Biological age in UK Biobank: biomarker composition and prediction of mortality, coronary heart disease and hospital admissions
(Short title: Biological age biomarker composition and outcome prediction)

Authors: Mei Sum Chan¹, Matthew Arnold¹,², Alison Offer¹, Imen Hammami¹, Marion Mafham¹, Jane Armitage¹,³, Rafael Perera⁴, Sarah Parish¹,³

Affiliations:
¹Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
²British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom
³MRC Population Health Research Unit, University of Oxford, Oxford, United Kingdom
⁴Nuffield Department of Primary Health Care Sciences, University of Oxford, Oxford, United Kingdom
Abstract

Background: Age is the strongest risk factor for most chronic diseases, and yet individuals may age at different rates biologically. A biological age formed from biomarkers may be a stronger risk factor than chronological age and understanding what factors contribute to it could provide insight into new opportunities for disease prevention.

Methods and findings: Among 480,019 UK Biobank participants aged 40-70 recruited in 2006-2010 and followed up for 6-12 years via linked death registry and secondary care records, a subpopulation of 141,254 (29.4%) non-smoking adults in good health and with no medication use or disease history at baseline were identified. Independent components of 72 biomarkers measured at baseline were characterised by principal component analysis. The Klemera Doubal method (KDM), which derived a weighted sum of biomarker principal components based on the strengths of their linear associations with chronological age, was used to derive sex-specific biological ages in this healthy subpopulation. The proportions of the overall biological and chronological age effects on mortality, coronary heart disease and age-related non-fatal hospital admissions (based on a hospital frailty index) that were explained by biological age were assessed using log-likelihoods of proportional hazards models.

Reduced lung function, reduced kidney function, slower reaction time, lower insulin-like-growth factor 1, lower hand grip strength and higher blood pressure were key contributors to biological age (explaining the highest percentages of its variance) in both men and women, while lower albumin, higher sex hormone-binding globulin and lower muscle mass in men, and higher liver enzymes, blood lipids and HbA1c in women were also important. Across both sexes, a 51-principal component biological age explained 66%, 80% and 63% of the age effects on mortality, coronary heart disease and hospital admissions, respectively. Restricting
the biological age to the 12-13 key biomarkers corresponding to the 10 most importantly
contributing principal components resulted in little change in these proportions for women,
but a reduction to 53%, 63% and 50%, respectively, for men.

Conclusions: This study identified that markers of impaired function in a range of organs
account for a substantial proportion of the apparent effect of age on disease and hospital
admissions. It supports a broader, multi-system approach to research and prevention of
diseases of ageing.

Main text

Introduction

Age is the strongest risk factor for most chronic diseases that limit healthy lifespan, yet
individuals may age biologically at different rates,\(^1\) illustrated by differential rates of disease
accumulation and frailty onset. Understanding the contributors to biological ageing could
lead to opportunities for early prevention of later life disease\(^2\) and limiting the disease burden
associated with growing elderly populations.

‘Biological age’ is often used to refer to a composite risk score estimated from measurements
of biomarkers of ageing,\(^1,3-5\) which may be constructed from different types and numbers of
biomarkers.\(^5-7\) Biological ages constructed from clinical biomarkers widely measured in
routine care were found to be more predictive of health outcomes than those constructed
using epigenetic ages\(^2,6\) and telomere length.\(^6\) The most important contributors to biological
age have varied between studies, in part due to differences in the set of biomarkers
available.\(^8,9\)
The main methods of estimating biological ages using clinical biomarkers are multiple linear regression, principal component analysis (PCA) and the Klemara Doubal method (KDM), which derives a weighted sum of biomarkers based on the strengths of their associations with chronological age\(^{10}\) and is favoured by comparison studies.\(^{9,11,12}\) KDM allows comparison of the predictive value of its biological age with that of chronological age for health outcomes, when chronological age has not been included as a constituent ‘biomarker’. However, KDM biological ages have typically included chronological age\(^{13}\) and the predictive values of biological ages have only occasionally been compared with those of chronological age\(^{13}\) or mortality-based scores.\(^{2,14}\) Multiple linear regression has been used in many studies\(^ {5,9,12}\) and both multiple linear regression and KDM have been extended to principal components of biomarkers instead of individual biomarkers.\(^ {11}\)

The UK Biobank is a richly phenotyped resource with 0.5 million participants\(^ {15}\) that provides an unrivalled opportunity to investigate earlier stages of ageing through biological, lifestyle and environmental factors, compared to previous studies of typically 100-10,000 participants with panels of fewer than 30 biomarkers.\(^ {5}\) A substantial middle aged, apparently healthy subpopulation of the UK Biobank can be identified, to reduce reverse causality from prior health or medication use affecting biomarker levels.

This study aims to: (1) investigate and implement reliable methods of estimating biological age in the UK Biobank, (2) identify the main biomarker determinants of biological age, and (3) investigate the relationship between biological and chronological age in the prediction of health outcomes.

**Methods**

**Study population**
The UK Biobank recruited 0.5 million participants across the UK aged 40-70 for baseline assessment in 2006-2010. Information on sociodemographic characteristics, self-reported health behaviours and medication was recorded. Linkage to Hospital Episode Statistics (HES) and national death registries provided prior and prospective information on secondary care outcomes and date and cause of death. Over 100 biomarkers and physical attributes were measured using various measurement devices, and blood and urine assays (S1 Appendix 1 and 2). This study was covered by the general ethical approval for UK Biobank studies from the NHS National Research Ethics Service on 17th June 2011 (Ref 11/NW/0382).

After data cleaning (S1 Figure 1), 480,019 participants, followed up for a median of 8.7 years for death registry records and 8.0 years for HES records were included (S1 Appendix 1). A composite measure of prior health, derived from chronic disease-related medication count, walking speed, number of HES episodes prior to recruitment, smoking status, previous age-related chronic disease and hip or wrist fracture was used to stratify the population into 4 groups: (1) healthy (generally non-smokers with no prior disease or medications), (2) some medications, (3) slightly unhealthy, and (4) poor health (S1 Appendix 1). This study focused on the 141,254 people in the healthy subpopulation, in order to reduce reverse causality, and contrasted results for this subpopulation with those for the poor health subpopulation and the whole population.

**Statistical Analyses**

Among 110 available biomarkers, a panel of 72 biomarkers was obtained after excluding biomarkers that: (1) were missing in >30% participants, (2) were measuring the same biological trait as another biomarker, or (3) had poor reproducibility, assessed through intra-individual correlations of <0.1 in the 19,335 individuals with a repeat measurement (S1 Appendix 2 and S1 Table 3). For each sex and prior health subpopulation, trends of each
biomarker with chronological age were visually assessed for linearity (S1 Appendix 3A), before using linear methods to estimate biological ages. To represent the biomarkers as linearly uncorrelated principal components, PCA with varimax rotation\textsuperscript{11} was carried out on the 72 biomarkers, which gave 51 principal components with eigenvalues >0.33. These principal components were characterised based on constituent biomarkers with the largest factor loadings (S1 Appendix 3B–D).

Biological ages were estimated using KDM,\textsuperscript{10} separately within each sex and prior health subpopulation (S1 Appendix 3C). A similar method, stepwise linear regression, was considered and the results for both methods were compared in the appendices (S1 Appendix 3C and 4). Biomarker principal components were ranked by their importance, measured by the proportion of variance in the biological ages that each component explained (S1 Appendix 3E).

Three health outcomes were constructed from HES and death records: (1) death from chronic disease (excluding: infectious diseases, pregnancy, congenital malformations and external causes),\textsuperscript{16} (2) fatal and non-fatal coronary heart disease (CHD) and (3) age-related non-fatal hospital admissions (S1 Appendix 2). These hospital admissions are the subset of those types of admissions in a published hospital frailty risk score\textsuperscript{17} that are age-related in the UK Biobank (S1 Table 6). The predictive powers of chronological age and biological ages for each health outcome were assessed using Cox proportional hazards models adjusted for Index of Multiple Deprivation 2010 quintile, smoker status, alcohol consumption and assessment centre, and computing Harrell’s C-indices (measures of statistical discrimination similar to the area under the receiver operating curve). Prediction of CHD and hospital admissions by biological age was compared with a benchmark of prediction by a mortality score similar to those proposed by previous studies\textsuperscript{2,14} and derived from stepwise Cox regression, using unadjusted Cox models (S1 Appendix 3F).
To investigate the relationship of the biological ages to chronological age, the proportion of variation in chronological age described by each biological age was estimated. The proportion of the overall biological and chronological age effect on mortality, CHD and hospital admission risk that was explained by each biological age was also estimated, by comparing the log-likelihoods from these Cox models (S1 Appendix 3G). Calibration of biological ages to chronological age and the risk calibration of biological ages with each health outcome was assessed (S1 Appendix 3H).

The statistical analysis was repeated on biomarkers corresponding to the 10 most important biomarker principal components in the analyses of the panel of 51 biomarker principal components among healthy participants of each sex. The predictive and explanatory power for this reduced biomarker panel was compared to the full panel. Guidelines for Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)\(^\text{18}\) were followed (S1 Table 14).
Results

Study characteristics

Of the 480,019 participants, 141,254 (29.4%) were in the healthy subpopulation (Table 1). During a median follow up period of 8.7 years for mortality and 8.0 years for CHD and hospital admissions (S1 Appendix 1), 1.7% of healthy, and 3.9% of all participants died from chronic diseases; 1.9% of healthy and 4.0% of all participants without CHD at baseline had a first CHD event; 16.0% of healthy and 23.1% of all participants who were not admitted to hospital for age-related reasons prior to baseline had been admitted with diagnoses of these conditions during follow up (S1 Table 7). Sociodemographic patterns and the proportion of participants healthy at baseline were similar between sexes.
### Table 1: Participant characteristics at baseline assessment in 2006-2010, for the Healthy and Poor health subpopulations and the whole UK Biobank population

|                      | Healthy subpopulation | Poor health subpopulation | Whole population |
|----------------------|-----------------------|---------------------------|------------------|
|                      | Persons   | Men     | Women    | Persons   | Men     | Women    | Persons   | Men     | Women    |
| Participants (n)     | 141,254   | 65,869  | 75,385   | 82,835   | 42,277  | 40,558   | 480,019   | 219,248 | 260,771 |
| Person-years at risk (millions) | 1.2       | 0.56    | 0.64     | 0.71     | 0.36    | 0.35     | 4.12      | 1.87    | 2.25    |
| Deaths during follow up (%) | 1.7       | 2.1     | 1.4      | 9.1      | 11.2    | 7.0      | 3.9       | 5.2     | 2.9     |
| Median age at baseline (years) | 56.0      | 55.7    | 56.4     | 62.2     | 62.5    | 61.8     | 58.3      | 58.8    | 58.0    |
| Age band at baseline in years (%) | 30.1      | 30.1    | 30.0     | 30.2     | 30.0    | 30.4     | 24.3      | 24.3    | 24.3    |
| Index of Multiple Deprivation quintile (%) | Q1 (least deprived) | 23.9 | 23.8 | 24.1 | 16.9 | 16.7 | 17.1 | 20.0 | 19.7 | 20.1 |
|                      | Q2        | 22.2    | 22.1    | 22.3    | 18.3    | 18.2    | 18.3     | 20.0    | 19.7    | 20.3    |
|                      | Q3        | 20.9    | 20.7    | 21.1    | 19.3    | 19.2    | 19.4     | 20.0    | 19.8    | 20.2    |
|                      | Q4        | 18.7    | 18.7    | 18.7    | 21.0    | 20.8    | 21.1     | 20.0    | 19.9    | 20.1    |
|                      | Q5 (most deprived) | 14.3 | 14.7 | 13.9 | 24.6 | 25.1 | 24.0 | 20.0 | 20.9 | 19.3 |
| Smoker status (%)    | Current   | 0.0     | 0.0     | 0.0     | 11.0    | 12.5    | 9.6      | 10.5    | 12.4    | 8.8      |
|                      | Previous  | 33.4    | 35.9    | 31.2    | 41.6    | 47.1    | 35.9     | 34.5    | 38.3    | 31.3     |
|                      | Never     | 66.6    | 64.1    | 68.8    | 46.6    | 39.7    | 53.7     | 54.5    | 48.8    | 59.4     |
| Alcohol consumption frequency (%) | Never | 5.5     | 4.5     | 6.5     | 11.5    | 8.8     | 14.2     | 8.0     | 6.3     | 9.5      |
|                      | Special occasions only | 8.8 | 5.6     | 11.6    | 14.1    | 9.3     | 19.1     | 11.5    | 7.3     | 15.0     |
|                      | One to three times a month | 10.5 | 8.6     | 12.3    | 10.8    | 9.1     | 12.6     | 11.1    | 8.9     | 13.0     |
|                      | Once or twice a week | 27.3    | 27.2    | 27.4    | 24.5    | 25.5    | 23.5     | 25.8    | 25.9    | 25.7     |
|                      | Three or four times a week | 26.8 | 29.5    | 24.5    | 19.8    | 23.1    | 16.4     | 23.1    | 26.1    | 20.5     |
|                      | Daily or almost daily | 21.0    | 24.7    | 17.7    | 19.0    | 23.8    | 14.0     | 20.3    | 25.3    | 16.1     |
|                      | No answer/missing | 0.0     | 0.0     | 0.0     | 0.3     | 0.3     | 0.3      | 0.2     | 0.2     | 0.2      |
| Body mass index in kg/m² (%) | <18.5    | 0.6     | 0.2     | 0.9     | 0.5     | 0.3     | 0.7      | 0.5     | 0.2     | 0.8      |
|                      | 18.5-24.9 | 41.7   | 32.6    | 49.6    | 24.5    | 19.6    | 29.7     | 32.8    | 25.2    | 39.2     |
|                      | 25-29.9   | 43.1    | 51.6    | 35.8    | 41.5    | 46.2    | 36.6     | 42.4    | 49.2    | 36.6     |
|                      | 30-39.9   | 14.1    | 15.3    | 13.0    | 30.1    | 31.4    | 28.7     | 22.4    | 24.0    | 21.0     |
|                      | 40+       | 0.5     | 0.3     | 0.7     | 3.4     | 2.5     | 4.3      | 1.9     | 1.3     | 2.4      |
Biomarker characteristics

The relationships of most candidate biomarkers to chronological age were broadly linear or flat (S1 Figure 3 and Figure 1). Several biomarkers displaying non-linear trends and differences by sex or by prior health are highlighted in Figure 1, which shows standardised biomarker levels for comparability between biomarkers. Lung function biomarkers, systolic blood pressure, cystatin C and reaction time had the strongest linear relationships with age, while diastolic blood pressure, body mass index (BMI) and low density lipoprotein cholesterol (LDL-C) had clear inverse U-shaped relationships with age that were attenuated in the healthy subpopulation. Diastolic blood pressure peaked or plateaued at different ages and BMI displayed different trends in the healthy and poor health subpopulation. Biomarkers that displayed substantially different trends between sexes were heel bone density, LDL-C, calcium, alkaline phosphatase and phosphate (S1 Figure 3).

Many biomarker principal components had a single biomarker strongly loaded onto them and were easily characterised. Multiple biomarkers were strongly loaded onto the adiposity, lung function, blood pressure and blood lipid principal components (S1 Figure 5). The coefficients of the biomarker principal components in the biological ages for the healthy subpopulation are listed in S1 Table 9. The estimated biological ages appeared stable in 10-fold cross validation as their prediction errors were small (S1 Appendix 4 and S1 Figure 6).
Biomarkers

Forced expiratory volume in 1s/height (FEV1/height)
Characteristics:
Linear decline with age, with the strongest relation to age of all biomarkers

Systolic blood pressure
Characteristics:
Broadly linear increase with age, but at a faster rate for healthy vs poor health individuals

Diastolic blood pressure
Characteristics:
Inverse U-shaped trend with age in poor health, peaking or plateauing at a later age for healthy vs poor health individuals
Biomarkers

Body mass index (BMI)

**Characteristics:**
Apparent differences in trends by prior health status: inverse U-shaped trend with age for poor health men, but broadly flat for healthy men.

Heel bone density

**Characteristics:**
No apparent difference in the level or trend of the biomarker with age between healthy and poor health individuals. Apparent differences in trends between men and women: a steeper decline with age is seen in women.

Low density lipoprotein cholesterol

**Characteristics:**
Apparent differences in trends and turning points, by prior health status and sex: levels in poor health individuals decline, and diverge from those of healthy individuals, at younger ages for men compared to women.

**Figure 1:** Biomarker-age trends (means and standard errors for each 2.5-year age group) for the healthy and poor health subpopulations and descriptions of their characteristics, for six selected biomarkers and by sex.
Predictive power and calibration of biological ages

The KDM biological ages (based on 51 principal components) were well calibrated as they matched healthy participants’ chronological ages on average (S1 Figure 7). The KDM ages were more predictive of CHD and hospital admissions than the benchmark mortality score (approximate increases in C-indices for CHD/hospital admissions: 0.135/0.111 in men, 0.109/0.068 in women), and the mortality score performed only slightly better than chance (C-indices ~0.5; S1 Table 12). KDM ages alone were not statistically significantly better than chronological age and did not statistically significantly supplement chronological age alone in predicting the health outcomes in the healthy subpopulation (increase in C-indices for mortality/CHD/hospital admissions: 0.007/0.006/0.002 in men, 0.002/0.023/0.001 in women; S1 Table 13). However, when estimated in unhealthier subpopulations, they supplemented chronological age alone in the prediction of mortality and CHD in these subpopulations (for the whole population, increase in C-indices: 0.031/0.015 in men, 0.014/0.028 in women; S1 Table 13).

The stepwise regression age was poorly calibrated as it was distributed across a narrower age range than chronological age on average (S1 Figure 7), and was not considered in further analyses. The prediction results and a risk calibration assessment are described in detail in S1 Appendix 4.

Biomarker importance in biological ages

In the KDM ages, reduced lung function featured most strongly in the healthy subpopulation (Figure 2), describing 12.4% (men) and 10.3% (women) of the variation in biological age (S1 Table 10). Higher cystatin C, slower reaction time, lower insulin-like growth factor-1 (IGF-1), lower hand grip strength, higher and higher blood pressure also featured strongly for both sexes; while lower albumin, higher sex hormone-binding globulin and lower muscle mass...
biomarkers featured strongly for men; and higher levels of alkaline phosphatase, LDL-C and apolipoprotein B and HbA1c for women. Multiple body systems were represented by these biomarkers: respiratory, renal, cardiovascular, musculoskeletal, endocrine, metabolic and immune, liver and nervous systems (S1 Table 5). When the analysis was restricted to the top 10 biomarker components corresponding to 13 biomarkers for men and 12 for women, forced expiratory volume in 1s/height, forced vital capacity/height, reaction time, IGF-1, cystatin C, hand grip strength/height, systolic and diastolic blood pressure in both sexes; albumin, sex hormone-binding globulin, fat-free mass, standing height and sitting height in men; and LDL-C, alkaline phosphatase, HbA1c and urea in women featured most strongly.

**Figure 2: Importances of the top 15 biomarker principal components in the Klemera Doubal (KDM) biological ages for healthy men (left) and women (right)**

**Relationship between biological and chronological age**

The KDM ages described 44.0% and 51.3% of the variation in chronological age for healthy men and women respectively. More importantly, with respect to the prediction of mortality, CHD and hospital admissions, and averaged across sexes, the KDM ages described 66%, 80% and 63% of the overall biological and chronological age effect respectively (Figure 3A and S1 Table 11). The proportion described by the KDM age is attributed to each constituent biomarker, broadly in the proportions of biomarker importances (Figure 2). Constructing the KDM age from the reduced panel of 12-13 biomarkers decreased the proportion explained by
biomarkers to 53%, 63% and 50% for each respective outcome in men, but made little
difference for women (Figure 3).
(A) Main analysis

![Chart showing contribution of biological and chronological age to health outcomes](chart1.png)

Contribution of either or both ages as a % of their total contribution

(B) Using the reduced biomarker panel

![Chart showing contribution of biological and chronological age to health outcomes](chart2.png)

Contribution of either or both ages as a % of their total contribution

Biological age alone  Chronological and biological age  Chronological age alone

Represented by biomarkers

**Figure 3:** Relative contribution of Klemara Doural (KDM) biological ages and chronological age in explaining each health outcome, in the (A) main analysis (top) and when (B) using the reduced biomarker panel (bottom), for healthy men (left) and women (right)

**Discussion**

This study found that biological ages consisting of markers of impaired function in a range of organs accounted for a substantial proportion of the apparent effect of chronological age on disease and hospital admissions.
Key biomarker determinants of biological ages

Lung, kidney, cognitive and liver function, IGF-1, hand grip strength and blood pressure were key contributors to biological ages for both sexes, while sex hormone-binding globulin and muscle mass in men, and cardiovascular function and HbA1c in women were also important.

The top-ranking biomarkers in this UK population generally matched those in a Singaporean study and slight variations by sex were seen in both populations. However, these lung and renal function biomarkers were not investigated in the study of Canadian, South Korean and Eastern European biological ages, which instead found that the top-ranking blood-based biomarkers varied by population and sex. Studies of ageing biomarkers also found that lung and renal biomarkers were top-ranking determinants of functional decline and variation in age-related traits.

The present study provides additional detail on the relative importance on ageing of biomarkers within body system groups, such as cystatin C over other renal biomarkers (creatinine and creatinine-based eGFR), as previous studies each assessed only one of these biomarkers.

Several key biomarkers in this study (blood pressure, blood lipids, height and lung function) have each been shown to be associated observationally, and in some cases causally, in randomised trials and Mendelian randomisation studies, with a range of age-related diseases (Table 2). Associations for other key determinants such as cystatin C and hand grip strength have been less extensively researched, and available studies have focused on mortality and cardiovascular outcomes. Blood pressure did not feature as strongly as the aforementioned biomarkers in our study, despite being well-established as a modifiable and causal risk factor of cardiovascular disease. Other cardiovascular biomarkers did not feature strongly for men, whereas for women, LDL-C and apolipoprotein B, causally linked to
atherosclerotic cardiovascular disease, were also important (Figure 2), and may have contributed to better prediction of CHD in women than in men (S1 Tables 12 and 13). Cohort effects in this population are difficult to disentangle, and may influence trends in body size.

Hence, height may be acting as a proxy for cohort effects.

Table 2: Examples of associations in published studies between the top 10 biomarker principal components of biological age identified in the present study and adverse health outcomes

| Key biomarker principal component in the present study | Evidence from meta-analyses of randomised trials | Evidence from Mendelian randomisation | Evidence from prospective studies |
|-------------------------------------------------------|-------------------------------------------------|---------------------------------------|----------------------------------|
| Lung function/height                                  | -                                               | Respiratory and autoimmune diseases   | All-cause, circulatory disease,   |
|                                                       |                                                 |                                       | respiratory and cancer mortality  |
| Cystatin C                                            | -                                               | (Not significantly associated with    | Cardiovascular disease, mortality |
|                                                       |                                                 | cardiovascular disease)               | and end-stage renal disease       |
| Reaction time test                                    | -                                               | -                                    | Mortality                        |
| IGF-1                                                 | Fracture risk                                   | (Not significantly associated with    | Mortality and heart failure and   |
|                                                       |                                                 | Alzheimer’s disease)                 | cognitive function)              |
| Hand grip strength/height                             | -                                               | Cardiovascular disease and            | Mortality, cardiovascular disease, |
|                                                       |                                                 | mortality                           | respiratory disease, cancer       |
| Blood pressure                                        | Mortality and cardiovascular disease            | Type 2 diabetes, Alzheimer’s disease | Vascular mortality (associations  |
|                                                       |                                                 |                                       | weaken with age)                  |
| Albumin (men only)                                    | -                                               | -                                    | Coronary heart disease           |
| Sex hormone-binding globulin (men only)               | -                                               | Type 2 diabetes                      | Type 2 diabetes                  |
| Muscle mass (men only)                                | -                                               | -                                    | Cancer mortality, physical       |
|                                                      |                                                 |                                       | disability                        |
| Height (men only)                                     | -                                               | Cardiovascular disease, hip fracture,| Mortality from cardiovascular   |
|                                                      |                                                 | intervertebral disc disorder,        | disease, liver disease, COPD,    |
|                                                      |                                                 | vasculitis, gastroesophageal reflux  | stomach and oral cancers, mental |
|                                                      |                                                 | disease                            | disorders                        |
| LDL-C and ApoB (women only)                           | Major vascular disease, vascular and all-cause  | Coronary heart disease               | Cardiovascular disease           |
|                                                      | mortality                                      |                                       |                                 |
| Alkaline phosphatase (women only)                     | -                                               | Type 2 diabetes (Associations with    | Osteosarcoma and cardiovascular   |
|                                                      |                                                 | IHD and T2D not robust after allowing | disease                           |
|                                                      |                                                 | for pleiotropy)                     |                                 |
| HbA1c (women only)                                    | (Not significantly associated with cardiovascular disease) | Coronary artery disease              | Mortality, cardiovascular disease, |
|                                                      |                                                 |                                       | cancer, diabetes                  |
| Urea (women only)                                     | -                                               | -                                    | Coronary heart disease           |
Relationships of biomarkers to age

Lung function biomarkers, systolic blood pressure, cystatin C and reaction time had the strongest linear relationships with age regardless of prior health, and therefore contributed substantially to variation in the KDM biological ages (Figure 2). The differences in trends for diastolic blood pressure, BMI and LDL-C by prior health were suggestive of reverse causal effects of disease or medication use on biomarker levels. BMI featured less strongly in the biological ages (the general adiposity component was 28th most important for men and 26th for women) potentially due to reverse causality, but has been causally linked to diseases. Therefore, BMI may be a modifiable risk factor that affects biological age.

Prediction in healthy versus unhealthy individuals

Knowledge of biological age is potentially more useful in apparently healthier individuals because in unhealthy individuals, disease and hospital admissions are diagnostic indicators of ageing, furthermore knowledge of risk of non-fatal outcomes is likely to provide a longer window for intervention and prevention than knowledge of mortality risk. Biological ages were better than the benchmark mortality score in predicting CHD and hospital admissions (S1 Table 12). The predictive value of a biological age varied by health status and was greater in unhealthy individuals (S1 Tables 12 and 13), likely reflecting the contribution of diagnostic indicators of ageing. Therefore, it is important to take into account the health and age profile of the population when comparing different studies.

The KDM permitted investigation of the relationship between biological and chronological ages with respect to predicting health outcomes, and automatically calibrated biological ages to chronological age. By contrast, the stepwise regression method resulted in a poorly-calibrated biological age in the UK Biobank despite its frequent use in biological ageing studies (where its risk calibration was not investigated). Comparison of an individual’s
biological age with their chronological age could provide a valuable means of communicating modifiable health risks, alongside their detailed biomarker profile. A prognostic biological age could also augment a national prevention programme promoting clinical biomarker screening in a middle-aged population, after causal factors underlying its constituents have been established. The most important biomarkers in biological age were measured via blood biochemistry measurements, spirometry and physical measurements, which can currently be administered quickly and simultaneously in clinical settings. For women, it may be suitable to measure just 12 key biomarkers across 7 body systems to assess biological ageing, as relatively little explanatory and predictive value was compromised. Despite the successful use of clinical risk prediction tools, ‘heart age’ and ‘lung age’ in clinical care, there is little evidence as yet of implementation of an overall biological age, and proposals to use it in drug development and clinical care may be longer term uses.

Strengths and limitations of this study

The estimation methods used assumed that biomarkers with the strongest linear relation to chronological age contribute most to biological age, but these biomarkers are not necessarily strongly linked to health outcomes. Analysis of key determinants of biological age were limited by the range of biomarkers available. Not all biomarker trends in the UK Biobank were linear (S1 Figure 3), but a previous study has shown that incorporating non-linearity was computationally complex and only slightly improved the accuracy of estimated biological age components. However, the epidemiological reliability of the present analyses was increased through stratification by prior health, use of biomarker principal components, cross validation and adherence to clinical risk prediction reporting guidelines (S1 Table 14).

Conclusions
A biological age consisting of clinical biomarkers reflecting functionality of a range of organs accounted for a substantial proportion of the effect of age on disease and hospital admissions in the UK Biobank. An overall biological age has potential to be used and evaluated as a broader-based approach to risk identification and prevention than individual biomarkers. Of the most important biomarkers contributing to the derived biological age, cardiometabolic biomarkers have well-studied causal associations with mortality and cardiovascular disease, but further research is needed to identify modifiable causal factors underlying all components, for a range of age-related diseases.

Acknowledgements

This research used the UK Biobank resource (application number 8835). We thank the participants of UK Biobank for their contribution to the resource. The work of MA was supported by core funding from: the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG/13/13/30194; RG/18/13/33946) and the National Institute for Health Research [Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust]. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Funding

The work was supported by grants to University of Oxford from the UK Medical Research Council through its funding of the MRC Population Health Research Unit (MRC_MC_U137686853) and the British Heart Foundation. MSC was funded by the Nuffield Department of Population Health. MA was funded by a British Heart Foundation Programme Grant (RG/18/13/33946). RP was funded by the NIHR Oxford Biomedical
Research Centre Program, the NIHR Collaboration for Leadership in Health Research and Care (CLARHC) Oxford, the NIHR Program for Applied Research, the NIHR Health Protection Research Unit Gastrointestinal Infections Group and the NIHR Diagnostic Evidence Co-operative. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests**

All authors have no competing interests to declare.

**Data Availability Statement**

The underlying data is open access through application to the UK Biobank, and materials and methods will be made freely available through the UK Biobank as part of this project.

**References**

1. Jylhava J, Pedersen NL, Hagg S. Biological Age Predictors. *EBioMedicine* 2017; **21**: 29-36.
2. 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**(12): e1002718.
3. Lara J, Cooper R, Nissan J, et al. A proposed panel of biomarkers of healthy ageing. *BMC Med* 2015; **13**: 222.
4. Zhang W, Jia L, Cai G, et al. Model Construction for Biological Age Based on a Cross-Sectional Study of a Healthy Chinese Han population. *Journal of Nutrition, Health & Aging* 2017; **21**(10): 1233-9.
5. Jia L, Zhang W, Chen X. Common methods of biological age estimation. *Clin Interv Aging* 2017; **12**: 759-72.
6. Belsky DW, Moffitt TE, Cohen AA, et al. Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing? *Am J Epidemiol* 2018; **187**(6): 1220-30.
7. Burkle A, Moreno-Villanueva M, Bernhard J, et al. MARK-AGE biomarkers of ageing. *Mech Ageing Dev* 2015; **151**: 2-12.
8. Mamoshina P, Kochetov K,Putin E, et al. Population specific biomarkers of human aging: a big data study using South Korean, Canadian and Eastern European patient populations. *J Gerontol A Biol Sci Med Sci* 2018.
9. Zhong X, Lu Y, Gao Q, et al. Estimating Biological Age in the Singapore Longitudinal Aging Study. *J Gerontol A Biol Sci Med Sci* 2019.
10. Klemera P, Doubal S. A new approach to the concept and computation of biological age. *Mech Ageing Dev* 2006; **127**(3): 240-8.
11. Cho IH, Park KS, Lim CJ. An empirical comparative study on biological age estimation algorithms with an application of Work Ability Index (WAI). *Mech Ageing Dev* 2010; **131**(2): 69-78.
12. Levine ME. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J Gerontol A Biol Sci Med Sci* 2013; **68**(6): 667-74.
13. Mitnitski A, Howlett SE, Rockwood K. Heterogeneity of Human Aging and Its Assessment. *J Gerontol A Biol Sci Med Sci* 2017; **72**(7): 877-84.
14. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* 2015; **386**(9993): 533-40.
15. Sudlow C, Gallagher 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(3): e1001779.

16. Yoo J, Kim Y, Cho ER, Jee SH. Biological age as a useful index to predict seventeen-year survival and mortality in Koreans. BMC Geriatr 2017; 17(1): 1.

17. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. Lancet 2018; 391(10132): 1775-82.

18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Predigosis Or Diagnosis (TRIPOD): The TRIPOD StatementThe TRIPOD Statement. Ann Intern Med 2015; 162(1): 55-63.

19. Newman AB, Sanders JL, Kizer JR, et al. Trajectories of function and biomarkers with age: the CHS All Stars Study. Int J Epidemiol 2016; 45(4): 1135-45.

20. Menni C, Kastenmüller G, Petersen AK, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol 2013; 42(4): 1111-9.

21. Shipilak MG, Sarnak MJ, Katz R, et al. Cystatin C and the Risk of Death and Cardiovascular Events among Elderly Persons. N Engl J Med 2005; 352(20): 2049-60.

22. Shipilak MG, Matsushita K, Arnljó J, et al. Cystatin C versus creatinine in determining risk based on kidney function. N Engl J Med 2013; 369(10): 932-43.

23. van der Laan SW, Fall T, Soumare A, et al. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. J Am Coll Cardiol 2016; 68(9): 934-45.

24. Farmer RE, Mathur R, Schmidt AF, et al. Associations Between Measures of Sarcopenic Obesity and Risk of Cardiovascular Disease and Mortality: A Cohort Study and Mendelian Randomization Analysis Using the UK Biobank. J Am Heart Assoc 2019; 8(13): e011638.

25. Ho FK, Celis-Morales CA, Petermann-Rocha F, et al. The association of grip strength with health outcomes does not differ if grip strength is used in absolute or relative terms: a prospective cohort study. Age Ageing 2019.

26. Etehad D, Emdin CA, Kiran A, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet 2016; 387(10022): 957-67.

27. Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J 2017; 38(32): 2459-72.

28. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet 2019; 51(3): 481-93.

29. Gupta RP, Strachan DP. Ventilatory function as a predictor of mortality in lifelong non-smokers: evidence from large British cohort studies. BMJ Open 2017; 7(7): e015381.

30. Shipley BA, Der G, Taylor MD, Deary IJ. Cognition and all-cause mortality across the entire adult age range: health and lifestyle survey. Psychosom Med 2006; 68(1): 17-24.

31. Barake M, Arabi A, Nakhoul N, et al. Effects of growth hormone therapy on bone density and fracture risk in relation to osteoporosis in the absence of growth hormone deficiency: a systematic review and meta-analysis. Endocrine 2018; 59(1): 39-49.

32. Williams DM, Karlsson IK, Pedersen NL, Hagg S. Circulating insulin-like growth factors and Alzheimer disease: a mendelian randomization study. Neurology 2018; 90(4): e291-e7.

33. Andreassen M, Raymond I, Kistorp C, Hildebrandt P, Faber J, Kristensen LO. IGFI as predictor of all cause mortality and cardiovascular disease in an elderly population. Eur J Endocrinol 2009; 160(1): 25-31.

34. Tumati S, Burger H, Martens S, van der Schouw YT, Aleman A. Association between Cognition and Serum Insulin-Like Growth Factor-1 in Middle-Aged & Older Men: An 8 Year Follow-Up Study. PLoS One 2016; 11(4): e0154450.

35. Aikens RC, Zhao W, Saleheen D, et al. Systolic Blood Pressure and Risk of Type 2 Diabetes: A Mendelian Randomization Study. Diabetes 2017; 66(2): 543.

36. Østergaard SD, Mukherjee S, Sharp SJ, et al. Associations Between Potentially Modifiable Risk Factors and Alzheimer Disease: A Mendelian Randomization Study. PLoS Med 2015; 12(6): e1001841.

37. Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002; 360(9349): 1903-13.

38. Danesh J, Collins R, Appleby P, Petro R. Association of Fibrinogen, C-reactive Protein, Albumin, or Leukocyte Count With Coronary Heart DiseaseMeta-analyses of Prospective Studies. JAMA 1998; 279(18): 1477-82.

39. Perry JRB, Weeden MN, Langenberg C, et al. Genetic evidence that raised sex hormone binding globulin (SHBG) levels reduce the risk of type 2 diabetes. Hum Mol Genet 2009; 19(3): 535-44.
40. Wang Q, Kangas AJ, Soininen P, et al. Sex hormone-binding globulin associations with circulating lipids and metabolites and the risk for type 2 diabetes: observational and causal effect estimates. *Int J Epidemiol* 2015; 44(2): 623-37.

41. Oppert JM, Charles MA, Thibult N, Guy-Grand B, Eschwege E, Ducimetiere P. Anthropometric estimates of muscle and fat mass in relation to cardiac and cancer mortality in men: the Paris Prospective Study. *Am J Clin Nutr* 2002; 75(6): 1107-13.

42. Janssen I, Heymsfield SB, Ross R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. *J Am Geriatr Soc* 2002; 50(5): 889-96.

43. Lai FY, Nath M, Hamby SE, Thompson JR, Nelson CP, Samani NJ. Adult height and risk of 50 diseases: a combined epidemiological and genetic analysis. *BMC Med* 2018; 16(1): 187.

44. Emerging Risk Factors Collaboration. Adult height and the risk of cause-specific death and vascular morbidity in 1 million people: individual participant meta-analysis. *Int J Epidemiol* 2012; 41(5): 1419-33.

45. Holmes MV, Asselbergs FW, Palmer TM, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* 2015; 36(9): 539-50.

46. Sniderman AD, Williams K, Contois JH, et al. A meta-analysis of low-density lipoprotein cholesterol, non-high-density lipoprotein cholesterol, and apolipoprotein B as markers of cardiovascular risk. *Circ Cardiovasc Qual Outcomes* 2011; 4(3): 337-45.

47. De Silva NMG, Borges MC, Hingorani A, et al. Liver Function and Risk of Type 2 Diabetes: Bidirectional Mendelian Randomization Study. *Diabetes* 2019; db181048.

48. Lai FY, Nath M, Hamby SE, Thompson JR, Nelson CP, Samani NJ. Adult height and risk of 50 diseases: a combined epidemiological and genetic analysis. *BMC Med* 2018; 16(1): 187.

49. Kunutsor SK, Bakker SJL, Kootstra-Ros JE, Gansevoort RT, Gregson J, Dullaart RPF. Serum Alkaline Phosphatase and Risk of Incident Cardiovascular Disease: Interrelationship with High Sensitivity C-Reactive Protein. *PLoS One* 2015; 10(7): e0132822.

50. Griffin SJ, Leaver JK, Irving GJ. Impact of metformin on cardiovascular disease: a meta-analysis of randomised trials among people with type 2 diabetes. *Diabetologia* 2017; 60(9): 1620-9.

51. Jiang H, Li J, Yu K, et al. Associations of estimated glomerular filtration rate and blood urea nitrogen with incident coronary heart disease: the Dongfeng-Tongji Cohort Study. *Sci Rep* 2017; 7(1): 9987.

52. Hyppönen E, Mulugeta A, Zhou A, Santhanakrishnan VK. A data-driven approach for studying the role of body mass in multiple diseases: a phenome-wide registry-based case-control study in the UK Biobank. *Lancet Digital Health* 2019; 1(3): e116-e26.

53. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature* 2018; 561(7721): 45-56.

54. Bio-Age. Bio-Age. http://bio-age.co.kr/ (accessed 1 August 2019).

55. Department of Health and Social Care. Advancing our health: prevention in the 2020s. London, 2019.

56. Spiegelhalter D. How old are you, really? Communicating chronic risk through ‘effective age’ of your body and organs. *BMCMed Inform Decis Mak* 2016; 16: 104.

57. Cohen AA, Milot E, Li Q, et al. Detection of a novel, integrative aging process suggests complex physiological integration. *PLoS One* 2015; 10(3): e0116489.

58. Pierson E, Koh P, Hashimoto T, et al. Inferring Multi-Dimensional Rates of Aging from Cross-Sectional Data. arXiv; 2018.