

RESEARCH PAPER

Retinal photograph-based deep learning predicts biological age, and stratifies morbidity and mortality risk

SIMON NUSINOVICI^{1,2,†}, TYLER HYUNGTAEK RIM^{1,2,†}, MARCO YU¹, GEUNYOUNG LEE³, YIH-CHUNG THAM^{1,2,4}, NING CHEUNG^{1,2}, CRYSTAL CHUN YUEN CHONG¹, ZHI DA SOH¹, SAHIL THAKUR¹, CHAN JOO LEE⁵, CHARUMATHI SABANAYAGAM^{1,2}, BYOUNG KWON LEE⁶, SUNGHA PARK⁷, SUNG SOO KIM⁸, HYEON CHANG KIM⁹, TIEN-YIN WONG^{1,2}, CHING-YU CHENG^{1,2,10}

¹Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

²Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore

³Medi Whale Inc., Seoul, South Korea

⁴Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

⁵Division of Cardiology, Severance Cardiovascular Hospital, Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea

⁶Division of Cardiology, Severance Cardiovascular Hospital, Gangnam Severance Hospital, Yonsei University Medical College of Medicine, Seoul, South Korea

⁷Division of Cardiology, Severance Cardiovascular Hospital and Integrated Research Center for Cerebrovascular and Cardiovascular Disease, Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea

⁸Department of Ophthalmology, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

⁹Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Korea

¹⁰Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

Address correspondence to: Tyler Hyungtaek Rim, Singapore Eye Research Institute, 20 College Road, The Academia, Level 6, Discovery Tower, Singapore 169856, Singapore. Tel.: +65 65767228; Email: tyler.rim@sneec.com.sg; Ching-Yu Cheng, Singapore Eye Research Institute, 20 College Road, The Academia, Level 6, Discovery Tower, Singapore 169856, Singapore. Tel.: +65 6576 7277; Email: chingyu.cheng@duke-nus.edu.sg

[†]Contributed equally

Abstract

Background: ageing is an important risk factor for a variety of human pathologies. Biological age (BA) may better capture ageing-related physiological changes compared with chronological age (CA).

Objective: we developed a deep learning (DL) algorithm to predict BA based on retinal photographs and evaluated the performance of our new ageing marker in the risk stratification of mortality and major morbidity in general populations.

Methods: we first trained a DL algorithm using 129,236 retinal photographs from 40,480 participants in the Korean Health Screening study to predict the probability of age being ≥ 65 years ('RetiAGE') and then evaluated the ability of RetiAGE to stratify the risk of mortality and major morbidity among 56,301 participants in the UK Biobank. Cox proportional hazards model was used to estimate the hazard ratios (HRs).

Results: in the UK Biobank, over a 10-year follow up, 2,236 (4.0%) died; of them, 636 (28.4%) were due to cardiovascular diseases (CVDs) and 1,276 (57.1%) due to cancers. Compared with the participants in the RetiAGE first quartile, those in the RetiAGE fourth quartile had a 67% higher risk of 10-year all-cause mortality (HR = 1.67 [1.42–1.95]), a 142% higher risk of CVD mortality (HR = 2.42 [1.69–3.48]) and a 60% higher risk of cancer mortality (HR = 1.60 [1.31–1.96]), independent of CA and established ageing phenotypic biomarkers. Likewise, compared with the first quartile group, the risk of CVD and cancer events in the fourth quartile group increased by 39% (HR = 1.39 [1.14–1.69]) and 18% (HR = 1.18 [1.10–1.26]), respectively. The best discrimination ability for RetiAGE alone was found for CVD mortality (c-index = 0.70,

sensitivity = 0.76, specificity = 0.55). Furthermore, adding RetiAGE increased the discrimination ability of the model beyond CA and phenotypic biomarkers (increment in c-index between 1 and 2%).

Conclusions: the DL-derived RetiAGE provides a novel, alternative approach to measure ageing.

Keywords: Deep learning, artificial intelligence, biological age, retinal photograph, mortality, cardiovascular disease, cancer, older people

Key Points

- We developed a retina-based biological age (termed RetiAGE) based on a deep learning algorithm trained using retinal photos.
- RetiAGE was associated with all-cause, cardiovascular disease and cancer mortality, and with cardiovascular and cancer events, independently of chronological age and phenotypic biomarkers.
- Furthermore, adding RetiAGE increased the discrimination ability of the model beyond chronological age and phenotypic biomarkers.
- This approach provides a novel, alternative approach to measure biological age using retinal photographs.

Introduction

Globally, the number of persons aged 80 years or over is projected to increase more than threefold between 2017 and 2050, reaching 425 million in 2050 [1]. This ageing population is likely to result in an increased prevalence of cardiovascular [2, 3] and chronic diseases [4, 5] with significant healthcare associated costs [6]. In this context, the identification of robust biomarkers for disease risk stratification could help implement early health interventions and limit the burden of these diseases.

Biological age (BA) can be defined as a quantity expressing the ‘true global state’ of ageing organism. Biomarkers of BA are of particular interest, because measurements of BA may better capture physiological changes associated with ageing process, compared with chronological age (CA). BA can thus be used to assess the general health status of individuals of the same CA. Different measurements can be used to estimate BA, including clinical biomarkers [7] (like total cholesterol and blood pressure or combination of several clinical biomarkers, such as ‘PhenoAge’ [8]), telomere length [9], DNA methylation [10], etc. For example, using physiological and blood biomarkers to estimate BA, studies found that individuals of the same CA varied on their BA by as much as 10 years above and below their CA [11]. Moreover, the estimated BA outperformed the CA in predicting frailty and mortality [11]. However, the invasive, high-cost and/or time-consuming nature of these measurements has limited their value as a clinically useful biomarker of BA.

The retina (fundus) of the eye represents a unique non-invasive window into the systemic health status. Changes in retinal vasculature, for example, may reflect a range of sub-clinical pathophysiological responses to hyperglycemia, hypertension and inflammation [12]. They are also associated with increased risk of several chronic and age-related diseases [13–17]. Furthermore, changes in the retina are associated with ageing. From middle age onwards, the geometrical complexity of the retinal vasculature is reduced [18] as well as the retinal vessel calibres [19]. Moreover, vessel calibres are

associated with carotid artery plaque and carotid artery intima-media thickness [20, 21]. More importantly, the retina is amenable to noninvasive imaging and rapid assessment with digital photography.

Deep learning (DL) is a subfield of machine learning and a leading methodology for extracting insights from unstructured data such as images. The flexibility of DL approaches makes them especially powerful at identifying patterns and has subsequently led to their rapid adoption within the medical imaging community. DL algorithms have been successfully applied to retinal photographs in predicting the risk of systemic diseases, such as anemia [22], chronic kidney diseases [23], estimating systemic biomarkers [24–26] and cardiovascular risk [27].

We hypothesised that BA could be predicted using DL on retinal images. Hence, in this study, we developed a retinal photograph-based DL algorithm to predict BA and determined the performance of this new BA marker in stratifying risk for mortality (all-cause, cardiovascular disease [CVD] and cancer) and disease events (CVD and cancer). Finally, we investigated the ability of the new BA marker to improve the discrimination of mortality and disease events beyond CA and established clinical biomarkers.

Methods

This study was approved by the Institutional Review Board (IRB) of Severance Hospital at Yonsei University College of Medicine in Seoul, Korea. The IRB waived the requirement to obtain informed consent. Because of its retrospective design and use of deidentified data (both image and clinical), this study was deemed exempt from IRB review by the IRB of SingHealth. In the UK Biobank study, written informed consent was obtained from the participants.

Overall study design

We provide here a summary of the materials and methods used for this study. A detailed version is available in

Table 1. Characteristics of the study population

	Korean Health Screening Study (<i>n</i> = 46,551)	UK Biobank Study (<i>n</i> = 56,301)
Characteristics and PhenoAGE variables and score		
Female, <i>n</i> (%)	21,134 (45.4%)	30,129 (53.5%)
CA (year), mean (SD)	53.8 (9.4)	57.1 (8.3)
Albumin (g/L), mean (SD)	44.7 (2.6)	45.7 (2.6)
Creatinine (umol/L), mean (SD)	69.4 (19.4)	73.2 (17.1)
Glucose (mmol/L), mean (SD)	5.5 (1.2)	5.1 (1.0)
C-reactive protein (mg/dL), mean (SD)	1.4 (4.9)	2.4 (4.2)
Lymphocyte percent, mean (SD)	33.8 (8.0)	29.3 (7.6)
Mean corpuscular cell volume (fL), mean (SD)	90.8 (4.7)	91.8 (4.5)
Red cell distribution width percent, mean (SD)	NA	13.5 (1.0)
Alkaline phosphatase (U/L), mean (SD)	65.7 (20.8)	83.5 (25.3)
White blood cell count (1,000 cells/uL), mean (SD)	5.7 (1.7)	7.0 (2.1)
PhenoAGE score	NA	51.3 (10.1)
Primary outcome: mortality		
Follow-up period (year), mean (SD)	4.2 (2.7–5.7)	9.4 (1.3)
All death, <i>n</i> (%)	194 (0.4%)	2,236 (4.0%)
CVD death, <i>n</i> (%)	23 (0.1%)	636 (1.1%)
Cancer death, <i>n</i> (%)	95 (0.2%)	1,276 (2.3%)
Secondary outcome: disease events		
CVD^a		
Follow-up (year), mean (SD)	NA	9.3 (1.4)
CVD events, <i>n</i> (%)	NA	1,255 (2.5%)
Cancer^b		
Follow-up (year), mean (SD)	NA	8.6 (2.3)
Cancer events, <i>n</i> (%)	NA	9,828 (20.3%)

Data are presented as *n*, *n* (%) of participants, mean (standard deviation [SD]). CVD = cardiovascular disease; NA = data not available; PhenoAGE = phenotypic age calculated based on clinical biomarkers (CA, albumin, creatinine, glucose, C-reactive protein [log], lymphocyte percent, mean [red] cell volume, red cell distribution width, alkaline phosphatase, white blood cell count) ^aAmong 49,493 participants without cancers at baseline ^bAmong 48,457 participants without CVDs at baseline

Appendix 1. In brief, we trained the DL algorithm to predict the probability for an individual of being ≥ 65 years old based on retinal photos using data from a health-screening centre in South Korea (Korean Health Screening study). We used a Visual Geometry Group (VGG), a classical deep convolutional neural network architecture with multiple layers that is widely used for image recognition [28]. The algorithm was trained to predict the likelihood of being old using a cut off value of 65 years old. No other information was used to train the algorithm. By doing so, we aimed at capturing patterns in the retina related to age by comparing an ‘older’ group with a ‘younger’ group in a broad and unspecific way. The algorithm was trained to pick up patterns that might occur in different parts of the retina and that might not be visible for human eyes. Furthermore, recognizing that 65 years old is an arbitrary cutoff, we also trained additional models by using 70 and 75 years old as the cutoff. We then assessed the association between this new marker (termed ‘RetiAGE’) in quartiles and mortality (all-cause, CVD and cancer related), and between RetiAGE and disease events (CVD and cancer) in the UK Biobank [29]. The flowchart of the study is presented in Appendix 7.

Statistical analyses

Cox proportional hazards model was used to estimate the hazard ratios (HRs) corresponding to the associations

between RetiAGE and the five outcomes. The Cox models were adjusted either on CA or on PhenoAGE, a phenotypic biomarker built using the following demographic and clinical data: CA, albumin, creatinine, glucose, c-reactive protein (log), lymphocyte percent, mean (red) cell volume, red cell distribution width, alkaline phosphatase and white blood cell count [8]. C-index was used to assess the discrimination of the Cox proportional hazards models [30]. The improvement of discrimination when adding RetiAGE to the risk model with either CA or PhenoAGE was assessed by testing the significance of the difference in c-index between the models with and without RetiAGE [31].

Results

Study population characteristics

In the Korean Health Screening study, the mean baseline age was 53.6 years (SD, 9.2) and 45.4% were female (Table 1). Among the 46,551 participants, 194 (0.4%) died during the 6-year follow-up. In the UK Biobank study, the mean baseline age was 57.1 years (SD, 8.3) and 46.5% were female (Table 1). Among the 56,301 participants, 2,236 (4.0%) died for all causes during the 10-year follow-up. Of them, 28.4% (636/2,236) were due to CVD-related causes and 57.1% (1,276/2,236) due to cancer related causes.

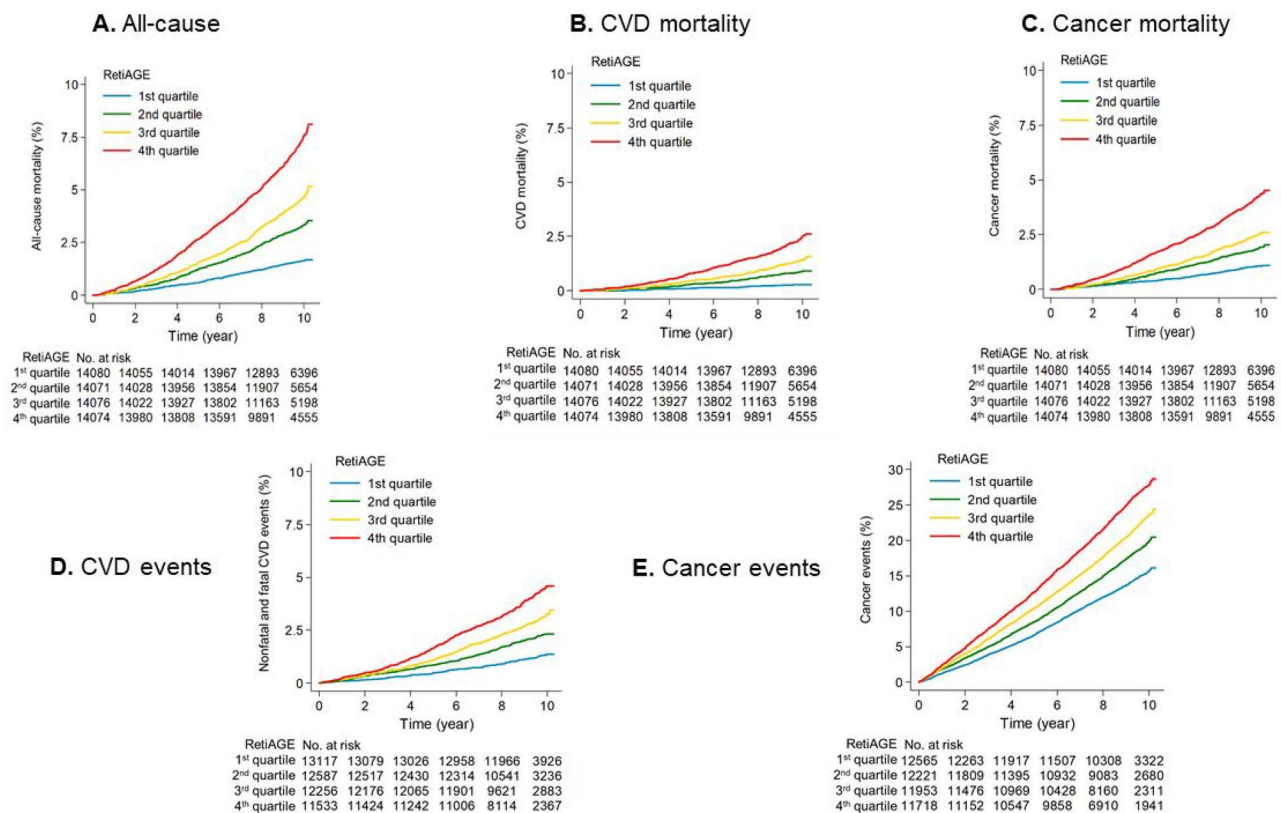


Figure 1. Kaplan–Meier estimates of mortality, CVD and cancer risks by RetiAGE quartiles in the UK Biobank study.

Performance of RetiAGE in predicting the probability of being ≥ 65 years old

The performance of RetiAGE in predicting the probability of being ≥ 65 years old in the internal testing set (derived from the Korean Health Screening study) was very good with an area under the receiver operating characteristic curve (AUROC) of 0.968 (95% confidence interval [CI]: 0.965–0.970) and an area under the precision-recall curve (AUPRC) of 0.83 (95% CI: 0.83–0.84) (Appendix 8). The characteristics of the developmental set for the DL algorithm training are shown in Appendix 2. The performance of RetiAGE in the UK Biobank study was moderate with an AUROC of 0.756 (0.753–0.759) and an AUPRC of 0.399 (0.388–0.410). Finally, the correlation between RetiAGE and CA was 0.62 (Spearman’s rank correlation coefficient, $P < 0.001$, Appendix 9A) and between RetiAGE and PhenoAGE was 0.56 (Spearman’s rank correlation coefficient, $P < 0.001$, Appendix 9B).

Relationship between DL-predicted RetiAGE score and mortality and diseases events

The distributions of RetiAGE and the corresponding quartile groups in the two studies are presented in Appendix 10. These distributions in the UK Biobank study are presented in Appendix 11 according to the CA and the survival status. In the UK Biobank study, the participants in the fourth RetiAGE quartile had highest all-cause (6.8% [$n = 952$] for

all-cause, 2.1% [$n = 301$] for CVD and 3.9% [$n = 543$] for cancer mortality) compared with those in the first quartile (1.6% [$n = 225$]), CVD (0.3% [$n = 37$]) and cancer mortality rates (1.0% [$n = 147$]) (Appendix 3).

Kaplan–Meier plots showed distinct mortality risk curves for the RetiAGE quartile groups (Figure 1A–C). The unadjusted HRs for participants in the fourth quartile group were 4.74 (95% CI: 4.10–5.48) for all-cause, 9.19 (95% CI: 6.53–12.93) for CVD and 4.11 (95% CI: 3.42–4.93) for cancer mortality, compared with those in the first quartile (Table 2). Adjustment on CA decreased the magnitude of effects, but the association remained significantly. After further adjustment on PhenoAGE (which includes CA and other established ageing biomarkers [8]), the HRs corresponding to the fourth quartile were 1.67 (95% CI: 1.42–1.95) for all-cause, 2.42 (95% CI: 1.69–3.48) for CVD and 1.60 (95% CI: 1.31–1.96) for cancer mortality. In addition to mortality, similar analyses were conducted with CVD events and cancer events (including fatal and non-fatal events). We consistently observed the association of disease risks with the RetiAGE quartile groups (Table 2 and Figure 1D and E). Compared with participants in the first quartile group, the risk of events for those in the fourth quartile was 39% and 18% higher for CVD (HR = 1.39 [1.14–1.69]) and cancer events (HR = 1.18 [1.10–1.26]), respectively, independent of PhenoAGE. Finally, in the Korean study, the HRs adjusted for CA were 2.03 (95% CI: 0.96–4.28) for the 2nd, 2.38 (95% CI: 1.05–5.41) for the 3rd

Table 2. Risk of mortality and morbidity associated with the quartiles of the deep-learning predicted age (RetiAGE score) in the UK Biobank study

RetiAGE	Events	Inc.	Unadj. HR (95%CI)	CA-adj. HR (95%CI)	PhenoAGE-adj. HR (95%CI)
All-cause mortality ^a					
1st quartile	225	1.6	1.00 (reference)	1.00 (reference)	1.00 (reference)
2nd quartile	447	3.3	2.06 (1.75, 2.42)	1.31 (1.10, 1.54)	1.26 (1.06, 1.48)
3rd quartile	612	4.6	2.89 (2.48, 3.37)	1.41 (1.19, 1.67)	1.32 (1.12, 1.55)
4th quartile	952	7.5	4.74 (4.10, 5.48)	1.82 (1.54, 2.15)	1.67 (1.42, 1.95)
HR trend, <i>P</i> for trend			1.62 (1.55–1.68), <i>P</i> < 0.001	1.21 (1.15–1.26), <i>P</i> < 0.001	1.17 (1.12–1.23), <i>P</i> < 0.001
CVD mortality ^a					
1st quartile	37	0.3	1.00 (reference)	1.00 (reference)	1.00 (reference)
2nd quartile	116	0.9	3.26 (2.25, 4.72)	1.87 (1.27, 2.74)	1.7 (1.16, 2.48)
3rd quartile	182	1.4	5.26 (3.69, 7.49)	2.21 (1.51, 3.22)	1.91 (1.32, 2.75)
4th quartile	301	2.4	9.19 (6.53, 12.93)	2.93 (2.01, 4.26)	2.42 (1.69, 3.48)
HR trend, <i>P</i> for trend			1.88 (1.74–2.04), <i>P</i> < 0.001	1.33 (1.22–1.46), <i>P</i> < 0.001	1.26 (1.16–1.38), <i>P</i> < 0.001
Cancer mortality ^a					
1st quartile	147	1.1	1.00 (reference)	1.00 (reference)	1.00 (reference)
2nd quartile	256	1.9	1.80 (1.47, 2.21)	1.16 (0.94, 1.43)	1.15 (0.93, 1.42)
3rd quartile	330	2.5	2.38 (1.96, 2.89)	1.19 (0.96, 1.48)	1.17 (0.95, 1.44)
4th quartile	543	4.3	4.11 (3.42, 4.93)	1.65 (1.34, 2.04)	1.60 (1.31, 1.96)
HR trend, <i>P</i> for trend			1.57 (1.49–1.65), <i>P</i> < 0.001	1.19 (1.12–1.26), <i>P</i> < 0.001	1.18 (1.11–1.25), <i>P</i> < 0.001
CVD events ^b					
1st quartile	168	1.3	1.00 (reference)	1.00 (reference)	1.00 (reference)
2nd quartile	271	2.3	1.74 (1.44, 2.11)	1.17 (0.96, 1.43)	1.14 (0.93, 1.39)
3rd quartile	358	3.2	2.43 (2.02, 2.92)	1.29 (1.06, 1.58)	1.23 (1.01, 1.50)
4th quartile	458	4.5	3.46 (2.90, 4.13)	1.48 (1.21, 1.82)	1.39 (1.14, 1.69)
HR trend, <i>P</i> for trend			1.48 (1.41–1.56), <i>P</i> < 0.001	1.14 (1.07–1.21), <i>P</i> < 0.001	1.11 (1.05–1.18), <i>P</i> < 0.001
Cancer events ^c					
1st quartile	1908	16.8	1.00 (reference)	1.00 (reference)	1.00 (reference)
2nd quartile	2,297	21.5	1.29 (1.22, 1.37)	1.07 (1.00, 1.14)	1.05 (0.98, 1.12)
3rd quartile	2,629	26.0	1.57 (1.48, 1.66)	1.13 (1.05, 1.20)	1.11 (1.04, 1.18)
4th quartile	2,994	31.6	1.93 (1.82, 2.04)	1.20 (1.12, 1.29)	1.18 (1.10, 1.26)
HR trend, <i>P</i> for trend			1.24 (1.22–1.26), <i>P</i> < 0.001	1.06 (1.04–1.09), <i>P</i> < 0.001	1.06 (1.04–1.09), <i>P</i> < 0.001

Inc = incidence per 1,000 person-years; CI = confidence interval; CVD = cardiovascular disease; HR = hazard ratio; Unadj. HR = unadjusted HR; CA-adj. HR = HR adjusted HR on chronological age; PhenoAGE-adj. HR = HR adjusted on PhenoAGE; PhenoAGE = phenotypic age calculated based on clinical biomarkers (CA, albumin, creatinine, glucose, C-reactive protein [log], lymphocyte percent, mean [red] cell volume, red cell distribution width, alkaline phosphatase, white blood cell count); RetiAGE = deep learning-based retinal biological age. ^a*n* = 56,301; ^b*n* = 49,493 for CVD; ^c*n* = 48,457

and 4.07 (95% CI: 1.70–9.74) for the 4th quartile group (Appendix 12).

The subgroup analysis by gender showed that RetiAGE performed better in males with higher magnitude of effects between RetiAGE and all-cause mortality (PhenoAGE-adjusted HR in the 4th quartile group = 1.79 [95% CI: 1.44–2.22] in males and 1.54 [95% CI: 1.21–1.95] in females) (Appendix 13). Moreover, to account for a possible reverse causality bias, we performed a sensitivity analysis by excluding participants that died within the first 2 years and observed similar findings (Appendix 14). Furthermore, we performed additional analyses on the age threshold considered for the DL algorithm training. Because 65 years old is an arbitrary cutoff, we also trained the DL algorithm using 70 and 75 years old and calculated the corresponding c-index values (Appendix 4). The results were similar and did not change the conclusion of the study. Finally, we further adjusted the models on vessel calibres [25] (Appendix 5) and found very similar results.

To localise the anatomy contributing to RetiAGE, saliency maps were generated (Figure 2). The saliency maps indicate that RetiAGE commonly focuses on the macula, optic disc and retinal vessels.

Improvement in predictive performance when adding the DL-predicted RetiAGE score to the risk models

Adding RetiAGE onto CA (model 2 versus model 1) or PhenoAGE (model 4 versus model 3) increased the discrimination around 1.5% for all mortality outcomes (Table 3). The highest increase in c-index was found for CVD mortality, with a difference in c-index up to 1.8%. Regarding CVD and cancer events, the differences in c-index after adding RetiAGE were within the same range (Table 3). Appendix 6 presents the sensitivities and specificities of the different risk models.

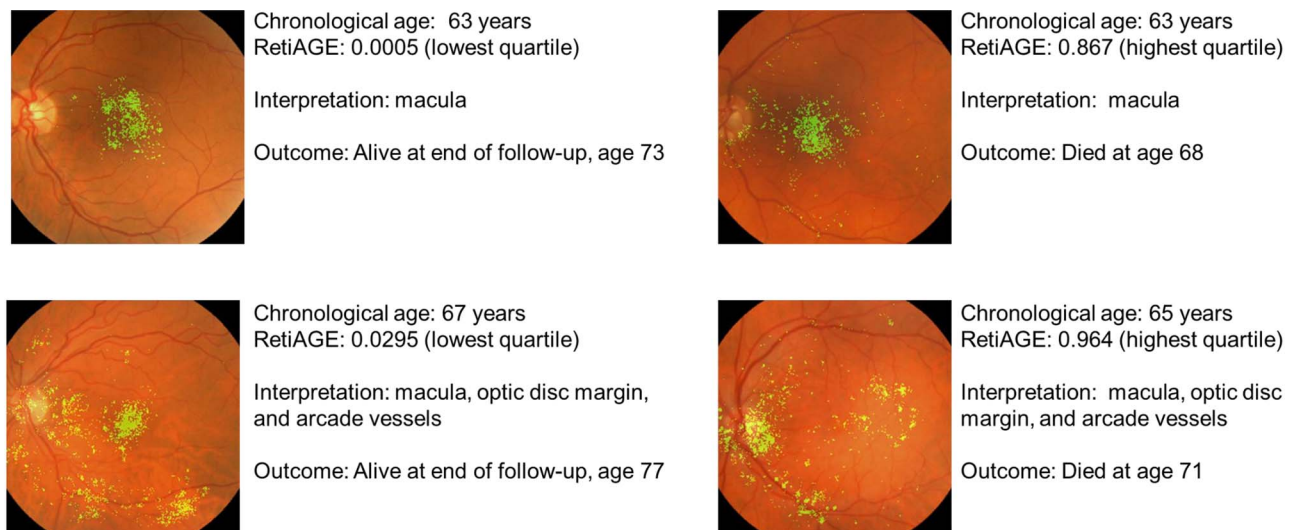


Figure 2. Saliency map localise anatomy contributing to RetiAGE

Table 3. Improvement in predictive performance (measured using c-index) when adding the deep learning predicted age (RetiAGE score) to the risk models in the UK Biobank study

	Model 0: RetiAGE	Model 1: CA	Model 2: CA + RetiAGE	Model 3: PhenoAGE	Model 4: PhenoAGE + RetiAGE
Primary outcome					
All-cause mortality	0.664 (0.653–0.675)	0.706 (0.696–0.716)	0.720 (0.709–0.730) ^a	0.737 (0.727–0.747)	0.750 (0.740–0.760) ^a
CVD mortality	0.702 (0.684–0.720)	0.742 (0.725–0.759)	0.760 (0.744–0.777) ^a	0.788 (0.773–0.802)	0.804 (0.790–0.819) ^a
Cancer mortality	0.657 (0.642–0.671)	0.696 (0.682–0.709)	0.709 (0.695–0.722) ^a	0.718 (0.705–0.731)	0.732 (0.718–0.745) ^a
Secondary outcome					
CVD event	0.646 (0.631–0.661)	0.691 (0.673–0.705)	0.701 (0.687–0.716) ^a	0.720 (0.706–0.733)	0.730 (0.716–0.744) ^a
Cancer event	0.601 (0.593–0.608)	0.629 (0.622–0.636)	0.637 (0.629–0.644) ^a	0.646 (0.639–0.654)	0.653 (0.646–0.661) ^a

The values in the table corresponded to the expressed as c-index with their 95% confidence intervals ^aSignificant difference between Model 1 and 2 ($P < 0.001$), and Model 3 and 4 ($P < 0.001$) based on DeLong's method. CVD = cardiovascular disease; RetiAGE = deep learning predicted biological age; PhenoAGE = phenotypic age calculated based on clinical biomarkers (CA, albumin, creatinine, glucose, C-reactive protein [log], lymphocyte percent, mean [red] cell volume, red cell distribution width, alkaline phosphatase, white blood cell count)

Discussion

We developed a retinal BA marker (RetiAGE) based on a DL algorithm trained using retinal photos from a large Korea dataset and demonstrated that this new marker can risk stratify for mortality and morbidity in the UK Biobank study, independently of CA and phenotypic biomarkers. RetiAGE corresponded to the probability of being older than 65 years old. People in the fourth quartile of RetiAGE (thus with a higher probability of being older) had a risk increased by 67% for all-cause mortality, 142% for CVD and 60% for cancer mortality; and by 39% for CVD events and 18% for cancer events over 10-year, compared with people in the first quartile. The best discrimination ability for RetiAGE alone was found for CVD mortality (c-index = 0.70, sensitivity = 0.76, specificity = 0.55). Furthermore, adding RetiAGE increased the discrimination ability of the model beyond CA and phenotypic biomarkers (increment in c-index between 1% and 2%). These results indicate that retinal marker of BA could be used as an alternative measurement of BA.

Our DL-predicted RetiAGE was associated with all-cause, CVD and cancer mortality, and with CVD and cancer events with moderate to high magnitude of effects (HRs corresponding to the highest quartile between 1.60 and 2.42 for mortality, and between 1.18 and 1.39 for disease events). These increased risks were similar to measurements of accelerated ageing related to oxidative stress (HR the fourth versus the first quartile = 1.56) and DNA methylation (HR = 1.71 for moderate and 2.92 for high epigenetic score) with regard to all-cause mortality during a 15-year follow-up period [32]. Moreover, similar associations were found for circulating biomarkers (alpha-1-acid glycoprotein, albumin, very low-density lipoprotein particles and citrate) with regard to all-cause (HR [per 1-SD increase] = 1.49), CVD (HR = 1.34) and cancer mortality (HR = 1.43), independently of conventional risk factors [33]. Compared with these measurements, our DL-predicted score based on retinal photos is simple and noninvasive. It is furthermore relatively cheap, usually charged \$20–30, compared with genetic tests that cost few hundred dollars. All these characteristics make our DL-predicted marker an appropriate and relevant screening tool

that could help early identify patients with a physiological deterioration possibly leading to diseases and increased risk of mortality.

In the context of ageing population with the rise of chronic diseases, provision of early personalised recommendations may have major public health benefits. For example, we found that RetiAGE alone had a quite good discrimination ability for CVD mortality (AUROC = 0.70), with 76% of the individuals that died within 10 years being correctly identified using this marker. Moreover, adding RetiAGE beyond CA and a phenotypic age score based on clinical biomarkers (PhenoAGE) allowed to further increase the predictive performance of the mortality risk models. The increases in discrimination were moderate, overall comprised between 1% and 1.8% increase in c-index, the maximum being found for CVD mortality. Adding RetiAGE beyond PhenoAGE increased the sensitivity by 9% for all-cause mortality and by 4% beyond CA for CVD events. However, these improvements came at the expense of decreases in specificity. Finally, we found that RetiAGE better risk stratifies in male compared with female. This is possibly due to differences in retinal vasculature between sex that are associated with systemic diseases. For example, retinal arteriolar vessel calibres are narrower in male [34]; and narrower arteriolar calibres are strongly associated with hypertension [35].

The c-index metric is known to be quite insensitive [36–38] and small increases around 1% might still be clinically meaningful [36]. For example, the increase we found when adding RetiAGE beyond CA to predict CVD mortality (c-index increment = 1.8%) was larger than the added value of HDL cholesterol to predict CVD risk beyond age, systolic blood pressure (SBP) and smoking (c-index increment = 1%) [38]. Despite this, HDL cholesterol is a strong risk factor of CVD risk and widely used in clinic to evaluate individual risks. C-reactive protein is another example of biomarker that is strongly associated with CV events but that do not improve the discrimination capability of the risk prediction model [37, 39]. Although our result seems promising, we need to confirm our results in other populations, and the clinical usefulness needs to be evaluated. Other DL algorithms have been used to predict BA from other kind of images or scans such as neuroimaging [40], facial images [41] or chest X-ray [42]. However, to the best of our knowledge, no study has yet investigated the association between these BA measurements and mortality. More research is thus needed to assess these associations and compare the usefulness of the different approaches using DL in mortality risk stratification.

Strengths of our study included a large Korean study for the development of our DL-predicted score, and a large study for validation (UK Biobank). The difference of the ethnicity between these two studies may explain the drop in the DL algorithm performance in predicting the probability of age being ≥ 65 years between the training dataset and the external one. However, in the latter one, we showed significant associations and improved predictive performance when adding RetiAGE in the mortality models, suggesting that our

new BA biomarker could be used in different populations. Moreover, we included in our analysis clinical biomarkers previously used to build a validated ageing biomarker ('PhenoAGE'), thus demonstrating the ability of our new biomarker in predicting mortality and morbidity related to CVD and cancer above and beyond these biomarkers. Finally, the similar results obtained after adjustment on vessel calibres along with the saliency maps show that RetiAGE did not only capture information in the retinal vasculature but also in other areas, such as macula or optic disc. This study has, however, limitations. Firstly, we trained the algorithm to predict the probability for an individual of being ≥ 65 years old based on retinal photos to capture retinal patterns associated with ageing process. However, because the training is only based on CA, the patterns might not be specific to poor health status. Secondly, we used a cut off to train the algorithm at 65 years old. Although frequently used, this cut off can be seen as arbitrary. We have thus performed sensitivity analyses using cut off at 70 and 75 years old. We found similar results that show that our approach is not dependant on the cut off value. Thirdly, in the UK Biobank study, the CVD and cancer statuses at baseline were self-reported and thus there might be recall bias. Fourthly, the unbalanced distribution of ethnicity did not allow to stratify the analyses on this factor. Finally, we only used good quality retinal photos for model training and validation. For example, in the Google's diabetic retinopathy screening study [43], 11.6% of the photos in a real-world prospective dataset, EyePACS-1, were ungradable. Therefore, our model performance may not be generalizable to real-world settings where clinical services are provided, such as diabetic retinopathy screening programs. The impact of ungradable photos on the performance would thus need to be evaluated.

In conclusion, we demonstrate here, using two large datasets from Korea and UK, that a DL algorithm applied on retinal photos can estimate BA and be used for the risk stratification of mortality and major morbidity related to CVD and cancer. Our approach provides a novel, alternative approach to measure ageing. The findings of the study highlight the usefulness of digital technology applied on retinal photos in the risk stratification of population health.

Supplementary Data: Supplementary data mentioned in the text are available to subscribers in Age and Ageing online.

Acknowledgements: The UK Biobank data were obtained from UK Biobank (application number 45925), and a full list of the IDs of gradable photographs and code are provided at https://github.com/medi-whale/UKBIOBAN_K_FUNDUS_Classifier. Data cannot be shared publicly due to the violation of patient privacy and lack of informed consent for data sharing. The Korean data were obtained from the Yonsei University, Department of Ophthalmology (contact Prof. SS Kim, semekim@yuhs.ac) for researchers who meet the criteria for access to confidential data.

Declaration of Conflicts of Interest: T.H.R. was a former scientific advisor and owns stock of Medi Whale. G.L. is an employee of Medi Whale and owns stock of Medi Whale. T.H.R., G.L., and T.Y.W. hold patents on a deep learning system in ophthalmology and these patents are not directly related to this study. C.Y.C. has received consulting fees from Medi Whale. T.Y.W. has received consulting fees from Allergan, Bayer, Boehringer-Ingelheim, Genentech, Merck, Novartis, Oxurion, Roche, and Samsung Bioepis. T.Y.W. is a co-founder of Plano and EyRis. S.P. received lecture fees from Pfizer, Boryoung, Hanmi, Dae-wong, Donga, Celltrion, Servier, Daiichi Sankyo and Dae-won. S.P. also received research grant from Daiichi Sankyo.

Declaration of Sources of Funding: This work was supported by grants from the National Medical Research Council, Singapore (NMRC/CIRG/1417/2015 and NMRC/CIRG/1488/2018 to C.C.Y.) and by the Healthy Longevity Catalyst Awards from National Medical Research Council, Singapore (MOH-HLCA21Jan-0004). This work was also supported by the Ministry of Trade, Industry and Energy and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (Project number. P0011929 to S.S.K), Korea; the Agency for Science, Technology, and Research (grant number A19D1b0095 to T.H.R.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. United Nations, Department of Economic and Social Affairs, population division. World Population Ageing 2017 - Highlights (ST/ESA/SER.A/397) 2017.
2. Odden MC, Coxson PG, Moran A, Lightwood JM, Goldman L, Bibbins-Domingo K. The impact of the aging population on coronary heart disease in the United States. *Am J Med* 2011; 124: 827–833.e5.
3. Stewart S, MacIntyre K, Capewell S, McMurray JJV. Heart failure and the aging population: an increasing burden in the 21st century? *Heart Br Card Soc* 2003; 89: 49–53.
4. Narayan KMV, Boyle JP, Geiss LS, Saaddine JB, Thompson TJ. Impact of recent increase in incidence on future diabetes burden: U.S., 2005–2050. *Diabetes Care* 2006; 29: 2114–6.
5. Boyle JP, Honeycutt AA, Narayan KMV et al. Projection of diabetes burden through 2050: impact of changing demography and disease prevalence in the U.S. *Diabetes Care* 2001; 24: 1936–40.
6. Caley M, Sidhu K. Estimating the future healthcare costs of an aging population in the UK: expansion of morbidity and the need for preventative care. *J Public Health* 2011; 33: 117–22.
7. Bae C-Y, Kang YG, Piao M-H et al. Models for estimating the biological age of five organs using clinical biomarkers that are commonly measured in clinical practice settings. *Maturitas* 2013; 75: 253–60.
8. Levine ME, Lu AT, Quach A et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 2018; 10: 573–91.
9. 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: 1220–30.
10. Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. *Aging Cell* 2015; 14: 924–32.
11. Zhong X, Lu Y, Gao Q et al. Estimating biological age in the Singapore longitudinal aging study. *J Gerontol Ser A* 2020; 75: 1913–20.
12. Ikram MK, Cheung CY, Lorenzi M et al. Retinal vascular Caliber as a biomarker for diabetes microvascular complications. *Diabetes Care* 2013; 36: 750–9.
13. McGeechan K. Meta-analysis: retinal vessel Caliber and risk for coronary heart disease. *Ann Intern Med* 2009; 151: 404.
14. Cheung CY, Sabanayagam C, Law AK et al. Retinal vascular geometry and 6 year incidence and progression of diabetic retinopathy. *Diabetologia* 2017; 60: 1770–81.
15. Sabanayagam C, Shankar A, Koh D et al. Retinal microvascular caliber and chronic kidney disease in an Asian population. *Am J Epidemiol* 2008; 169: 625–32.
16. Cheung CY, Tay WT, Mitchell P et al. Quantitative and qualitative retinal microvascular characteristics and blood pressure. *J Hypertens* 2011; 29: 1380–91.
17. Cheung CY, Ong YT, Ikram MK et al. Microvascular network alterations in the retina of patients with Alzheimer's disease. *Alzheimers Dement* 2014; 10: 135–42.
18. Azemin MZC, Kumar DK, Wong TY et al. Age-related rarefaction in the fractal dimension of retinal vessel. *Neurobiol Aging* 2012; 33: e1–194.e4.
19. Ikram MK, Ong YT, Cheung CY, Wong TY. Retinal vascular Caliber measurements: clinical significance, current knowledge and future perspectives. *Ophthalmologica* 2013; 229: 125–36.
20. Ikram MK, de Jong FJ, Vingerling JR et al. Are retinal arteriolar or Venular diameters associated with markers for cardiovascular disorders? The Rotterdam study. *Investig Ophthalmology Vis Sci* 2004; 45: 2129.
21. Klein R, Sharrett AR, Klein BEK et al. Are retinal arteriolar abnormalities related to atherosclerosis? The atherosclerosis risk in communities study. *Arterioscler Thromb Vasc Biol* 2000; 20: 1644–50.
22. Mitani A, Huang A, Venugopalan S et al. Detection of anaemia from retinal fundus images via deep learning. *Nat Biomed Eng* 2020; 4: 18–27.
23. Sabanayagam C, Xu D, Ting DSW et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit Health* 2020; 2: e295–302.
24. Poplin R, Varadarajan AV, Blumer K et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018; 2: 158–64.
25. Cheung CY, Xu D, Cheng C-Y et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng* 2021; 5: 498–508.
26. Rim TH, Lee G, Kim Y et al. Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. *Lancet Digit Health* 2020; 2: e526–36.

27. Rim TH, Lee CJ, Tham Y-C *et al.* Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *Lancet Digit Health* 2021; 3: e306–16.
28. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Conf. Pap. Int. Conf. Learn. Represent. ICLR* (2015).
29. Orimo H, Ito H, Suzuki T, Araki A, Hosoi T, Sawabe M. Reviewing the definition of ‘elderly’. *Geriatr Gerontol Int* 2006; 6: 149–58.
30. Pencina MJ, D’Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004; 23: 2109–23.
31. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med* 2015; 34: 685–703.
32. Gao X, Gao X, Zhang Y, Holleczeck B, Schöttker B, Brenner H. Oxidative stress and epigenetic mortality risk score: associations with all-cause mortality among elderly people. *Eur J Epidemiol* 2019; 34: 451–62.
33. Fischer K, Kettunen J, Würtz P *et al.* Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* 2014; 11: e1001606.
34. Wong TY, Islam FMA, Klein R *et al.* Retinal vascular Caliber, cardiovascular risk factors, and inflammation: the multi-ethnic study of atherosclerosis (MESA). *Investig Ophthalmology Vis Sci* 2006; 47: 2341.
35. Cheung CY, Ikram MK, Sabanayagam C, Wong TY. Retinal microvasculature as a model to study the manifestations of hypertension. *Hypertension* 2012; 60: 1094–103.
36. Pencina MJ, D’Agostino RB, D’Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27: 157–72.
37. Greenland P, O’Malley PG. When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase A2 and C-reactive protein for stroke risk. *Arch Intern Med* 2005; 165: 2454–6.
38. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115: 928–35.
39. Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N Engl J Med* 2002; 347: 1557–65.
40. Cole JH, Franke K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci* 2017; 40: 681–90.
41. Xia X, Chen X, Wu G, *et al.* Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle. *Nat Metab* 2020; 2: 946–57.
42. Karagyris A, Kashyap S, Wu JT, Sharma A, Moradi M, Syeda-Mahmood T. Age prediction using a large chest x-ray dataset. In: (eds. Hahn HK., Mori, K.) *Medical Imaging 2019: Computer-Aided Diagnosis*. Vol. 66. SPIE: California, United States, 2019. <https://doi.org/10.1117/12.2512922>.
43. Gulshan V, Peng L, Coram M *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316: 2402.

Received 2 August 2021; editorial decision 2 February 2022