npj Aging

https://doi.org/10.1038/s41514-026-00418-1

Article in Press

A multimodal ocular aging index reveals proteomic pathways and predicts incident age-related eye diseases

Received: 12 February 2026

Accepted: 18 May 2026

Cite this article as: Kai, J.-Y., Gong, S.-Y., Li, D.-L. *et al.* A multimodal ocular aging index reveals proteomic pathways and predicts incident age-related eye diseases. *npj Aging* (2026). https://doi.org/10.1038/s41514-026-00418-1

Jia-Yan Kai, Shi-Yi Gong, Dan-Lin Li, Carla Lanca, Andrzej Grzybowski, Chaofu Ke & Chen-Wei Pan

Title Page

A Multimodal Ocular Aging Index Reveals Proteomic Pathways and Predicts Incident Age-Related Eye Diseases

Jia-Yan Kai[1], Shi-Yi Gong[1], Dan-Lin Li[1], Carla Lanca[2, 3], Andrzej Grzybowski[4], Chaofu Ke[1], Chen-Wei Pan[1, *]

[1] School of Public Health, Suzhou Medical College of Soochow University, Suzhou, China;

[2] Division of Science, New York University Abu Dhabi, Abu Dhabi, UAE;

[3] Comprehensive Health Research Center (CHRC), Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Lisboa, Portugal;

[4] Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland.

Running head: Multimodal Ocular Aging Index

***Correspondence to:**

Chen-Wei Pan, MD, PhD.

School of Public Health, Suzhou Medical College of Soochow University

199 Ren Ai Road, Suzhou, China 215123

Tel: +86 0512 65883907 / Email: pcwonly@gmail.com

Word count: 3915

Abstract

Chronological age incompletely captures heterogeneity in biological aging. In this prospective study of 45,819 UK Biobank participants, we developed a multimodal ocular aging index (MOAI) by integrating ophthalmic phenotypes with plasma proteomic and metabolomic profiles using machine learning. The MOAI quantifies

divergence between ocular biological and chronological age. Over 13.80 years of follow-up, accelerated ocular aging was significantly associated with higher risks of incident age-related macular degeneration and cataract, even after adjustment for chronological age and established risk factors. Incorporation of the MOAI significantly improved risk reclassification beyond traditional predictors. Explainable modeling and pathway enrichment analyses identified inflammation-related proteins and pathways, including cytokine–cytokine receptor interactions and PI3K–Akt signaling, as key drivers of accelerated ocular aging. These findings establish a multimodal framework for quantifying organ-specific biological aging, link ocular aging to systemic inflammatory processes, and highlight the eye as a sensitive readout of aging biology with implications for healthspan.

Keywords: Ocular Aging; Multi-omics; Age-related macular degeneration; Cataract; UK Biobank

Introduction

An unprecedented global demographic transition towards an older population is underway, making age-related health conditions a major concern for public health.[1] Among these, age-related eye diseases, particularly age-related macular degeneration (AMD) and cataract, are the main causes of severe visual impairment and blindness worldwide, imposing substantial personal, societal, and economic burdens.[2] Crucially, while chronological age is the primary non-modifiable risk factor for these conditions, it is a poor predictor of individual disease risk and onset time.[3] There is significant variability in the rate of functional decline and disease susceptibility among individuals of the same age, suggesting that biological age of an organ system may be more informative than time-lived alone.[3] This underscores a central unmet need: the lack of robust, quantitative biomarkers capable of capturing the biological pace of tissue-specific aging, particularly for the eye, to enable early risk stratification and mechanistic insight.

The quest to quantify biological aging has led to the development of various "aging clocks," predominantly based on omics data like DNA methylation, which predict chronological age and deviations from it ("age acceleration") correlate with morbidity and mortality.[4] However, these systemic measures may not accurately reflect the aging trajectory of specific organs, which can diverge due to tissue-specific physiology, environmental exposures, and genetics.[5] The eye presents a uniquely advantageous model for studying organ-specific aging. It is the only organ where neural tissue, microvasculature, and connective structures can be directly and non-invasively imaged in vivo.[6] Moreover, its compartmentalized fluids (e.g., aqueous, vitreous) and accessibility make it a potential window into systemic and local biological processes.[7] Despite this potential, current approaches often rely on single modalities, such as structural imaging, lacking integration with deep molecular phenotyping to construct a comprehensive, biologically-grounded metric of ocular aging.[8, 9]

Notably, a significant gap exists between the rich, multi-modal data available in large biobanks and the development of integrative biomarkers of ocular aging. While UK Biobank offers an unparalleled resource combining detailed ophthalmic examinations, plasma proteomics, and metabolomics for tens of thousands of individuals,[10] these data streams have not been synergistically used to define a holistic "ocular age". Furthermore, the biological pathways underpinning accelerated ocular aging and its direct, prospective link to the incidence of major age-related eye diseases remain poorly characterized. Therefore, we aimed to develop and validate a “multimodal ocular aging index (MOAI)" using machine learning on integrated ophthalmic, proteomic, and metabolomic data from the UK Biobank. The MOAI is defined as a systemic biological age estimator that is specifically informative for ocular aging, rather than a pure tissue-specific clock. Our objectives were: (1) to construct a robust model predicting ocular

biological age and derive the MOAI representing individual aging rate; (2) to decipher the key protein drivers and enriched biological pathways defining accelerated ocular aging; and (3) to prospectively evaluate the association of the MOAI with the future risk of developing AMD and cataract. This study establishes a novel, biologically interpretable biomarker of ocular aging through the integration of multi-omics data, offering a new lens through which to view aging mechanisms and clinical risk.

Results

A total of 45,819 UK Biobank participants with multimodal data (ophthalmic examination, plasma proteomics, and metabolomics) were included in the final analysis cohort for model application and association studies (Figure 1). A subset of relatively healthy participants (n=10,582) was used for model development and internal validation via ten-fold cross-validation. Baseline characteristics of the participants in the training set (n=7,409) and validation set (n=3,173) are detailed in Table S1 and S2, with similar distributions for age and sex across both groups.

The XGBoost regression model, trained on the multimodal data, demonstrated good performance in predicting chronological age within the healthy development set, with the training set showing a mean Root Mean Square Error (RMSE) of 3.04, Mean Absolute Error (MAE) of 2.40 years, and R^2 of 0.86. In the validation set, the model showed RMSE of 3.08, MAE of 2.43 years, and R^2 of 0.85. The predicted ocular age showed a strong correlation with chronological age (Figure 2A), with a correlation coefficient of $r = 0.92$ ($P<0.00001$). In ablation analyses, the multimodal model achieved the best predictive performance, showing slightly lower RMSE compared with models based on single modalities alone (Table S3), indicating incremental value from integrating multiple data types.

When applied to the entire analysis cohort (n=45,819), the distribution of the MOAI generated was approximately normal with a mean of -0.25 years (standard deviation, SD=3.06) and a range from -22.03 to 24.08 years (Figure 2B). Baseline characteristics of the analysis cohort, stratified by quartiles of the MOAI, are presented in Table 1. Participants in the highest quartile (Q4, accelerated aging) were on average younger chronologically but exhibited a higher prevalence of current smoking and socio-economic deprivation compared to those in the lowest quartile (Q1, decelerated aging), highlighting the need for multivariate adjustment in subsequent analyses.

Over a median follow-up of 13.80 years, 832 (1.82%) incident AMD cases and 5,288 (11.59%) incident cataract cases were identified, corresponding to an incidence rate of 1.36 per 1,000 person-years (95% confidence interval [CI]: 1.27–1.46) for AMD and 8.99 per 1,000 person-years (95% CI: 8.75–9.24) for cataract. In the fully adjusted Cox proportional hazards model, a one-year increase in the MOAI was significantly associated with a higher risk of both diseases (hazard ratio [HR]=1.05, 95% CI 1.03-1.08; P=0.0001 for AMD; HR=1.02, 95% CI 1.00-1.03; P=0.004 for cataract) (Table 2). A clear dose-response relationship was observed when analyzed by quartiles of MOAI. Compared to individuals in Q1 (slowest aging), those in Q2 and Q3 showed non-significantly increased risks. However, participants in Q4 (accelerated aging) had a markedly and significantly elevated risk for both AMD (HR=1.53, 95% CI 1.23-1.91; P=0.0001) and cataract (HR=1.12, 95% CI 1.02-1.22; P=0.01) even after adjusting for chronological age and other covariates. The p-values for trend were highly significant (P=0.0005 for AMD, P=0.03 for cataract). These associations are visually supported by dose-response curves (Figure S1) and distinct separation in Kaplan-Meier survival curves across quartiles (Figure S2).

Although we did not observe significant differences in the area under the receiver operating characteristic (ROC)

curves (AUC) (Figure S3), adding the MOAI to a model incorporating all traditional risk factors significantly improved risk reclassification for both AMD and cataract, as evidenced by a positive and significant net reclassification index (NRI) (Table S4). The NRI was 12.8% (95% CI: 7.9%–17.8%; P<0.001) for AMD and 6.5% (95% CI: 4.7%–8.5%; P<0.001) for cataract. This reclassification primarily involved upward reclassification of individuals at high risk without predefined risk thresholds, shifting them into more accurate risk categories, which could potentially enable earlier interventions.

Subgroup analyses generally showed consistent positive associations between the MOAI and the risk of both AMD (Table S5) and cataract (Table S6). The observed associations remained robust in sensitivity analyses that excluded participants diagnosed with AMD or cataract within the first 2 years (Table S7) or 5 years (Table S8) of follow-up, reducing the likelihood of reverse causality. After adjusting for chronological age by residualizing the ocular age gap, the age-adjusted metric showed no significant correlation with chronological age ($r \approx 0$), and its associations with AMD and cataract remained materially unchanged (Table S9), supporting the robustness of our findings.

The final XG boost model incorporated a total of 666 features, including ocular variables such as average ganglion cell-inner plexiform layer (GC-IPL) thickness, spherical power, logMAR visual acuity, and average mean spherical equivalent (avMSE). Additionally, 13 metabolites were included in the model, covering key metabolic pathways such as glucose-lactate metabolism, fatty acids, and amino acids metabolism. The remaining features were predominantly proteins, including numerous inflammation-related biomarkers such as tumor necrosis factor receptor superfamily member 27 (TNFRSF27) and elastin. The SHapley Additive exPlanations (SHAP) summary plot identified the top 15 features contributing to the ocular age prediction, all of which were plasma proteins

(Figure 3A). A positive SHAP value indicates that a higher measured level of that protein was associated with an increased predicted ocular age. Notably, proteins like TNFRSF27, elastin, and latent-transforming growth factor beta-binding protein 2 were among the most influential. When aggregated by modality, proteomic features accounted for the vast majority of model contribution (approximately 99% of total SHAP importance), whereas metabolomic and ophthalmic features contributed modestly. KEGG pathway enrichment analysis of the proteins included by the machine learning model revealed significant overrepresentation of several critical biological pathways linked to ocular aging (Figure 3B). The most significantly enriched pathways included the cytokine-cytokine receptor interaction, and the PI3K-Akt signaling pathway, both of which are crucial in inflammation and cell signaling processes involved in aging.[11, 12]

Discussion

In this large-scale study of over 45,000 individuals from the UK Biobank, we developed and validated a novel multimodal metric of organ-specific biological aging, named the MOAI. By integrating ophthalmic measurements with plasma proteomic and metabolomic profiles using a machine learning framework, we defined an index that quantifies the divergence of an individual's ocular biological age from their chronological age. Our results suggest that the MOAI is linked to specific biological processes, particularly inflammatory pathways, and may serve as an independent predictor of incident AMD and cataract.

Various methods have been previously developed to assess biological aging.[4, 8, 13] One prominent approach involves the use of fundus images taken by optical coherence tomography (OCT). A recent cohort study in China applied deep learning techniques to predict retinal age from fundus images, which demonstrated that the retinal age gap

was significantly associated with mortality risk.[8] They also performed Mendelian randomization analysis and identified that factors such as anemia, glycated hemoglobin, and inflammation-related hemocytes may be linked to accelerated retinal aging.[9] Proteomics, another essential tool in aging research, has also been leveraged to construct aging clocks. A study in the USA developed a cellular proteomic age model using XG Boost, combining high-resolution proteomic profiling with single-cell transcriptomics to predict molecular aging in ocular diseases.[7] This model revealed that several inflammation-related proteins, including VEGF, ANGPT2, and MMP19, were closely linked to aging processes and played key roles in diseases such as uveitis and diabetic retinopathy.[7] These findings align with our results that inflammatory biomarkers, such as TNFRSF27 and elastin, may play a critical role in driving ocular aging and eye disease progression. Another widely used method is metabolomics, where metabolic biomarkers are utilized to predict biological age. For instance, a study using UK Biobank metabolomic data developed a metabolomic age that was significantly associated with the risk of AMD, with a hazard ratio of 1.02 (95% CI: 1.00-1.04).[14] The top 10 leading contributors of metabolomic age in this study included metabolites such as omega-3 fatty acids, docosahexaenoic acid (DHA), citrate, and various triglycerides.[14] Notably, several of these metabolites, including omega-3 fatty acids and DHA, were also included in our model, further validating the importance of these metabolic markers in aging-related processes. While these existing studies focus on single modalities, our approach brings together multiple omics layers, providing a more integrated framework for understanding ocular aging. Although the practical application of multimodal testing in clinical settings requires further exploration, the integration of ophthalmic, proteomic, and metabolomic data offers potential for more comprehensive insights into the biological mechanisms of aging. Future research will be necessary to evaluate the feasibility and clinical applicability of multimodal aging models, particularly in the context of personalized medicine and early disease detection.

The biological plausibility of the MOAI is supported by the protein signatures and pathway enrichments identified in our study. Our findings underscore the significant role of inflammation-related proteins, such as TNFRSF27, elastin, and GDF15, which have been shown to play crucial roles in aging and inflammatory processes. TNFRSF27, a member of the tumor necrosis factor receptor superfamily, is involved in immune signaling and may contribute to chronic inflammatory activation.[15] GDF15, a stress-responsive cytokine, has been linked to mitochondrial dysfunction and cellular aging.[16] Elastin, a key component of extracellular matrix, is essential for tissue elasticity, and its degradation is associated with age-related structural decline in ocular tissues.[17] These proteins are particularly relevant to the pathogenesis of AMD and cataract, where chronic inflammation and immune dysregulation are central.[18, 19] In AMD, for example, the activation of the inflammatory cascade in the retinal pigment epithelium (RPE) and choroid promotes oxidative stress, angiogenesis, and immune cell infiltration, which are hallmarks of AMD pathology.[18] Similarly, in cataract formation, inflammation in the lens could lead to the breakdown of lens proteins and the accumulation of cellular debris, accelerating cataract development.[20] Our results suggest that accelerated ocular aging, detectable in peripheral blood, may be characterized by a pre-clinical state of innate immune dysregulation that precedes clinical diagnosis. The use of SHAP values moves beyond a "black box" prediction, providing mechanistic hypotheses for what drives accelerated ocular aging at a molecular level. Furthermore, pathway enrichment analysis revealed significant overrepresentation of several key biological pathways, particularly cytokine-cytokine receptor interactions and the PI3K-Akt signaling pathway. These pathways are pivotal in regulating immune responses, cellular survival, and tissue remodeling.[11, 12] Notably, PI3K-Akt signaling plays a critical role in regulating autophagy and survival of retinal pigment epithelial cells.[21] Dysregulation of this pathway may impair autophagic clearance, leading to accumulation of cellular debris and

contributing to AMD pathogenesis.[21] Additionally, in the lens, altered PI3K-Akt signaling may disrupt lens epithelial cell homeostasis and promote protein aggregation, thereby facilitating cataract formation.[22] The dysregulation of the cytokine-cytokine receptor interactions pathway may further amplify inflammatory signaling in both retinal and lens tissues, exacerbating disease progression.[11] Disruption of these pathways may lead to retinal cell apoptosis, impaired vascular integrity, and lens cell dysfunction, all of which are key contributors to the progression of both AMD and cataracts.[23]

The clinical translation of our findings is twofold. First, for risk stratification, the MOAI offers an additional layer of information beyond chronological age. The consistent, dose-response relationship with disease risk, especially the markedly high hazard for individuals in the top quartile, suggests its utility in identifying high-risk subgroups for targeted monitoring or early intervention. Although no significant improvement in AUC was observed after the addition of MOAI, this may be due to the insensitivity of AUC to incremental information when baseline models already perform well. In contrast, NRI is more sensitive to changes in individual risk classification and provides complementary information beyond discrimination metrics. The significant improvement in NRI therefore suggests that MOAI enhances the ability to correctly reclassify individuals into more appropriate risk categories. The MOAI may be particularly useful in identifying individuals with accelerated biological aging not captured by chronological age, such as in younger individuals or those at early stages of disease. Second, our work illuminates potential intervention targets. The identified proteins and pathways, such as specific nodes highlighted in the SHAP analysis, represent candidate targets for pharmaceutical or lifestyle interventions aimed at slowing ocular aging itself. Moreover, the MOAI may serve as a sensitive intermediate endpoint in clinical trials testing geroprotective or disease-modifying therapies, potentially reducing the required trial duration and size compared to using hard

clinical endpoints like vision loss.

Several limitations of our study warrant consideration. First, the UK Biobank population is predominantly of White European ancestry and may not be representative of other ethnic groups where the prevalence and risk factors for age-related eye diseases can differ. External validation in diverse cohorts is essential. Second, while we utilized a high-throughput proteomic platform, it does not cover the full spectrum of the human proteome, and important protein drivers may have been missed. Third, the observational design cannot definitively establish causality between the accelerated aging phenotype and disease onset. Finally, participants in the 'relatively healthy' subset were classified as free from self-reported diseases at baseline, which may introduce misclassification bias due to undiagnosed or unreported conditions.

In conclusion, we have defined a biologically-grounded, multimodal biomarker of ocular aging that captures individual variability in the rate of aging within a critical sensory organ. The MOAI transcends chronological age by reflecting underlying dysregulation in immune and inflammatory pathways and provides a powerful, independent risk stratifier for the two most common age-related eye diseases. This study establishes a new paradigm for quantifying tissue-specific aging and paves the way for early prediction, mechanistic discovery, and the development of novel interventions aimed at promoting ocular health span.

Methods

Population and Study Design

The UK Biobank is a large, ongoing, prospective cohort study of over half a million residents aged 37-73 years

recruited across the UK in 2006-2010 at 22 assessment centers. This population-based study provides a wealth of phenotypic, genomic, and multi-omics data, making it a valuable resource for health research. At baseline, participants completed touchscreen questionnaires, underwent standardized physical measurements, and provided blood samples. Health outcomes are tracked primarily through linkage to hospital admissions and death registry records. The detailed study protocol and data can be accessed elsewhere (www.ukbiobank.ac.uk).[10] Ethical approval was obtained from the North West Multi-Centre Research Ethics Committee (REC reference: 11/NW/03820), with all participants providing written informed consent. The present study was conducted under application number 60651 following the Declaration of Helsinki.

Out of the initial subset of UK Biobank participants with available proteomic, clinical ophthalmic, and metabolomic data (n=53,073), we excluded individuals with major eye diseases such as AMD, cataract, or glaucoma at baseline (n=4,086), as well as those of non-European ancestry (n=3,168). This resulted in a final cohort of 45,819 individuals. Among this group, a subset of relatively healthy participants (n=10,582) was selected for training the machine learning model. These participants were defined as individuals free from any self-reported diseases at baseline. These self-reported conditions included a broad spectrum of systemic and non-ocular diseases (e.g., cardiovascular, metabolic, and neoplastic diseases), as detailed in Table S10. The definitions of baseline and incident eye diseases are provided in Table S11, and the overall study workflow is depicted in Figure 1.

Ophthalmic Assessment

The UK Biobank incorporated a set of ophthalmic examinations into its baseline assessment in 2009. These tests were designed to characterize multiple aspects of visual function and ocular structure in participants.[24] The range

of eye measures collected spans four principal domains: visual acuity, refractive error and corneal parameters, intraocular pressure (IOP) and corneal biomechanics, and retinal imaging via OCT. Visual acuity was assessed using calibrated logMAR charts under standardized conditions, recording the smallest letter size correctly identified at a specified viewing distance for each eye. Non-cycloplegic refractive status and keratometry were obtained using an autorefractor device (Tomey RC-5000). Key parameters included spherical power, cylindrical power, and several indices that capture corneal symmetry and regularity. IOP and corneal biomechanical properties were measured using an air-puff tonometer (Reichert Ocular Response Analyzer), capturing parameters such as corneal resistance factor and corneal-compensated IOP. High-resolution OCT scans were taken with the Topcon 3D OCT-1000 Mk2, providing 3D imaging of the retina and derived measures of retinal layer thickness and morphology. Data from the right eye were prioritized for analysis and measurements from the left eye were included only in cases where right-eye data were missing.

Blood Proteomics

The proteomic profiling in this study was carried out as part of the UK Biobank Proteomics Project, which collected blood samples in EDTA vacutainers (9ml) during the baseline assessment. Plasma was extracted immediately through centrifugation (10 minutes, 2,500g, 4°C) and was stored at −80°C for storage. A subset of 53,073 UK Biobank participants was randomly selected for proteomic analysis. Using Proximity Extension Assay (PEA) technology, there were 2,923 proteins initially quantified after rigorous quality control procedures, covering a broad spectrum of biological systems, including cardiometabolic, neurological, inflammatory, and oncological proteins. Four of the proteins (Cystatin-SN, Glioma pathogenesis-related protein 1, Procollagen C-endopeptidase enhancer 1, and Nucleophosmin) were excluded due to over 25% missing data, and a final set of 2,919 proteins was retained

for analysis. The PEA technology provides a high-throughput, antibody-based approach that is highly reproducible, and the proteomic data have been proven reliable in previous studies.[25, 26]

NMR Spectroscopy Metabolomic

Plasma metabolomic biomarkers were quantified by the nuclear magnetic resonance (NMR)-based platform from Nightingale Health, which employed standardized procedures for sample preparation, quality control, and data analysis. This approach has been shown to provide data with high repeatability and high correlation (r >0.9) to accredited clinical chemistry assays.[27] For this study, we used the baseline measurements from 117,121 randomly selected participants, which in total identified 168 circulating metabolites. These metabolites span a wide range of metabolic pathways, including glycolysis metabolites, amino acids, ketone bodies, lipoprotein lipids, and fatty acids.

Machine Learning Model for Ocular Age Prediction

In accordance with prior studies,[7, 13, 28] we assumed that ocular biological age equals chronological age in healthy individuals who experience normal ageing. Therefore, we used the subset of relatively healthy participants (n=10,582) to train the machine learning model, with chronological age as the reference variable representing the ground truth for ocular age. The first step in model development was to standardize all features (mean = 0, standard deviation = 1) to ensure comparability, except for chronological age and sex, which were retained in their original form. The dataset was then randomly divided into training and test sets in a 7:3 proportion. We tested multiple machine learning models, including ElasticNet, XGBoost, and Random Forest, and performed ten-fold cross-validation on the training set to evaluate model performance based on key metrics such as MAE and RMSE. The hyperparameters for each model were fine-tuned using grid search, and the model with the lowest RMSE on the

test set was chosen. Among the tested models, the XGBoost algorithm showed the best performance in predicting ocular age and was thus applied to build the optimal model using the following hyperparameters: nrounds=2666, eta=0.05, gamma=0.2, max_depth=3, subsample=0.7, colsample_bytree=0.5, and min_child_weight=5. Missing values in the input features were handled inherently by the XGBoost algorithm, which can accommodate missing data during tree construction without explicit imputation. For feature selection, feature importance was first evaluated within each fold of a 10-fold cross-validation procedure using XGBoost, and variables identified in each fold were recorded. We then counted the frequency with which each feature was selected across folds and retained features that appeared in at least a predefined number of folds ("cutoff"). The cutoff of 10 (i.e., features consistently selected in all folds) was chosen as it yielded the lowest error of the model during cross-validation. While more lenient cutoffs retained more features, they did not improve cross-validated RMSE (Figure S4). The final model was then applied to the full cohort of 45,819 individuals (analysis dataset) to estimate the ocular age for each participant and to further investigate the association between the MOAI and risks of age-related eye disorders. Moreover, we performed an ablation analysis to compare the performance of models built on (A) ophthalmic data only, (B) proteomic data only, (C) metabolomic data only, and (D) multimodal data, all using the XGBoost algorithm.

Multimodal Ocular Aging Index Definition

The divergence between the predicted ocular age and their chronological age was defined as the MOAI. This index quantifies the discrepancy between the biological aging of the eye and the individual's actual age. A positive MOAI suggested accelerated aging of the eye relative to the individual's actual age. In contrast, a negative MOAI reflected a slower rate of ocular aging.

Ascertainment of Age-related Eye Disorders

Major age-related eye disorders, including AMD and cataract, were identified through hospital inpatient records and national death registries, based on the International Classification of Diseases codes (ICD-10: H35.3 for AMD; H25, H26, H28 for cataract). The follow-up duration was determined from the date of attending baseline assessment to the earliest of death, the censoring date (19 December 2022), or the first occurrence of either eye disorder, whichever came first.

Statistical Analysis

Restricted cubic spline (RCS) analyses with four nodes were performed to explore the dose-response relationship between the MOAI and the risk of developing AMD and cataract. To assess the association between MOAI and the risk of incident AMD and cataract, Cox proportional hazards models were employed, with HRs along with their 95% CIs. We evaluated the proportional hazards assumption using the Schoenfeld residuals method, with no violations found. We ran two models: model 1 was adjusted for chronological age and sex; model 2 was adjusted for model 1 plus body mass index (BMI), socioeconomic status, education level, smoking status, physical activity, and alcohol consumption. Specific definitions of these covariates have been described elsewhere.[29] For categorical covariates, missing values were handled by introducing a separate category to retain all participants in the analysis, rather than performing imputation.

To assess the predictive ability of the MOAI for AMD and cataract, we plotted ROC curves for both the traditional predictive model and the model that included MOAI as an additional predictor for the 10-year risk of incident AMD

and cataract. The traditional model, in line with previous studies, included well-established risk factors such as chronological age, sex, BMI, socioeconomic status, education, alcohol consumption, smoking status, and physical activity. The predictive ability of both models was quantified by the AUC, where higher AUC values indicate better predictive performance. We also assessed the risk reclassification performance of the MOAI by calculating the continuous (category-free) net reclassification index (NRI) and integrated discrimination improvement (IDI). These metrics evaluate the extent to which the addition of MOAI improves risk prediction without relying on predefined risk categories.

To examine potential interaction effects, we added interaction terms in the Cox regression models, and subgroup analyses were conducted based on chronological age, sex, BMI, education level, physical activity, and smoking status. Sensitivity analyses were performed to examine the robustness of our results. To account for potential reverse causality, we conducted landmark analyses excluding participants who developed AMD or cataract within 2 or 5 years of recruitment. In addition, to address potential residual dependence on chronological age, we regressed the ocular age gap on chronological age and derived an age-adjusted ocular age gap using the residuals from this model. We then repeated the Cox regression analyses to evaluate the associations of the age-adjusted ocular age gap with AMD and cataract.

For the features included in the XG Boost model, we calculated SHAP values to assess the importance of each variable in predicting ocular age. We also summarized SHAP values by data modality (ophthalmic, proteomic, and metabolomic) to quantify their relative contributions to the model. Additionally, to investigate the biological pathways driving the ocular aging process, we performed KEGG pathway enrichment analysis on the proteins

included in the final model.

All statistical analyses were carried out using SAS 9.4 (SAS Institute Inc., Cary, North Carolina, USA) and R 4.3.3 (The R Foundation for Statistical Computing, Vienna, Austria). A P-value of < 0.05 (two-tailed) was considered statistically significant.

Data availability: The datasets generated and analysed during the current study are available in the UK Biobank repository, http://www.ukbiobank.ac.uk/.

Code Availability: The simulation codes of R version 4.3.3 for statistical analysis of this study are available at https://github.com/jiayankathryn-sudo/Multimodal-Ocular-Aging-Index.

Acknowledgements: This research was supported by the National Key R&D Program of China (2024YFC2510800, 2024YFC2510801). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would like to thank all participants and staff from the UK Biobank cohort.

Author Contributions: CWP had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conception and design: CWP; Acquisition, analysis, or interpretation of data: CFK, JYK, SYG, DLL, CL, and AG; Drafting of the manuscript: JYK; Critical revision of the manuscript for important intellectual content: CWP, CFK, CL, AG, and JYK.

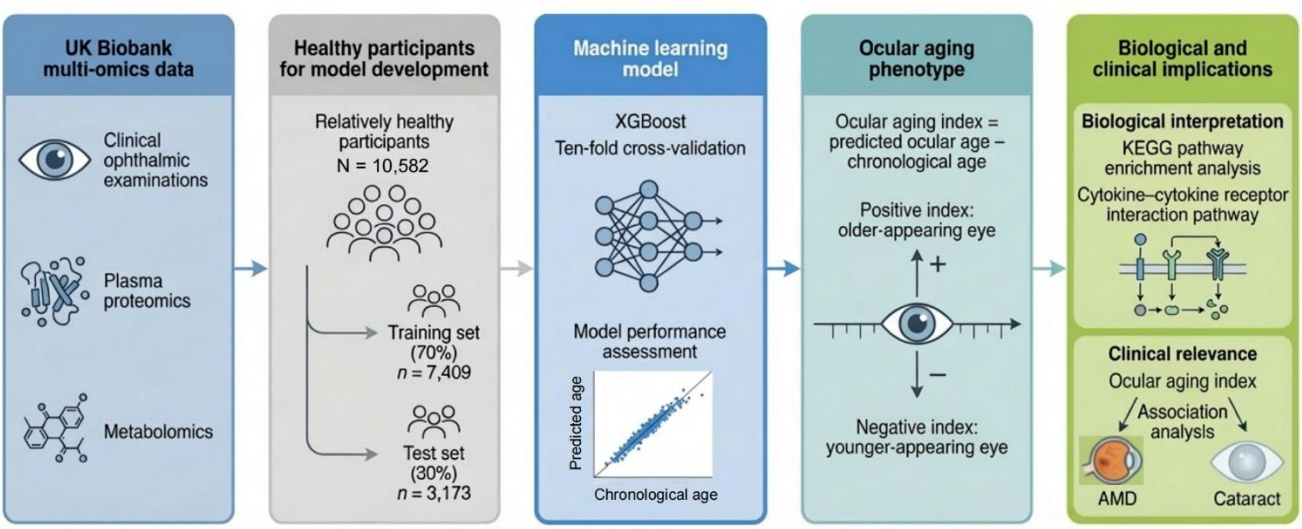
Competing Interests: The authors declare no competing financial or non-financial interests.

References

1. Kanasi E, Ayilavarapu S, Jones J. The aging population: demographics and the biology of aging. Periodontol 2000 2016;72(1):13-8.
2. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. Lancet Glob Health 2021;9(2):e144-e60.
3. Lowsky DJ, Olshansky SJ, Bhattacharya J, Goldman DP. Heterogeneity in healthy aging. J Gerontol A Biol Sci Med Sci 2014;69(6):640-9.
4. Duan R, Fu Q, Sun Y, Li Q. Epigenetic clock: A promising biomarker and practical tool in aging. Ageing Res Rev 2022;81:101743.
5. Yang JH, Hayano M, Griffin PT, et al. Loss of epigenetic information as a cause of mammalian aging. Cell 2024;187(5):1312-3.
6. Patton N, Aslam TM, MacGillivray T, et al. Retinal image analysis: concepts, applications and potential. Prog Retin Eye Res 2006;25(1):99-127.
7. Wolf J, Rasmussen DK, Sun YJ, et al. Liquid-biopsy proteomics combined with AI identifies cellular drivers of eye aging and disease in vivo. Cell 2023;186(22):4868-84.e12.
8. Zhu Z, Shi D, Guankai P, et al. Retinal age gap as a predictive biomarker for mortality risk. Br J Ophthalmol 2023;107(4):547-54.
9. Huang Y, Syed MG, Chen R, et al. Genomic determinants of biological age estimated by deep learning applied to retinal images. Geroscience 2025;47(2):2613-29.
10. Collins R. What makes UK Biobank special? Lancet 2012;379(9822):1173-4.
11. Bhol NK, Bhanjadeo MM, Singh AK, et al. The interplay between cytokines, inflammation, and antioxidants: mechanistic insights and therapeutic potentials of various antioxidants and anti-cytokine compounds. Biomedicine & Pharmacotherapy 2024;178:117177.
12. Glaviano A, Foo ASC, Lam HY, et al. PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. Mol Cancer 2023;22(1):138.
13. Li Y, Huang Q, Jiang J, et al. Large language model-based biological age prediction in large-scale populations. Nat Med 2025;31(9):2977-90.
14. Shang X, Liu J, Zhu Z, et al. Metabolomic age and risk of 50 chronic diseases in community-dwelling adults: A prospective cohort study. Aging Cell 2024;23(5):e14125.
15. Xu J. The role of tumor necrosis factor receptor superfamily in cancer: insights into oncogenesis, progression, and therapeutic strategies. NPJ Precis Oncol 2025;9(1):275.
16. Lee A, Vidhya SN, Hong A, et al. Growth differentiation factor-15 as a clinical biomarker of frailty, sarcopenia and functional decline: A systematic literature review. Ageing Res Rev 2026;118:103085.
17. Yeo GC, Mithieux SM, Weiss AS. The elastin matrix in tissue engineering and regeneration. Current Opinion in Biomedical Engineering 2018;6:27-32.
18. Datta S, Cano M, Ebrahimi K, et al. The impact of oxidative stress and inflammation on RPE degeneration in non-neovascular AMD. Prog Retin Eye Res 2017;60:201-18.
19. Li X, Du GL, Wu SN, et al. Association between systemic immune inflammation index and cataract incidence from 2005

to 2008. Sci Rep 2025;15(1):499.
20. Thompson B, Davidson EA, Chen Y, et al. Oxidative stress induces inflammation of lens cells and triggers immune surveillance of ocular tissues. Chemico-Biological Interactions 2022;355:109804.
21. Fang J, Huang Y, Li B, Du Y. Role and regulation of kinases in age-related macular degeneration. J Transl Med 2026;24(1):227.
22. Wang Y, Yang W, Zuo H, et al. Icariin interacts with IGFBP3 to alleviate diabetic cataract through PI3K/AKT signaling pathway. iScience 2025;28(7):112796.
23. Saddala MS, Lennikov A, Mukwaya A, Huang H. Transcriptome-Wide Analysis of CXCR5 Deficient Retinal Pigment Epithelial (RPE) Cells Reveals Molecular Signatures of RPE Homeostasis. Biomedicines 2020;8(6).
24. Chua SYL, Thomas D, Allen N, et al. Cohort profile: design and methods in the eye and vision consortium of UK Biobank. BMJ Open 2019;9(2):e025077.
25. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature 2023;622(7982):329-38.
26. Eldjarn GH, Ferkingstad E, Lund SH, et al. Large-scale plasma proteomics comparisons through genetics and disease associations. Nature 2023;622(7982):348-58.
27. Würtz P, Kangas AJ, Soininen P, et al. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. Am J Epidemiol 2017;186(9):1084-96.
28. Wang J, Knol MJ, Tiulpin A, et al. Gray Matter Age Prediction as a Biomarker for Risk of Dementia. Proc Natl Acad Sci U S A 2019;116(42):21213-8.
29. Kai JY, Dong XX, Li DL, et al. Association of Plasma Omega-3 Fatty Acids With POAG. Invest Ophthalmol Vis Sci 2025;66(12):3.

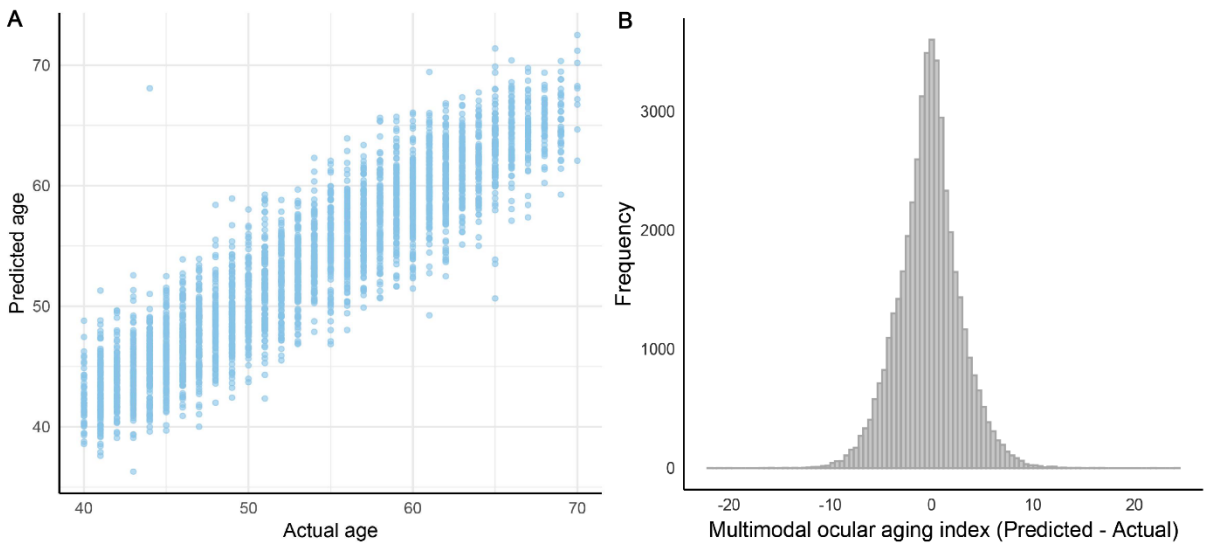
Figure 1. Overview of the study workflow.



Multi-modal data from the UK Biobank, including clinical ophthalmic examinations, plasma proteomics, and metabolomics, were integrated to calculate the multimodal ocular aging index using the XG boost model. The selection of XG boost models was based on ten-fold cross-validation. The XG boost model was then applied to

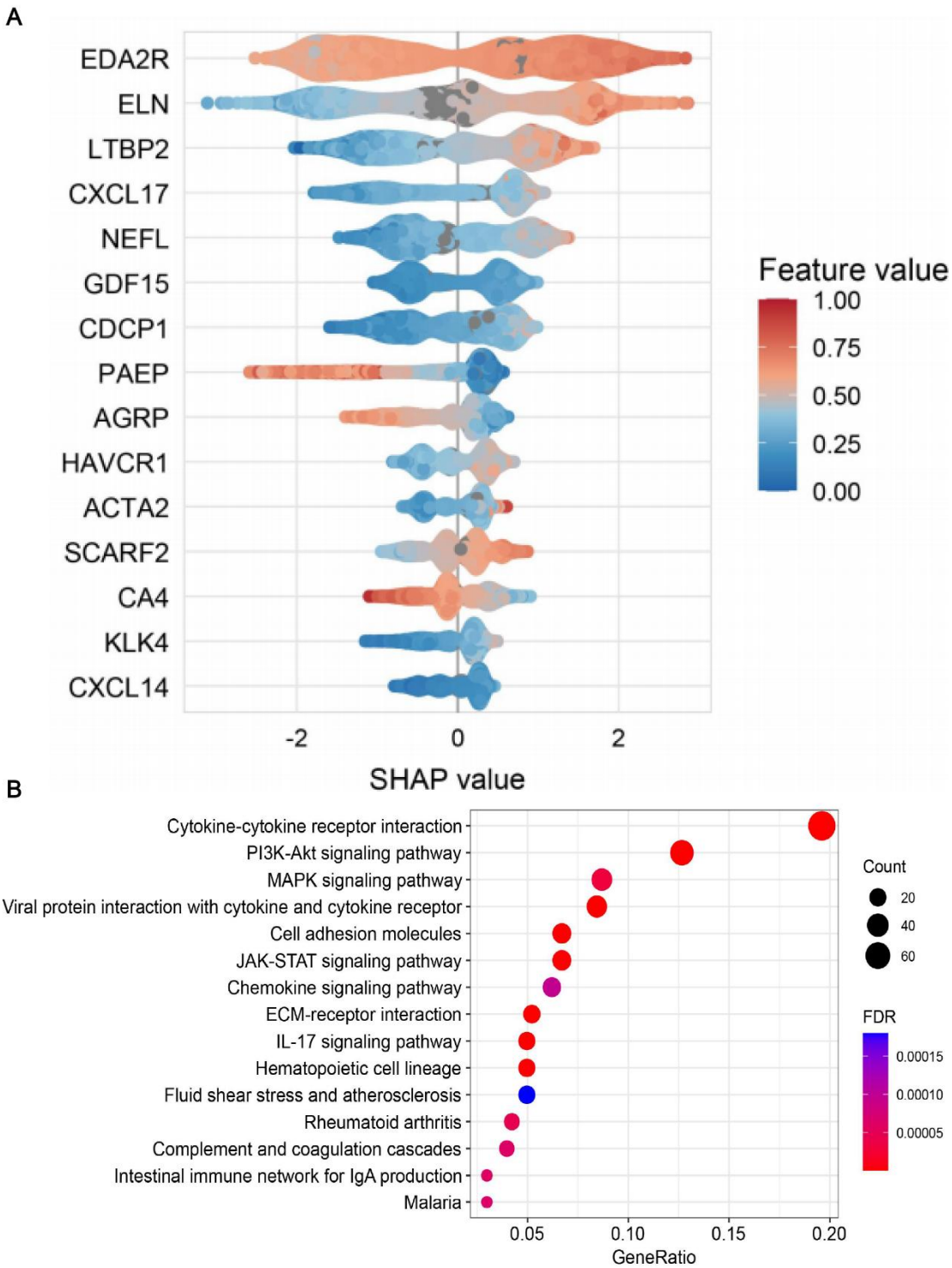
make ocular age predictions for participants in the analysis dataset. The multimodal ocular aging index was defined as the difference between predicted ocular age and chronological age, with positive values indicating accelerated ocular aging and negative values indicating decelerated ocular aging. Downstream analyses included pathway enrichment analysis to explore biological mechanisms and association analyses using Cox regression to evaluate relationships between the multimodal ocular aging index and age-related eye diseases, including age-related macular degeneration and cataract.

Figure 2. Summary of ocular age prediction.



(A) Performance of the XG boost model. Scatterplot depicting correlation of predicted ocular age with chronological age in relatively healthy participants. (B) Distribution of multimodal ocular aging index (mean=-0.25, SD=3.06, range: -22.03~24.08).

Figure 3. Machine learning–derived protein signatures and pathway enrichment underlying ocular aging.



(A) SHAP summary plot of the top 15 features contributing to ocular aging prediction, all of which were proteins. Each dot represents a SHAP value for a single observation, reflecting how much that feature contributed to the model output for that sample. The color gradient indicates the original feature value (blue = low, red = high). A positive SHAP value means that the feature increases the predicted risk (or outcome value), whereas a negative SHAP value means the opposite. (B) KEGG pathway enrichment analysis of proteins identified by the XG boost

model.

EDA2R: Tumor necrosis factor receptor superfamily member 27; ELN: Elastin; LTBP2: Latent-transforming growth factor beta-binding protein 2; CXCL17: C-X-C motif chemokine 17; NEFL: Neurofilament light polypeptide; GDF15: Growth/differentiation factor 15; CDCP1: CUB domain-containing protein 1; PAEP: Glycodelin; AGRP: Agouti-related protein; HAVCR1: Hepatitis A virus cellular receptor 1; ACTA2: Actin, aortic smooth muscle; SCARF2: Scavenger receptor class F member 2; CA4: Carbonic anhydrase 4; KLK4: Kallikrein-4; CXCL14: C-X-C motif chemokine 14.

Table 1. Baseline characteristics of study participants stratified by quartiles of multimodal ocular aging index.

Characteristics	Total	Quartiles of multimodal ocular aging index			
		Q1 (≤-2.05)	Q2 (-2.05~-0.21]	Q3 (-0.21~1.49]	Q4 (>1.49)
N	45,819	11,455	11,455	11,454	11,455
Age, years	56.75 (8.13)	61.75 (6.41)	57.87 (7.65)	54.85 (7.97)	52.54 (7.32)
Female sex, n (%)	24,706 (53.92)	5,953 (51.97)	6,249 (54.55)	6,302 (55.02)	6,202 (54.14)
BMI, kg/m^2	27.39 (4.75)	27.88 (4.76)	27.15 (4.49)	27.03 (4.66)	27.48 (5.03)
Education level, n (%)					
High	19,927 (43.49)	4,596 (40.12)	5,152 (44.98)	5,121 (44.71)	5,058 (44.16)
Low	25,256 (55.12)	6,681 (58.32)	6,121 (53.44)	6,188 (54.02)	6,266 (54.70)
Townsend score, n (%)					
Q1 (least deprived)	11,532 (25.17)	2,935 (25.62)	3,059 (26.70)	2,922 (25.51)	2,616 (22.84)
Q2	11,452 (24.99)	2,958 (25.82)	2,894 (25.26)	2,918 (25.48)	2,682 (23.41)
Q3	11,287 (24.63)	2,853 (24.91)	2,803 (24.47)	2,801 (24.45)	2,830 (24.71)
Q4 (most deprived)	11,493 (25.08)	2,697 (23.54)	2,684 (23.43)	2,804 (24.48)	3,308 (28.88)
Smoking status, n (%)					
Never smoked	24,496 (53.46)	5,911 (51.60)	6,255 (54.60)	6,380 (55.70)	5,950 (51.94)
Former smoker	16,326 (35.63)	4,447 (38.82)	4,050 (35.36)	3,861 (33.71)	3,968 (34.64)
Current smoker	4,851 (10.59)	1,058 (9.24)	1,109 (9.68)	1,190 (10.39)	1,494 (13.04)
Alcohol intake, n (%)					
Never	3,114 (6.80)	889 (7.76)	770 (6.72)	661 (5.77)	794 (6.93)
Special occasions only	4,906 (10.71)	1,318 (11.51)	1,149 (10.03)	1,108 (9.67)	1,331 (11.62)
One to three times a month	5,057 (11.04)	1,217 (10.62)	1,192 (10.41)	1,293 (11.29)	1,355 (11.83)
Once or twice a week	12,234 (26.70)	2,877 (25.12)	3,114 (27.18)	3,174 (27.71)	3,069 (26.79)
Three or four times a week	10,854 (23.69)	2,564 (22.38)	2,759 (24.09)	2,865 (25.01)	2,666 (23.27)
Daily or almost daily	9,619 (20.99)	2,582 (22.54)	2,462 (21.49)	2,348 (20.50)	2,227 (19.44)
Physical activity, n (%)					
Low	7,009 (15.30)	1,766 (15.42)	1,603 (13.99)	1,706 (14.89)	1,934 (16.88)
Moderate	18,500 (40.38)	4,472 (39.04)	4,649 (40.58)	4,800 (41.91)	4,579 (39.97)
High	11,410 (24.90)	2,781 (24.28)	2,934 (25.61)	2,900 (25.32)	2,795 (24.40)

Data are presented as No. (%) for categorical variables and mean (standard deviation) or median (interquartile range) for continuous variables.

BMI: body mass index.

Table 2. Associations of multimodal ocular aging index (MOAI) with age-related macular degeneration (AMD) and cataract.

		Model 1		Model 2	
Variable	N_{case}/N_{total}	HR (95% CI)	*P*	HR (95% CI)	*P*
AMD					
MOAI, per one year	832/45,644	1.05 (1.02, 1.08)	0.0003	1.05 (1.03, 1.08)	0.0001
MOAI					
Quantile 1	308/11,455	Reference		Reference	
Quantile 2	224/11,455	1.08 (0.91, 1.29)	0.36	1.12 (0.94, 1.33)	0.21
Quantile 3	159/11,454	1.12 (0.91, 1.36)	0.28	1.15 (0.95, 1.41)	0.16
Quantile 4	142/11,455	1.51 (1.21, 1.88)	0.0002	1.53 (1.23, 1.91)	0.0001
P for trend		0.001		0.0005	
Cataract					
MOAI, per one year	5288/45,644	1.01 (1.00, 1.02)	0.008	1.02 (1.00, 1.03)	0.004
MOAI					
Quantile 1	2007/11,455	Reference		Reference	
Quantile 2	1408/11,455	0.98 (0.91, 1.05)	0.53	1.00 (0.93, 1.07)	0.89
Quantile 3	1054/11,454	1.00 (0.93, 1.08)	0.95	1.02 (0.95, 1.11)	0.56
Quantile 4	834/11,455	1.12 (1.02, 1.22)	0.01	1.12 (1.02, 1.22)	0.01
P for trend		0.05		0.03	

Model 1: adjusted for actual age and sex.

Model 2: adjusted for covariates in model 1 plus body mass index, education level, socioeconomic status, smoking status, alcohol consumption, and physical activity.

HR: hazard ratio; CI: confidence interval; AMD: age-related macular degeneration.