

Biological aging markers and polygenic risk scores for mortality prediction: a multicohort study

Shayan Mostafaei

shayan.mostafaei@ki.se

Karolinska Institutet <https://orcid.org/0000-0002-1966-1306>

Jakob Lindh

Karolinska Institutet

Chenxi Qin

Karolinska Institutet

Jonathan K. L. Mak

Karolinska Institutet

Sara Hägg

Karolinska Institutet <https://orcid.org/0000-0002-2452-1500>

Research Article

Keywords: All-cause mortality, Biological aging measures, Polygenic risk score, Risk prediction, Cohort study

Posted Date: May 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9600666/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Chronological age is a strong predictor of mortality but does not fully capture heterogeneity in physiological decline. We evaluated whether clinically accessible biological aging (BA) measures and an aging-related polygenic risk score improve all-cause mortality prediction beyond chronological age in two independent cohorts: Swedish TwinGene (n = 9,617; median follow-up 16.70 years) and UK Biobank (n = 179,504; median follow-up 11.83 years). We studied three biomarker-based biological age estimates (PhenoAge, Klemmera–Doubal method, and homeostatic dysregulation), a frailty index, leukocyte telomere length, and multivariate aging polygenic risk scores. To focus on age-independent biological age estimates, we used age-adjusted residuals of biomarker-based biological age estimates in discrimination and prediction analyses. We assessed discrimination using receiver operating characteristic analyses and evaluated multivariable prediction using cross-validated ensemble models. Time-to-event associations were estimated using Cox proportional hazards models.

Chronological age showed strong univariate discrimination, with area under the ROC curve (AUC) 0.837 in TwinGene and 0.708 in UK Biobank. Among biological aging measures, PhenoAge residual had the highest discrimination (AUC: 0.874 in TwinGene; AUC: 0.624 in UK Biobank), whereas the polygenic risk score showed near-null discrimination (approximately 0.50 in both cohorts). In cross-validated ensemble prediction, adding biological aging measures to chronological age and covariates substantially improved discrimination in TwinGene (AUC: 0.936) and modestly improved discrimination in UK Biobank (AUC: 0.762), while adding the polygenic risk score without biological aging measures produced minimal change. In multivariable Cox models, PhenoAge residual remained independently associated with mortality in both cohorts, whereas the polygenic risk score was not.

Clinically biological aging measures, particularly PhenoAge, improve mortality prediction beyond chronological age, while the evaluated aging polygenic risk score adds little incremental predictive value. Cohort differences highlight the importance of evaluating transportability across populations, biomarker panels, and risk horizons.

Introduction

Accurate predicting all-cause mortality is a cornerstone of preventive medicine and public health. While chronological age (CA) is a strong predictor of mortality, it does not fully capture the heterogeneity in rates of physiological decline that shape individual health trajectories. Recent advances in aging research have therefore focused on biological aging (BA) measures that summarize multisystem dysregulation using clinically accessible data. Here we evaluate widely used and interpretable approaches including biomarker-based biological age estimates (Levine PhenoAge, Klemmera–Doubal method [KDM], and homeostatic dysregulation [HD]), along with leukocyte telomere length (TL) as a molecular marker. We also include a frailty index (FI), which is not a biomarker-derived “biological age” estimate but a validated clinical measure reflecting accumulation of age-related health deficits; we treat FI as a complementary aging-related construct to avoid conceptual confusion (1–4). These measures

have shown associations with morbidity and mortality in prior work (5–7), yet their comparative predictive performance and generalizability across cohorts remain incompletely characterized. (8).

A major barrier to comparing BA measures across settings is that algorithm inputs are not uniform across cohorts (9). Biomarker-based measures such as PhenoAge/KDM/HD may rely on different clinical panels depending on what is measured, raising concerns about harmonization, comparability, and transportability when the number and type of biomarkers differ. Because biomarker-based biological age estimates are strongly correlated with chronological age, we primarily used age-adjusted residual (“age-gap”) versions of these measures when models included chronological age (10). In addition, clinical relevance requires more than discrimination alone; calibration, reclassification, and decision-analytic measures help determine whether a model is reliable and potentially actionable in practice. Accordingly, we frame mortality prediction as both a performance and transportability problem—how well models work within a cohort and how well they generalize under distribution shifts between cohorts.

In parallel, polygenic risk scores (PRS) offer a way to quantify inherited susceptibility to aging-related outcomes (11), and could potentially complement BA measures (12). However, PRS performance for mortality prediction and its incremental value beyond chronological age and clinically accessible BA measures remain uncertain. The aim of this study was to systematically compare multiple clinically feasible BA measures and an aging-related PRS—individually and in combination with chronological age—using a discovery/validation multicohort framework. We evaluated and compared prediction performance for all-cause mortality in Swedish TwinGene (discovery) and UK Biobank (external validation), including performance across prespecified age and follow-up subgroups and key effect modifiers. Where biomarker availability allowed, we also considered including contemporary comparator measures (e.g., GOLD BioAge); however, key inputs required for some newer biological age algorithms were not consistently available across both cohorts, precluding harmonized computation and head-to-head comparison. We aimed to enhance transparency and reproducibility through versioned code and clearly specified evaluation procedures.

Results

Cohort characteristics and follow-up

The analytic samples included 9,617 participants in TwinGene and 179,504 in UKB (Fig. 1). Participants in UKB were younger at baseline (56.66 ± 8.00 years) than those in TwinGene (64.80 ± 8.01 years), with a similar proportion of men (UKB 46.9%; TwinGene 47.3%). Follow-up was shorter in UKB (median 11.83 years; IQR 11.16–12.49) than in TwinGene (median 16.70 years; IQR 15.60–17.80). Body mass index (BMI) was higher in the UKB (27.28 ± 4.66 kg/m²) than TwinGene (24.96 ± 3.28 kg/m²), and the proportion of current or former smokers was comparable (UKB 44.6%; TwinGene 43.0%). High/intermediate education was more common in UKB (80.2%) than TwinGene (56.6%). Consistent with the younger UKB baseline age and cohort differences, biomarker-based BA measures were lower on average in UKB (PhenoAge 47.71 ± 9.96 ; HD 2.23 ± 0.98) than TwinGene (PhenoAge 56.26 ± 9.05 ; HD

2.71 ± 0.99), while KDM was similar across cohorts (UKB 54.18 ± 9.36; TwinGene 55.82 ± 8.14). FI means were comparable (UKB 0.11 ± 0.07; TwinGene 0.12 ± 0.09), whereas adjusted telomere length was lower in UKB (0.83 ± 0.13) than TwinGene (1.05 ± 0.37). During follow-up, 3,090 deaths (32.1%) occurred in TwinGene and 14,492 deaths (8.1%) UKB (Table 1).

Table 1

Baseline characteristics and aging-related measures in the TwinGene and UK Biobank analytic cohorts

Characteristics	TwinGene	UK Biobank
No. of participants	9,617	179,504
No. of death	3,090	14,492
Age, years (mean ± SD)	64.80 ± 8.01	56.66 ± 8.00
Sex, men (n, %)	4,548 (47.3%)	84,202 (46.9%)
Follow up time, years (median; IQR)	16.70; 15.60–17.80	11.83; 11.16–12.49
BMI, kg/m ² (mean ± SD)	24.96 ± 3.28	27.28 ± 4.66
Alcohol consumption frequency, weekly (n, %)	Not available	112,549 (67.2%)
Smokers (current or former, n, %)	4145 (43%)	80,026 (44.6%)
Education level, high or intermediate (n, %)	5443 (56.6%)	143,962 (80.2%)
PhenoAge, years (mean ± SD)	56.26 ± 9.05	47.71 ± 9.96
KDM, years (mean ± SD)	55.82 ± 8.14	54.18 ± 9.36
HD (mean ± SD)	2.71 ± 0.99	2.23 ± 0.98
FI (mean ± SD)	0.12 ± 0.09	0.11 ± 0.07
Telomere length (mean ± SD)	1.05 ± 0.37	0.83 ± 0.13
PRS (mean ± SD)	0.21 ± 0.04	0.13 ± 0.11

Values are reported as mean ± SD, median (IQR), or n (%), as appropriate. The table summarizes the analytic samples included in the main integrated analyses (TwinGene n = 9,617; UK Biobank n = 179,504). Follow-up time is calculated from baseline assessment to death or administrative censoring. High/intermediate education is defined as ≥ 10 years of formal education. Smoking status includes current and former smokers. Alcohol consumption frequency is a self-reported UK Biobank variable and was not available in TwinGene. Biological aging measures include Phenotypic Age (PhenoAge), Klemm–Douglas Method (KDM), and Homeostatic Dysregulation (HD). Frailty Index (FI) is the Rockwood deficit-accumulation measure (reported here on its original scale as a proportion). Telomere length is the cohort-specific adjusted leukocyte telomere length measure. PRS values are shown as raw cohort-specific scores in this table; PRS were standardized within cohort (mean = 0, SD = 1) prior to regression and prediction analyses.

Correlations among aging measures

Across TwinGene (Fig. 2A) and UKB (Fig. 2B), correlations among aging measures ranged from weak to strong, with the strongest intercorrelations observed among the biomarker-based BA measures. CA correlated positively with PhenoAge and KDM in both cohorts, with a stronger CA–PhenoAge association in TwinGene. PhenoAge showed the strongest correlations with other BA measures, particularly with KDM and HD, and these PhenoAge correlations were higher in UKB (e.g., PhenoAge–KDM and PhenoAge–HD). PRS showed weak correlations with the non-genetic aging measures (including FI), despite mvAge including frailty-related GWAS components, consistent with FI reflecting a largely environmental and cohort- and measurement-dependent deficit accumulation phenotype. Telomere length showed weak negative correlations with chronological age and most BA measures, with slightly stronger negative associations in UKB. Overall, the correlation structure was similar across cohorts, with modest cohort differences in magnitude.

Univariate discrimination of predictors

Because ROC analyses treat mortality as a binary outcome and ignore censoring/time-to-event, we present these AUCs as a secondary comparison and prioritize survival models for time-to-death analyses. To reduce collinearity with CA and to focus on age-independent variation in biomarker-based biological age estimates, age-adjusted residuals for PhenoAge, KDM, and HD in all discrimination and prediction analyses. Specifically, within each cohort we fit a linear regression of each BA measure on CA at baseline (e.g., $\text{PhenoAge} \sim \text{CA}$, $\text{KDM} \sim \text{CA}$, $\text{HD} \sim \text{CA}$) and used the residuals as the predictor. According to univariate ROC analyses using age-adjusted residuals for the biomarker-based biological age estimates (Fig. 3), CA achieved an AUC of 0.837 (95% CI: 0.828–0.846) in TwinGene, significantly higher than the 0.708 (95% CI: 0.703–0.712) observed in UKB (DeLong $P = 5.23 \times 10^{-147}$). Among BA measures, PhenoAge_res showed the highest discrimination in TwinGene (AUC = 0.874, 95% CI: 0.866–0.883) but was lower in UKB (AUC = 0.624, 95% CI: 0.619–0.629; DeLong $P < 1 \times 10^{-300}$). KDM_res also performed better in TwinGene (AUC = 0.747, 95% CI: 0.736–0.758) than in UKB (AUC = 0.600, 95% CI: 0.595–0.605; DeLong $P = 7.98 \times 10^{-121}$), while HD_res showed modest discrimination in both cohorts (TwinGene AUC = 0.592, 95% CI: 0.580–0.605; UKB AUC = 0.565, 95% CI: 0.560–0.570; DeLong $P = 6.44 \times 10^{-5}$). FI had modest, similar performance (TwinGene AUC = 0.574, 95% CI: 0.562–0.587; UKB AUC = 0.587, 95% CI: 0.582–0.592; DeLong $P = 5.81 \times 10^{-2}$). Telomere length and PRS showed limited discrimination, with TL AUC 0.581 in TwinGene and 0.521 in UKB, and PRS AUC 0.510 and 0.503, respectively. Overall, CA remained the strongest single predictor in UKB, whereas PhenoAge_res provided the strongest univariate discrimination among BA measures in TwinGene (Table 2).

Table 2

Univariate predictive performance (AUCs) of aging and genetic measures for all-cause mortality in TwinGene and UK Biobank

Predictor	TwinGene AUC (95% CI)	UK Biobank AUC (95% CI)
Chronological age (CA)	0.837 (0.828–0.846)	0.708 (0.703–0.712)
PhenoAge residual (PhenoAge_res)	0.874 (0.866–0.883)	0.624 (0.619–0.629)
KDM residual (KDM_res)	0.747 (0.736–0.758)	0.600 (0.595–0.605)
HD residual (HD_res)	0.592 (0.580–0.605)	0.565 (0.560–0.570)
Frailty Index (FI)	0.574 (0.562–0.587)	0.587 (0.582–0.592)
Telomere Length	0.581 (0.569–0.593)	0.521 (0.516–0.526)
Polygenic Risk Scores (PRS_Z)	0.510 (0.498–0.522)	0.503 (0.499–0.508)
P-values (Others vs. CA)	$< 1 \times 10^{-9}$	$< 1 \times 10^{-5}$
AUCs (95% CIs) are from univariate ROC analyses for all-cause mortality in TwinGene (n = 9,617) and UK Biobank (n = 179,504). KDM: Klemmer-Doubal Method, HD: Homeostatic Dysregulation Index. Residualized biological age estimates (PhenoAge_res, KDM_res, HD_res) were computed within each cohort as residuals from linear regressions on chronological age at baseline (BA ~ CA). PRS was standardized within cohort (PRS_Z). Confidence intervals were computed using DeLong's method. P-values are from DeLong's test comparing the AUC of each BA measure versus chronological age within each cohort.		

Multivariable and ensemble prediction models

The predictive performance of multivariable models was evaluated in TwinGene and UKB using 10-fold cross-validated SuperLearner models based on out-of-fold predictions (Table 3; Fig. 4). In TwinGene, discrimination increased from CA alone (AUC 0.837, 95% CI 0.828–0.846) to CA + covariates (AUC 0.848, 95% CI 0.840–0.857), while adding PRS provided no meaningful improvement (AUC 0.848, 95% CI 0.841–0.857). Adding BA measures (PhenoAge residual, FI, and telomere length) yielded the best performance (AUC 0.936, 95% CI 0.931–0.942; paired DeLong $P = 1.50 \times 10^{-153}$ vs CA). In UKB, AUC similarly increased from 0.707 (95% CI 0.703–0.711) for CA to 0.744 (95% CI 0.740–0.748) with covariates, remained unchanged after adding PRS (AUC 0.744, 95% CI 0.740–0.748), and improved further with BA measures (AUC 0.762, 95% CI 0.758–0.766; paired DeLong $P = 4.81 \times 10^{-252}$ vs CA). Calibration of the full model showed generally good agreement between predicted and observed mortality risk within each cohort, with UKB predictions concentrated in the low-risk range consistent with its lower event rate and shorter follow-up (Fig. 4).

Table 3

Multivariable predictive accuracy for all-cause mortality using repeated 10-fold cross-validation in TwinGene and UK Biobank cohorts

Predictive Model	TwinGene AUC (95% CI)	UK Biobank AUC (95% CI)
CA	0.837 (0.828–0.846)	0.707 (0.703–0.711)
CA + Covariates	0.848 (0.840–0.857)	0.744 (0.740–0.748)
CA+ Covariates + PRS	0.848 (0.841–0.857)	0.744 (0.740–0.748)
Full model: CA+ Covariates + PRS+BA measures	0.936 (0.931–0.942)	0.762 (0.758–0.766)
Full model (repeated CV; mean $\pm$ SD across repeats)	0.936 $\pm$ 0.020	0.762 $\pm$ 0.015
P-values (Full model vs CA)	1.50 $\times$ 10 ⁻¹⁵³	4.81 $\times$ 10 ⁻²⁵²
Area under the receiver operating characteristic curve (AUC) with 95% confidence intervals (CIs) was estimated from 10-fold cross-validated out-of-fold predictions generated using CV.SuperLearner (binomial family) within each cohort. The full model included chronological age (CA), standardized polygenic risk score (PRS), age-adjusted PhenoAge residual, frailty index (FI), and telomere length, plus cohort-available covariates (TwinGene: sex, body mass index [BMI], smoking status, education years, and top 10 genetic principal components; UK Biobank: sex, BMI, smoking status, alcohol intake, education, and top 10 genetic principal components). Model comparisons versus CA used paired DeLong tests within cohort based on the out-of-fold predictions. To assess stability of the full model, the 10-fold cross-validation (CV) procedure was repeated 10 times using different random fold assignments, and the mean $\pm$ SD AUC across repeats is reported. <i>P</i>-values refer to the comparison of AUC between the full model and the CA-only model within each cohort.		

Time-to-event associations were evaluated using Cox proportional hazards models in TwinGene and UKB, using time-since-baseline as the primary timescale and then age-as-timescale sensitivity analyses. In univariate time-since-baseline models, CA was strongly associated with mortality (TwinGene HR = 1.150, 95% CI 1.144–1.155; UKB HR = 1.102, 95% CI 1.100–1.105). Age-adjusted biological age estimates were also positively associated with mortality, including PhenoAge residual (TwinGene HR = 1.305, 95% CI 1.297–1.312; UKB HR = 1.077, 95% CI 1.075–1.080), KDM residual (TwinGene HR = 1.156, 95% CI 1.151–1.161; UKB HR = 1.070, 95% CI 1.067–1.074), and HD residual (TwinGene HR = 1.300, 95% CI 1.267–1.334; UKB HR = 1.204, 95% CI 1.189–1.220). Frailty index predicted higher risk (TwinGene HR = 1.329, 95% CI 1.263–1.398; UKB HR = 2.990, 95% CI 2.840–3.148), while longer telomere length was inversely associated with mortality (TwinGene HR = 0.662, 95% CI 0.617–0.711; UKB HR = 0.179, 95% CI 0.156–0.204). PRS was not significantly associated with mortality in either cohort (TwinGene HR = 0.996, 95% CI 0.962–1.032; UKB HR = 0.999, 95% CI 0.993–1.026).

In multivariable time-since-baseline models adjusting for cohort-available covariates (TwinGene: sex, BMI, smoking, education; UKB: sex, BMI, smoking, alcohol, education) and including CA, PhenoAge residual remained independently associated with mortality in both cohorts (TwinGene HR = 1.263, 95% CI 1.255–1.271; UKB HR = 1.059, 95% CI 1.056–1.062), whereas PRS remained null (TwinGene HR = 0.991,

95% CI 0.957–1.026; UKB HR = 1.008, 95% CI 0.991–1.024). Frailty index and telomere length attenuated substantially in TwinGene after adjustment (Frailty Index HR = 1.012, 95% CI 0.962–1.065; Telomere Length HR = 0.975, 95% CI 0.878–1.083) but remained associated in UKB (Frailty Index HR = 1.739, 95% CI 1.646–1.838; Telomere Length HR = 0.684, 95% CI 0.597–0.782). Discrimination was high in TwinGene and moderate in UKB (C-index 0.913 [bootstrap 95% CI 0.908–0.917] and 0.749 [0.744–0.752], respectively) (Fig. 5A). Proportional hazards diagnostics indicated departures from proportionality (global tests $P < 2 \times 10^{-16}$ in both cohorts), so sensitivity analyses were performed: excluding early deaths in UKB (first 2 years) yielded very similar estimates (Frailty Index HR = 1.694; Telomere Length HR = 0.652; concordance 0.748). Using age as the timescale (left-truncated start–stop Cox; CA not included as a covariate) supported the same overall conclusions: PhenoAge residual remained strongly associated with mortality (TwinGene univariate HR = 1.148; multivariable HR = 1.147; UKB univariate HR = 1.080; multivariable HR = 1.063), PRS showed at most weak evidence (TwinGene univariate HR = 0.962, $P = 0.034$ but multivariable HR = 0.972, $P = 0.277$; UKB univariate HR = 1.002, $P = 0.157$ and multivariable HR = 1.014, $P = 0.098$), and frailty index remained associated (TwinGene univariate HR = 1.183; UKB multivariable HR = 1.676). In these age-timescale multivariable models, telomere length attenuated toward the null (TwinGene HR = 1.109, $P = 0.139$; UKB HR = 0.877, $P = 0.054$), and discrimination decreased as expected when age is absorbed into the baseline hazard (C-index 0.841 in TwinGene; 0.657 in UKB) (Fig. 5B).

Subgroup and interaction analyses

The comparative analysis of predictor performance for mortality across varying age groups and follow-up periods in the TwinGene and UKB cohorts showed a consistent ranking of predictors across age strata (< 50, 50–60, > 60 years), with PhenoAge showing the highest discrimination at each follow-up horizon (TwinGene AUC 0.978, 0.979, and 0.953 at 5, 10, and 15 years; UKB AUC 0.651, 0.632, and 0.544, respectively). KDM was the next-best discriminator, particularly in TwinGene (AUC 0.922, 0.885, 0.814), whereas PRS showed minimal predictive capacity in both cohorts (AUC ~ 0.50 across horizons). The predictive utility of HD, FI, and telomere length was more modest, with generally lower AUCs than PhenoAge and KDM. Predictor performance declined with longer follow-up in UKB across most measures (e.g., PhenoAge 0.651→0.544; KDM 0.617→0.558), due to the median length of follow-up being less than 12 years (**Fig S1**).

In the TwinGene, statistical analysis testing BA-by-BMI interactions showed statistically significant interactions for PhenoAge ($P = 0.0007$) and HD ($P < 0.001$) with BMI, whereas interactions for KDM, FI, telomere length, and PRS were not significant. In contrast, in UKB, BMI interactions were significant for PhenoAge, KDM, HD, and FI (all $P < 0.001$) but not for PRS ($P = 0.086$) or telomere length ($P = 0.301$) (**Fig S2**).

Sex-differences were observed for PhenoAge, KDM, and HD in mortality prediction in the TwinGene (PhenoAge $P = 0.011$, KDM $P = 0.041$, HD $P = 0.029$), while FI, telomere length, and PRS were not

significant. In UKB, sex interactions were significant for PhenoAge, KDM, HD, and FI (all $P < 0.001$) and for telomere length ($P = 0.020$), but not for PRS ($P = 0.635$) (**Fig S2**).

There was a significant effect of smoking status on mortality prediction for several BA measures in both cohorts. In the TwinGene, only PhenoAge showed evidence of interaction with smoking ($P = 0.003$), whereas KDM, HD, FI, telomere length, and PRS were not significant. In UKB, smoking interactions were significant for PhenoAge, KDM, HD, and FI (all $P < 0.001$) and telomere length ($P = 0.030$), but not PRS ($P = 0.764$) (**Fig S2**).

In the TwinGene, significant interactions were found between education years and PRS ($P = 0.035$), PhenoAge ($P = 0.022$), and telomere length ($P = 0.035$), whereas KDM, HD, and FI were not significant. In UKB, education interactions were observed for PRS ($P = 0.046$), PhenoAge ($P = 0.018$), and telomere length ($P = 0.041$), while KDM ($P = 0.101$), HD ($P = 0.193$), and FI ($P = 0.098$) did not reach significance (**Fig S2**).

Discussion

This multicohort study demonstrates that clinically accessible biological aging (BA) measures capture meaningful mortality risk information beyond chronological age (5, 13), and that a biomarker-based biological age estimate (PhenoAge) is consistently the strongest individual BA predictor across cohorts (1, 5). Our results reinforce that CA remains an essential benchmark, but also show that BA metrics capture complementary physiological information beyond CA, supporting their use in enhanced risk stratification frameworks (1, 5, 13). Importantly, our findings also underscore the limited incremental value of the specific longevity/healthspan PRS evaluated here for near-term mortality discrimination in these cohorts, emphasizing that genetic predisposition alone is insufficient to represent the dynamic, multi-system processes that shape mortality risk (11, 14–16).

PhenoAge consistently outperformed other BA markers in both cohorts, aligning with prior work showing that composite clinical-aging measures integrating multi-organ physiology often provide the strongest mortality signal (1, 5, 13). In TwinGene, PhenoAge achieved the highest discrimination among all evaluated predictors and remained strong across follow-up horizons, indicating that it captures system-level dysregulation relevant to both short- and longer-term mortality risk (1, 5, 17). In UKB, CA showed the highest univariate AUC, which is expected given the broader baseline age range and strong age-gradient in mortality risk (18), yet PhenoAge remained the strongest BA measure and contributed to improved discrimination when combined with other predictors. This clarifies that “PhenoAge is strongest” refers to the strongest BA measure, whereas “CA is strongest in UKB” refers to the strongest overall univariate predictor in that cohort.

Across BA measures, KDM showed the next strongest performance (especially in TwinGene), while HD, FI, and telomere length showed more modest and cohort-variable discrimination, consistent with the idea that these measures reflect different aging domains (e.g., homeostatic deviation, deficit accumulation, and cellular replicative biology) that do not always map linearly onto mortality risk in the

same way as composite clinical-aging indices (2, 17, 19, 20). In contrast, the mvAge-based PRS showed minimal discriminative power across analyses. This likely reflects a combination of (i) modest effect sizes of common variants, (ii) the distal relationship between inherited risk and near-term mortality events, and (iii) the reality that downstream exposures and physiology may dominate mortality risk trajectories, particularly in later life (11, 15, 16).

In multivariable prediction, integrating CA with BA measures and PRS improved model performance most clearly in TwinGene, supporting the concept that combining complementary biological signals can yield better discrimination than any single domain alone (21, 22). The comparatively smaller incremental gain in UKB is consistent with distribution shift (age structure, healthy volunteer bias, and follow-up structure) and the strong baseline predictive value of CA in that cohort (18, 23). These findings support the use of ensemble approaches (e.g., stacking) to integrate heterogeneous predictors, while also underscoring that the magnitude of benefit from BA integration is cohort-dependent and should be evaluated in the target population prior to clinical translation (24–26).

Sensitivity analyses across baseline age strata and fixed follow-up horizons (5/10/15 years) showed that the relative ranking of predictors was stable: PhenoAge remained strongest among BA measures, PRS remained weakest, and discrimination for several predictors attenuated with longer horizons, particularly in UKB, consistent with reduced effective sample size at longer follow-up and the cohort's shorter median follow-up (27). These results emphasize that prediction performance depends on the intended risk horizon, and that reporting horizon-specific discrimination improves interpretability for both clinical and public-health contexts (26, 27).

We also observed systematic effect modification by lifestyle and sociodemographic factors, with multiple BA-by-modifier interactions reaching statistical significance (1, 5, 13). In UKB, interaction evidence was especially strong for BA measures with BMI, sex, and smoking, while PRS and telomere length generally showed weaker and less consistent interaction signals. In TwinGene, interaction evidence was more selective. These findings support the conceptualization of BA measures as “exposure-sensitive” indicators, reflecting the cumulative physiological imprint of modifiable risk factors. At the same time, we interpret interaction *P*-values as primarily inferential/etiologic signals rather than direct evidence of clinically actionable subgroup thresholds and emphasize that future work should quantify interaction effect sizes and clinical utility metrics (e.g., risk reclassification and decision-curve analysis) before implementation (21, 24).

Key strengths include the two-cohort design (discovery + independent replication), large sample size (especially in UKB), long mortality follow-up in TwinGene, head-to-head comparison of multiple BA constructs and PRS, and complementary evaluation using both discrimination and time-to-event modeling (26). Limitations include restriction to European ancestry (28), potential cohort differences in biomarker availability and measurement, and the observational nature of the analysis (no causal inference) (24, 25). In addition, we could not benchmark against some newer clinically oriented biological-age algorithms (e.g., GOLD BioAge) because required biomarkers (e.g., albumin, alkaline

phosphatase, γ -glutamyl transferase, and complete blood count indices) were not available consistently across cohorts, preventing harmonized computation and comparison. Additionally, prediction performance may not transfer directly to settings with different baseline risk, lab platforms, or follow-up structure, highlighting the need for external validation in more diverse populations and clinical contexts.

In conclusion, clinically accessible BA measures, particularly the biomarker-based biological age estimate PhenoAge, provide strong and reproducible information for mortality prediction beyond chronological age, whereas the evaluated aging-related polygenic risk score adds little incremental value. These findings support continued development and validation of practical BA measures for risk stratification, with emphasis on transportability, harmonized computation, and clinical utility evaluation.

Materials and Methods

Study population

In this multicohort study, the Swedish TwinGene cohort, a sub-cohort derived from the Swedish Twin Registry, served as the discovery cohort. TwinGene is a population-based study that collected blood samples from 12,648 older Swedish twins between 2004 and 2008 (mean age 64.8 years [SD 7.9], 53% women), with a median follow-up of 16.5 years (IQR 15.2–17.3). For the present analyses, we used complete-case datasets defined separately for (i) biomarker-based aging measures, (ii) genetic data used for PRS, and (iii) the fully integrated set used for head-to-head model comparisons (see “Missingness and analysis sets”).

The UK Biobank (UKB) was used as the replication cohort. UKB is a population-based longitudinal study including ~500,000 individuals aged 37 to 73 years (mean age 56.7 years; 53% women), recruited across the UK between 2006 and 2010, with a median follow-up of 11.83 years (IQR 11.15–12.49). Participants were excluded if they had withdrawn consent (per the most recent UKB withdrawal list), lacked required genotypes for PRS construction, or had missingness in variables required for the relevant analysis set (29, 30). Therefore, the primary analyses were based on complete-case integrated datasets for genetic data and biological aging measures.

Although UKB has a substantially larger sample size, TwinGene was selected as the discovery cohort due to its longer follow-up duration and higher mortality rate, which improves power for time-to-event analyses. The study plan for participant selection in both cohorts is shown in **Fig 1**.

Ethics statement

The TwinGene study was approved by the Regional Ethics Review Board, Stockholm, Sweden and all participants have given their informed consent. The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee. Ethical approval for the study was granted by the Swedish Ethical Review Authority (Dnr. 2022-06634-01). All participants provided written informed consent. We

obtained fully de-identified data. Our study adheres to the tenets of the Declaration of Helsinki. This study followed the STROBE reporting guideline for observational studies.

Discovery cohort (TwinGene)

Genetic data: In TwinGene, genomic DNA from 9,896 participants was genotyped at Uppsala University using the Illumina HumanOmniExpress arrays (31). The sample included all dizygotic (DZ) twins and one twin from each monozygotic (MZ) twin pair to avoid redundancy. Individuals with DNA concentrations <20 ng/μL and a subset of 302 female MZ twin pairs previously included in a genome-wide association study were excluded. After quality control (QC), genotyping was successfully completed for 9,836 individuals. QC measures included the exclusion of SNPs with call rate $\leq 97\%$, minor allele frequency <1%, or Hardy–Weinberg equilibrium deviation ($P \leq 1 \times 10^{-7}$). Samples were excluded if genotyping success rate was <97%, sex discrepancies were detected, heterozygosity was abnormal (>3 SD from the mean), or cryptic relatedness was identified. No genotype imputation was performed in TwinGene; all analyses used directly genotyped SNPs (32).

Biological aging measures: We included 9,617 TwinGene participants with complete data to compute three biological age estimates: Klemmera–Doubal method (KDM) (2), PhenoAge (1), and homeostatic dysregulation (HD) (33). These measures were calculated using eight clinical biomarkers available in TwinGene: serum glucose (log-transformed; mmol/L), HbA1c (log-transformed), creatinine (log-transformed; μmol/L), cystatin C (mg/L), triglycerides (log-transformed; mg/dL), total cholesterol (mg/dL), LDL cholesterol (mg/dL), and HDL cholesterol (mg/dL), using the BioAge R package (34). KDM, PhenoAge, and HD were implemented using established BioAge package procedures; biomarkers were selected based on availability and low missingness ($\leq 20\%$), and correlation with chronological age ($|r| > 0.1$). KDM represents the age at which an individual's physiology corresponds to the average physiology in the reference population; PhenoAge is derived from a mortality score; and HD quantifies physiological deviation using Mahalanobis distance. Age-adjusted residuals of biological age estimates were created within each cohort by regressing each BA estimate on chronological age and retaining the residuals ("PhenoAge residual", "KDM residual", "HD residual"). In addition, the frailty index (FI) was calculated according to the Rockwood deficit accumulation model (19) and validated in TwinGene (35). To harmonize FI scaling across cohorts and improve numerical stability, we used a log-transformed FI. Adjusted telomere length (TL) was also included (36, 37). After harmonizing and integrating QC'ed genetic data with complete BA measures, FI, TL, and covariates, 9,617 TwinGene participants were retained for the primary integrated analyses. The distribution of clinical biomarkers is shown in **Fig S3**.

Outcome definition: The primary outcome was all-cause mortality. Time-to-event was defined as the interval from baseline assessment to death from any cause or censoring at end of follow-up, which concluded on February 1, 2024, in TwinGene. Deaths were identified through linkage to national registers.

Covariates: We included age at baseline, sex, body mass index (BMI), smoking status, education, and the top 10 principal components of genetic data in our analysis to adjust for potential confounding factors in

the TwinGene.

Replication cohort (UK Biobank)

Genetic data: This study used the extensive genetic resources of the UKB, incorporating imputed dosage data from 276,566 unrelated individuals of White British/European ancestry and approximately 96 million genetic variants (38). Quality control included excluding variants with minor allele frequency $<1\%$, imputation info score <0.8 , and deviation from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-10}$). Analyses were restricted to participants passing UKB sample QC and included in the primary integrated dataset after intersecting genotype availability with complete phenotype/covariate requirements (see “Missingness and analysis sets”).

Biological aging: To calculate BA measures, we used 331,699 UKB participants with complete 18-biomarker inputs required by the BA algorithms (as previously described) (6). We computed KDM (2), PhenoAge (1), and HD (33) using the BioAge R package (34). Details on the calculation and interpretation of these measures have been described previously (39). In addition, FI was calculated according to the Rockwood deficit accumulation model (19) and validated for the UKB (40). We used relative leucocyte telomere length (TL), provided as an adjusted T/S ratio correcting for technical parameters in the UKB (41). Age-adjusted residuals (“PhenoAge residual”, “KDM residual”, “HD residual”) were derived within UKB by regressing each measure on chronological age. After integrating BA measures with genetic data and UKB’s covariates, a total of 179,504 UKB participants were included for primary analyses.

Outcome definition: Time-to-event was defined as the interval from baseline assessment to all-cause mortality or censoring at end of follow-up on December 30, 2022, in UKB. Deaths were identified through linked national death registries.

Covariates: UKB’s covariates included baseline age, sex, BMI, alcohol consumption, smoking status, education, and the top 10 genetic principal components. Follow-up length was not entered as a covariate in Cox models; instead, fixed follow-up horizons (5/10/15 years) were used to define subgroup performance analyses.

Polygenic risk scores

We constructed polygenic risk scores (PRS) for aging-related traits using GWAS summary statistics from Rosoff et al. (2023) “mvAge”, a multivariate GWAS meta-analysis combining healthspan, lifespan, extreme longevity, frailty, and epigenetic age acceleration traits (~1.9 million participants of European ancestry) (11). We applied clumping and thresholding (C+T) in PLINK using an LD threshold $r^2=0.3$, a P -value threshold of 1×10^{-5} , and a 250 kb clumping window. Thresholds were chosen to balance SNP inclusion and predictive power; more stringent thresholds ($r^2 = 0.1$, $P < 5 \times 10^{-8}$) resulted in fewer variants and similarly low predictive performance. The PRS was calculated as the weighted sum of risk alleles

using GWAS effect sizes (42, 43). A full list of SNPs and weights is provided in **Table S1**. PRS was standardized within cohort prior to modeling.

Missingness and analysis sets

We defined predictor-specific available-case sets for descriptive/univariate analyses and a primary complete-case integrated set for head-to-head model comparisons.

- **TwinGene:** genotype data were available for 9,896 participants (9,836 after QC). The primary integrated analysis set included participants with complete data for BA measures (PhenoAge/KDM/HD), FI, TL, PRS, covariates, and survival outcome (n=9,617).
- **UK Biobank:** BA measures were computed among participants with complete inputs for the biomarker panel (as above), and the primary integrated set (n=179,504) was defined as the intersection of high-quality genotypes (unrelated European ancestry), complete BA measures, FI, TL, covariates, and survival outcome.

Statistical analysis

Variable scaling and transformations: To ensure comparability across predictors, continuous predictors were standardized to cohort-specific z-scores (mean=0, SD=1) before analysis. Age-adjusted residuals of biomarker-based biological age estimates (PhenoAge, KDM, and HD), leukocyte telomere length, and the polygenic risk scores were standardized. The frailty index was log-transformed and then standardized.

Univariate ROC and survival analyses: We assessed univariate mortality prediction using ROC AUC and Cox proportional hazards models. For survival analyses, hazard ratios (HRs) are reported per 1 SD increase in each standardized predictor, and model discrimination was summarized using Harrell's C-index. Subgroup AUCs were estimated within baseline age strata (<50, 50–60, >60 years) and within fixed follow-up horizons (5, 10, 15 years). For each horizon, the outcome was defined as death occurring within the horizon; participants were censored at the horizon for defining the binary endpoint used in ROC analyses.

Multivariable Cox models and timescale specifications: Multivariable Cox proportional hazards models were fitted in both cohorts including chronological age, PRS, PhenoAge residual, FI, and telomere length (and cohort-available covariates). Two timescale specifications were considered: (1) Primary models: time since baseline was used as the underlying timescale, with chronological age included as a covariate. (2) Sensitivity models: attained age was used as the underlying timescale in a left-truncated start–stop framework; chronological age was therefore not included as a covariate in these age-timescale models. Proportional hazards assumptions were evaluated using Schoenfeld residual diagnostics; when evidence of time-varying effects was observed, sensitivity analyses (including excluding early deaths) were performed.

Multivariable ensemble risk prediction: To integrate predictors into a single mortality prediction model, we used an ensemble approach implemented in CV.SuperLearner with 10-fold cross-validated out-of-fold predictions (44). We compared an age-only model (chronological age) versus a full model including chronological age, covariates, standardized PRS, and biological aging measures (PhenoAge residual, FI, TL). Discrimination was summarized using cross-validated AUC with DeLong 95% CIs, and calibration was assessed using grouped calibration plots and summary calibration statistics.

Interaction and effect-modification analyses: Effect modification by BMI, sex, smoking status, and education was evaluated by fitting regression models that included the main effects of each modifier and predictor and their interaction term, adjusting for cohort covariates. Interaction evidence is summarized as *P*-values for the interaction term (Wald and likelihood ratio tests), visualized as $-\log_{10}(P)$.

Contemporary comparator algorithms: We evaluated the feasibility of computing newer clinically oriented biological-age algorithms (e.g., GOLD BioAge). GOLD BioAge requires biomarkers including albumin, alkaline phosphatase, complete blood count indices, and γ -glutamyl transferase. These inputs were unavailable in the TwinGene analytic dataset and were not available consistently in the UKB analytic dataset used for head-to-head comparisons; therefore, GOLD BioAge was not computed.

Software and reproducibility: Analyses were performed in R, using packages including survival/survminer for survival modeling (45), pROC for ROC curves, ggplot2 for visualization, timeROC for time-dependent ROC where applicable, and SuperLearner for ensemble prediction (46-49). PRS was computed using PLINK (C+T). All analysis code is available on GitHub (<https://github.com/shayanmostafaei/Biological-Aging-and-PRS-for-Mortality-Prediction>).

Declarations

Acknowledgements: This research has been conducted using the UK Biobank Resource under Application Number 22224. We thank participants of the UK Biobank for their contribution to this work. The analysis of UKB genotypes was enabled by resources in project sens2017519 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725. We also acknowledge the Swedish Twin Registry for access to data. The Swedish Twin Registry is managed by Karolinska Institutet.

Author's contributions: Shayan Mostafaei, Jakob Lindh, and Sara Hägg contributed to the conceptualization and methodology of the study. Shayan Mostafaei and Jakob Lindh were responsible for software development, formal analysis, and data curation. Jonathan K. L. Mak and Chenxi Qin contributed to data preparation and investigation. Resources were provided by Sara Hägg. Shayan Mostafaei led the visualization and project administration. The original draft was written by Shayan Mostafaei and Sara Hägg, and all authors (Shayan Mostafaei, Jakob Lindh, Jonathan K. L. Mak, Chenxi Qin, and Sara Hägg) contributed to review and editing. All authors have read and approved the final manuscript.

Conflict of interest: The authors have declared that no competing interests exist.

Data availability: Data from the UK Biobank are available to bona fide researchers upon application via the UK Biobank Access Management System (<https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>). The TwinGene data are part of the Swedish Twin Registry and are not publicly available due to ethical and legal restrictions. Access to TwinGene data requires approval from the Swedish Ethical Review Authority and the Steering Committee of the Swedish Twin Registry. Researchers interested in accessing TwinGene data may contact the registry (str-research@meb.ki.se) for further information on the application process.

Generative AI and AI-assisted tools: We used ChatGPT (OpenAI) for language editing and to improve clarity of the manuscript text, and GitHub Copilot for assistance during R code development. All analyses, results, and interpretations were performed by the authors, who verified the correctness of the code and take full responsibility for the content of the manuscript.

References

1. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S et al (2018) An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 10(4):573
2. Klemmera P, Doubal S (2006) A new approach to the concept and computation of biological age. *Mech Ageing Dev* 127(3):240–248
3. Cohen AA, Milot E, Li Q, Bergeron P, Poirier R, Dusseault-Bélanger F et al (2015) Detection of a novel, integrative aging process suggests complex physiological integration. *PLoS ONE* 10(3):e0116489
4. Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A et al (2015) Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences*. ;112(30):E4104-E10
5. Liu Z, Kuo P-L, Horvath S, Crimmins E, Ferrucci L, Levine M (2018) A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: a cohort study. *PLoS Med* 15(12):e1002718
6. Mak JK, McMurran CE, Kuja-Halkola R, Hall P, Czene K, Jylhävä J et al (2023) Clinical biomarker-based biological aging and risk of cancer in the UK Biobank. *Br J Cancer* 129(1):94–103
7. Ho KM, Morgan DJ, Johnstone M, Edibam C (2023) Biological age is superior to chronological age in predicting hospital mortality of the critically ill. *Intern Emerg Med* 18(7):2019–2028
8. Liu Z, Chen X, Gill TM, Ma C, Crimmins EM, Levine ME (2019) Associations of genetics, behaviors, and life course circumstances with a novel aging and healthspan measure: Evidence from the Health and Retirement Study. *PLoS Med* 16(6):e1002827
9. Hastings WJ, Shalev I, Belsky DW (2019) Comparability of biological aging measures in the National Health and Nutrition Examination Study, 1999–2002. *Psychoneuroendocrinology* 106:171–178

10. Moon S-E, Yoon JW, Bae JH, Joo S, Kim YH, Lee BH et al (2025) Biological Age Estimation From the Age Gap Using Deep Learning Integrating Morbidity and Mortality: Model Development and Validation Study. *J Med Internet Res* 27:e71592
11. Rosoff DB, Mavromatis LA, Bell AS, Wagner J, Jung J, Marioni RE et al (2023) Multivariate genome-wide analysis of aging-related traits identifies novel loci and new drug targets for healthy aging. *Nat aging* 3(8):1020–1035
12. Bafei SEC, Shen C (2023) Biomarkers selection and mathematical modeling in biological age estimation. *npj Aging* 9(1):13
13. Jylhävä J, Pedersen NL, Hägg S (2017) Biological age predictors. *EBioMedicine* 21:29–36
14. Akeju O, Mens MM, Warmerdam R, Dijkema M, van den Biggelaar AH, Franke L et al (2024) Genetic Correlates of Biological Aging and the Influence on Prediction of Mortality. *Journals Gerontol Ser A: Biol Sci Med Sci* 79(4):glae024
15. Torkamani A, Wineinger NE, Topol EJ (2018) The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 19(9):581–590
16. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA et al (2017) 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 101(1):5–22
17. Cohen AA, Milot E, Li Q, Legault V, Fried LP, Ferrucci L (2014) Cross-population validation of statistical distance as a measure of physiological dysregulation during aging. *Exp Gerontol* 57:203–210
18. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T et al (2017) Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 186(9):1026–1034
19. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K (2008) A standard procedure for creating a frailty index. *BMC Geriatr* 8:1–10
20. Gao X, Zhang Y, Mons U, Brenner H (2018) Leukocyte telomere length and epigenetic-based mortality risk score: associations with all-cause mortality among older adults. *Epigenetics* 13(8):846–857
21. Pencina M, D’Agostino Sr R, D’Agostino R Jr, Vasan R Evaluating the added predictive ability of a new marker. from area under the ROC curve to reclassification and beyond. 2008:27
22. Van der Laan MJ, Polley EC, Hubbard AE (2007) Super learner
23. Van Alten S, Domingue BW, Galama T, Marees AT Reweighting the UK Biobank to reflect its underlying sampling population substantially reduces pervasive selection bias due to volunteering. *MedRxiv*. 2022:2022.05. 16.22275048.
24. EW S (2009) Clinical prediction models: a practical approach to development, validation, and updating. Springer, New York
25. Van Calster B, Vickers AJ (2015) Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making* 35(2):162–169

26. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Br Surg* 102(3):148–158
27. Heagerty PJ, Lumley T, Pepe MS (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56(2):337–344
28. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 51(4):584–591
29. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J et al (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 12(3):e1001779
30. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K et al (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562(7726):203–209
31. Ek WE, Reznichenko A, Ripke S, Niesler B, Zucchelli M, Rivera NV et al (2015) Exploring the genetics of irritable bowel syndrome: a GWA study in the general population and replication in multinational case-control cohorts. *Gut* 64(11):1774–1782
32. Liu W, Jiao X, Thutkawkorapin J, Mahdessian H, Lindblom A (2017) Cancer risk susceptibility loci in a Swedish population. *Oncotarget* 8(66):110300
33. Cohen AA, Milot E, Yong J, Seplaki CL, Fülöp T, Bandeen-Roche K et al (2013) A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. *Mech Ageing Dev* 134(3–4):110–117
34. Kwon D, Belsky DW (2021) A toolkit for quantification of biological age from blood chemistry and organ function test data: BioAge. *Geroscience* 43:2795–2808
35. Klein HE (2025) Is frailty index the most accurate predictor of CVD? *The American Journal of Managed Care (AJMC)*; [Available from: <https://www.ajmc.com/view/is-frailty-index-the-most-accurate-predictor-of-cvd->
36. Magnusson PK, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H et al (2013) The Swedish Twin Registry: establishment of a biobank and other recent developments. *Twin Res Hum Genet* 16(1):317–329
37. Chen Z, Chen Z, Jin X (2023) Mendelian randomization supports causality between overweight status and accelerated aging. *Aging Cell* 22(8):e13899
38. Gouveia C, Gibbons E, Dehghani N, Eapen J, Guerreiro R, Bras J (2022) Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK Biobank. *Sci Rep* 12(1):8404
39. Graf GH, Crowe CL, Kothari M, Kwon D, Manly JJ, Turney IC et al (2022) Testing Black-White disparities in biological aging among older adults in the United States: analysis of DNA-methylation and blood-chemistry methods. *Am J Epidemiol* 191(4):613–625
40. Mak JK, Kuja-Halkola R, Wang Y, Hägg S, Jylhävä J (2023) Can frailty scores predict the incidence of cancer? Results from two large population-based studies. *Geroscience* 45(3):2051–2064

41. Codd V, Denniff M, Swinfield C, Warner SC, Papakonstantinou M, Sheth S et al (2022) Measurement and initial characterization of leukocyte telomere length in 474,074 participants in UK Biobank. *Nat aging* 2(2):170–179
42. Privé F, Vilhjálmsdóttir BJ, Aschard H, Blum MG (2019) Making the most of clumping and thresholding for polygenic scores. *Am J Hum Genet* 105(6):1213–1221
43. Privé F, Arbel J, Vilhjálmsdóttir BJ (2020) LDpred2: better, faster, stronger. *Bioinformatics* 36(22–23):5424–5431
44. Dhillon A, Singh A, Bhalla VK (2024) HBS–STACK: hierarchical biomarker selection and stacked ensemble model for biomarker identification and cancer prediction on multi-omics. *Neural Comput Appl* 36(10):5413–5431
45. Therneau TM, Lumley T (2015) Package ‘survival’. *R Top Doc* 128(10):28–33
46. Polley E, LeDell E, Kennedy C, Lendle S, van der Laan M (2019) Package ‘SuperLearner’. CRAN
47. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C et al (2021) Package ‘pROC’. Package pROC
48. Chen T, He T, Benesty M, Khotilovich V (2019) Package ‘xgboost’. *R version* 90(1–66):40
49. RColorBrewer S, Liaw MA (2018) Package ‘randomforest’. University of California, Berkeley: Berkeley, CA, USA

Figures

Study flow chart

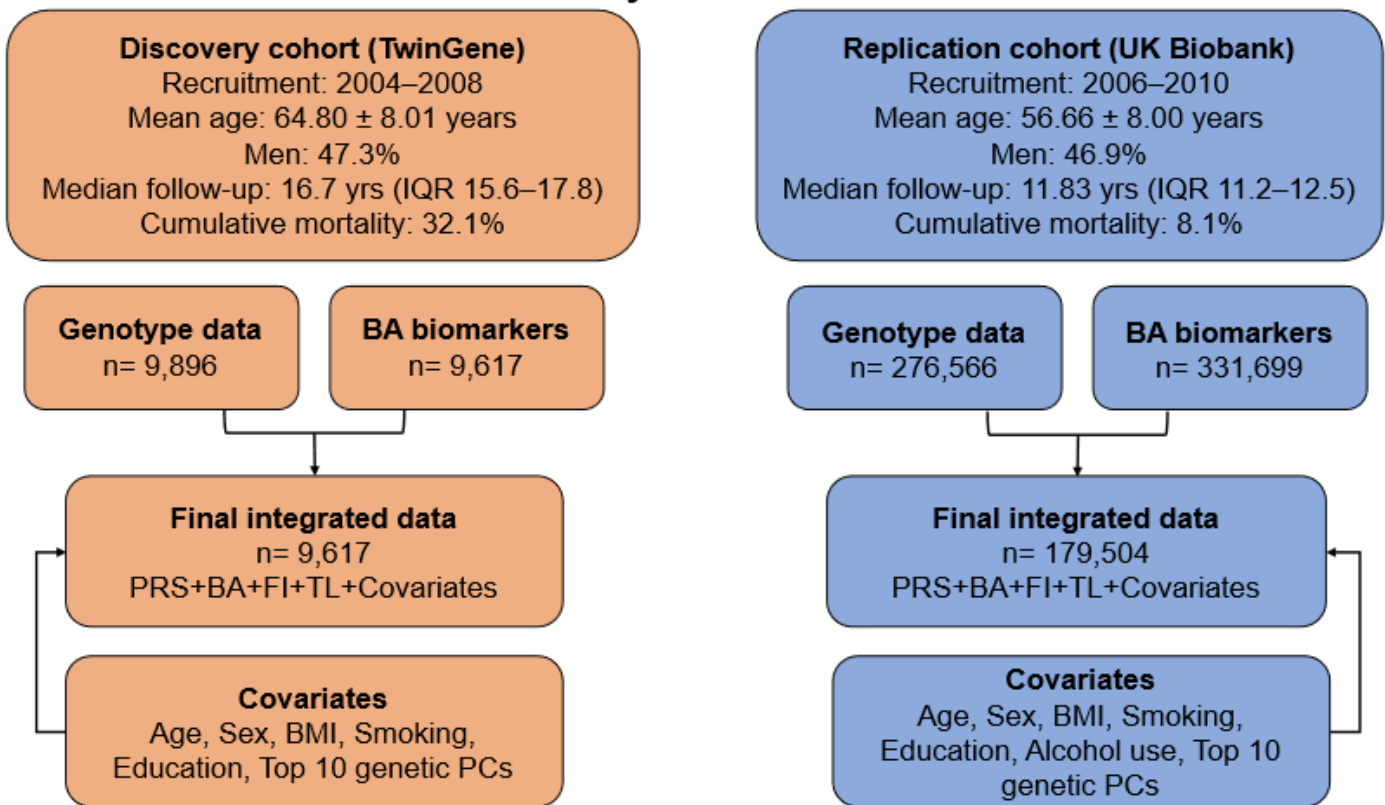
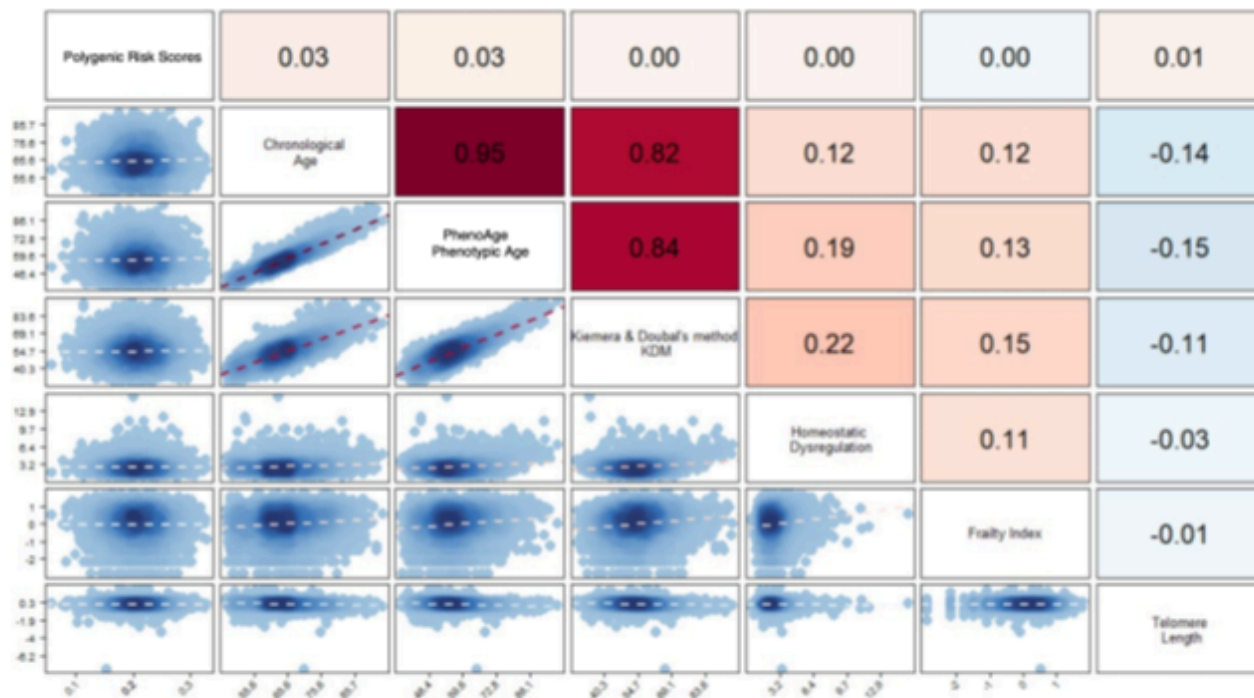
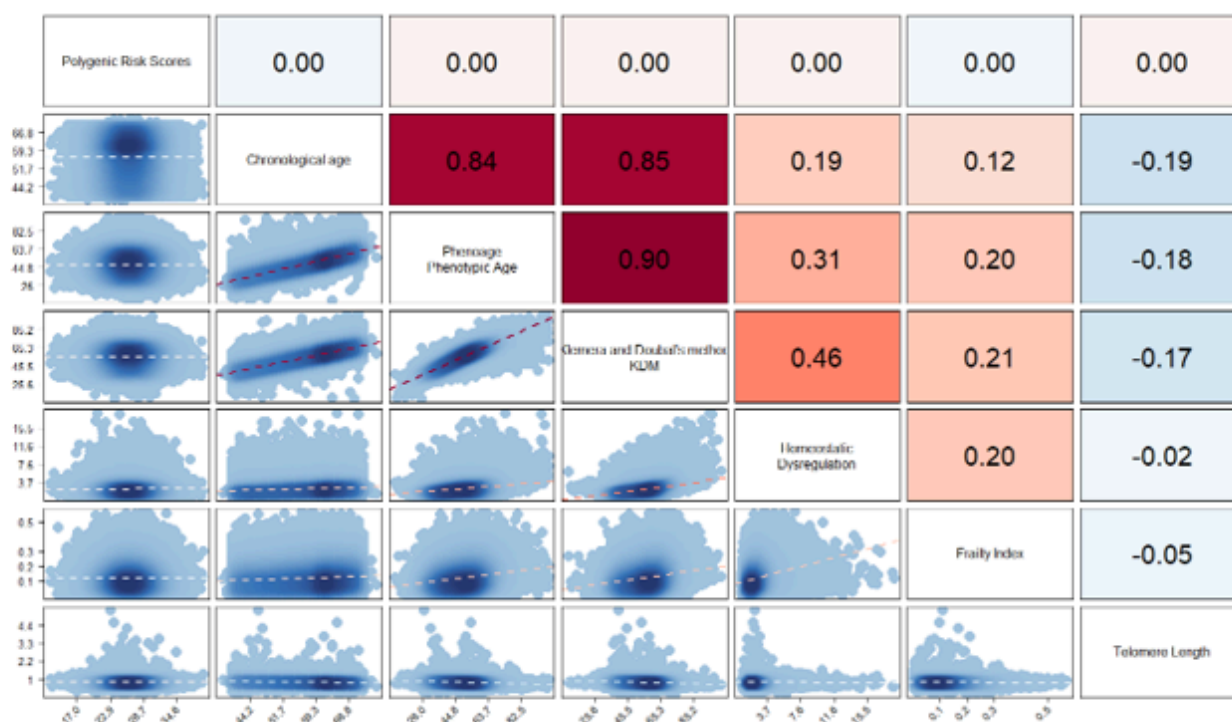


Figure 1

Study design and analytic sample construction in TwinGene and UK Biobank. TwinGene served as the discovery cohort and UK Biobank (UKB) as the external validation cohort. Boxes summarize recruitment periods, baseline age, follow-up, and cumulative mortality. For TwinGene, participants with genotype data passing quality control were eligible and the final analytic set (n=9,617) included individuals with complete data on biological aging biomarkers, frailty index (FI), telomere length (TL), polygenic risk score (PRS), and covariates. For UKB, the final analytic set (n=179,504) was defined as the intersection of unrelated European-ancestry participants with high-quality genotypes and participants with complete biomarker data required to compute BA measures, plus FI, TL, PRS, and covariates. Biomarker-based biological age estimates (PhenoAge, KDM, and HD) were analyzed primarily as age-adjusted residuals (PhenoAge_res, KDM_res, HD_res) to capture age-independent variation when models included chronological age. Cohort-specific covariates are listed in the bottom panels. Abbreviations: UKB, UK Biobank; BA, biological age estimates; PhenoAge, phenotypic age; KDM, Klemmera–Doubal method; HD, homeostatic dysregulation; FI, frailty index; TL, telomere length; PRS, polygenic risk score; PCs, principal components.



2A



2B

Figure 2

2A.

Correlations among aging measures in the Swedish TwinGene cohort. Scatterplot matrix of polygenic risk scores (PRS), chronological age (CA), Phenotypic Age (PhenoAge), Klemmera–Doubal method (KDM), homeostatic dysregulation (HD), frailty index (FI), and adjusted telomere length (TL) in TwinGene

(n=9,617). The lower triangle shows pairwise scatterplots; the upper triangle shows Pearson correlation coefficients (r), color-coded by direction and magnitude (red positive, blue negative).

Fig 2B.

Correlations among aging measures in the UK Biobank cohort. Scatterplot matrix of polygenic risk scores (PRS), chronological age (CA), Phenotypic Age (PhenoAge), Klemmera–Doubal method (KDM), homeostatic dysregulation (HD), frailty index (FI), and adjusted telomere length (TL) in UK Biobank (n=179,504). The lower triangle shows pairwise scatterplots; the upper triangle shows Pearson correlation coefficients (r), color-coded by direction and magnitude.

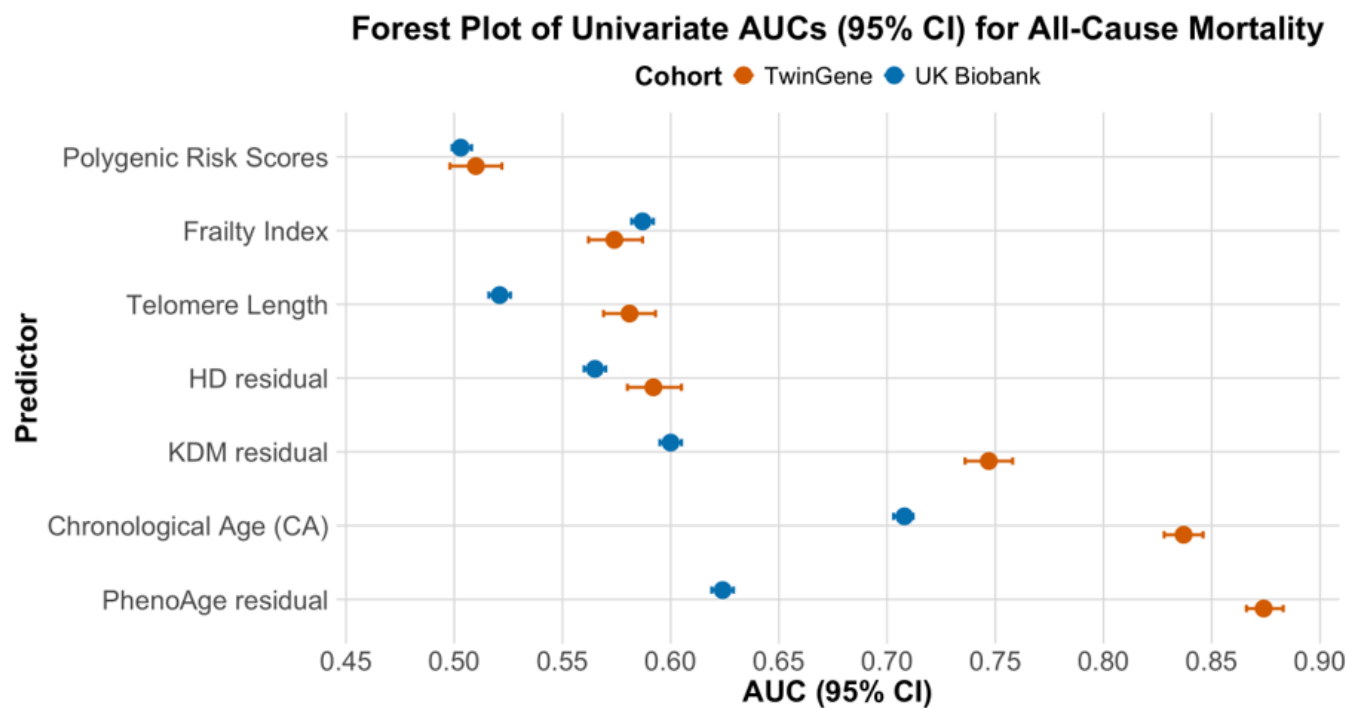


Figure 3

Univariate discrimination of aging and genetic predictors for all-cause mortality in TwinGene and UK Biobank. Forest plot showing the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals for each univariate predictor in TwinGene (n=9,617) and UK Biobank (n=179,504). Biomarker-based biological age estimates are shown as age-adjusted residuals (PhenoAge_res, KDM_res, HD_res), calculated within each cohort as residuals from linear regressions on chronological age (BA ~ CA). PRS was standardized within cohort (PRS_Z). Confidence intervals were computed using DeLong’s method.

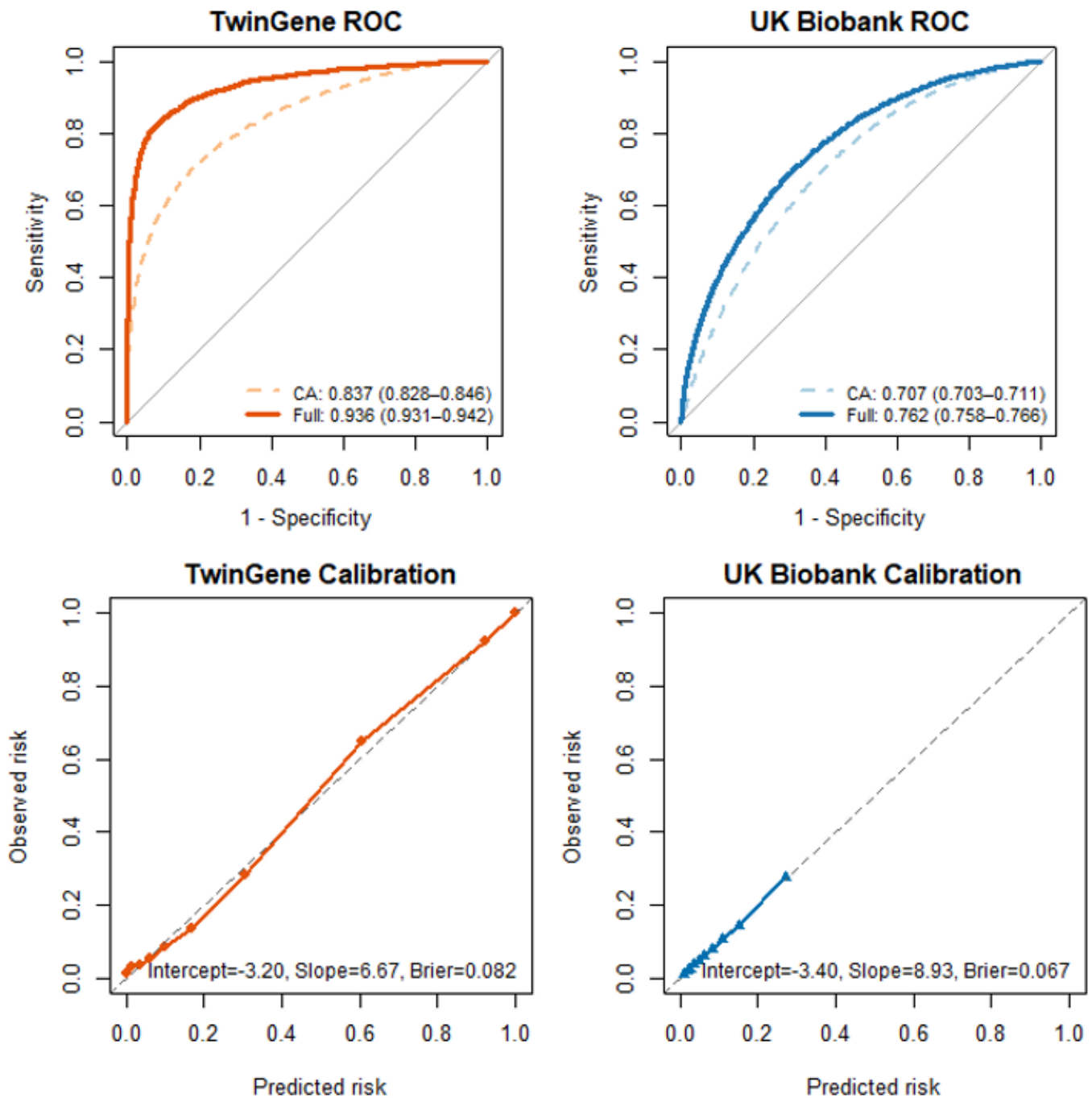


Figure 4

Discrimination and calibration of the age-only and full multivariable models for all-cause mortality in TwinGene and UK Biobank. Top row: ROC curves comparing the chronological age (CA) model versus the full model in TwinGene and UK Biobank. Curves are based on 10-fold cross-validated out-of-fold predictions from CV.SuperLearner; AUCs with 95% confidence intervals (DeLong) are shown in each panel. Bottom row: calibration plots for the full model in TwinGene and UK Biobank, showing observed versus predicted mortality risk across grouped predicted-risk bins (e.g., deciles); the dashed diagonal indicates ideal calibration. The full model corresponds to the final ladder model includes CA, covariates, standardized PRS, and biological aging measures (PhenoAge residual, frailty index, and telomere length).

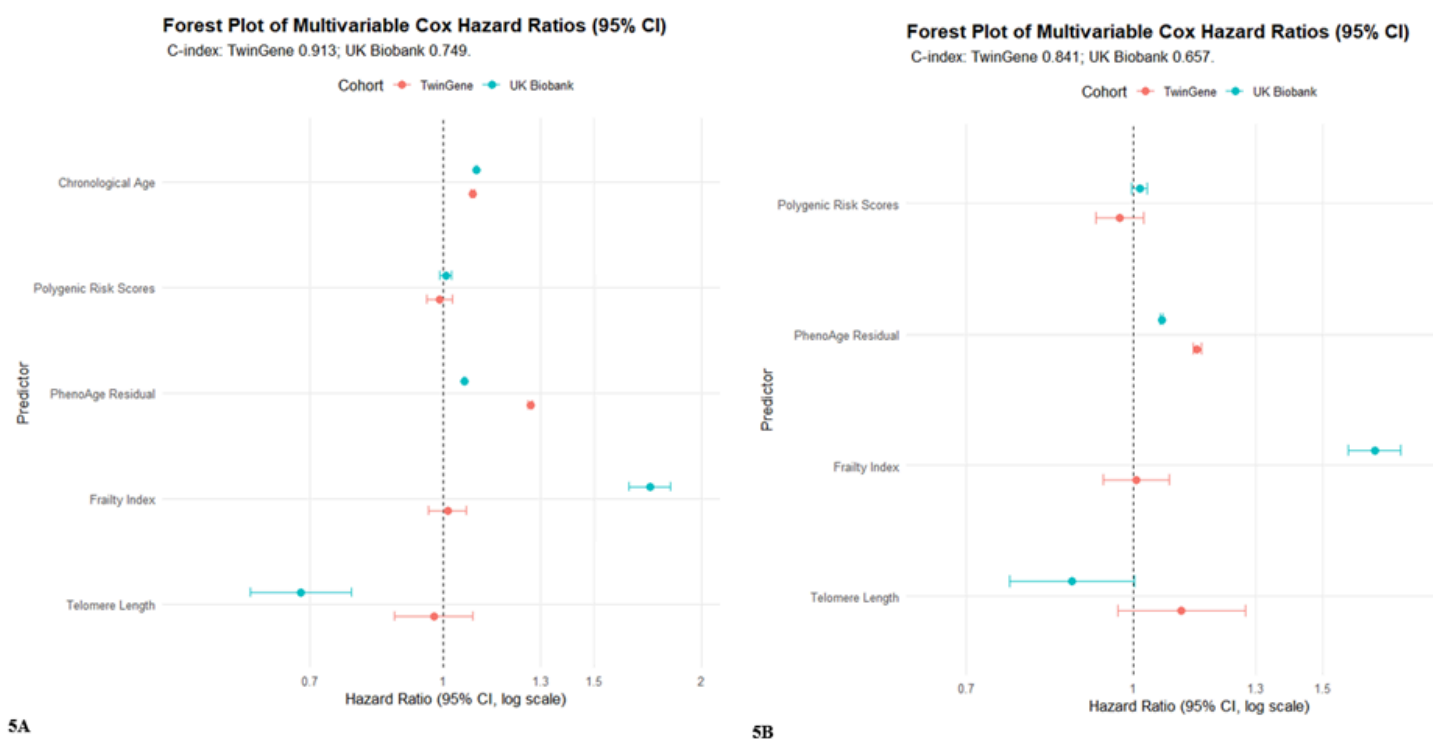


Figure 5

A.

Multivariable Cox hazard ratios for all-cause mortality (time-since-baseline as timescale) in TwinGene and UK Biobank.

Forest plot shows adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) from multivariable Cox proportional hazards models using time since baseline as the underlying timescale. Models included chronological age, standardized polygenic risk score, PhenoAge residual (age-adjusted), frailty index, and telomere length, with additional cohort-available covariates (TwinGene: sex, BMI, smoking, education; UK Biobank: sex, BMI, smoking, alcohol, education). Points indicate HRs and horizontal bars indicate 95% CIs; the dashed vertical line denotes HR=1. Model discrimination is summarized by Harrell's C-index (TwinGene 0.913; UK Biobank 0.749).

Fig 5B.

Multivariable Cox hazard ratios for all-cause mortality using age as the timescale in TwinGene and UK Biobank.

Forest plot shows adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) from Cox models using attained age as the underlying timescale (start–stop specification). Because age is the timescale, chronological age is not included as a predictor in these models. Predictors included standardized PRS,

PhenoAge residual (age-adjusted), frailty index, telomere length, and cohort-available covariates (TwinGene: sex, BMI, smoking, education; UK Biobank: sex, BMI, smoking, alcohol, education). Points indicate HRs and horizontal bars indicate 95% CIs; the dashed vertical line denotes HR=1. Model discrimination is summarized by Harrell's C-index (TwinGene 0.841; UK Biobank 0.657).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [STROBEChecklistMostafaeietalPLOSagingandHealth.docx](#)
- [SupplementalTablesandFigures.docx](#)