Journals of Gerontology: BIOLOGICAL SCIENCES
Cite journal as: J Gerontol A Biol Sci Med Sci. 2013 June;68(6):667–674
doi:10.1093/gerona/gls233

Advance Access publication December 3, 2012

Modeling the Rate of Senescence: Can Estimated Biological Age Predict Mortality More Accurately Than Chronological Age?

Morgan E. Levine

Davis School of Gerontology, University of Southern California, Los Angeles.

Address correspondence to Morgan E Levine, BA, University of Southern California Ethel Percy Andrus Gerontology Center, 3715 McClintock Avenue Los Angeles, CA 90089–0191. Email: canon@usc.edu

Marina Morozova
2nd year PhD Life Sciences student

NHANES III dataset, 1988-1994

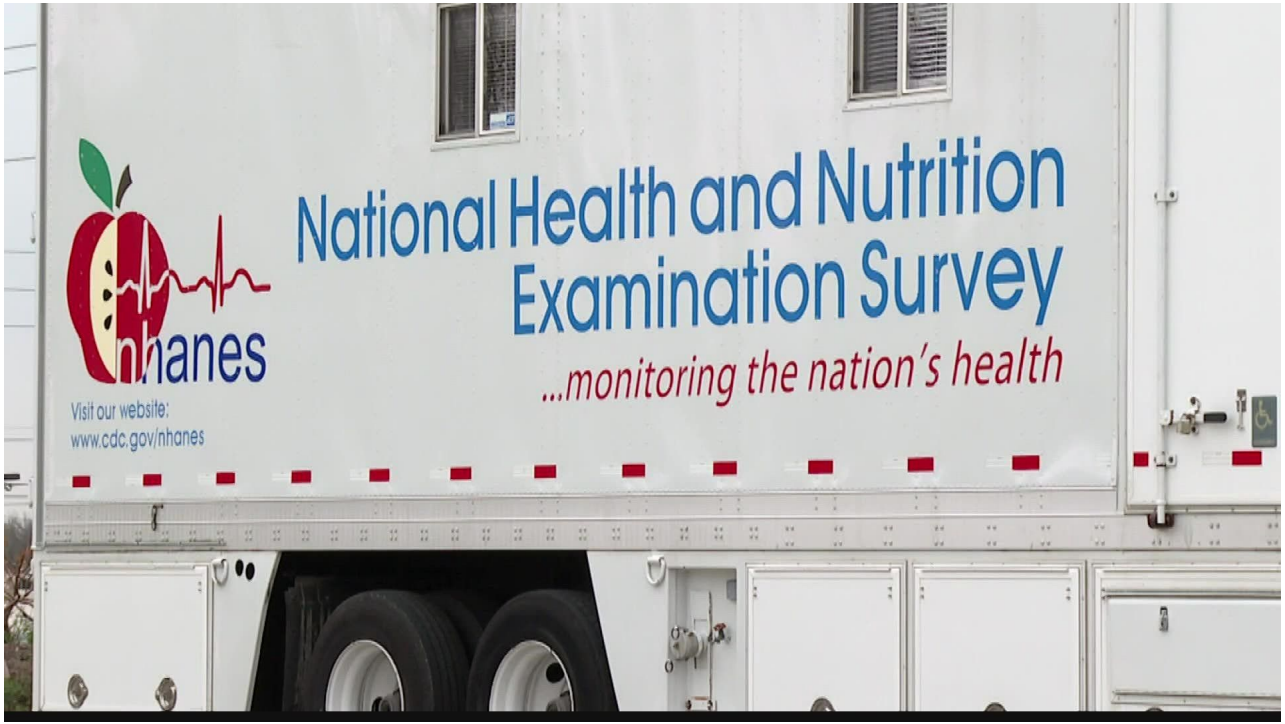
National Health and Nutrition Examination Surveys

Survey	*Dates*	*Ages*
NHES I	1960-62	18-79 years
NHES II	1963-65	6-11 years
NHES III	1966-70	12-17 years
NHANES I	1971-75	1-74 years
NHANES II	1976-80	6 mo.-74 years
HHANES	1982-84	6 mo.-74 years
NHANES III	1988-94	2 mo. +
NHANES	1999-	All ages

OP96S025

nhanes

National Health and Nutrition Examination Survey



https://wwwn.cdc.gov/nchs/nhanes/nhanes3/datafiles.aspx

33,994 persons

Household interviewing and examination appointments

At the beginning of the interviewing period, the field manager distributed segment (neighborhood) assignments to about 15 interviewers. They visited each household to administer a screening questionnaire that identified the household members, their ages and birth dates, and their racial and ethnic identities. Then according to the sampling instructions based on age, sex, and racial and ethnic identity contained on the screener, the interviewer selected sample persons from the household members.

Because of the effort to oversample the very young, the very old, black persons, and Mexican-Americans, interviewers had to screen many households. Although only about one

Interview at home (Household Adult/Youth Questionnaire, Family Questionnaire)

Schedule examination appointment

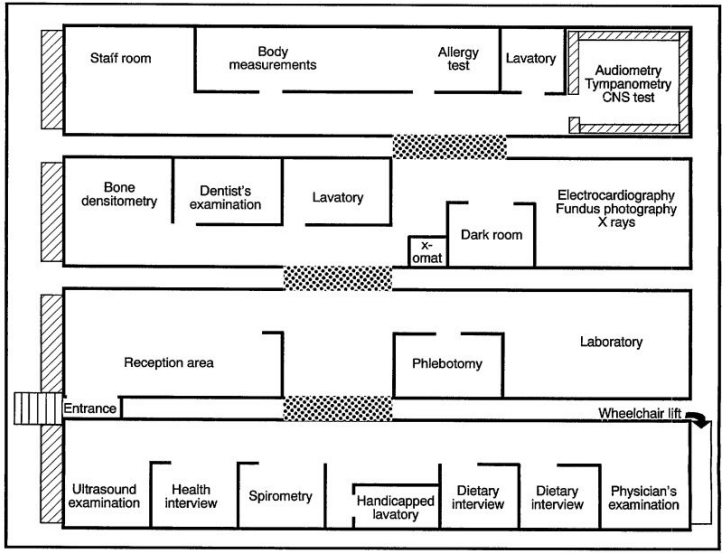


Figure 3. Floor plan of a mobile examination center used in the National Health and Nutrition Examination Survey: 1988–94

persuasion efforts from the interviewers. Household interviewers were very diligent about recontacting these people and succeeded in convincing many of them eventually to participate in the examination. These interviewers used a variety of techniques beyond the offer of remuneration, free transportation, and the results of the examination. Techniques included appealing to the sample person's sense of responsibility and community spirit. Eventually, about 14 percent of these people

Mobile Examination Center (MEC)

Follow-up National Death Index through December 31, 2006 (*archived*)

- Currently, 2019th National Death Index is available https://www.cdc.gov/nchs/data-linkage/mortality-public.htm
- It is Linked Mortality File that contains mortality follow-up data

```
SEQN ='NHANES Respondent Sequence Number'
ELIGSTAT = 'Eligibility Status for Mortality Follow-up'
MORTSTAT = 'Final Mortality Status'
UCOD_LEADING = 'Underlying Leading Cause of Death: Recode'
PERMTH_INT = 'Number of Person-Months of Follow-up from NHANES Interview date'
PERMTH_EXM = 'Number of Person-Months of Follow-up from NHANES Mobile Examination Center (MEC) Date'
```

UCOD_LEADING	Underlying Leading Cause of Death: Recode	Char	001	Diseases of heart (054-068)
			002	Malignant neoplasms (019-043)
			003	Chronic lower respiratory diseases (082-086)
			004	Accidents (unintentional injuries) (112-123)
			005	Cerebrovascular diseases (070)
			006	Alzheimer's disease (052)
			007	Diabetes mellitus (046)

Dataset overview

Assess Pearson's R and choose the best biomarkers
Age from 30 to 75 years old

1) Metabolic Function; 2) Cardiac Function;
3) Lung Function; 4) Kidney Function; 5) Liver Function;
6) Immune Function and Inflammation; 7) Cell Blood Count

	Biomarker	r	signif.
0	Serum C-reactive protein (mg/dL)	0.077	***
1	Serum creatinine (mg/dL)	0.146	***
2	Glycated hemoglobin: (%)	0.240	***
3	Serum albumin (g/dL)	-0.141	***
4	Serum cholesterol (mg/dL)	0.267	***
5	Cytomegalovirus optical density	-0.180	***
6	Serum blood urea nitrogen (mg/dL)	0.330	***
7	Serum alkaline phosphatase: SI (U/L)	0.180	***
8	Forced expiratory vol(FEV),.5 sec,max-ml	-0.480	***
9	Average K1 BP from household and MEC	0.492	***
10	Serum HDL cholesterol (mg/dL)	-0.001	
11	Hemoglobin (g/dL)	0.017	
12	Lymphocytes (percent of 100 cells)	-0.064	**
13	White blood cell count	-0.022	*
14	Hematocrit (%)	0.025	*
15	Red blood cell count	-0.050	***
16	Mononuclear percent (Coulter)	0.056	***
17	Granulocyte percent (Coulter)	0.043	***
18	Platelet count	-0.102	***
19	Pulse rate (beats/min) (age 5+ years)	0.039	***
20	Average K5 BP from household and MEC	0.022	*

	Full sample (N = 9717)	Aged 30–59 years (n = 6747)	Aged 60–75 years (n = 2858)
Age (y), M (std)	49.96, (13.34)	42.53, (8.17)	67.13, (4.31)
Female, %	52.85	54.26	49.51
Died, N	4105	1630	2399
Censored, N	5612	5117	459
Person-years, M (std)	22.74, (7.91)	25.41, (6.04)	16.52, (8.27)
Serum C-reactive protein (mg/dL), M (std)	0.50, (0.84)	0.46, (0.68)	0.59, (1.12)
Serum creatinine (mg/dL), M (std)	1.08, (0.41)	1.05, (0.34)	1.17, (0.52)
Glycated hemoglobin: (%), M (std)	5.66, (1.24)	5.51, (1.14)	6.01, (1.37)
Average K1 BP from household and MEC, M (std)	126.50, (18.64)	121.27, (15.83)	138.65, (19.13)
Serum albumin (g/dL), M (std)	4.12, (0.36)	4.15, (0.36)	4.06, (0.35)
Serum cholesterol (mg/dL), M (std)	211.07, (43.81)	205.10, (41.99)	224.82, (44.83)
Cytomegalovirus optical density, M (std)	1.20, (0.40)	1.24, (0.43)	1.10, (0.30)
Serum alkaline phosphatase: SI (U/L), M (std)	87.14, (32.85)	84.03, (31.38)	94.31, (35.24)
Forced expiratory vol(FEV),.5 sec,max-ml, M (std)	2317.83, (686.57)	2503.40, (638.39)	1890.65, (597.89)
Serum blood urea nitrogen (mg/dL), M (std)	14.24, (5.33)	13.17, (4.36)	16.73, (6.35)

Biological Age (BA) Estimates

1. Principal Component analysis (PCA)

2. Multiple linear regression

3. Klemera and Doubal's method

Principal Component analysis as BA

Principal component analysis.—The 10 biomarkers selected from the Pearson correlation were included in the PCA and run separately for men and women. Of the bio-

BAS estimates were then transformed to years by multiplying them by the standard deviation of CA and summing with mean CA, as shown in equation (8), for men, and equation (9) for women:

$$BA = (BAS \times 14.18) + 47.15 \quad (8)$$

$$BA = (BAS \times 13.92) + 47.75. \quad (9)$$

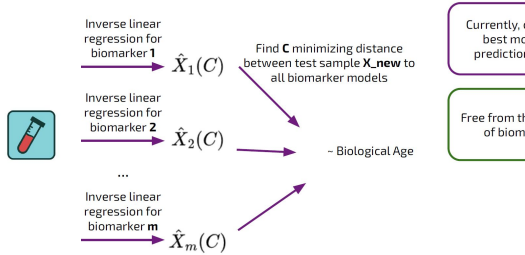
Finally, true BA (TBA) was calculated by adding z scores, calculated as $z = (y_i - \hat{y}) \times (1 - b)$, to BA values, where y_i is the individual's CA for the group, $\hat{y}$ is mean CA and b is the coefficient of BA regressed on CA.

Multiple linear regression

Two BA scores were calculated using sex-stratified MLR. The first incorporated all 10 biomarkers, and the second used only those selected by the PCA method. The results from the equations were then standardized so that the mean BA of participants of a given age was equal to CA.

Klemera and Doubal's method

Klemera-Doubal method [3] (Semi-supervised):



3. Klemera P., Doubal S. A new approach to the concept and computation of biological age // Mechanisms of ageing and development. – 2006. – Т. 127. – №. 3. – С. 240-248.

$$\mathrm{BA}_E = \frac{\sum_{j=1}^{m}\left(x_j - q_j\right)\left(\frac{k_j}{s_j^2}\right)}{\sum_{j=1}^{m}\left(\frac{k_j}{s_j}\right)^2} \qquad (2)$$

$$r_{char} = \frac{\sum_{j=1}^{m}\frac{r_j^2}{\sqrt{1-r_j^2}}}{\sum_{j=1}^{m}\frac{r_j}{\sqrt{1-r_j^2}}}$$

$$\mathrm{BA}_{EC} = \frac{\sum_{j=1}^{m}\left(x_j - q_j\right)\frac{k_j}{s_j^2} + \frac{\mathrm{CA}}{s_{\mathrm{BA}}^2}}{\sum_{j=1}^{m}\left(\frac{k_j}{s_j}\right)^2 + \frac{1}{s_{\mathrm{BA}}^2}}. \qquad (3)$$

$$s_{\mathrm{BA}}^2 = \left(\frac{\sum_{j=1}^{n}\left(\left(\mathrm{BA}_{Ei} - \mathrm{CA}_i\right) - \sum_{i=1}^{n}\left(\mathrm{BA}_{Ei} - \mathrm{CA}_i\right)/n\right)^2}{n}\right) - \left(\frac{1-r_{char}^2}{r_{char}^2}\right) \times \left(\frac{\left(\mathrm{CA}_{\max} - \mathrm{CA}_{\min}\right)^2}{12m}\right). \qquad (5)$$

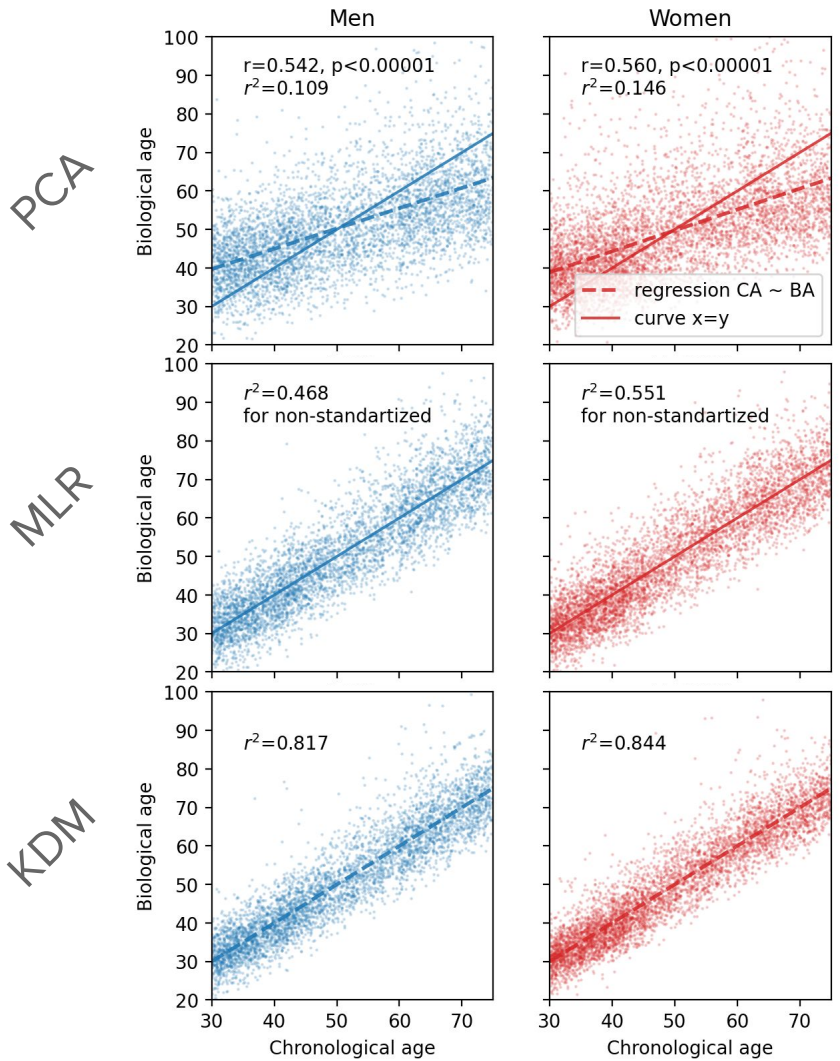
where x_j is a set of biomarker values, k_j is a slope, q_j is intercept

s_j i a root mean squared error of a biomarker regressed on chronological age

Comparison

	Mean (std)	Minimum	Maximum
Chronological Age	49.96, (13.34)	30.083333	74.916667
Principal Component Analysis	49.96, (12.95)	18.199931	211.367337
Multiple Linear Regression	49.96, (14.90)	14.347522	103.457585
Klemera and Doubal's method	49.96, (14.42)	17.852403	106.542431

Klemera and Doubal's method (KDM) gives the highest R^2 between biological and chronological ages



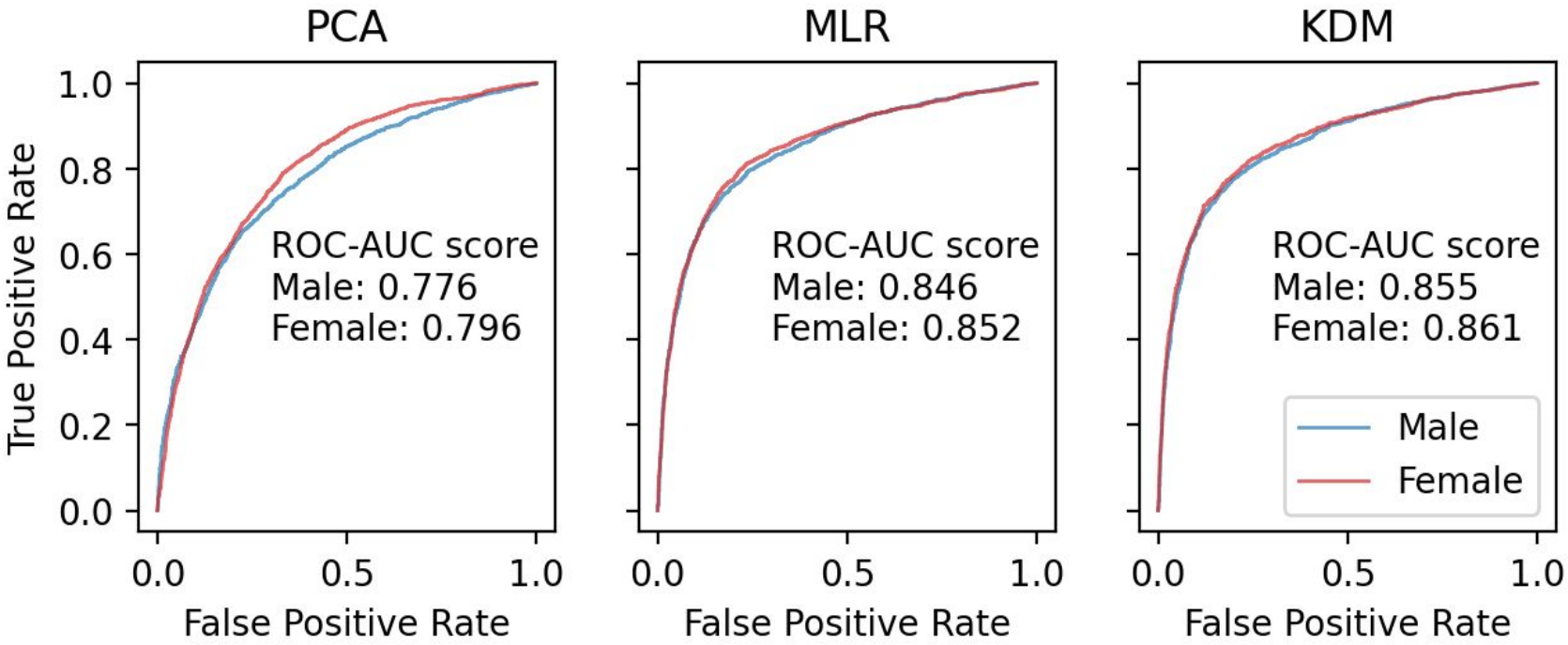
Biological Age Estimates and mortality

Stratify sample into 30–59 years old and 60–75 years old groups
For each group run Cox proportional hazard models

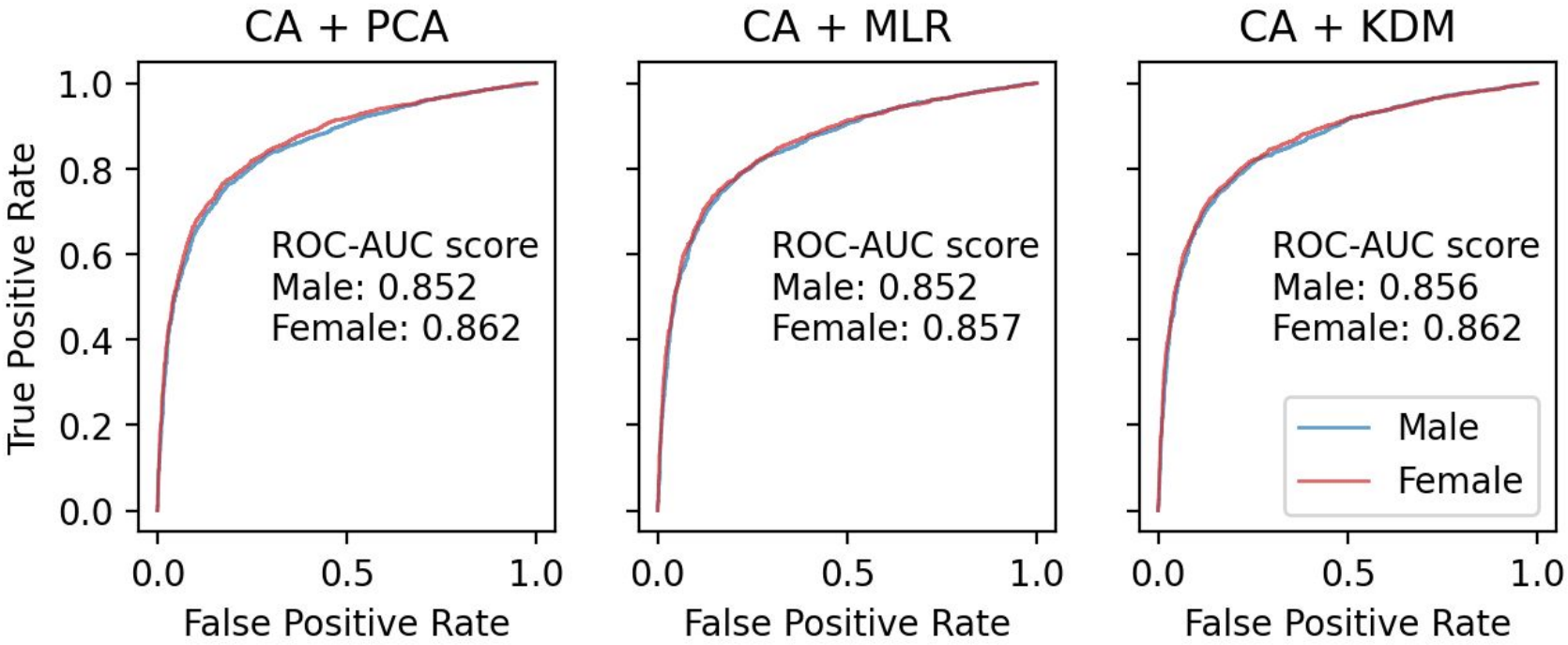
1. Chronological Age (CA)
2. Chronological Age + PCA Biological Age
3. Chronological Age + MLR Biological Age
4. Chronological Age + KDM Biological Age

All analyses were run controlling for sex

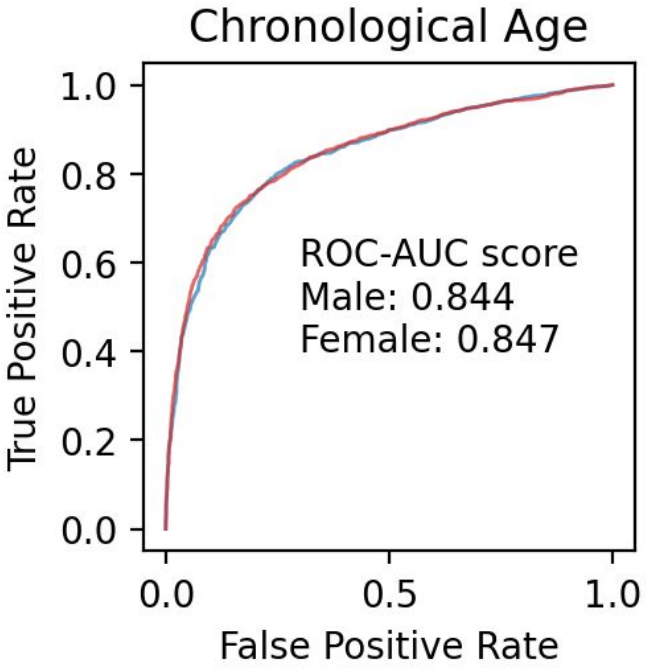
ROC-AUC for cumulative hazard of Biological ages

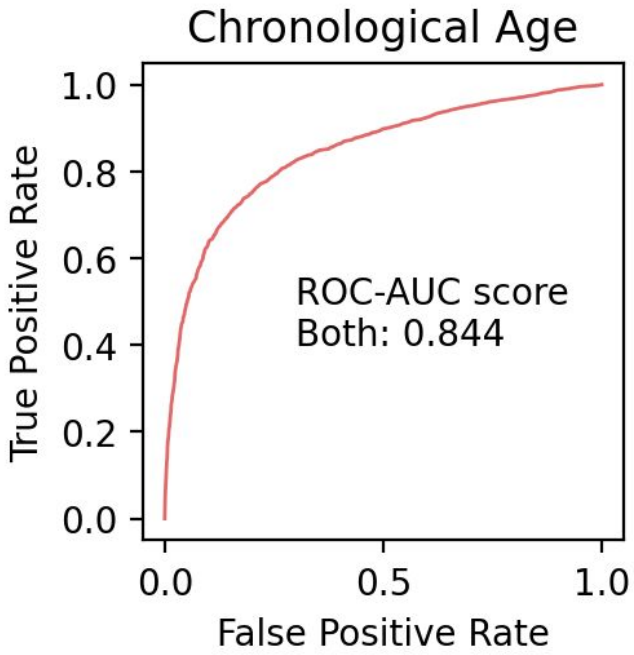


ROC-AUC for cumulative hazard of Biological and Chronological ages (CA)



ROC-AUC for cumulative hazard of Chronological age





ROC-AUC for cumulative hazard Stratified by age

	Full sample	Aged 30–59 years	Aged 60–75 years
Chronological Age	0.848	0.719	0.703
Principal Component Analysis	0.859	0.750	0.726
Multiple linear regression	0.857	0.740	0.721
Klemera and Doubal's method	0.861	0.751	0.729

Models predict better when full sample available

Biological age helps to improve predictions

Hazard Ratios

	CA HR (95% CI)	CA SE	BA HR (95% CI)	BA SE
CA + PCA	1.08 (1.08-1.08)***	0.002	1.03 (1.02-1.03)***	0.001
CA + MLR	1.04 (1.04-1.05)***	0.003	1.05 (1.05-1.05)***	0.002
CA + KDM	1.03 (1.02-1.03)***	0.003	1.07 (1.06-1.07)***	0.002

The more complex the model - the lower influence of Chronological Age (CA)

Conclusions

1. Biological age increases accuracy of predictions of mortality risks *although I expected more significant effects…*
2. Gender differences play a big part in survival model predictions
3. Klemera and Doubal's method gives the most accurate prediction of biological ages compared with Principal Component analysis and Multiple linear regression

Thank you for attention!

Questions?