

REVIEW ARTICLE OPEN



Biomarkers selection and mathematical modeling in biological age estimation

Solim Essomandan Clémence Bafei¹ and Chong Shen¹✉

Biological age (BA) is important for clinical monitoring and preventing aging-related disorders and disabilities. Clinical and/or cellular biomarkers are measured and integrated in years using mathematical models to display an individual's BA. To date, there is not yet a single or set of biomarker(s) and technique(s) that is validated as providing the BA that reflects the best real aging status of individuals. Herein, a comprehensive overview of aging biomarkers is provided and the potential of genetic variations as proxy indicators of the aging state is highlighted. A comprehensive overview of BA estimation methods is also provided as well as a discussion of their performances, advantages, limitations, and potential approaches to overcome these limitations.

npj Aging (2023)9:13; <https://doi.org/10.1038/s41514-023-00110-8>

INTRODUCTION

Biological aging is a process or set of processes that cause the viability of organs to deteriorate over time¹, while chronological aging refers to the passage of time. Biological age (BA), often used to measure the progress of biological aging, can indicate an aging individual's life expectancy and health status². Chronological age (CA) counts individuals' years lived, not their functional level, life experience, or vulnerability to aging-related diseases and disabilities. Consequently, people with the same CA might age differently³, more likely due to variations in their fundamental biological aging mechanisms⁴. BA is affected by various factors such as heredity, physical fitness, and external environmental stressors⁵.

BA predicts mortality and the occurrence of various age-associated disorders, such as cardiovascular diseases (CVD), diabetes, cancers, and Alzheimer's disease^{4,6,7}. BA is useful for clinical and community surveillance and evaluating interventions to delay or prevent aging-related disease and disability⁸. BA estimation methods measure aging biomarkers and integrate these values into years using the evolving mathematical models⁹. To date, there is not yet standardized or widely accepted aging biomarkers and mathematical model for quantifying BA¹⁰.

Methods proposed to quantify BA in units of years include multiple linear regression (MLR)¹¹, principal components analysis (PCA)¹², Hochschild method (HocM)¹³, Klemmer-Doubal method (KDM)^{2,9}, machine learning (ML) algorithms¹⁴, and epigenetic age calculators^{4,15–18}. Other methods, including homeostatic dysregulation (HD)¹⁹, allostatic load (AL)²⁰, and polygenic risk scores (PRS)²¹ algorithms have evolved to explore biological aging. In this review, we aim to overview BA estimation processes covering the biomarkers selection and the choice of mathematical models to calculate BA. The performance of the different mathematical models in predicting age-related adverse health outcomes is compared, noting their benefits, drawbacks, and potential solutions to overcome these limitations.

AGING BIOMARKERS

An aging biomarker is a physiological measure used alone or combined with other variables to detect, diagnose, or forecast the functional competence or function loss of any biological

component of a live organism in the absence of illness²². The American Federation of Aging Research (AFAR) recommends the following to select aging biomarkers²³: (1) a biomarker should predict the extent of aging and lifespan better than CA; (2) it should control the primary mechanism that underlies aging, not be the result of illnesses; (3) it can be tested multiple times without harming anyone; (4) it can be effective in humans and experimental animals. In other words, biomarkers for BA estimations must predict the best lifetime than CA. Nevertheless, biomarkers that strictly fulfill AFAR guidelines are not likely to exist²⁴.

Nakamura and Miyao²⁵ also proposed and tested four other statistical approaches for aging biomarkers selection. According to their standards, (i) an aging biomarker should correlate significantly with CA cross-sectionally; (ii) the longitudinal change of an aging biomarker with CA should concord with cross-sectional correlation; (iii) there should be a significant steadiness of individual variability in the measures; (iv) the part of change caused by aging should be proportional to the variation in lifespan among the concerned species.

Generally, researchers select two main types of aging biomarkers: those based on specific cellular level modification (cellular biomarkers) and patient-level combining data from multiple clinical tests (clinical biomarkers), according to their biological role in the aging process. Typical methods of BA estimation as suggested by Nakamura and Miyao²⁵ standard (i) correlate^{11,26,27} and/or regress^{28,29} candidates' biomarkers on CA to select the significant ones according to established cut-off-points of their effect sizes to determine aging biomarkers. However, in the Levine method, biomarkers are selected based on their relationship with death, while in HocM, biomarkers are selected based on their association with life and death factors^{4,13}.

Cellular biomarkers

Genomic biomarkers. BA can reflect the life expectancy of an individual³⁰. About a quarter of the variance in human longevity is due to genetic factors³¹. Telomeres are protective chromosomal ends of repeated DNA sequences bound by particular nucleoproteins and are lengthy at birth and diminish with each mitosis as age increases³². Telomere shortening is theorized to facilitate the physiological mechanism of aging³³, longevity³⁴, and death³⁵.

¹Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166 Jiangsu, China. ✉email: sc@njmu.edu.cn

Genome-wide association studies (GWAS) including Catalog lists 55 studies that revealed 676 genetic variants associated with aging, aging characteristics, lifespan, and longevity³⁶. The variants of apolipoprotein (APO)E, and Forkhead box O3 A(FOXO3A) have replicated almost constantly in various populations^{37,38}. APOE is a pleiotropic gene that has many functions including packaging and transporting low-density lipoprotein cholesterol³⁹. Its polymorphisms are determined by three alleles: $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ defined by combinations of genotypes of the single nucleotide polymorphisms (SNPs) rs7412 and rs429358. APOE- $\epsilon 2$ is associated with increased odds of longevity while $\epsilon 3$ and $\epsilon 4$ are associated with shorter life expectancy^{37,38}. Long-lived individuals often have a lower frequency of $\epsilon 4$, and a higher frequency of $\epsilon 2$ or a higher concentration of plasmatic APOE⁴⁰ whereas carriers of APOE- $\epsilon 4$ alleles often have abnormal lipid levels⁴¹ and are more at risk of Alzheimer's disease, diabetes, and CVD⁴². Those diseases might accelerate biological aging and increase the likelihood of death which could explain the detrimental effect of APOE- $\epsilon 4$ on longevity.

There are about 40 noncoding SNPs in FOXO3A that have been repeatedly linked to longevity in humans⁴³. FOXO3A is a component of the nutrient-sensing pathway connected to insulin and insulin growth factor (IGF)-1 regulates the cell response to oxidative stress and nutrition availability. It functions as a transcription factor on several homeostatic genes in response to diminished insulin or IGF-1 signaling⁴⁴. The variants may boost FOXO3A to battle oxidative stress by strengthening its links with upstream and downstream molecular partners. In animal models, changes that affect these signals can delay the aging process since this pathway controls many facets of cell homeostasis, including cell survival and proliferation⁴⁴. Another example of genetic variation associated with longevity is provided in the Supplementary information file.

Polygenic risk score as a predictor of longevity: One method of summarizing findings from GWAS for the prediction of complex variables, like longevity, is through polygenic risk scores (PRS). PRS are the weighted scores of independent variations that represent the genetic risk to develop a phenotypic characteristic. PRS for longevity can serve as a proxy indicator of biological aging as higher PRS for longevity might forecast a slower biological aging process and are associated with an increased likelihood of longevity⁴⁵ whereas high PRS of life-threatening diseases are often associated with decreased odds of longevity⁴⁶. High PRS for longevity is associated with an increased likelihood of healthy cognition aging and longer survival⁴⁷ and is negatively associated with metabolic syndrome⁴⁸ and brain aging process⁴⁹. All these strengthen the idea that genetic variation contributing to lifespan might regulate the biological aging process.

Epigenetic biomarkers. Epigenetics is rapidly being recognized as an important fragment of aging and longevity⁵⁰. Epigenetic modifications are theorized to facilitate the physiological mechanism of aging and are used as biomarkers to quantify BA^{33,51}. Biomarkers based on DNA methylation (DNAm) which is an epigenetic modification characterized by a chemical change of genomic nucleotide bases⁵² met the various elusive criteria for a cellular biomarker of aging: they apply to all sources of DNA and across the CA spectrum⁵³. Researchers apply ML algorithms to measure DNAm modification in multiple tissues, which enables the generation of highly accurate age estimators called epigenetic clocks. Numerous epigenetic clocks have emerged in recent years; the most popular ones are Hannum, Horvath, Levine, and Lu clocks, and genes that are associated with these four clocks' age acceleration (AA) have been identified. These genes include PIK3CB a gene related to human longevity, C1SD2 which has an important role in lifespan

regulation⁵⁴, TET2 involved in many aging/regenerative phenotypes⁵⁵, and IBA57 a protein-coding gene that is linked to several disorders of mitochondrial malfunction⁵⁶. The association of at least one epigenetic clock with aging and/or longevity-related genes suggests that epigenetic clocks might be valuable biomarkers of biological aging.

Transcriptional biomarkers. The expression of many genes exhibits age-related changes during growth and development⁵⁷ and studies that have constructed transcriptomic age which is a BA from transcriptomic sources using these genes have reported relatively good results (mean absolute error (MAE) = 4.7 and 7.8 years)^{57,58}. A previous study by Harries et al.⁵⁹ showed that the molecular pathways mostly involved in the processing and maturation of messenger RNA transcripts were strongly related to increasing CA, and identified a group of six transcripts to distinguish successfully young and old people. The transcriptomic age correlated strongly with CA but not to the same extent as epigenetic clocks age in two studies^{57,60}. Although the transcriptomic age was not a significant predictor of mortality in one study⁶⁰, it was significantly related to smoking, waist-to-hip ratio, and systolic blood pressure in another study⁵⁷. Furthermore, subjects with accelerated transcriptomic age had increased levels of total cholesterol and blood pressure⁵⁷ whereas subjects with younger transcriptomic age displayed younger biochemical indices such as IL-6, blood urea, and serum albumin than their counterparts with older BA⁶¹. All these highlight the relevance of transcriptomic biomarkers in reflecting the biological aging state.

Clinical biomarkers

Individual deteriorations in system integrity are theorized to be caused by molecular variations in aging, such as parameters that combine information from multiple clinical tests are the most used aging biomarkers to quantify BA³³. Those parameters can be biochemical, physical, mental, and functional parameters that continuously vary with age⁶². The potential of metabolites and proteins as biomarkers for BA estimation is provided in the Supplementary information file. Table 1 presents some parameters used to calculate BA across studies.

BA ESTIMATION METHODS

Multiple linear regression (MLR)

The MLR approach is widely used in small and large-scale studies^{26,63}. Using MLR, BA is assumed to be equal to the predicted CA of an individual and based on the relationship between the measured CA and several biomarkers¹¹. Durbina et al.⁶⁴ provided an upgrade of MLR because of the distorted estimation of BA at the regression edges (called systematic error) with the equation below:

$$\text{BAC} = \text{BA} + z \quad (1)$$

$$z = (\text{CA} - \text{Mean}_{\text{CA}}) \times (1 - b) \quad (2)$$

BA's simple linear regression coefficient on CA is b ; BAC is the resulting BA after correction or improvement. MLR is easy to perform and provides BA estimates that correlate significantly with CA among healthy individuals ($r > 0.80$)^{29,63} and predicts mortality among adults⁷. Nevertheless, MLR is criticized because each biomarker's weight effect depends on the strength of its link with CA. MLR's systematic error persists despite its correction approaches; there is also a risk of multicollinearity¹³, which can be overcome by excluding highly intercorrelated biomarkers in the same model. Also, studies comparing mathematical models reported MLR as not being the

Table 1. Parameters used to calculate BA across studies.

Clinical biomarkers	Groups	Parameters
	Musculoskeletal	Height ⁶ , waist circumference ⁶ , height and thigh circumference ²⁸ , gait-handgrip strength ²⁸ , knee extension strength ²⁸ , timed up and go ²⁸ , gait speed ²⁸ , chair rise ²⁸ , hands bones ⁶²
	Respiratory	Forced expiratory volume ^{25,28}
	Renal	Estimated glomerular filtration rate ^{6,28} , Albumin ^{26,63} , creatinine ^{26,63} , Urea nitrogen ^{6,26,28} , uric acid ⁶³
	Cardiovascular	Systolic blood pressure ^{6,25,63} , diastolic blood pressure ²⁸ , cholesterol ^{6,26,63} , triglycerides ^{6,26} , high-density lipoprotein cholesterol ⁶ , Low-density lipoprotein cholesterol ²⁶
	Hepatic	Aspartate aminotransferase ⁶ , gamma-glutamyl transpeptidase ⁶ , alkaline phosphatase ^{6,26,63} , bilirubin
	Red blood cells	Hemoglobin ^{6,28} , hematocrit ²⁶ , glycated hemoglobin ⁶³
	White blood cells and immune	Immunoglobulin M ²⁶ , β2-microglobulin ²⁶ , white blood cell count ⁶³ , C-reactive protein ⁶³
Cellular biomarkers	Genome	Telomere length ^{34,64} , (rs8176719, rs687621, rs643434, and rs505922) ⁶⁵ , (APOE-ε2, APOE-ε3, and APOE-ε4) ^{38,66} , (rs2069837 and rs2440012) ⁶⁷ , rs2802292, (rs2764264 and rs13217795)
	Epigenome	Hannum clocks ¹⁵ , Horvath clocks ¹⁶ , Levine clocks ⁴ , and Lu clocks ¹⁷
	Transcriptome	(LRRN3, CD27, GRAP, CCR6, VAMP5, CD248, D248, ABLIM1, LRRN3, CD27, GZMH, FAM102A, NELL2, SERPINE2, LEF1, CCR7 etc...) ^{57,59,61}

best for BA estimation because of the low power of its BA estimates to predict adverse health events compared to other models^{5,7,9,10,28}.

Principal component analysis (PCA) method

PCA selects the least possible physiological parameters to quantify BA¹² and is widely used to estimate BA^{6,10,12,65}. With PCA, BA estimates are built in five phases: correlation analysis, constancy analysis, redundancy analysis, PCA, and final equation construction⁶⁵. The last step, which consists of the equation construction, is built using the following equation⁶⁶:

$$BAS = \frac{a \times (X_1 - \text{Mean}_1)}{SD_1} + \frac{b \times (X_2 - \text{Mean}_2)}{SD_2} + \dots + \frac{m \times (X_m - \text{Mean}_m)}{SD_m} \quad (3)$$

BAS, Mean_m, SD_m, m, X_m are BA score, the mean, standard deviation, coefficient obtained by PCA, and the values of biomarker m, respectively. It is suggested to transform BAS to BA because the unit of BAS is not in a year using the following equation⁶⁶:

$$BA = (BAS \times SD_{CA}) + \text{Mean}_{CA} \quad (4)$$

To avoid systematic error, BA is corrected using the following equation:

$$BAc = BA + z \quad z = (2) \quad (5)$$

Increasing evidence shows that the BA estimates resulting from PCA can serve as an illness prediction tool to screen vulnerable populations and monitor the aging degree⁶⁷. As an example, BA estimates resulting from PCA predicted metabolic syndrome⁶⁸ and reduced survival likelihood⁶⁹ in different studies.

PCA is simple to understand and perform compared to ML, HocM, and KDM⁶⁷. PCA outclassed only MLR in predicting frail individuals²⁸ and has shown a good linear fit with CA with a mean regression slope of 0.96¹⁰ and 1⁵, respectively. However, MLR and PCA share the same correction method by using z-scores to correct the BAS, which has been found inappropriate because of the persistence of the distortion of the BA estimates at regression edges¹³. PCA limitations also include essential parameters exclusion⁷⁰, contributing to a loss of information⁷¹.

Hochschild's method (HocM)

HocM consists of assessing biomarkers based on their influence on lifespan and death. Hochschild⁷² chose physiological parameters according to their sensitivity to CA found in previous studies, "quality of life criterion", their ability to adapt to the devised instrument to quantify biomarkers, and their correlation with mortality. To consider the selected parameters as biomarkers of aging, they were correlated with the composite validation variable (CVV) based on eight death risk factors grouped in smoking status, diet, exercise level, lifespan, and education level⁷². Each risk factor was assigned a particular weight and the risk factors scores were standardized and combined into the mean validation variable (MVV) following the equation below:

$$MVV = \frac{\sum_{i=1}^n W_i SF_i}{\sum_{i=1}^n W_i} \quad (6)$$

where W_i and SF_i are the weight and the standardized scores of factor i and n is the total number of risk factors. CVV is obtained by standardizing MVV⁷². HocM was the first to use the reverse regression technic in the BA estimation process. In HocM, n linear regression equations of biomarkers on CA replace multiple linear regression of CA on biomarkers.

$$Y_i = d_i + e_i CA (+f_i K) \quad (7)$$

$i = 1 \sim n$, K is height and the term in brackets is added only when i designates a height-dependent biomarker such as forced expiratory volume and forced vital capacity. Stepwise mathematical procedures are performed to get the standardized BA (SBA), an intermediate BA product. The final phase of HocM is to convert the SBA to BA in years units. This is accomplished by employing a concept similar to Stanford-IQ Binet's index, which requires selecting an arbitrary mean and standard deviation⁷². There is a need for a cautious exploration of which subject's death risk factors will be included in CVV and a large sample size to apply HocM to a newly developed system⁹. HocM is a challenging multiple-stepwise approach that is not widely used for BA estimation.

Klemera and Doubal method (KDM)

KDM is a graphing method developed by Klemera and Doubal² to compute BA. KDM uses the reverse regression technique proposed in HocM for the BA estimation process in which CA is

Table 2. ML algorithms used to compute BA in some studies.

Studies	ML algorithms
Putin et al. ⁸⁰	Gradient boosting machine (GBM), random forest (RF), decision trees (DT), k-nearest neighbors (KNN), elastic net regression (ENR), Support vector machine (SVM)
Rahman et al. ⁸¹	Deep neural network (DNN), convolutional neural network (CNN), convolutional long short-term memory
Egorova et al. ⁸²	RF, SVM
Sagers et al. ⁸³	RF
Galibourg et al. ⁸⁴	AdaBoost, Bayesian ridge regression (BRR), DT, KNN, multi-layer perceptron (MLP), polynomial regression (POLYREG), RF, SVM, voting regressor
Wang et al. ⁷⁸	ENR, SVM, partial least squares, extreme gradient boosting (Xgboost)
Bae et al. ⁸⁵	POLYREG, SVM, Xgboost, DNN, RF
Yang et al. ⁸⁶	Xgboost, Catboost, Adaboost, Extra Trees, DNN, CNN, Light gradient boosting machine
Pyrkov et al. ¹⁴	CNN
Levine et al. ⁴ , Hannum et al. ¹⁵ , Horvath ¹⁶ , Belsky et al. ¹⁸	ENR

used as a standard biomarker, not an outcome to be predicted, enabling a substantial improvement in BA estimates' precision. The following equation was formulated using many mathematical steps²:

$$BA = \frac{\sum_{i=1}^n (x_i - g_i) \frac{h_i}{s_i^2} + \frac{E}{s_b^2}}{\sum_{i=1}^n \left(\frac{h_i}{s_i} \right)^2 + \frac{1}{s_b^2}} \quad (8)$$

E stand for CA, x_i is the value of biomarker i , g_i , h_i , s_i is the regression slope, the intercept, and the root mean squared error (RMSE) from a regression of i on CA, and s_b^2 is equal to the variance in CA explained by the n biomarkers. BA equations are built using computer-generated simulations; this advanced concept makes KDM an optimal method for BA estimation². Studies that have compared BA estimates resulting from KDM and other models in mortality prediction found that KDM was an excellent method to compute BA than the previous algorithms^{5,7,10,28} but recent algorithms such as HD and ML outperformed KDM in predicting various health outcomes including mortality^{20,73}.

Machine learning (ML) approaches for BA estimation

The ML approach, a generic artificial intelligence technique in which a computer learns data using a specially developed algorithm or without one is rapidly emerging and encourages biological aging and longevity research⁷⁴. ML can be supervised, unsupervised, and semi-supervised. Table 2 represents the ML algorithms used to compute BA in some studies. ML algorithms are broad and have provided relatively reliable BA estimates with prediction accuracy of 0.87–4.76 years and R^2 of 0.25–0.94^{73,75}. ML algorithms are useful for big data and estimating BA from non-conventional parameters, including radiograph⁷⁶ and magnetic resonance imaging results⁷⁷. Nevertheless, they are subjected to some limitations: there is not yet a specific algorithm in ML specially dedicated to BA estimation; also, many ML algorithms have a “black box” effect that prevents a systematic understanding of the relationships between features and labels, as well as how they are estimated⁷⁸. In this scenario, traditional statistical tools and comprehension of medical or biological topics are helpful. Also, there is a risk of overfitting with many biomarkers in the same model⁷⁹. Recently Yang et al.⁸⁰ suggested a composite ML-BA based on the stacking technique with a simple meta-model to avoid overfitting and improved the performance of the BA estimates in predicting different health outcomes⁸⁰. A brief description of the stacking method is provided in the Supplementary information file.

Epigenetic age calculators

First-generation epigenetic clocks. Hannum, Horvath, Levine, and Lu clocks are the most popular clocks that estimate epigenetic age, which is BA from DNA sources. Hannum clock is the first-reported trained and tested clock based on blood-derived DNA. Hannum's epigenetic age model was built using ENR analysis. It encompasses 71 CpGs chosen from the Illumina 450k array, which powerfully captures CA variations, partly driven by age-related shifts in blood cell composition¹⁵.

The Horvath clock was built using several tissues, including the blood data from Hannum as a likely “pan-tissue” master clock of CA, and focused on catching common changes independent of tissue type. The Horvath epigenetic age model was built using ENR analysis. It encompassed 353 CpGs on the previous Illumina 27k array¹⁶. Horvath's epigenetic age can be estimated using an online age calculator (<http://dnamage.genetics.ucla.edu/home>). A study reported Horvath's epigenetic age performing Hannum in correlation with CA⁸¹ whereas another study reported Hannum's epigenetic age as having the highest correlation with CA⁸².

Second-generation epigenetic clocks

DNAm PhenoAge: Levine et al.⁴ built a BA model using biomarkers from the whole blood sample of individuals in NHANES III. The hazard of death due to aging was regressed on 44 clinical parameters and CA in a Cox penalized regression model for choosing ten variables to include in their epigenetic age model called phenotypic age. The ten parameters were encompassed in a parametric proportional hazards model based on the Gompertz distribution. Based on that model, individuals' 10 years of death hazard were estimated. Subsequently, the death score was transformed into years⁴. Using ENR analysis, the phenotypic age estimates were regressed on DNAm data. The penalization parameter has been selected to lessen the cross-validated mean-square error rate, resulting in 513 CpGs. The PhenoAge is strongly correlated with CA ($r=0.94$) and outclasses earlier epigenetic clocks in predicting all-cause death, malignancies, healthspan, Alzheimer's disease⁴, and breast cancer⁸³.

DNAm GrimAge: DNAm GrimAge is a composite biomarker based on seven DNAm substitutes and an estimator of the self-reported number of cigarettes smoked yearly based on a DNAm developed by Lu et al.¹⁷. The GrimAge was constructed in two steps. The first was to define surrogate DNAm biomarkers for physiological stressors and risk variables and the second step consisted of integrating these biomarkers to create DNAm GrimAge, a single composite biomarker of longevity that is expressed in years using elastic net Cox regression¹⁷. The authors

further developed GrimAge 2 following the same steps as in the previous GrimAge with the difference that they added two plasmatic biomarkers: CRP and hemoglobin A1C which were log-transformed⁸⁴. The GrimAge 2 estimation approach is implemented in an online software (<https://dnamage.clockfoundation.org/>). The GrimAges outperformed the previously reviewed clocks and clinical biomarkers related to BA in predicting aging-related ailments and all-cause death^{60,85,86}.

Third-generation epigenetic clocks

Dunedin pace of aging (PA) methylation: Belsky et al.¹⁸ developed DNAm-related PA using the data from Dunedin cohort participants who had 18 biomarkers measured in three waves (age 26, 32, and 38) and who had DNAm data available at age 38. They applied ENR using the PA (please see the Supplementary Information file for PA estimation method description) as the criterion variable and all methylation probes found on both the Illumina 450k and EPIC arrays as predictors. A total of 46 CpG sites was associated with the longitudinal PA and the resulting Dunedin PA methylation was more strongly associated with aging-related ailments than PA¹⁸.

Dunedin Pace: Dunedin Pace measure differed from Dunedin PA methylation in four points: (1) PA analysis included data from 20 years of follow-up; (2) PA analysis included four-time points of measurement; (3) Dunedin Pace modeling excluded CpG sites for which probes having low reliability in blood datasets; and (4) the algorithm used to implement Dunedin Pace includes a normalization step that allows Dunedin Pace values for individual samples to be compared to the Dunedin⁸⁷. The final Dunedin Pace algorithm included 173 CPG sites. It has a high reproducibility with the intraclass correlation (ICC) greater than 0.87 across databases which enables it to be used as a reliable measurement tool for aging intervention trials⁸⁷.

Advantages and limitations of epigenetic clocks and age estimators: The epigenetic age estimates have yielded impressively robust correlations with CA with minimal variance and great consistency across multiple age groups and studies' populations⁸⁸. However, epigenetic age calculators have some limitations: instead of limiting the set to biologically important pathways, epigenetic age calculators take hundreds of CpGs as input. As a result, they are not as informative to identify intervention targets or molecular pathways. In addition, epigenetic clocks often have low reliability with the ICC in the poor to moderate range^{89,90}, resulting in up to 3 to 9 years difference between technical replicates and providing false positive or negative associations⁹⁰. To overcome this limitation, Higgins-Chen et al.⁹⁰ proposed and trained epigenetic clocks using principal components instead of individual CpGs which contributed significantly to clock reliability improvement with ICCs 0.990–0.998⁹⁰.

OTHER MEASURES OF BIOLOGICAL AGING

Homeostatic dysregulation (HD) algorithm

HD offers an alternative to exploring biological aging and has been used for that purpose across studies^{8,27,91,92}. Subjects' HD shows the variation of their physiology from a reference in good health based on a biomarker Mahalanobis distance (MD)⁹². MD is the multidimensional distance of an individual's biomarkers values from a single reference point⁹³. A higher HD score reflects a higher BA, risk of illness, incapacity, and death, while a smaller HD score reflects the opposite. HD does not assume any unidirectionality of aging biomarkers relationship with CA and does not incorporate CA into its calculation⁸.

Before calculating HD, a reference population should be defined; in some cases, the reference is the mean of all values, however, in other cases is theory-driven, such as the mean of values in a young, healthy population representing the maximum

of reproductive fitness. Then, biomarker values from the reference population should be standardized and used to calculate a biomarker variance-covariance matrix. The biomarker raw mean and SD, and the standardized-biomarker variance-covariance matrix should be applied within the Mahalanobis distance equation to obtain the HD algorithm⁹³:

$$HD_1 = \sqrt{(\bar{w} - \bar{y})^T \times S^{-1} \times (\bar{w} - \bar{y})} \quad (9)$$

\bar{w} is a vector of biomarkers values of a subject in the study; \bar{y} is a vector of biomarker means in the reference sample, and S is the reference sample variance-covariance matrix. If none of the parameters are correlated, the HD score is obtained by scaling each biomarker by its variance and adding the deviance squared for each biomarker⁹³.

$$HD_2 = \sqrt{\sum_{i=1}^n \frac{(w_i - y_i)^2}{SD^2(x_i)}} \quad (10)$$

$SD^2(x_i)$ is the variance of biomarker i , and n is the number of biomarkers. HD estimates have predicted morbidity, mortality, and functional deficit⁸. A study reported that HD estimates outperformed KDM in predicting mortality⁹⁴ whereas another study showed that KDM outperformed HD³³. The HD algorithm has limitations: the statistical distance depends on detecting a normal biological state, often calculated as the population average for all parameters, a reasonable imperfect estimate. Finding the ideal centroid is not easy; therefore, the dysregulation score could be systematically biased⁹¹.

Allostatic load algorithms

The AL is wear and tear on the body and brain caused by persistent overactivity or inactivity of biological systems that are often involved in response to environmental stimulus⁹⁵. While AL has been often used to explore how physiological stress influences different outcomes, Hasting et al.²⁰ demonstrated that AL and primary measures of biological aging such as KDM and HD share similarities in their fundamental concepts: Similar to biochemistry parameters of systemic BA, AL assessment incorporates systemic biomarkers to form a measure of cumulative dysregulation; furthermore, AL theory is theoretically relevant for BA processes because psychosocial stress can accelerate the progression of several aging signs. The authors further estimated AL scores using three methods as well as calculated two measures of biological aging using KDM and HD. The three different AL estimates had a moderate correlation with KDM-BA ($r = 0.34$ – 0.39) and a moderate to stronger correlation with HD estimates ($r = 0.46$ – 0.68), suggesting that the three estimates might share some similarities in reflecting the aging process. Subjects with higher AL and HD scores and KDM-BA performed worse on physical agility tests and reported being in worse physical and mental health. Through these, AL might be considered a proxy measure of biological aging. There are over 30 methods of AL estimation which were reviewed in a recent literature review⁹⁶.

Polygenic risk score algorithms

A polygenic risk score is a weighted score of independent variants that quantifies the genetic likelihood of developing a certain trait. PRS is typically calculated as follows:

$$PRS_i = \sum_j^M ES_j * D_{ji} \quad (11)$$

M is the number of variants in the score, the effect size (ES) of the variants j , and dosage (D_{ji}) is the number of copies of SNP j in the genotype of individual i . The ES is often gotten from a discovery GWAS²¹. Other variations of the classical method of PRS^{46,97} have been proposed and software was developed to facilitate PRS

Table 3. Summary of the BA estimation algorithms.

Methods	Merits	Limitations
MLR	MLR is easy to understand and perform BA estimates from MLR are relatively reliable and have good prediction power	-Each biomarker's weight effect depends on the power of its relationship with CA -Distortion of BA estimates at the regression edges (systematic error) -Risk of multicollinearity
PCA	PCA is easy to understand and perform It solves the major limitation of MLR PCA-BA estimates are effective in illness prediction to screen vulnerable populations and monitor the aging degree	-The exclusion of important parameters -Loss of information due to parameters dimensionality reduction in PCA -Inadequacy of the proposed correction
HocM	HocM avoids distortion of BA estimates at the regression edges (systematic error) HocM considers CA as an independent variable	There is limited information on HocM because that approach is not widely used HocM is difficult to be implemented in systems other than its author's own
KDM	It considers CA as a standard biomarker It solves the major limitations of MLR BA estimates resulting from KDM are highly predictive of illnesses, impairments, and mortality	
ML	ML is useful for big data and non-usual parameters	There is limited evidence that ML methods are better than conventional methods Risk of overfitting There is not yet an explicit algorithm in ML mainly dedicated to BA estimation
Epigenetic age calculators	Epigenetic age calculators provide BA estimates that correlated highly with CA Epigenetic age calculators use DNAm clocks as biomarkers	-Epigenetic age calculators take hundreds of CpGs as input instead of limiting the set to biologically important pathways limiting their ability to identify intervention targets or molecular pathways -DNAm clocks have low reproducibility -Risk of overfitting
HD	They are highly predictive of illnesses and mortality HD does not assume the unidirectionality of aging biomarkers' relationship HD does not incorporate CA into its calculation HD estimates are highly predictive of morbidity, mortality, and functional deficit	Finding the ideal centroid is not easy, making the dysregulation score biased systematically

calculation⁹⁸. The step-by-step construction of polygenic scores has been reviewed in a recent study⁹⁹.

SUMMARY OF BIOLOGICAL AGING ESTIMATION ALGORITHMS

Although each method has its advantages and limitations, all the BA estimates from the reviewed BA estimation methods have relatively good prediction power of various health outcomes than CA in many studies. Table 3 represents the summary of the BA estimation algorithms.

DISCUSSION

Biomarkers selection is an integral part of the BA estimation process, and the standards for biomarkers selection vary depending on BA estimation methods and researchers. Furthermore, many BA estimation approaches presume a linear relationship between biomarkers and CA. A linear model presupposes a constant increase or decrease of the biomarkers' values over time which might not accurately reflect the complex aging process taking place in the body. In reality, aging can occur at different rates for different individuals and in different parts of the body¹⁰⁰ making the assumption of linearity difficult to establish.

Additionally, many aging biomarkers such as telomere length or clinical parameters are influenced by a multitude of factors that are not necessarily linked directly to age, such as lifestyles, stress, and genetic predisposition. Thus, relying on a linear relationship between biomarkers and age may not accurately capture the true relationship between them. Furthermore, there is an increasing body of evidence showing a linear and non-linear variation of proteins, clinical biomarkers composites, and gene expression throughout the life course in healthy and general population^{75,101,102}. Lehallier et al.¹⁰² reported a non-linear variation of more than two thousand plasmatic proteins across the lifespan in both healthy humans and mice and in human replication cohorts. The authors further pinpointed hundreds of proteins varying in waves throughout the lifespan and summing the number of differentially expressed proteins at each age revealed three peaks at ages 34, 60, and 78. Additionally, there was a statistically significant overlap between proteins altering in the three age groups but the majority of proteins altering in old age were not identified by the linear model¹⁰². These results highlight the inadequacy of linear modeling to capture the complexity of biological aging across the lifetime. Thus, relying only upon a linear association assumption of parameters and CA to select aging biomarkers might exclude potentially valid aging

biomarkers that have a non-linear association with CA and result in inaccurate BA estimates. Therefore, both linear and non-linear associations of parameters and CA are worth exploring in the biomarker's selection process.

Progress in technology and easy data accessibility through open-access databases have contributed to an increased number of studies exploring biological aging. To date, the National Institute on Aging has listed over 20 publicly available databases for aging-related research¹⁰³, and their data have been extensively used in biological aging investigations. All these have not only enabled the identification of complex age-related patterns of potential aging biomarkers^{75,102} but have also raised some concerns about the data quality. There is a lack of standardized measurement methods making it difficult to compare and contrast results from different studies. For example, methods commonly used in large epidemiological studies for telomere length measurement are frequently associated with substantial measurement error¹⁰⁴. As a result, even when results are consistent and of high quality, they may be difficult to compare between laboratories. A recent study dropped telomere length from its list of biomarkers to be used for biological aging exploration because of the yet-unresolved field-wide controversies about its measurement¹⁰⁵; these highlight the need to develop standardized biomarkers measurement tools.

In this study, we have summarized current knowledge on aging biomarkers and BA estimation methods as well as discussed their performances in predicting numerous age-related diseases, their advantages, shortcomings, and potential solutions to overcome those shortcomings. The validity of the BA estimates highly depends on the choice of biomarkers and the method applied for its calculation. Previous studies have assessed the utility of different biomarkers and BA estimation methods have identified the best mathematical model based on the association between its BA estimates and CA and its capacity to predict various health outcomes compared to others. However, the constancy of the predictive ability of these BA estimates is inconclusive across multiple studies because the model that correlated the best with CA and/or predicts various outcomes in one study might fail in another study. Therefore, there is a need to clarify which parameter(s) and mathematical model might provide a trustworthy indicator of biological aging.

Received: 20 November 2022; Accepted: 8 May 2023;

Published online: 01 July 2023

REFERENCES

- Nakamura, E. & Miyao, K. Further evaluation of the basic nature of the human biological aging process based on a factor analysis of age-related physiological variables. *J. Gerontol. A Biol. Sci. Med. Sci.* **58**, 196–204 (2003).
- Klemera, P. & Doubal, S. A new approach to the concept and computation of biological age. *Mech. Ageing Dev.* **127**, 240–248 (2006).
- Mitnitski, A. B., Graham, J. E., Mogilner, A. J. & Rockwood, K. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatr.* **2**, 1 (2002).
- Levine, M. E. et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
- Jee, H. & Park, J. Selection of an optimal set of biomarkers and comparative analyses of biological age estimation models in Korean females. *Arch. Gerontol. Geriatr.* **70**, 84–91 (2017).
- Kang, Y. G. et al. Biological age as a health index for mortality and major age-related disease incidence in Koreans: National Health Insurance Service—Health screening 11-year follow-up study. *Clin. Interv. Aging* **13**, 429–436 (2018).
- Levine, M. E. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J. Gerontol. A Biol. Sci. Med. Sci.* **68**, 667–674 (2013).
- Parker, D. C. et al. Association of blood chemistry quantifications of biological aging with disability and mortality in older adults. *J. Gerontol. Ser. A* **75**, 1671–1679 (2020).
- Cho, I. H., Park, K. S. & Lim, C. J. An empirical comparative study on biological age estimation algorithms with an application of Work Ability Index (WAI). *Mech. Ageing Dev.* **131**, 69–78 (2010).
- Jee, H. Selection of a set of biomarkers and comparisons of biological age estimation models for Korean men. *J. Exerc. Rehabil.* **15**, 31–36 (2019).
- Dubina, T. L., Dyundikova, V. A. & Zhuk, E. V. Biological age and its estimation. II. Assessment of biological age of albino rats by multiple regression analysis. *Exp. Gerontol.* **18**, 5–18 (1983).
- Nakamura, E., Miyao, K. & Ozeki, T. Assessment of biological age by principal component analysis. *Mech. Ageing Dev.* **46**, 1–18 (1988).
- Hochschild, R. Improving the precision of biological age determinations. Part 1: a new approach to calculating biological age. *Exp. Gerontol.* **24**, 289–300 (1989).
- Pyrkov, T. V. et al. Extracting biological age from biomedical data via deep learning: too much of a good thing? *Sci. Rep.* **8**, 5210 (2018).
- Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
- Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
- Lu, A. T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
- Belsky, D. W. et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife* **9**, e54870 (2020).
- Cohen, A. A. et al. Statistical distance as a measure of physiological dysregulation is largely robust to variation in its biomarker composition. *PLoS ONE* **10**, e0122541 (2015).
- Hastings, W. J., Almeida, D. M. & Shalev, I. Conceptual and analytical overlap between allostatic load and systemic biological aging measures: analyses from the National Survey of Midlife Development in the United States. *J. Gerontol. A Biol. Sci. Med. Sci.* **77**, 1179–1188 (2022).
- Collister, J. A., Liu, X. & Clifton, L. Calculating polygenic risk scores (PRS) in UK Biobank: a practical guide for epidemiologists. *Front. Genet.* **13**, 818574 (2022).
- Baker, G. T. 3rd & Sprott, R. L. Biomarkers of aging. *Exp. Gerontol.* **23**, 223–239 (1988).
- Info Aging Guides. *Biomarkers of Aging an Introduction to Aging Science Brought to you by the American Federation for Aging Research.* (2016).
- Johnson, T. E. Recent results: biomarkers of aging. *Exp. Gerontol.* **41**, 1243–1246 (2006).
- Nakamura, E. & Miyao, K. A method for identifying biomarkers of aging and constructing an index of biological age in humans. *J. Gerontol. A Biol. Sci. Med. Sci.* **62**, 1096–1105 (2007).
- Li, X. et al. Application of biological age assessment of Chinese population in potential anti-ageing technology. *Immun. Ageing* **15**, 33 (2018).
- Liu, Z. Development and validation of 2 composite aging measures using routine clinical biomarkers in the Chinese population: analyses from 2 prospective cohort studies. *J. Gerontol. A Biol. Sci. Med. Sci.* **76**, 1627–1632 (2021).
- Zhong, X. et al. Estimating biological age in the Singapore Longitudinal Aging Study. *J. Gerontol. A Biol. Sci. Med. Sci.* **75**, 1913–1920 (2020).
- Cao, W. et al. Modelling biological age based on plasma peptides in Han Chinese adults. *Aging* **12**, 10676–10686 (2020).
- Lin, W. Y. Lifestyle factors and genetic variants on 2 biological age measures: evidence from 94,443 Taiwan Biobank Participants. *J. Gerontol. Ser. A* [glab251](https://doi.org/10.1093/gerona/glab251) <https://doi.org/10.1093/gerona/glab251> (2021).
- Deelen, J. et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432 (2014).
- Hamczyk, M. R., Nevado, R. M., Baretino, A., Fuster, V. & Andrés, V. Biological versus chronological aging: JACC focus seminar. *J. Am. Coll. Cardiol.* **75**, 919–930 (2020).
- Hastings, W. J., Shalev, I. & Belsky, D. W. Comparability of biological aging measures in the National Health and Nutrition Examination Study, 1999–2002. *Psychoneuroendocrinology* **106**, 171–178 (2019).
- Rehkopf, D. H. et al. Longer leukocyte telomere length in Costa Rica's Nicoya Peninsula: a population-based study. *Exp. Gerontol.* **48**, 1266–1273 (2013).
- Doroschuk, N. A. et al. An original biomarker for the risk of developing cardiovascular diseases and their complications: Telomere length. *Toxicol. Rep.* **8**, 499–504 (2021).
- The NHGRI-EBI Catalog of human genome-wide association studies. GWAS Catalog, trait: longevity. https://www.ebi.ac.uk/gwas/efotraits/EFO_0004300 (2023).
- Revelas, M. et al. Review and meta-analysis of genetic polymorphisms associated with exceptional human longevity. *Mech. Ageing Dev.* **175**, 24–34 (2018).
- Deelen, J. et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **10**, 3669 (2019).
- Raghavachari, N. The impact of Apolipoprotein E genetic variability in health and life span. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **75**, 1855–1857 (2020).

40. Muenchhoff, J. et al. Plasma apolipoproteins and physical and cognitive health in very old individuals. *Neurobiol. Aging* **55**, 49–60 (2017).
41. Martínez-Magaña, J. J. et al. Association between APOE polymorphisms and lipid profile in Mexican Amerindian population. *Mol. Genet. Genomic Med.* **7**, e958 (2019).
42. Polsinelli, A. J. et al. APOE ε4 is associated with earlier symptom onset in LOAD but later symptom onset in EOAD. *Alzheimers Dement.* <https://doi.org/10.1002/alz.12955> (2023).
43. Link, W. Introduction to FOXO biology. *Methods Mol. Biol.* **1890**, 1–9 (2019).
44. Caruso, C. et al. How important are genes to achieve longevity? *Int. J. Mol. Sci.* **23**, 5635 (2022).
45. Revelas, M. et al. Exceptional longevity and polygenic risk for cardiovascular health. *Genes* **10**, 227 (2019).
46. Hu, D. et al. Genetic trade-offs between complex diseases and longevity. *Aging Cell* **21**, e13654 (2022).
47. Tesi, N. et al. Polygenic risk score of longevity predicts longer survival across an age continuum. *J. Gerontol. A. Biol. Sci. Med. Sci.* **76**, 750–759 (2021).
48. Revelas, M. et al. High polygenic risk score for exceptional longevity is associated with a healthy metabolic profile. *GeroScience* **45**, 399–413 (2023).
49. Dong, C. et al. Parental life span and polygenic risk score of longevity are associated with white matter hyperintensities. *J. Gerontol. A. Biol. Sci. Med. Sci.* **77**, 689–696 (2022).
50. Brooks-Wilson, A. R. Genetics of healthy aging and longevity. *Hum. Genet.* **132**, 1323–1338 (2013).
51. Hillary, R. F. et al. Epigenetic measures of ageing predict the prevalence and incidence of leading causes of death and disease burden. *Clin. Epigenetics* **12**, 115 (2020).
52. He, X., Liu, J., Liu, B. & Shi, J. The use of DNA methylation clock in aging research. *Exp. Biol. Med.* **246**, 436–446 (2021).
53. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
54. Gibson, J. et al. A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLoS Genet.* **15**, e1008104 (2019).
55. McCartney, D. L. et al. Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome Biol.* **22**, 194 (2021).
56. Lin, W.-Y. Genome-wide association study for four measures of epigenetic age acceleration and two epigenetic surrogate markers using DNA methylation data from Taiwan Biobank. *Hum. Mol. Genet.* **31**, 1860–1870 (2022).
57. Peters, M. J. et al. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).
58. Holzschek, N. et al. Modeling transcriptomic age using knowledge-primed artificial neural networks. *NPJ Aging Mech. Dis.* **7**, 15 (2021).
59. Harries, L. W. et al. Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell* **10**, 868–878 (2011).
60. Li, X. et al. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *Elife* **9**, e51507 (2020).
61. Holly, A. C. et al. Towards a gene expression biomarker set for human biological age. *Aging Cell* **12**, 324–326 (2013).
62. Kalichman, L., Livshits, G. & Kobylansky, E. Association between morbidity and skeletal biomarkers of biological aging. *Hum. Biol.* **78**, 77–88 (2006).
63. Martínez de Toda, I., Vida, C., Díaz-Del Cerro, E. & De la Fuente, M. The immunity clock. *J. Gerontol. A. Biol. Sci. Med. Sci.* <https://doi.org/10.1093/gerona/glab136> (2021).
64. Dubina, T. L., AY, M. & Zhuk, E. V. Biological age and its estimation. III. Introduction of a correction to the multiple regression model of biological age in cross-sectional and longitudinal studies. *Exp. Gerontol.* **19**, 133–143 (1984).
65. Nakamura, E. & Miyao, K. Sex differences in human biological aging. *J. Gerontol. A. Biol. Sci. Med. Sci.* **63**, 936–944 (2008).
66. Jia, L., Zhang, W., Jia, R., Zhang, H. & Chen, X. Construction formula of biological age using the principal component analysis. *Biomed. Res. Int.* **2016**, 4697017 (2016).
67. Zhang, W. et al. Model construction for biological age based on a cross-sectional study of a healthy Chinese Han population. *J. Nutr. Health Aging* **21**, 1233–1239 (2017).
68. Bae, C.-Y. et al. Biological age and lifestyle in the diagnosis of metabolic syndrome: the NHIS health screening data, 2014–2015. *Sci. Rep.* **11**, 444 (2021).
69. Yoo, J., Kim, Y., Cho, E. R. & Jee, S. H. Biological age as a useful index to predict seventeen-year survival and mortality in Koreans. *BMC Geriatr.* **17**, 7 (2017).
70. Park, J., Cho, B., Kwon, H. & Lee, C. Developing a biological age assessment equation using principal component analysis and clinical biomarkers of aging in Korean men. *Arch. Gerontol. Geriatr.* **49**, 7–12 (2009).
71. Zhang, W.-G. et al. Construction of an integral formula of biological age for a healthy Chinese population using principle component analysis. *J. Nutr. Health Aging* **18**, 137–142 (2014).
72. Hochschild, R. Improving the precision of biological age determinations. Part 2: automatic human tests, age norms and variability. *Exp. Gerontol.* **24**, 301–316 (1989).
73. Wang, C. et al. A machine learning-based biological aging prediction and its associations with healthy lifestyles: the Dongfeng-Tongji cohort. *Ann. N. Y. Acad. Sci.* **1507**, 108–120 (2022).
74. Ashiqur Rahman, S. et al. Deep learning for biological age estimation. *Brief. Bioinform.* **22**, 1767–1781 (2021).
75. Sagers, L., Melas-Kyriazi, L., Patel, C. J. & Manrai, A. K. Prediction of chronological and biological age from laboratory data. *Aging* **12**, 7626–7638 (2020).
76. Galibourg, A. et al. Comparison of different machine learning approaches to predict dental age using Demirjian's staging approach. *Int. J. Legal Med.* **135**, 665–675 (2021).
77. Egorova, N., Liem, F., Hachinski, V. & Brodtmann, A. Predicted brain age after stroke. *Front. Aging Neurosci.* **11**, 348 (2019).
78. Gialluisi, A., Di Castelnuovo, A., Donati, M. B., de Gaetano, G. & Iacoviello, L. Machine learning approaches for the estimation of biological aging: the road ahead for population studies. *Front. Med.* **6**, 146 (2019).
79. Feditsev, A. et al. Markers of different machine learning approaches could serve as accurate non-invasive predictors of human biological and chronological age. *Aging* **9**, 1280–1292 (2017).
80. Yang, Q. et al. A machine learning-based data mining in medical examination data: a biological features-based biological age prediction model. *BMC Bioinformatics* **23**, 411 (2022).
81. Wang, T. et al. Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk. *Clin. Epigenetics* **12**, 5 (2020).
82. Armstrong, N. J. et al. Aging, exceptional longevity and comparisons of the Hannum and Horvath epigenetic clocks. *Epigenomics* **9**, 689–700 (2017).
83. Kresovich, J. K. et al. Methylation-based biological age and breast cancer risk. *J. Natl. Cancer Inst.* **111**, 1051–1058 (2019).
84. Lu, A. T. et al. DNA methylation GrimAge version 2. *Aging* **14**, 9484–9549 (2022).
85. McCrory, C. et al. GrimAge outperforms other epigenetic clocks in the prediction of age-related clinical phenotypes and all-cause mortality. *J. Gerontol. A. Biol. Sci. Med. Sci.* **76**, 741–749 (2021).
86. Föhr, T. et al. Does the epigenetic clock GrimAge predict mortality independent of genetic influences: an 18 year follow-up study in older female twin pairs. *Clin. Epigenetics* **13**, 128 (2021).
87. Belsky, D. W. et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *Elife* **11**, e73420 (2022).
88. Lim, U. & Song, M.-A. DNA methylation as a biomarker of aging in epidemiologic studies. *Methods Mol. Biol.* **1856**, 219–231 (2018).
89. Bose, M. et al. Evaluation of microarray-based DNA methylation measurement using technical replicates: the Atherosclerosis Risk In Communities (ARIC) Study. *BMC Bioinformatics* **15**, 312 (2014).
90. Higgins-Chen, A. T. et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. *Nat. Aging* **2**, 644–661 (2022).
91. Li, Q. et al. Homeostatic dysregulation proceeds in parallel in multiple physiological systems. *Aging Cell* **14**, 1103–1112 (2015).
92. Belsky, D. W., Huffman, K. M., Pieper, C. F., Shalev, I. & Kraus, W. E. Change in the rate of biological aging in response to caloric restriction: CALERIE Biobank Analysis. *J. Gerontol. A. Biol. Sci. Med. Sci.* **73**, 4–10 (2017).
93. Cohen, A. A. et al. A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. *Mech. Ageing Dev.* **134**, 110–117 (2013).
94. Jin, S. Y. et al. Association of lifestyle with mortality and the mediating role of aging among older adults in China. *Arch. Gerontol. Geriatr.* **98**, 104559 (2022).
95. McEwen, B. S. Stress, adaptation, and disease. Allostasis and allostatic load. *Ann. N. Y. Acad. Sci.* **840**, 33–44 (1998).
96. Carbone, J. T., Clift, J. & Alexander, N. Measuring allostatic load: approaches and limitations to algorithm creation. *J. Psychosom. Res.* **163**, 111050 (2022).
97. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
98. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, giz082 (2019).
99. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
100. Nie, C. et al. Distinct biological ages of organs and systems identified from a multi-omics study. *Cell Rep.* **38**, 110459 (2022).
101. Gheorghie, M. et al. Major aging-associated RNA expressions change at two distinct age-positions. *BMC Genomics* **15**, 132 (2014).
102. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).

103. National Institute on Aging. Publicly available databases for aging-related secondary analyses in the behavioral and social sciences. <https://www.nia.nih.gov/research/dbcsr/publicly-available-databases-aging-related-secondary-analyses-behavioral-and-social> (2013).
104. Thomas, S. Telomeres as Sentinels for Environmental Exposures, Psychosocial Stress, and Disease Susceptibility. <https://www.nia.nih.gov/research/dbcsr/telomeres-sentinels-environmental-exposures-psychosocial-stress-and-disease> (2017).
105. Elliott, M. L. et al. Disparities in the pace of biological aging among midlife adults of the same chronological age have implications for future frailty risk and policy. *Nat. Aging* **1**, 295–308 (2021).

ACKNOWLEDGEMENTS

This work was supported by grants from the National Key Research and Development Program of China (No. 2018YFC2000703-1), the National Natural Science Foundation of China (Grant No. 81872686), Jiangsu Provincial Fourth "333 Project", the Priority Academic Program Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine) and the Flagship Major Development of Jiangsu Higher Education Institutions.

AUTHOR CONTRIBUTIONS

C.S. designed the study. S.E.C.B. drafted the manuscript. C.S. revised the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41514-023-00110-8>.

Correspondence and requests for materials should be addressed to Chong Shen.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2023