

Quantification of biological age from biological markers using SAGE aging dataset

Juan Rivillas, School of Public Health, Department. Biostatistics and Epidemiology.
MRC Centre for Environment and Health. Email: j.rivillas-garcia20@imperial.ac.uk

Last updated: 20th Sep 2022.

Supervisors: Paolo Vineis, MD, MPH, FFPH, Faculty of Medicine, School of Public Health. Chair in Environmental Epidemiology. Email: p.vineis@imperial.ac.uk
Emilie Courtin, Ph.D. Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine LSHTM. Email: emilie.courtin@lshtm.ac.uk

Objective: To quantify biological age of individuals in the SAGE dataset.

In this document, I explain how I computed biological aging (BA) measures using the BioAge package developed by Kwon et al. (Kwon and Belsky 2021). The goals of this analysis were (a) to compute the BA with a set of working biomarkers to represent Colombian population; and (b) to compare different BA prediction measures to elucidate the most fitting and applicable for the SAGE dataset. To achieve this, BA were examined by associations between BA measures with one another, biomarkers, chronological age (CA), exposures, and outcomes (HTA, CVD, DM, obesity). A set of ten biomarkers (LDL, HDL, HbA1C, total cholesterol, glucose, triglycerides, systolic blood pressure, diastolic blood pressure, waist circumference, and body mass index) were selected by analysing relations between them and chronologic age in 434 healthy individuals among 2812 individuals with 60 years or older. The Klemmera-Doubal (KDM) statistical methods and Homeostatic Dysregulation (HD) BA were applied to obtain four different sets of BA algorithms. Preliminary results shows that BAA are strongly correlated variables (KDM=0.96 and HD=0.83). Nine out of ten biomarkers were correlated with KDM advance, except diastolic blood pressure, childhood adversity (household violence and migration) and lower childhood SEP. Hypertension, and elevated BMI are correlated. In the adjusted models by age and sex, Correlation analysis was consistent using alternative packages. In the regression analysis, the KDM advance shown a perfect linear relationship in model 2 (KDM+ CA), which means a higher predictor power. Conversely, it seems that biological age (KDM and HD) did not mediate relationship between low childhood SEP, adult SEP, and risk behaviours and outcomes in the models (not significance and lower prediction power in the models). Overall, this finding indicates that the KDM BA adequately correspond to the individual health status in the SAGE dataset. The BioAge R package facilitated the computation of biological age algorithms and tests different criteria to validate methods in the SAGE datasets.

This document is structured as follows:

- 1) Biological age and BA methods
 - Klemmera-Doubal method (KDM)
 - Homeostatic dysregulation (Mahalanobis distance MD)
 - 2) Methods and statistical analysis
 - 3) Results
 - Exploratory data analysis KDM BA and HD MD (Histograms and boxplots)
 - Correlation analysis BA measures (Scatter plots and correlation matrix)
 - Biomarkers and CA.
 - BA and CA
 - BA with biomarkers.
 - Associations between BA measures with one another.
 - Association BA measures with health outcomes.
 - Analyses tests socioeconomic patterning of biological aging algorithms.
 - 4) Conclusions
- References

1) Biological age and BA methods

The age of an individual can be estimated by either calculating chronological age (CA) or biological age (BA). Normal developmental phases and rates can be estimated by measuring elapsed time since birth (Jee et al., 2012). Although CA provide a simple, clear-cut method for estimating aging, CA does not provide adequate information on the rate of decline or physiological breakdown. The concept of biological aging was proposed to provide a reliable estimation of the degree of aging process (Levine 2013; Buslova 2017; Jee and Park 2017). Biological aging is the gradual and progressive decline in system integrity that occurs with advancing chronological age. Processes of biological aging begin with accumulation of cellular-level changes that increase vulnerability of tissues and organs to loss of function, ultimately causing disease, disability, and death (Kwon and Belsky 2021; Robinson et al. 2020). Although researchers agreed there is still no gold standard biomarker of aging, validation of BA has always been a controversial issue, as the term itself is an abstract concept, and the absence of the true value in the reality makes it difficult to evaluate the validity of the estimated BA (Cho, Park, and Lim 2010). Scholars have offered options how to deal with the validation issue developing and improving algorithmics for BA predictions.

There have been various attempts on assessing BA by a range of computational algorithms. The common assess to BA estimation is through measurements of various age-dependent markers as variables, and aggregating a selection of an optimal set of biomarkers into a value in units of years with some computational algorithms (Cho, Park, and Lim 2010; Jee and Park 2017; Kwon and Belsky 2021; Chadeau-Hyam et al. 2020; Karimi, Maryam; Castagne, R; Delpierre 2019). The algorithms include multiple linear regression (MLR) methods, principal component analysis (PCA), Hochschild method (Cho, Park, and Lim 2010), PhenoAge, Homeostatic Dysregulation (Kwon and Belsky 2021), Biological Health Score (BHS) (Karimi, Maryam; Castagne, R; Delpierre 2019; Chadeau-Hyam et al. 2020), and Klemmera and Doubal method (Kwon and Belsky 2021; Jee and Park 2017; Cho, Park, and Lim 2010), Allostatic Load, etc. They asserted that the validation of BA is ascertained by examining the validity of biomarkers from which the BA is estimated; they are construct validity and predictive validity. Construct validity refers to how well a candidate biomarker reflects the construct, and predictive validity refers to the usefulness of a biomarker (Ingram, 1988). KDM BA and HD methods have demonstrated links with morbidity, mortality, and health in older populations in previous publications (Jee and Park 2017; Buslova 2017; Kwon and Belsky 2021; Klemmera and Doubal 2006; Levine 2013).

Klemmera-Doubal method (KDM) Biological Age. An individual's KDM BA prediction corresponds to the chronological age at which individual physiology would be approximately normal. KDM BA older than chronological age indicates an advanced state of biological aging and increased risk. KDM BA younger than biological age indicates delayed biological aging and reduced risk for disease, disability, and mortality. KDM BA prediction model has been found to provide the most reliable and stable results for the practical assessment of BA (Cho, Park, and Lim 2010) and useful for studying a number of questions regarding the biology of aging (Levine 2013). Furthermore, when included with chronological age in a model, Klemmera and Doubal method had more robust predictive ability and caused chronological age to no longer be significantly associated with mortality.

Homeostatic Dysregulation. An individual's HD value corresponds to how different their physiology is from a healthy reference. Higher values of HD indicate an advanced state of biological aging and increased risk for disease, disability, and mortality. Lower values of homeostatic dysregulation indicate delayed biological aging and reduced risk. The Mahalanobis distance (MD) of biomarkers has been proposed as a proxy of physiological dysregulation. MD considers relation between biomarkers measurements as all of them are simultaneous play part in BA. The general idea is to measure how far a given markers deviates from the other vector in the data set (Flores-Guerrero et al. 2021). Although, a prediction of age calculated by Mahalanobis distance could be used only as rough estimation, evidence suggests that MD is able to provide information about physiological dysregulation (Buslova 2017).

The details of BA methods are discussed in the Table 1.

To date, an optimal working model which predicts BA with a set of working biomarkers has not been formulated to represent the Colombian population. Using the data obtained from the SABE aging survey, I examine Klemmera-Doubal method (KDM) Biological Age and Homeostatic Dysregulation, and ultimately, I propose the most suitable method for valid BA estimation to this dataset. Finding the most fitting and applicable BA estimation will be added as mediators to a set of analyses testing socioeconomic patterning of biological aging in my Ph.D. research.

Table 1. Definition and estimation of Klemra-Doubal method (KDM) Biological Age and Homeostatic Dysregulation.

Item	Klemra-Doubal method (KDM)	Homeostatic Dysregulation (HD)
Description	It has been named as a new approach to the concept and computation of biological age, which it was developed by mathematically modeling the BA estimation procedure.	In order to better represent an individual's degree of aging, a model must consider relation between biomarkers measurements as all of them are simultaneous play part in BA. The general idea is to measure how far a given vector deviates from the other vector. The distance gives understanding of similarity between them and considers the correlations of the data set.
Assumption or questions	The merits of the work are in its attempt to find the optimum way of computation for hypothetical BA estimation, and in its applicability for nonlinearity of certain biomarkers.	Higher values of HD indicate an advanced state of biological aging and increased risk for disease, disability, and mortality. Lower values of homeostatic dysregulation indicate delayed biological aging and reduced risk for disease, disability, and mortality.
Methods	<p>The KDM BA algorithm is derived from a series of regressions of individual biomarkers on chronological age in a reference population. The equation takes information from n number of regression lines of chronological age regressed on n biomarkers.</p> <p>Using the mathematically formulated definitions, the authors derived the optimum estimate of BA by minimizing the distance of point determined by values of biomarkers from one-dimensional line (or curve) determined by the regression functions in the space of all biomarkers (refer to their paper for elaborated descriptions).</p>	HD is computed as the Mahalanobis distance for a set of biomarkers relative to a reference sample.
Procedures	<p>The series of regressions take the form:</p> $BA_{EC} = -\frac{\sum_{i=1}^n (x_i - q_i) \frac{k_i}{s_i^2} + \frac{CA}{s_{BA}^2}}{\sum_{i=1}^n \left(\frac{k_i}{s_i^2} \right)^2 + \frac{1}{s_{BA}^2}}$ <p>where x is the value of biomarker i measured for an individual. For each biomarker i, the parameters k, q, and s are estimated from a regression of chronological age on the biomarker in the reference sample. k, q, and s are the regression intercept, slope, and root mean squared error, respectively. sBA is a scaling factor equal to the square root of the variance in chronological age explained by the biomarker set in the reference sample. CA is chronological age.</p>	<p>The Mahalanobis distance equation takes the form:</p> $D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}$ <p>where x is a multivariate observation (all the biomarker values for an individual) and μ is the equivalent-length vector of reference sample means for each variable. S is the reference sample variance-covariance matrix for the variables. If all variables are uncorrelated then this is equivalent to scaling each biomarker by its variance and then summing the squared deviance for an observation:</p> $D_M(\vec{x}) = \sqrt{\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma^2(x_i)}}$ <p>where n is the number of biomarkers and $\sigma^2(x_i)$ is the variance in the ith biomarker. In the hd_nhanes function, we specify the reference sample (healthy individuals aged 20-30 years for whom all user- selected biomarkers fall within the clinically normal range). For analysis, all biomarkers are standardized to have mean=0, SD=1 separately for men and women based on this reference sample. This approach computes HD relative to a young, healthy sample, following the approach we have used previously.</p>
Validation	Using the mathematical relations among BA, CA, and biomarkers, KDM BA defined and constructed hypothesis to demonstrate the validation of their work by computer-generated simulations. They provided two alternatives for the optimum estimate of BA, one without CA, and one with applying CA as another biomarker, but in a different way as the other biomarkers.	
Advantages	More robust predictive ability. Given the potential of BA to highlight heterogeneity, KDM may be useful for studying several questions regarding the biology of aging.	The Mahalanobis distance (MD) of biomarkers has been proposed as a proxy of physiological dysregulation.
Limitations	The outstanding mathematical model developed by Klemra and Doubal, however, requires some complicated calculations.	A prediction of age calculated by Mahalanobis distance could be used only as rough estimation, although evidence suggests that MD is able to provide information about physiological dysregulation.
Packages in R	The BioAge R package is an easy to install tool that can implement the Klemra-Doubal and homeostatic dysregulation methods following the Dwoon's approach used in previous work. The package includes datasets for training and testing algorithms from the US Health and Nutrition Examination Surveys (https://wwwn.cdc.gov/nchs/nhanes/Default.aspx). The package is available on GitHub (http://github.com/dayoonkwon/BioAge) and is licensed under the GNU General Public License v3.0.	
Overview recommended	An overview on KDM BA with two practical examples: i) Dayoon Kwon et al. (2021) using data from Nutrition Survey in the U.S. to estimate BA using BioAge package; ii) Haeng-Choi et al (2010) using Korean male data to estimate BA.	An overview on KDM BA with two practical examples: i) Dayoon Kwon et al. (2021) using data from Nutrition Survey in the U.S. to estimate BA using BioAge package; ii) Flores-Guerrero et al (2021) using MD as novel statistical proxy of homeostasis loss in the risk of diabetes.

2) Methods and statistical analysis

SABE ageing data

SABE is a nationally representative and cross-sectional survey conducted by the Ministry of Health and Social Protection of Colombia. The sample was a probabilistic, cluster, stratified and multi-stage. This population-based survey is the first wave of the ageing studies, and it provides information on health, anthropometric measures, blood samples, demographic characteristics, early-life inequality, environmental, lifestyles, and socioeconomic exposures SEP (Gomez et al., 2016) (Ministerio de Salud y Protección Social, 2018). The SABE survey is a cross-sectional study representing the non-institutionalized population of men and women at age 60 and older. The total sample size is 23,694 individuals and biomarkers sub-sample size is 4,092. For BA analysis, after applying wise deletion strategy individuals with incomplete biomarkers data were excluded leaving a new dataset with 2,812 individuals.

Statistical package

The BioAge is an excellent package for the quantification of biological aging based on analysis of chronological age and mortality risk: Klemmera-Doubal Biological Age, PhenoAge, and Homeostatic Dysregulation. In this analysis, KDM BA (*kdm_calc function*) and HD (*hd_calc function*) were computed, while Levine PhenoAge and its modifications were excluded because SABE does not provide mortality data. The BioAge R package is fully available for download at GitHub (<http://github.com/dayoonkwon/BioAge>). Additional packages were used to complete and validate hypotheses testing and dispersion (*ggbetweenstats*), correlation analysis (*correlation*), analysis of variance (*ANOVA*), ranked cross validation of correlations (*Lares*).

Estimation methods

The description of set of functions applied to generate biological age algorithms are described in the Table 1 and in-depth descriptions are published elsewhere (Kwon and Belsky 2021; Cho, Park, and Lim 2010; Klemmera and Doubal 2006; Jee and Park 2017; Buslova 2017). BioAge functions make it possible to train new algorithms using the NHANES datasets and then to project these algorithms onto SABE datasets (Original data, biomarkers subsample dataset after imputation, and case complete dataset).

The SABE's analyses proceed in six steps:

1. To check dispersion and looking for skew in BA (*using ggbetweenstats package*)
2. To test correlation among markers and CA (*using correlation package*)
3. To test association BA with CA (*using plot_ba function with age vs bioage vars*)
4. To test correlation among BA with one another (*using plot_baa function with BAA corplot*)
5. To test association BA with outcomes (*using table_health*)
6. Finally, a set of analyses tests socioeconomic patterning of BA (*using table_ses*)

There are two sets of “*plot_*” functions that were used to create comparative scatter plots using Pearson correlations: “*plot_ba*” function tests associations of chronological age with biological aging measures.

In the KDM BA and HD values were differenced from chronological age values and then standardized to have M=0, SD=1 separately for men and women. For linear regression analysis, outcome variables and coefficients on biological age variables were standardized to have M=0, SD=1 for analysis. Coefficients reported are interpretable on a Pearson's *r* scale and Spearman (categorical variables). “*plot_baa*” function tests associations among biological aging measures. In this function, KDM BA was computed as the difference between biological age and chronological age. KDM Biological Age measures were differenced from chronological age for analysis. These differenced values were then standardized to have M=0, SD=1 separately for men and women so that effect-sizes are denominated in terms of a sex-specific 1 SD unit increase in biological age advancement. Models included covariates for chronological age and sex.

The package tests associations of socioeconomic circumstances measures (childhood and adult SEP) with biological aging measures using linear regression. In these models, measures of socioeconomic circumstances are specified as independent variables and biological aging measures are specified as dependent variables. Socioeconomic circumstance measures are standardized to M=0, SD=1 for analysis so that effect-sizes are denominated in terms of a 1 SD unit improvement in socioeconomic circumstances. Using three set functions: “*table_*”, “*table_health*”, and “*table_ses*”. Linear regression is used to compute standardized beta coefficients (interpretable as Pearson's *r*) stratifying by gender, race, and age.

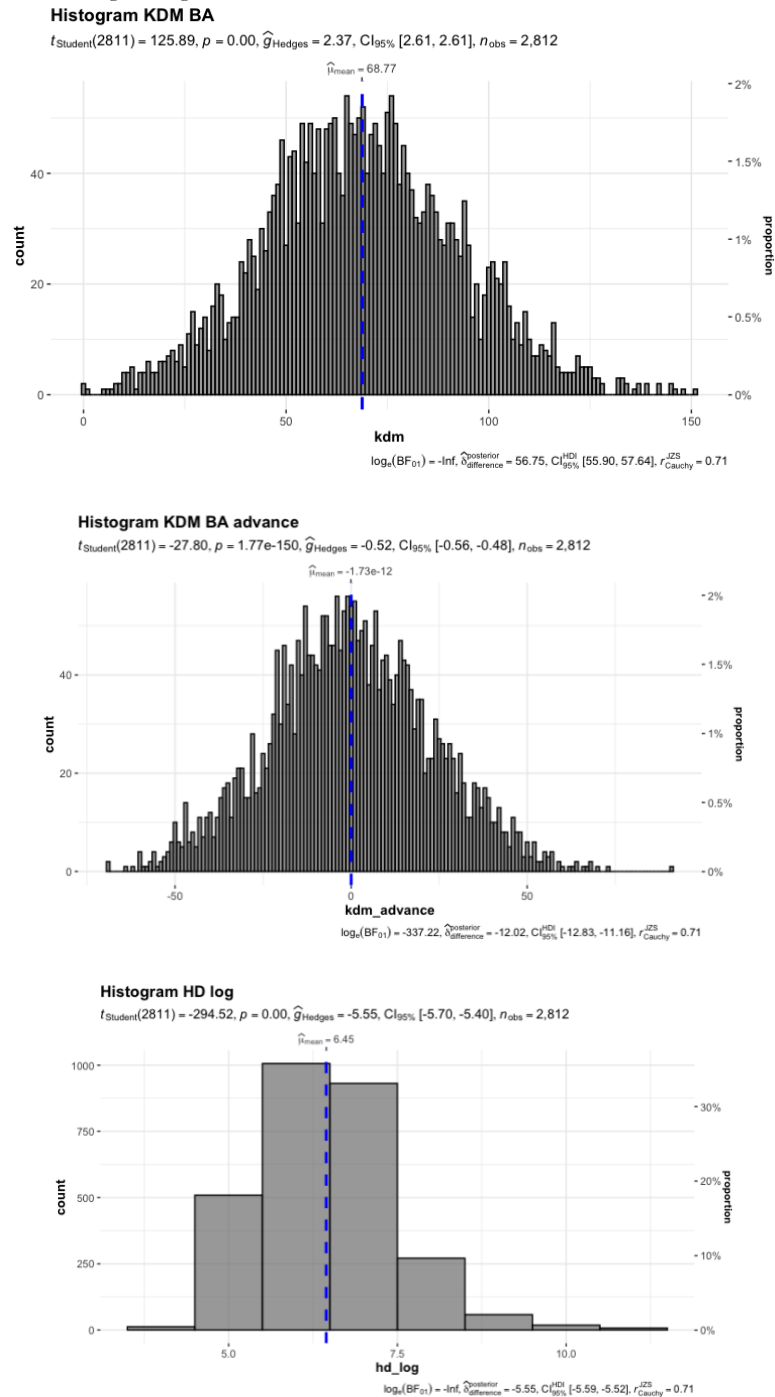
T-student and one-way ANOVA is used to compare three or more than three categorical groups to establish whether there is a difference between them, comparing the means of the samples. Scatter plots, histograms, violin plots, and correlation matrix were used to explore dispersion, skew, and correlations in the datasets. Hypothesis testing and matrix correlation of BA estimations are described in the code available at GitHub: https://github.com/juanrivillas/Childhood-SEP-and-Biologica-Age/blob/main/Scripts/hypothesis_testing.Rmd

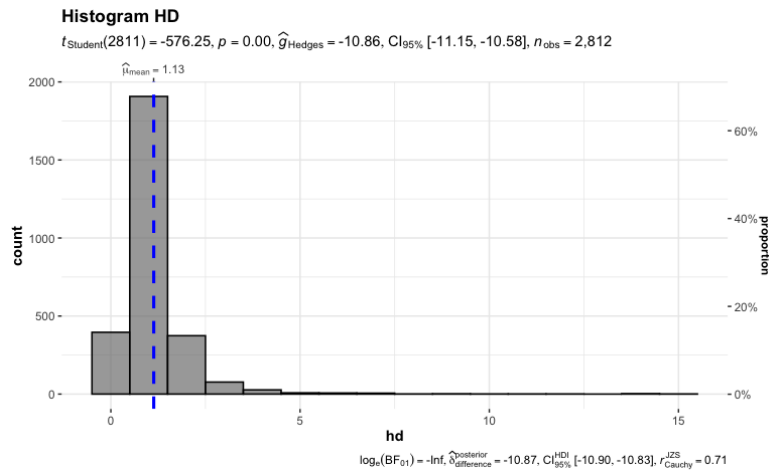
3) Results

Distribution of the biological age measures.

Figure 1 illustrates the distribution of four BA variables allowing to check if their means are significantly different from a specified value with a one-sample test (t student). KDM, KDM advance, and HD log show a symmetrical distribution. HD is positive skewed (mode < median < mean).

Figure 1. Distribution of Biological Age measures.





Associations biomarkers and CA

Descriptive statistics of the data Table 2 shows the mean and standard deviation (SD)s of the biomarker by diseases derived from the SABE dataset. Pearson correlations were used to assess the relationships of the ten potential biomarkers with age. Six biomarkers were significantly correlated with CA: LDL, total cholesterol, glycated hemoglobin, triglycerides, systolic blood pressure, and body mass index.

Table 2 Means and standard deviations of the biomarker's variables by disease and correlations with CA.

Biological marker	Diabetes				Hypertension				Cardiovascular disease				Pearson Correlation With CA
	Yes (n=469)		No (2,343)		Yes (n=1,503)		No (n=1,309)		Yes (n=385)		No (2,427)		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
LDL	120.44	35.75	127.94	35.35	123.31	36.39	130.57	34.09	118.41	70.20	128.00	34.95	***
HDL	42.75	12.01	45.44	13.18	44.78	12.68	45.24	13.42	43.86	13.07	45.17	13.01	
Total cholesterol	189.53	43.22	196.23	40.73	192.12	42.18	198.5	39.83	186.68	44.84	196.45	40.47	***
Glucose	138.45	68.63	94.01	20.45	103.70	38.14	98.80	36.61	105.54	43.25	100.76	36.48	*
HbA1c	13.94	1.72	14.03	1.63	13.86	1.65	14.19	1.62	13.91	1.66	14.03	1.64	***
Triglycerides	188.87	117.94	161.97	98.26	170.35	99.8	161.99	102.64	173.96	103.76	165.27	102.01	***
Body Mass Index	30.80	7.02	29.34	7.40	30.37	7.40	28.69	7.20	29.90	6.78	29.54	7.44	***
Waist circumference	97.37	10.75	92.35	10.67	94.78	10.73	91.37	10.70	94.96	10.82	92.91	10.83	
Systolic blood pressure	138.42	23.93	138.73	23.17	140.58	24.74	136.5	22.76	136.9	24.70	138.96	23.79	***
Diastolic blood pressure	83.10	23.88	82.92	23.66	83.85	24.59	81.92	22.58	85.00	26.07	82.62	23.28	

M = mean, SD = standard deviation

Note: ***p < .0001. **p < .01. *p < .05.

Associations BA and CA

The relationships between biological age and chronological age of four BA algorithm were examined using scatter plots along with fitted linear regression lines and 95% CI. SABE participants' KDM BA was correlated with chronological age (Pearson $r = 0.28$). SABE participants' KDM BA V2 (plus CA) was uncorrelated with chronological age and HD values were slightly negatively correlated with their chronological ages (HD Pearson $r = -0.052$ and HD log $r = -0.067$). Figure 3 presents associations of Klemere-Doubal method (KDM) and homeostatic dysregulation (HD) measures of biological age with CA among participants in the SABE ageing dataset using BioAge package. The figure plots values of the two biological aging measures against chronological age for men (blue) and women (pink) ($n = 2812$).

Figure 3 Associations of Klemere-Doubal's method (KDM) Biological Age and homeostatic dysregulation (HD) measures of biological age with chronological age among participants in the SABE ageing dataset.

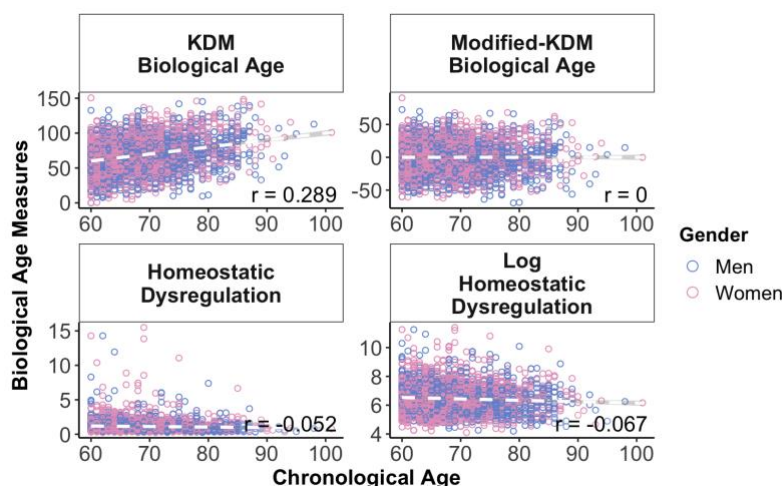
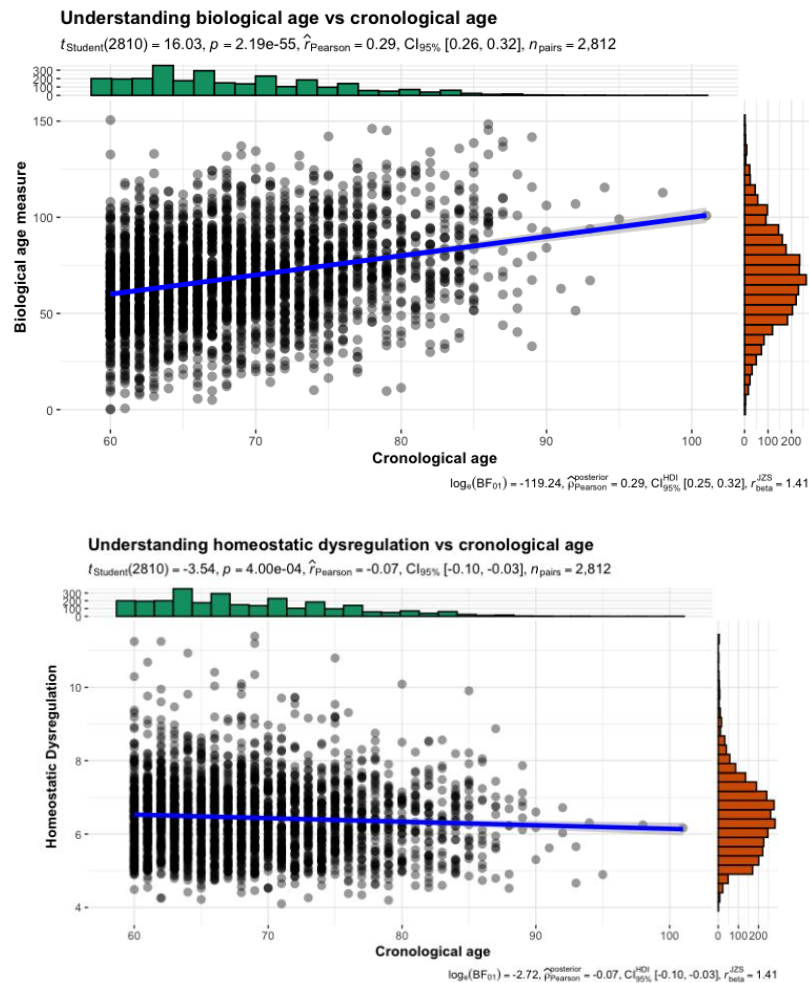


Figure 4 shows associations between BA and CA using alternative packages to double-test these associations (*ggbetweenstats* and *correlation* package).

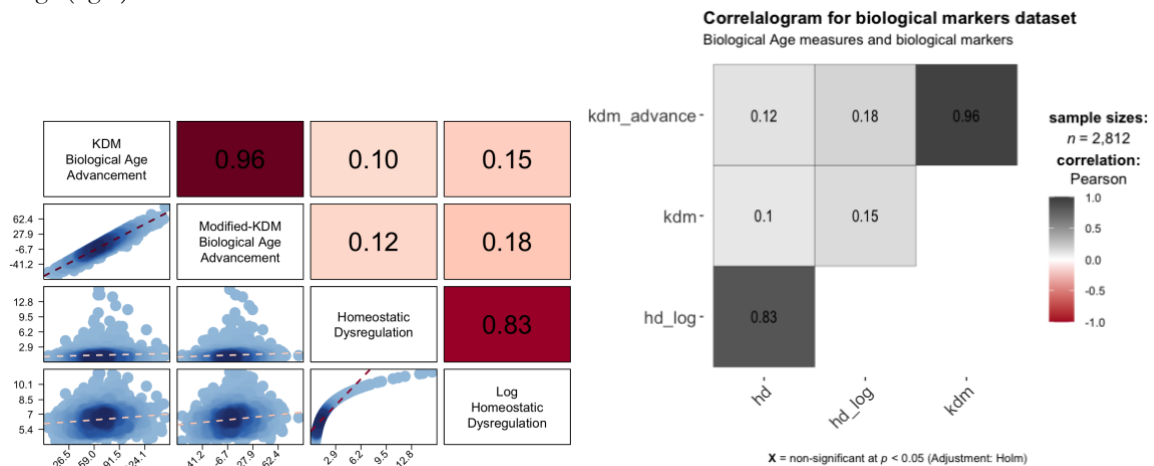
Figure 4 associations between KDM and HD with CA using *ggbetweenstats* package.



Associations between BA measures with one another.

Figure 5 illustrates association between BA and CA. A correlation plot (also referred as a correlogram) allows to highlight the variables that are most (positively and negatively) correlated. The correlogram represents the correlations for all pairs of variables. Positive correlations are displayed in blue and negative correlations in red. The intensity of the color is proportional to the correlation coefficient so the stronger the correlation (i.e., the closer to -1 or 1), the darker the boxes. A negative correlation implies that the two variables under consideration vary in opposite directions, that is, if one variable increases the other decreases and vice versa. A positive correlation implies that the two variables under consideration vary in the same direction, that is, if one variable increases the other increases and if one variable decreases the other decreases as well. The color legend on the right-hand side of the correlogram shows the correlation coefficients and the corresponding colors. Below KDM and KDM advance, and HD and HD log are strongly correlated variables (KDM=0.96 and HD=0.83, respectively). KDM has the higher the correlation, which means is a stronger estimation to reflect the BA of the individuals.

Figure 5. Correlation matrix of biological age and chronological age using *BioAge* package (left) and *ggbetweenstats* package (right).



Associations BA with biomarkers.

KDM BA was found to be correlated with all biomarkers (except diastolic blood pressure) and highly correlated with LDL (0.77) and total cholesterol (0.78). Correlation matrix using alternative packages and datasets are consistent with these associations.

Figure 5. Correlalogram for BA with biological markers dataset using BioAge package (top) and *ggbetweenstats* package (bottom).

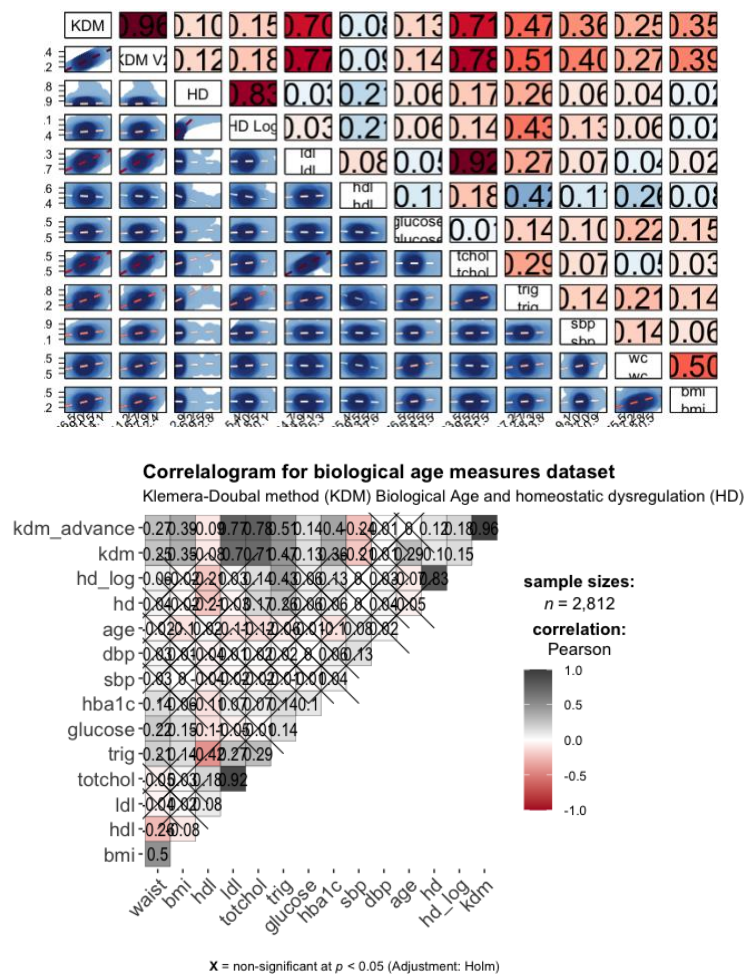
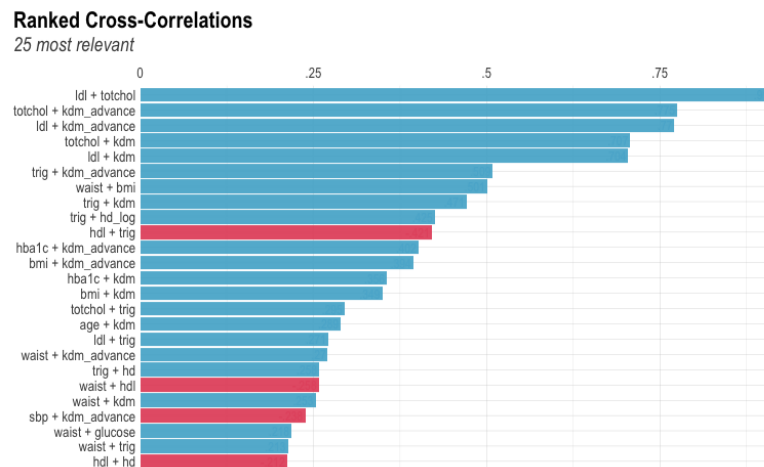


Figure 6 shows 25 most relevant correlations using correlation package, which are consistent with correlation matrix: KDM BA is highly correlated with LDL and total cholesterol.

Figure 6. Ranked cross-correlations (25 most relevant).



Analyses tests socioeconomic patterning of biological aging algorithms.

Table 4 reports Biological Age measures and biomarkers (means and SD) of the SABE participants by high-risk groups of ACEs and childhood SEP to examine differences. Cronological age is significant with lower childhood SEP and KDM BA had highly statistically significant result with high-risk childhood adversity and lower childhood SEP ($p < 0.001$). This means we can reject the null hypothesis. HD, HD log, and KDM v1 were not significant with childhood adversity and childhood SEP. Mean and SD of CA was 68.90 (6.97) in the high-risk childhood adversity group and 68.02 (6.59) in the low childhood SEP.

Figures 7 and 8 show the distribution of KDM advance (plus CA) by childhood adversity and childhood SEP groups and sex. There are not significant differences by sex.

Table 3 provides comparisons of the distribution between Chronological Age (CA) and estimates of Biological Age (BA) for the sample and by childhood SEP Group.

Tables 3 Biological Age measures and biomarkers of the SABE Participants included in analysis by high-risk groups of ACEs and childhood SEP

Biological markers	Adverse Childhood Experiences				<i>P</i> -value	Low childhood SEP				<i>P</i> -value
	Low-risk group		High-risk group			No		Yes		
	(n=2,378)		(n=434)			(n=1,075)		(1,737)		
	M	SD	M	SD		M	SD	M	SD	
Cronological age (CA)	68.90	6.97	68.02	6.59	0.01	69.61	7.22	68.24	6.67	<i>p</i> = .001
BA from Klemera-Doubal method KDM (10 biomarkers)	68.39	24.04	70.83	23.11	0.05	67.99	24.64	69.24	23.44	0.18
BA from KDM modified (10 biomarkers + CA)	-0.51	23.04	2.80	21.85	<i>p</i> = .001	-1.61	23.67	0.99	22.34	<i>p</i> = .001
BA from Homeostactic Dysregulation HD	1.12	0.99	1.19	1.03	0.19	1.11	0.97	1.14	1.01	0.57
BA from HD log	6.43	0.99	6.53	1.03	0.05	6.44	0.98	6.44	1.01	0.83
LDL	125.86	34.97	129.07	32.15	0.06	125.52	35.14	126.87	34.21	0.32
Serum high-density lipoproteins (mg/dL)	44.43	12.11	44.86	11.74	0.48	44.84	12.15	44.28	11.98	0.23
Total cholesterol (mg/dL)	194.13	40.37	197.79	36.37	0.06	193.29	40.12	195.57	39.65	0.14
Glycated haemoglobin (%)	13.94	1.63	14.14	1.51	0.01	13.84	1.65	14.04	1.58	<i>p</i> = .001
Glucose	94.83	13.45	94.81	11.35	0.98	94.53	12.84	95.01	13.33	0.34
Triglycerides	147.48	54.79	152.10	51.47	0.09	147.60	55.32	148.56	53.68	0.65
Body Mass Index	28.59	5.72	29.11	5.9	0.09	28.04	5.6	29.06	5.8	<i>p</i> = .001
Waist circumference (cm)	93.07	10.63	92.80	10.58	0.63	92.23	10.68	93.52	10.56	<i>p</i> = .001
Systolic blood pressure (mmHg)	138.98	23.95	137.06	23.72	0.12	139.02	24.29	138.48	23.7	0.56
Diastolic blood pressure	83.01	23.87	82.68	22.68	0.78	82.31	23.00	83.35	24.10	0.25

M = mean, SD = standard deviation

Figure 7. Boxplot to highlight the childhood adversity and SEP groups means.

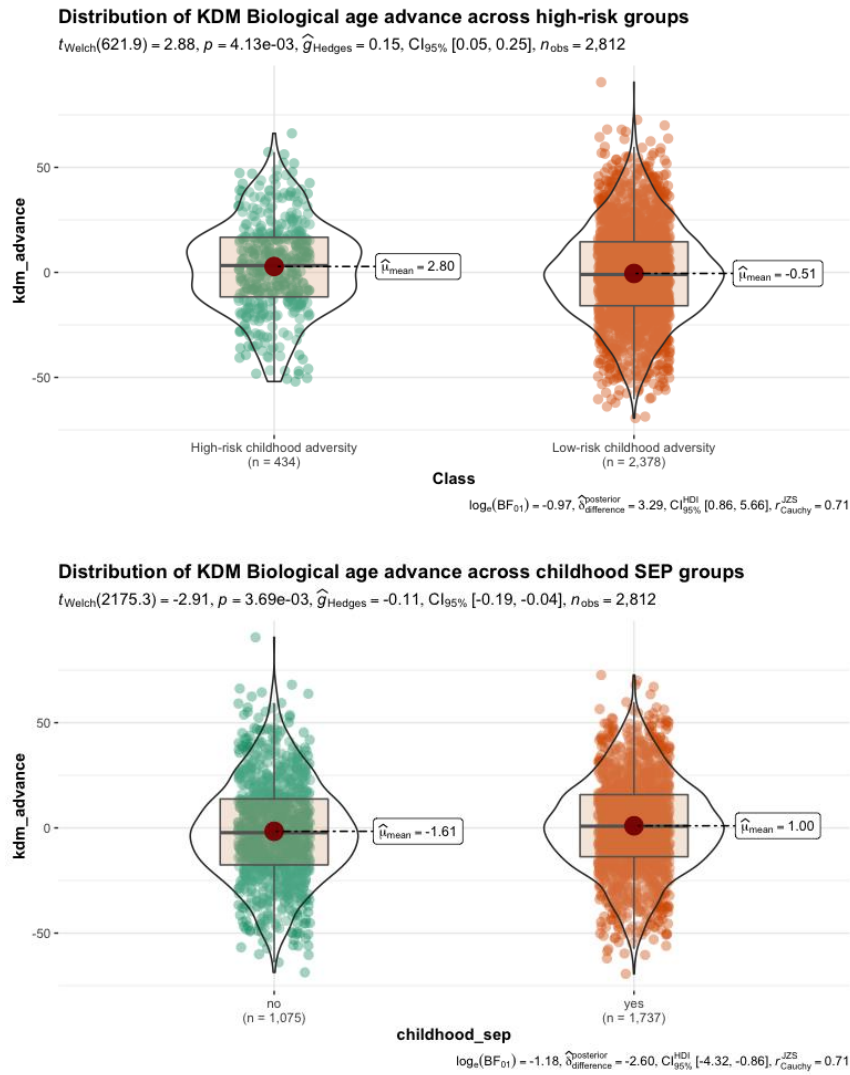
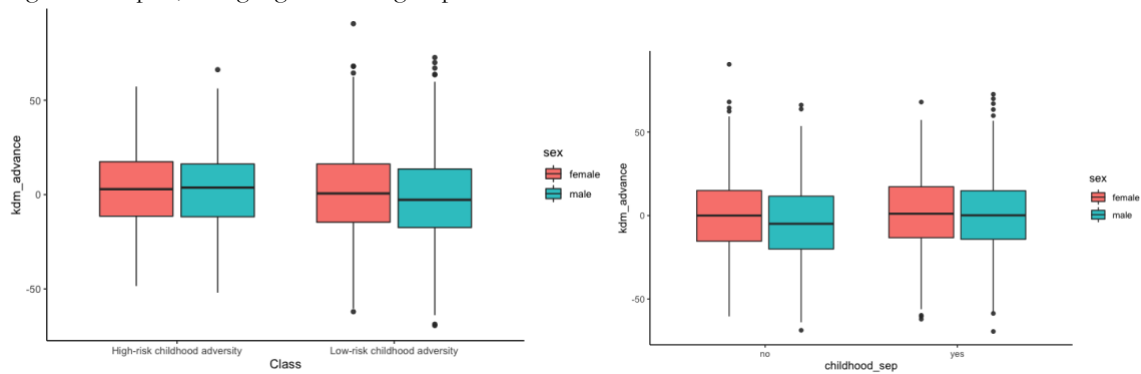


Figure 8 boxplot, to highlight the sex group means



Additional analysis:

- Correlation ACEs and outcomes
- Hypothesis testing's Childhood, adult SEP, and adult diseases (ANOVA results)

Figure 9 illustrate 25 most relevant correlation between several dimensions of ACEs and diseases. Now we check the top correlations of variables and values ranked in descending order. From the plot above we can extract some interesting insights:

- Domestic violence, food insecurity, poor health status, and emotional abuse are inversed correlated with low-risk group of childhood adversity.
- Poor food environment is commonly among lower childhood sep individuals and domestic violence in childhood.
- Individuals with cardiovascular diseases are commonly to report hypertension as well.
- Lower childhood SEP individuals have high risk of childhood adverse experiences.
- LDL is commonly among individuals with high cholesterol levels.
- KDM BA and V2 is correlated with LDL, total cholesterol, bmi, hba1c, and triglycerides.
- Raised body mass index and individuals with high waist size have homeostatic dysregulation.

Figure 9. 25 most relevant correlation between several dimensions of ACEs and diseases.

Ranked Cross-Correlations

25 most relevant

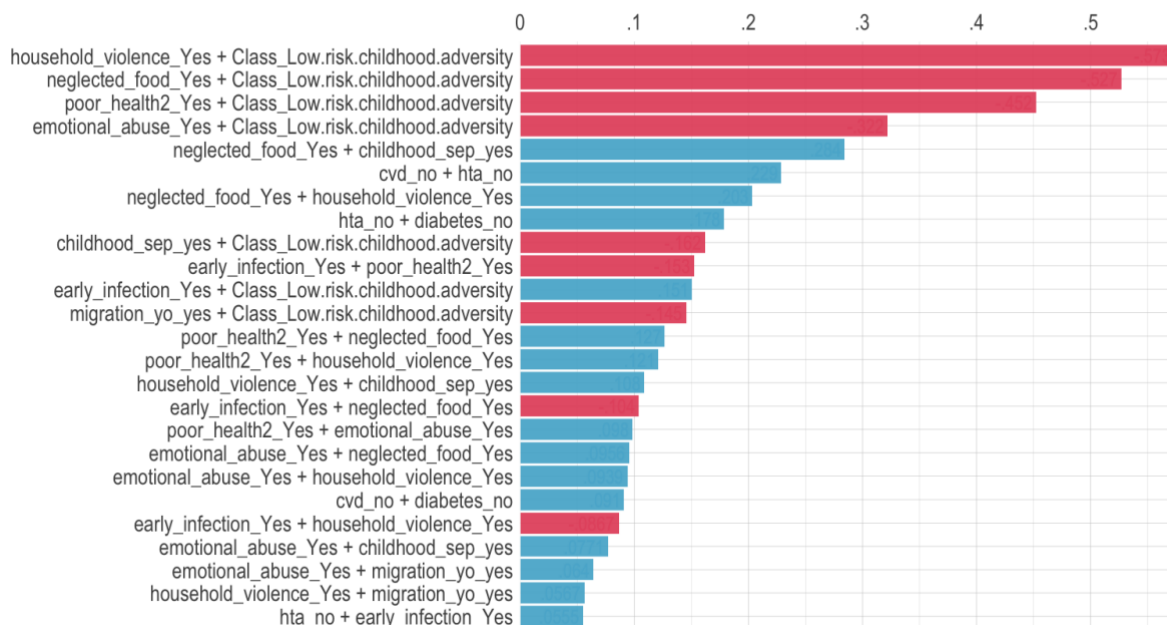


Table 4 reports Chi-squared test, significance level and explanation for each ACE measure and childhood SEP group indicating which main effects and interactions are significant.

Table 4. Hypothesis testing's Childhood, adult SEP, and adult diseases (ANOVA results)

Hypothesis	Chi-squared test	Significance	Result	Explanation
H1: ACE and CVD are independent of each other.	5.9617	0.01462	Rejected	The p-value is greater than 0.01. This shows moderate evidence against the null hypothesis in favour of the alternative: childhood adversity is significant with cardiovascular diseases.
H1: ACE and HTA are independent of each other.	0.21889	0.6399	Accepted	No significant. The p-value is greater than 0.1. This shows absent of evidence against the null hypothesis: data consistent with the null hypothesis (childhood adversity is not significant with hypertension).
H1: ACE and Diabetes are independent of each other.	2.0061	0.1567	Accepted	No significant. The p-value is greater than 0.1. This shows absent of evidence against the null hypothesis: data consistent with the null hypothesis (childhood adversity is not significant with diabetes).

Hypothesis	Chi-squared test	Significance	Result	Explanation
H2: Childhood SEP and CVD are independent of each other.	0.036043	0.8494	Accepted	No significant. The p-value is greater than 0.1. This shows absent of evidence against the null hypothesis: data consistent with the null hypothesis (childhood SEP is not significant with cardiovascular diseases)
H2: Childhood SEP and HTA are independent of each other.	0.30358	0.5816	Accepted	No significant. The p-value is greater than 0.1. This shows absent of evidence against the null hypothesis: data consistent with the null hypothesis (childhood SEP is not significant with Hypertension)
H2: Childhood SEP and Diabetes are independent of each other.	5.6304	0.01765	Rejected	The p-value is greater than 0.01. This shows moderate evidence against the null hypothesis in favour of the alternative: childhood SEP is significant with diabetes
H3: Adult SEP and diabetes		0.7223	Accepted	No significant
Adult SEP and CVD		0.05191		The p-value is greater than 0.05. This shows low evidence against the null hypothesis in favour of the alternative: adult SEP is significant with cardiovascular diseases.
Health insurance and HTA		0.0006571	Rejected	Significant. The p value is less than 0.001. This shows very strong evidence against the null hypothesis in favour of the alternative: health insurance is significant with hypertension.
Health insurance and CVD		0.001537	Rejected	Significant. The p value is greater than 0.001. This shows strong evidence against the null hypothesis in favour of the alternative: health insurance is significant with cardiovascular diseases.

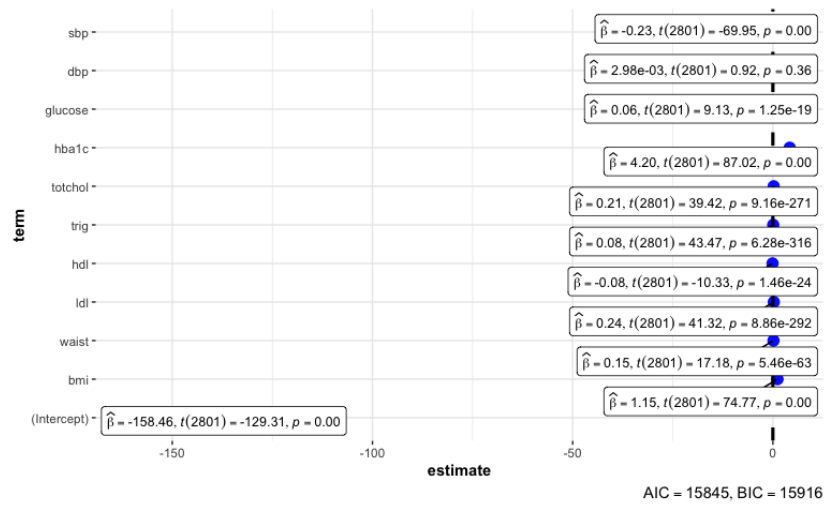
Regression model results are reported in Table 5 and intercorrelations of KDM advance with mediators and outcomes are shown in Figure 10. The regression model included sex and age, Residual Standard Error, Multiple R-Squared, and Adjusted R-squared to assess the goodness of fit of the model. Adjusted R-Squared measures the strength of the linear relationship between the predictor variables and the response variable. A multiple R-squared of 1 indicates a perfect linear relationship while a multiple R-squared of 0 indicates no linear relationship. Said differently, R-squared gives an idea of the proportion of the variance in the response variable that can be explained by the predictor variables. In SABE datasets, the Adjusted R-squared is 0.69 for model 1 (KDM + biomarkers) and there is a perfect linear relationship in model 2 (KDM+ CA). This indicates that 69% and 100% of the variance in KD can be explained by the predictors in the model (biological markers and chronological age). In this same example, childhood SEP and outcomes were significant but with lower predictor values (8% to both).

Table. 5 Regression model results

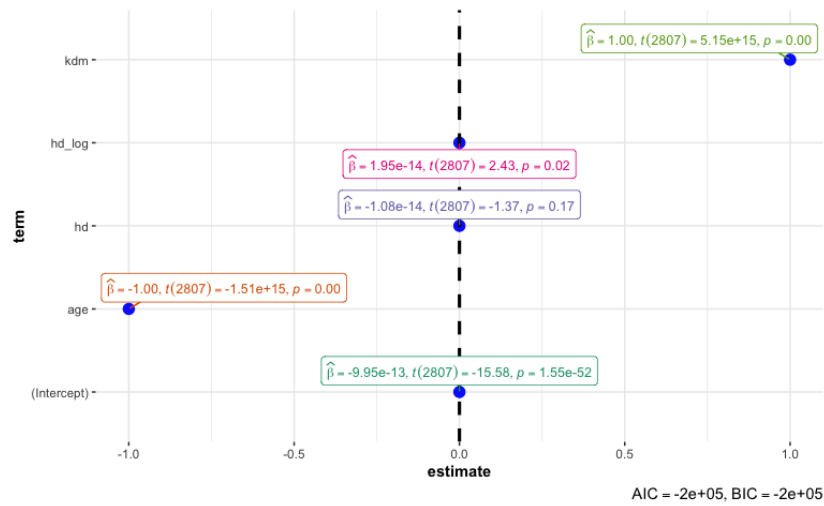
Model/coefficients	Residual standard errors	df	Adjusted R-squared	P value	AIC	BIC
M1 KDM BA + biomarkers	4.04	2801	0.6989	< .001***	15845	15916
M2 KDM BA + BA measures	0.001	2807	1	< .001***	-2e+05	-2e-05
M3 KDM BA + ACES	22.79	2805	0.08703	< .001***	25564	25629
M4 KDM BA + childhood SEP	22.84	2809	0.004296	0.0008706	25565	25601
M5 KDM BA + adult SEP	22.89	2804	-0.0003549	0.5396	25998	25652
M6 KDM BA + social policies	22.89	2805	-0.00001	0.4348	25589	25654
M7 KDM BA + risk behaviors	22.89	2807	0.0003014	0.3036	25342	25396
M8 KDM BA + diseases	21.86	2807	0.08827	< .001***	25334	25370

Data are presented P values of KMD biological age as continuous variable (per 1 log unit increment). All models adjusted for age and sex. Model 1. KDM advance and ten biomarkers. Model 2. KDM advance, ten biomarkers, and chronological age. Model 3 KDM advance plus six childhood adversities (emotional abuse, food insecurity, migration, domestic violence, early-life chronic infection, and poor health status). Model 5. KDM advance and education, occupational status, and current neighborhood deprivation. Model 6. KDM advance and health insurance and cash transfer benefits. Model 7. KDM advance and smoking, alcohol consumption, and raised body mass index. Model 8. KDM advance and outcomes (cardiovascular diseases, hypertension, diabetes, and obesity).

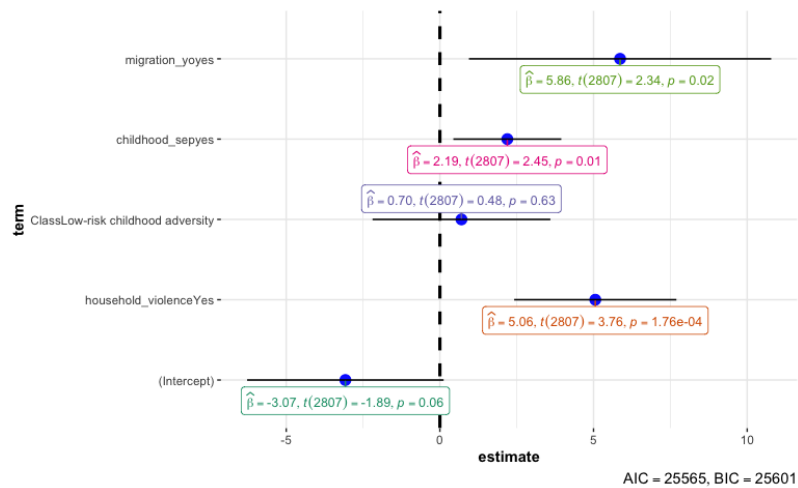
Figure 10. KDM advance with mediators and outcomes.
BA and biological markers

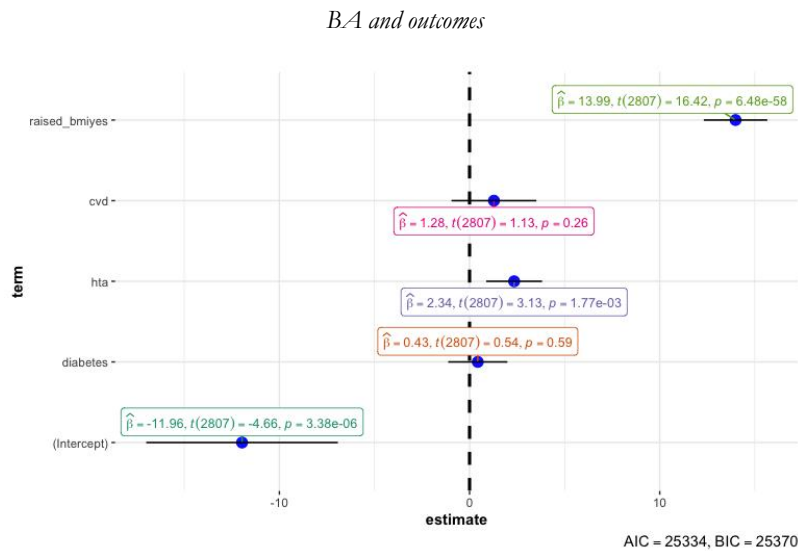
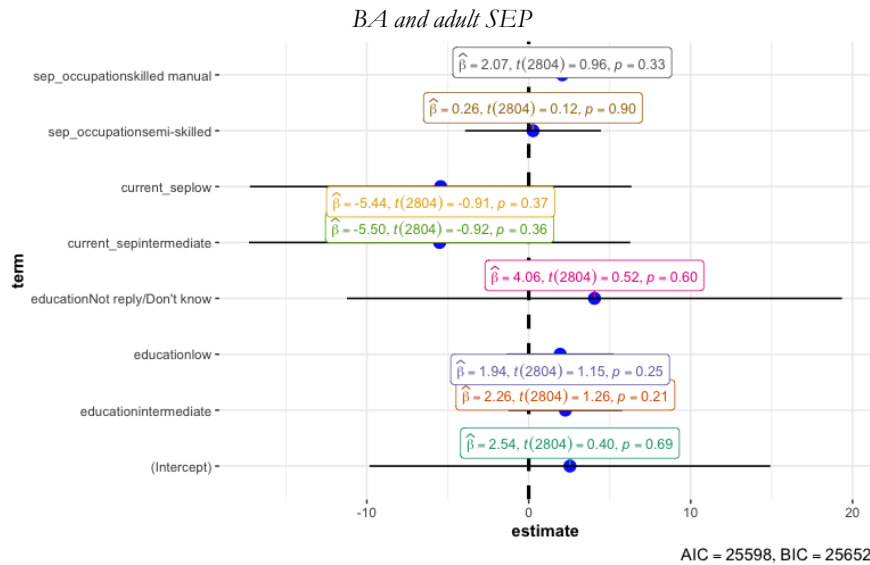


BA and comparison with one another



BA among childhood SEP groups





5) Conclusions

The Klemera-Doubal's method (KDM advance) turned out to be strongly correlated with biomarkers, BA measures, and chronological age, indicating that the estimates of the methods adequately correspond to the health status of the individuals. KDM advance was slightly correlated with diseases, childhood adversity, and childhood SEP. Although KDM showed a correlation with childhood adversity (household violence and migration), and childhood socioeconomic position than Homeostatic Dysregulation. Correlation analysis was consistent using alternative packages. Nine out of ten biomarkers were correlated with KDM advance, except diastolic blood pressure. In the regression analysis, the KDM advance shown a perfect linear relationship in model 2 (KDM+ CA), which means a higher predictor power.

References

- Buslova, Anastasiia. 2017. 'Prediction of Biological Age', 18. <http://lup.lub.lu.se/student-papers/record/8917091>.
- Chadeau-Hyam, Marc, Barbara Bodinier, Roel Vermeulen, Maryam Karimi, Verena Zuber, Raphaële Castagné, Joshua Elliott, et al. 2020. 'Education, Biological Ageing, All-Cause and Cause-Specific Mortality and Morbidity: UK Biobank Cohort Study'. *EClinicalMedicine*. <https://doi.org/10.1016/j.eclinm.2020.100658>.
- Cho, Il Haeng, Kyung S. Park, and Chang Joo Lim. 2010. 'An Empirical Comparative Study on Biological Age Estimation Algorithms with an Application of Work Ability Index (WAI)'. *Mechanisms of Ageing and Development* 131 (2): 69–78. <https://doi.org/10.1016/j.mad.2009.12.001>.
- Flores-Guerrero, Jose L., Marco A. Grzegorzczak, Margery A. Connelly, Erwin Garcia, Gerjan Navis, Robin P.F. Dullaart, and Stephan J.L. Bakker. 2021. 'Mahalanobis Distance, a Novel Statistical Proxy of Homeostasis Loss Is Longitudinally Associated with Risk of Type 2 Diabetes'. *EBioMedicine* 71: 103550.

- <https://doi.org/10.1016/j.ebiom.2021.103550>.
- Jeon, Haemi, and Jaehyun Park. 2017. 'Selection of an Optimal Set of Biomarkers and Comparative Analyses of Biological Age Estimation Models in Korean Females'. *Archives of Gerontology and Geriatrics* 70: 84–91. <https://doi.org/10.1016/j.archger.2017.01.005>.
- Karimi, Maryam; Castagne, R; Delplanque, C. 2019. 'Early-Life Inequalities and Biological Ageing: A Multisystem Biological Health Score Approach in Understanding Society'. *Epidemiology Community Health* 73:693.
- Klemera, Petr, and Stanislav Doubal. 2006. 'A New Approach to the Concept and Computation of Biological Age'. *Mechanisms of Ageing and Development* 127 (3): 240–48. <https://doi.org/10.1016/j.mad.2005.10.004>.
- Kwon, Dayoon, and Daniel W. Belsky. 2021. 'A Toolkit for Quantification of Biological Age from Blood Chemistry and Organ Function Test Data: BioAge'. *GeroScience* 43 (6): 2795–2808. <https://doi.org/10.1007/s11357-021-00480-5>.
- Levine, Morgan E. 2013. 'Modeling the Rate of Senescence: Can Estimated Biological Age Predict Mortality More Accurately than Chronological Age?' *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 68 (6): 667–74. <https://doi.org/10.1093/gerona/gls233>.
- Robinson, Oliver, Marc Chadeau Hyam, Ibrahim Karaman, Rui Climaco Pinto, Mika Ala-Korpela, Evangelos Handakas, Giovanni Fiorito, et al. 2020. 'Determinants of Accelerated Metabolomic and Epigenetic Aging in a UK Cohort'. *Aging Cell* 19 (6): 1–13. <https://doi.org/10.1111/acel.13149>.
- Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. *Journal of Open Source Software*, 6(61), 3167, doi:10.21105/joss.03167