OPEN

Blood Test–Based Age Acceleration Is Inversely Associated with High-Volume Sports Activity

VENCEL JUHÁSZ[1], ANNA ORSZÁG[2], DOROTTYA BALLA[1], LILIÁNA SZABÓ[1], NÓRA SYDÓ[1,3], ORSOLYA KISS[1,3], EMESE CSULAK[1], MÁTÉ BABITY[1], ZSÓFIA DOHY[1], RÉKA SKODA[1], DÁVID BECKER[1], BÉLA MERKELY[1,3], ANDRÁS BENCZÚR[2], HAJNALKA VÁGÓ[1,3], and CSABA KEREPESI[2]

[1]Heart and Vascular Centre, Semmelweis University, Budapest, HUNGARY; [2]Institute for Computer Science and Control (SZTAKI), Hungarian Research Network (HUN-REN), Budapest, HUNGARY; [3]Department of Sports Medicine, Semmelweis University, Budapest, HUNGARY

ABSTRACT

JUHÁSZ, V., A. ORSZÁG, D. BALLA, L. SZABÓ, N. SYDÓ, O. KISS, E. CSULAK, M. BABITY, Z. DOHY, R. SKODA, D. BECKER, B. MERKELY, A. BENCZÚR, H. VÁGÓ, and C. KEREPESI. Blood Test–Based Age Acceleration Is Inversely Associated with High-Volume Sports Activity. *Med. Sci. Sports Exerc.*, Vol. 56, No. 5, pp. 868–875, 2024. **Purpose:** We develop blood test–based aging clocks and examine how these clocks reflect high-volume sports activity. **Methods:** We use blood tests and body metrics data of 421 Hungarian athletes and 283 age-matched controls (mean age, 24.1 and 23.9 yr, respectively), the latter selected from a group of healthy Caucasians of the National Health and Nutrition Examination Survey (NHANES) to represent the general population (n = 11,412). We train two age prediction models (i.e., aging clocks) using the NHANES dataset: the first model relies on blood test parameters only, whereas the second one additionally incorporates body measurements and sex. **Results:** We find lower age acceleration among athletes compared with the age-matched controls with a median value of −1.7 and 1.4 yr, $P < 0.0001$. BMI is positively associated with age acceleration among the age-matched controls ($r = 0.17$, $P < 0.01$) and the unrestricted NHANES population ($r = 0.11$, $P < 0.001$). We find no association between BMI and age acceleration within the athlete dataset. Instead, age acceleration is positively associated with body fat percentage ($r = 0.21$, $P < 0.05$) and negatively associated with skeletal muscle mass (Pearson $r = -0.18$, $P < 0.05$) among athletes. The most important blood test features in age predictions were serum ferritin, mean cell volume, blood urea nitrogen, and albumin levels. **Conclusions:** We develop and apply blood test–based aging clocks to adult athletes and healthy controls. The data suggest that high-volume sports activity is associated with slowed biological aging. Here, we propose an alternative, promising application of routine blood tests. **Key Words:** AGING CLOCK, ATHLETES, BIOLOGICAL AGE, BLOOD TEST, PHYSICAL ACTIVITY

Aging has a significant effect on health, the economy, and society, yet its biological mechanisms remain poorly understood. Importantly, there is a distinction between biological age and chronological age. Biological age can deviate from chronological age, with higher or lower values indicating accelerated or slowed aging, respectively. Recently, epigenetic clocks, which are machine learning models predicting an individual's age based on epigenetic markers, have emerged as the gold standard tools for measuring the aging process (1–6). Epigenetic clocks have demonstrated accelerated aging in individuals afflicted with aging-related diseases and outperformed chronological age in predicting mortality. Accordingly, it is thought that epigenetic clocks effectively measure biological age (7–12). Certain dietary and lifestyle factors influence biological age as determined by epigenetic clocks. For instance, factors such as consuming fish, poultry, fruits, and vegetables and engaging in regular exercise have been associated with slower epigenetic aging. Conversely, factors such as red meat consumption, high blood pressure, high BMI, and smoking have been linked to accelerated epigenetic aging (6,13,14).

Address for correspondence: Hajnalka Vágó, M.D., Ph.D., 68 Városmajor St, Budapest H-1122, Hungary; E-mail: vago.hajnalka@semmelweis.hu; Csaba Kerepesi, Kende u. 13-17. Budapest, H-1111, Hungary; E-mail: kerepesi@sztaki.hu
Vencel Juhász, Anna Ország, Hajnalka Vágó, and Csaba Kerepesi contributed equally and are shared first or senior authors.
Submitted for publication September 2023.
Accepted for publication December 2023.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.acsm-msse.org).

0195-9131/24/5605-0868/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000003380

Over the recent years, various methods have emerged to assess the biological age of young athletes. McClean et al. (15) used hand–wrist imaging to predict biological age in pediatric athletes. They found that ECG interpretation governed by biological age provided excellent diagnostic accuracy and performed better in detecting an underlying cardiac condition than chronological age. Another application involves measuring the telomere length, which seems to carry clinical importance as early as childhood. Skilton et al. (16) found that arterial wall thickness, a marker of vascular disease risk, was associated with decreased telomere length, a marker of aging, already at the age of 8 yr. Spólnicka et al. (17) examined DNA methylation of young elite athletes (mean age 24.8 yr) and matched controls (mean age 24.2 yr) by using epigenetic clocks based on five CpG sites. They found an increased epigenetic age acceleration in athletes, which contradicts epidemiological data showing longer life span of athletes. The age predictions were based on only two genes, TRIM59 and KLF14, which are linked to anticancer and anti-inflammatory effects potentially contributing to the increased life expectancy of athletes. The authors concluded that intense physical training might have a complex influence on the aging process.

Here, we aim to investigate whether standard blood tests are suitable for determining the biological age of athletes and the general population. Standard blood tests offer advantages because of their cost-effectiveness, regular availability, and widespread accessibility in significant quantities. Blood test–based age prediction models were developed using deep learning, and the age acceleration of these clocks was associated with all-cause mortality and smoking status (18–20). A recent study using the National Health and Nutrition Examination Survey (NHANES) database for training and testing discovered that factors such as being male, having a lower socioeconomic status, exposure to tobacco, leading a sedentary lifestyle, experiencing obesity, or having a systemic disease were linked to a higher deviation in personalized physiological age (PPA) (21).

In this study, we train and test aging clocks based on blood tests and body measurements using the publicly available NHANES database. Additionally, we apply the clocks on a dataset comprising 421 Hungarian athletes previously collected in another study (22).

METHODS

Participants and measurements. We studied 421 Hungarian athletes of Caucasian ethnicity, ranging in age from 14 to 56 yr (hereafter referred to as the *athlete dataset*). The athletes included individuals engaged in leisure ($n = 56$), competitive ($n = 211$), and elite-level ($n = 154$) sports activities (categorized according to the 2020 ESC guidelines [23]), and their blood samples were collected at a tertiary cardiovascular center. The recruitment strategy did not involve any active solicitation or intervention. Professional athletes visited the clinic for mandatory testing before being permitted to return to play after SARS-CoV-2 infection, whereas leisure athletes requested testing on their own initiative. The laboratory testing was carried out approximately 24 d after the athletic subjects' SARS-CoV-2 infection. The detailed results of these tests have been published by Juhász et al. (22). Note that written informed consent was obtained from the participants beforehand.

We included the following 36 blood test measurements in our calculations: red blood cell count (million cells per microliter), red blood cell distribution width (%), white blood cell count (1000 cells per microliter), segmented neutrophil count (1000 cells per microliter), lymphocyte count (1000 cells per microliter), basophil count (1000 cells per microliter), eosinophil count (1000 cells per microliter), monocyte count (1000 cells per microliter), platelet count (1000 cells per microliter), mean cell hemoglobin (pg), mean cell hemoglobin concentration ($g \cdot dL^{-1}$), mean cell volume (fL), hemoglobin ($g \cdot dL^{-1}$), hematocrit (%), iron in refrigerated serum ($\mu mol \cdot L^{-1}$), transferrin saturation (%), total iron binding capacity ($\mu mol \cdot L^{-1}$), ferritin ($\mu g \cdot L^{-1}$), C-reactive protein ($mg \cdot dL^{-1}$), gamma-glutamyl transferase (GGT, $IU \cdot L^{-1}$), alkaline phosphatase ($U \cdot L^{-1}$), alanine transaminase (ALT or GPT, $U \cdot L^{-1}$), aspartate-aminotransferase (AST or GOT, $U \cdot L^{-1}$), total protein ($g \cdot L^{-1}$), albumin ($g \cdot dL^{-1}$), LDH ($U \cdot L^{-1}$), uric acid ($\mu mol \cdot L^{-1}$), total bilirubin ($\mu mol \cdot L^{-1}$), LDL cholesterol ($mmol \cdot L^{-1}$), HDL cholesterol ($mmol \cdot L^{-1}$), triglycerides ($mmol \cdot L^{-1}$), creatinine in refrigerated serum ($\mu mol \cdot L^{-1}$), blood urea nitrogen ($mmol \cdot L^{-1}$), sodium ($mmol \cdot L^{-1}$), potassium ($mmol \cdot L^{-1}$), and glucose in serum ($mmol \cdot L^{-1}$). We also measured weight, height, BMI, body fat percentage, skeletal muscle mass (SMM), and the reported weekly training hours of the athletes. Body composition was derived from a bioimpedance-based method using Inbody 770 (InBody, Cerritos, CA).

From the NHANES database, we selected Caucasian individuals who self-reported being in good, very good, or excellent health status. We further refined our selection to individuals for whom all 36 blood measurements, along with height, weight, and BMI data, were available. For better modeling accuracy, we only included records with age information provided in months. After applying these criteria, our search strategy yielded a total of 11,412 patient records from the period spanning 2001 to 2010 with an age range from 12 to 79.2 yr. This dataset is referred to as the *NHANES full dataset*. Details of the measurements used in the study are shown in Table 1.

Machine learning-based age prediction. In constructing aging clocks based on blood tests, we used machine learning algorithms to train a regression model. In this setup, chronological age served as the outcome, whereas predictors included blood test parameters, body measurements, and sex. First, we split the 11,412 samples of the NHANES full dataset into a training set of 9243 samples, a validation set of 1027 samples, and a test set of 1142 samples, hereafter referred to as the *NHANES training*, *validation*, and *test datasets*. We refrained from employing imputation methods; rather, we kept missing values intact in the data. Missing value distributions of the datasets are available in Supplemental Figure 1 (Supplemental Digital Content 1, The distribution of missing values in the different datasets, http://links.lww.com/MSS/C991).

We deployed LightGBM, an efficient implementation of the gradient boosting decision tree algorithm (24). We trained two age predictors by using the NHANES training dataset. The first

TABLE 1. Descriptive statistics of the measurements used in the athlete dataset, the age-matched NHANES test dataset, and the NHANES full dataset.

	Athlete Dataset			Age-Matched NHANES Test Dataset			NHANES Full Dataset		
	n	Mean	SD	*n*	Mean	SD	*n*	Mean	SD
Age	421	24.1	8.8	283	23.9	8.7	11,412	41.8	20.2
Number of females	167 (40%)	N/A	N/A	149 (53%)	N/A	N/A	5,772 (51%)	N/A	N/A
Number of males	254 (60%)	N/A	N/A	134 (47%)	N/A	N/A	5,640 (49%)	N/A	N/A
Weight (kg)	421	77.1	15.2	280	75.9	20.7	11,309	78.0	20.3
Standing height (cm)	421	179.6	9.9	281	170.6	20.1	11,314	169.6	9.9
Body mass index ($kg \cdot m^{-2}$)	421	23.7	3.2	280	26.1	6.4	11,283	27.0	6.2
SMM (kg)	121	34.9	7.4	N/A	N/A	N/A	N/A	N/A	N/A
Body fat percentage (%)	125	19.2	6.8	N/A	N/A	N/A	N/A	N/A	N/A
Weekly training hours	420	13.2	6.5	N/A	N/A	N/A	N/A	N/A	N/A
Albumin ($g \cdot dL^{-1}$)	421	49.5	3.1	259	43.4	4.5	10,900	43.0	3.6
Alkaline phosphatase ($U \cdot L^{-1}$)	421	86.3	51.6	61	85.6	55.5	2,418	84.2	56.4
Alanine transaminase (ALT/GPT, $U \cdot L^{-1}$)	421	19.3	14.6	258	21.3	12.7	10,846	24.0	24.5
Aspartate-aminotransferase (AST/GOT, $U \cdot L^{-1}$)	421	21.6	7.3	258	23.0	6.9	10,845	24.7	11.4
Basophil count (1000 cells per microliter)	421	0.0	0.0	263	0.0	0.1	10,976	0.0	0.1
Blood urea nitrogen ($mmol \cdot L^{-1}$)	421	5.2	1.3	259	3.9	1.4	10,900	4.5	1.8
C-reactive protein ($mg \cdot dL^{-1}$)	421	0.1	0.6	263	0.3	0.6	10,949	0.3	0.7
Creatinine in refrigerated serum ($\mu mol \cdot L^{-1}$)	421	77.7	14.1	259	72.2	15.6	10,900	76.7	22.2
Eosinophil count (1000 cells per microliter)	421	0.1	0.1	263	0.2	0.1	10,976	0.2	0.2
Ferritin ($\mu g \cdot L^{-1}$)	421	94.1	88.6	160	58.2	82.3	5,105	75.7	105.3
Gamma-glutamyl transferase (GGT, $IU \cdot L^{-1}$)	421	15.6	14.4	259	18.3	16.2	10,900	23.7	30.4
Glucose in serum ($mmol \cdot L^{-1}$)	421	5.1	3.3	259	4.8	0.6	10,900	5.2	1.4
HDL cholesterol ($mmol \cdot L^{-1}$)	421	1.5	0.4	200	1.3	0.4	8,497	1.4	0.4
Hematocrit (%)	421	43.1	3.3	263	42.4	4.6	11,000	42.4	4.1
Hemoglobin ($g \cdot dL^{-1}$)	421	14.8	1.2	263	14.5	1.6	11,000	14.5	1.4
Iron in refrigerated serum ($\mu mol \cdot L^{-1}$)	421	20.8	7.1	259	16.5	7.2	10,894	16.3	6.7
LDH ($U \cdot L^{-1}$)	421	284.1	55.8	197	119.9	21.7	8,426	128.6	28.0
LDL cholesterol ($mmol \cdot L^{-1}$)	421	2.8	0.8	129	2.6	0.7	5,068	2.9	0.9
Lymphocyte count (1000 cells per microliter)	421	2.0	0.5	263	2.2	0.6	10,976	2.1	1.1
Mean cell hemoglobin concentration ($g \cdot dL^{-1}$)	421	34.2	1.4	263	34.2	0.8	11,000	34.2	0.8
Mean cell hemoglobin (pg)	421	30.6	13.5	263	30.5	1.9	11,000	30.8	1.9
Mean cell volume (fL)	421	87.0	3.8	263	89.2	4.7	11,000	90.0	4.7
Monocyte count (1000 cells per microliter)	421	0.5	0.1	263	0.6	0.2	10,976	0.6	0.2
Platelet count (1000 cells per microliter)	421	242.6	53.0	263	260.8	53.6	11,000	263.0	65.9
Potassium ($mmol \cdot L^{-1}$)	421	4.3	0.3	259	4.0	0.3	10,900	4.0	0.3
Red blood cell count (million cells per microliter)	421	5.0	0.5	263	4.8	0.6	11,000	4.7	0.5
Red blood cell distribution width (%)	421	12.1	0.9	263	12.4	0.7	11,000	12.5	0.9
Segmented neutrophil count (1000 cells per microliter)	421	3.5	1.3	263	4.6	1.8	10,976	4.4	1.7
Sodium ($mmol \cdot L^{-1}$)	421	138.8	1.7	259	138.8	2.2	10,899	139.1	2.2
Total bilirubin ($\mu mol \cdot L^{-1}$)	421	11.0	7.3	259	13.5	7.1	10,895	13.3	5.4
Total iron binding capacity ($\mu mol \cdot L^{-1}$)	421	69.7	10.7	119	68.0	11.2	4,087	66.8	11.6
Total protein ($g \cdot L^{-1}$)	421	71.8	4.2	259	71.7	5.4	10,892	71.0	4.6
Transferrin saturation (%)	421	30.5	11.6	119	24.3	11.7	4,086	24.8	11.8
Triglycerides ($mmol \cdot L^{-1}$)	421	1.2	0.7	259	1.4	1.0	10,895	1.6	1.4
Uric acid ($\mu mol \cdot L^{-1}$)	421	297.8	73.9	259	315.2	87.0	10,899	317.5	80.9
White blood cell count (1000 cells per microliter)	421	6.2	1.5	263	7.6	2.1	11,000	7.3	2.3

The first 10 rows contain age, sex, body metrics, and weekly training hours, followed by 36 laboratory blood test measurements.

predictor solely encompassed variables related to blood test parameters (referred to as clock 1), whereas the second one incorporated body measurements (height, weight, and BMI) and sex in addition (referred to as clock 2). We performed model parameter optimization using the NHANES validation set. We selected the models with the lowest mean absolute error (MAE) obtained between chronological age and predicted age. For both clocks, the optimal model was constructed by using depth 5 trees with 10 leaves, a learning rate of 0.2, and 100 boosting iterations.

To maintain the integrity of the comparison with athletes and to prevent any overlap with data used for model training or parameter optimization, additional experiments were conducted exclusively on the NHANES test set, serving as an independent testing set. Out of the NHANES test set, we selected a control group whose age distribution matched the age distribution of the athlete dataset, which we hereafter refer to as the *age-matched NHANES test dataset*. By age matching, we sampled 283 of the 1142 individuals in the NHANES test dataset, see Table 1.

Predicted age, delta age, and age acceleration. We examined three distinct aging-related metrics that have the potential to capture the difference between chronological and biological age. These metrics encompass predicted age, delta age, and age acceleration.

The predicted age is the output of the age prediction models (clock 1 and clock 2) based on blood test parameters. We hypothesize that predicted age is a better estimate of biological age than chronological age. We calculated delta age (or age gap) by subtracting the chronological age from the predicted age of each individual (i.e., delta age = predicted age − chronological age). This measurement represents the difference between biological age and chronological age. A higher delta age indicates accelerated aging when comparing two individuals with the same chronological age.

Age acceleration is defined by Thompson et al. (25) as the deviation of the individual's biological age from the trend. Age acceleration is calculated for an individual as the residual

TABLE 2. Performance of the blood test–based age prediction models.

Models		NHANES Test Dataset		Age-Matched NHANES Test Dataset		Athlete Dataset	
Model Name	Training Parameters	MAE (yr)	Pearson *r*	MAE (yr)	Pearson *r*	MAE (yr)	Pearson *r*
Clock 1	Laboratory parameters	9.33	0.80	9.63	0.54	9.29	0.44
Clock 2	Laboratory parameters, body metrics, sex	8.93	**0.82**	9.45	0.55	8.40	0.46

after fitting a regression line between predicted and chronological age. In our experiments, we relied on the regression line of the NHANES test dataset as a reference for both the NHANES test dataset and the athlete dataset. In our statistical analysis, we preferred age acceleration over delta age because age acceleration is independent of age by definition. In essence, our investigation focuses on the deviation from the usual aging pattern, irrespective of an individual's chronological age. This approach is deemed a more dependable measure for comparing aging-related effects across diverse age groups and populations.

Statistical analysis. We used the Python package SciPy (26) for statistical calculations. We calculated two-sided *t*-tests to compare two independent groups. A P value less than 0.05 was considered significant. For comparing repeated test results, a one-sample *t*-test was calculated for the differences (popmean = 0) after we adjusted the predicted age of the second measurement by the elapsed time. We evaluated correlations by the Pearson correlation coefficient (r). We used the following notation for P values: ns for $P > 0.05$, * $0.05 > P \geq 0.01$, ** $0.01 > P \geq 0.001$, *** $0.001 > P \geq 0.0001$, and **** $0.0001 \geq P$.

RESULTS

Performance of blood test–based age prediction is comparable for athletes and controls. To assess the biological age of athletes, we developed two aging clocks (clocks 1 and 2) using a large cohort of healthy Caucasian people from the NHANES database. We trained clock 1 on only blood tests, whereas clock 2 was trained using blood tests, body measurements (height, weight, BMI), and sex. We tested both clocks on three datasets: (i) the NHANES test dataset, (ii) the age-matched NHANES test dataset, and (iii) the athlete dataset (Table 2, Fig. 1). The prediction performance for the athlete dataset was comparable with the age-matched NHANES test dataset by both clocks (Table 2). The athlete dataset comprised 421 subjects, delineated as follows: 154 elite athletes (37%), 211 competitive athletes (50%), and 56 leisure athletes (13%). Among these athletes, 88% engaged in a minimum of 6 h of weekly training volume, contributing to an average weekly training duration of 13.2 h. The majority were involved in mixed sports disciplines ($n = 322$, 76%), followed by endurance ($n = 53$, 13%), power ($n = 38$, 9%), and skill ($n = 8$, 2%) types.

Blood test–based age acceleration is inversely associated with high-volume sports activity. We found a significantly lower age acceleration in the athlete dataset compared with the age-matched NHANES dataset by both aging clocks (clock 1: median = −1.7, IQR = [−8.9, 6.14], vs median = 1.4, IQR = [−3.8, 7.4]; clock 2: median = −1.7, IQR = [−7.6, 4.4], vs median = 1.7, IQR = [−2.8, 7.5] yr, $P < 0.0001$), Figures 2A and 2B. Our findings suggest that blood test–based age acceleration is inversely associated with high-volume sports activity.

We also examined the associations between BMI and blood test–based age acceleration. We observed that BMI was positively associated with clock 1 age acceleration for both the NHANES test dataset ($r = 0.113$, $P < 0.001$) and the age-matched NHANES test dataset ($r = 0.168$, $P < 0.01$), Figures 2C and 2D. By contrast, within the athlete dataset, we observed no association between BMI and age acceleration; however, age acceleration based on both aging clocks was positively associated with body fat percentage ($r = 0.208$ and $P < 0.05$ for clock 1, as well as $r = 0.259$ and $P < 0.01$ for clock 2), Figures 2E and 2F. Accordingly, SMM is negatively associated with the clock 1 age acceleration within the athlete dataset ($r = -0.182$, $P < 0.05$), Figure 2E. In summary, BMI is positively associated with blood test–based age acceleration in the general population; however, athletes exhibit distinct patterns. In their case, body fat percentage may be a more appropriate indicator of aging.

We further investigated the possible sex differences in blood test–based age acceleration. Age acceleration was higher in female athletes compared with male athletes (median = 2.9, IQR = [−2.8, 9.4], vs median = −5.9, IQR = [−12.1, 2.3] yr, $P < 0.0001$), Figure 2I. By contrast, in the age-matched

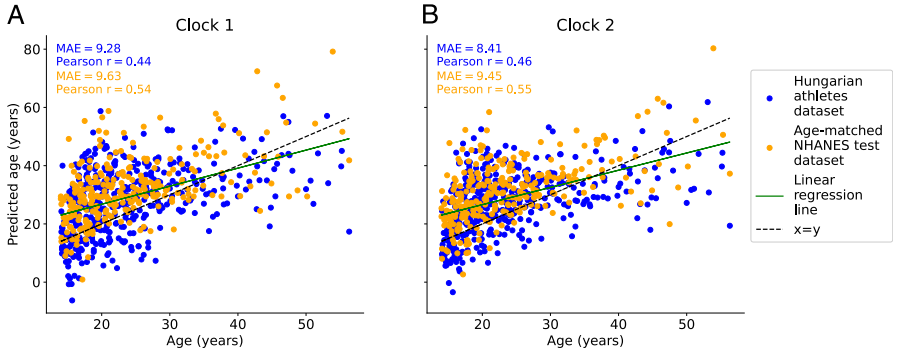

FIGURE 1—Chronological and predicted age over the age-matched NHANES test and the athlete dataset. A. Age prediction of clock 1 (blood test only). B. Age prediction of clock 2 (blood test, body metrics, and sex).

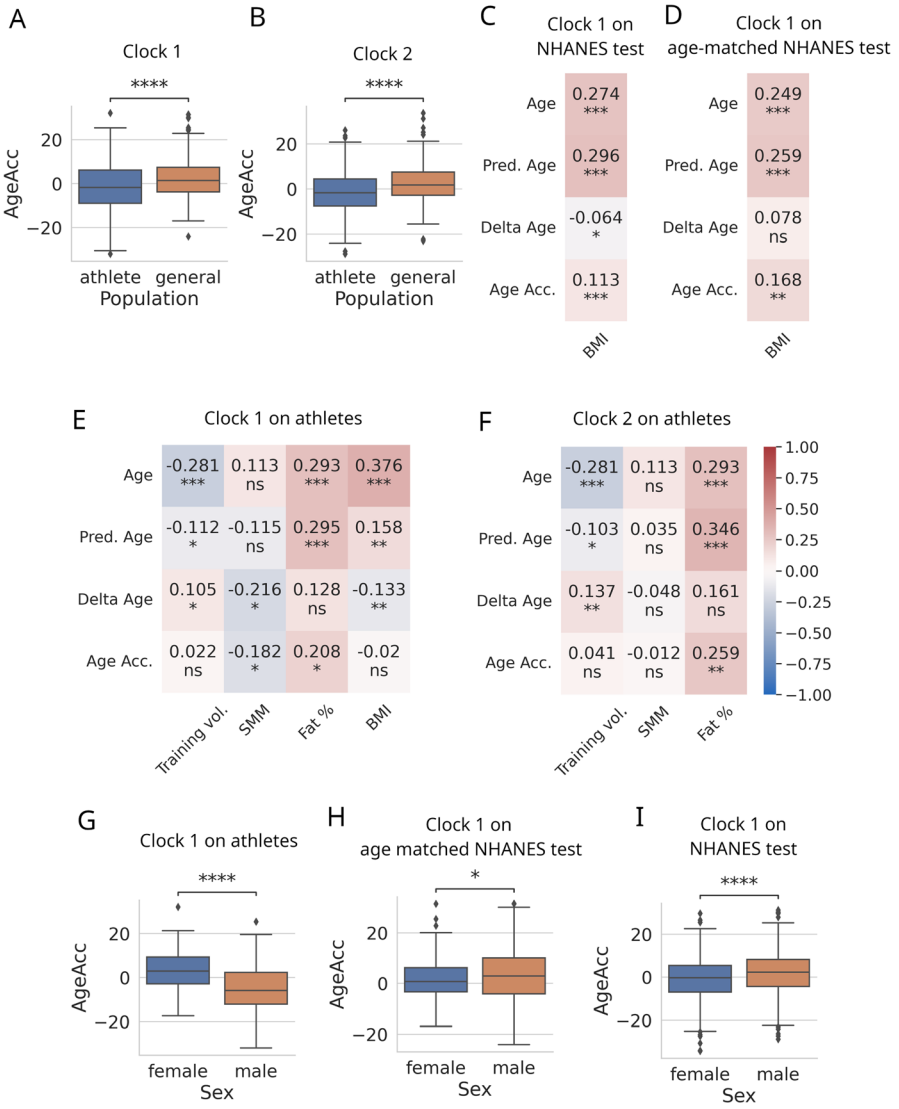

FIGURE 2—Inverse association of blood test–based age acceleration and high-volume sports activity. A–B. Differences in age acceleration between the athlete dataset (athlete population) and the age-matched NHANES dataset (general population) evaluated by clocks 1 and 2, respectively. C–F. Pearson correlation coefficient of age, predicted age, delta age, and age acceleration against training volume, SMM, fat percentage, and BMI over the different datasets. G–I. Sex difference in age acceleration over the three datasets. BMI and sex difference are analyzed only for clock 1, as clock 2 used both as training variables. Pred, predicted; Acc, acceleration; vol, volume.

NHANES test database, we observed that the mean age acceleration of females was lower compared with males (median = 0.7, IQR [−3.4, 6.2], vs median = 2.9, IQR [−4.1, 10.1], $P < 0.05$), Figure 2H. The same association was observed in the NHANES test dataset, which represents a general population with a broader age range (females: median = −0.3, IQR [−6.9, 5.4] vs males: median = 2.3, IQR [−4.3, 8.2], $P < 0.0001$), Figure 2I.

We hypothesized that the observed sex differences influence the decreased age acceleration in athletes, as depicted in Figures 2A and 2B. Therefore, we separated the athlete dataset and the age-matched NHANES dataset by sex and found that the effect remained in males but did not remain in females (Supplemental Fig. 2, Supplemental Digital Content 1, http://links.lww.com/MSS/C991). In other words, the age acceleration of female athletes was not lower than that of female controls.

Feature importance analysis. We applied the Shapley Additive Explanation method (27) to examine the most important features used by clocks 1 and 2. We investigated the effect of the 10 most important features on the predicted age when clocks 1 and 2 were applied separately to the athlete dataset and the age-matched NHANES test dataset (Fig. 3). Serum ferritin levels, mean cell volume, blood urea nitrogen, and serum albumin are among the essential variables in the age prediction in the NHANES dataset, with the first three positively and the fourth negatively associated with laboratory test–based age. Another observation was that for clock 2, BMI is a more important variable in the age-matched NHANES dataset than in the athlete dataset. Sex was the sixth most important feature when testing in the general population, but it was not among the 10 most important ones (14th) for athletes.

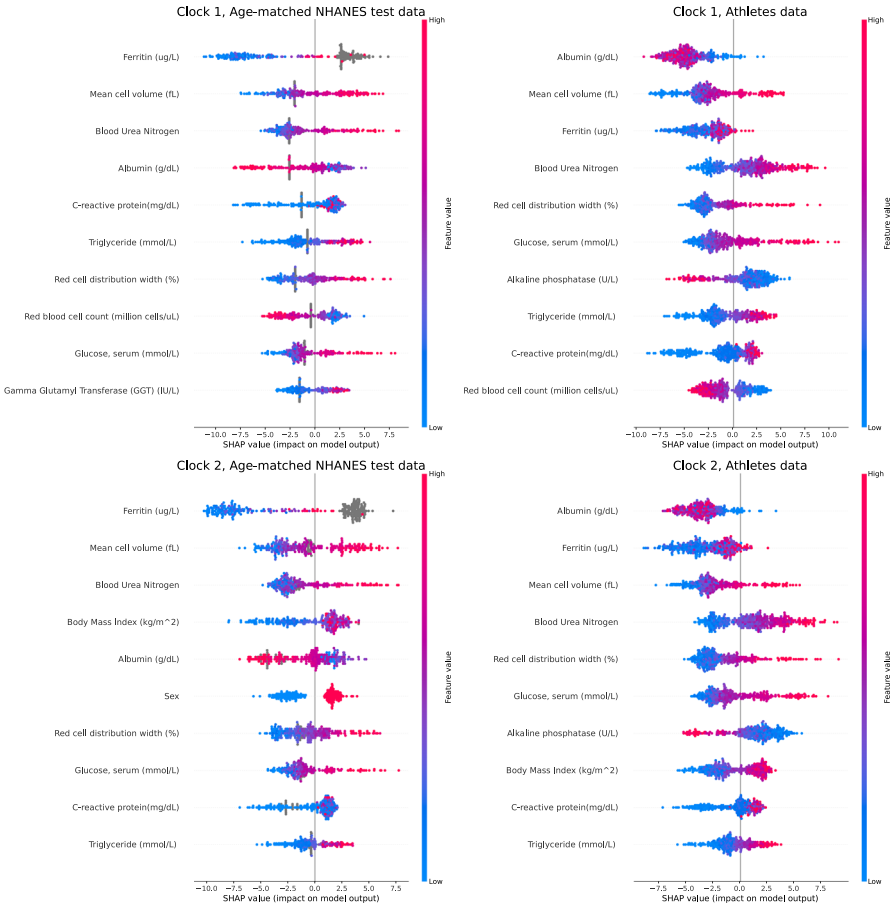

FIGURE 3—The most important features of the two aging clocks evaluated on the age-matched NHANES test and the athlete datasets. Each dot indicates a test sample. *Dots* concentrating on the left or right side of the panels indicate a negative or positive effect on the predicted age, respectively. *Blue dots* indicate low, and *red dots* indicate high values of the given variable for the given test sample (*gray dots* indicate missing values).

Repeated test analysis. In our experiments, we use the laboratory testing data from athletes for a secondary purpose, primarily aimed at investigating asymptomatic, mild, or moderately symptomatic SARS-CoV-2 infections. Subsequently, for certain individuals, a follow-up examination was conducted after their recovery. Among the athlete dataset, complete repeated blood tests were available for 20 individuals, with a median time difference of 107 d between the samplings. To evaluate the potential effect of recent SARS-CoV-2 infection on blood test–based age, we analyzed the variance in predicted age between the two measurements, accounting for the elapsed time. The difference was not significant by using one-sample *t*-tests (see Methods). We also examined the consistency of age acceleration between the two measurements. The age acceleration of the first measurement was strongly associated with the age acceleration of the second measurement for both clock 1 ($r = 0.7231$, $P = 0.00031$) and clock 2 ($r = 0.6184$, $P = 0.00365$).

DISCUSSION

We presented two aging clocks trained on blood tests, body measurements, and sex data in different setups and demonstrated a stable performance of the clocks across the examined general and athletic populations. After matching the ethnicity (considering only Caucasians), health status (considering only reportedly healthy individuals), and age between the NHANES database and the Hungarian athlete datasets, we observed no considerable difference in the accuracy of the clocks across the groups (Table 2). Interestingly, we observed a decrease in performance (a remarkable reduction of r and a slight increase in MAE for both clocks) after narrowing the age range of the NHANES test dataset by age matching to the athlete dataset. This observation indicates that age distribution may influence age prediction performance evaluations.

A previous study examined the performance of blood test–based age prediction models across different populations (18). When the models were trained and tested on the same population, the best performance was $r = 0.84$ and MAE = 6.25 (East Europe). However, the performance decreased remarkably as the models were trained and tested across different populations. For example, in the case when a model was trained on blood tests and sex variables of the Canadian population and tested on the East European population, the performance degraded to $r = 0.52$ and MAE = 9.68. Here, we reached similar performances when we trained on the Caucasian American population and tested on the Caucasian Hungarian population.

Blood test–based age acceleration was significantly lower in the athlete dataset for both clocks, which underlines the

beneficial effect of regular physical activity. An intriguing finding is that age acceleration is lower in athletic males compared with athletic females (Fig. 2G), but this phenomenon is reversed in nonathletes (Figs. 2H and 2I). In other words, performing high-volume physical activity has a greater effect on biological age in young adult males than in females. The sex-separated analysis (Supplemental Fig. 2, Supplemental Digital Content 1, http://links.lww.com/MSS/C991) also supports this idea.

In the general population, BMI was positively associated with increased biological aging measures. By contrast, when the same clock was applied to athletes, BMI was not associated with age acceleration. This suggests that using BMI is more appropriate in nonathletes, but other measures are necessary in athletes to elucidate associations between biological age and body composition. Indeed, in a cohort of 2435 middle-aged, nonathlete subjects, an increased accelerometer-measured physical activity was associated with lower epigenetic age, albeit BMI attenuated this effect (28). On the other hand, Garrido-Chamorro et al. (29,30) found that BMI alone is unsuitable for reporting the quality of body composition in athletes despite showing a good association with body fat content.

By contrast, in athletes, SMM is inversely associated with age acceleration, indicating a beneficial effect on aging. Additionally, body fat percentage is positively associated with age acceleration, suggesting an adverse effect on aging in athletes. These two findings again underscore the importance of applying advanced body composition parameters in athletes within and beyond clinical environments.

Being highly active and engaging in regular physical activity can lead to a wide range of physiological changes in the body, some of which are reflected in laboratory markers (31). These changes may have implications not only in the context of daily diagnostics for clinicians working with athletes but also in the estimation of biological age. Our feature importance analysis revealed several clinically meaningful laboratory parameters (such as serum ferritin, blood urea nitrogen, mean cell volume, and albumin levels) that are important in age prediction by our blood test–based clocks.

The estimation of biological age has become an important tool in the assessment of athletes and physically active individuals in the last decade. In young adolescent athletes, biological maturation has been mainly measured using noninvasive methods, including skeletal age, pubertal status, and peak height velocity (32,33). Sex differences in the rate of decline in physical activity observed during adolescence and the biological age measured using peak height velocity were closely associated with the commencement of the reduction in exercise volume (34).

During adulthood, biological age has other meaningful implications in sports. The optimal or "best" age for different sports and even positions within one discipline (e.g., goalkeeper vs striker in soccer) is not readily definable. At the same time, determining when to retire from a professional athletic career is highly dependent on the individual's performance and, intuitively, biological age. Chronological age may also lack appropriateness in predicting the beginning of a decline in physical performance in athletes approaching the end of their professional careers (35).

In contemporary times, assessing the health status of competitive athletes gains increasing attention, which involves performing comprehensive and complex sports medicine investigations. Laboratory tests are widely available and performed in a variety of clinical settings. Our results suggest that it is practical to define and exploit the potential added value of blood tests in biological age estimation, as long as epigenetic testing is not routinely performed and remains expensive. In addition to its recognized prognostic significance, using the degree and direction of changes in biological age as an outcome variable could emerge as a promising marker for enhancing physical activity-based lifestyle interventions and extending its applicability beyond. Nevertheless, blood test–based biological age measuring models need further improvement and research to increase their accuracy and define applicability in athletes and other populations.

Limitations. Laboratory blood testing in the athlete dataset was carried out after recovery from an asymptomatic, mild, or moderately symptomatic SARS-CoV-2 infection. Although none of the subjects was hospitalized before due to COVID-19 and had no limiting symptoms upon presenting for blood testing, the recent infection might have caused some alterations in their results. On the other hand, we did not find a significant difference in the predicted age in those subjects who returned to a repeated laboratory testing median 107 d later. Furthermore, age accelerations showed a strong association between the two examined time points with both clocks.

In the absence of directly comparable training volume-related variables, we were not able to contrast the physical activity of the NHANES dataset and the athlete dataset. The lack of a high-volume control group from the same geographical area limited the direct comparison of the Hungarian athlete data to a local control group. We relied on data from the NHANES database in the United States, which provided a comparable set of matched variables, such as ethnicity, health status, and age. Although we achieved the same age prediction performance for the two matched datasets, there still can be some batch effect influencing the results. In addition, the training and testing data contain only Caucasians limiting usage and applicability to other ethnic groups.

CONCLUSIONS

We developed two distinct blood test–based aging clocks demonstrating stable performance and applied them to athletes and healthy controls. Our findings suggest that high-volume sports activity may slow down the aging process. Because laboratory tests are readily accessible, blood test–based aging clocks offer a convenient method for predicting the biological age of diverse populations including athletes. This approach also carries the potential to become an instrument for adjusting training programs and clinical treatments to biological age instead of chronological age.

The authors thank their assistants at the Semmelweis University Heart and Vascular Centre outpatient clinic for their indispensable help with the blood tests.

The authors declare no conflicts of interest to be reported. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. The results of the present study do not constitute endorsement by the American College of Sports

This study was supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory, Hungary. Project no. TKP2021-NKTA-46 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme.

Age prediction, delta age, age acceleration, and other data used to create Figures 1 and 2 are available in the Supplemental Data (Supplemental Digital Content 2, http://links.lww.com/MSS/C990). Python codes for the application of clocks 1 and 2 models are available on GitHub (https://github.com/annaorszag/BloodTestAgeClock).

C.K., H.V., D.B., and A.B. conceived and supervised the study concept and design. A.O. acquired the NHANES datasets and performed the data analysis, including machine learning. V.J., N.S., O.K., E.C., M.B., and H.V. carried out the clinical examination of the athletes and created the registry presented in this study (data acquisition). V.J. acquired the athlete dataset and contributed to the conceptualization of the study and analysis. V.J., C.K., and A.B. interpreted the data. C.K., V.J., A.B., and A.O. wrote the original draft of the manuscript. H.V., D.B., and B.M. contributed to project management, funding acquisition, and manuscript revision. D.B., L.S., Z.D., and R.S. contributed to data acquisition and edited the manuscript. V.J. and A.O. contributed equally and are shared first authors. H.V. and C.K. contributed equally and are shared senior authors.

REFERENCES

1. Bell CG, Lowe R, Adams PD, et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol*. 2019;20(1):249.
2. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
3. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359–67.
4. Horvath S, Oshima J, Martin GM, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)*. 2018;10(7):1758–75.
5. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*. 2019;11(2):303–27.
6. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 2018;10(4):573–91.
7. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 2018;19(6):371–84.
8. Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)*. 2015;7(12):1130–42.
9. Horvath S, Garagnani P, Bacalini MG, et al. Accelerated epigenetic aging in Down syndrome. *Aging Cell*. 2015;14(3):491–5.
10. Maierhofer A, Flunkert J, Oshima J, Martin GM, Haaf T, Horvath S. Accelerated epigenetic aging in Werner syndrome. *Aging (Albany NY)*. 2017;9(4):1143–52.
11. Marioni RE, Shah S, McRae AF, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol*. 2015;16(1):25.
12. Marioni RE, Shah S, McRae AF, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian birth cohort 1936. *Int J Epidemiol*. 2015;44(4):1388–96.
13. Quach A, Levine ME, Tanaka T, et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)*. 2017;9(2):419–46.
14. Jokai M, Torma F, Mcgreevy KM, et al. DNA methylation clock DNAmFitAge shows regular exercise is associated with slower aging and systemic adaptation. *Geroscience*. 2023;45(5):2805–17.
15. McClean G, Riding NR, Pieles G, et al. Prevalence and significance of T-wave inversion in Arab and Black paediatric athletes: should anterior T-wave inversion interpretation be governed by biological or chronological age? *Eur J Prev Cardiol*. 2019;26(6):641–52.
16. Skilton MR, Nakhla S, Ayer JG, et al. Telomere length in early childhood: early life risk factors and association with carotid intima-media thickness in later childhood. *Eur J Prev Cardiol*. 2016;23(10):1086–92.
17. Spólnicka M, Pośpiech E, Adamczyk JG, et al. Modified aging of elite athletes revealed by analysis of epigenetic age markers. *Aging (Albany NY)*. 2018;10(2):241–52.
18. Mamoshina P, Kochetov K, Putin E, et al. Population specific biomarkers of human aging: a big data study using South Korean, Canadian, and Eastern European patient populations. *J Gerontol A Biol Sci Med Sci*. 2018;73(11):1482–90.
19. Mamoshina P, Kochetov K, Cortese F, et al. Blood biochemistry analysis to detect smoking status and quantify accelerated aging in smokers. *Sci Rep*. 2019;9(1):142.
20. Putin E, Mamoshina P, Aliper A, et al. Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging (Albany NY)*. 2016;8(5):1021–30.
21. Bernard D, Doumard E, Ader I, et al. Explainable machine learning framework to predict personalized physiological aging. *Aging Cell*. 2023;22(8):e13872.
22. Juhász V, Szabó L, Pavlik A, et al. Short and mid-term characteristics of COVID-19 disease course in athletes: a high-volume, single-center study. *Scand J Med Sci Sports*. 2023;33(3):341–52.
23. Pelliccia A, Sharma S, Gati S, et al. 2020 ESC Guidelines on sports cardiology and exercise in patients with cardiovascular disease. *Eur Heart J*. 2021;42(1):17–96.
24. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;3149–57.
25. Thompson MJ, Chwialkowska K, Rubbi L, et al. A multi-tissue full lifespan epigenetic clock for mice. *Aging (Albany NY)*. 2018;10(10):2832–54.
26. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261–72.
27. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56–67.
28. Spartano NL, Wang R, Yang Q, et al. Association of accelerometer-measured physical activity and sedentary time with epigenetic markers of aging. *Med Sci Sports Exerc*. 2023;55(2):264–72.
29. Ode JJ, Pivarnik JM, Reeves MJ, Knous JL. Body mass index as a predictor of percent fat in college athletes and nonathletes. *Med Sci Sports Exerc*. 2007;39(3):403–9.
30. Garrido-Chamorro RP, Sirvent-Belando JE, Gonzalez-Lorenzo M, Martin-Carratala ML, Roche E. Correlation between body mass index and body composition in elite athletes. *J Sports Med Phys Fitness*. 2009;49(3):278–84.
31. Banfi G, Colombini A, Lombardi G, Lubkowska A. Metabolic markers in sports medicine. In: *Advances in Clinical Chemistry* vol. 56. Academic Press Inc; 2012. pp. 1–54.
32. Malina RM, Rogol AD, Cumming SP, Coelho E, Silva MJ, Figueiredo AJ. Biological maturation of youth athletes: assessment and implications. *Br J Sports Med*. 2015;49(13):852–9.
33. Johnson A, Doherty PJ, Freemont A. Investigation of growth, development, and factors associated with injury in elite schoolboy footballers: prospective study. *BMJ*. 2009;338:b490.
34. Cairney J, Veldhuizen S, Kwan M, Hay J, Faught BE. Biological age and sex-related declines in physical activity during adolescence. *Med Sci Sports Exerc*. 2014;46(4):730–5.
35. Senefeld JW, Hunter SK. Are masters athletic performances predictive of human aging in men and women? *Mov Sports Sci*. 2019;104:5–12.