# Exposure accumulation drives age-dependent disease architectures and polygenic risk scores

Xilin Jiang[1,2,3,4,5,*], Alkes L. Price[3,4,6], John Danesh[1,2,7,8,9,10], Gil McVean[11], Michael Inouye[1,2,7,9,12,13], Sarah Urbut[4,14], Arun Durvasula[15,16,17,*]

1. British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge UK
2. Victor Phillip Dahdaleh Heart and Lung Research Institute University of Cambridge, Cambridge UK
3. Department of Epidemiology, Harvard T.H. Chan School of Public Health Boston MA USA
4. Broad Institute, Cambridge, MA USA
5. Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge UK
6. Department of Biostatistics, Harvard T.H. Chan School of Public Health Boston MA USA
7. British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK
8. National Institute for Health and Care Research Blood and Transplant Research Unit in Donor Health and Behaviour, University of Cambridge, Cambridge, UK
9. Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK
10. Department of Human Genetics, Wellcome Sanger Institute, Hinxton, UK
11. Ellison Institute of Technology, Oxford, UK
12. Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
13. Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia
14. Cardiology Division, Massachusetts General Hospital, Boston, MA USA.
15. Division of Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA
16. Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA
17. Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California, USA

* Correspondence: xj262@medschl.cam.ac.uk (XJ); arun.durvasula@med.usc.edu (AD)

**Abstract**

Our understanding of the dependence of the genetic and environmental architecture of common diseases on age is incomplete. Here, we use longitudinal data to quantify age-dependent genetic and environmental components of complex traits and diseases. When applying our approach to 16 UK Biobank quantitative traits (average $N$=180K), we found environmental variance can accumulate with age, which leads to decreasing heritability with age, following an exposure accumulation model. We find heritability decreases with age for 5 traits, including systolic blood pressure and lung function (FEV1/FVC), with an average change of −17.8% with age per 10-years. When analyzing complex diseases, we found evidence that environmental variance in disease liability also accumulates with age for 5 of 9 diseases, with an average decrease of heritability −18.1% per 10 years. We found a liability threshold model with exposure accumulation explains 86% of decreasing PRS prediction power with age in 9 UK Biobank complex diseases (compared to 65% by the liability-threshold model alone, $p$ = 3e-11). Finally, we show that both genetic and non-genetic predictors have decreasing prediction accuracy with age in 61 predictor-disease pairs, consistent with simulations. Our results suggest ascertaining younger cohorts is more powerful for training both genetic and non-genetic risk prediction models.

## Introduction

The extent to which genetics and environmental exposures interact with age when contributing to phenotypic variation in a population is unclear. Previous studies have found evidence that environmental exposures can account for more phenotypic variation in older individuals, potentially due to the accumulation of exposures as people age[1–3]. However, measuring the relevant environmental exposures throughout an individual's lifetime remains a significant challenge. On the other hand, gene-age interaction, which can arise from multiple sources[4], including proportional amplification of genetic variance[5] and different age profiles of genetic effects across genotypes[6], could drive age-dependent phenotypic variation. Jointly studying the age profiles of genetic and non-genetic components of complex traits can elucidate the architectures of age-dependent complex traits.

Recent work has shown age-varying prediction accuracy for polygenic risk scores (PRS), but whether this observation is driven by gene-age interactions or other mechanisms is unknown[6–10]. Significant challenges remain in studying age-dependent disease architectures. First, disease liability is not observed and performing gene-age interaction tests on binary disease observations may not be informative of interactions on the liability scale[11]. Second, although methods exist reduce bias in modeling time-to-event data due to censoring, these methods often have assumptions that are not met in empirical data[6].

Here, we estimated the age-dependent profiles of genetic and environment components of complex traits and disease liability. We show that the impact of environmental exposures accumulates with age for a wide range of complex traits and diseases, which causes a decrease in heritability with age. We show that exposure accumulation, under a liability-threshold model[12,13], explains decreasing prediction accuracy of PRS prediction accuracy for 9 diseases in UK Biobank. Finally, we show by simulation and analytical derivation that a liability-threshold model predicts that both genetic and environmental factors will have decreased prediction accuracy for older cases, which we validate in empirical data using 61 predictor-disease pairs.

## Results

### *Overview of methods*

We study age-dependent trait architecture under an exposure accumulation (EA) model where liability is linear in genetic, environment, and age quintiles[14], with environmental variance accumulating with age.

The EA model can be formalized under the following:

$$L = G + E + age + \sum_{a=0}^{age} E_a, \quad (1)$$

where $L$ denotes liability, $G$ is the genetic component of liability, $E$ is the environmental component, $age$ is the quintile of age, and $E_a$ is the contribution of age to liability. $\sum_{a=0}^{age} E_a$ is the term arising from the EA model. We study architecture across age quintiles to relax statistical assumptions on the relationship between age and $L$. We evaluate evidence for 4 additional models that can cause an age-dependent architecture: proportional amplification, age-dependent G, exposure accumulation with decay, and non-proportional amplification (**Supplementary Table 1**).

We evaluate evidence of the EA model by estimating age-varying heritability ($\Delta h^2 = h^2_{older} - h^2_{younger}$), which is negative under the EA model. We estimate $\Delta h^2$ by estimating the heritability for each age quintile with BOLT-REML[15] and fitting a model where heritability changes linearly with age. We estimate heritability with age and sex as covariates to account for within age quintile variation due to age and sex. We test for non-zero $\Delta h^2$ with a likelihood ratio test between the model with age term and a null model where heritability is constant with age. We also evaluate the environmental correlation between age quintiles ($\rho(E_{younger}, E_{older})$), which is expected to be less than 1.0 under the EA model. We use phenotypic correlations ($\rho_l$), which is an upper bound of $\rho(E_{younger}, E_{older})$ under the EA model, to evaluate evidence of $\rho(E_{younger}, E_{older}) < 1$. We account for error-in-measurement of $\rho_l$ by computing the denoised phenotypic correlation ($\tilde{\rho}_l$) for individuals with three phenotype measurements at different age points. We estimated $\tilde{\rho}_l$ as the ratio of first-third visit $\rho_l$ and first-second visit $\rho_l$. The ratio cancels the impact of measurement error. We divide $log(\tilde{\rho}_l)$ by the average age difference between the second and third visits to estimate $\tilde{\rho}_l$ between individuals with one year of age difference (see **Methods**).

We estimated $\Delta h^2$ and $\tilde{\rho}_l$ for 16 quantitative traits across 180K individuals in the UK Biobank (**Supplementary Table 2**); the sample size reflects a restriction to individuals of British ancestry with diagnostic records from both primary care data and hospital inpatient data, which we imposed to mitigate the impact of missing diagnoses. The quantitative traits in this study were selected following ref. [16] (**Methods**). We use *cross-sectional design* to estimate $\Delta h^2$, where we compare individuals with different ages at the baseline visit. We evaluate the impact of recruitment bias that can arise from a cross-sectional design in a secondary analysis. We use *longitudinal measurement* to estimate $\tilde{\rho}_l$, where we compare multiple measurements at different age points of the same individual.

We estimate $\Delta h^2$ for the estimated disease liability of 9 diseases (**Supplementary Table 3**). We use 36 quantitative traits (**Supplementary Table 4**) to estimate disease liability at baseline age by predicting binary disease phenotypes from the quantitative traits. We do not estimate $\widetilde{\rho_l}$ for estimated disease liability because we do not have longitudinal measurements for all quantitative traits that were used to estimate disease liability.

We study the impact of age-dependent disease architecture on disease prediction using simulations. We extract age-at-diagnosis for each patient using the first diagnosis in either primary care data or hospital inpatient data to evaluate age-dependent disease prediction accuracy in empirical data. We construct PRS and non-genetic predictors using five-fold cross validation without age stratification. We have publicly released summary data of age-specific heritability, phenotypic correlation, and genetic correlation estimates are publicly available publicly available (see **Data Availability**). We have also released software to compute age-dependent heritability and prediction accuracy (and **Code Availability**).

*Age-dependent genetic architecture of quantitative traits*

We estimated $\Delta h^2$ w.r.t. age of 16 quantitative traits (**Supplementary Table 2**) to evaluate the evidence supporting the EA model. First, we found 5 of the 16 traits have significantly decreasing heritability with age ($p$<0.05/9, FEV1/FVC, LDL cholesterol, total cholesterol, diastolic blood pressure, and systolic blood pressure) with an average 10-year decrease of 17.8% in $h^2$ (**Figure 1A**, **Supplementary Table 5, Supplementary Figure 1**). Second, we computed genetic and environmental variance of each age quintile using heritability and phenotypic variance estimates (**Figure 1B, Supplementary Figure 2, Supplementary Figure 3, Supplementary Table 6, Supplementary Table 7**). We estimated phenotypic variance for males and females separately to account for sex differences[5,11], which provides estimates for 32 trait-sex pairs. We found $\Delta var(E)$ is significantly larger than $\Delta var(G)$ for 10 trait-sex pairs (red dots in **Figure 1B**), supporting the EA model. We found $var(E)$ and $var(G)$ vary proportionally with age in traits that do not have significant age-varying heritability (grey dots in **Figure 1B**), supporting a proportional amplification model that has been observed in GxE studies[5,11]. The EA model will move traits above the diagonal line in **Figure 1B**, while the proportional amplification model will move traits along the diagonal line. Taken together, we conclude that FEV1/FVC, LDL cholesterol, total cholesterol, systolic blood pressure, and diastolic blood pressure show evidence for the EA model.

Next, we estimated genetic and phenotypic correlations between age groups to evaluate additional evidence in support of the EA model. $\rho(G_{younger}, G_{older})$ is expected to equal 1.0 and $\rho(E_{younger}, E_{older})$ is expected to be less than 1.0 under the EA model. We found the genetic correlation, $\rho(G_{younger}, G_{older})$, was close to 1.0 for most traits, with exceptions of HbA1c ($p$=1e-4) and glucose ($p$=3e-4) (**Figure 2A, Supplementary Table 8**). A genetic correlation less

171 than 1.0 supports a model where genetic components have different age profiles
172 (**Supplementary Table 1**), which is consistent with our previous work that showed that genetic
173 variants have distinct age profiles in conveying common disease risk[4]. We computed the
174 denoised phenotypic correlation $\widetilde{\rho_l}$ to evaluate the environmental correlation across age
175 ($\rho(E_{younger}, E_{older})$) (**Methods**). For the 11 traits that have longitudinal data for 3 visits, 7 have
176 $\widetilde{\rho_l}$ significantly less than 1.0 (**Figure 2B, Supplementary Table 9**), supporting the EA model. $\widetilde{\rho_l}$
177 has high correlation with $\Delta h^2$ w.r.t. age ($\rho$=0.61 across 11 traits, **Figure 2C**), which is consistent
178 with the EA model. Specifically, the 3 traits (Systolic BP, Diastolic BP, and FEV1/FVC) that
179 show evidence supporting the EA model in **Figure 1** have the largest difference between $\widetilde{\rho_l}$ and
180 1. Our analysis of the genetic and environmental correlation between young and old individuals
181 provides orthogonal evidence in support of EA for systolic blood pressure, diastolic blood
182 pressure, and FEV1/FVC.
183
184 We performed six secondary analyses to test the robustness of our results. First, we tested
185 whether age-dependent recruitment bias of individuals with different health status could explain
186 decreasing heritability estimates. We created two age-matched samples (mean age: 55.9 vs. 56.6)
187 with high and low numbers of prevalent disease diagnoses (mean diagnoses: 6.8 vs 34.6) and
188 estimated the heritability of 16 quantitative traits in the two groups. We observed minimal
189 heritability differences between the two groups, which suggests that biased recruitment of
190 healthier older individuals in the UK Biobank[17] does not contribute to the decreasing heritability
191 estimates with age (**Supplementary Figure 4**). Second, we evaluated the amplification models[5]
192 by estimating changes in genetic variance w.r.t. age for 32 trait-sex pairs (**Supplementary**
193 **Figure 2, Supplementary Table 7**). We found 25 traits with age-varying genetic variance, align
194 with previous GxSex and GxE studies that report amplifications[5,11]. Third, we tested an
195 alternative model where exposure accumulation decays with age by identifying traits with $\Delta h^2$
196 close to zero but a denoised phenotypic correlation less than 1.0 across age bins. We found
197 evidence support this model for BMI and body WHR among the traits that we can
198 estimate $\widetilde{\rho_l}$ (i.e. those with multiple measurements) (**Supplementary Table 9**). Fourth, we tested
199 a non-proportional amplification model by identifying traits with both genetic correlation and
200 denoised phenotypic correlation close to 1 but age-varying heritability (**Supplementary Table**
201 **9**). We found no such traits among the traits that we can estimate $\widetilde{\rho_l}$. Fifth, we estimated genetic
202 correlation between the top and bottom age quantiles (median age difference = 22.0) without
203 meta-analyzing across age bins and found smaller $\rho\left(G_{young}, G_{old}\right)$ compared with the meta-
204 analyzed results with larger standard errors (**Supplementary Figure 5**). Sixth, we tested whether
205 using age in years instead of age quintiles can explain more variance in quantitative traits. We
206 found age quintiles explains similar $R^2$ as age in years (**Supplementary Figure 6A**), suggesting
207 age quintiles capture the majority of age-related variation while allowing us to use standard
208 methods to compute heritability and genetic correlations. We found adding an additional age$^2$
209 covariate to the age in years explains more quantitative trait variance (mean $R^2$= 0.024 across 16
210 traits) than the model with only linear age in years (mean $R^2$= 0.022 across 16 traits).

Taken together, these results support exposure accumulation as the primary driver of age-dependent architecture in quantitative traits.

*Age-dependent genetic architecture of estimated disease liability*

We next investigated whether the liability of 9 diseases (**Supplementary Table 3**) is consistent with the EA model. We estimated disease liability by applying LASSO[18] to 36 quantitative traits (**Supplementary Table 4**) measured at baseline and predicting disease status in the entire population (i.e. across all age bins). We use the value predicted by LASSO as estimated liability to approximate disease liability (**Methods**).

We hypothesized that estimated liability is highly correlated with the true disease liability at the baseline age, thus can be used to study the age-dependent architecture of disease liability. We confirmed that the estimated liability achieves high accuracy across nine diseases, with an average liability-scale correlation of 0.603 (s.d. = 0.206 across nine diseases, **Supplementary Table 10**) with case-control status. We applied the same methods as in the quantitative trait analysis to the estimated liabilities to estimate $\Delta h^2$ and $\rho(G_{younger}, G_{older})$. We validated that $h^2$ of estimated liability is similar to the liability-scale $h^2$of the respective diseases (**Supplementary Figure 7**). We found the heritability of the predicted liability significantly decreases for 5 of the 9 predicted disease liabilities (**Figure 3A**, **Supplementary Table 11**, mean $\Delta h^2$ = -18.1%±2.7% for every 10 years of age), which supports the EA model of disease liability. We observe genetic correlations across age bins ($\rho(G_{younger}, G_{older})$) are significantly different from 1 for coronary artery disease (CAD) and T2D (**Figure 3B**, **Supplementary Table 12**), which aligns with previous observations of heterogeneous age profiles in CAD[6] and subtypes of T2D[19–21]. Taken together, we determined the estimated liabilities of asthma, hypertension, high lipid, COPD, and high cholesterol follow the EA model.

*Simulations implicate factors contributing to age-dependent disease PRS accuracy*

We hypothesized that the following 3 factors could contribute to age-dependent disease PRS accuracy: (i) age-dependent liability under a liability threshold model[6,12]; (ii) prediction of incident cases (instead of prevalent cases)[22]; and (iii) exposure accumulation under our new EA model. We refer to these three hypotheses as factors (i) - (iii)

Second, the intuition of factor (ii) is that excluding prevalent cases at baseline will create negative correlation between the predictable component and non-predictable component of disease liability, which reduces prediction accuracy (see **Supplementary Note)**. We note that there are two ways to define cases and controls at a specific age point (**Figure 4A**). First, *prevalent case association* defines cases as individuals who are diagnosed with the disease

before the age point and controls as individuals who have not had a diagnosis of the disease at least until the age point; prevalent case association is a common choice in constructing genetic risk scores[16,23–26]. Second, *incident case prediction* defines cases as individuals who are diagnosed for the first time within a time window after the age point and controls as individuals who are not diagnosed at least until the end of the time window; incident case prediction is a common choice in clinical risk prediction[27,28]. We report results from both case-control definitions but focus on the latter because incident case prediction has more clinical utility.

Third, the intuition of factor (iii) is that PRS prediction accuracy decreases with age when heritability decreases with age (see **Supplementary Note**). Under the proposed EA model, heritability will decrease due to exposure accumulation, causing decreasing PRS accuracy with age.

We performed simulations to evaluate the contribution of factors (i)-(iii) to age-dependent prediction accuracy (**Figure 4B-C and Supplementary Figure 8A**). Our simulation framework modeled liability increasing linearly with age. The age of onset is determined when the liability of an individual reaches a fixed threshold. We evaluate prediction accuracy using the observed $R^2$ with 1:1 case-control ratio ($R^2{}_{obs}$), area under the ROC Curve (AUC), liability-scale $R^2$ ($R^2{}_{liab}$), and log odds ratio ($logOR$). Further details are provided in the **Methods** section. $R^2{}_{obs}$, AUC, and $logOR$ all decrease with age for both prevalent case association and incident case prediction, while $R^2{}_{liab}$ decreases with age for incident case prediction but not for prevalent case association, supporting the contribution of factor (i) to age-dependent prediction accuracy. Incident case prediction accuracy decreases more than the prevalent case association accuracy, supporting the contribution of factor (ii). We found that prediction accuracy under the EA model decreases more than under linear liability-threshold model, supporting the contribution of factor (iii). We estimated $R^2{}_{liab}$ under a normal distribution assumption[29] in empirical data analyses. We evaluated the accuracy of this approximation using simulations and found it was accurate (**Supplementary Figure 8B**).

We show in the **Supplementary Note** that the contribution of factor (i) to age-dependent prediction accuracy depends on the $R^2{}_{liab}$ of the predictor to the underlying $L$. To validate this hypothesis, we simulated predictors with varying $R^2{}_{liab}$ to the underlying $L$, representing a perfect predictor ($R^2{}_{liab} = 99\%$), a strong predictor ($R^2{}_{liab} = 50\%$), a clinical score ($R^2{}_{liab} = 20\%$), and a PRS ($R^2{}_{liab} = 5\%$). Prediction accuracy decreased more steeply with age for weaker predictors, while the difference is negligible between the clinical risk score and PRS (**Figure 4C, Supplementary Figure 8A**).

Our analyses show that prediction accuracy decreases with age under the liability-threshold model (factor (i)) using both analytical results and simulations, while the decreasing slope is

290 steeper when predicting incident cases (factor (ii)), predicting phenotypes that follow the EA
291 model (factor (iii)), and the predictors have low overall $R^2{}_{liab}$.
292
293 *Age-dependent prediction accuracy of PRS is consistent with liability-threshold model and*
294 *prediction of incident diseases*
295
296 We next evaluated whether the age-dependent disease PRS accuracy is consistent with factor (i)
297 and (ii) in empirical data. We computed PRS using BOLT-REML with fivefold cross validation
298 for 9 diseases without age stratification. We divided the cases into age quintiles based on the
299 earliest diagnosis age of each patient and created sample sets with a 1:1 case-control ratio to
300 ensure each age bin has the same sample size and case-control ratio. We computed 4 metrics to
301 measure incident case prediction accuracy: log hazard ratio ($logHR$), log odds ratio ($logOR$), the
302 effect size of PRS in a linear regression on case-control status, and $R^2{}_{obs}$. We also compute
303 $R^2{}_{obs}$ for prevalent case association using the median age of each age quintile as the age index.
304 We report the change of the 4 metrics ($logHR$, $logOR$, linear regression effects, and $R^2{}_{obs}$) by
305 fitting a model where the metric changes linearly with age. Briefly, we test for non-zero changes
306 in each metric with a likelihood ratio test between the model with age term and a null model
307 where the metric is constant with age (**Methods**). We report the 10-year change of each metric
308 relative to the estimates in the entire population.
309
310 We found PRS accuracy decreases with age across all diseases we studied, consistent with factor
311 (i). We started by evaluating the $logHR$ of each PRS to account for censoring. We found that the
312 $logHR$ of each PRS significantly decreased with age for 6 of the 9 diseases ($p<0.05/9$, **Figure**
313 **5A**, **Supplementary Table 10**), which aligns with a previous study that used lead GWAS
314 associations and observed decreasing effect sizes with age[6]. On average, the $logHR$ of each PRS
315 in the oldest quintile was 51%±5% of the youngest quintile. For example, the $logHR$ of the PRS
316 for predicting hypertension was 0.39±0.01 for the youngest quintile of cases (median($age_{onset}$) =
317 47.1) and 0.16±0.01 for the oldest quintile of cases (median($age_{onset}$) = 72.2). The $logHR$ of the
318 PRS for predicting T2D is 0.32±0.02 for the youngest quintile of cases (median($age_{onset}$) = 47.7)
319 and 0.22±0.02 for the oldest quintile of cases (median($age_{onset}$) = 72.8). These observations were
320 consistent across different prediction accuracy metrics of PRS ($R^2{}_{obs}$, $logOR$, and linear
321 regression effects, **Supplementary Figure 9, Supplementary Figure 10,** and **Supplementary**
322 **Table 13**). We focus on $R^2{}_{obs}$ in later sections because it is on the same scale as $h^2$. We found
323 the average 10-year decrease for incident case prediction accuracy ($R^2{}_{obs}$) was greater than the
324 10-year decrease in accuracy for prevalent case association (**Figure 5B, Supplementary Table**
325 **14**), consistent with simulations of factor (ii) in **Figure 4B** (orange curves versus purple curves,
326 respectively). For example, $R^2{}_{obs}$ for PRS predicting hypertension decreases 50.0%±1.7% every
327 10 years of age in incident case prediction and 26.8%±3.3% for prevalent case association. $R^2{}_{obs}$
328 for PRS predicting T2D decreases 25.1%±4.2% every 10 years of age in incident case prediction

329 and no longer decreases for prevalent case association (slope= -2.9%±4.2% per 10 years). Our
330 analyses indicate that PRS prediction accuracy generally decreases significantly with age.
331
332 *Exposure accumulation drives age-dependent disease PRS accuracy*
333
334 We next evaluated whether the liability-threshold model (factor (i)), for incident case prediction
335 (factor (ii)), and the EA model (factor (iii)) can explain age-dependent disease PRS accuracy in
336 empirical data. We focus on incident case prediction so that all reported age-dependent
337 prediction accuracy contains the contribution of factor (ii). We create quantitative risk scores
338 (QRS) using quantitative traits to predict diseases. Because quantitative traits contain both
339 genetic and non-genetic information, QRS does not contain the impact of the EA model.
340 Therefore, age-dependent QRS prediction accuracy represents the contribution from factor (i)
341 and factor (ii). We used the same 36 baseline quantitative traits to create QRS for each disease
342 and add Gaussian noise so that QRS and PRS have the same $R^2{}_{liab}$ in the entire population
343 (**Methods**). We compute the proportion of observed decrease in PRS accuracy with age that can
344 be explained by each model to evaluate the contribution of the factor (i)-(iii) (**Methods**).
345
346 We found the liability-threshold model (factor (i)) and EA models (factor (iii)) together explain
347 the observed age-dependent prediction accuracy of PRS for incident case prediction (factor (ii)).
348 First, under the liability-threshold model for incident case prediction (factor (i)+(ii)), predictors
349 using genetic data and predictors using non-genetic data (*i.e.* measurements of quantitative traits
350 and environmental exposures) are expected to have the same age-dependent prediction accuracy,
351 if they explain the same amount of variance in the liability ( $R^2{}_{liab}$) (**Figure 4C**). We used the
352 change of QRS $R^2{}_{obs}$ with age to estimate the contribution from factor (i) + (ii) (**Methods**). We
353 found 65.1% of the change of PRS $R^2{}_{obs}$ with age are explained by these two factors (**Figure**
354 **6A**, **Supplementary Table 15** and **Supplementary Table 16**). The change of QRS $R^2{}_{obs}$ with
355 age is significantly different from the change of PRS $R^2{}_{obs}$ with age for hypertension ($p$=6e-8),
356 asthma ($p$ = 9e-13), and CAD ($p$ = 7e-7). We determined that factor (i) + (ii) explains most, but
357 not all of the decreasing PRS prediction accuracy. Second, under the EA model (factor (iii)), the
358 heritability of disease liability decreases with age, which causes PRS (which only captures
359 genetic variance) accuracy to decrease with age. We found $\Delta h^2$ of disease liability is not large
360 enough to fully explain the decrease in PRS prediction accuracy (**Figure 6B**), suggesting the EA
361 model in disease liability (factor (ii)) is insufficient to explain all of decreasing PRS prediction
362 accuracy. Lastly, we combine the effect of EA model with liability-threshold model by
363 computing the product of QRS $R^2{}_{obs}$ and $h^2$ of predicted liability as the expected prediction
364 accuracy of PRS, which captures the contribution of factors (i), (ii), and (iii). We determined that
365 85.8% of the observed age profiles of PRS is explained by the factor (i)+(ii)+(iii) (**Figure 6C,**
366 **Supplementary Table 17**). A model with factors (i)+(ii)+(iii) explains significantly more
367 observed $\Delta R^2{}_{obs}(PRS)$ with age than the model without factor (iii) (P = 3e-11). We note that
368 CAD cannot be explained by our models ($p$ = 7e-7 for x and y difference in **Figure 6C**). CAD

369 also has $\rho(G_{younger}, G_{older})$ < 1.0 (**Figure 3B**), suggesting that age-dependent genetic
370 heterogeneity might cause the deviation from the liability-threshold + EA model. Taken together,
371 our results suggest that the liability-threshold (factor (i)), incident case prediction (factor(ii)), and
372 the EA model (factor (iii)) explain age-dependent prediction accuracy of PRS.
373
374 *Age-dependent accuracy of non-genetic predictors of disease*
375
376 We investigated whether the liability-threshold model applies to age-dependent accuracy of both
377 genetic and non-genetic predictors. We constructed risk predictors using the 36 quantitative
378 measurements as non-genetic variables to further validate the liability-threshold model as a
379 general model of disease. We note these 36 quantitative measurements contain both genetic and
380 non-genetic (but predominantly non-genetic), but we use "non-genetic" to distinguish from
381 analysis that focus on G. For each disease, we constructed 4 predictors with varying prediction
382 power: (1) "QRS" predictor which is a LASSO predictor that uses all 36 quantitative
383 measurements, (2) A "Top 3" predictor that is a joint linear predictor trained using the top three
384 variables ranked by the $R^2{}_{liab}$ with target disease, (3) "single (strong)" predictors which are the
385 quantitative measurements that rank 1st-3rd by the $R^2{}_{liab}$ with target disease, (4) "single
386 (weak)" predictors which are the quantitative measurements that rank 4th and 5th. We trained all
387 predictors using five-fold cross validation and estimated their prediction accuracy of diseases in
388 each age quintile based on baseline UK Biobank visit. We then computed the change of $R^2{}_{obs}$
389 w.r.t. age for each predictor using the same method as the previous sections.
390
391 Predictors with greater prediction power in the entire population have less age-dependent
392 prediction accuracy (**Figure 7A**, **Supplementary Table 18,** and **Supplementary Table 19**), as
393 predicted by simulation results (**Figure 4C, Supplementary Note**). Relative change of $R^2_{obs}$
394 w.r.t. age is -47.6%±1.3% for "single (weak)" predictors (avg($R^2{}_{liab}$) = 9.8%) and -23.6%±1.1%
395 for "QRS" predictors (avg($R^2{}_{liab}$) = 32.3%), averaging across the 9 diseases. We further expand
396 to 10 predictors for each disease (QRS, "Top 3", each of the top 5 predictors, "top 2", PRS, PRS
397 + QRS, restricting to predictors with $R^2{}_{liab}$ > 5% for adequate power) and found strong positive
398 correlations between relative change of $R^2_{obs}$ w.r.t. age and $R^2{}_{liab}$ across disease-predictor pairs
399 (**Figure 7B**, slope = 0.64, $p$ =1.2e-5). We conclude that predictors that use environment variables
400 have decreasing accuracy with age and the effect is stronger for predictors with low overall
401 prediction accuracy, suggesting combining genetic and non-genetic predictors can increase
402 prediction power in older populations.
403
404 **Discussion**
405
406 We systematically quantified age-profiles of genetic and environmental variation to understand
407 how complex disease architecture and polygenic risk score (PRS) accuracy vary with age. We
408 found exposure accumulation is common in quantitative traits and complex diseases. We found

409 the age-dependent PRS accuracy can be explained by a combination of a liability-threshold
410 model and the exposure accumulation model. Our analysis was enabled by a novel approach to
411 study estimated disease liability. Finally, we found both genetic and non-genetic predictors
412 decrease in prediction accuracy with age in empirical data and simulations.
413
414 Our study represents an advance over previous studies that investigate age-dependent effects.
415 First, we identify changes in genetic and environmental variance across 16 complex quantitative
416 traits and 9 diseases. While previous work has studied age-dependent effects, they have focused
417 on single variant effects[3] or have studied a limited set of phenotypes[2]. Second, to the best of our
418 knowledge, our work is the first to systematically quantify how environmental variance in
419 complex traits accumulates with age using longitudinal data. Third, our work is the first to
420 provide a quantitative explanation for the decrease in prediction accuracy for genetic and non-
421 genetic predictors. While other studies have observed this phenomenon, they have not been able
422 to explain it[6,9]. We note, however, that CAD was an exception to the EA model and, while we
423 speculate that this may be due to age-dependent genetic heterogeneity, this is something of a
424 mystery to be resolved in subsequent studies.
425
426 Our study has several implications. First, the EA model suggest GWAS study can increase power
427 by accounting for age-dependent heritability. In particular, we hypothesize that mixed effect
428 model can increase power by modelling age-specific random effects. Second, the EA model
429 suggests when incorporating PRS in clinical risk models, PRS should be weighted differently
430 with age to increase power. Third, our results suggest all predictors (*i.e.* genetic or non-genetic)
431 of disease have age-dependent accuracy, while predictors with stronger prediction power in the
432 entire population will have more robust prediction accuracy at older age. This motivates
433 constructing predictors that combine multiple data modalities to achieve maximal prediction
434 accuracy across age. Fourth, the observed exposure accumulation suggests genetic effects
435 contribute more to phenotypic variation at early ages, suggesting ascertaining environment
436 variables is critical to accurately predict phenotypes at later ages. Lastly, our results suggest
437 recruiting younger cases can improve power for discovering both genetic and environment
438 biomarkers for diseases. All of these implications motivate directions for future research.
439
440 Our study has several limitations. First, our modeling is not able to detect uniformly decreasing
441 genetic variance which can cause decreasing $h^2$ with age and $\rho\left(G_{young}, G_{old}\right)$ close to 1.
442 However, we believe this is not likely because we found evidence of accumulating
443 environmental variance in the $\widetilde{\rho_l}$ analyses which explains the decreasing $h^2$ (**Figure 2B**).
444 Second, the age range of our study sample is restricted, with baseline visit age between 40 to 79
445 years of age. Future analyses with a wider age distribution may discover stronger effects. Third,
446 we use a cross-sectional analysis, which is susceptible to recruitment bias. We tested whether
447 age-dependent recruitment bias of healthier individuals could explain decreasing heritability
448 estimates with age and observed minimal heritability differences between the two groups,

449 suggesting a minimal effect of biased recruitment. Fourth, our analyses uses only individuals of
450 European individuals. While it is of interest to perform the analysis on individuals of other
451 ancestries, currently available data of other genetic ancestry might not have sufficient power for
452 age-stratified analyses. Fifth, our analyses use the UK Biobank data set. It is of interest to
453 replicate the analysis in other cohorts such as All of Us[30]. Sixth, majority of our findings rely on
454 the validity of liability-threshold model, although liability-threshold model has been widely
455 accepted and used[12,13,31–37]. Despite these limitations, our work identifies exposure accumulation
456 as a key aspect age-dependent disease architecture, and our results explain the decreasing genetic
457 and non-genetic disease prediction accuracy with age.
458
459 **Methods**
460 Study population in UK Biobank
461 We used the disease diagnosis information from 180,026 individuals whose primary care data
462 (General Practice, GP) was available and of British Isle ancestry. The primary care data are
463 coded in readv2 which were mapped to SNOMED CT
464 (https://digital.nhs.uk/services/terminology-and-classifications/read-codes), SNOMED CT codes
465 were then mapped to ICD-10 codes (using NHS digital's code mapping https://digital.nhs.uk/),
466 which were combined with ICD-10 codes from Hospital Episode Statistics (HES) and mapped to
467 PheCodes [38]. We extracted the earliest diagnosis age using the earliest records from the
468 combined primary care and HES data, which are used as age labels in our analysis. The choice of
469 focusing on individuals with both primary data and HES data was to get the best age at diagnosis
470 for each disease. When combining GP and HES data to extract diagnosis, correlation between
471 binary and quantitative traits increased compared to only using HES data to extract disease
472 (**Supplementary Figure 11**), suggesting that combining GP and HES data capture diagnoses
473 more accurately as correlation decreases when variables are measured with noise.
474
475 Selection of quantitative and binary traits
476 We extracted 36 quantitative traits measured at baseline visit of UK Biobank based on ref. [16]. As
477 our goal is to study disease liability, we selected 9 diseases using phecodes that have
478 interpretable underlying quantitative traits from the 36 quantitative traits (e.g. Hb1Ac for T2D,
479 BMI for obesity, LDLc for hypercholesterolemia) and prevalence higher than 5%. Disease status
480 and age-at-diagnosis were extracted from GP+HES data described in the previous section.
481
482 We corrected blood biochemistry measurement for cholesterol, hypertensive, and diabetic
483 medications: cholesterol (+1 s.d.), LDL-c (+ 1 s.d.), and triglyceride levels (+0.5 s.d.) were
484 corrected for lipid lowering medication[39]; systolic blood pressure (+15 mm Hg) and diastolic
485 blood pressure (+ 10 mm Hg) were corrected for hypertension medication[40]; HbA1c level is
486 corrected for non-insulin diabetic medication (+ 10.929) and insulin drugs (+ 1.3 x 10.929)[41].
487 The respective changes for individuals who are taking the medication are in the parentheses after
488 each trait; s.d. is the standard deviation computed in the entire population. We verified that the

correlation between the corrected quantitative traits and corresponding binary traits is higher compared to the correlation between uncorrected quantitative traits and the same set of binary traits. $\rho(LDL, hypercholesterolemia)$: 0.188 post-medication correction vs -0.005 pre-correction. $\rho(Systolic\ blood\ pressure, hypertension)$: 0.519 post-medication correction vs 0.425 pre-correction. $\rho(HbA1c, T2D)$: 0.565 post-medication correction vs 0.524 pre-correction. Both corrected and uncorrected values were included in the final 36 quantitative traits.

In our analyses, smoking status is defined as an ordinal variable with the following values: never (0), previous (1), current (2) (**Supplementary Table 4**).

To select quantitative traits that are most relevant to the nine diseases, we further computed the correlation between the 36 quantitative traits and the nine disease traits and selected 16 of the 36 quantitative traits based on liability scale correlation [29]; the included quantitative traits are among the top 3 correlated traits for at least one of the nine binary diseases.

Heritability and genetic correlation

We estimated heritability using the REML function of BOLT-LMM[15] version 2.4.1. 180,026 individuals were divided into 5 age bins based on the baseline visit age. We performed REML analysis for each age bin using sex and baseline visit age as covariates; we use the same 536,492 HapMap3 SNPs that have moderate LD with each other which is the same as the PRS section, see below. We also computed the heritability of 16 quantitative traits and nine diseases in the entire population.

We estimated genetic correlation between age bins for the 16 quantitative traits using LDscore regression[42]. First, we used the mixed-effect model to compute summary statistics for each age bin using BOLT-LMM version 2.4.1. We used the same set of variants and individuals as the heritability estimation. Since BOLT-LMM is a mixed effect model that controls population stratification, we constrained the heritability intercept of LDSC to 1.0; we constrained covariance intercept to be 0 as there is no sample overlap between different age bins. We run cross-trait LDSC for each age-bin pair (10 pairs in total for 5 age bins) to estimate the genetic correlation and its standard error.

Testing for heritability changes across age bins

We tested the age-dependency of heritability using heritability estimates with standard error across the five age bins. We used likelihood ratio tests to identify non-constant heritability across age bins. We tested two nested models: $H_0: h^2 \sim N(0 \times age + b, s.e.(h^2))$ and $H_1: h^2 \sim N(a \times age + b, s.e.(h^2))$. We estimated $b$ under $H_0$ and $a, b$ under $H_1$ using maximum likelihood estimators across the five heritability estimates. For each age quintile, we use the median age to fit the model. The slope $a$ estimated under $H_1$ are the per year change of $h^2$. We use Monte-Carlo sampling of $h^2$ for each age quintile from the point estimates and s.e.

529 and estimate the parameters of each Monte-Carlo sample. The s.d. of $a$ estimates across Monte-
530 Carlo samples are the s.e. of the estimates. We note this approach relies on the fact that samples
531 are non overlapping across age quintiles.
532
533 In all analyses, we report the relative change of heritability compared to the heritability estimate
534 in the entire population.
535
536 Testing for genetic correlation less than 1 between age bins
537 The quantity of interest is the difference between genetic correlation and 1: $1 - \rho_g$. First, we
538 scale genetic correlation using $\rho_{normalised} = exp(10 \times ln\, \rho_{\Delta age} / \Delta age)$ to account for the
539 varying age differences of age-bin pairs; the age differences are computed using the median of
540 each age quintile. Scaling on the exponential scale aligns with a Wiener process of liability w.r.t
541 age (see below). For 5 age bins, we can estimate 10 genetic correlations between pairs of age
542 bins; these 10 genetic correlations have four degrees of freedom. Therefore, we use inverse-
543 variance weighting to meta-analyse four of the 10 scaled $\rho_g$ (we use $\rho_g$ of neighboring age-bin
544 pairs 1-2, 2-3, 3-4, and 4-5) to compute one 10-year $r_g$ for each trait; standard errors of per-10
545 year estimates are used for inverse-variance weights.
546
547 Computing phenotypic correlation across age points
548 Under the assumption that random variance accumulates linearly w.r.t age (as in a Wiener
549 process), the correlation between liability across age bins is a function on the multiplicity scale
550 $\rho_{\Delta age} = e^{-\beta\, \Delta age}$. Therefore, we computed the expected 10-year genetic and phenotypic
551 correlation by $exp(10 \times ln\, \rho_{\Delta age} / \Delta age)$ to make the correlations comparable for age bin pairs
552 with different age differences. We computed phenotype correlations across individuals who have
553 three visits to the UK Biobank. We remove the impact of measurement error by computing a
554 ratio of two correlations $\rho_{\Delta age} = \rho_l(t_1, t_3)/\rho_l\,(t_1, t_2)$, which cancels out the measurement
555 error, and we use the average age difference between second and third visit ($\Delta age = t_3 - t_2$) to
556 scale the estimate.
557
558 Testing the impact of selection in recruitment process
559 To test for whether the age-dependent heritability changes can be attributed to selected
560 recruitment of healthier individuals at older age groups[17], we computed the number of past
561 diagnosis records as an indicator of health status and created samples with matched age but
562 different past diagnosis. We used health records from both GP and HES, and divided the target
563 population into two groups, "healthy" versus "unhealthy", using the median number of prevalent
564 records (median=16). We then use inverse probability weights to sample 45,267 individuals from
565 both groups, resulting in two samples that have matched age distributions (mean: 55.9 vs. 56.6,
566 respectively) but different numbers of past diseases (mean: 6.8 vs 34.6, respectively); in
567 comparison, top and bottom age quintile has an average age of 67.5 and 45.3 and an average of

24.2 and 18.7 of past diseases, respectively. We compute heritability of the 16 quantitative for both groups using the same procedure as for age quintiles.

Simulation of linear liability-threshold model

We simulated genetic variance G and environment variance E from normal distribution with mean 0 and variance equal to $h^2$ and $(1-h^2)$ respectively. Liability $L$ equals the sum of $G$ and $E$ for each individual. To simulate liability increasing with age, we take 2% of the simulated individuals with the highest liability as the youngest disease group, top 2-4% of liability as the second age group and repeated the procedure to obtain five groups of individuals with increasing onset age. Each disease threshold corresponds to the disease cut off at an age point. Prevalent cases are defined as the individuals that have liabilities higher than the threshold at a specific age point; incident cases are defined as individuals in the bin near the liability threshold at a specific age point (i.e. all individuals with liability near the corresponding liability); controls are defined as all individuals excluding prevalent cases (i.e. all individuals have liability lower than the corresponding liability).We simulated 100,000 individuals, with $h^2$=0.3 and using five age bins of 2% of individuals, which matches the disease prevalence and sample size in empirical analyses.

We note our simulation uses percentiles of individuals to represent cases with increasing age, which aligns with the empirical analysis (see below) where we divided cases into quintiles based on the diagnosis age. We note under the $\mathrm{PA_{LT}}$ model, the liability percentile of each individual does not change, which means this model is invariant to proportional amplification.

Simulation of EA liability-threshold model

We simulated initial liability L using the same procedure as the linear liability-threshold model. For each age bin, we sample cumulative environmental variance $E_a$ from normal distribution with 0 mean and 0.05 variance. We then cumulatively add these sampled environmental variances to the liability when moving to an older age quintile, which will result in 20% more accumulated variance in the fifth age quintile. In empirical data, we found ~20% decrease in heritability over 20 years for quantitative traits, which motivates our choice of simulation parameters.

We defined prevalent cases as individuals who have got disease at any prior age bin or the current age bin, and incident cases as those who get disease within the current age bin. For each age bin, we take the top 2% individuals who have not prevalent diseases to simulate that liability increases with age. We note this simulation procedure makes the number of cases equal to previous simulation of the linear model so the metrics in **Figure 4B** are comparable.

PRS construction

We obtained 536,492 HapMap3 SNPs using pruning parameters of window size=50 kb, step size=5 kb, and $r^2$ threshold=0.8. We then used BOLT-LMM to estimate PRS, using genetic sex and age at baseline as covariates, using 5-fold cross validation for the 9 diseases and 16 quantitative traits. PRS are scaled to have mean=0 and sd=1 in the entire population.

PRS incident case prediction and prevalent case association

For incident case prediction, we first divided cases into five age quintiles based on their first age-at-diagnosis; for each age quintile, we define case control status based using interval-censor procedure [8], where cases are defined as those who are diagnosed within the age interval and controls are defined as those who remained healthy till the end of age interval or lost to follow up within the age interval. We then down sampled the controls for each age bin to have the same number as the cases. As a result, for a given disease, each age interval has exactly the same number of cases and controls which ensures the odds ratio and hazard ratio are comparable across each age bin. When evaluating PRS using linear, logistic, and Cox regression models, we included genetic sex and top five genetic PCs as covariates. When evaluating PRS using observed-scale $R^2$, we first regress the outcome on genetic sex and top five genetic PCs, then compute Pearson's correlation between the residual and PRS; the square of Pearson's correlation is the observed-scale $R^2$ with 1:1 case control ratio.

For prevalent cases association, we used the five median case age of the respective age intervals as the baseline age points and considered all individuals who have been followed up past the baseline age point. The prevalence cases are defined as those who have at least one diagnosis record before the baseline age point. We sampled controls to be the same number as cases for each age point, so that for a given disease, the effect sizes of linear, logistic, and Cox regression across five baseline age points are comparable. The regression models and observed scale $R^2$ were computed using the same covariates as in incident case association.

Predicted disease liability

We use 36 continuous traits measured at baseline to predict the 9 binary disease traits; the resulting predictor is the *predicted disease liability*. Cases are defined as the union of both incident and prevalent disease. We use LASSO linear model (instead of logistic regression as we assume a liability-threshold model) to train the prediction models with 10-fold cross validation; the regularization coefficients are chosen by the cross-validation with both 1se (default setting of the R glmnet package[43,44]) and 2se (stronger regularization). We chose the predicted liability using the 2se model as it has a similar liability-scale correlation (average=0.60 across 9 diseases) with the binary trait to the 1se model (average=0.61 across 9 diseases), while being more explainable and less susceptible to overfitting. The models were fitted in the entire population without including age as a covariate so that the predicted disease liability represents the approximate disease liability at the age of baseline visit.

647 We used the same procedure as the quantitative traits analysis to estimate heritability and genetic
648 correlation for predicted liability. We used BOLT-REML to estimate heritability. We used
649 BOLT-LMM to compute GWAS summary statistics then use cross-trait LDSC to estimate
650 genetic correlation of each age-bin pair (10 pairs in total for 5 age bins) to estimate the genetic
651 correlation and its standard error. We did not constrain intercepts for cross-trait LDSC to avoid
652 potential bias population relatedness.
653
654 Incident case prediction with non-genetic predictors
655 For environment variable predictors, we divided incident cases into five age bins using the age at
656 first diagnosis visit. We note that in contrast to PRS prediction accuracy, we excluded
657 individuals who have prevalent cases before baseline visits as environment variables are
658 measured at baseline visit. The definition of incident cases and controls for each age quintile,
659 down sampling of control is the same as the PRS incident case prediction. We computed
660 observed-scale $R^2$ for each environment predictor by regression of sex and top five genetic PCs
661 from the outcomes and computed Pearson's correlation between the outcome residual and the
662 predictor; the square of Pearson's correlation is the observed-scale $R^2$ with 1:1 case control ratio.
663
664 Estimating age-dependent prediction accuracy under the linear liability-threshold model
665 We use rank normal transformation of predicted liability to create a *quantitative risk score*
666 *(QRS)*. QRS combines both G and E information which is different from PRS that only captures
667 G information; under the linear liability-threshold model, QRS and PRS has the sample age-
668 dependent prediction accuracy profile when their liability-scale $R^2$ are equivalent across the
669 population (**Figure 4C**). We performed the following 6 steps to use QRS to estimate expected
670 age-dependency of PRS under the liability-threshold model. (1) QRS are rank normal
671 transformed to have similar distributions as PRS. (2) QRS are adjusted for age at measurement
672 and genetic sex; PRS are adjusted for first 5 genetic PCs and genetic sex. Adjustment of
673 covariates ensures QRS and PRS are comparable under the linear liability-threshold model for
674 each age bin, as QRS will change with age while PRS does not change with age. (3) We
675 computed the liability-scale $R^2$ of QRS and PRS using all individuals with primary care
676 information (N = 215,162); cases are defined as the union of prevalent cases and incident cases.
677 We computed observed-scale $R^2$ for both QRS and PRS and transformed observed-scale $R^2$ to
678 liability-scale $R^2$ using equation 14 of ref [29] (4) We added gaussian noise (mean = 0) to QRS so
679 that the new QRS has the same liability-scale $R^2$ as PRS. (5) We computed QRS observed-scale
680 $R^2$ for each age bin using the same case control definition as incident case prediction of PRS. We
681 note a caveat that some QRS are measured after disease diagnosis. Our choice is based on two
682 reasons: first, under the linear liability-threshold model, QRS (age and sex adjusted) represents
683 the liability that remains constant across age; second, observed-scale $R^2$ for QRS and PRS are
684 computed using the same samples, making them comparable. We use QRS to predict incident
685 cases at baseline visit and found similar results, see below. (6) We performed Monte-Carlo
686 sampling to account for two sources of uncertainty: (i) the process to add noise to QRS and (ii)

control sampling within each age bin. We reported mean and standard deviation of 50 Monte-Carlo sampling estimates for each parameter.

We define the proportion of observed decrease in PRS accuracy explained by each model as:

$$\tau = \sum_{diseases} \max(\Delta R^2_{QRS}, \Delta R^2_{PRS}) / \sum_{diseases} \Delta R^2_{PRS}.$$

The max in the numerator reflects the constraint that the proportion of age-dependent PRS accuracy that can be explained is less than 100%. Therefore, diseases with more decrease in QRS accuracy than PRS accuracy are assigned same value in numerator and denominator. This choice is justified by empirical data as we do not observe significantly more decrease in QRS accuracy than PRS accuracy at a nominal *p*-value of 0.05, except for CAD.

Estimating age-dependent prediction accuracy under the EA + liability-threshold model

We multiplied the estimated slope from QRS with the slope of $h^2$ of predicted liability to estimate the expected decrease under the EA + liability-threshold model. Under the EA model PRS prediction accuracy is proportional to $h^2$ of disease liability at each age point. The EA + liability-threshold model estimates capture both the age-dependency from the linear liability-threshold model (captured by QRS) and the EA model (captured by the $\Delta h^2$ of predicted liability). The expected change is computed as 1 - (1 + $\Delta R^2{}_{QRS}$) ×(1 + $\Delta h^2{}_{predicted-liability}$). We use Monte-Carlo sampling of $\Delta R^2{}_{QRS}$ and $\Delta h^2{}_{predicted-liability}$ to estimate the standard error of the estimator; we omit the sampling covariance between $\Delta R^2{}_{QRS}$ and $\Delta h^2{}_{predicted-liability}$ as they were estimated from approximately independent information (*i.e.* $\Delta R^2{}_{QRS}$ were estimated using age quintiles of cases and baseline measurements, while $\Delta h^2{}_{predicted-liability}$ uses genetic information).

Non-genetic predictors

We analyzed 12 predictors including 11 environment predictors and PRS. We computed the relative change of observed-scale $R^2$ compared to the observed-scale $R^2$ estimate in the entire population. Observed-scale $R^2$ in the entire population were transformed to liability-scale $R^2$ using equation 14 of ref [29]. When computing the relationship between liability-scale $R^2$ (in the entire population) to the change of observed-scale $R^2$, we only include 70 of 108 predictors that have liability-scale $R^2$ larger than 5% to ensure there is sufficient power to detect age-dependency.

**Code Availability**

Code to replicate all analyses in the manuscript is available at https://github.com/Xilin-Jiang/GxAge

**Data Availability**

Summary data in the manuscript is available at https://github.com/Xilin-Jiang/GxAge

**Acknowledgements**

Use of the UK Biobank data was performed under application number 7439. This research was funded by Wellcome Trust early-career award 227566/Z/23/Z (X.J.) and the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R35GM16046 (A.D.). M.I. is supported by the Munz Chair of Cardiovascular Prediction and Prevention and the NIHR Cambridge Biomedical Research Centre (NIHR203312) as well as by the UK Economic and Social Research 878 Council (ES/T013192/1). J.D. is supported by a British Heart Foundation Personal Chair. This work was supported by core funding from the British Heart Foundation (RG/F/23/110103), NIHR Cambridge Biomedical Research Centre (NIHR203312), BHF Chair Award (CH/12/2/29428), Cambridge BHF Centre of Research Excellence (RE/24/130011) and by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and the Wellcome Trust. The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or the National Institutes of Health.

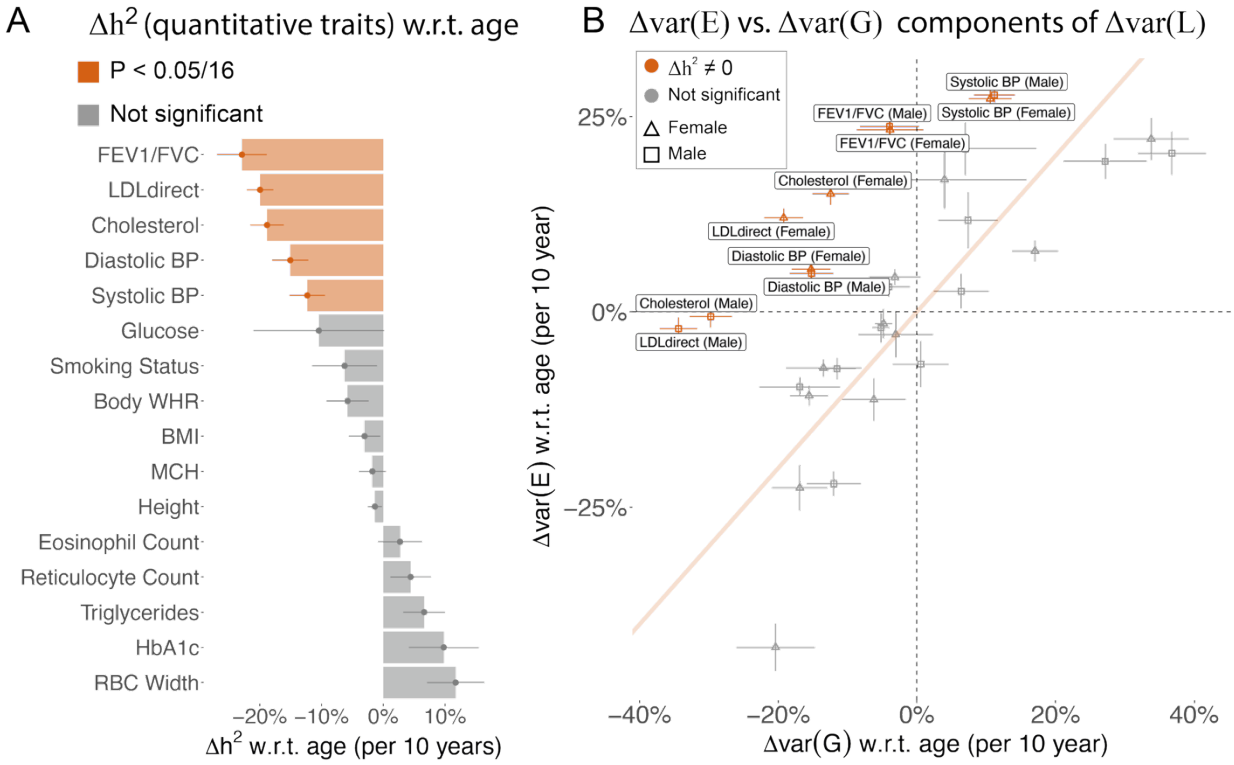



**Figure 1. Age-dependent variance of quantitative traits**. (A) Heritability changes per 10-years for 16 quantitative traits. Red bars show significant heritability changes across age; error bars show standard error via Monte-Carlo sampling. (B) 10-year change for genetic and environment variance for the 16 quantitative traits. Traits with significant age-dependent heritability are highlighted in red. Traits in grey show evidence of proportional amplification of genetic and environmental variance. Error bars are standard error via Monte-Carlo sampling. Numerical results are provided in **Supplementary Table 5 and Supplementary Table 7**. We note that no estimates are significantly outside the plausible range and that allowing estimates outside of the plausible range is necessary to ensure unbiasedness of estimates.

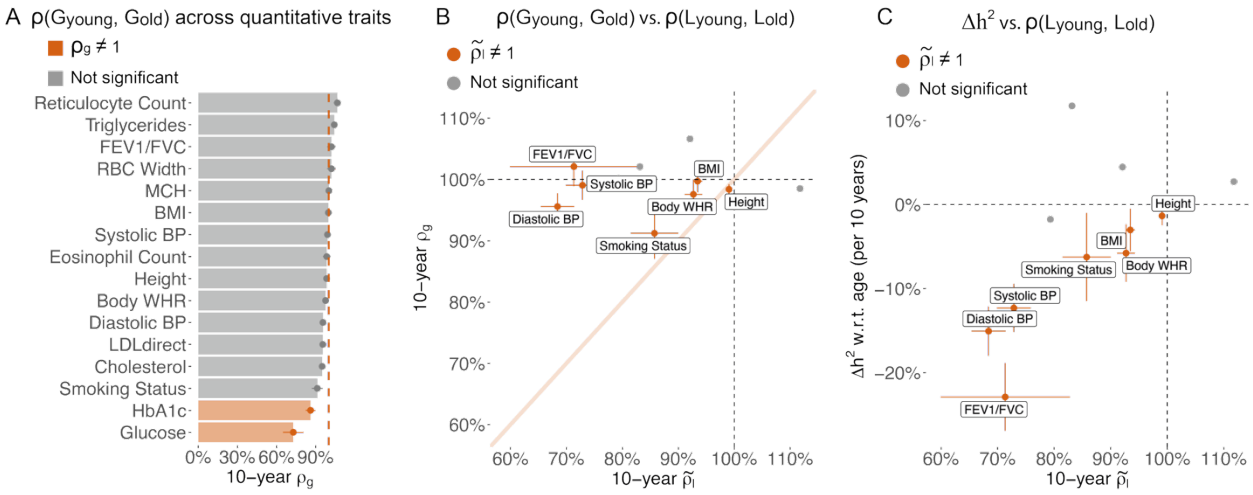



**Figure 2. Genetic liability is more correlated across young and older age groups than environmental liability.** (A) 10-year genetic correlation between younger and older age bins for 16 quantitative traits. We report the genetic correlation per 10-year age difference. (B) Scatter plot of 10-year genetic correlation versus denoised 10-year phenotypic correlation for the 11 quantitative traits that have three measurements per individual. We found no genetic correlation or phenotypic correlation that are significantly higher than 1 at nominal *p*-value = 0.05. (C) Scatter plot of heritability changes per 10-years versus denoised 10-year phenotypic correlation for the 11 quantitative traits that have three measurements per individual. Traits with denoised 10-year phenotypic correlation significantly different from 1 are shown in red and labeled in panel B and C; non-significant traits are shown in grey. Error bars for phenotypic correlation are standard error via bootstrapping over individuals; error bars for genetic correlation are standard error after meta-analysis across 4 age-bin pairs. Numerical results are provided in **Supplementary Table 8-9**. We note that no estimates are significantly outside the plausible range and that allowing estimates outside of the plausible range is necessary to ensure unbiasedness of estimates.

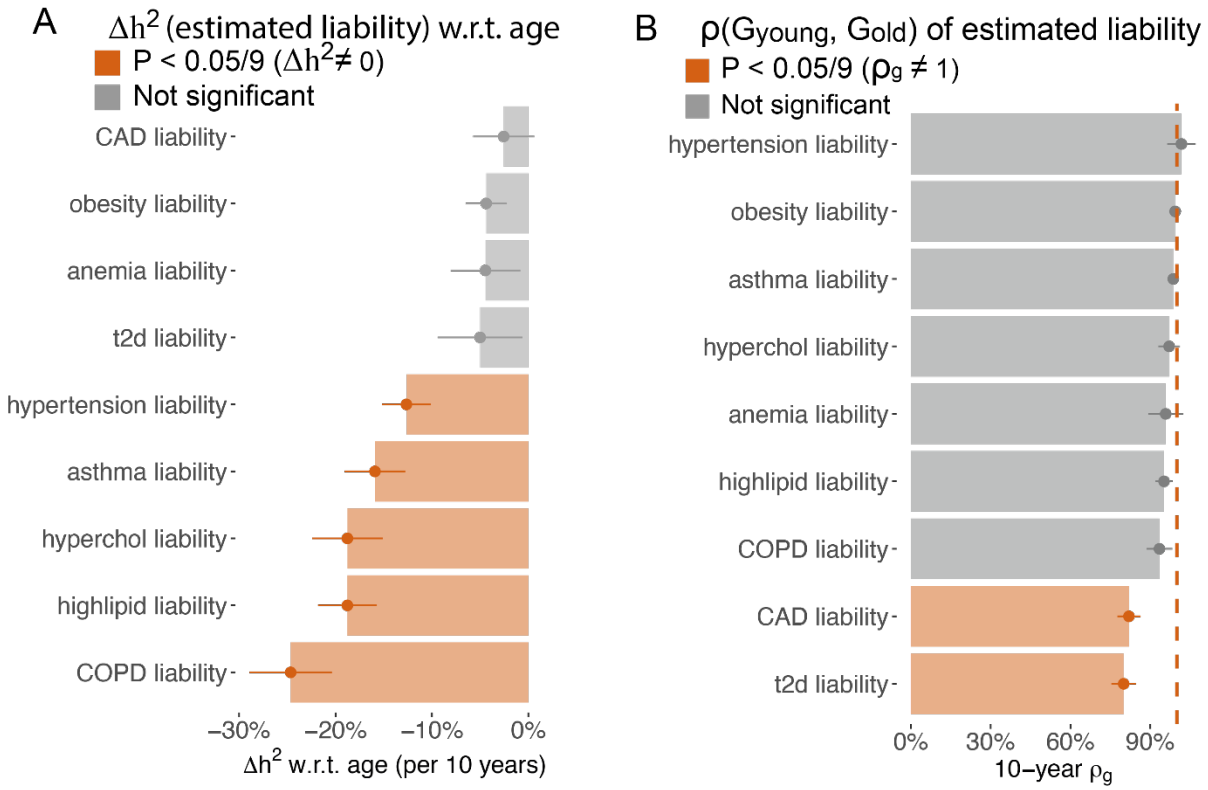
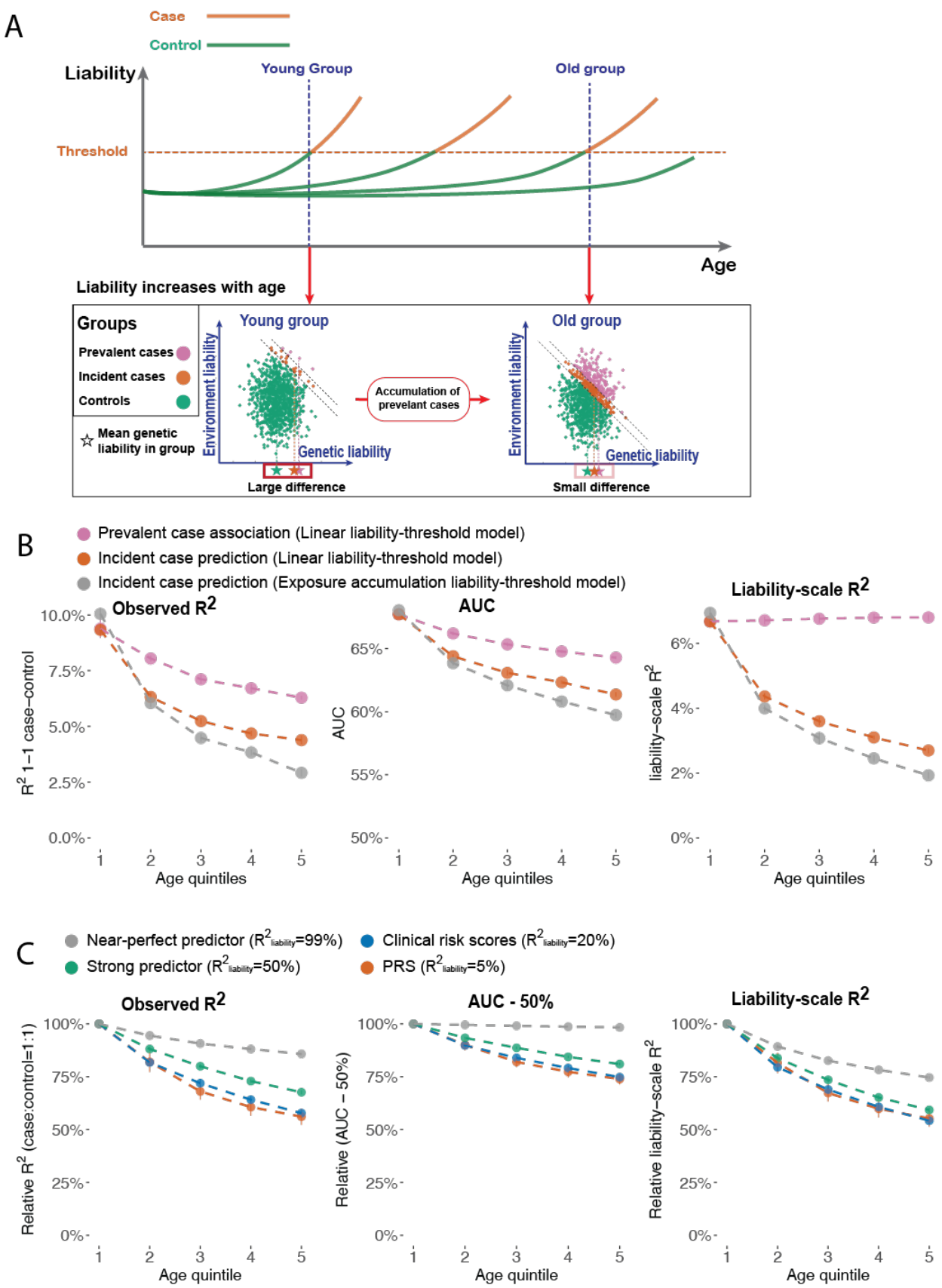



**Figure 3. Age-dependent genetic architecture of estimated liability supports the EA model.** (A) 10-year heritability change for estimated liability of 9 diseases; error bars show standard error via Monte-Carlo sampling. (B) 10-year genetic correlation for the estimated liability of 9 diseases. Error bars are standard error via meta-analysis of 10 age bin pairs. Numerical results are provided in **Supplementary Table 11-12**.


A
Case
Control
Liability
Young Group
Old group
Threshold
Age
Liability increases with age
Groups
Prevalent cases
Incident cases
Controls
Mean genetic liability in group
Young group
Environment liability
Genetic liability
Large difference
Accumulation of prevelant cases
Old group
Environment liability
Genetic liability
Small difference
B
Prevalent case association (Linear liability-threshold model)
Incident case prediction (Linear liability-threshold model)
Incident case prediction (Exposure accumulation liability-threshold model)
Observed R2
R2 1–1 case–control
10.0%
7.5%
5.0%
2.5%
0.0%
Age quintiles
AUC
AUC
65%
60%
55%
50%
Age quintiles
Liability-scale R2
liability–scale R2
6%
4%
2%
0%
Age quintiles
C
Near-perfect predictor (R2liability=99%)
Clinical risk scores (R2liability=20%)
Strong predictor (R2liability=50%)
PRS (R2liability=5%)
Observed R2
Relative R2 (case:control=1:1)
100%
75%
50%
25%
0%
Age quintile
AUC - 50%
Relative (AUC – 50%)
Age quintile
Liability-scale R2
Relative liability–scale R2
Age quintile

786

Figure 4. **Liability-threshold model and the EA model explains age-dependent disease prediction accuracy in simulations.** (A) Schematic figure explaining the decreasing disease prediction accuracy with age. In the top panel, each line represents the liability trajectory of an individual. The panel below shows the joint distribution of genetic and environmental liability at two age points. At later age points, cases are more similar to controls compared to earlier age points. (B) Prevalent case association accuracy (purple), incident case prediction accuracy (red), and Incident case prediction under the EA model (grey) across different age quintiles of cases. Prediction accuracy is measured by observed-scale $R^2$ in 1:1 case-control matched samples, AUC, and liability scale $R^2$ (C) Incident case prediction accuracy of predictors with different liability-scale $R^2$.

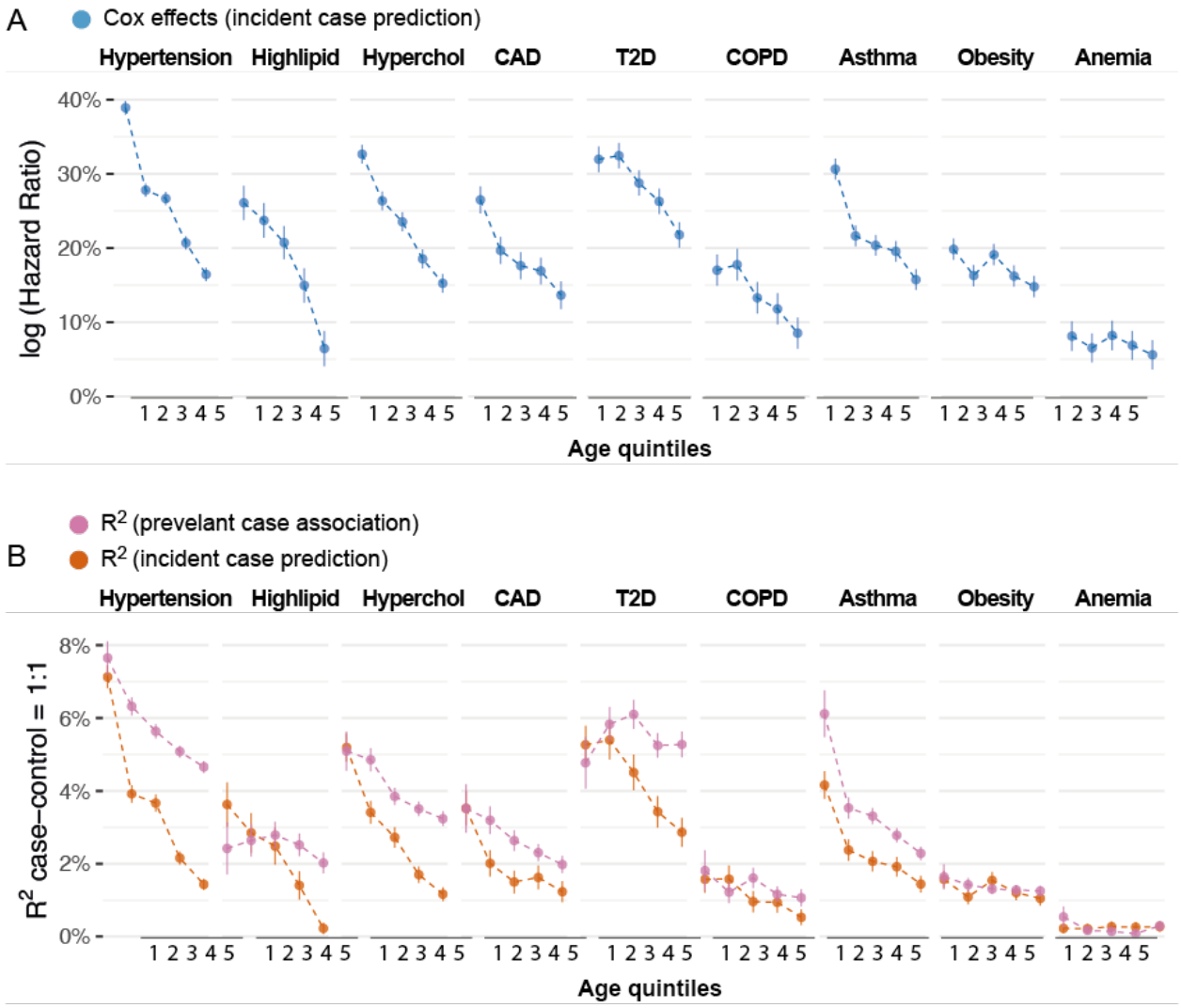



**Figure 5: Age-dependent incident case prediction accuracy in 9 diseases.** (A) Cox log hazard ratio of PRS on 9 diseases across 5 age quintiles. Each age bin has the same number of cases and 1:1 case-control ratio. Bars show standard error. (B) Age-dependent incident case prediction compared to prevalent case association accuracy in 9 diseases. Numerical results are provided in **Supplementary Table 13-14**.

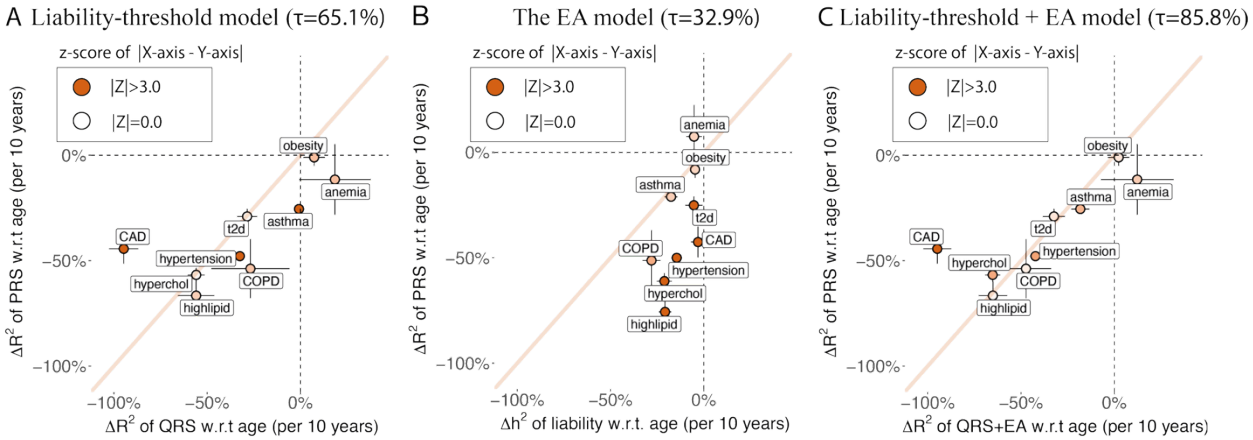



Figure 6: **Liability-threshold model and the EA model explain age-dependent PRS prediction accuracy.** (A) 10-year change in QRS prediction $R^2$ under the liability-threshold model versus 10-year change in PRS prediction $R^2$ for 9 disease traits. (B) 10-year heritability change in estimated disease liability versus 10-year change in PRS prediction $R^2$ for corresponding disease traits. (C) 10-year change in QRS prediction $R^2$ under the liability-threshold+EA model versus 10-year change in PRS prediction $R^2$ for 9 disease traits. Diagonal line has slope equals to 1. Bars show the se and color shows the z-score of difference in X and Y values of each dot. $\tau$ is the proportion of observed decrease in PRS accuracy explained by each model (**Methods**). Numerical results are provided in **Supplementary Table 15-17**. We note that no estimates are significantly outside the plausible range and that allowing estimates outside of the plausible range is necessary to ensure unbiasedness of estimates.

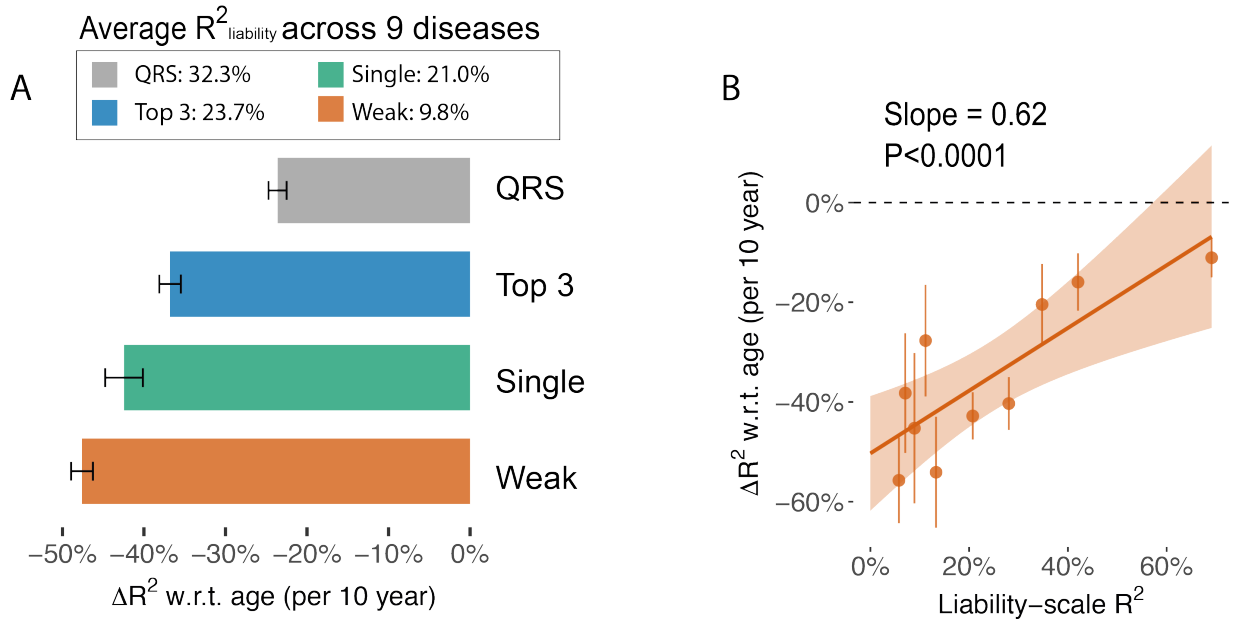



Figure 7. **Risk models with higher liability-scale $R^2$ are less attenuated when predicting late-onset diseases.** (A) Age-dependent prediction accuracy of predictors averaged across 9 diseases. X-axis shows the per-10-year change of prediction $R^2$ computed in 1-1 case control matching groups. QRS: predicted liability using 36 quantitative traits. Top 3: linear predictors use the top three quantitative traits that have the highest correlation with the diseases. Single (Strong): using a single quantitative trait to predict incident diseases, averaged across the top three traits as in the "Top 3" group; Single (Weak): the average between 4th and 5th predictor of each disease. (B) Scatter plot comparing the liability-scale $R^2$ (without age stratification) with the change of $R^2$ w.r.t. age across predictor-disease pairs. Predictor-disease pairs are grouped into 10 deciles based on liability-scale $R^2$ for visualization. The red line shows a regression line with confidence interval. Numerical results are provided in **Supplementary Table 18-19**.

**References:**


1. Simino, J. *et al.* Gene-age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. *Am. J. Hum. Genet.* **95**, 24–38 (2014).
2. Robinson, M. R. *et al.* Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat. Genet.* **49**, 1174–1181 (2017).
3. Winkler, T. W. *et al.* Genetic-by-age interaction analyses on complex traits in UK Biobank and their potential to identify effects on longitudinal trait change. *Genome Biol.* **25**, 300 (2024).
4. Miao, J. *et al.* PIGEON: a statistical framework for estimating gene-environment interaction for polygenic traits. *Nat. Hum. Behav.* 1–15 (2025).
5. Zhu, C. *et al.* Amplification is the primary mode of gene-by-sex interaction in complex human traits. *Cell Genomics* **3**, (2023).
6. Jiang, X., Holmes, C. & McVean, G. The impact of age on genetic risk for common diseases. *PLoS Genet.* **17**, e1009723 (2021).
7. Sakaue, S. *et al.* Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. *Nat. Med.* **26**, 542–548 (2020).
8. Jukarainen, S. *et al.* Genetic risk factors have a substantial impact on healthy life years. *Nat. Med.* **28**, 1893–1901 (2022).
9. Thompson, D. J. *et al.* A systematic evaluation of the performance and properties of the UK Biobank Polygenic Risk Score (PRS) Release. *PLoS One* **19**, e0307270 (2024).
10. Argentieri, M. A. *et al.* Integrating the environmental and genetic architectures of aging and mortality. *Nat. Med.* **31**, 1016–1025 (2025).
11. Durvasula, A. & Price, A. L. Distinct explanations underlie gene-environment interactions in the UK Biobank. *Am. J. Hum. Genet.* **112**, 644–658 (2025).
12. Zaitlen, N. *et al.* Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet.* **8**, e1003032 (2012).
13. Pedersen, E. M. *et al.* Accounting for age of onset and family history improves power in genome-wide association studies. *Am. J. Hum. Genet.* **109**, 417–432 (2022).
14. Aschard, H., Zaitlen, N., Lindström, S. & Kraft, P. Variation in predictive ability of common genetic variants by established strata: the example of breast cancer and age. *Epidemiology* **26**, 51–58 (2015).
15. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
16. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat. Genet.* **54**, 450–458 (2022).
17. Fry, A. *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
18. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the royal statistical society series b-methodological* **58**, 267–288 (1996).

19. Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Med.* **15**, e1002654 (2018).
20. Jiang, X. *et al.* Age-dependent topic modeling of comorbidities in UK Biobank identifies disease subtypes with differential genetic risk. *Nat. Genet.* **55**, 1854–1865 (2023).
21. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* 1–11 (2024).
22. Celentano, D. D., Szklo, M. & Farag, Y. *Gordis epidemiology*. (Elsevier - Health Sciences Division, 2024).
23. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
24. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
25. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).
26. Zheng, Z. *et al.* Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nat. Genet.* **56**, 767–777 (2024).
27. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur. Heart J.* **42**, 2439–2454 (2021).
28. Patel, A. P. *et al.* A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat. Med.* **29**, 1793–1803 (2023).
29. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).
30. Genomic data in the All of Us Research Program. *Nature* 1–7 (2024).
31. Dempster, E. R. & Lerner, I. M. Heritability of threshold characters. *Genetics* **35**, 212–236 (1950).
32. Wray, N. R. & Goddard, M. E. Multi-locus models of genetic risk of disease. *Genome Med.* **2**, 10 (2010).
33. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
34. Witte, J. S., Visscher, P. M. & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. *Nat. Rev. Genet.* **15**, 765–776 (2014).
35. Hujoel, M. L. A., Gazal, S., Loh, P.-R., Patterson, N. & Price, A. L. Liability threshold modeling of case-control status and family history of disease increases association power. *Nat. Genet.* **52**, 541–547 (2020).
36. Hujoel, M. L. A., Loh, P.-R., Neale, B. M. & Price, A. L. Incorporating family history of disease improves polygenic risk scores in diverse populations. *Cell Genom.* **2**, 100152 (2022).
37. Pedersen, E. M. *et al.* ADuLT: An efficient and robust time-to-event GWAS. *Nat. Commun.* **14**, 5553 (2023).
38. Wu, P. *et al.* Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform* **7**, e14325 (2019).

39. Kofink, D. *et al.* Statin Effects on Metabolic Profiles: Data From the PREVEND IT (Prevention of Renal and Vascular End-stage Disease Intervention Trial). *Circ. Cardiovasc. Genet.* **10**, (2017).
40. Wu, J. *et al.* A summary of the effects of antihypertensive medications on measured blood pressure. *Am. J. Hypertens.* **18**, 935–942 (2005).
41. Esposito, K., Chiodini, P., Bellastella, G., Maiorino, M. I. & Giugliano, D. Proportion of patients at HbA1c target <7% with eight classes of antidiabetic drugs in type 2 diabetes: systematic review of 218 randomized controlled trials with 78 945 patients. *Diabetes Obes. Metab.* **14**, 228–233 (2012).
42. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
43. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
44. Tay, J. K., Narasimhan, B. & Hastie, T. Elastic net regularization paths for all generalized linear models. *J. Stat. Softw.* **106**, 1–31 (2023).