

Development and Characterization of Proteomic Aging Clocks in the Atherosclerosis Risk in Communities (ARIC) Study

Authors: Shuo Wang¹, Zexi Rao², Rui Cao², Anne H. Blaes³, Josef Coresh⁴, Corinne E. Joshi^{4,5}, Benoit Lehallier⁶, Pamela L. Lutsey⁷, James S. Pankow⁷, Sanaz Sedaghat⁷, Weihong Tang⁷, Bharat Thyagarajan¹, Keenan A. Walker⁸, Peter Ganz⁹, Elizabeth A. Platz^{4,5}, Weihua Guan^{2*}, Anna Prizment^{1*}

¹Department of Laboratory Medicine and Pathology, Medical School, University of Minnesota, Minneapolis, MN

²Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN

³Division of Hematology, Oncology and Transplantation, Medical School, University of Minnesota, Minneapolis, MN

⁴Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

⁵Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD

⁶Alkermest Inc, San Carlos, CA, United States

⁶Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA

⁷Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN

⁸Laboratory of Behavioral Neuroscience, Intramural Research Program, National Institute on Aging, Baltimore, MD

⁹Division of Cardiology, Zuckerberg San Francisco General Hospital and Department of Medicine, University of California, San Francisco, CA

*Both Weihua Guan and Anna Prizment are senior authors.

Corresponding author: Anna Prizment, email: prizm001@umn.edu, Department of Laboratory Medicine and Pathology, University of Minnesota

1 **Abstract**

2 Biological age may be estimated by proteomic aging clocks (PACs). Previous published PACs were
3 constructed either in smaller studies or mainly in White individuals, and they used proteomic measures
4 from only one-time point. In the Atherosclerosis Risk in Communities (ARIC) study of about 12,000
5 persons followed for 30 years (around 75% White, 25% Black), we created de novo PACs and compared
6 their performance to published PACs at two different time points. We measured 4,712 plasma proteins by
7 SomaScan in 11,761 midlife participants, aged 46-70 years (1990-92), and 5,183 late-life participants,
8 aged 66-90 years (2011-13). All proteins were log₂-transformed to correct for skewness. We created de
9 novo PACs by training them against chronological age using elastic net regression in two-thirds of
10 healthy participants in midlife and late life and compared their performance to three published PACs. We
11 estimated age acceleration (by regressing each PAC on chronological age) and its change from midlife to
12 late life. We examined their associations with mortality from all-cause, cardiovascular disease (CVD),
13 cancer, and lower respiratory disease (LRD) using Cox proportional hazards regression in all remaining
14 participants irrespective of health. The model was adjusted for chronological age, smoking, body mass
15 index (BMI), and other confounders. The ARIC PACs had a slightly stronger correlation with
16 chronological age than published PACs in healthy participants at each time point. Associations with
17 mortality were similar for the ARIC and published PACs. For late-life and midlife age acceleration for the
18 ARIC PACs, respectively, hazard ratios (HRs) per one standard deviation were 1.65 and 1.38 (both
19 $p < 0.001$) for all-cause mortality, 1.37 and 1.20 (both $p < 0.001$) for CVD mortality, 1.21 ($p = 0.03$) and 1.04
20 ($p = 0.19$) for cancer mortality, and 1.46 and 1.68 (both $p < 0.001$) for LRD mortality. For the change in age
21 acceleration, HRs for all-cause, CVD, and LRD mortality were comparable to those observed for late-life
22 age acceleration. The association between the change in age acceleration and cancer mortality was
23 insignificant. In this prospective study, the ARIC and published PACs were similarly associated with an
24 increased risk of mortality and advanced testing in relation to various age-related conditions in future
25 studies is suggested.

26 **Introduction**

27 In the United States the average human life expectancy has increased by 30 years during the 20th century.
28 This increased life expectancy has given rise to the number of individuals living with age-related diseases
29 and disabilities and has inevitably led to an increased risk of mortality, reduced health span, lower quality
30 of life, and increased healthcare costs in the United States. Research is needed to understand the
31 biological mechanisms of aging as we develop and target preventions and interventions that prolong
32 healthy lifespan (1, 2).

33 An individual's extent of aging, i.e., how far they are into the aging process, cannot be
34 sufficiently measured by chronological age as individuals develop physiological dysregulations at
35 different chronological ages (1, 2). To better understand the extent of aging, researchers introduced a term
36 called "biological age" to capture how far individuals are into their aging process independent of
37 chronological age. Biological age, according to the definition proposed by Baker and Sprott, is
38 characterized by the "biological parameter[s] of an organism, either alone or in some multivariate
39 composite that will, in the absence of disease, better predict functional capability at some late age than
40 will chronological age" (3).

41 To estimate a person's biological age, researchers have developed metrics called aging clocks
42 using epigenetic, transcriptomic, metabolomic, proteomic, and other biomarkers (4). Aging clocks are
43 strongly correlated with chronological age in healthy individuals. However, in individuals with
44 comorbidities or predisposing conditions, aging clocks deviate from chronological age because these
45 conditions impact levels of age-associated biomarkers (5, 6). Studies show that aging clocks may be used
46 to identify individuals who have a positive deviation of biological age from their chronological age
47 (called age acceleration) that may predict their future risk of age-related conditions (5-7). In addition,
48 aging clocks may also track the effectiveness of anti-aging interventions in clinical trials (5, 8-10).

49 The most studied aging clocks are epigenetic clocks, such as the Horvath clock, Hannum clock,
50 DNAm PhenoAge, and GrimAge (11-14). However, there is a lack of understanding of the underlying
51 mechanisms of aging-related changes in DNA methylation sites. It remains unclear what aspects of aging

52 those clocks reflect (15). Recently, new assays that measure thousands of proteins in a small blood sample
53 simultaneously have been developed. For instance, the SomaScan assay, a modified aptamer-based
54 technology (16-18). These assays make it possible to construct proteomic aging clocks (PACs) (5-7, 19).
55 The strength of PACs is that they include proteomic-based biomarkers, an intermediate phenotype that is
56 most proximal to age-related diseases, and thus may provide more accurate information on aging and age-
57 related pathologies (5, 20). Importantly, proteins serve as a target in 96% of FDA-approved drugs (21).
58 Therefore, in addition to predicting biological age and risk of diseases, proteins comprising PACs, if
59 causal, hold promise as targets of anti-aging drugs. Targeting age-related processes or pathological
60 manifestations instead of a single disease is advantageous as this approach may simultaneously reduce
61 the development or progression of multiple age-related diseases and potentially prolong healthy
62 lifespan.

63 Several PACs have been developed using SomaScan assays, such as the PACs created by
64 Lehallier [2020] (N = 3,301, aged 18-76 years) (6), Tanaka [2018] (N = 240, aged 22-93 years) (5), and
65 Sathyan [2020] (N = 1,025, aged 65-95 years) (19). The descriptions of those published PACs, including
66 the number of proteins used to construct those PACs, are presented in **S1 Table**. Although those
67 published PACs showed high correlations with chronological age, they were developed either in relatively
68 small studies or in studies included individuals of European descent (5, 6, 19, 22). However, proteins
69 associated with age and age-related diseases vary by race and socioeconomic status (23-25). Moreover,
70 previously published PACs were constructed using a one-time measure. Thus, it is necessary to develop
71 PACs in a large longitudinal study of diverse individuals and examine if the change in PACs over time is
72 associated with mortality independent of chronological age, smoking, and other lifestyle factors and
73 behaviors.

74 In this study, we developed new PACs in participants followed from midlife and late life and
75 examined their associations with mortality within a large population-based prospective cohort of White
76 and Black, men and women, in the Atherosclerosis Risk in Communities (ARIC) study. In ARIC, about
77 5,000 plasma proteins were measured using the SomaScan assay (v.4) from plasma samples collected at

78 two different times (20 years apart). We aimed to compare the midlife and late-life ARIC PACs
79 developed in healthy participants (without major age-associated diseases) with the published Lehallier's,
80 Tanaka's, and Sathyan's PACs. Comparing their correlation with chronological age and their associations
81 with mortality from all-cause, cardiovascular disease (CVD), cancer, and lower respiratory disease
82 (LRD). In addition, using protein data measured at two different time points, we examined whether the
83 change in PACs from midlife to late life was associated with premature mortality.

84 **Methods**

86 ***Study population***

87 This study included White and Black, men and women, participants of the ongoing ARIC study
88 (RRID: SCR_021769), which was initiated in 1987 (26, 27). At Visit 1 (1987-89), 15,792 volunteers aged
89 45-64 years were recruited from four U.S. study centers, Washington County, Maryland; the northwest
90 suburbs of Minneapolis, Minnesota; Jackson, Mississippi; and Forsyth County, North Carolina.
91 Participants in the Minnesota and Maryland centers were primarily White and the recruitment in
92 Mississippi was restricted to Black residents. ARIC was approved by institutional review boards at each
93 participating center and all study participants provided written informed consent. To date, nine visits have
94 been completed (26). ARIC participants have received follow-up telephone calls annually from 1987 to
95 2012 and semi-annually after 2012, with response rates of 90%-99% for the annual follow-up calls and
96 83%-90% for semi-annual follow-up calls among living participants who have not withdrawn consent to
97 be contacted (27). There is also continuous surveillance of local hospitals and linkage to the National
98 Death Index (NDI).

99 ***Plasma collection***

100 In this study, we used plasma samples collected at Visit 2 (1990-92) from 11,761 participants
101 aged 46-70 years (midlife) and at Visit 5 (2011-13) from 5,183 participants aged 66-90 years (late life).
102 The blood sample collection, processing, and storage in ARIC was designed to minimize the spontaneous
103 biochemical reactions after blood collection and is consistent with the recommended practice for

104 proteomics data analysis in epidemiological studies (16, 28, 29). After venipuncture, blood samples were
105 put immediately in an ice water bath. Centrifugation was performed within 10 min after venipuncture at
106 room temperature (15-25 °C). After centrifugation, the aliquots were stored at –80 °C within 90 min from
107 venipuncture and were unthawed before this analysis.

108 *Protein measurement and quality control*

109 Plasma samples were analyzed using a SOMAmer (Slow Off-rate Modified Aptamers) based
110 capture array called SomaScan® by Somalogic, Inc. (Boulder, CO, USA) (18, 30-32). The SomaScan
111 platform uses single-stranded modified DNA-based aptamers to capture conformational protein epitopes.
112 The description of the SomaScan assay and the data normalization process have been described
113 previously (16, 17, 32).

114 Among the 5,284 available aptamers, we excluded aptamers with a Bland-Altman coefficient of
115 variation (CVBA) greater than 50% or a variance of less than 0.01 on the log scale, or binding to mouse
116 Fc-fusion, contaminants, or non-proteins (33). After the exclusion, 4,955 aptamers were included (at Visit
117 2 and Visit 5) which corresponded to 4,712 proteins. About 5% of proteins had more than one aptamer
118 binding to the same protein. Each aptamer was treated as a variable in the construction of PACs. The
119 CVBA for split samples was 6% at Visit 2 and 7% at Visit 5. Protein measures were reported as relative
120 fluorescent units (RFU) and were log₂-transformed to correct for skewness.

121 *Identifying healthy participants*

122 In this study, we created the midlife (Visit 2) and late-life (Visit 5) ARIC PACs in “healthy
123 participants” defined as participants without major age-associated diseases that are linked to premature
124 mortality. Specifically, abnormal kidney function (i.e., estimated glomerular filtration rate (eGFR) less
125 than 60 mL/min/1.73m²), cancer, chronic obstructive pulmonary disease (COPD), CVD (heart failure,
126 definite or probable stroke, or coronary heart disease (34, 35)), diabetes, and hypertension (or
127 uncontrolled hypertension for late-life participants at Visit 5). The definitions and assessments of these
128 major diseases in ARIC and the detailed process of identifying healthy participants are described in the **S1**
129 **Appendix**. We identified 4,489 midlife healthy participants at Visit 2 (38.2% of all Visit 2 participants,

130 **Fig 1)** and 945 late-life healthy participants at Visit 5 (18.2% of all Visit 5 participants, **Fig 2).**

131 *Assessment of mortality and other characteristics of interest*

132 Deaths were ascertained through annual (semi-annual since 2012) follow-up telephone calls to
133 participants or their proxies, surveillance of local hospitals, state records, and linkage to NDI through
134 December 31, 2017 for participants in Mississippi or through December 31, 2019 for participants in other
135 centers (36). All-cause mortality was defined as death resulting from any cause. CVD mortality, cancer
136 mortality, and LRD mortality were defined based on the underlying cause of death using *International*
137 *Classification of Diseases, Ninth Revision*, codes (ICD-9 codes) 390–459 or *International Classification*
138 *of Diseases, Tenth Revision*, codes (ICD-10 codes) I00–I99 for CVD deaths, ICD-9 codes 140-239 or
139 ICD-10 codes C00-C97 for cancer deaths, and ICD-9 codes 466 and 480-519 or ICD-10 codes J10-J98
140 for LRD deaths.

141 Other characteristics of interest included demographic, lifestyle, and medical characteristics.
142 Namely chronological age, sex, race, study center, education, smoking status, pack-years of smoking,
143 alcohol intake, body mass index (BMI), physical activity, aspirin use, hormone replacement therapy
144 (HRT) in females (only at Visit 2; this variable is not available at Visit 5), and eGFR (26). Education
145 attainment was collected at Visit 1. Physical activity was collected at Visit 1 (used as physical activity at
146 Visit 2 in this study), and Visit 5. The other variables listed above were collected at both Visit 2 and Visit
147 5. Detailed procedures for assessing these characteristics are described in the **S1 Appendix**.

148 *Statistical analysis*

149 *Development of PACs*

150 To construct ARIC PACs in midlife (Visit 2) and late life (Visit 5), we randomly selected two-
151 thirds of healthy participants at each visit and used them as the training set at the corresponding visit. The
152 remaining one-third of healthy participants were used as the test set (**Fig 1 and Fig 2**). We utilized the
153 training set to train PACs against chronological age and obtain the appropriate hyperparameter values and
154 weight for each aptamer: $chronological\ age = \beta_0 + \sum_{i=1}^n \beta_i \times aptamer_i$, where $aptamer_i$ is the level

155 of the *ith* aptamer. We used the test set to examine the Pearson correlation (r) between PAC and
156 chronological age and median absolute error (MAE) to validate each PAC.

157 *Construction of midlife PACs in the Visit 2 training set*

158 Using the Visit 2 training set, we constructed the midlife ARIC PAC using elastic net regression
159 ($\alpha=0.5$) and with log2-transformed proteins. Lambda value was selected based on 10-fold cross-
160 validation. We chose elastic net regression because it combines the penalties from both Lasso and Ridge
161 regressions, and most previous aging clocks, including PACs and epigenetic clocks, were constructed
162 using elastic net regression. Using the Visit 2 proteomics data, we also trained four other midlife PACs by
163 applying different penalized regression methods and various protein transformations (described in **S2**
164 **Table**). For instance, one of the created PACs accounted for the potential nonlinear associations between
165 proteins and chronological age by including both the square term and cubic term of each aptamer. Those
166 four PACs were strongly correlated ($r \geq 0.97$) with the midlife ARIC PAC that was constructed using the
167 simplest protein transformation (**S3 Table**). Therefore, the simplest ARIC PAC was used for further
168 investigation.

169 In addition to the ARIC PAC, we also computed three published PACs in midlife: Lehallier's (6),
170 Tanaka's (5), and Sathyan's PACs (19). In our study, we computed Sathyan's PAC using the published
171 weights (19). For Lehallier's and Tanaka's PACs, we had to estimate ARIC weights specific to these
172 PACs using Ridge regression in the training set because ARIC did not include all the aptamers reported in
173 these PACs (**S1 Table**). The lambda value for Ridge regression was selected based on 10-fold cross-
174 validation.

175 *Construction of late-life PACs in the Visit 5 training set*

176 Because hypertension is one of the most common conditions in older persons in the United States
177 (37), to construct the late-life ARIC PAC, we additionally included participants with controlled
178 hypertension as healthy participants. Controlled hypertension was defined as the measured diastolic blood
179 pressure being below 90 and the measured systolic blood pressure being below 140 while the participant

180 is on medication (38). Adding these participants increased the number of healthy participants by 95%
181 (462 participants) but did not change the PAC's performance as shown in **S4 Table**.

182 Using the Visit 5 training set, we constructed the late-life ARIC PAC using elastic net regression
183 ($\alpha=0.5$), the same approach as for the midlife ARIC PAC. Lambda value was selected based on 10-
184 fold cross-validation. In addition to the late-life ARIC PAC, we computed the late-life Lehallier's and
185 Tanaka's PACs using ARIC weights estimated using the Visit 5 training set by applying Ridge regression
186 as discussed above and we computed the Sathyan's PAC using the published weights.

187 *Internal validation of PACs and examining associations with mortality*

188 In the remaining 8,768 participants at Visit 2 (**Fig 1**) and 4,553 participants at Visit 5 (**Fig 2**) after
189 excluding the training set, we computed PACs at the corresponding visits using the weighted sum of
190 proteins determined in the training set. We internally validated each PAC in the test set of healthy
191 participants at the corresponding visits by computing the Pearson correlation between PAC and
192 chronological age at that visit and MAE.

193 In all the remaining participants at each visit, to capture the PACs' effects independent of
194 chronological age, we created age acceleration for each PAC as residuals after regressing PAC on
195 chronological age (39). Demographic, lifestyle, and medical characteristics were examined across
196 quartiles of age acceleration as mean (SD) or percentage (%). To further investigate PACs, we examined
197 the associations between PACs and mortality. We used Cox proportional hazards regression to calculate
198 hazard ratios (HRs) and 95% confidence intervals (CIs) for mortality from all-cause, CVD, cancer, and
199 LRD with age acceleration. For the associations with CVD mortality, cancer mortality, and LRD
200 mortality, deaths from other causes were treated as competing events using the Fine and Gray method (40,
201 41). We modeled age acceleration as a continuous variable because there was no evidence of nonlinearity
202 observed when we applied cubic splines. For each participant, the total person-years were determined
203 from the date of blood collection (at Visit 2 or Visit 5, depending on the analysis) until death, censoring,
204 or the end of follow-up (either December 31, 2017 for participants from Mississippi or December 31,
205 2019 for participants from other centers), whichever occurred first. The proportional hazards assumption,

206 examined by the graphical methods using log-log survival curves with age acceleration dichotomized at
207 the median, was not violated in any regression models. The model was adjusted for chronological age,
208 sex, joint terms for race and study center (Black participants from Mississippi; Black participants from
209 any other centers; White participants from Maryland; White participants from North Carolina; and White
210 participants from Minnesota), education, BMI, smoking status, pack-years of smoking, alcohol intake,
211 physical activity, HRT (at Visit 2 only), diabetes, hypertension, and prevalent CVD, and eGFR (fully-
212 adjusted model). These variables were associated with either age acceleration or risk of mortality. To
213 confirm these variables as potential confounders, we computed the magnitude of R squared by regressing
214 age acceleration for both the midlife and late-life ARIC PACs on these variables at the corresponding
215 visits in the model adjusted for chronological age (**S6 Table**). We did not adjust for aspirin use because
216 aspirin use had no association with midlife or late-life age acceleration for ARIC PACs and aspirin use
217 explained <0.0015 of variance in both midlife and late-life age acceleration (**S6 Table**). In this study, we
218 found that HRs (95% CIs) for mortality were the same in the age-adjusted and fully-adjusted models.
219 Thus, we reported results for the fully adjusted model.

220 We also examined whether the change in age acceleration from midlife (Visit 2) to late life (Visit
221 5), computed as the age acceleration for the late-life ARIC PAC minus the age acceleration for the midlife
222 ARIC PAC, was associated with all-cause mortality and cause-specific mortality types using Cox
223 proportional hazard regression. For each participant, the total person-years was determined from Visit 5
224 date until death, censoring, or the end of follow-up. For this analysis, we additionally adjusted for midlife
225 age acceleration. Also, we examined whether the associations with mortality were modified by midlife
226 age acceleration (continuous variable) using a multiplicative term between the change in age acceleration
227 and midlife age acceleration. For the change in age acceleration, we only examined the change based on
228 the ARIC PACs because the ARIC and published PACs showed similar associations with all mortality
229 types at each visit.

230 In addition to studying the associations with mortality, we examined if midlife lifestyle and
231 medical characteristics (Visit 2) were associated with late-life age acceleration (Visit 5). This analysis

232 was conducted using multivariable linear regression and midlife participants' characteristics including:
233 chronological age, sex, race, education, BMI, smoking status, pack-years of smoking, alcohol intake,
234 physical activity (at Visit 1), HRT use, diabetes, hypertension, CVD, and eGFR were included into the
235 model simultaneously.

236 Finally, we tested whether or not the exclusion of the training set influenced the associations
237 between PACs and mortality. We examined this by comparing the associations for Sathyan's PAC in all
238 participants and in participants after excluding the training set. We used Sathyan's PAC rather than other
239 published PACs because all the proteins reported in Sathyan's PAC were measured in ARIC and we were
240 able to calculate Sathyan's PAC using published weights.

241 *Exploratory analyses*

242 In an exploratory analysis, we examined whether sex, race, or chronological age (in tertiles)
243 modified the associations of age acceleration with all-cause mortality, CVD mortality, and cancer
244 mortality by including a multiplicative term between age acceleration and the variable of interest in the
245 corresponding models. We did not examine LRD mortality due to the limited number of LRD deaths.

246 In the second exploratory analysis, we examined the association between age acceleration for the
247 midlife ARIC PAC and the 10-year risk of death as this may be important for clinical screening. We
248 tested the 10-year risk for midlife PAC only, because the follow-up period starting from late life was less
249 than 10 years. Here we examined the midlife ARIC PAC only, because ARIC and published PACs
250 showed similar associations with all mortality types.

251 In this study, PACs were constructed using R (version 4.1.2, package "glmnet"), and all the other
252 analyses were performed using SAS 9.4 (RRID: SCR_008567).

253 **Results**

254 *Midlife PACs*

255 After excluding the Visit 2 training set, the remaining participants at Visit 2 (midlife) were on
256 average 58.1±5.7 years old, 54.6% were female, and 27.1% were identified as Black.

257 *Pearson correlation coefficients between PACs and chronological age in midlife*

258 Elastic net regression selected 788 aptamers for the midlife ARIC PAC (**Table 1**). In the Visit 2
259 test set, the midlife ARIC PAC was correlated with chronological age ($r=0.80$, MAE=2.19 years, **Table 1**
260 **and Fig 3a**). Of the three midlife published PACs, Lehallier's PAC ($r=0.76$, **Table 1 and Fig 3b**) had a
261 slightly higher correlation with chronological age than Tanaka's ($r=0.66$) and Sathyan's PACs ($r=0.58$)
262 (**S5 Table and S1 Fig**). The midlife ARIC PAC was strongly correlated with the midlife Lehallier's
263 ($r=0.89$), Tanaka's ($r=0.77$), and Sathyan's PACs ($r=0.71$) (**S3 Table**).

264 *Distributions of midlife characteristics across quartiles of midlife age acceleration*

265 Distributions of midlife characteristics (Visit 2) across quartiles of midlife age acceleration are
266 shown in **Table 2** and **S7 Table**. Among the 8,768 participants in midlife (all Visit 2 participants after
267 excluding the Visit 2 training set), the range of age acceleration was from -14.0 to +24.2 years for the
268 midlife ARIC PAC. The distributions of characteristics including HRT use, prevalent CVD, and eGFR
269 were in the same direction across age acceleration for the midlife ARIC and published PACs (**Table 2**
270 **and S7 Table**). However, the distributions of gender, race, education, BMI, current smoking, aspirin use,
271 prevalent hypertension, and prevalent diabetes were different across different PACs (**Table 2 and S7**
272 **Table**). The difference may be because different PACs capture different aspects of aging.

273 *Association between midlife age acceleration and mortality*

274 Among the 8,768 participants at Visit 2, 5,294 died by 2019 with a median follow-up of 23.8
275 years. Age acceleration for the midlife ARIC PAC and published PACs showed associations of similar
276 magnitude with all mortality types (**Table 3 and S8 Table**). For the midlife ARIC PAC, a one SD (SD =
277 2.94 years) increase in age acceleration was associated with a 38% increased risk of all-cause mortality
278 [95% CI: 1.34-1.42], a 20% increased risk of CVD mortality [95% CI: 1.14-1.27], and a 36% increased
279 risk of LRD mortality [95% CI: 1.22-1.51] (**Table 3**). Neither age acceleration for the midlife ARIC PAC
280 nor published PACs was associated with cancer mortality (**Table 3**).

281 *Late-life PACs*

282 After excluding the Visit 5 training set, the remaining participants at Visit 5 (late life) were on
283 average 76.5±5.3 years old, 56.3% were female, and 19.7% were identified as Black.

284 *Pearson correlation coefficients between PACs and chronological age in late life*

285 Elastic net regression selected 135 aptamers for the late-life ARIC PAC (**Table 1**). In the Visit 5
286 test set, the late-life ARIC PAC was correlated with chronological age ($r=0.71$, $MAE=2.36$ years, **Table 1**
287 **and Fig 4a**). The late-life Lehallier's PAC had a correlation of 0.63 with chronological age (**Table 1 and**
288 **Fig 4b**) and the late-life Tanaka's and Sathyan's PACs had correlations of 0.59 and 0.69 with
289 chronological age, respectively (**S5 Table and S2 Fig**). In the Visit 5 test set, the late-life ARIC PAC was
290 strongly correlated with the late-life Lehallier's ($r = 0.84$), Tanaka's ($r = 0.79$), and Sathyan's PACs ($r =$
291 0.84) (**S9 Table**).

292 *Distribution of characteristics in late life across quartiles of late-life age acceleration*

293 Distribution of late-life characteristics (Visit 5) across quartiles of late-life age acceleration are
294 shown in **Table 4 and S10 Table**. Among the 4,553 participants in late life (all Visit 5 participants after
295 excluding the Visit 5 training set), the range of age acceleration was from -7.5 to +17.0 years for the late-
296 life ARIC PAC. The distributions of characteristics including having a college-level education, physical
297 activity, prevalent CVD, and eGFR were in the same direction across age acceleration for the late-life
298 ARIC and published PACs (**Table 4 and S10 Table**). However, the percentages of White participants,
299 never smokers, and never drinkers, and the prevalences of hypertension and diabetes were different
300 across different PACs (**Table 4 and S10 Table**).

301 *Association between late-life age acceleration and mortality*

302 Among the 4,553 participants at Visit 5, 1,123 died by 2019 with a median follow-up of 6.53
303 years. Age acceleration for the late-life ARIC and three published PACs were similarly associated with all
304 mortality types (**Table 5 and S11 Table**). For the late-life ARIC PAC, a one SD ($SD=2.61$ years)
305 increase in age acceleration was associated with an increased risk of all-cause mortality [HR (95% CI) =
306 1.65 ($1.52-1.79$)], CVD mortality [HR (95% CI) = 1.37 ($1.18-1.58$)], cancer mortality [HR (95% CI) =
307 1.21 ($1.02-1.44$)], and LRD mortality [HR (95% CI) = 1.68 ($1.32, 2.12$)] (**Table 5**).

308 *Associations of the change in age acceleration from midlife to late life with mortality*

309 The median timespan between Visit 2 and Visit 5 was 20.8 years, ranging from 18.6 to 23.5
310 years. Among the 2,707 participants who survived up to Visit 5 (after excluding the training sets at Visit 2
311 and Visit 5), the midlife and late-life ARIC PACs were correlated with each other ($r=0.69$) and 48.4% of
312 participants had a greater age acceleration in late life compared to midlife. In the fully adjusted model
313 (additionally adjusted for midlife age acceleration), the change in age acceleration from midlife to late life
314 was associated with all-cause mortality, CVD mortality, and LRD mortality, but not cancer mortality.
315 HRs (95% CIs) per one SD of the change in age acceleration were 1.71 (1.52-1.94) for all-cause
316 mortality, 1.38 (1.13-1.68) for CVD mortality, 1.46 (1.05-2.04) for LRD mortality, and 1.30 (0.98-1.71)
317 for cancer mortality (**Table 5**). Midlife age acceleration did not modify the associations between the
318 change in age acceleration and all-cause mortality (p-interaction for the multiplicative term=0.26), CVD
319 mortality (p-interaction=0.64), LRD mortality (p-interaction=0.26), or cancer mortality (p-
320 interaction=0.58).

321 *Association between midlife participants' characteristics and late-life age acceleration*

322 In the multivariable analysis of midlife participants' characteristics (Visit 2), we found that being
323 current smokers, never drinkers, or having diabetes, hypertension, CVD, a higher BMI, higher pack-years
324 of smoking, lower eGFR or lower physical activity in midlife were associated with higher late-life age
325 acceleration (**Table 6**).

326 *Comparison of the associations between age acceleration and mortality in the full cohort and the cohort* 327 *subset after excluding the training set*

328 The magnitudes of associations of age acceleration for Sathyan's PAC in both midlife and late
329 life with mortality in all participants at each visit were comparable to the magnitudes of those associations
330 in participants after excluding the training set (**S12 Table**).

331 *Proteins included in PACs*

332 There are 49 common aptamers included in both the midlife and late-life ARIC PACs, accounting
333 for 6.2% of all proteins in the midlife ARIC PAC and 36.4% of all proteins in the late-life ARIC PAC (**S3**
334 **Fig**). Four proteins were found in common across both the midlife and late-life ARIC PACs as well as the

335 three published PACs: pleiotrophin (PTN), A disintegrin and metalloproteinase with thrombospondin
336 motifs 5 (ADAMTS-5), macrophage metalloelastase (MMP12), and cell adhesion molecule-related/down-
337 regulated by oncogenes (CDON).

338 We also identified 20 proteins in each ARIC PAC (midlife and late-life) based on the largest
339 absolute weights of their constituting aptamers (**S13 Table**). We found six proteins whose corresponding
340 aptamers had the largest absolute weights in both ARIC PACs: transgelin (TAGL), WNT1-inducible-
341 signaling pathway protein 2 (WISP-2), chordin-like protein 1 (CRDK1), collagen alpha-1(XV) chain
342 (COF1), complement component C1q receptor (C1QR1), and pleiotrophin (PTN).

343 *Exploratory analyses*

344 *Associations between age acceleration and mortality stratified by sex, race, and chronological age*

345 The results for the associations between midlife PACs and mortality stratified by sex, race, and
346 chronological age are presented in **S14 Table** and **S4 Fig**. Notably, chronological age (in tertiles)
347 statistically modified the associations of age acceleration for both the midlife ARIC and three published
348 PACs with CVD mortality (p-interactions<0.01), and the association was strongest among participants
349 aged 47–54 years (first tertile) (**S14 Table** and **S4 Fig**).

350 The results for the associations between late-life PACs and mortality stratified by sex, race, and
351 chronological age are presented in **S15 Table** and **S5 Fig**. Sex statistically modified the association
352 between age acceleration and cancer mortality (p-interactions≤0.04) for the late-life ARIC PAC as well as
353 the three published PACs, and the association was stronger and significant in women for all PACs (**S15**
354 **Table** and **S5 Fig**). In addition, chronological age (in tertiles) significantly modified the association
355 between age acceleration and CVD mortality (p-interaction=0.04) for the late-life ARIC PAC but not the
356 published PACs (**S15 Table** and **S5 Fig**).

357 *Association between age acceleration for midlife ARIC PAC and 10-year risk of death*

358 Among the 8,768 participants in midlife, a total of 1,137 participants died within 10 years,
359 including 430 deaths attributed to CVD, 434 to cancer, and 85 to LRD. In the fully adjusted model, a one
360 SD (SD=2.94 years) increase in age acceleration for the midlife ARIC PAC was associated with an

361 increased risk of all-cause mortality [HR (95% CI)=1.49 (1.41-1.58)], CVD mortality [HR (95% CI)=1.47
362 (1.33-1.62)], cancer mortality [HR (95% CI)=1.21 (1.09-1.34)], and LRD mortality [HR (95% CI)=1.95
363 (1.60-2.38)].

364 **Discussion**

365 In a large prospective community-based study of White and Black individuals, the ARIC study,
366 we tested three published PACs (5, 6, 19) and constructed and validated *de novo* PACs in midlife (46-70
367 years) and late life (66-90 years), using 4,955 aptamers measured by the SomaScan assay (v.4). Both the
368 midlife and late-life ARIC PACs were developed in healthy participants and were strongly correlated with
369 chronological age. Correlations between chronological age and the ARIC PACs were 0.80 in midlife and
370 0.71 in late life, which were slightly stronger compared to the correlations between chronological age and
371 the three published PACs (Lehallier's, Tanaka's, and Sathyan's) respectively ($r=0.58-0.76$ in midlife and
372 $r=0.59-0.69$ in late life). All the HRs for the associations with mortality, including mortality from all-
373 cause, CVD, cancer, and LRD, were very similar for the ARIC and published PACs in midlife and late
374 life, respectively. Notably, the associations with all-cause mortality, CVD mortality, and LRD mortality
375 were significant at each visit but stronger in late life than in midlife, and the associations with cancer
376 mortality were significant in late life only. The change in age acceleration from midlife to late life had
377 associations of similar magnitude with all-cause mortality and CVD mortality when compared to the
378 associations for the late-life ARIC PAC. The HR estimate for LRD mortality was slightly lower for the
379 change in age acceleration compared to the late-life ARIC PAC, but the confidence intervals for these two
380 estimates largely overlapped. The change in age acceleration was not associated with cancer mortality.

381 In midlife we applied different penalized regressions and various transformations of proteins to
382 develop five *de novo* ARIC PACs, including a PAC that accounted for non-linear associations between
383 proteins and chronological age. These five PACs were highly correlated with each other. Thus, among
384 these five PACs, we selected the midlife ARIC PAC, constructed using the simplest protein
385 transformation, i.e., log₂ transformation without any further transformation. We selected the PAC with
386 the simplest protein transformation because, if validated, it would be easier to use this PAC in future

387 studies. We also constructed the late-life ARIC PAC using the same method as employed for the midlife
388 ARIC PAC. In our study, elastic net regression selected 788 aptamers for the midlife ARIC PAC and 135
389 aptamers for the late-life ARIC PAC. The smaller number of aptamers for the late-life ARIC PAC may be
390 because of the smaller training set at Visit 5 (N=630) compared to the Visit 2 training set (N=2,993).
391 With a larger training set, penalized regressions have more power to select more aptamers. This is in
392 agreement with Sathyan's PAC of 162 proteins, which was developed using the same SomaScan assay as
393 in our study with a training set of 500 participants (19).

394 We compared associations of midlife and late-life ARIC and published PACs with mortality.
395 Although different PACs included different proteins, the age acceleration for both ARIC and published
396 PACs showed comparable associations with mortality at each time point. Our findings for all-cause
397 mortality in midlife participants were similar to the findings in the InCHIANTI study (N=459,
398 chronological age: 21-98 years) by Tanaka et al. In their study, they reported a significant association
399 between age acceleration for Tanaka's PAC and all-cause mortality after adjusting for chronological age,
400 sex, and study site [HR (95% CI) per 1 SD = 1.29 (1.11-1.50)] (22). Our findings suggest that PACs
401 consisting of different proteins may be used for predicting mortality. It will be important to understand
402 this phenomenon in future studies.

403 In our study, for both ARIC and published PACs, their late-life age acceleration showed stronger
404 associations with all mortality types than midlife age acceleration. This may be because PACs that used
405 proteins measured in late life capture information for some biological function closer to mortality than
406 proteins measured in midlife. It is also possible that the longer follow-up of up to 29.9 years since midlife
407 introduced regression dilution bias (42), resulting in weaker associations with midlife PACs. Potential
408 regression dilution bias may also explain our stronger findings in the analysis of the midlife ARIC PAC
409 and mortality with a follow-up period restricted to 10 years compared to a total follow-up of up to 29.9
410 years until 2019. The association between the midlife ARIC PAC and cancer mortality became significant
411 when participants were followed for a maximum of 10 years. The associations with all-cause mortality,
412 CVD mortality, and LRD mortality were stronger compared to the associations observed for participants

413 followed up to 2019. In summary, our results underscored the potential of PACs to predict both biological
414 age and mortality in midlife and late life. Our results also suggested that PACs may be useful to predict
415 the 10-year risk of death in a clinical context.

416 Our findings showed that midlife individuals who were current smokers (compared to never
417 smokers), as well as those with higher (vs. lower) BMI, lower (vs. higher) eGFR, and age-related
418 diseases, such as CVD, hypertension, and diabetes in midlife, were associated with higher age
419 acceleration in late life. In addition, a larger change in age acceleration from midlife to late life was
420 associated with an increased risk of all-cause mortality, CVD mortality, and LRD mortality. Future
421 studies should incorporate multiple time points in applying PACs to model the change in age acceleration
422 over time.

423 The strengths of this population-based observational study include its prospective design with a
424 follow-up of more than 20 years and detailed demographic and lifestyle information. Furthermore, the
425 ARIC cohort includes a diverse sample comprising both White and Black individuals, while previous
426 studies of PACs either had small sample sizes or included mainly White individuals (5, 6, 19). Also, we
427 compared multiple PACs regarding their correlation with chronological age and their associations with
428 mortality. In addition, with the availability of proteomics data from two distinct visits, we were able to
429 examine the association between the midlife to late-life change in age acceleration and mortality.
430 Moreover, we adjusted for a broader range of confounders while previous studies of PACs only
431 adjusted for demographic factors (19, 22). Our study has several possible limitations. First, the
432 possibility of protein degradation during long-term storage cannot be excluded. However, the blood
433 samples were frozen right after their collection and have never been thawed reducing the possibility of
434 degradation. Further, no evidence of protein degradation across two visits in ARIC was shown by the
435 similar precision of the assay from a split duplicate analysis at both visits (CVBA = 6% at Visit 2 and 7%
436 at Visit 5) (16). Second, ARIC measured proteins in plasma, rather than other tissues, which limited the
437 generalizability of our PACs to proteins from other tissues.

438 In conclusion, we developed de novo midlife and late-life PACs in a diverse population of White
439 and Black individuals and showed that these PACs were associated with mortality risk. The magnitude of
440 these associations is similar to the associations observed for previously published PACs, both in midlife
441 and late life. Moreover, the change in age acceleration from midlife to late life showed comparable
442 associations with mortality as the late-life PAC. Future studies are recommended to investigate the
443 potential use of these PACs as biomarkers for biological age and risk stratification for age-related
444 disease. If validated in external studies, these PACs may serve as surrogate endpoints in clinical
445 trials of anti-aging interventions and inform physicians about the implementation of anti-aging
446 lifestyle and therapeutic interventions.

447

448 **Acknowledgment**

449 The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds
450 from the National Heart, Lung, and Blood Institute; National Institutes of Health; Department of Health
451 and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003,
452 75N92022D00004, 75N92022D00005). SomaLogic Inc. conducted the SomaScan assays in exchange for
453 the use of ARIC data. This work was supported in part by NIH/NHLBI grant R01 HL134320. Cancer data
454 in ARIC are also supported by the National Cancer Institute (U01 CA164975 and NU58DP007114).
455 Cancer incidence data have been provided by the Maryland Cancer Registry, Center for Cancer
456 Surveillance and Control, Department of Mental Health and Hygiene, 201 W. Preston Street, Room 400,
457 Baltimore, MD 21201. We acknowledge the State of Maryland, the Maryland Cigarette Restitution Fund,
458 and the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention
459 (CDC) for the funds that helped support the availability of the cancer registry data. The authors thank the
460 staff and participants of the ARIC study for their important contributions. This study was also supported
461 by R01CA267977 and R21AG079242. Dr. Lutsey was partially supported by NIH/NHLBI K24
462 HL159246. Dr. Walker is supported by the National Institute on Aging's Intramural Research Program.

463 This study was funded, in part, by the National Institute on Aging Intramural Research Program. The
464 content of this work is solely the responsibility of the authors and does not necessarily represent the
465 official views of the National Institutes of Health.

References

1. Sebastiani P, Thyagarajan B, Sun F, Schupf N, Newman AB, Montano M, et al. Biomarker signatures of aging. *Aging Cell*. 2017;16(2):329-38.
2. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 2018;19(6):371-84.
3. Baker GT, Sprott RL. Biomarkers of aging. *Exp Gerontol*. 1988;23(4-5):223-39.
4. Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. *EBioMedicine*. 2017;21:29-36.
5. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, et al. Plasma proteomic signature of age in healthy humans. *Aging Cell*. 2018;17(5):e12799.
6. Lehallier B, Shokhirev MN, Wyss-Coray T, Johnson AA. Data mining of human plasma proteins generates a multitude of highly predictive aging clocks that reflect different aspects of aging. *Aging Cell*. 2020;19(11):e13256.
7. Johnson AA, Shokhirev MN, Wyss-Coray T, Lehallier B. Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev*. 2020;60:101070.
8. Sierra F, Hadley E, Suzman R, Hodes R. Prospects for life span extension. *Annu Rev Med*. 2009;60:457-69.
9. Ferrucci L, Gonzalez-Freire M, Fabbri E, Simonsick E, Tanaka T, Moore Z, et al. Measuring biological aging in humans: A quest. *Aging Cell*. 2020;19(2):e13080.
10. Johnson AA, Stolzing A. The role of lipid metabolism in aging, lifespan regulation, and age-related disease. *Aging Cell*. 2019;18(6):e13048.
11. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
12. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359-67.
13. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 2018;10(4):573-91.
14. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*. 2019;11(2):303-27.
15. Galkin F, Mamoshina P, Aliper A, de Magalhães JP, Gladyshev VN, Zhavoronkov A. Biohorology and biomarkers of aging: Current state-of-the-art, challenges and opportunities. *Ageing Res Rev*. 2020;60:101050.
16. Tin A, Yu B, Ma J, Masushita K, Daya N, Hoogeveen RC, et al. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *The journal of applied laboratory medicine*. 2019;4(1):30-9.
17. Candia J, Daya GN, Tanaka T, Ferrucci L, Walker KA. Assessment of variability in the plasma 7k SomaScan proteomics assay. *Sci Rep*. 2022;12(1):17147.
18. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One*. 2010;5(12):e15004.
19. Sathyan S, Ayers E, Gao T, Weiss EF, Milman S, Verghese J, et al. Plasma proteomic profile of age, health span, and all-cause mortality in older adults. *Aging Cell*. 2020;19(11):e13250.

20. Zaghlool SB, Kühnel B, Elhadad MA, Kader S, Halama A, Thareja G, et al. Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits. *Nat Commun.* 2020;11(1):15.
21. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nature reviews Drug discovery.* 2017;16(1):19-34.
22. Tanaka T, Basisty N, Fantoni G, Candia J, Moore AZ, Biancotto A, et al. Plasma proteomic biomarker signature of age predicts health and life span. *Elife.* 2020;9.
23. Morris JC, Schindler SE, McCue LM, Moulder KL, Benzinger TLS, Cruchaga C, et al. Assessment of Racial Disparities in Biomarkers for Alzheimer Disease. *JAMA Neurol.* 2019;76(3):264-73.
24. Tahmasebi H, Asgari S, Hall A, Higgins V, Chowdhury A, Thompson R, et al. Influence of ethnicity on biochemical markers of health and disease in the CALIPER cohort of healthy children and adolescents. *Clin Chem Lab Med.* 2020;58(4):605-17.
25. Stanley S, Vanarsa K, Soliman S, Habazi D, Pedroza C, Gidley G, et al. Comprehensive aptamer-based screening identifies a spectrum of urinary biomarkers of lupus nephritis across ethnicities. *Nat Commun.* 2020;11(1):2197.
26. Wright JD, Folsom AR, Coresh J, Sharrett AR, Couper D, Wagenknecht LE, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol.* 2021;77(23):2939-59.
27. Joshu CE, Barber JR, Coresh J, Couper DJ, Mosley TH, Vitolins MZ, et al. Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer. *Cancer Epidemiol Biomarkers Prev.* 2018;27(3):295-305.
28. Tworoger SS, Hankinson SE. Collection, processing, and storage of biological samples in epidemiologic studies: sex hormones, carotenoids, inflammatory markers, and proteomics as examples. *Cancer Epidemiol Biomarkers Prev.* 2006;15(9):1578-81.
29. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, et al. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics.* 2005;5(13):3262-77.
30. Gold L, Walker JJ, Wilcox SK, Williams S. Advances in human proteomics at high scale with the SOMAscan proteomics platform. *New biotechnology.* 2012;29(5):543-9.
31. Kim CH, Tworoger SS, Stampfer MJ, Dillon ST, Gu X, Sawyer SJ, et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci Rep.* 2018;8(1):8382.
32. Candia J, Cheung F, Kotliarov Y, Fantoni G, Sellers B, Griesman T, et al. Assessment of Variability in the SOMAscan Assay. *Sci Rep.* 2017;7(1):14248.
33. Rooney MR, Chen J, Ballantyne CM, Hoogeveen RC, Tang O, Grams ME, et al. Comparison of Proteomic Measurements Across Platforms in the Atherosclerosis Risk in Communities (ARIC) Study. *Clin Chem.* 2023;69(1):68-79.
34. Bell EJ, Lutsey PL, Windham BG, Folsom AR. Physical activity and cardiovascular disease in African Americans in Atherosclerosis Risk in Communities. *Med Sci Sports Exerc.* 2013;45(5):901-7.
35. Folsom AR, Yatsuya H, Nettleton JA, Lutsey PL, Cushman M, Rosamond WD, et al. Community prevalence of ideal cardiovascular health, by the American Heart Association definition, and relationship with cardiovascular disease incidence. *J Am Coll Cardiol.* 2011;57(16):1690-6.

36. Rooney MR, Tang O, Pankow JS, Selvin E. Glycaemic markers and all-cause mortality in older adults with and without diabetes: the Atherosclerosis Risk in Communities (ARIC) study. *Diabetologia*. 2021;64(2):339-48.
37. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation*. 2015;131(4):e29-322.
38. Lu J, Lu Y, Krumholz HM, Jiang L. Prevalence and control of hypertension - Authors' reply. *Lancet*. 2018;392(10155):1306.
39. Kresovich JK, Xu Z, O'Brien KM, Weinberg CR, Sandler DP, Taylor JA. Methylation-Based Biological Age and Breast Cancer Risk. *J Natl Cancer Inst*. 2019;111(10):1051-8.
40. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*. 1999;94(446):496-509.
41. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391-400.
42. Clarke R, Shipley M, Lewington S, Youngman L, Collins R, Marmot M, et al. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *Am J Epidemiol*. 1999;150(4):341-53.

Figure 1. Study population at Visit 2 (midlife, 1990-1992, the chronological age of participants is 46-70 years); ARIC. The midlife ARIC PAC was constructed in a group of health participants in the training set and its association with mortality was examined in all remaining participants irrespective of health.

Figure 2. Study population at Visit 5 (late life, 2011-2013, the chronological age of participants is 66-90 years); ARIC. The late-life ARIC PAC was constructed in a group of health participants in the training set and its association with mortality was examined in all remaining participants irrespective of health.

Figure 3. Pearson correlation (r) between the midlife ARIC and Lehallier's PACs and chronological age in the Visit 2 test set of healthy participants. The x-axis depicts chronological age. The y-axis represents proteomic aging clock (PAC). (A) The midlife ARIC PAC was constructed using healthy participants from ARIC. The correlation between the midlife ARIC PAC and chronological age was 0.80. (B) Lehallier's PAC was computed using ARIC weights obtained from Ridge regression based on proteins available in ARIC. The correlation between midlife Lehallier's PAC and chronological age was 0.76.

Figure 4. Pearson correlation (r) between the late-life ARIC and Lehallier's PACs and chronological age in the Visit 5 test set of healthy participants, ARIC. The x-axis depicts chronological age. The y-axis represents proteomic aging clock (PAC). (A) The late-life ARIC PAC was constructed using healthy participants from ARIC. The correlation between the late-life ARIC PAC and chronological age was 0.71. (B) Lehallier's PAC was computed using ARIC weights obtained from Ridge regression based on proteins available in ARIC. The correlation between late-life Lehallier's PAC and chronological age was 0.63.

Table 1. Pearson correlation between the midlife and late-life ARIC and Lehallier’s PACs and chronological age and MAE, ARIC

| <i>The midlife ARIC and published PACs; Visit 2 (N = 2,993 in training set, N = 1,496 in test set)</i> | | |
|--|--------------------|---------------------------|
| | midlife ARIC PAC | midlife Lehallier’s PAC |
| Number of aptamers in PAC | 788 | 415 |
| Hyperparameter value (lambda) | 0.11 | 1.03 |
| Correlation in the training set ^a | 0.92 | 0.82 |
| Correlation in the test set ^a | 0.80 | 0.76 |
| MAE in the training set ^a | 1.50 | 2.13 |
| MAE in the test set ^a | 2.19 | 2.39 |
| <i>The late-life ARIC and published PACs; Visit 5 (N = 630 in training set, N = 315 in test set)</i> | | |
| | late-life ARIC PAC | late-life Lehallier’s PAC |
| Number of aptamers in PAC | 135 | 415 |
| Hyperparameter value (lambda) | 0.46 | 4.42 |
| Correlation in the training set ^a | 0.84 | 0.84 |
| Correlation in the test set ^a | 0.71 | 0.63 |
| MAE in the training set ^a | 1.47 | 1.71 |
| MAE in the test set ^a | 2.36 | 2.37 |

Abbreviations: MAE – median absolute error; PAC – proteomic ageing clock.

^aAmong healthy participants at Visit 2 and Visit 5, we randomly selected two-thirds of healthy participants at each visit and used them as the training set at the corresponding visit; the remaining one-third of healthy participants at each visit was used as the test set at the corresponding visit.

Table 2. Midlife participants' characteristics across quartiles of age acceleration for the midlife ARIC and Lehallier's PACs; ARIC

| | midlife ARIC PAC | | | | P-value ^c | midlife Lehallier's PAC | | | | P-value ^c |
|--|------------------|-----------------|-----------------|-----------------|----------------------|-------------------------|-----------------|-----------------|-----------------|----------------------|
| | Q1 (N=2,192) | Q2 (N=2,192) | Q3 (N=2,192) | Q4 (N=2,192) | | Q1 (N=2,192) | Q2 (N=2,192) | Q3 (N=2,192) | Q4 (N=2,192) | |
| Age acceleration, years | -14.0 to -1.9 | -1.8 to -0.2 | -0.3 to +1.7 | +1.8 to +24.2 | | -15.1 to -2.0 | -1.9 to -0.2 | -0.1 to +1.8 | +1.9 to +26.5 | |
| Chronological age, years (SD) | 58.3 (5.8) | 57.9 (5.8) | 57.8 (5.7) | 58.3 (5.6) | 0.07 | 58.2 (5.8) | 58.1 (5.8) | 58.1 (5.6) | 58.1 (5.6) | 0.88 |
| Female, % | 55.6 | 51.6 | 55.4 | 55.8 | 0.01 | 53.2 | 53.2 | 55.5 | 56.5 | 0.06 |
| White, % | 72.8 | 75.7 | 75.2 | 67.9 | <0.0001 | 69.8 | 75.6 | 74.2 | 72.2 | <0.0001 |
| Education, % | | | | | | | | | | |
| Less than high school | 21.8 | 21.8 | 22.9 | 30.0 | | 22.0 | 22.9 | 24.9 | 26.6 | |
| High school/vocational | 41.1 | 40.9 | 43.2 | 40.6 | <0.0001 | 39.8 | 42.4 | 40.9 | 42.6 | <0.0001 |
| College | 37.1 | 37.4 | 33.9 | 29.3 | | 38.2 | 34.6 | 34.1 | 30.7 | |
| BMI, kg/m ² (SD) | 28.2 (5.1) | 28.3 (5.1) | 28.3 (5.6) | 28.6 (6.3) | 0.13 | 28.5 (5.3) | 28.4 (5.4) | 28.2 (5.5) | 28.4 (5.9) | 0.58 |
| Smoking status, % | | | | | | | | | | |
| Current smoker | 22.3 | 23.2 | 22.4 | 24.2 | | 21.6 | 23.3 | 23.6 | 23.4 | |
| Former smoker | 36.6 | 38.4 | 39.1 | 39.6 | 0.06 | 36.8 | 38.3 | 38.5 | 39.9 | 0.05 |
| Never smoker | 41.1 | 38.4 | 38.6 | 36.2 | | 41.5 | 38.4 | 37.8 | 36.5 | |
| Pack-years of smoking among ever smokers, pack-years (SD) | 28.5 (22.8) | 30.4 (23.4) | 29.9 (22.0) | 32.5 (24.1) | 0.0003 | 28.4 (22.3) | 29.5 (23.1) | 30.3 (23.0) | 32.9 (23.7) | <0.0001 |
| Alcohol intake, % | | | | | | | | | | |
| Current drinker | 57.4 | 58.2 | 54.9 | 48.3 | | 54.8 | 56.8 | 55.9 | 51.2 | |
| Former drinker | 21.6 | 20.9 | 20.8 | 26.5 | <0.0001 | 22.7 | 20.4 | 21.2 | 25.3 | 0.002 |
| Never drinker | 20.9 | 20.9 | 24.3 | 25.2 | | 22.5 | 22.6 | 22.7 | 23.5 | |
| Physical activity ^a , scores (SD) | 2.48 (0.8) | 2.45 (0.8) | 2.42 (0.8) | 2.33 (0.8) | <0.0001 | 2.46 (0.8) | 2.44 (0.8) | 2.41 (0.8) | 2.38 (0.8) | 0.03 |
| Aspirin use in the preceding two weeks, % | 51.1 | 51.2 | 54.5 | 52.7 | 0.07 | 48.7 | 53.1 | 52.4 | 55.3 | 0.0002 |
| Ever user of hormone replacement therapy (females only), % | 50.2 | 47.0 | 41.2 | 36.3 | <0.0001 | 48.8 | 45.9 | 42.7 | 37.3 | <0.0001 |
| Diabetes ^b , % | 15.1 | 16.9 | 20.1 | 29.2 | <0.0001 | 17.0 | 16.4 | 20.2 | 27.8 | <0.0001 |
| Hypertension ^b , % | 42.6 | 46.2 | 49.4 | 55.8 | <0.0001 | 44.9 | 48.4 | 47.4 | 53.4 | <0.0001 |
| CVD ^b , % | 11.8 | 13.9 | 16.4 | 21.5 | <0.0001 | 11.8 | 13.7 | 15.6 | 22.8 | <0.0001 |
| eGFR, mL/min/1.73 m ² (SD) | 98.0 (11.1) | 97.3 (12.6) | 96.3 (13.3) | 90.9 (19.6) | <0.0001 | 97.3 (12.1) | 97.2 (12.5) | 95.8 (14.1) | 92.3 (19.2) | <0.0001 |

Abbreviations: PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

^aPhysical activity at Visit 1 was assessed using a leisure-time sports index that ranged from 1 to 5. We assumed physical activity scores remained the same at Visit 1 and Visit 2. We reported physical activity scores with two decimal places to illustrate the trend more effectively.

^bAll diseases were prevalent diseases.

^cP-values were calculated using chi-square tests for categorical variables and using ANOVA tests for continuous variables.

Table 3. The associations between age acceleration for the midlife ARIC and Lehallier's PACs and mortality; ARIC (1990-2019)

| | No. of participants | No. of deaths | Total person-years | HR (95% CI) ^a per one SD of age acceleration | |
|--|---------------------|---------------|--------------------|---|---|
| | | | | midlife ARIC PAC (SD=2.94 years) | midlife Lehallier's PAC (SD=3.00 years) |
| All-cause mortality | 8,768 | 5,294 | 182,630 | 1.38 (1.34, 1.42) | 1.34 (1.30, 1.38) |
| CVD mortality (Fine and Gray model) | 8,768 | 1,734 | 182,630 | 1.20 (1.14, 1.27) | 1.19 (1.13, 1.25) |
| Cancer mortality (Fine and Gray model) | 8,768 | 1,516 | 182,630 | 1.04 (0.98, 1.10) | 1.05 (0.99, 1.12) |
| LRD mortality (Fine and Gray model) | 8,768 | 522 | 182,630 | 1.36 (1.22, 1.51) | 1.30 (1.17, 1.45) |

Abbreviations: PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; LRD – lower respiratory disease; eGFR – estimated glomerular filtration rate; SD – standard deviation; HR – Hazard ratio; CI – confidence interval.

^aThe model was adjusted for chronological age, gender, joint terms for race and study center (Black participants from Mississippi; Black participants from any other centers; White participants from Maryland; White participants from North Carolina; and White participants from Minnesota), education, BMI, smoking status, pack-years of smoking, alcohol intake, physical activity (at Visit 1), hormone replacement therapy, diabetes, hypertension, CVD, and eGFR at Visit 2.

Table 4. Visit 5 participants' characteristics across quartiles of age acceleration for late-life ARIC and Lehallier's PACs; ARIC

| | late-life ARIC PAC | | | | P-value ^c | late-life Lehallier's PAC | | | | P-value ^c |
|---|--------------------|-----------------|-----------------|-----------------|----------------------|---------------------------|-----------------|-----------------|-----------------|----------------------|
| | Q1 (N=1,138) | Q2 (N=1,138) | Q3 (N=1,139) | Q4 (N=1,138) | | Q1 (N=1,138) | Q2 (N=1,138) | Q3 (N=1,139) | Q4 (N=1,138) | |
| Age acceleration | - 7.5 to -1.8 | -1.7 to -0.2 | -0.1 to +1.4 | +1.5 to +17.0 | | -9.1 to -1.8 | -1.7 to -0.2 | -0.1 to 1.5 | 1.6 to 14.4 | |
| Chronological age, years (SD) | 76.9 (5.1) | 75.9 (5.0) | 76.1 (5.3) | 76.7 (5.4) | <0.0001 | 76.7 (5.2) | 76.3 (5.1) | 76.0 (5.2) | 76.8 (5.5) | 0.0009 |
| Female, % | 57.7 | 57.5 | 57.1 | 52.9 | 0.06 | 62.2 | 59.8 | 53.9 | 49.5 | <0.0001 |
| White, % | 75.9 | 83.0 | 82.5 | 79.6 | <0.0001 | 78.6 | 80.5 | 81.9 | 80.1 | 0.25 |
| Education, % | | | | | | | | | | |
| <High school | 13.7 | 11.7 | 15.2 | 17.0 | | 11.3 | 13.3 | 14.8 | 18.2 | |
| High school/vocational | 41.3 | 44.4 | 42.1 | 42.6 | 0.01 | 41.7 | 42.8 | 42.7 | 43.2 | <0.0001 |
| College | 44.9 | 43.8 | 42.7 | 40.4 | | 47.0 | 43.9 | 42.5 | 38.6 | |
| BMI, kg/m ² (SD) | 29.1 (4.9) | 28.8 (5.2) | 28.6 (5.7) | 28.6 (6.7) | 0.15 | 28.8 (5.0) | 28.6 (5.3) | 29.2 (6.2) | 28.7 (6.1) | 0.19 |
| Smoking status, % | | | | | | | | | | |
| Current smoker | 4.0 | 5.1 | 6.8 | 10.0 | | 3.6 | 6.7 | 6.4 | 9.2 | |
| Former smoker | 54.5 | 56.8 | 50.1 | 50.8 | <0.0001 | 54.4 | 50.3 | 54.1 | 53.6 | <0.0001 |
| Never smoker | 41.4 | 38.1 | 43.2 | 39.1 | | 42.0 | 43.0 | 39.5 | 37.2 | |
| Pack-years of smoking among ever smokers, pack-years (SD) | 10.9 (16.3) | 11.9 (19.0) | 12.7 (21.8) | 15.4 (22.3) | <0.0001 | 10.5 (16.8) | 12.1 (21.2) | 12.2 (19.0) | 16.0 (22.2) | <0.0001 |
| Alcohol intake, % | | | | | | | | | | |
| Current drinker | 53.2 | 50.6 | 49.1 | 46.1 | | 53.0 | 50.6 | 49.0 | 46.6 | |
| Former drinker | 28.3 | 30.8 | 28.4 | 31.0 | 0.01 | 28.7 | 29.0 | 29.7 | 31.2 | 0.12 |
| Never drinker | 18.5 | 18.6 | 22.5 | 22.8 | | 18.3 | 20.4 | 21.3 | 22.2 | |
| Physical activity ^a , scores (SD) | 2.70 (0.8) | 2.64 (0.8) | 2.56 (0.8) | 2.39 (0.8) | <0.0001 | 2.71 (0.8) | 2.64 (0.8) | 2.54 (0.8) | 2.39 (0.8) | <0.0001 |
| Aspirin use in the preceding two weeks, % | 68.8 | 69.8 | 71.0 | 73.0 | 0.14 | 69.0 | 69.7 | 71.3 | 72.6 | 0.23 |
| Diabetes ^b , % | 40.5 | 36.3 | 34.0 | 39.2 | 0.01 | 32.4 | 35.3 | 38.5 | 43.8 | <0.0001 |
| Hypertension ^b , % | 75.5 | 74.9 | 75.6 | 82.1 | <0.0001 | 72.7 | 76.2 | 77.0 | 82.2 | <0.0001 |
| CVD ^b , % | 19.9 | 24.7 | 29.8 | 38.4 | <0.0001 | 19.6 | 23.3 | 28.4 | 41.5 | <0.0001 |
| eGFR, mL/min/1.73 m ² (SD) | 77.9 (13.8) | 73.6 (14.9) | 70.6 (16.7) | 59.9 (20.2) | <0.0001 | 76.2 (14.3) | 73.6 (15.9) | 70.5 (16.8) | 61.6 (20.3) | <0.0001 |

Abbreviations: PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

^aPhysical activity was assessed using a leisure-time sports index that ranged from 1 to 5. We reported physical activity scores with two decimal places to illustrate the trend more effectively.

^bAll the diseases were prevalent diseases.

^cP-values were calculated using chi-square for categorical variables and using ANOVA for continuous variables.

Table 5. The associations of age acceleration for the late-life ARIC and Lehallier’s PACs and the change in age acceleration from midlife to late life with mortality; ARIC (2011-2019)

| | No. of participants | No. of deaths | Total person-years | HR (95%CI) ^a per one SD of age acceleration | |
|---|---------------------|---------------|--------------------|---|---|
| | | | | late-life ARIC PAC (SD = 2.61 years) | late-life Lehallier’s PAC (SD = 2.54 years) |
| All-cause mortality | 4,553 | 1,123 | 29,356 | 1.65 (1.52, 1.79) | 1.58 (1.46, 1.72) |
| CVD mortality (Fine and Gray model) | 4,553 | 348 | 29,356 | 1.37 (1.18, 1.58) | 1.38 (1.19, 1.62) |
| Cancer mortality (Fine and Gray model) | 4,553 | 278 | 29,356 | 1.21 (1.02, 1.44) | 1.19 (1.02, 1.40) |
| LRD mortality (Fine and Gray model) | 4,553 | 128 | 29,356 | 1.68 (1.32, 2.12) | 1.57 (1.21, 2.03) |
| The change in age acceleration from midlife to late life^b | | | | | |
| | No. of participants | No. of deaths | Total person-years | HR (95%CI) per one SD of change in age acceleration (SD = 2.91 years) | |
| All-cause mortality | 2,707 | 736 | 17,081 | 1.71 (1.52, 1.94) | |
| CVD mortality (Fine and Grey model) | 2,707 | 239 | 17,081 | 1.38 (1.13, 1.68) | NA ^c |
| Cancer mortality (Fine and Grey model) | 2,707 | 172 | 17,081 | 1.30 (0.98, 1.71) | |
| LRD mortality (Fine and Gray model) | 2,707 | 94 | 17,081 | 1.46 (1.05, 2.04) | |

Abbreviations: PAC – proteomic aging clock; SD – standard deviation; BMI – body mass index; CVD – cardiovascular disease; LRD – lower respiratory disease; eGFR – estimated glomerular filtration rate. HR – Hazard ratio; CI – confidence interval.

^a The model was adjusted for chronological age, gender, joint terms for race and study center (Black participants from Mississippi; Black participants from any other centers; White participants from Maryland; White participants from North Carolina; and White participants from Minnesota), education, BMI, smoking status, pack-years of smoking, alcohol intake, physical activity, diabetes, hypertension, CVD, and eGFR at Visit 5.

^b The associations for the change in age acceleration was examined among the 2,707 participants who survived until Visit 5 after excluding the training sets at Visit 2 and at Visit 5 and the model was additionally adjusted for midlife age acceleration.

^c The associations between the change in age acceleration and mortality were examined using the ARIC PACs only because the ARIC PACs and published PACs showed similar associations with all mortality types.

Table 6. Association^a between midlife participants' characteristics and late-life age acceleration, i.e., age acceleration for the late-life ARIC PAC; ARIC

| Midlife participants' Characteristics ^b | Coefficients | P-value or P-trend ^c |
|--|--------------|---------------------------------|
| Chronological age | -0.03 | <0.0001 |
| Male | 0.05 | 0.65 |
| Black | -0.67 | <0.0001 |
| Education | | |
| <High school | 0 | |
| High school/vocational | -0.21 | 0.27 |
| College | -0.16 | |
| BMI | 0.04 | <0.0001 |
| Smoking status | | |
| Never smoker | 0.00 | |
| Former smoker | -0.25 | <0.0001 |
| Current smoker | 0.32 | |
| Pack-years of smoking | 0.01 | 0.0001 |
| Alcohol intake | | |
| Never drinker | 0 | |
| Former drinker | -0.14 | 0.04 |
| Current drinker | -0.27 | |
| Physical activity | -0.08 | 0.13 |
| Hormone replacement therapy | | |
| Female never user | 0 | |
| Female ever user | -0.23 | 0.04 |
| Male | 0 | |
| Hypertension | 0.45 | <0.0001 |
| CVD | 0.45 | 0.006 |
| Diabetes | 1.04 | <0.0001 |
| eGFR | -0.03 | <0.0001 |

Abbreviations: PAC – proteomic aging clock; BMI – body mass index; CVD – cardiovascular disease; eGFR – estimated glomerular filtration rate.

^aThe association was examined among the 4,553 participants who had information on the late-life ARIC PAC (after excluding the Visit 5 training set), and participants' characteristics were included into model simultaneously.

^bPhysical activity at Visit 1 was assessed using a leisure-time sports index that ranged from 1 to 5. We assumed that physical activity scores remained the same at Visit 1 and Visit 2. All the other characteristics were collected at Visit 2.

^cP-value for continuous variables and P-trend for categorical variables

Supporting information

S1 Appendix: Assessment of diseases and characteristics of interests, as well as procedures for identifying healthy participants.

S1 Fig: Pearson correlation (r) between midlife Tanaka's and Sathyan's proteomic aging clocks (PACs) and chronological age in healthy participants of the Visit 2 test set, ARIC. The x-axis depicts chronological age. The y-axis represents proteomic aging clock (PAC). (A) Tanaka's PAC was computed using ARIC weights obtained from Ridge regression based on proteins available in ARIC. The correlation between midlife Tanaka's PAC and chronological age was 0.66. (B) Sathyan's PAC was calculated using the published weights. The correlation between midlife Sathyan's PAC and chronological age was 0.58.

S2 Fig: Pearson correlation (r) between late-life Tanaka's and Sathyan's proteomic aging clocks (PAC) and chronological age in healthy participants of the Visit 5 test set, ARIC. The x-axis depicts chronological age. The y-axis represents proteomic aging clock (PAC). (A) Tanaka's PAC was computed using ARIC weights obtained from Ridge regression based on proteins available in ARIC. The correlation between late-life Tanaka's PAC and chronological age was 0.59. (B) Sathyan's PAC was calculated using the published weights. The correlation between late-life Sathyan's PAC and chronological age was 0.69.

S3 Fig: Overlap of aptamers included in the midlife and late-life ARIC proteomic aging clocks (PACs). The gray circle shows the aptamers included in the midlife ARIC PAC and the yellow circle shows the aptamers included in the late-life ARIC PAC.

S4 Fig: Association between age acceleration for the midlife ARIC PAC and mortality stratified by race, gender, and chronological age (in tertiles); ARIC (1990-2019)

S5 Fig: Association between age acceleration for the late-life ARIC PAC and mortality stratified by race, gender, and chronological age (in tertiles); ARIC (2011-2019)

S1 Table: Description of the ARIC and published proteomic aging clocks (PACs)

S2 Table: Description of the de novo ARIC proteomic aging clocks (PACs) constructed in middle-aged healthy participants; ARIC

S3 Table: Pearson correlation coefficients between the de novo ARIC proteomic aging clocks (PACs) constructed in middle-aged healthy participants and midlife published PACs among the Visit 2 test set of healthy participants

S4 Table: Including/excluding participants with controlled hypertension^a for healthy participants at Visit 5 to construct proteomic aging clocks (PACs) using elastic net regression

S5 Table: Pearson correlation and median absolute error (MAE) between the midlife and late-life Tanaka's and Sathyan's proteomic aging clocks (PACs) and chronological age, ARIC

S6 Table: R squared after regressing age acceleration for the midlife and late-life ARIC proteomic aging clocks (PACs) on covariates at the corresponding visits

S7 Table: Visit 2 participants' characteristics across quartiles of age acceleration for midlife Tanaka's and Sathyan's PACs; ARIC

S8 Table: The association between age acceleration for midlife Tanaka's and Sathyan's PACs and mortality; ARIC (1990-2019)

S9 Table: Pearson correlation coefficients between the late-life ARIC and published PACs in the Visit 5 test set of healthy participants

S10 Table: Visit 5 participants' characteristics across quartiles of age acceleration for late-life Tanaka's and Sathyan's PACs; ARIC

S11 Table: The associations of age acceleration for the late-life Tanaka's and Sathyan's PACs with mortality; ARIC (2011-2019)

S12 Table: The associations of age acceleration for midlife and late-life Sathyan's PAC with mortality in all ARIC participants

S13 Table: Top 20 proteins with the largest absolute weight in the midlife and late-life ARIC PACs

S14 Table: The association between age acceleration for the midlife ARIC and published PACs and mortality stratified by sex, race, and chronological age (in tertiles); ARIC (1990-2019)

S15 Table: The associations of age acceleration for the late-life ARIC and published PACs with mortality stratified by sex, race, and chronological age (in tertiles); ARIC (2011-2019)

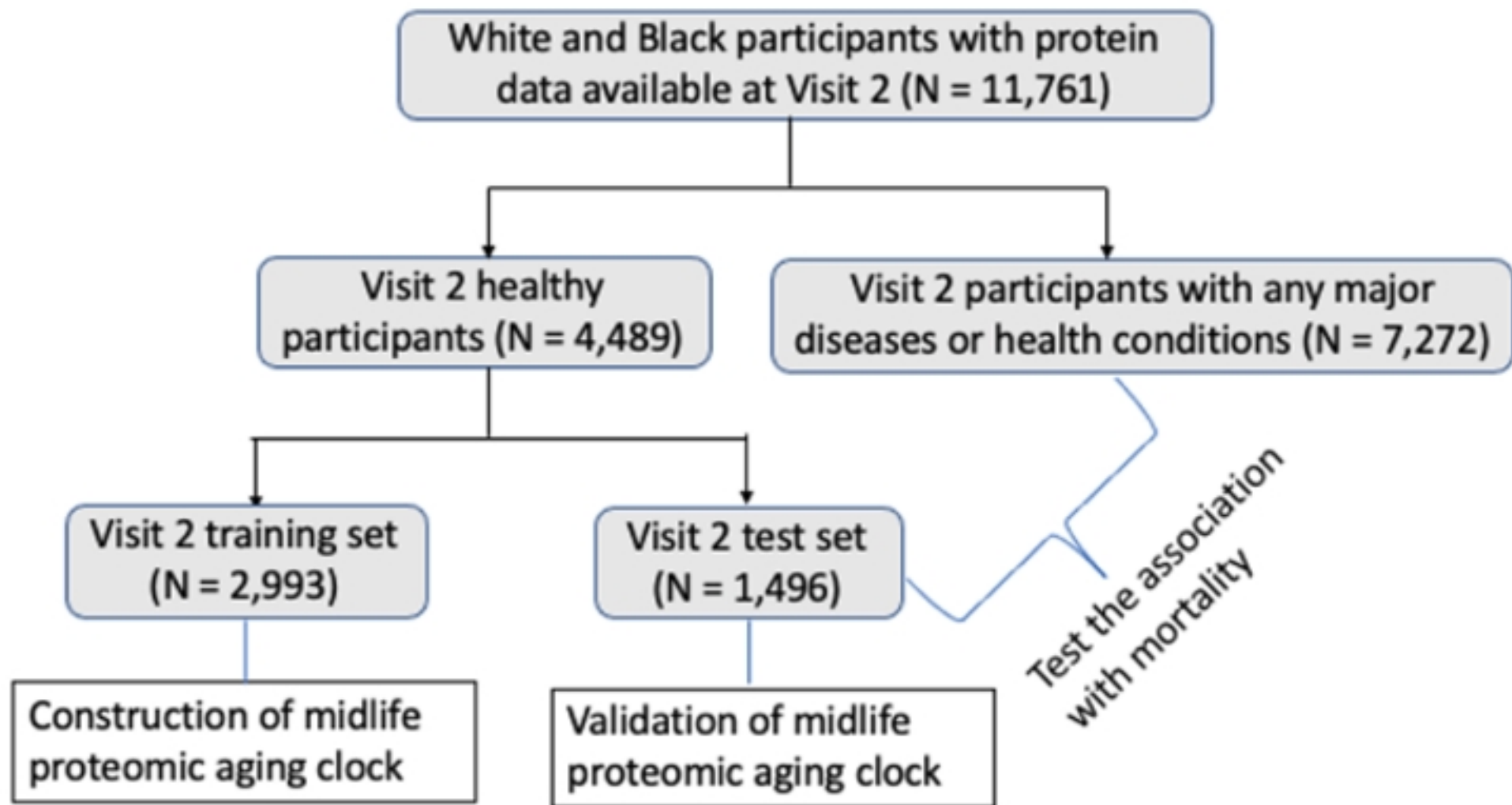


Fig 1

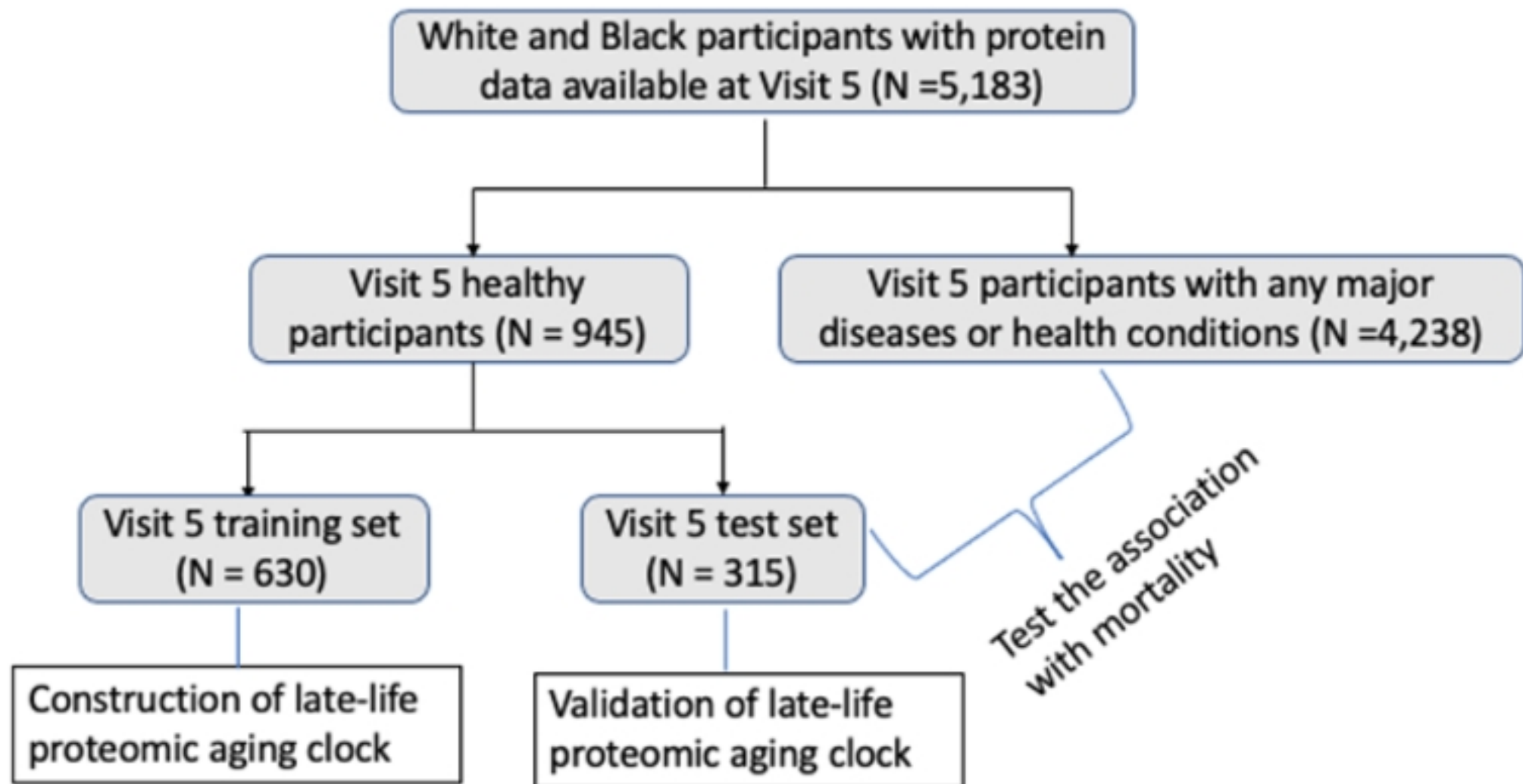


Fig 2

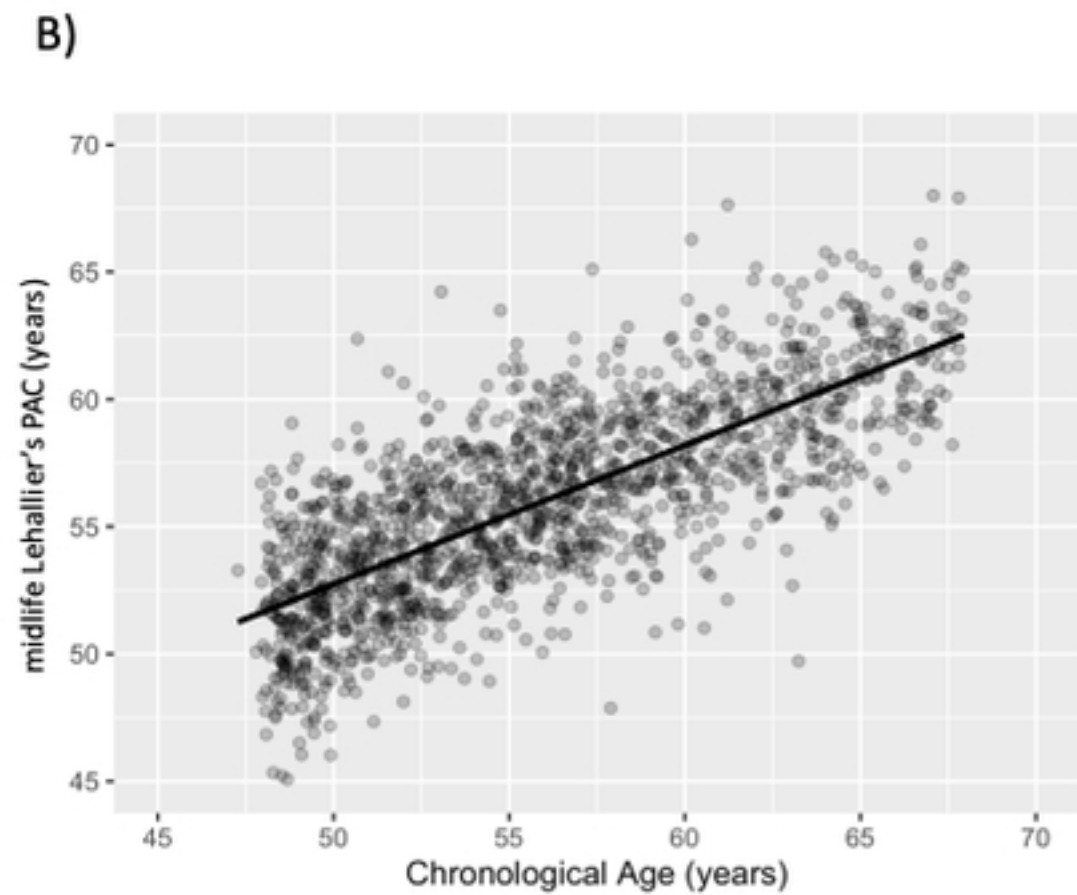
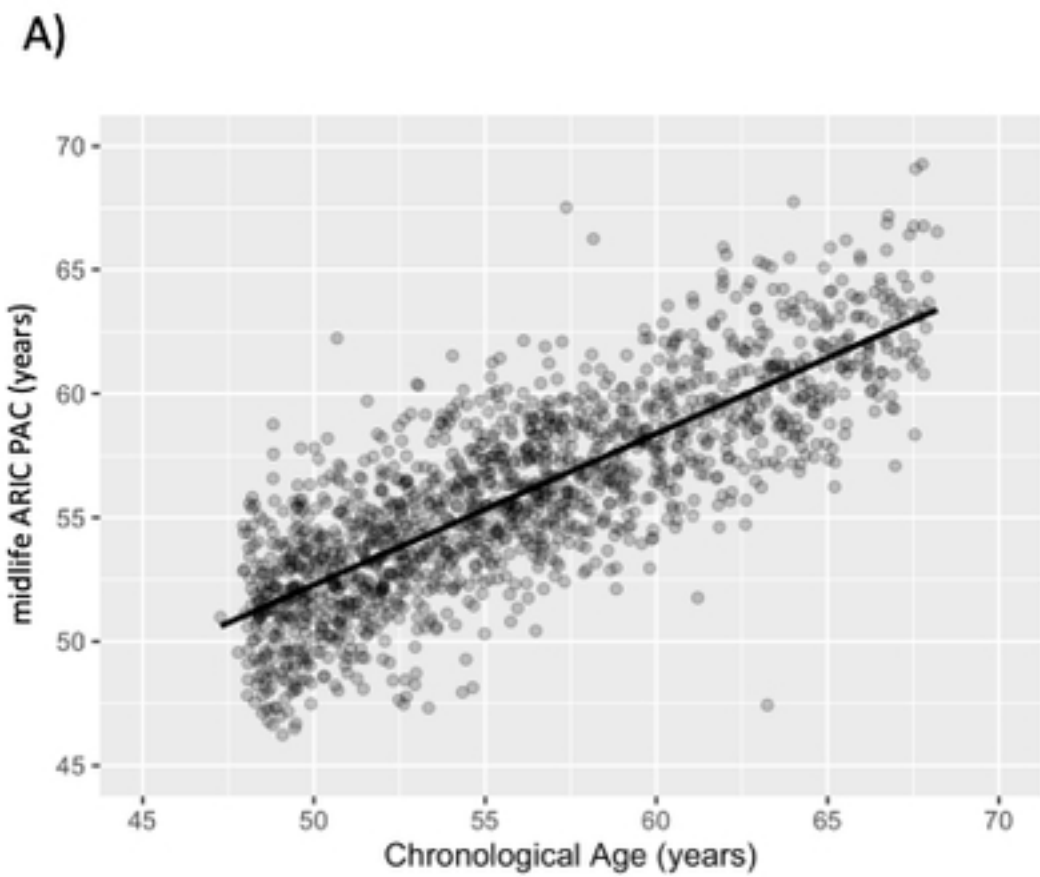
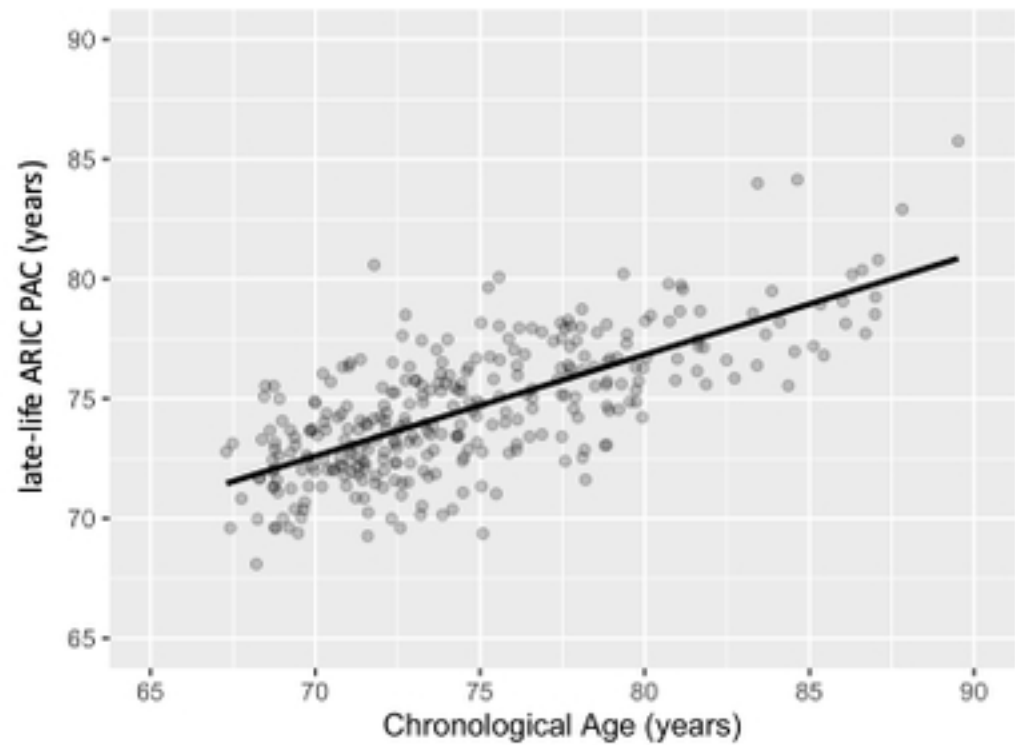


Fig 3

A)



B)

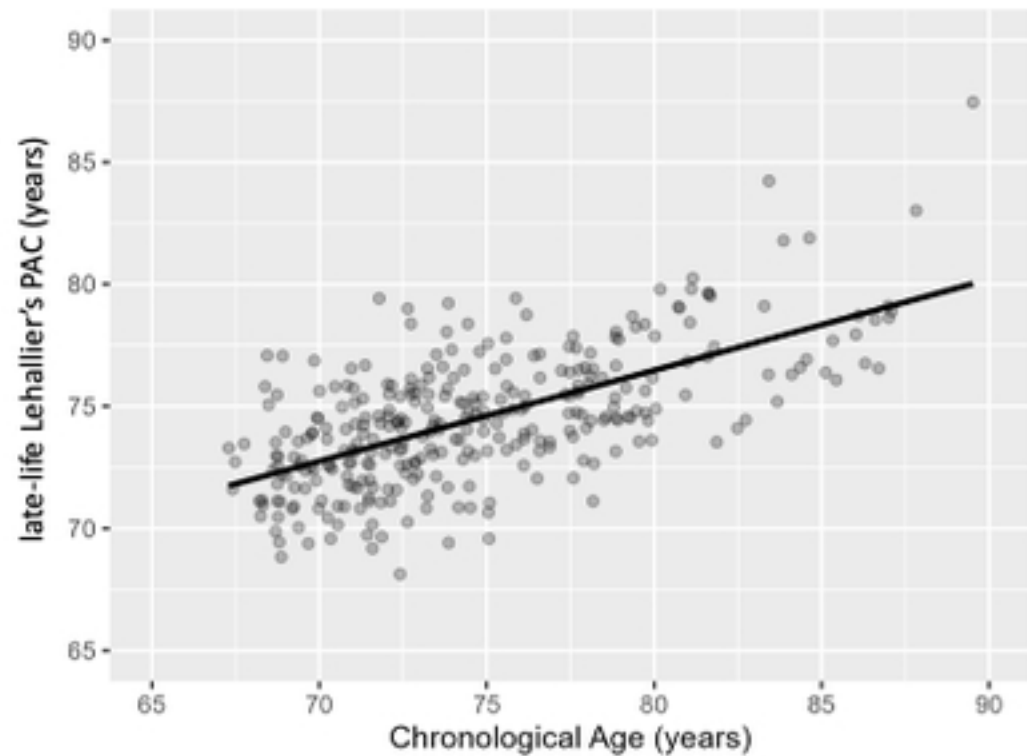


Fig 4