunedinPACE trajectories between the two groups.

We acknowledge limitations. There is no gold standard measure of biological aging[37]. We analyzed several measures which represent the current state-of-the-art in DNAm quantification of biological aging[38]. Nevertheless, these measures are acknowledged to be incomplete summaries of biological changes that occur with aging and to have technical limitations[39,40]. Treatment effects on aspects of biological aging not captured by the DNAm measures are not included in effect estimates; measurement error due to technical limitations of DNAm assays may bias effect estimates towards the null. Treatment effect estimates may therefore represent a lower-bound of the true impact of CALERIE intervention on biological aging. The measures we studied summarize biological aging in general and do not isolate system-specific aging processes[41]. However, CR has diverse effects across multiple biological systems[42,43]. Our general measures of biological aging thus provide a reasonable test of cross-system impacts. On average, trial participants did not achieve the prescribed dose of 25% CR and some control group participants reduced their caloric intake. Despite this imperfect adherence, treatment group participants experienced substantial and sustained weight loss and related changes in body and tissue composition, broad improvement in cardiometabolic health and a slowing of aging-related physiological changes[26,31,44,45]. Our dose–response and TOT analyses indicated that participants who achieved higher doses of CR experienced more pronounced reductions in DunedinPACE. The CALERIE Trial sample does not represent the general population and treatment effects may not generalize beyond the population of healthy volunteers recruited to participate. CALERIE follow-up is, so far, limited to the end of the intervention period. Whether treatment and any slowing in biological aging that resulted from it translated to long-term clinical benefit is currently unknown.

Within the context of these limitations, our findings have implications for future geroscience research. Aging biology research has identified multiple therapies with potential to improve healthy lifespan in humans. A barrier to advancing translation of these therapies through human trials is that intervention studies run for months or years, but human aging takes decades to cause disease[46–48]. New measurements that summarize biological changes occurring with aging have potential to overcome this challenge; measurements to quantify biological aging that both predict future disease, disability and mortality and can detect changes in aging processes over short timescales have potential to function as surrogate endpoints for intervention effects on healthy lifespan[38,49]. The methods proposed to quantify biological aging analyzed in this study are predictive of aging-related health decline and mortality. However, until this study, none had been tested in a randomized controlled trial of a geroscience-based intervention[49]. Our findings highlight DunedinPACE as a measure with potential utility in future trials. DunedinPACE has high test–retest reliability and shows strong associations with healthspan endpoints in validation analyses[24,29]. Ultimately, establishing DunedinPACE and other DNAm measures of aging as surrogate endpoints for geroscience will require evidence that changes in DNAm measures account for intervention effects on primary healthy-aging endpoints, including incidence of chronic disease and mortality[18–20]. The evidence reported from CALERIE suggests that DunedinPACE may be helpful in identifying short-term interventions worthy of long-term follow-up to generate such evidence.

CALERIE was a 24-month, intensive behavioral intervention to deliver a therapy proven to slow aging in animal models. Although treatment effect sizes were small, even modest slowing of the pace of aging can have profound effects on population health[11–13]. Future trials, especially those considering less-intensive or shorter-term interventions, such as intermittent fasting[50], should plan for larger samples to ensure adequate statistical power. Further, efforts to forecast potential benefits from interventions designed to delay aging may best serve policy makers and planners if they work from assumptions of modest intervention effects.

## Methods

We conducted new DNAm assays of stored blood biospecimens collected from the CALERIE Phase 2 randomized controlled trial and merged these data with existing secondary data from the trial. The assays of the biospecimens were conducted blind to the conditions of the trial. Details of trial design and the collection of other trial data were reported previously[10,26].

### Study design and participants

CALERIE Phase 2 was a multi-center, randomized controlled trial conducted at three clinical centers in the United States[10] (ClinicalTrials.gov Identifier: NCT00427193). It aimed to evaluate the time-course effects of 25% CR (that is, intake 25% below the individual's baseline level) over a 2-yr period in healthy adults (men aged 21–50 yr, premenopausal women aged 21–47 yr) with BMI in the normal weight or slightly overweight range (BMI 22.0–27.9 kg $m^{-2}$). The study protocol was approved by Institutional Review Boards at three clinical centers (Washington University School of Medicine, St Louis, MO, USA; Pennington Biomedical Research Center, Baton Rouge, LA, USA; Tufts University, Boston, MA, USA) and the coordinating center at Duke University (Durham, NC, USA). All study participants provided written, informed consent. Nongenomic data were obtained from the CALERIE Biorepository (https://calerie.duke.edu/apply-samples-and-data-analysis).

### Randomization and masking

After baseline testing, participants were randomly assigned at a ratio of 2:1 to a CR behavioral intervention or to an AL control group. Randomization was stratified by site, sex and BMI. A permuted block randomization technique was used.

### Procedures

Study procedures were published previously[10,21,26] and are described here in brief. Participants in the CR group were prescribed a 25% restriction in calorie intake based on energy requirements estimated from two DLW measurement periods at baseline. Participants were provided three meals per day for 27 d to familiarize themselves with portion sizes for a 25% reduced calorie intake; meals included eating plans modified to suit various cultural preferences. Participants also received instruction on the essentials of CR. Finally, participants were provided with intensive group and individual behavioral counseling sessions once a week, with 24 group and individual counseling sessions over the first 24 weeks of the intervention. Adherence to the CR intervention was estimated in real time by the degree to which individual weight change followed a predicted weight loss trajectory (15.5% weight loss at 1 yr followed by weight loss maintenance). The precise level of CR achieved was quantified retrospectively by calculating energy intake during the CR intervention and comparing it with baseline energy intake. Energy intake during the 2-yr trial was quantified from total daily energy expenditure (assessed during 2-week DLW periods every 6 months) and changes in body composition (that is, fat mass and fat-free mass). Participants assigned to the AL group continued on their regular diets; they received no specific dietary intervention or

counseling. They had quarterly contact with study investigators to complete the assessments.

### Quantification of %CR

Mean %CR was calculated at each of the follow-up timepoints as percentage decrease in energy intake relative to baseline using the equation $\%CR_{mean} = (1 - EI_{mean}/EI_{BL}) \times 100$ (ref. [21]). $EI_{BL}$ was defined as total energy expenditure (TEE) at preintervention baseline and $EI_{mean}$ was defined as the average of TEE across all follow-up visits through the visit at which %CR was calculated. TEE was measured by the DLW method during two consecutive 2-week periods at baseline and during 2-week periods at months 6, 12, 18 and 24 in the CR group[10,44].

### DNAm data

DNA extracted from blood samples was obtained from the CALERIE Biorepository at the University of Vermont. DNAm data were generated by the Kobor Lab at the University of British Columbia and processed by the Genomic Analysis and Bioinformatics Shared Resource at Duke University. Illumina Infinium Methylation EPIC BeadChip arrays were used to assay genome-wide DNAm data from banked DNA samples extracted from blood collected at the baseline, 12-month and 24-month follow-ups. The EPIC array quantifies DNAm levels at >850,000 CpG sites across all known genes, regions and key regulatory regions. Briefly, 750-ng extracted DNA samples were bisulfite converted using the EZ DNA Methylation kit (Zymo Research), and 160 ng of the converted DNA was used as input for the EPIC arrays (Illumina). EPIC arrays were processed according to the manufacturer's instructions and scanned using the Illumina iScan platform. To the extent possible, baseline, 12-month and 24-month samples from the same individual were processed in the same array batch and on the same BeadChip to minimize batch effects; CR treatment and AL control participants were included on all chips. Quality control and normalization analyses were performed using the methylumi (v.2.32.0)[51] Bioconductor (v.2.46.0)[52] package for the R statistical programming environment (v.3.6.3). Probes were considered missing in a sample if they had detection *P* values >0.05 and were excluded from the analysis if they were missing in >5% of sample. Normalization to eliminate systematic dye bias in 2-channel probes was carried out using the methylumi default method. Following quality control and normalization, DNAm data for 828,613 CpGs were available for $n = 595$ samples (baseline $n = 214$; 12 months $n = 193$; 24 months $n = 188$). Additional batch correction was performed by residualizing DNAm measurements for PCs estimated from array control-probe beta values[53]. Cell count estimation was performed using the Houseman equation via the *minfi* and *FlowSorted.Blood.EPIC* R packages[28,54].

### DNAm clocks and pace-of-aging measures

DNAm clocks are algorithms that combine information from DNAm measurements across the genome to quantify variation in biological age[55].

The first-generation DNAm clocks were developed from machine-learning analyses comparing samples from individuals of different chronological age. These clocks were highly accurate in predicting the chronological age of new samples and also showed some capacity for predicting differences in mortality risk, although effect sizes tend to be small and inconsistent across studies[56–58]. We analyzed the first-generation clocks proposed by Horvath (Horvath clock) and Hannum et al. (Hannum clock)[56,57].

The second-generation DNAm clocks were developed with the goal of improving quantification of biological aging by focusing on differences in mortality risk instead of on differences in chronological age[22,23]. These clocks also include an intermediate step in which DNAm data are fitted to physiological parameters. The second-generation clocks are more predictive of morbidity and mortality as compared with the first-generation clocks[59] and are proposed to have improved potential for testing impacts of interventions to slow aging[14]. We analyzed the second-generation clocks proposed by Levine et al. (PhenoAge clock) and Lu et al. (GrimAge clock)[22,23].

A limitation of several DNAm clocks is that when residualized for chronological age, values show only moderate test–retest reliability across technical replicates. Test–retest reliability is a critical feature of measurements used to evaluate the impact of intervention because change from preintervention to postintervention cannot be distinguished from technical noise unless reliability is high. To improve technical reliability, Higgins-Chen and colleagues developed a new computational method that retrained DNAm clocks using DNAm PCs[25]. The resulting 'PC clocks' demonstrate exceptional test–retest reliability across technical replicates.

A third generation of DNAm measures of aging are referred to as pace-of-aging measures. In contrast to first- and second-generation DNAm clocks, which aim to quantify how much aging has occurred up to the time of measurement, pace-of-aging measures aim to quantity how fast the process of aging-related deterioration of system integrity is proceeding. We analyzed the newest pace-of-aging measure, DunedinPACE, which is shorthand for 'Pace of Aging Computed from the Epigenome'[24]. DunedinPACE was developed by modeling within-individual multi-system physiological change across four timepoints in same-age individuals in the Dunedin Study 1972–1973 birth cohort[60,61], when participants were aged 26, 32, 38 and 45 yr. DunedinPACE was developed from analysis of a pace-of-aging composite of slopes of aging-related change in the following physiological measures: ApoB100/ApoA1 ratio, BMI, blood urea nitrogen, high-sensitivity C-reactive protein, cardiorespiratory fitness, dental caries experience, total cholesterol, forced expiratory volume in 1 second, forced expiratory volume in 1 second/fixed vital capacity ratio, estimated glomerular filtration rate, hemoglobin A1C, high-density lipoprotein cholesterol, leptin, lipoprotein(a), mean arterial pressure, mean periodontal attachment loss, triglycerides, waist-to-hip ratio and white blood cell count. Slopes of change were estimated from four repeated measurements collected over a period of two decades. This physiological pace-of-aging composite is described in detail in ref. [61]. The DunedinPACE DNAm algorithm was derived from elastic net regression of the physiological pace-of-aging composite on Illumina EPIC array DNAm data derived from blood samples collected at the age 45 follow-up assessment. The set of CpG sites included in the DNAm dataset used to develop the DunedinPACE algorithm was restricted to those showing acceptable test–retest reliability as determined in the analysis in ref. [62]. The DunedinPACE DNAm algorithm is described in detail in ref. [24].

Our primary analysis focused on the PC versions of the PhenoAge and GrimAge second-generation clocks and DunedinPACE, all of which show exceptional test–retest reliability in technical replicates. We report results for both original and PC versions of DNAm clocks in the Supplementary Information.

### Analysis

Analysis included all participants with available DNAm data at trial baseline and at least one follow-up timepoint.

We computed change scores for all aging measures by comparing values at the 12-month and 24-month follow-up assessments with baseline values (that is, 12-month change = 12-month value − baseline; 24-month change = 24-month value − baseline). We conducted analyses of these change scores to test the hypothesis that CR slows biological aging using two complementary approaches: (1) we conducted ITT analysis which compared change scores between participants randomized to CR intervention and the AL control group; (2) we conducted TOT analysis using IV methods to estimate the effect of CR on change scores.

In ITT analysis, we tested the effect of randomization to CR versus AL on aging measure change scores using repeated-measures ANCOVA implemented under mixed models, following the approach used in past CALERIE analysis[26]. The model included terms for treatment condition (CR or AL), follow-up time, an interaction term modeling heterogeneity in the

treatment effect between the 12- and 24-month follow-ups, the baseline level of the aging measure and the following pretreatment covariates: chronological age, sex, race/ethnicity (Black, White, Other), BMI stratum at randomization (normal weight (22.0–24.9 kg m$^{-2}$) and overweight (25.0–27.9 kg m$^{-2}$)) and study site. Models were fitted using the Stata software's 'mixed' command. Details of estimation and calculation of confidence intervals are reported in Stata's documentation of the command[63].

In TOT analysis, we tested the effect of the CR intervention on aging measure change scores using IV regression implemented using a two-stage least squares approach[64]. The first-stage regression modeled CR treatment dose as a function of randomization condition (CR versus AL) and pretreatment characteristics (chronological age, sex, race/ethnicity, BMI, study site and baseline value of the biological aging measure). The model instruments were randomization condition and interactions of randomization condition with sex and pretreatment values of BMI and the biological aging measure. The second-stage regression modeled aging measure change scores as a function of the CR treatment dose estimated from the first-stage regression and pretreatment covariates. Separate models were fitted for the 12- and 24-month follow-ups. IV regression models were fitted using the Stata 16.0 software's 'ivregress' command. Details of estimation and calculation of confidence intervals are reported in Stata's documentation of the command[65]. TOT models are described in detail below.

In ITT and TOT analyses, effect sizes were scaled in standardized units according to the distribution of the aging measures at pretreatment baseline. For the DNAm clocks, clock ages were differenced from chronological ages and standard deviations for these age-difference values were used for scaling. For DunedinPACE, the standard deviations of the original values were used for scaling. Treatment effects denominated in these standardized units are interpreted as Cohen's *d*.

### Specification of TOT regression models

We tested TOT effects using two-stage least squares IV regression. IV regression is a method commonly used to reduce the impact of confounding in association analysis. It can also be applied to account for contamination/nonadherence in randomized trials[64]. Under conditions of nonadherence, traditional ITT analysis can result in a biased estimate of the treatment effect and an IV estimator can provide a complement[66]. In CALERIE, adherence was imperfect; the average CR achieved in the treatment group was roughly half the prescribed dose of 25% (ref. [10]). The ITT estimate may therefore underestimate the effect of CR on biological aging.

In our analysis, we used IV regression to estimate the effect of 20% CR on change in measures of biological aging. We focused on a CR dose of 20% instead of the 25% dose prescribed in the trial because few individuals achieved 25% CR, especially through the 24-month follow-up. The 20% CR level represented the 75th percentile of the treatment group CR distribution at 12-month follow-up and the 87th percentile of the treatment group CR distribution at 24-month follow-up.

The IV approach we used involved two related regressions. The first regression modeled observed treatment dose (%CR relative to baseline) on pretreatment characteristics and the instrument of randomization condition. The second regression modeled the outcomes (changes in measures of biological aging) as functions of the predicted treatment dose estimated by the first regression and pretreatment covariates.

We developed our IV regression model by first modeling intervention group participants' achieved CR treatment dose as a function of pretreatment covariates: chronological age, sex, BMI, study site. We fitted a saturated regression model including interactions among all pretreatment characteristics and additional covariate adjustment for race/ethnicity, which was included only as a main effect. (Race/ethnicity was omitted from the interaction terms because there was insufficient site- and sex-specific variation in race/ethnicity to fit models.) This analysis identified sex, baseline BMI and their interaction as statistically significant predictors of CR dose at the alpha = 0.05 level.

Next, we parameterized our IV regression specifying the first stage to include the 'instruments' of intervention group and interactions of intervention group with sex, pretreatment BMI and a three-way interaction between intervention condition, sex and pretreatment BMI. The base first-stage regression took the form

$$\%\mathrm{CR}_t = a + \mathrm{CR} + \mathrm{CR} \times \mathrm{sex} + \mathrm{CR} \times \mathrm{BMI}_{\mathrm{baseline}} + \mathrm{CR} \times \mathrm{sex} \times \mathrm{BMI}_{\mathrm{baseline}} + X + e \quad (1)$$

in which $\%\mathrm{CR}_t$ is the %CR relative to baseline achieved at time $t$ (either 12- or 24-month follow-up), $\mathrm{BMI}_{\mathrm{baseline}}$ is pretreatment BMI, $X$ is a matrix of all pretreatment covariates, $a$ is a model intercept and $e$ is the error term. Results from this first-stage regression were then included in the second-stage model:

$$\mathrm{Delta\ BA}_t = a + \%\mathrm{CR}_t + \mathrm{X} + e \quad (2)$$

in which $\%\mathrm{CR}_t$ is %CR predicted from equation (1). For final TOT analysis, we included a further instrument in the first-stage regression consisting of the interaction between the baseline level of the aging measure and the CR treatment group. Sensitivity analysis involving re-estimating the IV regression models omitting this final instrument did not change results.

Supplementary Fig. 1 plots predicted values of %CR based on our base first-stage model (that is, the model in equation (1)).

### Statistics and reproducibility

We conducted new DNAm assays of stored blood biospecimens collected from the CALERIE Phase 2 randomized controlled trial and merged these data with existing secondary data from the trial. The assays of the biospecimens were conducted blind to the conditions of the trial. After baseline testing, $n = 220$ participants were randomly assigned at a ratio of 2:1 to a CR behavioral intervention or to an AL control group. Randomization was stratified by site, sex and BMI. A permuted block randomization technique was used. No statistical methods were used to predetermine sample sizes; we analyzed data from all participants for whom blood DNAm data were available at baseline and at least one follow-up timepoint ($N = 197$; CR $n = 128$, AL $n = 69$). Participants had mean age of 38 yr (s.d. = 7), 70% were women and 77% were white; there were no differences in age, sex or race/ethnicity between AL and CR at baseline (Table 1). Data met model assumptions. Normality of outcome variables was evaluated by visual inspection of distributions and the Shapiro–Wilk test[67]. Equality of variances was evaluated according to the tests proposed by Brown and Forsythe[68] and Markowski and Markowski[69]. Models used to test ITT and TOT effects were fitted with heteroskedasticity-robust standard errors. Normality of distribution of error terms was evaluated by visual inspection of histograms of residuals and the Shapiro–Wilk test.

### DNAm clocks

DNAm clock measures of aging are algorithms that estimate biological age, the state of an organism's biology represented as the age at which that state would be typical in a reference population. The clocks we analyzed were developed to predict mortality risk. The age values computed by the clock algorithms correspond to the age at which predicted mortality risk would be approximately normal in the reference population used to develop the clock. We computed clock values based on versions of the clock algorithms developed from DNAm PCs (sometimes referred to as 'PC clocks')[18,21].

**PhenoAge clock.** The PhenoAge clock was based on analysis of nine blood chemistry markers, age and mortality data from the US National Health and Nutrition Examination Surveys ($n = 9{,}926$ participants aged 18 yr and older; 23 yr of mortality follow-up); DNAm and blood chemistry data from the Invecchiare in Chianti (InCHIANTI) Study ($n = 912$

participants aged 21–100 yr); and the US Health and Retirement Study ($n$ = 3,593 participants aged 51–100 yr)[19].

**GrimAge clock.** The GrimAge clock was based on analysis of eight plasma protein markers, smoking pack years, age, sex and mortality data from the Framingham Heart Study Offspring and Gen3 Cohorts ($n$ = 2,751 participants aged 24–92 yr)[47–49].

### Pace of aging

Pace-of-aging measures estimate the rate of biological aging, defined as the rate of decline in overall system integrity. Pace-of-aging values correspond to the years of biological aging experienced during a single calendar year. A value of 1 represents the typical pace of aging in a reference population; values above 1 indicate faster pace of aging; values below 1 indicate slower pace of aging.

**DunedinPACE.** Based on analysis of pace of aging in the Dunedin Study ($n$ = 817 participants examined at ages 26, 32, 38 and 45 yr)[24], pace of aging was measured from within-person change over time in 19 blood chemistry and organ function test metrics of system integrity[24]. DNAm was measured at age 45 yr.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All data, including DNA methylation data, are available for academic research purposes from the CALERIE Biorepository: https://calerie.duke.edu. Instructions for applying for data access are detailed at https://calerie.duke.edu/samples-data-access-and-analysis. Applications for some types of data may require IRB oversight. Guidelines are available at https://calerie.duke.edu/sites/default/files/2022-08/calerie_ancillary_study_guidelines_revised_042921.pdf. Registration to obtain an application form can be completed at https://calerie.duke.edu/database-submission-form. Phenotypic data used in the primary analysis were obtained from the analysis datasets 'subject1', 'visits', 'ivrsrand', 'clwtvis' and 'pctcr'. Additional data were obtained from 'teerq', 'rmrresid', 'vitalsa', 'oclabflt'. Source data for Fig. 2 and Supplementary Fig. 3 are provided in Supplementary Information.

## Code availability

Code is available from GitHub: https://github.com/danbelsky/CALERIE_ClocksAnalysis.

## References


1. Kaeberlein, M. Longevity and aging. *F1000Prime Rep.* **5**, 5 (2013).
2. Kennedy, B. K. et al. Geroscience: linking aging to chronic disease. *Cell* **159**, 709–713 (2014).
3. Campisi, J. et al. From discoveries in ageing research to therapeutics for healthy ageing. *Nature* **571**, 183–192 (2019).
4. Speakman, J. R. & Mitchell, S. E. Caloric restriction. *Mol. Aspects Med.* **32**, 159–221 (2011).
5. Maegawa, S. et al. Caloric restriction delays age-related methylation drift. *Nat. Commun.* **8**, 539 (2017).
6. Hahn, O. et al. Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biol.* **18**, 56 (2017).
7. Petkovich, D. A. et al. Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metab.* **25**, 954–960.e6 (2017).
8. Anderson, R. M. & Weindruch, R. The caloric restriction paradigm: implications for healthy human aging. *Am. J. Hum. Biol.* **24**, 101–106 (2012).
9. Mattison, J. A. et al. Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. *Nature* **489**, 318–321 (2012).
10. Ravussin, E. et al. A 2-year randomized controlled trial of human caloric restriction: feasibility and effects on predictors of health span and longevity. *J. Gerontol. A* **70**, 1097–1104 (2015).
11. Scott, A. J., Ellison, M. & Sinclair, D. A. The economic value of targeting aging. *Nat. Aging* **1**, 616–623 (2021).
12. Kaeberlein, M. It is time to embrace 21st-century medicine. *Public Policy Aging Rep.* **29**, 111–115 (2019).
13. Goldman, D. P. et al. Substantial health and economic returns from delayed aging may warrant a new focus for medical research. *Health Aff.* **32**, 1698–1705 (2013).
14. Fahy, G. M. et al. Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, e13028 (2019).
15. Chen, L. et al. Effects of vitamin D3 supplementation on epigenetic aging in overweight and obese African Americans with suboptimal vitamin D status: a randomized clinical trial. *J. Gerontol. A* **74**, 91–98 (2019).
16. Sae-Lee, C. et al. Dietary intervention modifies DNA methylation age assessed by the epigenetic clock. *Mol. Nutr. Food Res.* **62**, 1800092 (2018).
17. Colchero, F. et al. The long lives of primates and the 'invariant rate of ageing' hypothesis. *Nat. Commun.* **12**, 3666 (2021).
18. Fleming, T. R. & Powers, J. H. Biomarkers and surrogate endpoints in clinical trials. *Stat. Med.* **31**, 2973–2984 (2012).
19. Prentice, R. L. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat. Med.* **8**, 431–440 (1989).
20. Justice, J. N. et al. A framework for selection of blood-based biomarkers for geroscience-guided clinical trials: report from the TAME Biomarkers Workgroup. *GeroScience* **40**, 419–436 (2018).
21. Racette, S. B. et al. Approaches for quantifying energy intake and %calorie restriction during calorie restriction interventions in humans: the multicenter CALERIE study. *Am. J. Physiol. Endocrinol. Metab.* **302**, E441–E448 (2012).
22. Levine, M. E. et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
23. Lu, A. T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
24. Belsky, D. W. et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11**, e73420 (2022).
25. Higgins-Chen, A. T. et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. *Nat. Aging* **2**, 644–661 (2022).
26. Kraus, W. E. et al. 2 years of calorie restriction and cardiometabolic risk (CALERIE): exploratory outcomes of a multicentre, phase 2, randomised controlled trial. *Lancet Diabetes Endocrinol.* **7**, 673–683 (2019).
27. Benjamin, D. J. et al. Redefine statistical significance. *Nat. Hum. Behav.* **2**, 6–10 (2018).
28. Salas, L. A. et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* **19**, 64 (2018).
29. Sugden, K. et al. Association of pace of aging measured by blood-based DNA methylation with age-related cognitive impairment and dementia. *Neurology* **99**, e1402–e1413 (2022).
30. Hillary, R. F. et al. Epigenetic measures of ageing predict the prevalence and incidence of leading causes of death and disease burden. *Clin. Epigenetics* **12**, 115 (2020).
31. Belsky, D. W., Huffman, K. M., Pieper, C. F., Shalev, I. & Kraus, W. E. Change in the rate of biological aging in response to caloric restriction: CALERIE Biobank Analysis. *J. Gerontol. A* **73**, 4–10 (2017).

32. Kwon, D. & Belsky, D. W. A toolkit for quantification of biological age from blood chemistry and organ function test data: BioAge. *Geroscience* **43**, 2795–2808 (2021).
33. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
34. Spadaro, O. et al. Caloric restriction in humans reveals immunometabolic regulators of health span. *Science* **375**, 671–677 (2022).
35. Redman, L. M. et al. Metabolic slowing and reduced oxidative damage with sustained caloric restriction support the rate of living and oxidative damage theories of aging. *Cell Metab.* **27**, 805–815.e4 (2018).
36. Anthonisen, N. R. et al. The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical trial. *Ann. Intern. Med.* **142**, 233–239 (2005).
37. Ferrucci, L. et al. Measuring biological aging in humans: a quest. *Aging Cell* **19**, e13080 (2020).
38. Kritchevsky, S. B. & Justice, J. N. Testing the geroscience hypothesis: early days. *J. Gerontol. A* **75**, 99–101 (2020).
39. Bell, C. G. et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* **20**, 249 (2019).
40. Belsky, D. W. et al. Eleven telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: do they measure the same thing? *Am. J. Epidemiol.* **187**, 1220–1230 (2018).
41. Ahadi, S. et al. Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nat. Med.* **26**, 83–90 (2020).
42. Ma, S. et al. Caloric restriction reprograms the single-cell transcriptional landscape of *Rattus norvegicus* aging. *Cell* **180**, 984–1001.e22 (2020).
43. Dorling, J. L., Martin, C. K. & Redman, L. M. Calorie restriction for enhanced longevity: the role of novel dietary strategies in the present obesogenic environment. *Ageing Res. Rev.* **64**, 101038 (2020).
44. Das, S. K. et al. Body-composition changes in the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE)-2 study: a 2-y randomized controlled trial of calorie restriction in nonobese humans. *Am. J. Clin. Nutr.* **105**, 913–927 (2017).
45. Shen, W. et al. Effect of 2-year caloric restriction on organ and tissue size in nonobese 21- to 50-year-old adults in a randomized clinical trial: the CALERIE study. *Am. J. Clin. Nutr.* **114**, 1295–1303 (2021).
46. Moffitt, T. E., Belsky, D. W., Danese, A., Poulton, R. & Caspi, A. The longitudinal study of aging in human young adults: knowledge gaps and research agenda. *J. Gerontol. A* **72**, 210–215 (2017).
47. Sierra, F. et al. Moving geroscience from the bench to clinical care and health policy. *J. Am. Geriatr. Soc.* **69**, 2455–2463 (2021).
48. Justice, J. et al. Frameworks for proof-of-concept clinical trials of interventions that target fundamental aging processes. *J. Gerontol. A Biol. Sci. Med. Sci.* **71**, 1415–1423 (2016).
49. Justice, J. N. & Kritchevsky, S. B. Putting epigenetic biomarkers to the test for clinical trials. *eLife* **9**, e58592 (2020).
50. Longo, V. D. et al. Intermittent and periodic fasting, longevity and disease. *Nat. Aging* **1**, 47–59 (2021).
51. Davis, S. et al. methylumi: handle Illumina methylation data. R package version 2.32.0 (2021).
52. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
53. Lehne, B. et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* **16**, 37 (2015).
54. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
55. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
56. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
57. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
58. Chen, B. H. et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* **8**, 1844–1865 (2016).
59. Levine, M. E. Assessment of epigenetic clocks as biomarkers of aging in basic and population research. *J. Gerontol. A* **75**, 463–465 (2020).
60. Belsky, D. W. et al. Quantification of biological aging in young adults. *Proc. Natl Acad. Sci. USA* **112**, E4104–E4110 (2015).
61. Elliott, M. L. et al. 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 (2021).
62. Sugden, K. et al. Patterns of reliability: assessing the reproducibility and integrity of DNA methylation measurement. *Patterns* **1**, 100014 (2020).
63. *Stata Multilevel Mixed-Effects Reference Manual* (StataCorp, 2013).
64. Sussman, J. B. & Hayward, R. A. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. *Brit. Med. J.* **340**, c2073 (2010).
65. Baum, C. F., Schaffer, M. E. & Stillman, S. Instrumental variables and GMM: estimation and testing. *Stata J.* **3**, 1–31 (2003).
66. Bang, H. & Davis, C. E. On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? *Stat. Med.* **26**, 954–964 (2007).
67. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965).
68. Brown, M. B. & Forsythe, A. B. Robust tests for the equality of variances. *J. Am. Stat. Assoc.* **69**, 364–367 (1974).
69. Markowski, C. A. & Markowski, E. P. Conditions for the effectiveness of a preliminary test of variance. *Am. Stat.* **44**, 322–326 (1990).


## Acknowledgements

'CALERIE' is a registered trademark. This research was supported by grant no. R01AG061378 (to D.W.B.). R.W. received additional support from the American Brain Foundation (to R.W.). C.F.P. and V.B.K. received additional support from grant no. P30AG028716 (to C.F.P. and V.B.K.); V.B.K. received additional support from grant no. R01AG054840 (to V.B.K.); K.M.H., W.E.K. and C.F.P. received additional support from grant no. R33AG070455 (to K.M.H., W.E.K. and C.F.P.); M.S.K. received additional support from the CIHR grant no. RN439810 (to M.S.K.); S.B.R., M.B. and S.K.D. received support from grant no. R01 AG071717 (to S.B.R., M.B. and S.K.D.); I.S. and W.J.H. received support from grant no. R03AG071549 (to I.S.). B.S. received support from grant no. U01AG060906 (to B.S.). We thank the CALERIE Research Network no. R33AG070455 for their assistance in this project and the Dunedin Study no. R01AG032282 for facilitating early access to the DunedinPACE DNA methylation algorithm. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper. R.W. completed work on this project while affiliated with the Butler Columbia Aging Center. She is now in the Department of Neurology at the Columbia University Irving Medical Center.

## Author contributions

D.W.B. designed the research. D.W.B., R.W., C.P.R., D.L.C., K.M.H., M.S.K., M. Kothari, G.H.G., V.B.K., W.E.K., D.T.S.L., C.F.P., M.E.R., M.B., S.K.D., L.F., W.J.H., M. Kebbe, D.C.P., S.B.R., I.S. and B.S. conducted

the research. C.P.R., D.L.C., M.S.K. and D.T.S.L. prepared the DNA methylation datasets. D.W.B., R.W. and C.P.R. analyzed the data. D.W.B. and R.W. wrote the first draft of the paper. D.W.B. and C.P.R. wrote the revised draft of the paper. All authors contributed critical review of the paper.

## Competing interests

D.L.C. and D.W.B. are listed as inventors on a Duke University and University of Otago invention, DunedinPACE, that was licensed to a commercial entity. The other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s43587-022-00357-y.

**Correspondence and requests for materials** should be addressed to D. W. Belsky.

**Peer review information** *Nature Aging* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023

[1]Butler Columbia Aging Center, Columbia University Mailman School of Public Health, New York, NY, USA. [2]Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA. [3]Duke Molecular Physiology Institute and Department of Medicine, Duke University School of Medicine, Durham, NC, USA. [4]Department of Medical Genetics, Edwin S.H. Leong Healthy Aging Program, Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, British Columbia, Canada. [5]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA. [6]Center on Aging and Development, Biostatistics and Bioinformatics, Duke University, Durham, NC, USA. [7]Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, USA. [8]Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA. [9]Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. [10]Department of Biobehavioral Health, Pennsylvania State University, State College, PA, USA. [11]Pennington Biomedical Research Center, Baton Rouge, LA, USA. [12]Department of Medicine, Duke University School of Medicine, Durham, NC, USA. [13]Program in Physical Therapy and Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA. [14]College of Health Solutions, Arizona State University, Phoenix, AZ, USA. [15]Buck Institute for Research on Aging, Novato, CA, USA. [16]These authors contributed equally: R. Waziry, C. P. Ryan. ✉e-mail: Daniel.Belsky@Columbia.edu

## SUPPLEMENTARY TABLES

**Table S1. DNA-methylation clock and Pace of Aging values for CALERIE™ participants in the Ad Libitum (AL) and Caloric Restriction (CR) groups at baseline, 12-month, and 24-month assessments.** Panel A shows values for original and PC-based versions[18] of first-generation clocks. Panel B shows values for original and PC-based versions of second-generation clocks. Panel C shows values for DunedinPACE. Values for DNA methylation clocks are age-difference scores computed by subtracting participants' chronological ages from their clock ages. DunedinPACE values are not adjusted for chronological age. 95% Confidence intervals are reported in brackets. Follow-up DNA methylation data for the n=69 AL and n=128 CR participants in the analysis sample were available for n=65 AL and n=125 CR participants at 12-months and n=68 AL and 117 CR participants at 24-months. The positive offset observed for the DNAm clocks is a common feature of observations made using EPIC arrays[27].

**Panel A. First-Generation Clocks**

| | | Ad Libitum (n=69) | | Calorie Restriction (n=128) | |
|---|---|---|---|---|---|
| | | Mean | 95% CI | Mean | 95% CI |
| Horvath Clock | Baseline | -2.08 | [-2.94 , -1.22] | -2.06 | [-2.75 , -1.38] |
| | 12mo | -2.44 | [-3.37 , -1.51] | -2.09 | [-2.79 , -1.39] |
| | 24mo | -2.38 | [-3.28 , -1.48] | -2.16 | [-2.89 , -1.43] |
| PC Horvath Clock | Baseline | 9.06 | [8.07 , 10.04] | 9.20 | [8.50 , 9.90] |
| | 12mo | 8.71 | [7.66 , 9.77] | 9.02 | [8.34 , 9.71] |
| | 24mo | 8.58 | [7.55 , 9.61] | 8.74 | [8.00 , 9.49] |
| Hannum Clock | Baseline | 3.87 | [3.22 , 4.52] | 4.17 | [3.67 , 4.67] |
| | 12mo | 3.59 | [2.81 , 4.36] | 3.98 | [3.45 , 4.50] |
| | 24mo | 3.75 | [3.05 , 4.46] | 3.98 | [3.43 , 4.52] |
| PC Hannum Clock | Baseline | 13.44 | [12.58 , 14.29] | 13.64 | [13.04 , 14.24] |
| | 12mo | 13.24 | [12.28 , 14.19] | 13.46 | [12.86 , 14.06] |
| | 24mo | 13.11 | [12.22 , 14.00] | 13.29 | [12.64 , 13.94] |
| Skin & Blood Clock | Baseline | 1.39 | [0.67 , 2.12] | 1.60 | [1.13 , 2.06] |
| | 12mo | 1.73 | [1.03 , 2.44] | 1.69 | [1.24 , 2.15] |
| | 24mo | 1.92 | [1.16 , 2.69] | 1.68 | [1.17 , 2.19] |
| PC Skin & Blood Clock | Baseline | 6.44 | [5.34 , 7.53] | 6.92 | [6.16 , 7.69] |
| | 12mo | 6.00 | [4.80 , 7.19] | 6.69 | [5.95 , 7.44] |
| | 24mo | 5.71 | [4.57 , 6.85] | 6.31 | [5.50 , 7.12] |

SUPPLEMENTARY TABLES

## Table S1 Continued

**Panel B. Second-Generation Clocks**

| | | Ad Libitum (n=69) | | Calorie Restriction (n=128) | |
|---|---|---|---|---|---|
| | | Mean | 95% CI | Mean | 95% CI |
| PhenoAge Clock | Baseline | -10.39 | [-11.53 , -9.24] | -10.14 | [-11.03 , -9.24] |
| | 12mo | -10.59 | [-11.82 , -9.35] | -10.10 | [-11.04 , -9.16] |
| | 24mo | -10.99 | [-12.12 , -9.86] | -10.12 | [-11.00 , -9.24] |
| PC PhenoAge Clock | Baseline | 2.48 | [1.36 , 3.61] | 2.66 | [1.81 , 3.51] |
| | 12mo | 2.40 | [1.24 , 3.55] | 2.36 | [1.53 , 3.19] |
| | 24mo | 2.23 | [1.12 , 3.35] | 2.51 | [1.67 , 3.36] |
| GrimAge Clock | Baseline | 4.85 | [4.14 , 5.56] | 4.78 | [4.32 , 5.24] |
| | 12mo | 4.63 | [3.93 , 5.33] | 4.33 | [3.88 , 4.78] |
| | 24mo | 4.25 | [3.61 , 4.90] | 4.06 | [3.59 , 4.54] |
| PC GrimAge Clock | Baseline | 11.51 | [10.88 , 12.14] | 11.85 | [11.34 , 12.35] |
| | 12mo | 11.23 | [10.53 , 11.93] | 11.42 | [10.92 , 11.92] |
| | 24mo | 10.79 | [10.16 , 11.43] | 11.08 | [10.56 , 11.60] |

**Panel C. Pace of Aging**

| | | | | | |
|---|---|---|---|---|---|
| Dunedin PACE | Baseline | 0.94 | [0.92 , 0.96] | 0.95 | [0.93 , 0.96] |
| | 12mo | 0.95 | [0.93 , 0.97] | 0.93 | [0.92 , 0.94] |
| | 24mo | 0.95 | [0.93 , 0.97] | 0.93 | [0.92 , 0.95] |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text.

SUPPLEMENTARY TABLES

**Table S2. Technical reliability and within-individual stability of DNAm clocks and DunedinPACE in GSE55763 and the CALERIE™ trial.** Panel A shows values for original and PC-based versions[18] of first-generation clocks. Panel B shows values for original and PC-based versions of second-generation clocks. Panel C shows values for DunedinPACE. The left side of the table reports technical (test-retest) reliability across technical replicates from n=36 samples measured in duplicate in GSE55763. The right side of the table reports within-individual stability across repeated measurements at baseline, 12months, and 24 months in the CALERIE™ trial. Reliabilities are calculated as intra-class correlation coefficients (ICCs). Pearson r correlation are also reported.

| | **Technical Reliability (GSE55763)** | | | **Within-Individual Stability (CALERIE™)** | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | AL | | | CR | |
| | ICC | 95% CI | | **r** | ICC | 95% CI | **r** | ICC | 95% CI |
| **Panel A. First-Generation Clocks** | | | | | | | | | |
| Horvath Clock | 0.824 | [0.68 , 0.97] | Bl-12mo | 0.79 | 0.78 | [-0.62 , 2.29] | 0.78 | 0.79 | [0.71 , 0.85] |
| | | | Bl-24mo | 0.74 | 0.71 | [0.57 , 0.81] | 0.82 | 0.83 | [0.77 , 0.88] |
| PC Horvath Clock | 0.969 | [0.95 , 0.99] | Bl-12mo | 0.97 | 0.97 | [0.95 , 0.98] | 0.96 | 0.96 | [0.94 , 0.97] |
| | | | Bl-24mo | 0.97 | 0.96 | [0.93 , 0.97] | 0.97 | 0.96 | [0.94 , 0.97] |
| Hannum Clock | 0.853 | [0.74 , 0.96] | Bl-12mo | 0.84 | 0.83 | [0.74 , 0.89] | 0.79 | 0.78 | [0.71 , 0.84] |
| | | | Bl-24mo | 0.82 | 0.82 | [0.72 , 0.88] | 0.80 | 0.79 | [0.72 , 0.85] |
| PC Hannum Clock | 0.986 | [0.98 , 0.99] | Bl-12mo | 0.95 | 0.95 | [0.91 , 0.97] | 0.91 | 0.91 | [0.88 , 0.94] |
| | | | Bl-24mo | 0.92 | 0.92 | [0.87 , 0.95] | 0.93 | 0.93 | [0.90 , 0.95] |
| Skin & Blood Clock | 0.882 | [0.75 , 1.01] | Bl-12mo | 0.83 | 0.83 | [0.73 , 0.89] | 0.76 | 0.76 | [0.68 , 0.83] |
| | | | Bl-24mo | 0.82 | 0.81 | [0.71 , 0.88] | 0.85 | 0.85 | [0.79 , 0.89] |
| PC Skin & Blood Clock | 0.978 | [0.96 , 0.99] | Bl-12mo | 0.96 | 0.96 | [0.93 , 0.97] | 0.95 | 0.95 | [0.93 , 0.96] |
| | | | Bl-24mo | 0.95 | 0.94 | [0.90 , 0.96] | 0.96 | 0.95 | [0.93 , 0.97] |
| **Panel B. Second-Generation Clocks** | | | | | | | | | |
| PhenoAge Clock | 0.727 | [0.55 , 0.91] | Bl-12mo | 0.74 | 0.75 | [0.62 , 0.84] | 0.74 | 0.74 | [0.65 , 0.81] |
| | | | Bl-24mo | 0.62 | 0.62 | [0.45 , 0.75] | 0.81 | 0.81 | [0.74 , 0.87] |
| PC PhenoAge Clock | 0.987 | [0.98 , 1.00] | Bl-12mo | 0.85 | 0.85 | [0.77 , 0.91] | 0.82 | 0.82 | [0.76 , 0.87] |
| | | | Bl-24mo | 0.79 | 0.80 | [0.69 , 0.87] | 0.89 | 0.89 | [0.84 , 0.92] |
| GrimAge Clock | 0.96 | [0.93 , 0.99] | Bl-12mo | 0.91 | 0.92 | [0.87 , 0.95] | 0.83 | 0.80 | [0.72 , 0.85] |
| | | | Bl-24mo | 0.87 | 0.85 | [0.77 , 0.91] | 0.86 | 0.80 | [0.72 , 0.86] |
| PC GrimAge Clock | 0.988 | [0.97 , 1.00] | Bl-12mo | 0.93 | 0.91 | [0.86 , 0.95] | 0.91 | 0.86 | [0.81 , 0.90] |
| | | | Bl-24mo | 0.87 | 0.83 | [0.74 , 0.89] | 0.93 | 0.88 | [0.83 , 0.92] |
| **Panel C. Pace of Aging** | | | | | | | | | |
| DunedinPACE | 0.964 | [0.93 , 1.00] | Bl-12mo | 0.85 | 0.84 | [0.75 , 0.90] | 0.77 | 0.75 | [0.67 , 0.82] |
| | | | Bl-24mo | 0.84 | 0.83 | [0.74 , 0.89] | 0.80 | 0.79 | [0.71 , 0.85] |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text.

SUPPLEMENTARY TABLES

**Table S3. Change scores for DNA-methylation measures of aging at 12-and 24-month follow-ups in CALERIE™-Trial Ad Libitum and Caloric Restriction group participants.** Panel A shows change scores for original and PC-based versions[18] of first-generation clocks. Panel B shows change scores for original and PC-based versions of second-generation clocks. Panel C shows change scores for DunedinPACE. The table shows change-scores computed relative to baseline for the aging measures at 12- and 24-month follow-ups in the AL and CR groups. Change scores are scaled in units of the aging measures' standard deviations at pre-treatment baseline. Change scores can be back-converted into the original units of the measures by multiplying by the standard deviation. The standard deviations are as follows: Horvath Clock = 3.84; PC Horvath Clock = 4.07; Hannum Clock = 2.85; PC Hannum Clock = 3.50; Skin & Blood Clock = 2.81; PC Skin & Blood Clock = 4.46; PhenoAge Clock = 5.04; PC PhenoAge Clock = 4.84; GrimAge Clock = 2.76; PC GrimAge Clock = 2.82; DunedinPACE = 0.09. For the clocks, change scores are computed from age-difference values (i.e. clock age – chronological age) at baseline and the 12- and 24-month follow-ups.

| | | Ad Libitum (n=69) | | Calorie Restriction (n=128) | |
|---|---|---|---|---|---|
| | | Mean | 95% CI | Mean | 95% CI |
| **Panel A. First-Generation Clocks** | | | | | |
| Horvath Clock | 12mo | -0.08 | [-0.23 , 0.07] | -0.02 | [-0.14 , 0.10] |
| | 24mo | -0.10 | [-0.27 , 0.07] | -0.06 | [-0.17 , 0.05] |
| PC Horvath Clock | 12mo | -0.08 | [-0.14 , -0.02] | -0.07 | [-0.12 , -0.02] |
| | 24mo | -0.15 | [-0.21 , -0.08] | -0.11 | [-0.15 , -0.06] |
| Hannum Clock | 12mo | -0.11 | [-0.25 , 0.04] | -0.09 | [-0.21 , 0.03] |
| | 24mo | -0.06 | [-0.21 , 0.09] | -0.06 | [-0.18 , 0.06] |
| PC Hannum Clock | 12mo | -0.07 | [-0.15 , 0.02] | -0.07 | [-0.14 , 0.01] |
| | 24mo | -0.13 | [-0.23 , -0.03] | -0.09 | [-0.16 , -0.02] |
| Skin & Blood Clock | 12mo | 0.14 | [-0.01 , 0.29] | 0.03 | [-0.08 , 0.14] |
| | 24mo | 0.17 | [0.01 , 0.33] | 0.00 | [-0.09 , 0.10] |
| PC Skin & Blood Clock | 12mo | -0.10 | [-0.17 , -0.02] | -0.08 | [-0.14 , -0.03] |
| | 24mo | -0.19 | [-0.26 , -0.11] | -0.12 | [-0.17 , -0.07] |
| **Panel B. Second-Generation Clocks** | | | | | |
| PhenoAge Clock | 12mo | 0.00 | [-0.17 , 0.17] | 0.00 | [-0.14 , 0.13] |
| | 24mo | -0.14 | [-0.34 , 0.06] | 0.03 | [-0.09 , 0.14] |
| PC PhenoAge Clock | 12mo | -0.02 | [-0.15 , 0.11] | -0.06 | [-0.17 , 0.04] |
| | 24mo | -0.08 | [-0.22 , 0.07] | -0.02 | [-0.11 , 0.06] |
| GrimAge Clock | 12mo | -0.07 | [-0.18 , 0.04] | -0.19 | [-0.28 , -0.09] |
| | 24mo | -0.23 | [-0.36 , -0.11] | -0.23 | [-0.33 , -0.14] |
| PC GrimAge Clock | 12mo | -0.09 | [-0.19 , 0.00] | -0.17 | [-0.24 , -0.09] |
| | 24mo | -0.26 | [-0.37 , -0.14] | -0.22 | [-0.29 , -0.15] |
| **Panel C. Pace of Aging** | | | | | |
| DunedinPACE | 12mo | 0.13 | [-0.01 , 0.27] | -0.18 | [-0.29 , -0.06] |
| | 24mo | 0.12 | [-0.01 , 0.26] | -0.12 | [-0.23 , -0.01] |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text.

SUPPLEMENTARY TABLES

**Table S4. Effect-size estimates for intent-to-treat (ITT) and effect of treatment-on-the-treated (TOT) analysis testing the effect of CALERIE™ intervention on DNA methylation measures of aging.** Panel A shows treatment effects for original and PC-based versions[18] of first-generation clocks. Panel B shows treatment effects for original and PC-based versions of second-generation clocks. Panel C shows treatment effects for DunedinPACE. Intent to treat (ITT) effect estimates are reported on the left side of the table. Effect of treatment-on-the-treated (TOT) effect estimates are reported on the right side of the table. ITT analysis was conducted using mixed models to compare change-scores for DNA methylation measures of aging between CR and AL groups. Models included covariates for sex, race/ethnicity, study site, and pre-treatment baseline chronological age and body-mass index stratum. TOT analysis was conducted using instrumental variables regression to test associations of CR dose with aging-measure change-scores. Instruments included treatment group and interactions between treatment group and sex, body-mass index, and baseline levels of the aging measures. Model covariates were the same as in ITT analysis. TOT effect estimates are reported for a 20% dose of CR. This level of CR represents the 75$^{th}$ percentile of the treatment group CR distribution at 12-months follow-up and the 87$^{th}$ percentile of the treatment group CR distribution at 24-months follow-up. Effect-sizes are scaled in terms of standard deviations of the measures at pre-intervention baseline so they can be interpreted on a Cohen's d scale. Effect-sizes can be back-converted into the original units of the measures by multiplying the Cohen's d effect-size by the standard deviation. All tests reported were two-sided. The standard deviations are as follows: Horvath Clock = 3.84; PC Horvath Clock = 4.07; Hannum Clock = 2.85; PC Hannum Clock = 3.50; Skin & Blood Clock = 2.81; PC Skin & Blood Clock = 4.46; PhenoAge Clock = 5.04; PC PhenoAge Clock = 4.84; GrimAge Clock = 2.76; PC GrimAge Clock = 2.82; DunedinPACE = 0.09.

**Panel A. First Generation Clocks**

| | | Intent-to-Treat (ITT): Difference between CR and AL | | | Effect of Treatment on the Treated (TOT): Estimated Effect of 20% CR | | |
|---|---|---|---|---|---|---|---|
| | | Treatment Effect (Cohen's d) | 95% CI | p-value | Treatment Effect (Cohen's d) | 95% CI | p-value |
| Horvath | 12mo | 0.07 | [-0.11 , 0.25] | 0.457 | 0.14 | [-0.10 , 0.39] | 0.249 |
| Clock | 24mo | 0.05 | [-0.14 , 0.24] | 0.599 | 0.10 | [-0.22 , 0.42] | 0.526 |
| PC Horvath | 12mo | 0.01 | [-0.06 , 0.09] | 0.771 | 0.05 | [-0.06 , 0.15] | 0.375 |
| Clock | 24mo | 0.04 | [-0.04 , 0.12] | 0.341 | 0.05 | [-0.08 , 0.19] | 0.438 |
| Hannum | 12mo | 0.03 | [-0.15 , 0.21] | 0.758 | 0.09 | [-0.15 , 0.33] | 0.484 |
| Clock | 24mo | 0.00 | [-0.18 , 0.18] | 0.986 | -0.02 | [-0.35 , 0.31] | 0.909 |
| PC Hannum | 12mo | 0.00 | [-0.11 , 0.11] | 0.964 | 0.02 | [-0.14 , 0.17] | 0.845 |
| Clock | 24mo | 0.04 | [-0.08 , 0.15] | 0.501 | 0.03 | [-0.16 , 0.23] | 0.753 |
| Skin & Blood | 12mo | -0.09 | [-0.26 , 0.08] | 0.316 | -0.06 | [-0.30 , 0.18] | 0.618 |
| Clock | 24mo | -0.14 | [-0.32 , 0.04] | 0.139 | -0.19 | [-0.49 , 0.10] | 0.197 |
| PC Skin & | 12mo | 0.02 | [-0.07 , 0.11] | 0.647 | 0.05 | [-0.07 , 0.17] | 0.390 |
| Blood Clock | 24mo | 0.06 | [-0.02 , 0.15] | 0.156 | 0.07 | [-0.07 , 0.22] | 0.314 |

## Table S4 Continued

**Panel B. Second Generation Clocks**

| | | Intent-to-Treat (ITT): Difference between CR and AL | | | Effect of Treatment on the Treated (TOT): Estimated Effect of 20% CR | | |
|---|---|---|---|---|---|---|---|
| | | Effect (Cohen's d) | 95% CI | p-value | Effect (Cohen's d) | 95% CI | p-value |
| PhenoAge | 12mo | 0.04 | [-0.17 , 0.24] | 0.719 | -0.02 | [-0.29 , 0.26] | 0.889 |
| Clock | 24mo | 0.16 | [-0.03 , 0.36] | 0.104 | 0.21 | [-0.09 , 0.51] | 0.174 |
| PC | 12mo | -0.03 | [-0.19 , 0.12] | 0.663 | -0.08 | [-0.31 , 0.14] | 0.473 |
| PhenoAge | 24mo | 0.05 | [-0.11 , 0.20] | 0.550 | 0.00 | [-0.24 , 0.25] | 0.977 |
| GrimAge | 12mo | -0.10 | [-0.23 , 0.04] | 0.170 | -0.14 | [-0.34 , 0.06] | 0.164 |
| Clock | 24mo | -0.01 | [-0.14 , 0.12] | 0.898 | -0.09 | [-0.32 , 0.14] | 0.425 |
| PC GrimAge | 12mo | -0.04 | [-0.16 , 0.07] | 0.461 | -0.13 | [-0.30 , 0.05] | 0.154 |
| Clock | 24mo | 0.05 | [-0.07 , 0.17] | 0.432 | 0.01 | [-0.17 , 0.19] | 0.902 |
| **Panel C. Pace of Aging** | | | | | | | |
| Dunedin | 12mo | **-0.29** | **[-0.45 , -0.13]** | **4.83E-04** | **-0.43** | **[-0.67 , -0.19]** | **3.69E-04** |
| PACE | 24mo | **-0.25** | **[-0.41 , -0.09]** | **0.002** | **-0.40** | **[-0.67 , -0.12]** | **0.004** |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text. Bolded entries fall below the significance threshold of $p < 0.005$.

**Table S5. Analysis of CALERIE™ intervention effects with treatment-group stratified by dose of calorie restriction.** Panel A shows treatment effects for original and PC-based versions[18] of first-generation clocks. Panel B shows treatment effects for original and PC-based versions of second-generation clocks. Panel C shows treatment effects for DunedinPACE. The table reports effect-size estimates from analysis using the same regression model used to estimate ITT effects (**Table S4**) with the treatment group stratified by the percent calorie restriction achieved by participants in the CR-intervention group. We used a threshold value of 10% CR relative to baseline to stratify the treatment group. The <10% CR group included n=36 participants at 12-months follow-up and n=48 participants at 24-months follow-up. The >10% CR group included n=89 participants at 12-months follow-up and n=69 participants at 24-months follow-up. All tests reported were two-sided.

**Panel A. First-Generation Clocks**

| | | | Treatment Effect (Cohen's d) | 95% CI | p-value |
|---|---|---|---|---|---|
| Horvath Clock | <10% CR | 12mo | 0.13 | [-0.13 , 0.38] | 0.331 |
| | | 24mo | 0.07 | [-0.15 , 0.29] | 0.542 |
| | >10% CR | 12mo | 0.05 | [-0.15 , 0.24] | 0.641 |
| | | 24mo | 0.04 | [-0.17 , 0.24] | 0.740 |
| PC Horvath Clock | <10% CR | 12mo | 0.00 | [-0.10 , 0.11] | 0.963 |
| | | 24mo | 0.06 | [-0.04 , 0.16] | 0.227 |
| | >10% CR | 12mo | 0.02 | [-0.06 , 0.10] | 0.713 |
| | | 24mo | 0.02 | [-0.06 , 0.10] | 0.599 |
| Hannum Clock | <10% CR | 12mo | -0.01 | [-0.25 , 0.23] | 0.945 |
| | | 24mo | -0.02 | [-0.22 , 0.18] | 0.857 |
| | >10% CR | 12mo | 0.04 | [-0.15 , 0.24] | 0.660 |
| | | 24mo | 0.02 | [-0.19 , 0.22] | 0.874 |
| PC Hannum Clock | <10% CR | 12mo | -0.01 | [-0.17 , 0.16] | 0.942 |
| | | 24mo | 0.07 | [-0.08 , 0.22] | 0.347 |
| | >10% CR | 12mo | 0.01 | [-0.11 , 0.12] | 0.914 |
| | | 24mo | 0.02 | [-0.10 , 0.14] | 0.767 |
| Skin & Blood Clock | <10% CR | 12mo | -0.09 | [-0.30 , 0.13] | 0.428 |
| | | 24mo | -0.18 | [-0.39 , 0.03] | 0.095 |
| | >10% CR | 12mo | -0.09 | [-0.28 , 0.10] | 0.350 |
| | | 24mo | -0.11 | [-0.31 , 0.09] | 0.283 |
| PC Skin & Blood Clock | <10% CR | 12mo | 0.00 | [-0.12 , 0.12] | 1.000 |
| | | 24mo | 0.09 | [-0.02 , 0.20] | 0.105 |
| | >10% CR | 12mo | 0.03 | [-0.06 , 0.12] | 0.527 |
| | | 24mo | 0.05 | [-0.04 , 0.14] | 0.316 |

## Table S5 Continued

**Panel B. Second-Generation Clocks**

| | | | Treatment Effect (Cohen's d) | 95% CI | p-value |
|---|---|---|---|---|---|
| PhenoAge Clock | <10% CR | 12mo | 0.17 | [-0.10 , 0.44] | 0.219 |
| | | 24mo | 0.10 | [-0.13 , 0.34] | 0.392 |
| | >10% CR | 12mo | -0.02 | [-0.24 , 0.21] | 0.888 |
| | | 24mo | 0.20 | [-0.01 , 0.41] | 0.062 |
| PC PhenoAge Clock | <10% CR | 12mo | -0.07 | [-0.30 , 0.16] | 0.540 |
| | | 24mo | 0.05 | [-0.13 , 0.24] | 0.580 |
| | >10% CR | 12mo | -0.02 | [-0.19 , 0.14] | 0.811 |
| | | 24mo | 0.04 | [-0.12 , 0.21] | 0.591 |
| GrimAge Clock | <10% CR | 12mo | -0.12 | [-0.29 , 0.05] | 0.167 |
| | | 24mo | -0.07 | [-0.24 , 0.09] | 0.383 |
| | >10% CR | 12mo | -0.09 | [-0.24 , 0.07] | 0.266 |
| | | 24mo | 0.04 | [-0.10 , 0.18] | 0.598 |
| PC GrimAge Clock | <10% CR | 12mo | -0.06 | [-0.22 , 0.10] | 0.492 |
| | | 24mo | 0.05 | [-0.09 , 0.19] | 0.454 |
| | >10% CR | 12mo | -0.04 | [-0.17 , 0.09] | 0.539 |
| | | 24mo | 0.04 | [-0.08 , 0.17] | 0.496 |
| **Panel C. Pace of Aging** | | | | | |
| Dunedin PACE | <10% CR | 12mo | -0.19 | [-0.41 , 0.04] | 0.099 |
| | | 24mo | -0.14 | [-0.34 , 0.06] | 0.166 |
| | >10% CR | 12mo | **-0.33** | **[-0.51 , -0.15]** | **2.31E-04** |
| | | 24mo | **-0.33** | **[-0.50 , -0.15]** | **2.63E-04** |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text. Bolded entries fall below the significance threshold of $p < 0.005$.

**Table S6. Cell-count-adjusted effect-size estimates for intent-to-treat (ITT) and effect of treatment-on-the-treated (TOT) analysis testing the effect of CALERIE™ intervention on DNA methylation measures of aging.** Panel A shows treatment effects for original and PC-based versions[18] of first-generation clocks. Panel B shows treatment effects for original and PC-based versions of second-generation clocks. Panel C shows treatment effects for DunedinPACE. Intent to treat (ITT) effect estimates are reported on the left side of the table. Effect of treatment-on-the-treated (TOT) effect estimates are reported on the right side of the table. Models are the same as those reported in **Table S4**, with addition of covariates for change scores of estimated counts of monocytes, lymphocytes (CD4, CD8, natural killer, and b-cells), and neutrophils computed using the R package FlowSorted.Blood.EPIC. All tests reported were two-sided.

| | | Intent-to-Treat (ITT): Difference between CR and AL | | | Effect of Treatment on the Treated (TOT): Estimated Effect of 20% CR | | |
|---|---|---|---|---|---|---|---|
| | | Treatment Effect (Cohen's d) | 95% CI | p-value | Treatment Effect (Cohen's d) | 95% CI | p-value |
| **Panel A. First-Generation Clocks** | | | | | | | |
| Horvath Clock | 12mo | 0.11 | [-0.07 , 0.29] | 0.230 | 0.16 | [-0.09 , 0.42] | 0.198 |
| | 24mo | 0.08 | [-0.11 , 0.27] | 0.429 | 0.15 | [-0.17 , 0.47] | 0.356 |
| PC Horvath Clock | 12mo | 0.04 | [-0.02 , 0.10] | 0.197 | 0.07 | [-0.02 , 0.15] | 0.123 |
| | 24mo | 0.03 | [-0.03 , 0.09] | 0.333 | 0.08 | [-0.03 , 0.19] | 0.156 |
| Hannum Clock | 12mo | 0.06 | [-0.12 , 0.23] | 0.513 | 0.09 | [-0.15 , 0.34] | 0.454 |
| | 24mo | -0.01 | [-0.18 , 0.17] | 0.947 | -0.05 | [-0.36 , 0.25] | 0.725 |
| PC Hannum Clock | 12mo | 0.06 | [-0.01 , 0.14] | 0.097 | 0.09 | [-0.02 , 0.19] | 0.097 |
| | 24mo | 0.05 | [-0.04 , 0.14] | 0.269 | 0.09 | [-0.06 , 0.23] | 0.236 |
| Skin & Blood Clock | 12mo | -0.09 | [-0.26 , 0.09] | 0.322 | -0.10 | [-0.34 , 0.13] | 0.384 |
| | 24mo | -0.12 | [-0.31 , 0.06] | 0.187 | -0.20 | [-0.50 , 0.10] | 0.191 |
| PC Skin & Blood Clock | 12mo | 0.04 | [-0.02 , 0.11] | 0.206 | 0.07 | [-0.02 , 0.16] | 0.111 |
| | 24mo | 0.04 | [-0.03 , 0.11] | 0.263 | 0.08 | [-0.03 , 0.19] | 0.138 |
| **Panel B. Second-Generation Clocks** | | | | | | | |
| PhenoAge Clock | 12mo | 0.04 | [-0.14 , 0.22] | 0.664 | 0.05 | [-0.21 , 0.30] | 0.724 |
| | 24mo | 0.10 | [-0.08 , 0.27] | 0.280 | 0.14 | [-0.13 , 0.42] | 0.308 |
| PC PhenoAge | 12mo | 0.01 | [-0.06 , 0.08] | 0.746 | 0.02 | [-0.08 , 0.11] | 0.685 |
| | 24mo | 0.01 | [-0.07 , 0.09] | 0.846 | 0.03 | [-0.10 , 0.15] | 0.692 |
| GrimAge Clock | 12mo | -0.08 | [-0.20 , 0.04] | 0.213 | -0.10 | [-0.28 , 0.07] | 0.252 |
| | 24mo | -0.05 | [-0.18 , 0.08] | 0.470 | -0.10 | [-0.32 , 0.13] | 0.396 |
| PC GrimAge Clock | 12mo | -0.03 | [-0.08 , 0.02] | 0.248 | -0.05 | [-0.12 , 0.03] | 0.210 |
| | 24mo | 0.00 | [-0.06 , 0.05] | 0.884 | -0.03 | [-0.12 , 0.06] | 0.544 |
| **Panel C. Pace of Aging** | | | | | | | |
| Dunedin PACE | 12mo | **-0.26** | **[-0.42 , -0.10]** | **1.29E-03** | **-0.36** | **[-0.59 , -0.13]** | **0.002** |
| | 24mo | **-0.24** | **[-0.40 , -0.07]** | **0.005** | -0.30 | [-0.58 , -0.02] | 0.033 |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text. Bolded entries fall below the significance threshold of $p < 0.005$.

**Table S7 Means for DNA-methylation measures of aging over follow-up for women and men in the CALERIE™ Trial.** Data are the same as reported in **Table S1** disaggregated by sex. Panel A shows values for original and PC-based versions[18] of first-generation clocks. Panel B shows values for original and PC-based versions of second-generation clocks. Panel C shows values for DunedinPACE.

**Panel A. First-Generation Clocks**

| | | Women | | | | Men | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ad Libitum (n=48) | | Calorie Restriction (n=89) | | Ad Libitum (n=21) | | Calorie Restriction (n=39) | |
| | | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI |
| Horvath Clock | Baseline | -2.36 | [-3.30 , -1.43] | -2.44 | [-3.30 , -1.58] | -1.43 | [-3.30 , 0.45] | -1.21 | [-2.29 , -0.12] |
| | 12mo | -2.39 | [-3.37 , -1.41] | -2.45 | [-3.34 , -1.57] | -2.54 | [-4.68 , -0.40] | -1.22 | [-2.29 , -0.15] |
| | 24mo | -2.40 | [-3.39 , -1.41] | -2.59 | [-3.51 , -1.68] | -2.34 | [-4.29 , -0.40] | -1.23 | [-2.35 , -0.10] |
| PC Horvath Clock | Baseline | 8.74 | [7.62 , 9.87] | 8.83 | [8.02 , 9.63] | 9.77 | [7.77 , 11.77] | 10.04 | [8.68 , 11.40] |
| | 12mo | 8.50 | [7.25 , 9.75] | 8.66 | [7.86 , 9.47] | 9.20 | [7.18 , 11.21] | 9.88 | [8.57 , 11.19] |
| | 24mo | 8.25 | [7.04 , 9.45] | 8.53 | [7.66 , 9.41] | 9.33 | [7.36 , 11.31] | 9.20 | [7.76 , 10.63] |
| Hannum Clock | Baseline | 3.54 | [2.71 , 4.37] | 4.19 | [3.61 , 4.76] | 4.62 | [3.71 , 5.54] | 4.13 | [3.10 , 5.16] |
| | 12mo | 3.41 | [2.39 , 4.43] | 3.96 | [3.32 , 4.59] | 3.98 | [2.95 , 5.02] | 4.02 | [3.08 , 4.97] |
| | 24mo | 3.54 | [2.71 , 4.37] | 4.06 | [3.40 , 4.71] | 4.23 | [2.90 , 5.57] | 3.80 | [2.80 , 4.81] |
| PC Hannum Clock | Baseline | 12.94 | [11.95 , 13.93] | 13.34 | [12.63 , 14.05] | 14.57 | [12.96 , 16.18] | 14.32 | [13.21 , 15.44] |
| | 12mo | 12.79 | [11.64 , 13.94] | 13.19 | [12.47 , 13.92] | 14.27 | [12.57 , 15.96] | 14.09 | [13.05 , 15.14] |
| | 24mo | 12.53 | [11.49 , 13.57] | 13.17 | [12.37 , 13.97] | 14.41 | [12.81 , 16.01] | 13.55 | [12.43 , 14.68] |
| Skin & Blood Clock | Baseline | 1.60 | [0.83 , 2.37] | 1.60 | [0.99 , 2.21] | 0.92 | [-0.73 , 2.57] | 1.58 | [0.94 , 2.23] |
| | 12mo | 2.02 | [1.21 , 2.83] | 1.74 | [1.18 , 2.30] | 1.06 | [-0.32 , 2.45] | 1.58 | [0.80 , 2.36] |
| | 24mo | 2.29 | [1.47 , 3.11] | 1.70 | [1.01 , 2.38] | 1.10 | [-0.54 , 2.74] | 1.66 | [1.01 , 2.30] |
| PC Skin & Blood Clock | Baseline | 6.44 | [5.10 , 7.78] | 6.90 | [6.02 , 7.79] | 6.43 | [4.50 , 8.36] | 6.97 | [5.47 , 8.46] |
| | 12mo | 6.04 | [4.55 , 7.53] | 6.64 | [5.77 , 7.52] | 5.90 | [3.83 , 7.96] | 6.82 | [5.38 , 8.25] |
| | 24mo | 5.61 | [4.19 , 7.04] | 6.44 | [5.50 , 7.39] | 5.92 | [3.98 , 7.86] | 6.01 | [4.44 , 7.58] |

**Panel B. Second-Generation Clocks**

| | | Women | | | | Men | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ad Libitum (n=48) | | Calorie Restriction (n=89) | | Ad Libitum (n=21) | | Calorie Restriction (n=39) | |
| | | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI |
| PhenoAge Clock | Baseline | -9.56 | [-10.88 , -8.23] | -9.50 | [-10.58 , -8.43] | -12.28 | [-14.34 , -10.22] | -11.59 | [-13.15 , -10.04] |
| | 12mo | -10.10 | [-11.63 , -8.57] | -9.34 | [-10.41 , -8.28] | -11.70 | [-13.73 , -9.67] | -11.91 | [-13.73 , -10.10] |
| | 24mo | -10.33 | [-11.73 , -8.93] | -9.38 | [-10.39 , -8.37] | -12.47 | [-14.23 , -10.71] | -11.71 | [-13.35 , -10.08] |
| PC PhenoAge Clock | Baseline | 2.90 | [1.64 , 4.15] | 2.96 | [1.94 , 3.99] | 1.53 | [-0.81 , 3.87] | 1.98 | [0.44 , 3.52] |
| | 12mo | 2.83 | [1.49 , 4.16] | 2.63 | [1.67 , 3.58] | 1.41 | [-0.87 , 3.69] | 1.73 | [0.06 , 3.41] |
| | 24mo | 2.32 | [1.02 , 3.61] | 3.14 | [2.13 , 4.15] | 2.05 | [-0.15 , 4.25] | 1.15 | [-0.31 , 2.62] |
| GrimAge Clock | Baseline | 4.65 | [3.73 , 5.58] | 4.98 | [4.42 , 5.54] | 5.29 | [4.33 , 6.25] | 4.31 | [3.51 , 5.10] |
| | 12mo | 4.44 | [3.51 , 5.38] | 4.43 | [3.87 , 5.00] | 5.06 | [4.19 , 5.93] | 4.07 | [3.36 , 4.79] |
| | 24mo | 3.97 | [3.16 , 4.79] | 4.33 | [3.72 , 4.94] | 4.88 | [3.88 , 5.88] | 3.50 | [2.80 , 4.19] |
| PC GrimAge Clock | Baseline | 12.10 | [11.38 , 12.82] | 12.81 | [12.24 , 13.38] | 10.15 | [9.08 , 11.21] | 9.65 | [9.00 , 10.29] |
| | 12mo | 11.69 | [10.85 , 12.54] | 12.35 | [11.81 , 12.88] | 10.17 | [9.00 , 11.35] | 9.22 | [8.50 , 9.94] |
| | 24mo | 11.28 | [10.54 , 12.03] | 12.11 | [11.55 , 12.67] | 9.70 | [8.61 , 10.79] | 8.85 | [8.15 , 9.55] |

**Panel C. Pace of Aging**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dunedin PACE | Baseline | 0.94 | [0.92 , 0.96] | 0.95 | [0.93 , 0.96] | 0.94 | [0.92 , 0.96] | 0.95 | [0.93 , 0.96] |
| | 12mo | 0.95 | [0.93 , 0.97] | 0.93 | [0.92 , 0.94] | 0.95 | [0.93 , 0.97] | 0.93 | [0.92 , 0.94] |
| | 24mo | 0.95 | [0.93 , 0.97] | 0.93 | [0.92 , 0.95] | 0.95 | [0.93 , 0.97] | 0.93 | [0.92 , 0.95] |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text.

**Table S8. Change scores for DNA-methylation measures of aging at 12-and 24-month follow-ups for women and men in the CALERIE™ Trial.** Data are the same as reported in **Table S3** disaggregated by sex. Panel A shows values for original and PC-based versions[18] of first-generation clocks. Panel B shows values for original and PC-based versions of second-generation clocks. Panel C shows values for DunedinPACE.

**Panel A. First-Generation Clocks**

| | | Women | | | | Men | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ad Libitum (n=48) | | Calorie Restriction (n=89) | | Ad Libitum (n=21) | | Calorie Restriction (n=39) | |
| | | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI |
| Horvath Clock | 12mo | -0.02 | [-0.21 , 0.17] | 0.00 | [-0.15 , 0.16] | -0.21 | [-0.46 , 0.04] | -0.07 | [-0.26 , 0.12] |
| | 24mo | -0.04 | [-0.24 , 0.17] | -0.08 | [-0.22 , 0.07] | -0.24 | [-0.53 , 0.05] | -0.03 | [-0.20 , 0.13] |
| PC Horvath Clock | 12mo | -0.16 | [-0.15 , 0.01] | -0.05 | [-0.10 , 0.01] | -0.09 | [-0.20 , 0.01] | -0.12 | [-0.21 , -0.03] |
| | 24mo | 0.00 | [-0.24 , -0.08] | -0.06 | [-0.11 , -0.01] | -0.11 | [-0.21 , 0.00] | -0.21 | [-0.29 , -0.12] |
| Hannum Clock | 12mo | -0.07 | [-0.25 , 0.11] | -0.07 | [-0.21 , 0.07] | -0.18 | [-0.44 , 0.07] | -0.13 | [-0.35 , 0.09] |
| | 24mo | -0.02 | [-0.19 , 0.15] | -0.03 | [-0.17 , 0.11] | -0.14 | [-0.45 , 0.18] | -0.13 | [-0.37 , 0.11] |
| PC Hannum Clock | 12mo | -0.06 | [-0.17 , 0.05] | -0.04 | [-0.12 , 0.05] | -0.07 | [-0.22 , 0.07] | -0.14 | [-0.27 , 0.00] |
| | 24mo | -0.17 | [-0.29 , -0.05] | -0.04 | [-0.12 , 0.04] | -0.05 | [-0.20 , 0.11] | -0.21 | [-0.34 , -0.09] |
| Skin & Blood Clock | 12mo | 0.15 | [-0.04 , 0.33] | 0.06 | [-0.07 , 0.19] | 0.12 | [-0.13 , 0.38] | -0.04 | [-0.27 , 0.18] |
| | 24mo | 0.22 | [0.03 , 0.40] | 0.00 | [-0.12 , 0.12] | 0.06 | [-0.27 , 0.39] | 0.01 | [-0.15 , 0.17] |
| PC Skin & Blood Clock | 12mo | -0.10 | [-0.19 , 0.00] | -0.07 | [-0.13 , -0.01] | -0.10 | [-0.21 , 0.02] | -0.11 | [-0.22 , 0.00] |
| | 24mo | -0.22 | [-0.32 , -0.12] | -0.09 | [-0.15 , -0.03] | -0.11 | [-0.21 , -0.01] | -0.21 | [-0.29 , -0.12] |

**Panel B. Second-Generation Clocks**

| | | Women | | | | Men | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ad Libitum (n=48) | | Calorie Restriction (n=89) | | Ad Libitum (n=21) | | Calorie Restriction (n=39) | |
| | | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI | Mean | 95% CI |
| PhenoAge Clock | 12mo | -0.06 | [-0.25 , 0.12] | 0.04 | [-0.13 , 0.21] | 0.14 | [-0.23 , 0.50] | -0.11 | [-0.31 , 0.08] |
| | 24mo | -0.18 | [-0.45 , 0.08] | 0.04 | [-0.10 , 0.18] | -0.04 | [-0.28 , 0.21] | -0.01 | [-0.20 , 0.18] |
| PC PhenoAge Clock | 12mo | -0.05 | [-0.22 , 0.12] | -0.04 | [-0.16 , 0.08] | 0.05 | [-0.13 , 0.24] | -0.12 | [-0.30 , 0.07] |
| | 24mo | -0.16 | [-0.34 , 0.03] | 0.05 | [-0.06 , 0.15] | 0.11 | [-0.11 , 0.32] | -0.17 | [-0.32 , -0.02] |
| GrimAge Clock | 12mo | -0.07 | [-0.21 , 0.07] | -0.21 | [-0.32 , -0.10] | -0.08 | [-0.26 , 0.10] | -0.14 | [-0.32 , 0.05] |
| | 24mo | -0.27 | [-0.42 , -0.13] | -0.21 | [-0.33 , -0.09] | -0.15 | [-0.36 , 0.06] | -0.28 | [-0.43 , -0.12] |
| PC GrimAge Clock | 12mo | -0.13 | [-0.25 , -0.01] | -0.16 | [-0.25 , -0.07] | -0.01 | [-0.17 , 0.15] | -0.18 | [-0.33 , -0.03] |
| | 24mo | -0.30 | [-0.44 , -0.15] | -0.18 | [-0.27 , -0.10] | -0.16 | [-0.32 , 0.00] | -0.29 | [-0.41 , -0.17] |

**Panel C. Pace of Aging**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dunedin PACE | 12mo | 0.07 | [-0.05 , 0.31] | -0.16 | [-0.29 , -0.02] | 0.14 | [-0.08 , 0.36] | -0.22 | [-0.46 , 0.02] |
| | 24mo | 0.00 | [-0.08 , 0.23] | -0.09 | [-0.22 , 0.04] | 0.23 | [-0.04 , 0.51] | -0.20 | [-0.40 , 0.01] |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text.

**Table S9. Effect-estimates from intent-to-treat (ITT) analysis for women and men in the CALERIE™ Trial.** The table shows results from ITT analysis stratified by sex and p-values for tests of sex differences in treatment effects. Tests of sex differences were conducted by including a product term testing interaction between treatment group and sex in the ITT models reported in **Table S4**. Panel A shows treatment effects for original and PC-based versions[18] of first-generation clocks. Panel B shows treatment effects for original and PC-based versions of second-generation clocks. Panel C shows treatment effects for DunedinPACE. All tests reported were two-sided.

**Panel A. First-Generation Clocks**

| | | Women (AL n=48, CR n=89) | | | Men (AL n=21, CR n=39) | | | |
|---|---|---|---|---|---|---|---|---|
| | | Treatment Effect (Cohen's D) | 95% CI | p-value | Treatment Effect (Cohen's D) | 95% CI | p-value | p-value for test of sex difference |
| Horvath Clock | 12mo | 0.03 | [-0.18 , 0.25] | 0.763 | -0.03 | [-0.26 , 0.19] | 0.787 | 0.282 |
| | 24mo | 0.15 | [-0.19 , 0.48] | 0.389 | 0.23 | [-0.10 , 0.56] | 0.178 | 0.505 |
| PC Horvath Clock | 12mo | 0.01 | [-0.08 , 0.11] | 0.758 | 0.09 | [-0.01 , 0.18] | 0.073 | 0.231 |
| | 24mo | 0.00 | [-0.12 , 0.13] | 0.943 | -0.07 | [-0.20 , 0.06] | 0.274 | 0.082 |
| Hannum Clock | 12mo | 0.01 | [-0.22 , 0.24] | 0.934 | -0.01 | [-0.22 , 0.19] | 0.895 | 0.741 |
| | 24mo | 0.07 | [-0.22 , 0.36] | 0.630 | 0.04 | [-0.33 , 0.40] | 0.843 | 0.961 |
| PC Hannum Clock | 12mo | 0.02 | [-0.12 , 0.16] | 0.796 | 0.12 | [-0.02 , 0.27] | 0.101 | 0.125 |
| | 24mo | -0.03 | [-0.20 , 0.14] | 0.722 | -0.14 | [-0.32 , 0.04] | 0.127 | 0.069 |
| Skin & Blood Clock | 12mo | -0.09 | [-0.30 , 0.13] | 0.434 | -0.20 | [-0.42 , 0.02] | 0.068 | 0.583 |
| | 24mo | -0.10 | [-0.40 , 0.21] | 0.534 | 0.00 | [-0.35 , 0.35] | 0.999 | 0.347 |
| PC Skin & Blood Clock | 12mo | 0.02 | [-0.09 , 0.13] | 0.721 | 0.12 | [0.01 , 0.24] | 0.037 | 0.256 |
| | 24mo | 0.02 | [-0.12 , 0.17] | 0.747 | -0.06 | [-0.19 , 0.06] | 0.313 | 0.033 |

**Panel B. Second-Generation Clocks**

| | | Women (AL n=48, CR n=89) | | | Men (AL n=21, CR n=39) | | | |
|---|---|---|---|---|---|---|---|---|
| | | Treatment Effect (Cohen's D) | 95% CI | p-value | Treatment Effect (Cohen's D) | 95% CI | p-value | p-value for test of sex difference |
| PhenoAge Clock | 12mo | 0.13 | [-0.11 , 0.37] | 0.300 | 0.21 | [-0.05 , 0.46] | 0.112 | 0.201 |
| | 24mo | -0.17 | [-0.54 , 0.20] | 0.361 | 0.07 | [-0.22 , 0.35] | 0.640 | 0.494 |
| PhenoAge Clock | 12mo | 0.00 | [-0.19 , 0.20] | 0.996 | 0.18 | [-0.02 , 0.37] | 0.072 | 0.054 |
| | 24mo | -0.11 | [-0.37 , 0.14] | 0.385 | -0.24 | [-0.48 , 0.01] | 0.059 | 0.083 |
| GrimAge Clock | 12mo | -0.12 | [-0.28 , 0.05] | 0.172 | 0.06 | [-0.10 , 0.22] | 0.488 | 0.585 |
| | 24mo | -0.04 | [-0.27 , 0.19] | 0.723 | -0.15 | [-0.38 , 0.08] | 0.193 | 0.098 |
| GrimAge Clock | 12mo | 0.00 | [-0.14 , 0.14] | 0.988 | 0.12 | [-0.03 , 0.26] | 0.117 | 0.089 |
| | 24mo | -0.14 | [-0.35 , 0.07] | 0.187 | -0.10 | [-0.30 , 0.09] | 0.292 | 0.558 |

**Panel C. Pace of Aging**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dunedin PACE | 12mo | -0.24 | [-0.44 , -0.05] | 0.015 | -0.15 | [-0.34 , 0.04] | 0.114 | 0.166 |
| | 24mo | -0.39 | [-0.69 , -0.09] | 0.010 | -0.46 | [-0.74 , -0.17] | 0.002 | 0.344 |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text.

**Table S10. Effect-estimates from treatment-effect-on-the-treated (TOT) analysis for women and men in the CALERIE™ Trial.** The table shows results from TOT analysis stratified by sex and p-values for tests of sex differences in treatment effects. Tests of sex differences were conducted by including a product term testing interaction between treatment group and sex in the TOT models reported in **Table S4**. Panel A shows treatment effects for original and PC-based versions[18] of first-generation clocks. Panel B shows treatment effects for original and PC-based versions of second-generation clocks. Panel C shows treatment effects for DunedinPACE. All tests reported were two-sided.

**Panel A. First-Generation Clocks**

| | | Women (AL n=48, CR n=89) | | | Men (AL n=21, CR n=39) | | |
|---|---|---|---|---|---|---|---|
| | | Treatment Effect (Cohen's D) | 95% CI | p-value | Treatment Effect (Cohen's D) | 95% CI | p-value |
| Horvath Clock | 12mo | 0.05 | [-0.29 , 0.38] | 0.776 | 0.25 | [-0.10 , 0.61] | 0.154 |
| | 24mo | -0.09 | [-0.49 , 0.30] | 0.646 | 0.43 | [-0.04 , 0.89] | 0.072 |
| PC Horvath Clock | 12mo | 0.09 | [-0.07 , 0.24] | 0.267 | 0.03 | [-0.10 , 0.16] | 0.604 |
| | 24mo | 0.16 | [-0.03 , 0.34] | 0.099 | -0.08 | [-0.31 , 0.14] | 0.463 |
| Hannum Clock | 12mo | 0.07 | [-0.29 , 0.42] | 0.707 | 0.15 | [-0.21 , 0.50] | 0.417 |
| | 24mo | 0.05 | [-0.36 , 0.45] | 0.823 | -0.01 | [-0.57 , 0.56] | 0.979 |
| PC Hannum Clock | 12mo | 0.08 | [-0.15 , 0.30] | 0.512 | -0.05 | [-0.25 , 0.15] | 0.600 |
| | 24mo | 0.20 | [-0.07 , 0.48] | 0.151 | -0.22 | [-0.52 , 0.07] | 0.131 |
| Skin & Blood Clock | 12mo | -0.06 | [-0.41 , 0.29] | 0.742 | 0.01 | [-0.37 , 0.38] | 0.963 |
| | 24mo | -0.33 | [-0.78 , 0.12] | 0.151 | -0.12 | [-0.63 , 0.39] | 0.651 |
| PC Skin & Blood Clock | 12mo | 0.08 | [-0.09 , 0.26] | 0.349 | 0.03 | [-0.15 , 0.20] | 0.774 |
| | 24mo | 0.17 | [-0.04 , 0.38] | 0.113 | -0.09 | [-0.29 , 0.12] | 0.396 |

**Panel B. Second-Generation Clocks**

| | | Women (AL n=48, CR n=89) | | | Men (AL n=21, CR n=39) | | |
|---|---|---|---|---|---|---|---|
| | | Treatment Effect (Cohen's D) | 95% CI | p-value | Treatment Effect (Cohen's D) | 95% CI | p-value |
| PhenoAge Clock | 12mo | 0.16 | [-0.20 , 0.53] | 0.385 | -0.26 | [-0.63 , 0.11] | 0.165 |
| | 24mo | 0.31 | [-0.15 , 0.76] | 0.185 | 0.07 | [-0.38 , 0.52] | 0.765 |
| PC PhenoAge Clock | 12mo | 0.01 | [-0.32 , 0.34] | 0.966 | -0.17 | [-0.49 , 0.15] | 0.303 |
| | 24mo | 0.30 | [-0.05 , 0.65] | 0.092 | -0.40 | [-0.76 , -0.04] | 0.028 |
| GrimAge Clock | 12mo | -0.27 | [-0.55 , 0.01] | 0.056 | -0.08 | [-0.36 , 0.21] | 0.597 |
| | 24mo | 0.01 | [-0.32 , 0.34] | 0.952 | -0.45 | [-0.81 , -0.08] | 0.017 |
| PC GrimAge Clock | 12mo | -0.02 | [-0.25 , 0.21] | 0.855 | -0.25 | [-0.51 , 0.01] | 0.061 |
| | 24mo | 0.22 | [-0.03 , 0.47] | 0.091 | -0.30 | [-0.58 , -0.03] | 0.030 |

**Panel C. Pace of Aging**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Dunedin PACE | **12mo** | **-0.42** | **[-0.73 , -0.11]** | **0.007** | **-0.59** | **[-0.93 , -0.25]** | **0.001** |
| | 24mo | -0.32 | [-0.69 , 0.04] | 0.081 | **-0.76** | **[-1.21 , -0.30]** | **0.001** |

Table entries for measurements included in primary analysis (PC PhenoAge, PC GrimAge, DunedinPACE) are highlighted in black text. Bolded entries fall below the significance threshold of $p < 0.005$.

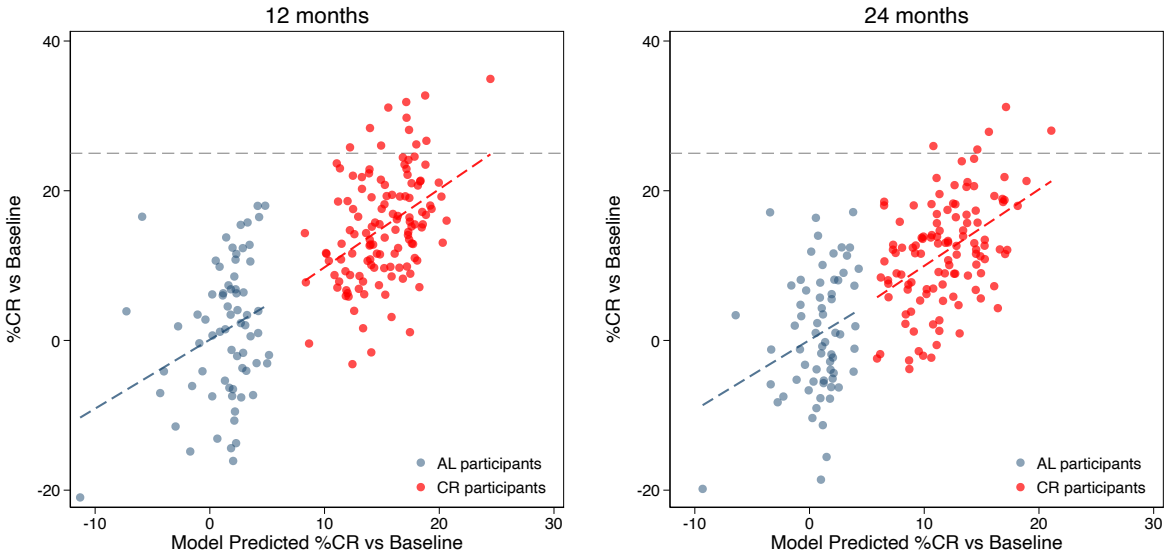



**Figure S1. Actual %CR plotted against Predicted CR.** The figure plots the measured % CR vs baseline (y-axis) against the model-predicted CR (x-axis) at the 12- (left) and 24-month (right) follow-up assessments. Values for ad-libitum (AL) control-group participants are plotted in blue. Values for caloric restriction (CR) intervention-group participants are plotted in red. The intervention target of 25% CR is noted with a dashed gray line intersecting the y-axis. Predicted values are derived from the base model for the first stage of IV regression analysis described in equation (1). The model includes terms for intervention condition, sex, pre-treatment baseline body-mass index, interactions among these variables, and covariates for pre-treatment baseline chronological age, study site, and race/ethnicity.

**Figure S2. Associations of DNA methylation measures of aging with chronological age at pre-intervention baseline in the CALERIE™ Trial.** Figures show scatter plots of the DNA methylation measures of aging (Y-axis) against chronological age (X-axis) for women (blue circles) and men (purple triangles) and fitted regression slopes. Correlations with chronological age are as follows: Horvath Clock r=0.87, PC Horvath Clock r=0.84, Hannum Clock r=0.92, PC Hannum Clock r=0.88, Skin & Blood Clock r=0.94, PC Skin & Blood Clock r=0.80, PhenoAge Clock r=0.84, PC PhenoAge Clock r=0.85, GrimAge Clock r=0.93, PC GrimAge Clock r=0.92, DunedinPACE r=0.15.

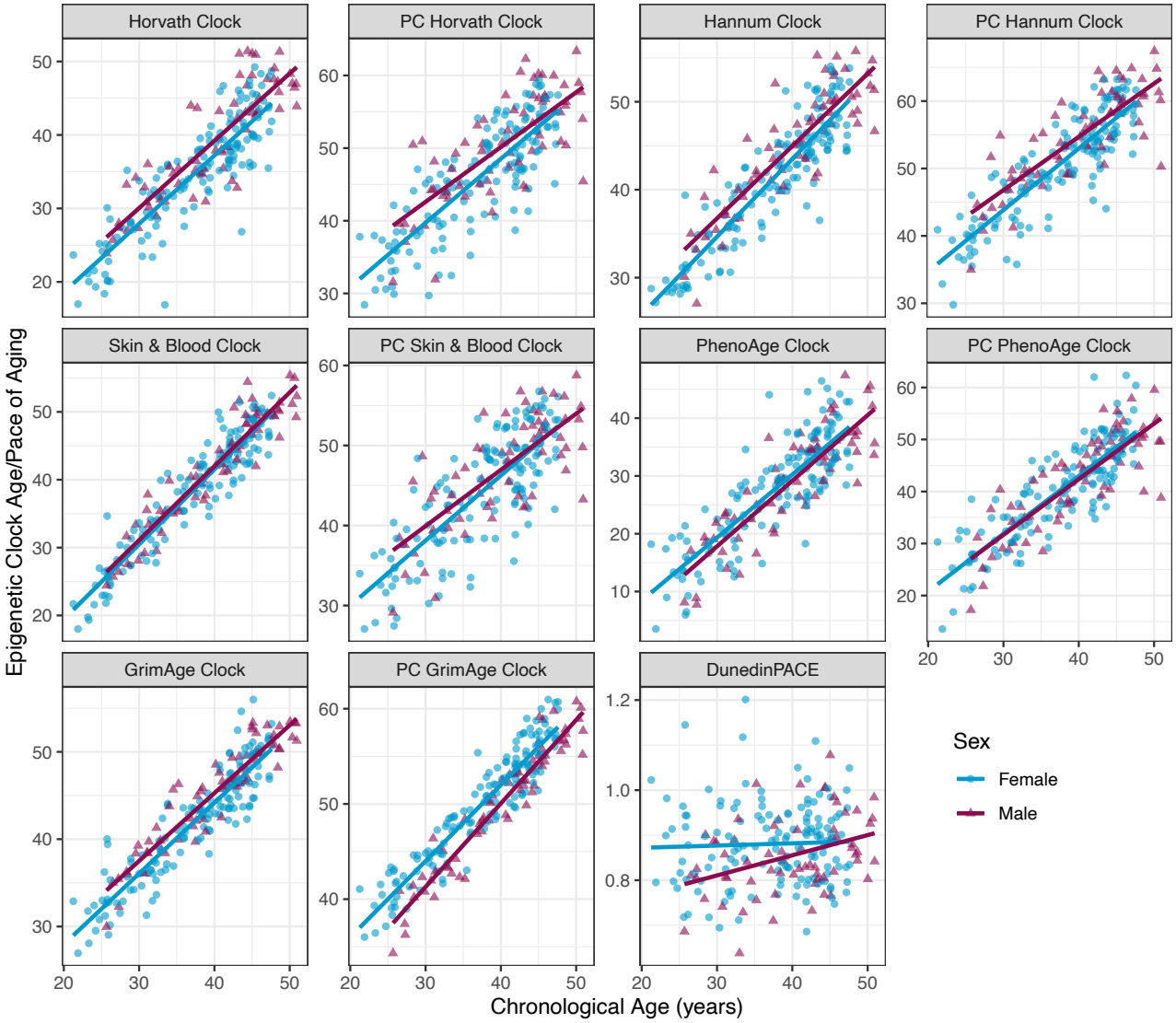
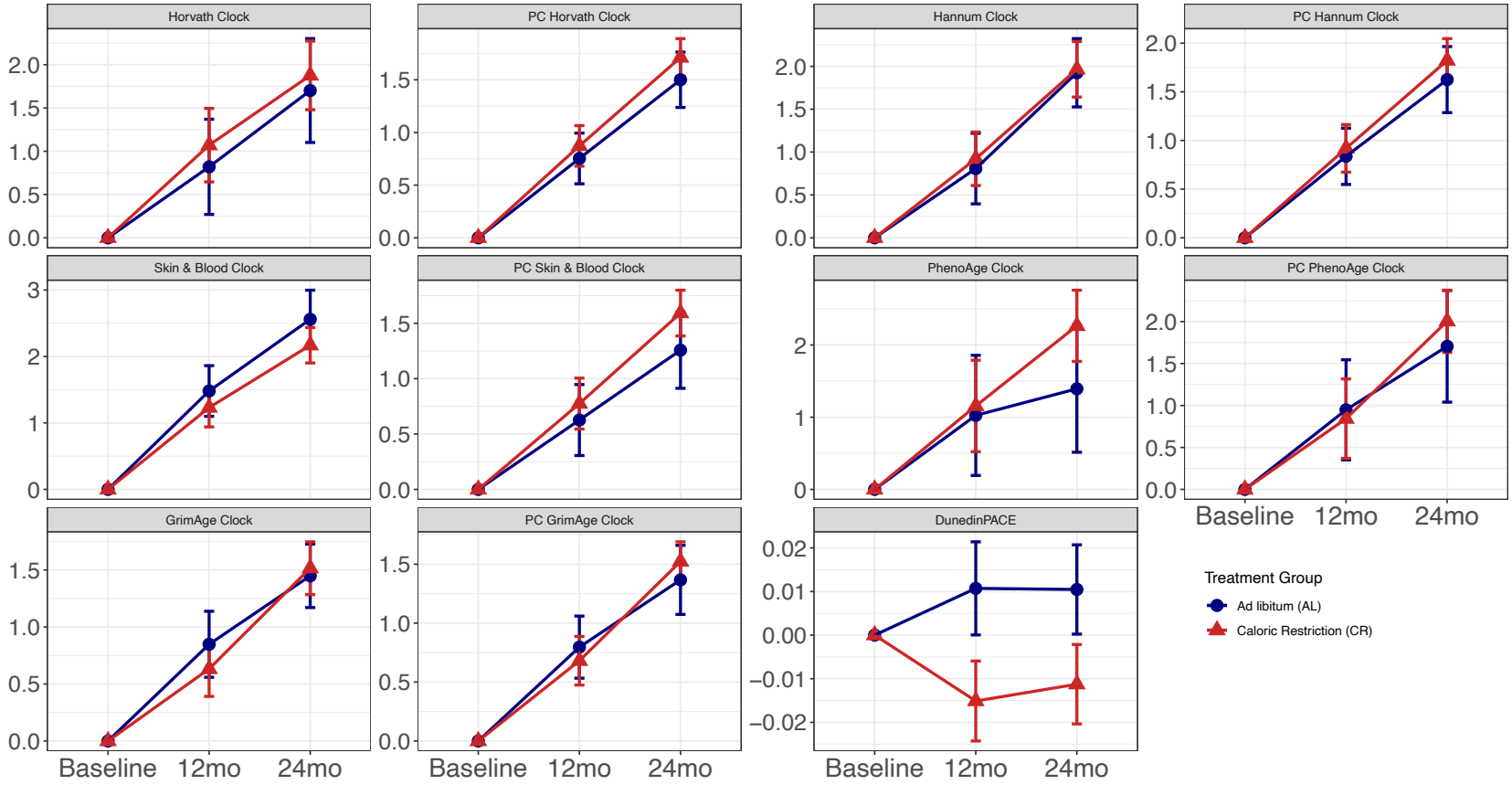


**Figure S3. Change from baseline to 12- and 24-month follow-up in DNA methylation measures of aging in the ad libitum (AL) and caloric restriction (CR) groups in the CALERIE™ Trial.** Values for the AL control group (n = 69 participants) are graphed in blue. Values for the CR treatment group (n = 128 participants) are graphed in red. Panel A mean values of change from baseline and 95% confidence intervals estimated from mixed models at the 12- and 24-month follow-ups for the AL and CR groups. There is no confidence interval estimated for baseline because change from baseline is exactly zero at this timepoint. Panel B shows box plots of the observed values of the measures at baseline and 12- and 24-month follow-ups. The boxes show the interquartile range; the whiskers show 1.5x the interquartile range; the center line shows the median; individual participant data are plotted as dots and connected with lines. Change scores are denominated in "years" of DNA methylation age. For the clocks, expected change under the null hypothesis is 1 year at 12-months follow-up and 2 years at 24 months follow-up.

**Panel A**


Horvath Clock
PC Horvath Clock
Hannum Clock
PC Hannum Clock
Skin & Blood Clock
PC Skin & Blood Clock
PhenoAge Clock
PC PhenoAge Clock
GrimAge Clock
PC GrimAge Clock
DunedinPACE
Baseline
12mo
24mo
Treatment Group
Ad libitum (AL)
Caloric Restriction (CR)

## Figure S3 Continued

### Panel B

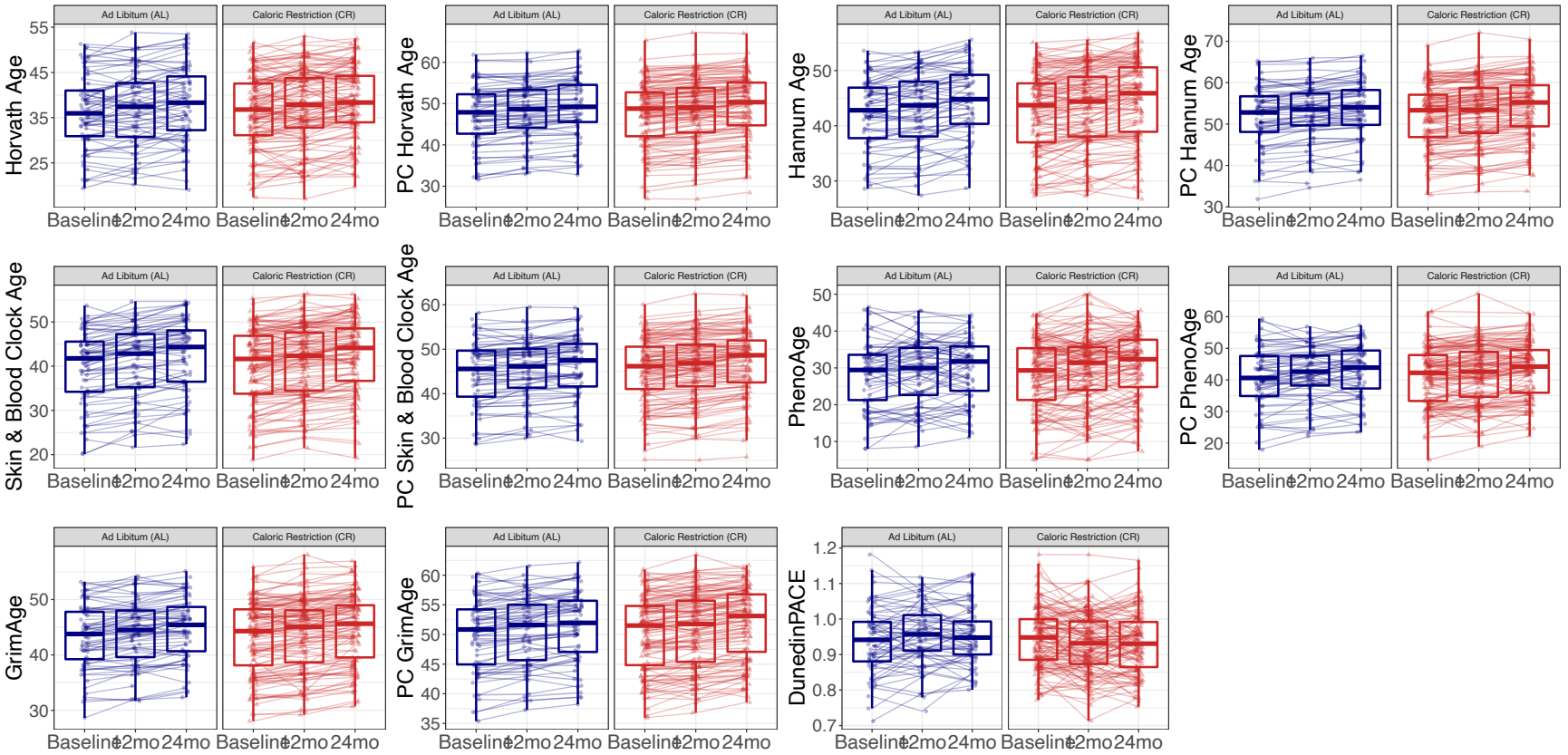


**Figure S4. Association of DunedinPACE with clinical measures modified by CALERIE™ intervention.** To evaluate whether intervention-induced changes in DunedinPACE overlapped with intervention effects on clinical measures, we conducted analysis of clinical measures altered by CALERIE™ intervention as reported in prior publications by our group[2,28]. Panel A shows effect-sizes (and 95% confidence intervals) for associations between DunedinPACE and each of the clinical measures at pre-intervention baseline (n = 197). Effect-sizes are graphed as Pearson r to be comparable across clinical measures with different scales. Panel B graphs estimates of total effects of CALERIE™ treatment on each of the clinical measures (blue bars) and indirect effects mediated through DunedinPACE (red bars). Effect estimates were generated according to the equations published by Imai et al.[29,29] implemented with the Stata package 'medeff'. Treatment effects are reported on a standard deviation scale (Cohen's d) so they are comparable between clinical measures. Confidence intervals did not overlap zero for estimates of mediation percent for log CRP, HDL, Triglycerides, log Triglycerides, Insulin AUC, SBP, DBP, Waist Circumference, Metabolic Syndrome, KDM BA, PhenoAge. For estimates of the adjusted causal mediation effect (ACME), confidence intervals overlapped zero for all outcomes with the exception of log-CRP and PhenoAge.

**Panel A.**

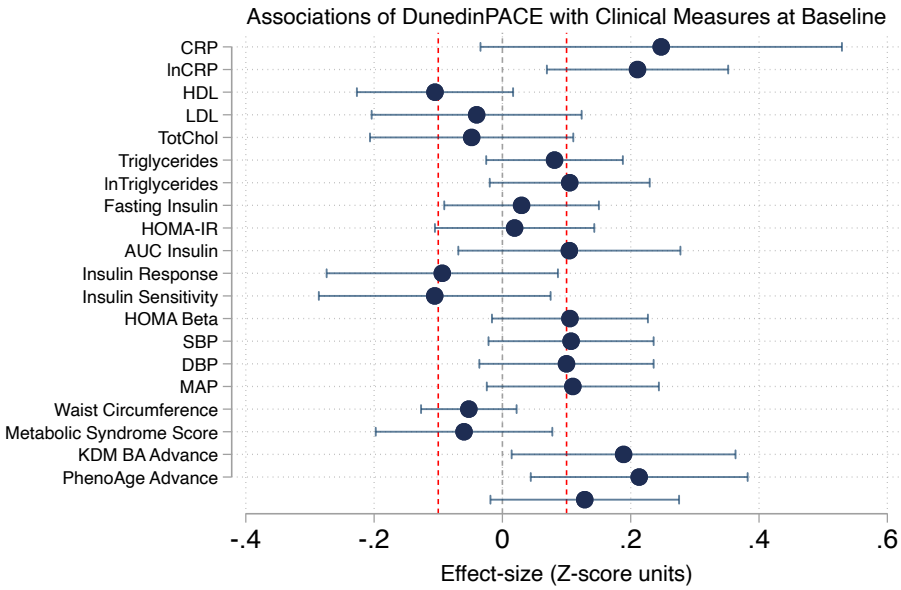
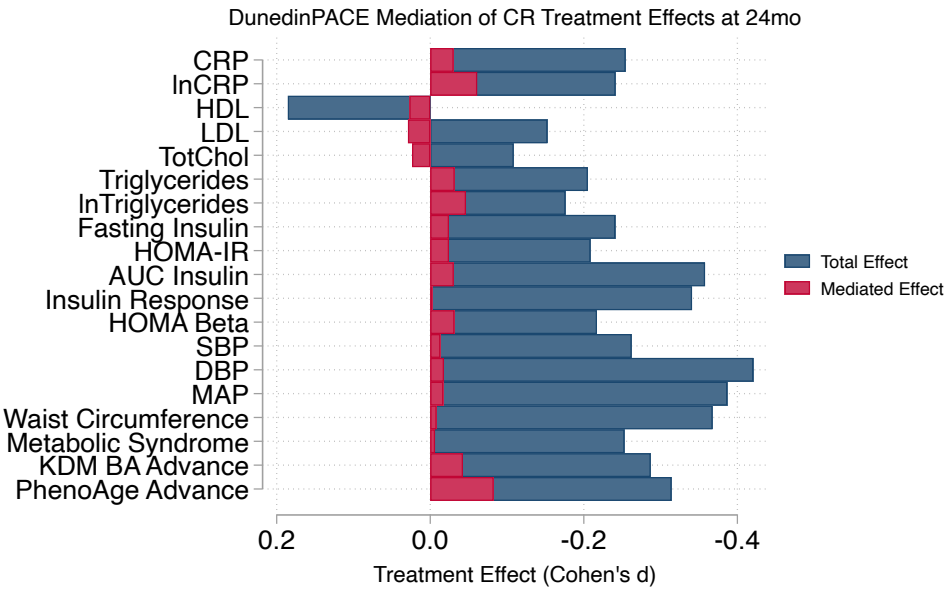


**Figure S4 Continued**

**Panel B.**



## REFERENCES

1. Ravussin, E. *et al.* A 2-Year Randomized Controlled Trial of Human Caloric Restriction: Feasibility and Effects on Predictors of Health Span and Longevity. *J. Gerontol. A. Biol. Sci. Med. Sci.* **70**, 1097–1104 (2015).
2. Kraus, W. E. *et al.* 2 years of calorie restriction and cardiometabolic risk (CALERIE): exploratory outcomes of a multicentre, phase 2, randomised controlled trial. *Lancet Diabetes Endocrinol.* **0**, (2019).
3. Racette, S. B. *et al.* Approaches for quantifying energy intake and %calorie restriction during calorie restriction interventions in humans: the multicenter CALERIE study. *Am. J. Physiol. - Endocrinol. Metab.* **302**, E441–E448 (2012).
4. Das, S. K. *et al.* Body-composition changes in the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE)-2 study: a 2-y randomized controlled trial of calorie restriction in nonobese humans. *Am. J. Clin. Nutr.* **105**, 913–927 (2017).
5. Davis, S. *et al.* methylumi: Handle Illumina methylation data. (2021) doi:10.18129/B9.bioc.methylumi.
6. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
7. Lehne, B. *et al.* A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* **16**, 37 (2015).
8. Salas, L. A. *et al.* An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* **19**, 64 (2018).

9. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
10. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
11. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
12. Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
13. Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* **8**, 1844–1865 (2016).
14. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
15. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
16. Levine, M. E. Assessment of Epigenetic Clocks as Biomarkers of Aging in Basic and Population Research. *J. Gerontol. Ser. A* **75**, 463–465 (2020).
17. Fahy, G. M. *et al.* Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, e13028 (2019).
18. Higgins-Chen, A. T. *et al.* A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. *Nat. Aging* **2**, 644–661 (2022).
19. Belsky, D. W. *et al.* DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11**, e73420 (2022).

20. Belsky, D. W. *et al.* Quantification of biological aging in young adults. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4104-4110 (2015).

21. Elliott, M. L. *et al.* 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 (2021).

22. Sugden, K. *et al.* Patterns of Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement. *Patterns* **1**, 100014 (2020).

23. StataCorp, L. Stata multilevel mixed-effects reference manual. *Coll. Stn. TX StataCorp LP* **9**, (2013).

24. Sussman, J. B. & Hayward, R. A. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. *BMJ* **340**, c2073 (2010).

25. Baum, C. F., Schaffer, M. E. & Stillman, S. Instrumental variables and GMM: Estimation and testing. *Stata J.* **3**, 1–31 (2003).

26. Bang, H. & Davis, C. E. On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? *Stat. Med.* **26**, 954–964 (2007).

27. McEwen, L. M. *et al.* Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array. *Clin. Epigenetics* **10**, 123 (2018).

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

29. Imai, K., Keele, L. & Yamamoto, T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat. Sci.* **25**, 51–71 (2010).

nature portfolio

Corresponding author(s): Daniel W Belsky

Last updated by author(s): 12-10-2021

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | DNA methylation data: Quality control and normalization analyses were performed using the methylumi (v. 2.32.0) 30 Bioconductor (v 2.46.0) 31 package for the R statistical programming environment (v 3.6.3). |
| Data analysis | Cell count estimation was performed using the Houseman Equation via the minfi (1.44.0) and FlowSorted.Blood.EPIC (2.2.0) R packages. DNA methylation measures of aging were computed using the software hosted by the Horvath Lab (epigenetic clocks) and R code publicly availabel on GitHub (Pace of Aging Measures; https://github.com/danbelsky). Analyses were conducted using the Stata Software (version 15.0). Stata code used to conduct analysis will be made available on GitHub upon acceptance of the article for publication. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All non-genomic data are available from the CALERIE biorepository (calerie.duke.edu). Genomic data are available from corresponding author until March 2024, at which point they will be available from the CALERIE biorepository.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | N=197 |
| Data exclusions | Data for n=23 participants were excluded due to missing or low quality DNA methylation data. |
| Replication | CALERIE is the only trial of long-term CR in healthy, non-obese humans. No replication was undertaken. |
| Randomization | After baseline testing, participants were randomly assigned at a ratio of 2:1 to a CR behavioral intervention or to an ad libitum control group (AL). Randomization was stratified by site, sex, and BMI. A permuted block randomization technique was used. |
| Blinding | All study personnel doing assessments were masked to treatment assignment. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☐ | ☒ Human research participants |
| ☐ | ☒ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Human research participants

Policy information about studies involving human research participants

| | |
|---|---|
| Population characteristics | Of 238 eligible individuals, CALERIE randomized N=220 participants to treatment (145 CR-intervention and 75 AL-control; Figure 1). Blood DNAm data were generated at baseline and at least one follow-up timepoint for n=197 participants (128 CR and 69 AL). Of this analysis sample, n=105 (82%) of CR participants and n=59 (86%) of AL participants had DNAm data available from all three time points (baseline, 12-months, and 24-months). Participants had mean age of 38 years (SD=7), 70% were women, and 77% were white; there were no differences in age, sex, race/ethnicity between AL and CR at baseline (Table 1). |
| Recruitment | Our study involved analysis of stored biospecimens and analysis of already-collected de-identified secondary data. Briefly, the details of the parent study's recruitment are as follows: CALERIE's age range (21–50 years) was selected to be comparable to the life stage when many "adult onset" CR studies showing sub- stantial effects on life span and aging were begun in rodents. The body mass index (BMI) range (22.0 ≤ BMI < 28 kg/m2) was selected to examine CR's effects in both normal weight and moderately over- weight persons. Exclusion criteria (detailed in Rochon J, Bales CW, Ravussin E, et al.; CALERIE Study Group. Design and conduct of the CALERIE study: comprehensive assessment of the long-term effects of reducing intake of energy. J Gerontol A Biol Sci Med Sci. 2011;66:97–108. doi:10.1093/gerona/glq168) included significant medical conditions (eg, cardiovascular disease or diabetes), abnormal laboratory markers (eg, elevated potassium, or below-normal hemoglobin levels), present or potential psychiatric or behavioral problems (eg, eating disorders or depressive symptoms), regular use of medications except oral contraceptives, current smoking, a high level of regular physical activity, and pregnancy. See Stewart TM, Bhapkar M, Das S, et al.; CALERIE Study Group. Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy Phase 2 (CALERIE Phase 2) screening and recruitment: methods and results. Con- temp Clin Trials. 2013;34:10–20. doi:10.1016/j.cct.2012.08.011 for more details. |

| | |
|---|---|
| Ethics oversight | Oversight of our study was performed by the Institutional Review Board of Columbia University Irving Medical Center AAAS2948. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| | |
|---|---|
| Clinical trial registration | NCT00427193 |
| Study protocol | https://calerie.duke.edu/protocols-procedures |
| Data collection | CALERIE™ Phase-2, was a multi-center, randomized controlled trial conducted at three clinical centers in the USA10 (ClinicalTrials.gov Identifier: NCT00427193). It aimed to evaluate the time-course effects of 25% CR (i.e. intake 25% below the individual's baseline level) over a 2-year period in healthy, adults (men aged 21–50 y, premenopausal women aged 21–47 y) with BMI in the normal weight or slightly overweight range (BMI 22.0-27.9 kg/m2). The study protocol was approved by Institutional Review Boards at three clinical centers (Washington University School of Medicine, St Louis, MO, USA; Pennington Biomedical Research Center, Baton Rouge, LA, USA; Tufts University, Boston, MA, USA) and the coordinating center at Duke University (Durham, NC, USA). All study participants provided written informed consent. |
| Outcomes | The primary outcomes were three DNA methylation measures of aging (PC PhenoAge, PC GrimAge, and DunedinPACE). We selected these based on published evidence of measurement reliability and criterion validity (i.e. associations with aging-related morbidity and mortality). Secondary outcomes were the original versions of the PhenoAge and GrimAge DNA methylation clocks, and the original and PC versions of DNA methylation clocks proposed by Horvath (multi-tissue and Skin & Blood) and Hannum et al. |