ClinicalAge, a framework for modelling ageing trajectories from longitudinal electronic health records in the UK: a model development and validation study

Jie Lian, Zhengxian Fan*, Wei Fan, Shishir Rao, Qianqian Yang, Guyu Zeng, Ben Omega Petrazzini, Nouman Ahmed, Zeinab Bidel, Milad Nazarzadeh, Ron Do, Andrew Clegg, Malgorzata Wamil, Kazem Rahimi*

Summary

Background Ageing is a key determinant of morbidity and mortality. In this study, we aimed to develop and assess a new ageing biomarker, ClinicalAge, derived from routinely collected longitudinal electronic health record (EHR) data.

Methods In this model development and validation study, we used UK Biobank data for model development and UK Clinical Practice Research Datalink (CPRD) Aurum data for model validation. The UK Biobank is a longitudinal population-based cohort of approximately 500 000 individuals aged 40–69 years at recruitment (March 16, 2006, to Oct 1, 2010), with linked EHR from the November 2022 UK Biobank data release. CPRD Aurum is a primary care EHR database containing anonymised longitudinal data from over 30 million patients across the UK, including diagnoses, medication, procedures, and laboratory results. A transformer-based deep learning model was trained from longitudinal EHR data to generate ClinicalAge using the validated biomarker PhenoAge as reference. We used a UK Biobank cohort for model development and internal validation, and three CPRD Aurum cohorts for external validation, longitudinal trajectory analysis, and comparison with the electronic frailty index. ClinicalAge was evaluated for associations with 17 functional measures, ten biological ageing biomarkers, 31 ageing-related diseases (11 cancer diagnoses and 20 non-cancer diseases), and all-cause, cancer, and cardiovascular mortality. We further examined longitudinal ClinicalAge trajectories and investigated clinical and disease-specific factors associated with accelerated ageing.

Findings The UK Biobank cohort included 320 891 participants (median chronological age 58·0 years [IQR 51·0–63·0]; median PhenoAge 56·0 years [48·6–62·0]; 173 067 [53·9%] were female; and 147 824 [46·1%] were male) and the CPRD cohort included 4 728 376 individuals (51·0 years [44·0–60·0]; 2 293 243 [48·5%] were female and 2 435 133 [51·5%] were male). ClinicalAge had an R^2 score of 0·819 with PhenoAge and showed a significant association with the 17 functional measures and ten biological ageing biomarkers. It was also significantly associated with the incidence of all 20 non-cancer diseases in both datasets, as well as nine (82%) of 11 cancers in UK Biobank (seven positive associations and two negative) and ten (91%) of 11 cancers (eight positive associations and two negative) in CPRD. Moreover, ClinicalAge was significantly associated with all-cause mortality (UK Biobank adjusted hazard ratio [HR] 1·23 [95% CI 1·21–1·24]; CPRD adjusted HR 1·36 [1·36–1·36]), cancer mortality (UK Biobank adjusted HR 1·12 [1·10–1·14]; CPRD adjusted HR 1·17 [1·16–1·17]), and cardiovascular mortality (UK Biobank adjusted HR 1·33 [1·29–1·36]; CPRD adjusted HR 1·36 [1·35–1·36]), after adjustment for clinical covariates and lifestyle factors. Midlife was identified as a key window for divergence, with early acceleration in individuals developing diabetes, hypertension, depression, or lifestyle disorders.

Interpretation ClinicalAge is a non-invasive biomarker that leverages routinely collected EHR data to quantify ageing across the life course. It shows strong associations with ageing-related biomarkers, functional decline, disease risk, and mortality. Its interpretability and anticipatory signal offer a practical tool for the early identification of individuals at risk of accelerated ageing, enabling targeted preventive strategies such as lifestyle modification, cardiometabolic control, and mental health support.

Funding None.

Copyright © 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Lancet Digit Health 2026

Published Online
https://doi.org/10.1016/j.landig.2026.101060

*Contributed equally

Deep Medicine, Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, UK (J Lian PhD, Z Fan DPhil, W Fan PhD, S Rao DPhil, Q Yang MSc, G Zeng PhD, B O Petrazzini BSc, N Ahmed MSc, Z Bidel MSc, M Nazarzadeh DPhil, M Wamil PhD, Prof K Rahimi MD); **School of Computer Science, University of Auckland, Auckland, New Zealand** (W Fan); **Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA** (B O Petrazzini, Prof R Do PhD); **Academic Unit for Ageing and Stroke Research, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK** (Prof A Clegg MD)

Correspondence to:
Dr Jie Lian, Deep Medicine, Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford OX2 0EW, UK
jie.lian@wrh.ox.ac.uk

Introduction

Ageing represents a fundamental biological process marked by a gradual loss of physiological integrity, increasing the risk of chronic diseases, disability, and mortality.[1–3] Although chronological age is widely used in clinical practice, it often does not capture the substantial variability in health status among individuals of the same chronological age. This variation has prompted the

Research in context

Evidence before this study

We searched PubMed and Google Scholar for articles published between Jan 1, 2010, and Jan 1, 2025, using search terms "biological age", "ageing clocks", "electronic health records", "PhenoAge", "frailty index", and "deep learning and ageing". The search was limited to the English language. Biological ageing has been widely studied using molecular biomarkers such as DNA methylation clocks, proteomics, telomere length, and composite measures (eg, PhenoAge). These models show strong associations with morbidity, mortality, and functional decline, but require specialised biospecimens and laboratory assays, limiting clinical scalability. Parallel work such as the electronic frailty index (eFI) has shown the potential of electronic health records (EHR)-based ageing tools, but most frailty measures are static and threshold-based. Few studies have leveraged longitudinal EHR data with deep learning to estimate a continuous ageing measure, and, to our knowledge, none have validated such a model across millions of individuals and a broad range of biological, functional, and clinical outcomes and trajectory analyses.

Added value of this study

We developed ClinicalAge, an ageing measurement that leverages routinely collected longitudinal EHR and a transformer-based model to learn age-related patterns directly from clinical data. This framework provides an interpretable and scalable approach to monitor ageing trajectories at scale. Importantly, ClinicalAge bridges the gap between molecular ageing clocks and clinical risk assessment by revealing modifiable, disease-linked aspects of ageing that emerge decades before manifestation of major health outcomes. ClinicalAge provides a framework for studying ageing trajectories at population scale and for identifying potentially modifiable patterns of accelerated ageing.

Implications of all the available evidence

Taken together with existing evidence on ageing measures, ClinicalAge provides a practical framework for integrating ageing assessment into routine health care and population health research. Because it relies on routinely collected EHR data, it could be implemented without additional biomarker collection. ClinicalAge complements existing ageing markers by capturing dynamic clinical processes that precede disease onset. Further research should evaluate transportability across health systems, clinical utility, fairness across population subgroups, and whether ClinicalAge-guided prevention strategies improve health outcomes.

development of informative indicators of ageing that capture risk and functional decline. Such ageing markers, especially those responsive to medical or lifestyle interventions, offer crucial advantages: they enable early identification of individuals at elevated risk of disease and functional decline, support accurate stratification of disease vulnerability, and facilitate personalised, preventive care tailored to an individual's ageing trajectory rather than their chronological age.

Biomarkers of ageing are commonly grouped into clinical non-invasive markers, molecular biomarkers, and composite approaches. Clinical measures, such as grip strength, offer advantages for large-scale use by avoiding invasive procedures, albeit with modest predictive performance.[4] By contrast, molecular biomarkers measurable in peripheral blood, such as telomere length,[5] inflammatory markers,[6] DNA methylation,[7,8] protein markers,[9] or metabolomic markers,[10] offer mechanistic insight into ageing processes. While valuable, many of these biomarkers were originally developed or validated in broad biomedical contexts rather than specifically for ageing research, and their wide application in clinical practice is often limited by high cost, invasiveness, and technical complexity. To address these limitations, composite indices integrating both clinical and molecular measures, such as PhenoAge[11] and GrimAge2,[12] have shown improved performance in predicting age-related clinical outcomes. Comparative studies have also evaluated multiple biological age algorithms, including PhenoAge, within large population cohorts, showing that different measures capture overlapping yet distinct aspects of ageing biology.[13] However, these algorithms still have similar barriers to implementation due to their reliance on specialised testing and data inputs.

Beyond technical limitations, many existing methods rely on chronological age as the reference, which can obscure clinically relevant deviations in ageing trajectories. By comparison, phenotypical ageing measures, which have been shown to outperform chronological age in predicting mortality and life expectancy,[14] stand out as a more informative alternative reference framework. In addition, most models are based on single-timepoint data, overlooking the progressive and dynamic character of ageing over time.[15–17] Together, these limitations underscore the need for ageing measures that are clinically informative and sensitive to change.

To address the limitations of static biomarkers and the scalability challenges of existing approaches, the increasing availability of electronic health records (EHRs) presents a timely and underused opportunity. Large-scale EHRs, such as the UK Clinical Practice Research Datalink (CPRD),[18] provide rich longitudinal clinical data, such as diagnoses, procedures, and medications, enabling prediction of disease risk, ageing risk (electronic frailty index [eFI]),[19] and mortality beyond conventional clinical risk factors. Complementary resources such as the UK Biobank[20] further link EHRs to genetic and biochemical data, offering a platform to connect biological measures with detailed clinical phenotypes.

Building on this foundation, we hypothesise that longitudinal EHR-derived health trajectories can serve as a

non-invasive proxy for ageing, capturing cumulative physiological decline and its clinical manifestations over time. We introduce ClinicalAge, an ageing measurement that applies deep learning to integrate EHR data into a composite measure of ageing. By leveraging the breadth and temporal depth of clinical information in the EHR, ClinicalAge aims to predict ageing-associated outcomes, identify factors driving accelerated ageing, and provide a scalable tool for translational and population health research.

Methods

Study design

For this model development and validation study, we used two large-scale UK datasets: UK Biobank and CPRD Aurum. The UK Biobank is a longitudinal population-based cohort of approximately 500 000 individuals aged 40–69 years at recruitment (March 16, 2006, to Oct 1, 2010),[20] with a linked EHR from the November 2022 UK Biobank data release. CPRD Aurum is a primary care EHR database containing anonymised longitudinal data from over 30 million patients with no age restriction across the UK, including diagnoses, medication, procedures, and laboratory results.[21] Although both draw on UK data, the two datasets are maintained under distinct governance with no individual-level linkage overlap. Data on sex (male or female) and ethnicity (Asian, Black, other, or White) were obtained from the respective databases.

We developed and internally validated a transformer-based model, based on the preprint paper by Vaswani and colleagues,[22] in the UK Biobank to estimate an ageing score from longitudinal EHR data. Each individual's EHR history was represented as a chronological sequence of coded clinical events, including diagnoses, procedures, and medication, tokenised using a custom vocabulary.[23,24] Our model learns from the presence, absence, and timing of events within each patient's EHR, treating missingness as an informative signal rather than a deficiency requiring correction through imputation. As for the training reference, we used age acceleration derived from PhenoAge,[11] a widely validated biomarker of ageing, which has been shown accurately to predict morbidity, mortality, and functional decline.

We randomly sampled the data at the participant level into 80% for training (including 10% validation to determine the hyperparameter), and 20% for internal testing. The CPRD Aurum dataset was reserved for independent external validation (figure 1A; appendix pp 3–7).

The model output, representing the estimated gap of clinical age score (ClinicalAge) from chronological age, was defined as the ClinicalAge gap (figure 1A). A larger ClinicalAge gap indicates accelerated ageing, whereas a smaller or negative gap reflects slower or resilient ageing. Subsequently, ClinicalAge gap was analysed for its associations with established biological ageing biomarkers, functional measures, incident ageing-related disease, cause-specific mortality, and ageing trajectories (figure 1B).

Ethical approval for UK Biobank was obtained from the North West Multi-centre Research Ethics Committee (reference 11/NW/0382), and all participants provided written informed consent. The present analyses were conducted under UK Biobank application number 83942. CPRD data were accessed under approval from the independent scientific advisory committee through CPRD's Research Data Governance Process. All data were anonymised before access, and individual patient consent was not required.

Participants

We included four cohorts in this study: cohort one, UK Biobank development and internal validation; cohort two, CPRD Aurum external validation; cohort three, CPRD Aurum trajectory cohort; and cohort four, the CPRD Aurum eFI[19] cohort.

For cohort one, we included UK Biobank participants with complete data on the nine blood biomarkers required to compute PhenoAge, who also passed quality control (appendix p 8). PhenoAge is a validated biomarker of biological ageing derived from nine standard blood biomarkers: albumin, creatinine, glucose, C-reactive protein, lymphocyte percentage, mean cell volume, red cell distribution width, alkaline phosphatase, and white blood cell count (appendix p 8). Participants were aged 40–69 years at recruitment time, had at least one recorded visit, and had uninterrupted follow-up thereafter. The age range of 40–69 years reflected the recruitment design of UK Biobank, at which PhenoAge biomarkers were collected, which enrolled participants exclusively within this interval.

For cohort two (CPRD validation), we sampled individuals by assigning a random baseline date between Jan 1, 2006, and Jan 1, 2010, mirroring the UK Biobank enrolment period. Inclusion criteria required an age of 40–69 years at baseline and at least 1-year previous general practice (GP) registration.

For cohort three (trajectory analyses), we restricted the analyses to a subset of cohort two with EHR data available from age 40 years onwards and associated follow-up. Follow-up continued until participants reached age 65 years or their CPRD records ended (eg, due to deregistration or practice withdrawal), at which point data were censored.

Cohort four comprised individuals in CPRD Aurum with a recorded eFI score, using the date of eFI assessment as baseline and applying the same age and registration criteria as for cohort two. The eFI is based on the cumulative deficit model and comprises 36 equally weighted health deficits recorded in primary care EHRs; it has been extensively validated in UK populations for predicting outcomes such as hospitalisation, care home admission, and mortality.[19,25]

See Online for appendix

Association with ageing biomarkers, functions, incidence of disease, and survival

We first evaluated the cross-sectional association between ClinicalAge gap and 17 individual functional measures

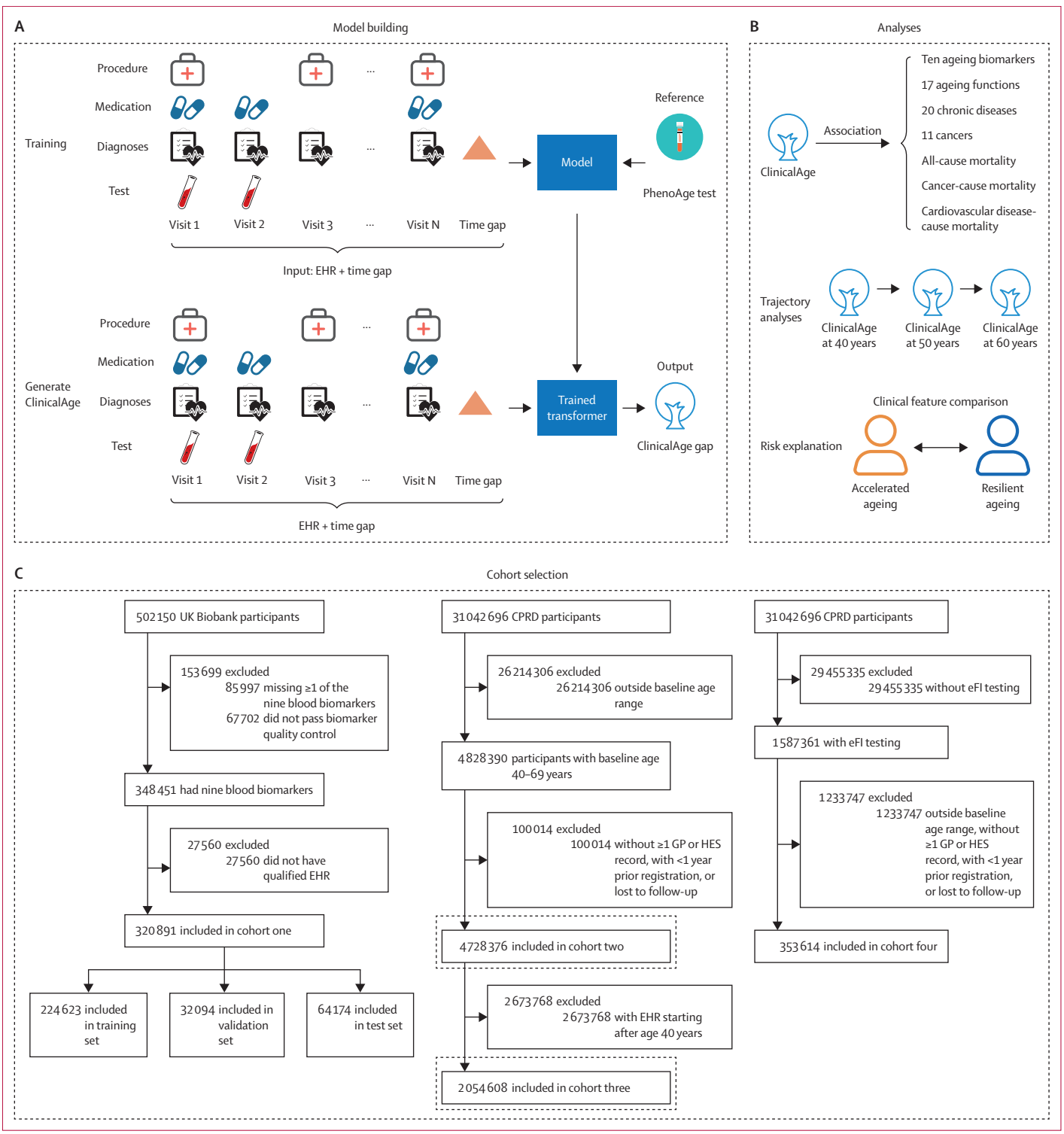

Figure 1: **Overview of the model and cohorts**
(A) Overview of the ClinicalAge model framework. A transformer-based deep learning model was trained on UK Biobank data to estimate the ClinicalAge gap from longitudinal EHRs. (B) Analyses overview, examining associations with functional measures, ageing biomarkers, disease incidence, and mortality, as well as trajectory analyses and comparisons of accelerated ageing patterns. (C) Cohort selection. Participants in the UK Biobank with blood biomarker data were used for model development and internal validation. CPRD Aurum was used for external validation and trajectory analyses, including a subset with electronic frailty index (eFI) scores. CPRD=Clinical Practice Research Datalink. EHR=electronic health record. GP=general practice. HES=hospital episode statistics.

(appendix p 17) in cohort one, spanning physical domains, cognitive domains, and frailty (appendix pp 14–16), following established protocols.[9,10] Similarly, we examined its cross-sectional relationship with ten established biomarkers of biological ageing in cohort one (appendix pp 9, 18), including leukocyte telomere length[5] and insulin-like growth factor 1 (IGF-1),[26] alongside clinical biochemistry markers.

To assess the predictive utility of ClinicalAge gap for ageing-related disease risk, we fitted Cox proportional hazards models to estimate 10-year risk of 31 age-related conditions (11 cancer diagnoses and 20 non-cancer diseases) that represent major causes of death or highly prevalent morbidities in ageing adults in cohorts one and two (appendix p 10). Disease outcomes were defined using EHR diagnosis codes;[27–29] the full list of conditions and their code definitions are provided in the appendix (pp 19–51).

We further evaluated the association between ClinicalAge gap and 10-year incidence of all-cause mortality, cancer-specific mortality, and cardiovascular disease mortality using four models in cohorts one and two: (1) unadjusted; (2) adjusted for baseline PhenoAge; (3) adjusted for baseline chronological age and sex; and (4) fully adjusted for baseline chronological age, sex, BMI, systolic blood pressure, HDL and LDL cholesterol, smoking status, index of multiple deprivation (IMD; 2015 version), and ethnicity.[9,10] The IMD is an area-level composite measure of socioeconomic deprivation in the UK, combining seven domains ranked within each country of the UK. Participants were categorised into national quintiles based on the distribution within each country.

To compare predictive value of ClinicalAge with eFI, we conducted parallel analyses in cohort four. This cohort was randomly partitioned into 80% training and 20% validation sets. We computed C-indices for models one, three, and four, with 95% CIs from 1000 bootstrap resamples. Calibration was assessed by deciles of predicted risk using observed risks and a calibration slope.

ClinicalAge trajectory and interpretability

To examine longitudinal trajectories of the ClinicalAge gap and identify clinical features associated with accelerated or decelerated ageing, we conducted a series of trajectory analyses in cohort three. ClinicalAge gap was computed at age milestones (age 40, 45, 50, 55, 60, and 65 years) using data accumulated up to each age. At each milestone, participants were categorised into four groups based on the distribution of ClinicalAge gap: group one (ClinicalAge gap less than or equal to mean minus SD; ie, clinically younger); group two (ClinicalAge gap more than mean minus SD but less than or equal to the mean; ie, slightly clinically younger or average); group three (ClinicalAge gap more than the mean but less than or equal to mean plus SD; ie, slightly clinically older); and group four (ClinicalAge gap more than mean plus SD; ie, clinically older). Two primary trajectory types were examined: (1) participants transitioning from group one to group four, and (2) those who consistently remained in either group one or group four throughout follow-up. To explore clinical drivers underlying differences in ageing rate, we compared the relative frequencies of ICD-coded diagnoses across groups, with differences visualised using heat maps.

Statistical analysis

Continuous variables are reported as medians with IQRs and means with SDs, and categorical variables as counts and percentages. Model performance was assessed by comparing ClinicalAge with PhenoAge using coefficient of determination (R^2), mean absolute error, and mean absolute percentage error to assess alignment. Agreement between ClinicalAge and PhenoAge was evaluated using Bland–Altman analysis. Associations between ClinicalAge gap and outcomes were tested using linear regression models. Cox proportional hazards models estimated hazard ratios (HRs) and 95% CIs for disease incidence and mortality, with time-to-event defined from baseline to first event, death, or censoring. Additional analyses included plotting Kaplan–Meier survival curves, fitting Fine–Gray models to account for competing risks, and assessing the covariate linearity assumption (appendix p 10).

For analyses of the ten ageing biomarkers and 17 functional measures, only participants with non-missing outcomes were included. For disease and mortality association analyses, missing values in baseline covariates (fully adjusted model) were addressed using multiple imputation by chained equations. Five imputed datasets were generated and combined using Rubin's rules to account for uncertainty. Statistical significance was defined as $p<0·05$ (two-sided). Complete-case sensitivity analyses were conducted (appendix p 10).

All analyses were conducted using Python, version 3.8, with relevant packages detailed in the appendix (p 10). Our study satisfies all five CODE-EHR minimum framework standards (with three achieving the preferred level; appendix pp 99–101) and is reported in accordance with TRIPOD+AI (appendix pp 97–98).

Role of the funding source

There was no funding source for this study.

Results

Cohort construction is summarised in figure 1C. A total of 320 891 participants from the UK Biobank were included in cohort one (table 1), representing 63·9% of UK Biobank patients (appendix p 11). The median chronological age was 58·0 years (IQR 51·0–63·0), the median PhenoAge was 56·0 years (IQR 48·6–62·0), 173 067 (53·9%) participants were female, and 147 824 (46·1%) participants were male. During 10 years of follow-up in 320 891 participants, 13 468 (4·2%) participants died (7662 [2·4%] cancer-related deaths and 2731 [0·9%] cardiovascular disease-related deaths). Participants had a mean number of recorded health-care visits of 22·3 (SD 32·4). In cohort two, 4 728 376 participants from CPRD Aurum were included. The median chronological age

	Cohort one: UK Biobank (n=320 891)	Cohort two: CPRD (n=4 728 376)
Chronological age, years	58·0 (51·0 to 63·0)	51·0 (44·0 to 60·0)
PhenoAge, years	56·0 (48·6 to 62·0)	NA
ClinicalAge, years	56·4 (48·8 to 61·8)	49·6 (42·3 to 58·3)
ClinicalAge gap, years	−1·8 (−2·4 to −1·2)	−1·9 (−2·4 to −1·3)
Sex		
Female	173 067 (53·9%)	2 293 243 (48·5%)
Male	147 824 (46·1%)	2 435 133 (51·5%)
BMI (kg/m^2)	26·9 (24·3 to 30·0); n=319 788	27·7 (24·5 to 31·6); n=1 643 909
HDL cholesterol (mmol/L)	1·4 (1·2 to 1·7); n=320 801	1·4 (1·2 to 1·7); n=1 215 192
LDL cholesterol (mmol/L)	3·5 (3·0 to 4·1); n=320 401	3·0 (2·3 to 3·7); n=903 743
Systolic blood pressure (mm Hg)	139·0 (126·0 to 152·0); n=299 851	130·0 (120·0 to 140·0); n=2 691 406
Index of multiple deprivation		
Class 1 (least deprived)	103 110/273 524 (37·7%)	1 120 241 (23·7%)
Class 2	75 069/273 524 (27·4%)	1 000 320 (21·2%)
Class 3	38 404/273 524 (14·0%)	916 741 (19·4%)
Class 4	22 561/273 524 (8·2%)	876 214 (18·5%)
Class 5 (most deprived)	34 380/273 524 (12·6%)	814 860 (17·2%)
Missing	47 367	0
Smoking		
Current	31 115/319 366 (9·7%)	807 382/2 655 723 (30·4%)
Never	174 797/319 366 (54·7%)	1 270 243/2 655 723 (47·8%)
Previous	113 454/319 366 (35·5%)	578 098/2 655 723 (21·8%)
Missing	1525	2 072 653
Ethnicity		
Asian	6283/319 493 (2·0%)	147 791/3 533 137 (4·2%)
Black	3856/319 493 (1·2%)	110 949/3 533 137 (3·1%)
Other	4156/319 493 (1·3%)	70 440/3 533 137 (2·0%)
White	305 198/319 493 (95·5%)	3 203 957/3 533 137 (90·7%)
Missing	1398	1 195 239

Data are median (IQR), n (%), or n/N (%). Where there were missing data, the sample size is shown. CPRD=Clinical Practice Research Datalink. NA=not applicable.

Table 1: **Baseline characteristics of participants in UK Biobank and CPRD**

was 51·0 years (IQR 44·0–60·0), 2 293 243 (48·5%) participants were female, and 2 435 133 (51·5%) participants were male. Over the 10-year follow-up of 4 728 376 participants, 344 807 (7·3%) participants died (157 626 [3·3%] from cancer and 132 692 [2·8%] from cardiovascular disease). Participants had a mean number of recorded health-care visits of 35·3 (SD 35·8). Baseline characteristics were not tabulated for cohort three as this cohort is a subset of cohort two.

Cohort four (eFI study) included 353 614 participants (appendix p 52), with a median age of 65·0 years (IQR 63·0–67·0) and a median follow-up of 32·0 months (IQR 15·0–42·0). Within this subgroup, 9329 (2·6%) participants died (4165 [1·2%] cancer-related and 4100 [1·2%] cardiovascular disease-related deaths).

In cohort one, ClinicalAge aligned closely with PhenoAge (R^2=0·819, 95% CI 0·816–0·821; mean absolute error 3·01, 95% CI 2·99–3·02; mean absolute percentage error 5·54%, 95% CI 5·51–5·57). Bland–Altman agreement was high (mean bias 0·03; limits −7·43 to 7·49; appendix p 68), and both ClinicalAge gap and PhenoAge gap had approximately normal distributions with slight left skew (appendix p 69). ClinicalAge also showed a high degree of alignment with chronological age (R^2= 0·933, 95% CI 0·932–0·934). Across CPRD and UK Biobank, women had lower median ClinicalAge gaps than men, and higher deprivation (IMD) was associated with higher medians and wider IQRs (appendix pp 53–54). To confirm that this metric was not confounded by chronological age itself, we performed residualisation quality checks, which showed negligible age dependence (appendix p 11).

Among the 17 functional measures (figure 2A, appendix pp 56–57), increasing ClinicalAge gap was associated with poor self-rated health, slow walking pace, frequent tiredness, and insomnia. Higher ClinicalAge gaps were also linked to higher frailty scores, elevated systolic blood pressure, increased arterial stiffness, slower reaction time, and higher BMI. Conversely, ClinicalAge gap was negatively associated with grip strength, bone mineral density, lung function, and fluid intelligence score, reflecting physical and cognitive performance declines.

In cohort one, the ClinicalAge gap was significantly associated with all ten established biomarkers of biological ageing (figure 2B, appendix p 55), including PhenoAge components (albumin, C-reactive protein [CRP], and

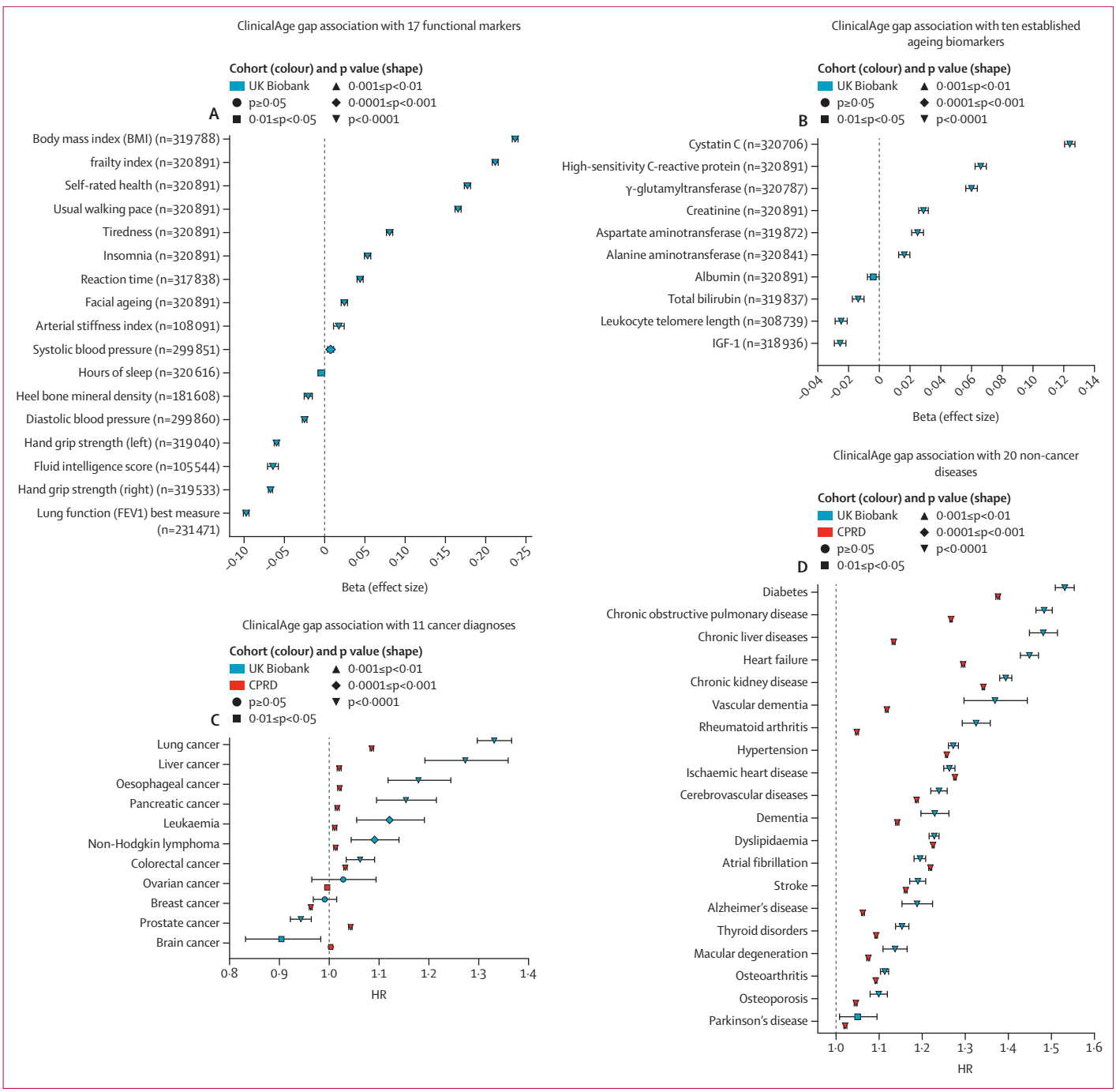

Figure 2: **Validation of ClinicalAge and associations with ageing biomarkers, function, and clinical outcomes**
(A) Association between ClinicalAge gap and 17 functional ageing measures in UK Biobank (cohort one). (B) Association between ClinicalAge gap and ten established biomarkers of biological ageing in UK Biobank (cohort one). (C) HRs for the association between ClinicalAge gap and 10-year incidence of 11 cancer diseases in UK Biobank (cohort one) and CPRD (cohort two). (D) HRs for the association between ClinicalAge gap and 10-year incidence of 20 non-cancer diseases in UK Biobank (cohort one) and CPRD (cohort two). (A) and (B) display standardised beta coefficients from linear regression models adjusted for model four covariates (UK Biobank only, biomarkers unavailable in CPRD Aurum), with 95% CIs. For BMI and systolic blood pressure outcome, the corresponding outcomes variable was excluded from the covariate set. (C) and (D) report HRs per 1-year increase in ClinicalAge gap from Cox models adjusted for model three, with 95% CIs. CPRD=Clinical Practice Research Datalink. HR=hazard ratio. IGF-1=insulin-like growth factor 1.

creatinine). Strong positive associations were observed with cystatin C (standardised beta 0·124, 95% CI 0·121 to 0·127; p<0·0001), high-sensitivity CRP (0·066, 0·062 to 0·070; p<0·0001), and γ-glutamyl transferase (0·060, 0·056 to 0·064; p<0·0001). Negative associations with IGF-1 (–0·026, –0·029 to –0·022; p<0·0001) and leukocyte

telomere length (–0·025, –0·029 to –0·021; p<0·0001) were observed.

Associations between ClinicalAge gap and 10-year risk of 31 age-related diseases in cohort one and cohort two adjusted for model three are shown in figure 2C and 2D. In cohort one, the strongest associations were observed for type 2 diabetes, chronic kidney disease, chronic obstructive pulmonary disease, and heart failure. Elevated ClinicalAge gap was positively associated with the incidence of all 20 non-cancer diseases examined. Among cancer outcomes, significant associations were identified for nine (82%) of 11 cancers; positive associations were identified for seven (64%) of 11 cancers and negative associations were observed for brain and prostate cancer (appendix pp 58–59). Sensitivity Fine–Gray competing risk analyses for incident disease outcomes yielded results consistent with the primary analyses (appendix pp 60–61). Other related results can be found in the appendix (pp 69–72).

In cohort two, associations broadly mirrored those in UK Biobank, supporting the generalisability of ClinicalAge gap across clinical populations (appendix p 62). ClinicalAge gap was positively associated with all 20 non-cancer conditions adjusted by model three. Among cancers, significant associations were identified for ten (91%) of 11 cancers; eight (73%) of 11 cancers showed positive associations, whereas two negative associations were observed for breast and ovarian cancer.

Increased ClinicalAge gap was strongly associated with 10-year mortality risk across all causes of death in both cohort one and cohort two (table 2). In cohort one, each 1-year increase in ClinicalAge gap was associated with an HR of 1·40 (95% CI 1·38–1·41) for all-cause mortality in model one. This association remained robust after full adjustment for age, sex, lifestyle, and clinical covariates (model four HR 1·23, 95% CI 1·21–1·24). Similar results were observed for cancer-specific mortality (model four HR 1·12, 95% CI 1·10–1·14) and cardiovascular disease mortality (model four HR 1·33, 95% CI 1·29–1·36), with all associations remaining statistically significant at p<0·0001. Associations persisted after adjusting for PhenoAge (model two), suggesting complementary information with PhenoAge. Sensitivity analyses (Fine–Gray for cause-specific mortality, spline checks for linearity, and Kaplan–Meier curves) supported the primary findings and are reported in UK Biobank (appendix pp 64–65, 73–75).

In cohort two, the ClinicalAge gap showed similar associations with mortality outcomes (table 2). The HR for all-cause mortality in the unadjusted model was 1·63 (95% CI 1·62–1·63; p<0·0001) and remained elevated after full adjustment (model four HR 1·36, 95% CI 1·36–1·36; p<0·0001). Also significant were cancer-cause mortality (model four HR 1·17, 95% CI 1·16–1·17; p<0·0001) and cardiovascular disease mortality (model four HR 1·36, 95% CI 1·35–1·36; p<0·0001). Kaplan–Meier curves showed similar patterns to those in cohort one (appendix pp 76–78).

In cohort four, the ClinicalAge gap consistently outperformed eFI in predicting all-cause, cancer-specific, and cardiovascular disease mortality. For all-cause mortality, the ClinicalAge gap had a higher C-index than eFI in the validation set across models (model one 0·73 [95% CI 0·72–0·73] *vs* 0·68 [0·67–0·69]; model four 0·75 [0·74–0·76] *vs* 0·69 [0·68–0·70]). Similarly, for cancer-related mortality, ClinicalAge gap showed better prediction performance than eFI (model four 0·72 [0·70–0·73] *vs* 0·68 [0·67–0·70]). Calibration varied across models. ClinicalAge gap model three had calibration slopes closest to 1 across all three mortality outcomes. Full discrimination and calibration results for models one, three, and four are reported in the appendix (pp 65–66, 79–81).

We included 2 054 608 participants in cohort three, with a median follow-up of 15·5 years (IQR 11·5–20·5). Trajectory analyses showed dynamic but structured shifts in ClinicalAge gap categories across midlife (figure 3A). The four trajectory groups reflect distinct patterns of ageing progression derived from longitudinal EHRs: group one (ClinicalAge gap less than or equal to mean minus SD; ie, clinically younger); group two (ClinicalAge gap more than mean minus SD but less than or equal to the mean; ie, slightly clinically younger or average); group three (ClinicalAge gap more than the mean but less than or equal to mean plus SD; ie, slightly clinically older); and group four (ClinicalAge gap more than mean plus SD; ie, clinically older).

At age 40 years, individuals in group one (clinically younger; ie, ClinicalAge gap ≤–2·63) were most likely to remain clinically younger at age 45 years: 124 674 (55·2%) of 226 050 participants remained in group one, whereas only 1877 (0·8%) of 226 050 participants transitioned to group four (clinically older). Conversely, among individuals in group four at age 40 years, 140 111 (63·4%) of 221 014 participants remained in group four by age 45 and only 255 (0·1%) of 221 014 participants shifted to group one. A similar pattern was observed from ages 45 years to 50 years (figure 3A). Overall, group membership became increasingly stable with age, with the clinically older group (group four) showing the greatest persistence. Thresholds for each ClinicalAge gap group at each age milestone are reported in the appendix (p 67).

To identify early risk factors for unfavourable ClinicalAge trajectories, we examined individuals who transitioned from group one to group four over 5 years and 10 years (figure 3B; appendix pp 82–85). Between ages 40 years and 45 years, the most common newly recorded conditions in this worsening subgroup were type 2 diabetes (863 [46·0%] of 1877 participants) and tobacco-use disorders (432 [23·0%] of 1877 participants). Between 45 years and 50 years, patterns were similar, with an increased prevalence of diabetes, hypertension, tobacco-use disorders, nerve compression or entrapment, and major depressive disorder.

For a 25-year perspective, figure 3C and the appendix (p 86) compare individuals who persistently remained in group one versus group four from ages 40 years to 65 years. Notably, 5004 women and only 86 men remained in the

persistently clinically younger group, although comorbidity patterns were broadly similar by sex. Among men who persistently remained in group four, the proportions of men with type 2 diabetes, essential hypertension, and tobacco-use disorders were 70, 65, and 40 percentage points higher, respectively, than those among men persistently in group one. Among women, the corresponding differences for type 2 diabetes and essential hypertension were 77 and 68 percentage points, respectively. Compared with those persistently in group one, individuals persistently in group four carried substantially higher burdens between ages 40 years and 65 years of metabolic and chronic physical disease.

We compared ClinicalAge gap trajectories from age 40 years to 65 years between individuals diagnosed with 31 ageing-related diseases and the general population without that disease (figure 4; appendix pp 87–96). Among the non-cancer diseases, diabetes consistently showed some of the largest deviations. Individuals with diabetes had markedly higher ClinicalAge gaps than individuals without diabetes from as early as age 40 years, with the gap widening steadily across midlife and almost reaching 2 years by age 65 years.

Across non-cancer diseases, individuals generally showed higher ClinicalAge gaps compared with the general population at all ages. These differences were already evident at age 40 years among individuals diagnosed before age 50 years and persisted through age 65 years. By contrast, for all cancers combined, ClinicalAge gaps diverged from those of cancer-free peers among individuals diagnosed with cancer after age 50 years, and individuals diagnosed before age 50 years generally had ageing trajectories that remained closer to cancer-free peers between ages 40 years and 65 years.

Beyond these group-level patterns, we performed an exploratory individual-level analysis using diabetes as an exemplar, given its consistently strong signal across our analyses. 5-year encounter-level trajectories of randomly sampled individuals revealed smoothly evolving and clinically interpretable patterns, with ClinicalAge gaps progressively increasing as the disease advanced (appendix pp 12–13).

	Cohort one: UK Biobank (n=320 891)		Cohort two: CPRD (n=4 728 376)	
	HR (95% CI)	p value	HR (95% CI)	p value
All-cause mortality				
Model one	1·40 (1·38–1·41)	<0·0001	1·63 (1·62–1·63)	<0·0001
Model two	1·22 (1·20–1·23)	<0·0001	NA	NA
Model three	1·29 (1·28–1·31)	<0·0001	1·46 (1·46–1·46)	<0·0001
Model four	1·23 (1·21–1·24)	<0·0001	1·36 (1·36–1·36)	<0·0001
Cancer-cause mortality				
Model one	1·26 (1·24–1·28)	<0·0001	1·42 (1·42–1·43)	<0·0001
Model two	1·11 (1·09–1·12)	<0·0001	NA	NA
Model three	1·18 (1·17–1·20)	<0·0001	1·27 (1·26–1·27)	<0·0001
Model four	1·12 (1·10–1·14)	<0·0001	1·17 (1·16–1·17)	<0·0001
Cardiovascular disease-cause mortality				
Model one	1·59 (1·56–1·63)	<0·0001	1·86 (1·86–1·87)	<0·0001
Model two	1·37 (1·34–1·40)	<0·0001	NA	NA
Model three	1·44 (1·41–1·47)	<0·0001	1·65 (1·64–1·65)	<0·0001
Model four	1·33 (1·29–1·36)	<0·0001	1·36 (1·35–1·36)	<0·0001

HRs are per 1-year increase in ClinicalAge gap. Model one: unadjusted. Model two: adjusted for baseline PhenoAge. Model three: adjusted for baseline chronological age and sex. Model four: fully adjusted for baseline chronological age, sex, BMI, systolic blood pressure, HDL and LDL cholesterol, smoking status, index of multiple deprivation, and ethnicity. Traumatic causes of death (ICD-10 V01–Y98) accounted for 3·0% of deaths in UK Biobank and 1·0% of deaths in UK CPRD. CPRD=Clinical Practice Research Datalink. HR=hazard ratio. NA=not applicable.

Table 2: **Associations between ClinicalAge gap and 10-year mortality risk in UK Biobank and CPRD**

Discussion

We present ClinicalAge, an ageing measurement constructed from routinely collected EHRs. We used a transformer-based model trained on longitudinal EHR data; it summarises dynamic health trajectories into a continuous ageing metric. It offers a pragmatic complement to molecular ageing clocks (eg, DNA methylation and proteomics),[8,9] capturing temporal physiological changes without biospecimen assays. It is non-invasive and potentially easily scalable across health systems, enabling population-level ageing surveillance.

ClinicalAge provides an interpretable measure that can be embedded into clinical workflows to identify individuals at risk of accelerated ageing. Its construct validity is supported by strong associations with established biological ageing markers (eg, creatinine, IGF-1, albumin, telomere length) and functional measures (eg, frailty, grip strength, and fluid intelligence). Compared with existing indices such as the eFI,[19] ClinicalAge captures longitudinal ageing dynamics and shows better discrimination across mortality outcomes. Traditional indices such as the Charlson[30] and Elixhauser[31] scores, which quantify the burden of coexisting medical conditions, do not explicitly model longitudinal ageing trajectory. Whereas models such as Delphi-2M[32] predict the occurrence and timing of diseases, ClinicalAge learns ageing dynamics to provide an interpretable, transferable, and easily communicated measure for clinical use.

ClinicalAge captures modifiable signals of ageing. It is sensitive to behavioural, social, and health system interactions that shape physiological ageing trajectories. Individuals with accelerated ClinicalAge showed increased prevalence of type 2 diabetes, hypertension, obesity, tobacco use, and depression—most conditions with well-established preventive and therapeutic interventions.[33] Its anticipatory signal could enable early, targeted action, particularly in midlife (ages 40–50 years) when trajectories begin to diverge: those transitioning from slow to fast agers during this window carried an increased burden of cardiometabolic and mental health conditions. Lifestyle factors (eg, tobacco and alcohol-use disorders) were especially influential in midlife, whereas chronic diseases (eg, diabetes and hypertension) became more prominent over longer horizons. The model's sensitivity to metabolic and renal dysfunction likely reflects the PhenoAge training target, which includes biomarkers such as creatinine.

ClinicalAge reveals marked differences across diseases. Chronic diseases such as Alzheimer's disease, chronic



(Figure 3 continues on next page)

kidney disease, chronic obstructive pulmonary disease, and heart failure were strongly associated with elevated ClinicalAge gap, often detectable within the decade before diagnosis, consistent with hallmarks of ageing (eg, chronic inflammation, metabolic dysfunction, and multisystem burden).[1,34] While it was positive for most cancers, inverse associations appeared for brain, prostate, and breast cancer. For prostate and breast cancer, age-related screening and competing risk or ascertainment effects likely explain the inverse pattern rather than slowed biological ageing; the inverse association for brain cancer was small and could reflect survival bias.

Another key finding is that non-cancer chronic diseases are associated with sustained, lifelong divergence in the ClinicalAge gap, whereas, for all cancers combined, diagnoses after age 50 years were associated with later-life divergence. This difference suggests fundamentally distinct patterns of ageing burden: while non-cancer disease reflects cumulative, progressive damage, cancer-related ageing could be treatment-related, and reversible in individuals who remained alive during follow-up following a cancer diagnosis.

Our model suggested a preclinical ageing burden, consistent with the concept of age-related vulnerability, whereby reduced physiological reserve increases susceptibility to health stressors.[35] Recognising and targeting this burden could help shift health care from a disease-centred paradigm to an ageing-centred paradigm.

This study has limitations. Our findings, validated in two UK cohorts, might not fully generalise to other populations or health systems, and EHR quality and completeness vary across settings. EHRs undercapture domains such as resilience, cognition, and social determinants, as these are not routinely assessed or consistently recorded in a structured format. Medication data were largely limited to GP prescriptions, with hospital-administered treatments incompletely recorded, which could affect estimates for some cancers. Overall cancer incidence is lower in UK Biobank than in the general population,[36] although results were consistent in the nationally representative CPRD Aurum cohort, supporting robustness of ClinicalAge. Trajectory analyses required continuous records from age 40 years until age 65 years or the end of available records, potentially reducing representativeness and biasing later trajectories towards individuals with sustained follow-up. Using PhenoAge as the training target could accentuate metabolic or renal signals; alternative ageing targets could yield different models and warrant evaluation. Finally,



Figure 3: **Trajectory patterns and clinical drivers of ageing in cohort three**

(A) Transitions in ClinicalAge gap groups across milestones (ages 40 years to 65 years). (B) Top 15 ICD-coded clinical features among individuals whose ClinicalAge gap shifted from group one to group four between ages 40 years to 45 years and 45 years to 50 years. (C) Heatmap of the most distinguishing clinical features in participants who persistently remained in group one versus group four from ages 40 years to 65 years, stratified by sex. Group one (ClinicalAge gap less than or equal to mean minus SD; ie, clinically younger); group two (ClinicalAge gap more than mean minus SD but less than or equal to the mean; ie, slightly clinically younger or average); group three (ClinicalAge gap more than the mean but less than or equal to mean plus SD; ie, slightly clinically older); and group four (ClinicalAge gap more than mean plus SD; ie, clinically older).

although surveillance and detection bias are inherent to EHRs, our age-anchored design, which recomputes ClinicalAge at each milestone using all history, reduces long-run differential bias.

This work should be regarded as a proof-of-concept study. Future studies will extend this framework by integrating lifestyle, behavioural, and environmental data, including physical activity, diet, and alcohol use patterns, to improve understanding of the determinants and modifiability of divergent ageing trajectories and to evaluate their clinical actionability. Moreover, future work will expand the age range to include adults younger than 40 years and adults older than 65 years, enabling assessment of ageing dynamics across the full adult lifespan.

ClinicalAge could be embedded in EHRs as a dynamic measurement for clinicians and patients. By modelling longitudinal trajectories, it provides trajectory-level specificity, distinguishing stable, accelerating, and resilient ageing within individuals, which is expressed as an

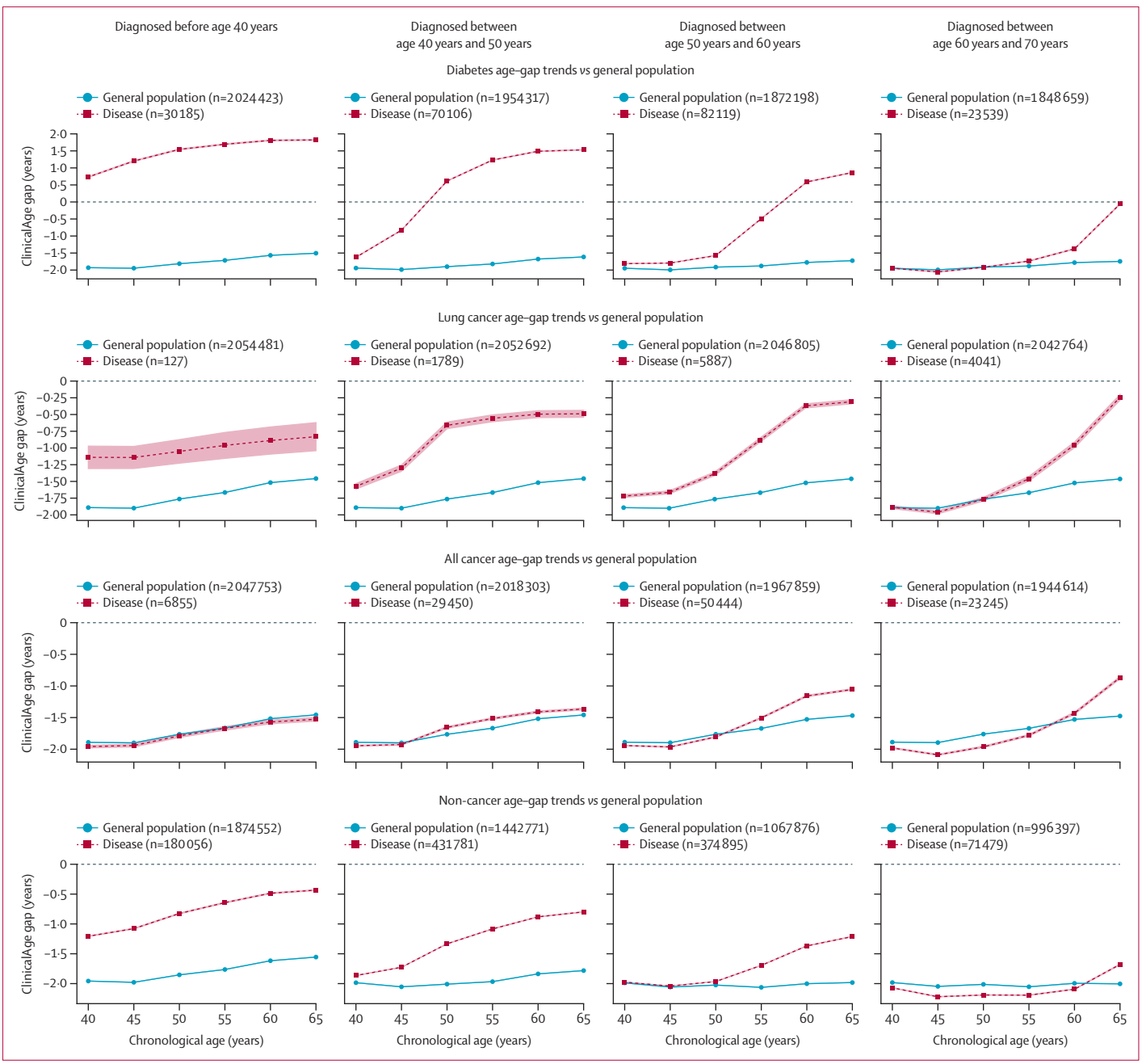

Figure 4: **Disease-specific ageing trajectories based on ClinicalAge gap**

Age-gap trends comparing individuals diagnosed with specific diseases to the general population (without this disease) from ages 40 years to 65 years. Each panel shows the mean ClinicalAge gap by chronological age for participants stratified by disease onset window: before age 40 years and age 40–50 years, 50–60 years, and 60–70 years. Results are shown for diabetes, lung cancer, all cancers combined, and all non-cancer chronic diseases. Red lines represent disease groups; blue lines represent general population controls. Shaded areas indicate 95% CIs.

interpretable age gap (years). Periodic updates could support timing-sensitive prevention (eg, a person aged 45 years whose ClinicalAge gap increases from +1 year to +4 years between the ages of 40 years and 45 years triggering an early review of metabolic risk factors, mood, and medication burden), enable proactive population stratification and recall prioritisation (eg, flagging those with the highest or fastest-rising gaps for early follow-up), and act as an adjunct to long-term-condition clinics (eg, persistently high gaps in people with diabetes or kidney disease prompting care intensification or multidisciplinary review). ClinicalAge can also facilitate risk communication and shared decision making by translating complex longitudinal histories into a simple age-gap summary. Its interpretability

and potentially low implementation cost make it suitable for both population monitoring and individualised longitudinal care.

In conclusion, ClinicalAge is an interpretable and generalisable measure of ageing derived entirely from routinely collected clinical data. It could bridge ageing science with practical implementation, enabling early risk stratification, precision prevention, and the integration of ageing science into routine health care. As ageing becomes the dominant determinant of morbidity and mortality worldwide, EHR-based tools such as ClinicalAge could support a new era of population-level ageing surveillance and intervention.

Contributors

JL contributed to conceptualisation, data curation, formal analysis, investigation, methodology, project administration, supervision, validation, visualisation, and writing of the original draft, as well as review and editing of the manuscript. ZF contributed to conceptualisation, data curation, formal analysis, investigation, methodology, validation, visualisation, and writing of the original draft, as well as review and editing of the manuscript. WF and SR contributed to investigation, methodology, and writing of the original draft, as well as review and editing of the manuscript. QY and GZ contributed to investigation, validation, visualisation, and writing of the original draft, as well as review and editing of the manuscript. BOP, NA, ZB, MN, RD, AC, and MW contributed to investigation and writing of the original draft, as well as review and editing of the manuscript. KR contributed to conceptualisation, funding acquisition, project administration, resources, supervision, validation, and writing of the original draft, as well as review and editing of the manuscript. All authors had full access to all data used in the study and had final responsibility for the decision to submit for publication. JL and ZF directly accessed and verified the underlying data.

Declaration of interests

KR reports grants paid to his institution from the National Institute for Health Research, Medical Research Council (MRC), British Heart Foundation, the EU, Roche, and the Novo Nordisk Oxford Big Data Partnership. He reports royalties or licenses from Lucem Health (personal and institution); personal fees from Radcliffe Cardiology (speaker); and personal fees for his role as Editor-in-Chief of Heart. He also serves on a Medtronic advisory board for renal denervation (institution). SR reports support from the Medical Research Council and Oxford University Hospitals National Health Service (NHS) Trust (paid to institution). He reports consulting fees from Lucem Health (personal) and honoraria from BMJ Heart journal (personal). ZF reports grants from the British Heart Foundation. MN reports receiving reimbursement and honoraria from AstraZeneca, Nemysis, and Albus Health outside the submitted work. He also serves as a statistical adviser for the Heart journal. AC reports support from a National Institute for Health and Care Research (NIHR) Research Professorship award, the NIHR Applied Research Collaboration Yorkshire & Humber, the NIHR Leeds Biomedical Research Centre, the Cross-NIHR Collaboration in Multiple Long-Term Conditions, and Health Data Research UK, outside the submitted work. AC led the development and UK implementation of the electronic frailty index (eFI), which is licensed to suppliers of electronic health record systems at no cost, provided no premium charge is applied to the end NHS user. AC reports consulting activity for the Geras Centre for Aging Research and support for attending a meeting from the Australia and New Zealand Society of Geriatric Medicine. AC has served on trial steering committees and data monitoring and ethics committees for NIHR trials and holds leadership or committee roles in the global Ageing Research Trialists collaborative, National Institute for Health and Care Excellence guideline development, NIHR funding committees, the Dunhill Medical Trust, and the MRC. All other authors declare no conflict of interests.

Data sharing

Access to Clinical Practice Research Datalink (CPRD) data, including UK primary care records and linked datasets such as hospital episode statistics, is subject to approval through CPRD's Research Data Governance Process (https://www.cprd.com/research-applications). The UK Biobank data used in this study were obtained under approved application number 83942 and are available to qualified researchers through UK Biobank's data access procedures (https://www.ukbiobank.ac.uk/use-our-data/). The code used for model training will be publicly available at: https://github.com/SereneLian/ClinicalAge.

Acknowledgments

During the preparation of this work, the authors used ChatGPT (GPT-4o and GPT-5 Thinking; OpenAI) for code debugging, code checking, language proofreading, and grammar checking. The authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

References

1 López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell* 2013; **153:** 1194–217.
2 Niccoli T, Partridge L. Ageing as a risk factor for disease. *Curr Biol* 2012; **22:** R741–52.
3 Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature* 2018; **561:** 45–56.
4 Cooper R, Kuh D, Hardy R, the Mortality Review Group, and the FALCon and HALCyon Study Teams. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. *BMJ* 2010; **341:** c4467.
5 Blackburn EH, Epel ES, Lin J. Human telomere biology: a contributory and interactive factor in aging, disease risks, and protection. *Science* 2015; **350:** 1193–98.
6 Franceschi C, Campisi J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J Gerontol A Biol Sci Med Sci* 2014; **69** (suppl 1)**:** S4–9.
7 Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* 2018; **19:** 371–84.
8 Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)* 2019; **11:** 303–27.
9 Argentieri MA, Xiao S, Bennett D, et al. Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nat Med* 2024; **30:** 2450–60.
10 Rutledge J, Oh H, Wyss-Coray T. Measuring biological age using omics data. *Nat Rev Genet* 2022; **23:** 715–27.
11 Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 2018; **10:** 573–91.
12 Lu AT, Binder AM, Zhang J, et al. DNA methylation GrimAge version 2. *Aging (Albany NY)* 2022; **14:** 9484–549.
13 Mak JKL, McMurran CE, Kuja-Halkola R, et al. Clinical biomarker-based biological aging and risk of cancer in the UK Biobank. *Br J Cancer* 2023; **129:** 94–103.
14 Liu Z, Kuo P-L, Horvath S, Crimmins E, Ferrucci L, Levine M. A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: a cohort study. *PLoS Med* 2018; **15:** e1002718.
15 Belsky DW, Caspi A, Corcoran DL, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* 2022; **11:** e73420.
16 Brinkley TE, Justice JN, Basu S, et al. Research priorities for measuring biologic age: summary and future directions from the Research Centers Collaborative Network Workshop. *Geroscience* 2022; **44:** 2573–83.
17 Teschendorff AE, Horvath S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. *Nat Rev Genet* 2025; **26:** 350–68.
18 Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44:** 827–36.
19 Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing* 2016; **45:** 353–60.

20 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12:** e1001779.

21 Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019; **48:** 1740–1740g.

22 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv* 2017; published online June 12. https://doi.org/10.48550/arXiv.1706.03762 (preprint).

23 Rao S, Li Y, Mamouei M, et al. Refined selection of individuals for preventive cardiovascular disease treatment with a transformer-based risk model. *Lancet Digit Health* 2025; **7:** 100873.

24 Fan Z, Mamouei M, Li Y, Rao S, Rahimi K. Identification of heart failure subtypes using transformer-based deep learning modelling: a population-based study of 379 108 individuals. *EBioMedicine* 2025; **114:** 105657.

25 Williams DM, Jylhävä J, Pedersen NL, Hägg S. A frailty index for UK Biobank participants. *J Gerontol A Biol Sci Med Sci* 2019; **74:** 582–87.

26 Junnila RK, List EO, Berryman DE, Murrey JW, Kopchick JJ. The GH/IGF-1 axis in ageing and longevity. *Nat Rev Endocrinol* 2013; **9:** 366–76.

27 Conrad N, Misra S, Verbakel JY, et al. Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. *Lancet* 2023; **401:** 1878–90.

28 Conrad N, Verbeke G, Molenberghs G, et al. Autoimmune diseases and cardiovascular risk: a population-based study on 19 autoimmune diseases and 12 cardiovascular diseases in 22 million individuals in the UK. *Lancet* 2022; **400:** 733–43.

29 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017; **357:** j2099.

30 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987; **40:** 373–83.

31 Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998; **36:** 8–27.

32 Shmatko A, Jung AW, Gaurav K, et al. Learning the natural history of human disease with generative transformers. *Nature* 2025; **647:** 248–56.

33 Vodovotz Y, Barnard N, Hu FB, et al. Prioritized research for the prevention, treatment, and reversal of chronic disease: recommendations from the Lifestyle Medicine Research Summit. *Front Med (Lausanne)* 2020; **7:** 585744.

34 Franceschi C, Capri M, Monti D, et al. Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech Ageing Dev* 2007; **128:** 92–105.

35 Whitson HE, Cohen HJ, Schmader KE, Morey MC, Kuchel G, Colon-Emeric CS. Physical resilience: not simply the opposite of frailty. *J Am Geriatr Soc* 2018; **66:** 1459–61.

36 Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186:** 1026–34.