

*Sources of Inequality at Birth: The Interplay Between
Genes and Parental Socioeconomic Status*

*Pietro Biroli, Nicolau Martin-Bassols, Andries T. Marees,
Hans van Kippersluis, Cornelius A. Rietveld, Pia Arce,
Kevin Thom, Stephanie von Hinke, Jeremy Vollen, Titus
Galama*

Paper No: 2026-006

**CESR-SCHAEFFER
WORKING PAPER SERIES**

The Working Papers in this series have not undergone peer review or been edited by USC. The series is intended to make results of CESR and Schaeffer Center research widely available, in preliminary form, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. CESR-Schaeffer working papers can be cited without permission of the author so long as the source is clearly referred to as a CESR-Schaeffer working paper.

Sources of Inequality at Birth: The Interplay Between Genes and Parental Socioeconomic Status^{*}

Pietro Biroli^{1,§} Nicolau Martin-Bassols^{1,§} Andries T. Marees²
Hans van Kippersluis^{3,4} Cornelius A. Rietveld^{3,4} Pia Arce⁵
Kevin Thom⁶ Stephanie von Hinke⁷ Jeremy Vollen⁸
Titus Galama^{2,3,9,†}

§ Shared first author. † Last author, Corresponding author.

Abstract

The start of a human’s life can be characterized by two lotteries: that of your genes (nature) and the family you were born into (nurture). These set in motion a trajectory, from birth onward, in health and human capital. Leveraging three longitudinal social-science data sets, we systematically analyze the relationship between an individual’s genotype, the socioeconomic status (SES) of the families they grew up in, and their realized traits in adulthood. We proxy an individual’s genetic predisposition by polygenic indexes (PGIs) and family SES by a latent factor of parental education and father’s (former) occupational status. We then investigate how PGIs, parental SES, and their interaction contribute to later-life outcomes across a range of forty-five socioeconomic, anthropometric, health, behavioral, and personality traits. We find strong genetic and socioeconomic associations with these phenotypes, but no evidence of sizable gene-environment interactions.

Keywords— Gene-by-Environment Interplay; Genoeconomics; Polygenic indices; Social Science Genetics; ESSGN

^{*}Research reported in this publication was supported by NORFACE DIAL funding (462-16-100), the National Institute On Aging of the National Institutes of Health (RF1055654, R56AG058726, R01AG078522, and R01AG079554), the Dutch National Science Foundation (016.VIDI.185.044), and the European Social Science Genetics Network (ESSGN) under the European Union’s Horizon 2020 research and innovation programme’s Marie Skłodowska-Curie grant agreement (ESSGN 101073237). The views expressed in this publication are those of the authors and do not necessarily reflect those of the European Union, MSCA Horizon Europe, or ESSGN. Neither the European Union, nor the granting authority or ESSGN can be held responsible for them.

¹University of Bologna; ²Vrije Universiteit Amsterdam; ³Tinbergen Institute; ⁴Erasmus University Rotterdam; ⁵University of Zurich; ⁶University of Iowa; ⁷University of Bristol; ⁸Northwestern University; ⁹University of Southern California.

1 Introduction

Families play an essential role in shaping life chances, as evidenced by the strong intergenerational transmission of advantage^{1–11}. Indeed, the start of a human’s life can be characterized by two lotteries: one assigning genes (nature), and the other assigning a rearing environment (nurture)^{12,13}. Both operate through familial ties, and each has a substantial influence on life chances. Twin studies suggest that for a wide variety of traits, some 30 to 70 percent of the variation between individuals can be attributed to differences in genetics¹⁴. However, heritability estimates in this range imply that while genes matter, environments also play an important role. One particularly important environmental factor is parental socioeconomic status (SES), the social position that parents occupy based on education, occupation, income, and related characteristics. Parental SES is strongly tied to a family’s economic resources, and their ability to invest in the development and well-being of their children. A large literature studies how familial SES shapes early-life conditions and the subsequent adult outcomes of children^{15–20}.

A central question is whether the two “lotteries” over genes and family SES interact in shaping realized outcomes. The literature on gene-environment (G×E) interactions explores whether certain environments reduce or exacerbate genetic advantage^{21–24}. There are strong theoretical motivations to investigate this in the specific context of parental SES. In economics, child endowments (including genes) and parental resources are often considered distinct inputs in the production of children’s outcomes⁴. Yet, it is not clear whether they act as substitutes or complements. That is, parental resources could compensate for less advantageous endowments, or they could multiply the benefits of advantageous endowments. In the behavioral genetics literature, a prominent theory is the Scarr–Rowe hypothesis^{25,26}, which posits that genetic influences on cognitive ability are more fully expressed in advantageous environments. Empirically evaluating these hypotheses is important. Many redistributive social policies attempt to mitigate disparities in familial SES. The presence of gene-SES interactions suggest that such policies may also moderate gradients induced by the genetic lottery.

No firm consensus has emerged from the empirical literature evaluating gene-SES interactions. Empirical tests using twin and sibling designs have produced heterogeneous findings across cohorts and national contexts^{26–30}. In some settings, genetic variance appears to be lower under socioeconomic disadvantage. In others, such moderation is weak or absent. These results have been interpreted as reflecting complementarity between genetic endowments and environmental resources

in some contexts, and substitutability in others.

One path forward in this literature involves the use of observed measures of genetic factors constructed from molecular genetic data. The growing availability of molecular genetic data, and the development of new statistical methods for exploring $G \times E$ interactions allow researchers to explore these questions in larger and richer social-science datasets^{24,31}. Genome-wide association studies (GWASs) have robustly identified relationships between specific single-nucleotide polymorphisms (SNPs; a common measure of genetic variation) and a large and growing number of health and human-capital related outcomes^{32–35}. These molecular-level associations can be aggregated into Polygenic Indices (PGIs) that can account for a substantial proportion of the variation in their targeted phenotypes^{36,37}.

A growing literature explores interactions between PGIs and a variety of environmental exposures, including adverse conditions during childhood, neighborhood characteristics, and policies like cigarette taxes or compulsory schooling laws^{38–44}. While many of these results provide evidence specifically on gene-SES interactions, the robustness of the evidence varies depending on the traits studied, and the aspect of SES that is being measured. Although Bierut et al. (2023)⁴⁴ find evidence that childhood financial adversity exacerbates genetic risk for smoking, the review by Pasmán et al. (2019)⁴⁵ could not draw firm conclusions about genetic interactions with related environments (e.g., childhood trauma and parental education) for a wider range of substance use. Similarly, for psychiatric traits, such as major depressive disorder, interactions between PGIs and environmental variables such as childhood trauma are modest and sometimes produce conflicting results (see Assary et al. 2018⁴⁶ for a review). More robust findings exist for physical health outcomes, with social deprivation exacerbating the genetic risk of obesity^{39,47–49}. In the case of educational attainment (EA), there is some evidence that the relationship between an EA PGI and college completion is stronger among higher SES families^{50–52}.

There thus exists some evidence that $G \times SES$ interactions may shape health- and human capital-related outcomes. However, the fragmented nature of the evidence leaves open the question of whether these interactions constitute a general feature of genetic associations or are instead specific to particular traits and environments. This uncertainty is driven by three features of the literature. First, the strength of the evidence varies markedly across domains and measures of SES, with more robust findings for some behavioral and physical health outcomes than for psychiatric disorders^{38,40–44,46}. This cross-domain heterogeneity suggests that $G \times E$ associations may be highly

trait-specific. Second, the range of outcomes analyzed remains limited, as most studies focus on a small set of traits without systematically assessing whether similar interaction patterns emerge elsewhere. Results may therefore also reflect publication bias. Third, existing studies are typically conducted within single datasets, making it difficult to determine whether reported interactions are stable across samples or instead reflect particular environments.

Our study seeks to address these challenges by systematically estimating and comparing the magnitude of the interaction between a consistently measured measure of childhood environment (parental SES) – proxied by a latent factor of years of education of the father, years of education of the mother, and the occupational status of the father – and a set of consistently measured genetic predispositions (measured by PGIs) on the outcome of interest measured in adulthood, for 45 traits related to health and human capital, in three longitudinal social science data sets: the Health and Retirement Study (HRS), the Wisconsin Longitudinal Study (WLS), and the English Longitudinal Study of Ageing (ELSA) ^{53–55}.

2 Results

Using a mega-analysis that pools data from three datasets (HRS, WLS, and ELSA), we explore how genetic predispositions PGI , socioeconomic background SES , and their interaction $PGI \times SES$ contribute to later-life outcomes running 45 OLS regressions, one for each trait, in each of the three datasets, as:

$$Y_i^s = \alpha^s + \beta^s PGI_i^s + \gamma^s SES_i^s + \rho^s [PGI_i^s \times SES_i^s] + \mathbf{X}_i' \delta_s + \varepsilon_i^s,$$

where Y_i^s are the different socioeconomic, anthropometric, health, behavioral, and personality-related traits indexed by s (where s ranges over 45 different outcomes); PGI_i^s are the outcome specific polygenic indices (each PGI is constructed to predict its phenotype); and SES_i^s (parental SES) is a proxy for the SES of the family one was born into. We control for X_i , a vector containing an indicator for the dataset (HRS, WLS, and ELSA), the first ten ancestry-specific principal components of the genetic data, age and age squared (in years), sex, region of birth, interactions between age and sex, and interactions of all baseline controls with the PGI and parental SES.

The main coefficients of interest are β^s , capturing the extent to which the PGIs are predictive of their respective outcomes, γ^s , denoting the relation between parental SES and the outcomes, and

ρ^s , estimating the interaction between the PGI and parental SES. The estimated coefficients from the mega-analysis pooling all three datasets are plotted in Figure 1 for all 45 traits. We find strong genetic and socioeconomic associations: both β^s and γ^s are large in magnitude and statistically different from zero for most outcomes. Specifically, genetic associations β^s (blue triangles) are positive and significant (except for Type I diabetes and smoking cessation, which are close to zero). SES associations γ^s (red squares) exhibit a different pattern, with qualitatively adverse conditions such as depressive symptoms and anxiety consistently presenting negative associations with the SES measure, and qualitatively advantageous outcomes such as educational attainment and household income presenting positive associations, showing that parental SES is associated with better health and higher SES in later life. These findings underscore the effective predictive power of both PGIs and the parental SES measure. In contrast, we find no evidence of economically meaningful interactions between the two (orange circles): the estimated ρ^s coefficients are small in magnitude and centered around zero.^a

To better illustrate this, Figure 2 plots the distribution of the PGI, β^s , and interaction PGI \times SES, ρ^s , coefficients across the 45 different outcomes, pooled across the three datasets. The PGI coefficients are always positive, with an estimated magnitude of around 10% of a standard deviation ($\beta^s = 0.1$), while the coefficients of the PGI \times SES interactions are always close to zero, $\rho^s \approx 0$. This suggests that the statistical interaction between genes and parental SES is not substantial.

Moreover, as shown in Figure 3, the coefficients ρ^s for the PGI \times SES interactions (orange circles) are virtually all statistically indistinguishable from zero. Only five interaction terms are significantly different from zero at the 95% confidence level, namely those for risk-taking behavior, cholesterol, severe obesity, anxiety, and depressive symptoms. None of these findings remain significant after Bonferroni adjustment for multiple hypothesis testing. In contrast, both the PGI (blue triangles) and parental SES (red squares) coefficients remain statistically significant after Bonferroni correction, with exceptions in only six (parental SES) and eight (PGI) instances, respectively.

To rule out the concern that the estimated interaction terms ρ^s are close to zero due to limited predictive power of PGIs, we next select traits with relatively large genetic associations ($\beta^s > 0.05$). Figure 4 presents the distribution of the corresponding PGI (β^s) and PGI \times SES interaction (ρ^s) coefficients. The interaction estimates remain small and tightly clustered around zero, indicating the near-zero interactions are not due to weak PGIs. Even for the PGIs with the strongest predictive

^aThe full set of PGI, SES, and PGI \times SES estimates with standard errors for all 45 phenotypes is reported in Supplementary Table A.1. Corresponding dataset-specific estimates for the HRS, the WLS, and ELSA are reported in Appendix Figure A.1. Results across the datasets are highly consistent.

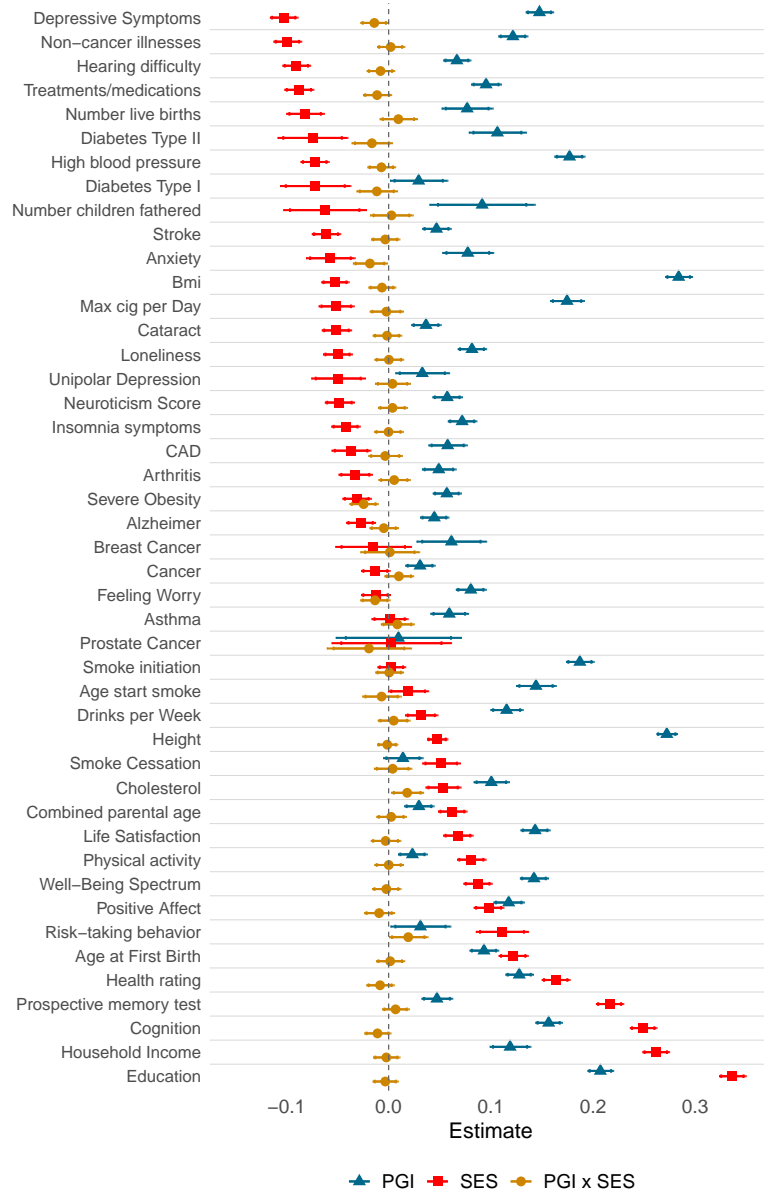


Figure 1: Mega-analysis: Associations between PGIs, parental SES, and PGI \times SES across phenotypes

Plot of the coefficients estimated from the OLS regression model specified in Equation 1, showing the associations along with their 95% confidence intervals, between polygenic indices (PGIs; triangles), parental socioeconomic status (SES; squares), and their interaction (PGI \times SES; circles) across 45 phenotypes. The mega-analysis pools all three datasets, the HRS, the WLS, and ELSA. Control variables include an indicator for each dataset, 10 ancestry-specific principal components of the genetic data, age and age squared, sex, region of birth, and interactions between age and sex.

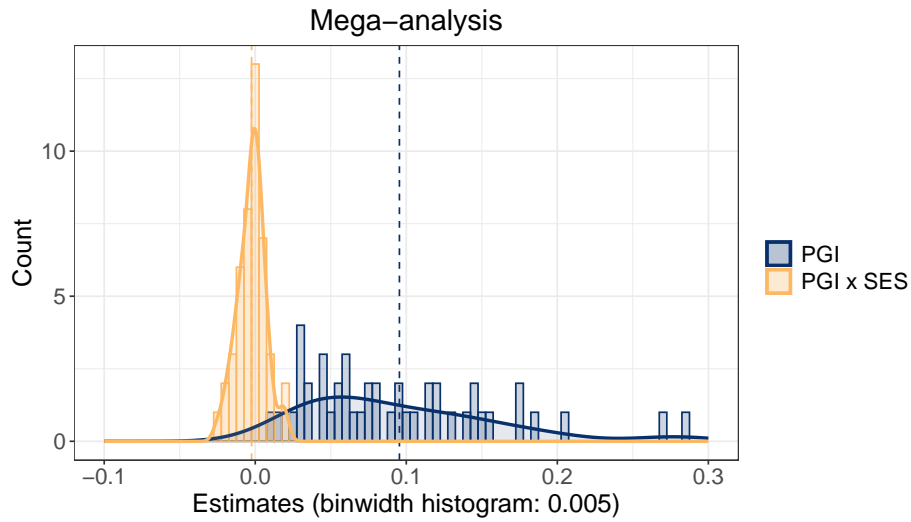


Figure 2: Density and histogram of PGI and PGI \times SES coefficients across 45 phenotypes. Smoothed density and overlaid histogram representing the distribution of estimated coefficients from the OLS regression model specified in Equation 1, based on 45 estimated coefficients, one for each phenotype. These coefficients are pooled from the Health and Retirement Study (HRS), the Wisconsin Longitudinal Study (WLS), and the English Longitudinal Study of Ageing (ELSA). The dashed vertical line represents the average coefficient for PGI and PGI \times SES across the different phenotypes.

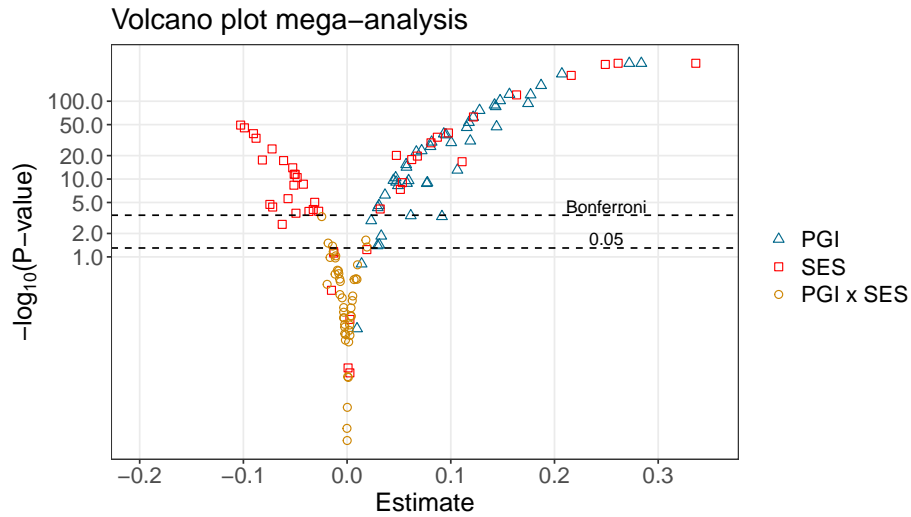


Figure 3: Volcano plot of PGI, parental SES, and PGI \times SES coefficients across 45 phenotypes. Estimated coefficients from the OLS regression model specified in Equation 1 plotted against their corresponding $-\log_{10}(\text{p-values})$, pooling the HRS, the WLS, and ELSA. The horizontal reference lines indicate conventional significance thresholds and Bonferroni-adjusted cutoffs for multiple hypothesis testing.

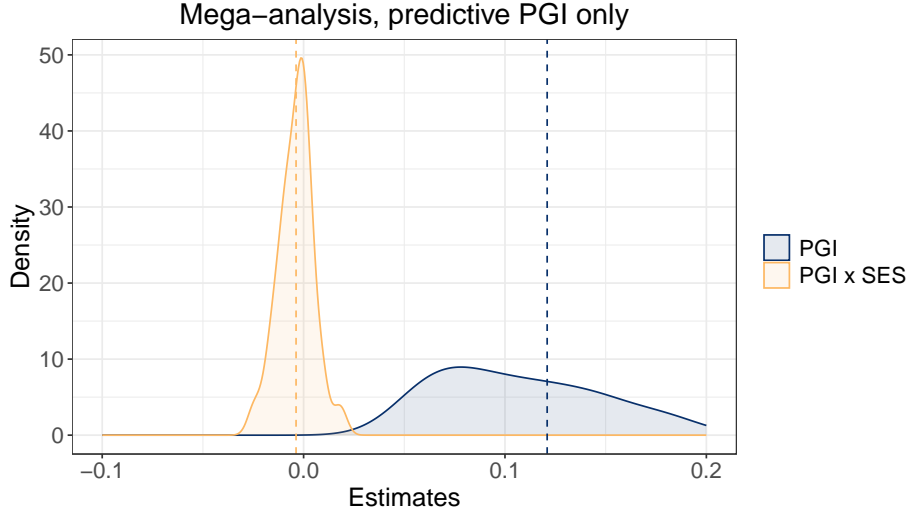


Figure 4: Histogram of estimated coefficients for PGI and PGI \times SES

Smoothed density plot of the estimated coefficients for PGIs and their interaction with SES (PGI \times SES). Pooled OLS regression model specified in Equation 1, limited to PGIs with greater predictive power, where coefficients (β) are greater than 0.05.

power, such as educational attainment, the interaction remains close to zero. For education, the estimated PGI main effect is $\beta^s = 0.207$ ($p < 0.001$), the parental SES main effect is $\gamma^s = 0.336$ ($p < 0.001$), and the PGI \times SES interaction is $\rho^s = -0.003$ ($p = 0.598$) — small and statistically indistinguishable from zero. Imposing an even more stringent criterion of restricting the analysis to phenotypes with large main associations for both the genetic predisposition and for parental SES ($\beta^s > 0.10$ and $\gamma^s > 0.10$), we end up with only six traits (13.3%) that meet this threshold. Yet even within this subset, the estimated PGI \times SES interactions remain statistically indistinguishable from zero.

All three datasets sample individuals in middle- and older-age, but while the HRS is nationally representative of the US population older than 50, the WLS is representative of Wisconsin’s high school class of 1957 and of similarly aged white Americans who completed high school, and the ELSA is representative of the English population older than 50. These differences in sampling and contexts could lead to heterogeneity in the interaction terms, potentially averaging out to near-zero associations. Figure 5 however, suggests this is unlikely. Across all datasets, PGIs exhibit similar predictive power, centered around $\beta = 0.1$. Importantly, the estimated interaction terms are consistently centered around zero, albeit with greater confidence intervals due to smaller sample

size, indicating that the absence of substantial interactions is not specific to any particular sample.

3 Discussion

Despite an extensive literature showing evidence of $G \times E$ interactions, the question of their existence, for which traits, genotypes and environments, remains far from resolved, with numerous ambiguous and even conflicting results, e.g.,^{38–50}. We approached this question by systematically exploring $PGI \times SES$ interactions in three longitudinal social-science data sets containing information on 45 human capital-related phenotypes, 45 related genotypes, and parental SES. Utilizing these data, we assessed the predictive power of parental SES, of PGIs, and of the linear interaction between parental SES and PGIs across the 45 phenotypes. At a 95% confidence level, PGIs significantly predict 43 of 45 phenotypes (95.6%) and parental SES significantly predicts 38 of 45 phenotypes (84.4%), whereas statistically significant $PGI \times SES$ interaction associations are observed for only 5 of 45 phenotypes (11.1%). None of these interaction terms remain statistically significant after correcting for multiple hypothesis testing. Importantly, this finding is not driven by weak genetic or weak SES gradients. Restricting the analysis to phenotypes with statistically significant and economically meaningful main associations, we still find no statistically significant $PGI \times SES$ interactions. Taken together, these findings provide little evidence of meaningful linear $G \times E$ interaction between PGIs and parental SES.

In other words, the predictive power of the PGIs we use does not appear to systematically differ between individuals born into high and low SES families. Thus, using a DNA-based measure of genetic predispositions, we do not find support for the Scarr-Rowe hypothesis. Previous $PGI \times E$ interaction studies have shown conflicting results regarding the Scarr-Rowe hypothesis. US-based studies confirmed the hypothesis⁵⁰, whereas European-based studies found the interactions to be nonexistent or even reversed⁵⁶. This difference between US and European settings is in line with twin research that showed that Scarr-Rowe interactions are non-existing in countries with broader welfare policies³⁰. However, in the present study, exploiting both US and UK data, we find negligible differences across datasets.

There are two main interpretations of our work. First, the lack of significant $PGI \times SES$ interactions in a systematic and comprehensive analysis could indicate that previous results were affected by ‘publication bias’, the tendency for studies with significant results to be published more frequently than studies with insignificant results⁵⁷. This bias can skew the overall literature on a

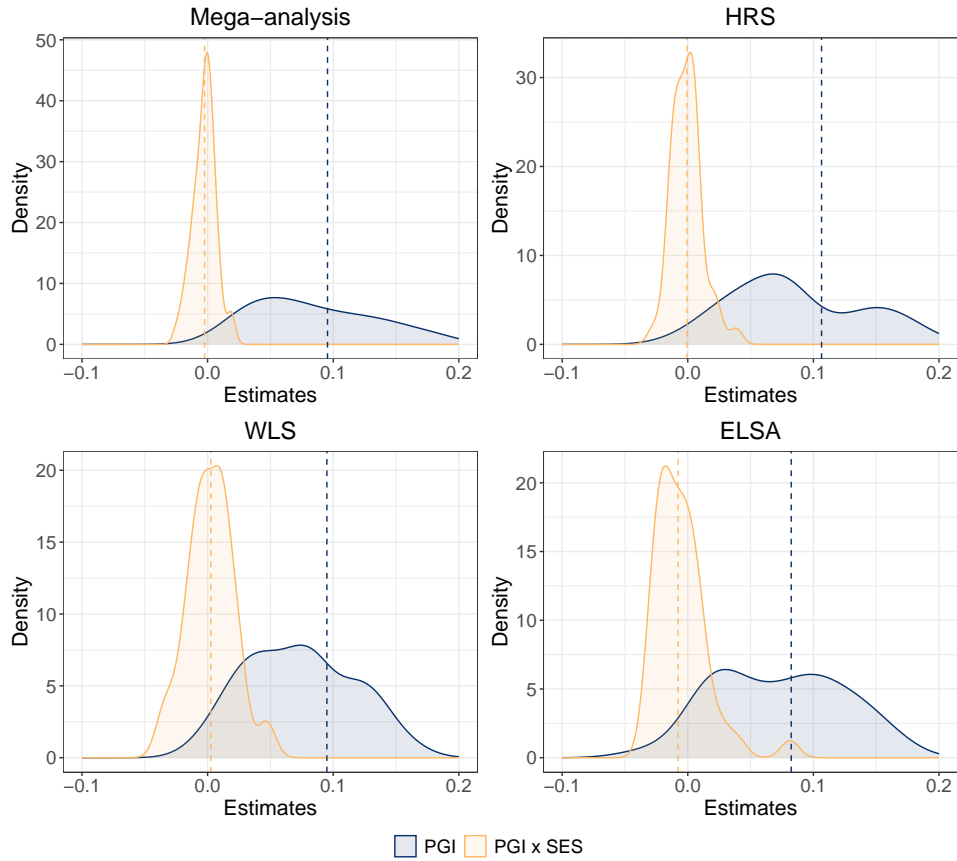


Figure 5: Histogram of PGI and $\text{PGI} \times \text{SES}$ coefficients across 45 phenotypes for different databases

Smoothed density of the estimated coefficients for the PGI and $\text{PGI} \times \text{SES}$ from the OLS regression model specified in Equation 1, presented separately for each dataset: the HRS, the WLS, and ELSA. The density plots illustrate the distribution of estimated coefficients within each dataset, allowing for a comparison of their spread and central tendency across the three samples. The dashed vertical line represents the average coefficient for PGI and $\text{PGI} \times \text{SES}$.

topic, as the published research may not accurately represent the true effect size or the range of results. This has been found to be a prevalent issue in the social sciences^{58,59}. In the young field of social genomics, this bias may have led scholars, consciously or unconsciously, to concentrate on a select group of phenotypes and genetic measures that present G×E interactions, thereby neglecting a broader spectrum of relevant lifetime outcomes. In an aggregate manner, this phenomenon could have contributed to the impression that G×E interactions are the norm, while in reality, they may be the exception. This misperception could skew both theoretical understanding and practical applications, leading to an overestimation of the prevalence and significance of G×E interactions in the literature. This interpretation, consistent with our results, would suggest that overall, genes (nature) and parental SES (nurture) contribute to human capital formation^{14,60–62}, but mostly in an independent manner.

However, despite using some of the largest GWASs available, it could be that our genetic measures simply have too much measurement error⁶³. Moreover, the parental SES measure used in the current study might not be optimal: it is possible that the size of a PGI×SES interaction is not related to general parental SES, but to rather more extreme or specific situations such as severe financial distress or an abusive family environment. Studies have shown that childhood maltreatment and traumatic circumstances can have large effects on children, lasting into adulthood^{64–66}. Epigenetic studies even show clear biological mechanisms explaining life course differences between children with and without a history of maltreatment^{67–69}.

Not finding systematic PGI×SES may also be the result of incorrectly specifying the underlying model: human capital may be a more complex function of both nature and nurture, not captured by a simple model with a linear interaction term. Therefore, the PGI may fail to capture the specific measures of genetic variation underlying the interplay between genetic variation and SES. Potentially, methods such as the variance GWAS^{70,71} or PIGEON⁷² may elucidate more complex interaction mechanisms, as such approaches utilize more direct measures of variance in phenotypic outcomes associated with SNPs. Finally, parental SES and genetic inheritance are not independent in a between-family setting. This gene-environment correlation (r_{GE}) may bias the estimation of PGI×SES interactions²⁴.

Another possibility is that G×E estimates are sensitive to outcome definition, as alternative specifications group individuals into substantively distinct categories that may be differentially affected by both SES and genetic factors. This is illustrated by the interaction with educational

attainment, where our results diverge from those of Papageorge and Thom (2020)⁵⁰ and Ghirardi and Bernardi (2025)⁵², despite relying on the same data (HRS), phenotype (education), and genetic measure (EA PGI). The discrepancy appears to arise from differences in how the outcome is constructed: our analysis uses years of education as a continuous measure, whereas Papageorge and Thom (2020)⁵⁰ and Ghirardi and Bernardi (2025)⁵² focus on dichotomous outcomes (e.g., college or high school completion). Replicating their approach, we find a positive and statistically significant interaction between the education PGI and SES when using a college completion indicator ($\rho = 0.01$, $p = 0.007$). In contrast, when high school completion is used as the outcome, the interaction is negative and statistically significant ($\rho = -0.06$, $p = 0.000$). This pattern suggests that college completion is a highly selective margin along which SES amplifies genetic effects, whereas high school completion reflects a less selective threshold where SES attenuates them. Consequently, aggregating these distinct margins into a linear measure such as years of education may mask offsetting effects and produce null estimates. More broadly, these results highlight the sensitivity of G×E estimates to outcome definition and underscore the importance of adopting outcome-specific, theoretically grounded measures for which interactions are substantively meaningful.

These potential interpretations are not mutually exclusive and likely explain different aspects of our results. Yet, notwithstanding the methodological limitations, the evidence presented here meaningfully shifts prior expectations by suggesting that G×E interactions may be less widespread across major life-course outcomes than previously assumed, with genetic and socio-economic factors more often acting independently. Our findings also underscore the necessity of advancing genetic measures with enhanced predictive capabilities and identifying specific environmental factors influencing genetic expression. Furthermore, it emphasizes the importance of developing models and functional forms that best elucidate the interactions between genes and the environment.

4 Methods

4.1 Data

We use three longitudinal social science datasets, the Health and Retirement Study (HRS), the Wisconsin Longitudinal Study (WLS), and the English Longitudinal Study of Ageing (ELSA), containing rich genotypic and phenotypic information. We systematically analyze the relationship between the offspring’s (henceforth, the child’s) genetic propensity towards a trait (genotype), the socioeconomic status (SES) of the families they grew up in when they were children, and their actual trait in adulthood (phenotype).

The HRS is a longitudinal household survey providing rich data on about 26,000 individuals, representative of the U.S. population over the age of 50. Up to 13 waves of data per respondent are available, from 1992–2016. We use the publicly available HRS core survey and linked genetic data for the years 2006, 2008, and 2010. HRS core surveys are conducted biennially using a combination of face-to-face and telephone interviewing.

The WLS is a long-term study of a sample of 10,317 men and women who graduated from Wisconsin high schools in 1957. Survey data were collected from the original respondents or their parents in 1957, 1964, 1975, 1992, 2004, and 2011.

ELSA is a study on the dynamics of health, social, well-being, and economic circumstances in the English population aged 50 and older. More than 18,000 people have taken part in the study since it started in 2002, with the same people re-interviewed every two years.

4.2 Phenotypes

Phenotypes were selected from the GWAS catalog and the GWAS Atlas^{73,74}, versions May-30-2019 and April-17-2019, respectively. We selected the most relevant and powerful GWASs for various trait ‘domains/families’. All included GWASs passed the following exclusion criteria:

- Number of SNPs $> 450,000$.
- Over 80 percent of samples of EUR population.
- SNP-based heritability Z-score > 2 .
- All relevant data fields available (e.g., beta or odds ratio, effect, and non-effect alleles).
- GWAS with the largest N for the trait of interest.

- The exact or highly comparable phenotype must be available in the HRS, WLS, and ELSA.

Ultimately, we selected 45 traits (phenotypes), with corresponding GWASs. An overview of these 45 GWASs is provided in Supplementary Table B.1.

We defined each trait across HRS, WLS and ELSA, constructing 45 phenotypes that are as similar as possible to the definition of the corresponding GWAS. In cases where the phenotypes differed from the GWAS definition, we harmonized the phenotypes as much as possible. Phenotypes between studies were sometimes measured with different questions. This might partially explain the differences in phenotypic correlations between studies. Some phenotypes with important differences in measurement between surveys are Cholesterol, Hearing difficulty, Osteoarthritis, and Type I and II Diabetes.

4.3 Childhood SES

We measure the Socioeconomic Status (SES) of individuals in their childhood through an estimate of parental SES. We estimate a General structural equation model, with observed and latent variables. For the observed variables we used:

- Mother’s years of education
- Father’s years of education
- Father’s occupation

These variables exist in each of the HRS, WLS, and ELSA, with only minor differences. Father’s occupation variables are defined as 0 for “Other”, 1 for “Farmers”, 2 for “Services + Operators + Admin/Clerical”, 3 for “Sales”, 4 for “Management”, and 5 for “Professional”. These categories were assigned based on an analysis of income and education levels in the 3 studies. ELSA does not include the option “Farmer”, and therefore that category is excluded only for ELSA. The years of education reported by the father and mother vary between surveys. In HRS and ELSA the minimum number of years of parental education reported is 0 and the maximum is 14 for ELSA and 17 or more for HRS. In WLS the minimum is 7 years and the maximum 18. From the general structural equation model, we calculate a standardized SES latent factor.

The distribution of the SES factor for ELSA is much bigger than for the other two studies due to the presence of a higher rate of respondents that reported 0 years of maternal and/or paternal education. We therefore imputed 7 years of education for such respondents and included a

dummy variable to identify these individuals in the structural model to obtain a narrower and more comparable distribution of the SES factor.

4.4 Polygenic Index

PGIs were constructed using GWAS results that leave HRS, WLS, and ELSA out for reasons of statistical independence. We selected GWASs from the NHGRI-EBI GWAS Catalog⁷⁵ and the GWAS Atlas⁷⁴, restricting to phenotypes available in all three datasets. We applied criteria for sufficient statistical power, including GWAS sample size, whether a superior GWAS was available for the same phenotype, the number of genome-wide significant hits, SNP-based heritability, and phenotype prevalence in the population, selecting the most relevant phenotype within each domain.

We generated two types of PGI for each phenotype: Full SNPs PGI and Top SNPs PGI. PGIs were constructed using PLINK as weighted sums of allele counts using GWAS effect sizes as weights, without linkage disequilibrium pruning or clumping. The Full SNPs PGI aggregates all available SNPs, capturing genetic associations across the whole genome (p -value ≤ 1). The Top SNPs PGI aggregates only SNPs that reach genome-wide significance in the GWAS (p -value $\leq 5 \times 10^{-8}$). To account for multiple hypothesis testing across approximately one million independent SNPs in the human genome (adjacent SNPs are often in linkage disequilibrium, i.e., inherited together), the commonly used threshold for genome-wide statistical significance is $p < 5 \times 10^{-8}$ (i.e., 0.05/1,000,000).

Genetic quality control steps for the target datasets (HRS, WLS, and ELSA) were conducted following guidelines recommended in the literature⁷⁶.

4.5 Statistical Methods

We estimate the relationship between phenotypes, polygenic indices (PGIs), socioeconomic status (SES), and the interaction between SES and PGI using pooled data from HRS, ELSA, and WLS. The model is estimated by OLS on the pooled sample and includes dataset fixed effects, as well as controls for age, sex, and the first 10 genetic principal components. Prior to estimation, phenotypes are standardized to have mean zero and variance one. The Full SNPs PGI and the SES factor are also standardized to have mean zero and variance one. As a robustness check, we also estimated the models separately for each dataset and using the top SNPs PGI aggregating the genome-wide significant SNPs. PGIs restricted to genome-wide significant SNPs generally lacked power and are therefore not presented.

References

- [1] Blau, P. M. & Duncan, O. D. *The American Occupational Structure* (Wiley, New York, 1967).
- [2] Jencks, C. *et al.* *Inequality: A reassessment of the effect of family and schooling in America* (Basic Books, New York, 1972).
- [3] Bourdieu, P. & Passeron, J.-C. *Reproduction in Education, Society and Culture* (Sage, 1977).
- [4] Becker, G. S. & Tomes, N. Human capital and the rise and fall of families. *Journal of Labor Economics* **4**, 1–47 (1986). URL <http://www.jstor.org/stable/2534952>.
- [5] Black, S. E., Devereux, P. J. & Salvanes, K. G. Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American Economic Review* **95**, 437–449 (2005). URL <http://pubs.aeaweb.org/doi/10.1257/0002828053828635>.
- [6] Mare, R. D. A multigenerational view of inequality. *Demography* **48**, 1–23 (2011).
- [7] Solon, G. Intergenerational income mobility in the united states. *American Economic Review* **82**, 393–408 (1992).
- [8] Shonkoff, J. P. & Phillips, D. A. *From neurons to neighborhoods: The science of early childhood development* (National Academy Press, Washington, DC, 2000).
- [9] Heckman, J. J. Skill formation and the economics of investing in disadvantaged children. *Science* **312**, 1900–1902 (2006).
- [10] Corak, M. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives* **27**, 79–102 (2013).
- [11] Chetty, R., Friedman, J. N., Saez, E., Turner, N. & Yagan, D. Income segregation and intergenerational mobility across colleges in the united states. *Quarterly Journal of Economics* **135**, 1567–1633 (2020). URL <https://academic.oup.com/qje/article/135/3/1567/5741707>.
- [12] Harden, K. P. *The genetic lottery: Why DNA matters for social equality* (Princeton University Press, 2021).
- [13] Houmark, M. A., Ronda, V. & Rosholm, M. The nurture of nature and the nature of nurture: How genes and investments interact in the formation of skills. *American Economic Review* **114**, 385–425 (2024).

- [14] Polderman, T. J. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* **47**, 702–709 (2015).
- [15] Barker, D. J. *et al.* Fetal nutrition and cardiovascular disease in adult life. *Lancet* **341**, 938–941 (1993).
- [16] Galobardes, B., Lynch, J. W. & Smith, G. D. Is the association between childhood socioeconomic circumstances and cause-specific mortality established? update of a systematic review. *Journal of Epidemiology & Community Health* **62**, 387–390 (2008).
- [17] Cunha, F., Heckman, J. J., Lochner, L. & Masterov, D. V. Interpreting the evidence on life cycle skill formation. In *Handbook of the Economics of Education*, vol. 1, 697–812 (Elsevier, 2006).
- [18] Almond, D. & Currie, J. Human capital development before age five. In Ashenfelter, O. & Card, D. (eds.) *Handbook of Labor Economics*, vol. 4B, 1315–1486 (Elsevier, 2011).
- [19] Aizer, A. & Currie, J. The intergenerational transmission of inequality: Maternal disadvantage and health at birth. *Science* **344**, 856–861 (2014).
- [20] Almond, D., Currie, J. & Duque, V. Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature* **56**, 1360–1446 (2018).
- [21] Haldane, J. B. S. The interaction of nature and nurture. *Annals of Eugenics* **13**, 197–205 (1946). URL <http://www.ncbi.nlm.nih.gov/pubmed/20282564>.
- [22] Plomin, R., DeFries, J. C. & Loehlin, J. C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin* **84**, 309 (1977).
- [23] McAllister, K. *et al.* Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *American Journal of Epidemiology* **186**, 753–761 (2017).
- [24] Biroli, P. *et al.* The economics and econometrics of gene–environment interplay. *Review of Economic Studies* **93**, 144–180 (2025).
- [25] Scarr-Salapatek, S. Race, social class, and iq. *Science* **174**, 1285–1295 (1971).

- [26] Rowe, D. C., Jacobson, K. C. & Van den Oord, E. J. Genetic and environmental influences on vocabulary iq: Parental education level as moderator. *Child development* **70**, 1151–1162 (1999).
- [27] Turkheimer, E. N., Haley, A., Waldron, M., D’Onofrio, B. M. & Gottesman, I. I. Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* **14**, 623–628 (2003). URL <http://pss.sagepub.com/content/14/6/623.full>.
- [28] Harden, K. P., Turkheimer, E. & Loehlin, J. C. Genotype by environment interaction in adolescents’ cognitive aptitude. *Behavior Genetics* **37**, 273–283 (2007).
- [29] Hanscombe, K. B. *et al.* Socioeconomic status (ses) and children’s intelligence (iq): In a uk-representative sample ses moderates the environmental, not genetic, effect on iq. *PLoS ONE* **7**, e30320 (2012).
- [30] Tucker-Drob, E. M. & Bates, T. C. Large cross-national differences in gene \times socioeconomic status interaction on intelligence. *Psychological science* **27**, 138–149 (2016).
- [31] Domingue, B. W., Trejo, S., Armstrong-Carter, E. & Tucker-Drob, E. M. Interactions between polygenic scores and environments: Methodological and conceptual challenges. *Sociological Science* **7**, 465–486 (2020). URL <http://dx.doi.org/10.15195/v7.a19>.
- [32] Hill, W. D. *et al.* Molecular genetic contributions to social deprivation and household income in uk biobank. *Current Biology* **26**, 3083–3089 (2016).
- [33] Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**, 1112–1121 (2018).
- [34] Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of european ancestry. *Human Molecular Genetics* **27**, 3641–3649 (2018).
- [35] Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *American Journal of Human Genetics* **110**, 179–194 (2023).
- [36] Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236–1241 (2015).

- [37] Becker, J. *et al.* Resource profile and user guide of the polygenic index repository. *Nature Human Behaviour* **5**, 1744–1758 (2021).
- [38] Meyers, J. L. *et al.* Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit neighborhood health study. *Translational Psychiatry* **3**, e290 (2013). URL <http://dx.doi.org/10.1038/tp.2013.63><http://www.nature.com/doifinder/10.1038/tp.2013.63>.
- [39] Barcellos, S. H., Carvalho, L. S. & Turley, P. Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences* **115**, E9765–E9772 (2018).
- [40] Domingue, B. W., Conley, D., Fletcher, J. & Boardman, J. D. Cohort effects in the genetic influence on smoking. *Behavior Genetics* **46**, 31–42 (2016).
- [41] Fletcher, J. M. Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. *PloS ONE* **7**, e50576 (2012). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3515624{&}tool=pmcentrez{&}rendertype=abstract>.
- [42] Schmitz, L. & Conley, D. The long-term consequences of Vietnam-era conscription and genotype on smoking behavior and health. *Behavior Genetics* **46**, 43–58 (2016). URL <http://link.springer.com/10.1007/s10519-015-9739-1>.
- [43] Treur, J. L. *et al.* Testing familial transmission of smoking with two different research designs. *Nicotine and Tobacco Research* **20**, 836–842 (2018).
- [44] Bierut, L., Biroli, P., Galama, T. J. & Thom, K. Challenges in studying the interplay of genes and environment. a study of childhood financial distress moderating genetic predisposition for peak smoking. *Journal of Economic Psychology* **98**, 102636 (2023).
- [45] Pasman, J. A., Verweij, K. J. & Vink, J. M. Systematic review of polygenic gene–environment interaction in tobacco, alcohol, and cannabis use. *Behavior Genetics* **49**, 349–365 (2019).
- [46] Assary, E., Vincent, J. P., Keers, R. & Pluess, M. Gene-environment interaction and psychiatric disorders: Review and future directions. In *Seminars in cell & developmental biology*, vol. 77, 133–143 (Elsevier, 2018).

- [47] Tyrrell, J. *et al.* Gene-obesogenic environment interactions in the UK Biobank study. *International Journal of Epidemiology* **46**, 559–575 (2017). URL <https://doi.org/10.1093/ije/dyw337>. <https://academic.oup.com/ije/article-pdf/46/2/559/24172372/dyw337.pdf>.
- [48] Amin, V. *et al.* Gene-environment interactions between education and body mass: Evidence from the UK and Finland. *Social Science & Medicine* **195**, 12–16 (2017).
- [49] Hoang, C. T., Amin, V., Behrman, J. R., Kohler, H.-P. & Kohler, I. V. Heterogenous trajectories in physical, mental and cognitive health among older americans: Roles of genetics and life course contextual factors. *SSM-Population Health* **23**, 101448 (2023).
- [50] Papageorge, N. W. & Thom, K. Genes, education, and labor market outcomes: evidence from the health and retirement study. *Journal of the European Economic Association* **18**, 1351–1399 (2020).
- [51] Ghirardi, G., Gil-Hernández, C. J., Bernardi, F., van Bergen, E. & Demange, P. Interaction of family SES with children’s genetic propensity for cognitive and noncognitive skills: No evidence of the Scarr-Rowe hypothesis for educational outcomes. *Research in Social Stratification and Mobility* **92**, 100960 (2024).
- [52] Ghirardi, G. & Bernardi, F. Compensating or boosting genetic propensities? Gene-family socioeconomic status interactions by educational outcome selectivity. *Social Science Research* **129**, 103174 (2025).
- [53] Sonnega, A. *et al.* Cohort profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology* **43**, 576–585 (2014). URL <http://www.ncbi.nlm.nih.gov/pubmed/24671021><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3997380><https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyu067>.
- [54] Herd, P., Carr, D. & Roan, C. Cohort profile: Wisconsin longitudinal study (wls). *International Journal of Epidemiology* **43**, 34–41 (2014).
- [55] Steptoe, A., Breeze, E., Banks, J. & Nazroo, J. Cohort profile: The English Longitudinal Study of Ageing. *International Journal of Epidemiology* **42**, 1640–1648 (2013). URL <https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dys168>.

- [56] Isungset, M. A. *et al.* Social and genetic associations with educational performance in a scandinavian welfare state. *Proceedings of the National Academy of Sciences* **119**, e2201869119 (2022).
- [57] Sterling, T. D. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American statistical association* **54**, 30–34 (1959).
- [58] Franco, A., Malhotra, N. & Simonovits, G. Publication bias in the social sciences: Unlocking the file drawer. *Science* **345**, 1502–1505 (2014).
- [59] Peplow, M. Social sciences suffer from severe publication bias. *Nature* **10** (2014).
- [60] Turkheimer, E. Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science* **9**, 160–164 (2000).
- [61] Hout, M., Brooks, C. & Manza, J. The persistence of classes in post-industrial societies. *International Sociology* **8**, 259–277 (1993).
- [62] Siponen, S. M., Ahonen, R. S., Savolainen, P. H. & Hämeen-Anttila, K. P. Children’s health and parental socioeconomic factors: a population-based survey in finland. *BMC Public Health* **11**, 1–8 (2011).
- [63] Van Kippersluis, H. *et al.* Overcoming attenuation bias in regressions using polygenic indices. *Nature Communications* **14**, 4473 (2023).
- [64] Norman, R. E. *et al.* The long-term health consequences of child physical abuse, emotional abuse, and neglect: A systematic review and meta-analysis. *PLoS Medicine* **9**, e1001349 (2012).
- [65] Gardner, M. J., Thomas, H. J. & Erskine, H. E. The association between five forms of child maltreatment and depressive and anxiety disorders: A systematic review and meta-analysis. *Child Abuse & Neglect* **96**, 104082 (2019).
- [66] D’arcy-Bewick, S. *et al.* Childhood abuse and neglect, and mortality risk in adulthood: A systematic review and meta-analysis. *Child Abuse & Neglect* **134**, 105922 (2022).
- [67] Teicher, M. H., Samson, J. A., Anderson, C. M. & Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nature Reviews Neuroscience* **17**, 652–666 (2016).

- [68] Cowell, R. A., Cicchetti, D., Rogosch, F. A. & Toth, S. L. Childhood maltreatment and its effect on neurocognitive functioning: Timing and chronicity matter. *Development and Psychopathology* **27**, 521–533 (2015).
- [69] Teicher, M. H. & Samson, J. A. Annual research review: Enduring neurobiological effects of childhood abuse and neglect. *Journal of Child Psychology and Psychiatry* **57**, 241–266 (2016).
- [70] Johnson, R., Sotoudeh, R. & Conley, D. Polygenic scores for plasticity: a new tool for studying gene–environment interplay. *Demography* **59**, 1045–1070 (2022).
- [71] Schmitz, L. L., Goodwin, J., Miao, J., Lu, Q. & Conley, D. The impact of late-career job loss and genetic risk on body mass index: Evidence from variance polygenic scores. *Scientific Reports* **11**, 1–15 (2021).
- [72] Miao, J. *et al.* Pigeon: a statistical framework for estimating gene–environment interaction for polygenic traits. *Nature human behaviour* **9**, 1654–1668 (2025).
- [73] Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**, D1001–D1006 (2013). URL <https://doi.org/10.1093/nar/gkt1229>.
<https://academic.oup.com/nar/article-pdf/42/D1/D1001/3646160/gkt1229.pdf>.
- [74] Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339–1348 (2019).
- [75] Buniello, A. *et al.* The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**, D1005–D1012 (2019).
- [76] Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research* **27**, e1608 (2018).
- [77] Bulik-Sullivan, B. K. *et al.* Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291–295 (2015).
- [78] Baselmans, B. M. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics* **51**, 445–451 (2019).
- [79] Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics* **50**, 668–681 (2018).

- [80] Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343–352 (2019).
- [81] Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature genetics* **51**, 245–257 (2019).
- [82] Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature genetics* **48**, 1462–1472 (2016).
- [83] Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics* **51**, 237–244 (2019).
- [84] Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer’s disease risk. *Nature genetics* **51**, 404–413 (2019).
- [85] Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular psychiatry* **21**, 1391–1399 (2016).
- [86] Zengini, E. *et al.* Genome-wide analyses using uk biobank data provide insights into the genetic architecture of osteoarthritis. *Nature genetics* **50**, 549–558 (2018).
- [87] Demenaïs, F. *et al.* Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nature genetics* **50**, 42–53 (2018).
- [88] Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
- [89] Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
- [90] Neale, B. M. & *et al.* Rapid gwas of thousands of phenotypes in the uk biobank. <http://www.nealelab.is/uk-biobank> (2018). Neale Lab UK Biobank GWAS.
- [91] Hoffmann, T. J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nature genetics* **50**, 401–413 (2018).

- [92] Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics* **50**, 912–919 (2018).
- [93] Deelen, J. *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nature communications* **10**, 3669 (2019).
- [94] Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature genetics* **49**, 1385–1391 (2017).
- [95] Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature genetics* **47**, 381–386 (2015).
- [96] Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature genetics* **50**, 1505–1513 (2018).
- [97] Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics* **46**, 1173–1186 (2014).
- [98] Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature genetics* **51**, 394–403 (2019).
- [99] Day, F. R., Ong, K. K. & Perry, J. R. Elucidating the genetic basis of social interaction and isolation. *Nature communications* **9**, 2457 (2018).
- [100] Klimentidis, Y. C. *et al.* Genome-wide association study of habitual physical activity in over 377,000 uk biobank participants identifies multiple variants including *cadm2* and *apoe*. *International journal of obesity* **42**, 1161–1176 (2018).
- [101] Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature genetics* **50**, 928–936 (2018).
- [102] Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics* **50**, 524–537 (2018).
- [103] Smith, J., Ryan, L., Sonnega, A. & Weir, D. Psychosocial and lifestyle questionnaire 2006–2016 documentation report core section LB. Tech. Rep., Survey Research Center Institute for Social

- Research University of Michigan Ann Arbor, Michigan (2017). URL https://hrs.isr.umich.edu/sites/default/files/biblio/HRS2006-2016SAQDocumentation_{_}07.06.17.pdf.
- [104] Lachman, M. E. & Weaver, S. L. The Midlife Development Inventory (MIDI) personality scales: Scale construction and scoring. Tech. Rep., Psychology Department, Brandeis University (1997). URL <https://www.brandeis.edu/psychology/lachman/pdfs/midi-personality-scales.pdf>.
- [105] Kimball, M. S., Sahm, C. R. & Shapiro, M. D. Imputing risk tolerance from survey responses. *Journal of the American Statistical Association* **103**, 1028–1038 (2008).
- [106] Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics* (2019). URL <http://www.nature.com/articles/s41588-018-0307-5>.
- [107] Stephan, Y., Sutin, A. R., Kornadt, A., Caudroit, J. & Terracciano, A. Higher IQ in adolescence is related to a younger subjective age in later life: Findings from the Wisconsin Longitudinal Study. *Intelligence* **69**, 195–199 (2018).
- [108] Center for Demography of Health and Aging. Documentation of scales in Wisconsin Longitudinal Study. Tech. Rep., University of Wisconsin (1992).
- [109] Savard, J. R. *et al.* Wisconsin Longitudinal Study user’s guide. Tech. Rep., University of Wisconsin (1957). URL <http://www.ssc.wisc.edu/wlsresearch>.
- [110] Gale, C. R., Allerhand, M., Sayer, A. A., Cooper, C. & Deary, I. J. The dynamic relationship between cognitive function and walking speed: The English Longitudinal Study of Ageing. *Age* **36**, 9682 (2014).
- [111] Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. The satisfaction with life scale. *Journal of Personality Assessment* **49**, 71–75 (1985).
- [112] Van de Velde, S., Levecque, K. & Bracke, P. Measurement equivalence of the CES-D 8 in the general population in Belgium: A gender perspective. *Archives of Public Health* **67**, 15 (2009). URL <https://archpublichealth.biomedcentral.com/articles/10.1186/0778-7367-67-1-15>.

- [113] Karim, J., Weisz, R., Bibi, Z. & ur Rehman, S. Validation of the eight-item Center for Epidemiologic Studies Depression Scale (CES-D) among older adults. *Current Psychology* **34**, 681–692 (2015).
- [114] Steffick, D. E. *et al.* HRS/AHEAD cocumentation report documentation of affective functioning measures in the Health and Retirement Study. Tech. Rep., University of Michigan (2000).
- [115] Hughes, M. E., Waite, L. J., Hawkley, L. C. & Cacioppo, J. T. A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on Aging* **26**, 655–672 (2004).
- [116] Akaneya, Y. *et al.* Ephrin-a5 and epha5 interaction induces synaptogenesis during early hippocampal development. *PloS ONE* **5**, e9697 (2010).
- [117] Lin, P. *et al.* A new statistic to evaluate imputation reliability. *PloS ONE* **5** (2010).
- [118] Weir, D. *et al.* Quality control report for genotypic data. *Health and Retirement Study* (2012).
- [119] 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

Appendix

A Extended Results



Figure A.1: Associations between PGIs, parental SES, and PGI \times SES across phenotypes

Plot of the coefficients estimated from the OLS regression model specified in Equation 1, showing the associations, along with their 90% and 95% confidence intervals, between polygenic indices (PGIs), parental socioeconomic status (SES), and their interaction (PGI \times SES) across 45 phenotypes. The estimates are obtained from a mega-analysis that pools all three datasets—the HRS, the WLS, and ELSA—and from separate analyses for each individual database. Control variables include an indicator for each dataset, 10 ancestry-specific principal components, age and age squared, sex, region of birth, and interactions between age and sex.

Table A.1 reports the estimates and standard errors for the PGI, parental SES, and the PGI \times SES interaction across the 45 phenotypes, obtained from the mega-analysis pooling HRS, WLS, and ELSA.

Table A.1: Associations between PGIs, parental SES, and PGI \times SES across phenotypes

Phenotype	PGI (β^s)	SES (γ^s)	PGI \times SES (ρ^s)
Depressive Symptoms	0.148*** (0.007)	-0.103*** (0.007)	-0.014* (0.007)
Non-cancer illnesses	0.122*** (0.007)	-0.099*** (0.007)	0.002 (0.007)
Hearing difficulty	0.067*** (0.007)	-0.090*** (0.007)	-0.008 (0.007)
Treatments / medications taken	0.095*** (0.007)	-0.088*** (0.007)	-0.011 (0.007)
Number of live births (female)	0.077*** (0.013)	-0.082*** (0.009)	0.010 (0.009)
Diabetes Type II	0.106*** (0.014)	-0.074*** (0.017)	-0.016 (0.010)
High blood pressure	0.177*** (0.007)	-0.072*** (0.007)	-0.007 (0.007)
Diabetes Type I	0.029* (0.014)	-0.072*** (0.018)	-0.012 (0.010)
Number of children fathered (male)	0.091*** (0.026)	-0.063** (0.021)	0.003 (0.011)
Stroke	0.047*** (0.007)	-0.061*** (0.007)	-0.003 (0.007)
Anxiety	0.077*** (0.013)	-0.057*** (0.012)	-0.018* (0.008)
BMI	0.284*** (0.007)	-0.052*** (0.007)	-0.007 (0.007)
Max cig per Day	0.175*** (0.008)	-0.051*** (0.009)	-0.002 (0.008)
Cataract	0.037*** (0.007)	-0.051*** (0.007)	-0.002 (0.007)
Loneliness	0.082*** (0.007)	-0.050*** (0.007)	0.000 (0.007)
Unipolar Depression	0.033* (0.013)	-0.049*** (0.013)	0.004 (0.009)
Neuroticism Score	0.057*** (0.007)	-0.048*** (0.007)	0.004 (0.007)
Insomnia symptoms	0.072*** (0.007)	-0.042*** (0.007)	-0.000 (0.007)
Coronary Artery Disease (CAD)	0.058*** (0.010)	-0.037*** (0.010)	-0.003 (0.008)
Arthritis	0.049*** (0.008)	-0.033*** (0.008)	0.005 (0.008)
Severe Obesity	0.057*** (0.007)	-0.031*** (0.007)	-0.024*** (0.007)
Alzheimer	0.045*** (0.007)	-0.027*** (0.007)	-0.005 (0.007)
Breast Cancer	0.061*** (0.017)	-0.015 (0.019)	0.001 (0.015)
Cancer	0.031*** (0.007)	-0.013 (0.007)	0.010 (0.007)
Feeling Worry	0.080*** (0.007)	-0.013 (0.007)	-0.013 (0.007)
Asthma	0.059*** (0.009)	0.001 (0.009)	0.009 (0.008)
Prostate Cancer	0.010 (0.031)	0.003 (0.030)	-0.019 (0.021)
Smoke initiation	0.187*** (0.007)	0.003 (0.007)	0.001 (0.007)
Age start smoke	0.144*** (0.010)	0.019 (0.010)	-0.007 (0.010)
Drinks per Week	0.115*** (0.008)	0.032*** (0.008)	0.005 (0.008)
Height	0.272*** (0.005)	0.048*** (0.005)	-0.001 (0.005)
Smoke Cessation	0.014 (0.010)	0.051*** (0.009)	0.004 (0.009)
Cholesterol	0.100*** (0.009)	0.053*** (0.009)	0.018* (0.008)
Combined parental age	0.030*** (0.007)	0.062*** (0.007)	0.002 (0.007)

Continued on next page

Table A.1 – *Continued from previous page*

Phenotype	PGI (β^s)	SES (γ^s)	PGI \times SES (ρ^s)
Life Satisfaction	0.143*** (0.007)	0.068*** (0.007)	-0.003 (0.007)
Physical activity	0.023** (0.007)	0.081*** (0.007)	0.000 (0.007)
Well-Being Spectrum	0.142*** (0.007)	0.087*** (0.007)	-0.002 (0.007)
Positive Affect	0.117*** (0.008)	0.098*** (0.007)	-0.009 (0.007)
Risk-taking behavior	0.031* (0.015)	0.111*** (0.013)	0.019* (0.010)
Age at First Birth	0.093*** (0.007)	0.122*** (0.007)	0.002 (0.007)
Health rating	0.128*** (0.007)	0.163*** (0.007)	-0.008 (0.007)
Prospective memory test	0.047*** (0.008)	0.216*** (0.007)	0.007 (0.007)
Cognition	0.157*** (0.007)	0.249*** (0.007)	-0.011 (0.007)
Household Income	0.119*** (0.010)	0.261*** (0.007)	-0.002 (0.007)
Education	0.207*** (0.006)	0.336*** (0.007)	-0.003 (0.006)

Note: Estimated coefficients from the OLS regression model specified in [Equation 1](#), showing the associations between polygenic indices (PGIs), parental socioeconomic status (SES), and their interaction (PGI \times SES) across 45 phenotypes, for the mega-analysis pooling HRS, WLS, and ELSA. Standard errors are reported in parentheses, and phenotypes are sorted in ascending order of the estimated parental SES coefficient, matching the ordering of [Figure 1](#). Control variables include an indicator for each dataset, 10 ancestry-specific principal components, age and age squared, sex, region of birth, and interactions between age and sex. Significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

B Phenotype selection

B.1 GWAS catalog and criteria for selection

To select the most relevant and powerful GWASs for this study we used the following methodology. First, we consulted the GWAS Catalog (<https://www.ebi.ac.uk/gwas/downloads>; version of 30 May 2019)⁷³, which contained 776 GWASs. Second, we divided all GWASs into 11 major trait families based on the literature⁷⁴. Third, we used the following criteria to filter out GWASs not suited for our study:

- Low genome-wide significant SNP count (Association count < 6).
- More relevant trait available.
- Low prevalence.
- Not in HRS.
- Non-European sample.
- Larger GWAS (for this trait) exists.
- Other study/measure used.

We made the filter steps quantifiable when possible. However, in determining the relevance of a specific trait to our study we had to rely on subjective assessment. We labeled a trait as “More relevant GWAS available” when, for instance, instead of “coronary artery disease, type I diabetes mellitus” we used the main “coronary artery disease” and main “type I diabetes” phenotype GWAS.

Given that we had three datasets to test the PGI in, the trait also had to be present in those datasets. Therefore, for this initial filtering we filtered out traits when they were not in the HRS, by labeling them “Not in HRS”. This filter step was based on a Google search of the terms “Health and Retirement Study” in combination with the trait.

B.2 Selection methodology in the GWAS Catalog

After the first filtering step leaving only potentially relevant GWASs, we conducted a rigorous selection process. All remaining GWASs were subjected to a comparison within their respective trait families. Generally, we included the GWASs of a unique trait with the largest sample size within a given trait domain. For instance, we included the largest GWASs within the domain ‘Metabolic’, which were, among others, diabetes type 2 and BMI. The power of a GWAS was based on the sample size and the GWAS association count. Furthermore, we included GWASs with lower GWAS association counts if these GWASs provided interesting phenotypic information beyond the information provided by the most powerful GWASs within a trait family. For example, we included the phenotype ‘obesity’ because we assessed that this phenotype provides information on top of what ‘BMI’ provides.

Furthermore, since we also consulted the GWAS Atlas (discussed in the next section), we also assessed whether the selected GWAS from the GWAS Catalog was a better choice than those available from the GWAS Atlas.

B.3 Selection methodology in the GWAS Atlas

We consulted the GWAS Atlas (Watanabe et al. (2019)⁷⁴; <https://atlas.ctglab.nl/>, version of 17 April 2019). Regarding filter and selection steps, we used similar principles as described in the previous sections. However, since the GWAS Atlas provides more detailed information regarding the GWASs, we were able to apply more rigorous filtering steps. The filtering criteria for the GWAS Atlas were:

- Number of SNPs > 450,000.
- SNP-based heritability Z-score > 2.
- Contains EUR population.
- Low genome-wide significant SNP count (Association count < 6).
- Selected nonbinary traits with >50,000 European individuals with non-missing phenotypes; selected binary traits for which available cases and controls were both >10,000 and the total sample size was >50,000.

Similar to the selection process in the GWAS Catalog, we selected the largest and most relevant GWASs of the trait domains as indicated by the GWAS Atlas⁷⁴. Finally, we compared the selected GWASs from the GWAS Atlas with those of the GWAS Catalog. The final choice regarding inclusion was documented and based on the sample size of the GWAS and its relevance, and assessed independently by two researchers.

Note that, in addition to published GWASs, the GWAS Atlas also contained GWASs specifically conducted for the GWAS Atlas project⁷⁴. We prioritized previously published GWASs and only included GWAS-Atlas-specific GWASs if the phenotype was important to our study and there was no equivalent previously published GWAS available in the GWAS Catalog or GWAS Atlas.

B.4 Processing of selected GWASs

Subsequently, summary statistics of the GWASs selected from the GWAS Atlas and GWAS Catalog were downloaded. The summary statistics that had no information regarding the number of SNPs and SNP-h2 Z-score were put through a final round of filtering using LDSC⁷⁷, to ensure that only well-powered GWASs were included with a sufficient number of SNPs. Next, we included only those phenotypes that are available in the HRS, ELSA, and WLS.

After downloading and reviewing the summary statistics and reading the papers of the selected GWASs, we made the following minor adaptations:

- We revised the selection of phenotypes to improve consistency and data availability. Specifically, the social interaction phenotype was replaced with a loneliness phenotype derived from the same study, as both summary statistics and corresponding variables are available.
- The phenotype subjective well-being was replaced by four related constructs—positive affect, neuroticism, life satisfaction, and depressive symptoms—based on the multivariate GWAS by Baselmans et al. (2019)⁷⁸.
- We excluded the phenotypes unipolar depression and Major Depressive Disorder 2 (MDD2) from Wray et al. (2018)⁷⁹, as more recent and comprehensive GWAS results are available^{78,80}.
- For parental longevity, we retained only the combined parental age at death (z-scored). Alternative measures (e.g., father’s age at death, mother’s age at death, and indicators of extreme longevity) were excluded due to redundancy. The combined measure was preferred because it maximizes sample size (by including individuals with only one deceased parent) and exhibits the highest explanatory power (R^2).
- The phenotype extreme/severe obesity was excluded due to duplication with the obese vs. thin phenotype, as both definitions are effectively equivalent. Additionally, a continuous BMI measure is already included.
- For general risk-taking behavior, we use the more recent GWAS summary statistics from Karlsson Linnér et al. (2019)⁸¹.
- For age at first birth, we use the pooled GWAS summary statistics rather than sex-specific estimates, as this maximizes sample size and maintains consistency with our main analyses, which are not stratified by gender.
- Regarding subjective well-being and its related constructs (e.g., neuroticism and life satisfaction), we rely on multivariate GWAS results rather than univariate summary statistics.
- We established the following criteria for selecting variables in the Health and Retirement Study (HRS):
 - For general intelligence, we prioritize the cognition score (cognitionScore) over the Mental Status Index (mentalStatus) and Word Recall Index (wordRecall), provided that missingness is not excessive.
- When exact phenotype matches are unavailable in HRS, we use proxy variables:
 - General risk-taking behavior: we use behavioral measures (e.g., gambling games), if available, rather than self-reported risk preferences, due to closer conceptual alignment with GWAS measures.
 - Sleep-related traits: GWAS measures are based on sleep duration (hours slept), which is not directly available in HRS. Instead, we use survey questions capturing early awakening and

perceived restfulness. However, given the availability of an insomnia variable, these proxies are excluded from the final analysis.

- Physical activity: GWAS measures of moderate-to-vigorous physical activity (MVPA) are based on weighted weekly minutes of activity. In HRS, we construct a binary indicator (actModVig) equal to 1 if respondents engage in moderate or vigorous physical activity at least once per week, and 0 otherwise.
- As a general rule for selecting GWAS results, we prioritize multivariate methods such as MTAG when available, as they increase statistical power while distinguishing between related phenotypes. However, we do not use Genomic SEM or similar approaches that produce latent-factor GWAS, in order to maintain interpretability of individual phenotypes.

In total, our final selection includes 45 GWASs. A complete overview of the included studies is provided in Supplementary Table B.1.

Table B.1: GWAS references for phenotypes used in the analysis

Phenotype	Reference
Age at First Birth	Barban et al. (2016) ⁸²
Age start smoke	Liu et al. (2019) ⁸³
Alzheimer	Jansen et al. (2019) ⁸⁴
Anxiety	Otowa et al. (2016) ⁸⁵
Arthritis	Zengini et al. (2018) ⁸⁶
Asthma	Demenaïs et al. (2018) ⁸⁷
BMI	Locke et al. (2015) ⁸⁸
Breast Cancer	Michailidou et al. (2017) ⁸⁹
Cancer	Neale Lab UKB (2018) ⁹⁰
Cataract	Neale Lab UKB (2018) ⁹⁰
Cholesterol	Hoffmann et al. (2018) ⁹¹
Cognition	Savage et al. (2018) ⁹²
Combined parental age	Deelen et al. (2019) ⁹³
Coronary Artery Disease (CAD)	Nelson et al. (2017) ⁹⁴
Depressive Symptoms	Baselmans et al. (2019) ⁷⁸
Diabetes Type I	Onengut-Gumuscu et al. (2015) ⁹⁵
Diabetes Type II	Mahajan et al. (2018) ⁹⁶
Drinks per Week	Liu et al. (2019) ⁸³
Education	Lee et al. (2018) ³³
Feeling Worry	Otowa et al. (2016) ⁸⁵

Continued on next page

Table B.1 – *Continued from previous page*

Phenotype	Reference
Health rating	Neale Lab UKB (2018) ⁹⁰
Hearing difficulty	Neale Lab UKB (2018) ⁹⁰
Height	Wood et al. (2014) ⁹⁷
High blood pressure	Neale Lab UKB (2018) ⁹⁰
Household Income	Neale Lab UKB (2018) ⁹⁰
Insomnia symptoms	Jansen et al. (2019) ⁹⁸
Life Satisfaction	Baselmans et al. (2019) ⁷⁸
Loneliness	Day et al. (2018) ⁹⁹
Max cig per day	Liu et al. (2019) ⁸³
Neuroticism Score	Baselmans et al. (2019) ⁷⁸
Non-cancer illnesses	Neale Lab UKB (2018) ⁹⁰
Number of children fathered (male)	Barban et al. (2016) ⁸²
Number of live births (female)	Neale Lab UKB (2018) ⁹⁰
Physical activity	Klimentidis et al. (2018) ¹⁰⁰
Positive Affect	Baselmans et al. (2019) ⁷⁸
Prostate Cancer	Schumacher et al. (2018) ¹⁰¹
Prospective memory test	Neale Lab UKB (2018) ⁹⁰
Risk-taking behavior	Karlsson Linnér et al. (2019) ⁸¹
Severe Obesity	Locke et al. (2015) ⁸⁸
Smoke cessation	Liu et al. (2019) ⁸³
Smoke initiation	Liu et al. (2019) ⁸³
Stroke	Malik et al. (2018) ¹⁰²
Treatments / medications taken	Neale Lab UKB (2018) ⁹⁰
Unipolar Depression	Wray et al. (2018) ⁷⁹
Well-Being Spectrum	Baselmans et al. (2019) ⁷⁸

C Data Cleaning: HRS

We constructed most phenotypes using RAND HRS Longitudinal files. RAND HRS Longitudinal Files collect, clean, and rename variables from Core and Exit Interviews of the HRS. Other variables were extracted directly from the yearly study datasets. In this case, the variables used to create the phenotypes are not necessarily present in all rounds.

Some general data cleaning was required on HRS variables. Afterwards, we merged a panel with all waves through a unique identifier.

C.1 Height (`height`)

The height variable is present in all rounds and is part of the RAND files (`rheight`).

C.2 Body Mass Index (`bmi`)

We extract the BMI variable directly from the RAND files. Values over 70 were set to missing, since they are considered to be data entry errors.

C.3 Obese vs. thin (`obesitySevere`)

We created the extreme obesity variable based on the definition from the UKB GWAS. Severe obesity is a dummy variable equal to 1 if BMI is greater than or equal to 40 kg/m^2 , and 0 otherwise.

C.4 General intelligence latent factor (`cognitionScore`)

The total cognition score is calculated by HRS as the sum of the Total Recall Index and the Mental Status Index.

The total recall index (`RwTR20`, `R1TR40`, `RwATR20`, `RwHTR40`) is available in all waves and summarizes the immediate and delayed word recall tasks. In Waves 1 and 2H, the recall wordlist contained 20 words, while in all other waves it contained 10. Thus the scores range from 0 to 40 in Waves 1 and 2H and from 0 to 20 in other waves.

The mental status index (`RwMSTOT`, `R2AMSTOT`) sums scores from counting, naming, and vocabulary tasks and is available for Wave 2A and from Wave 3 forward. This reflects the absence of some of these tests in Waves 1 and 2H.

A total cognition score (`RwCOGTOT`, `R2ACGTOT`) sums the total recall and mental status indices. Because the mental status index is missing for Waves 1 and 2H, the total cognition index is also missing for these waves. The total cognition score ranges from 0 to 35.

Further detail on these variables can be found in the RAND HRS Longitudinal File Codebook.

C.5 Life satisfaction (`lifeSatisfaction`)

Life satisfaction variables are not included in RAND files and were extracted directly from the leave-behind section from 2004 onwards. Variables extracted are **LB003A–LB003E**.

We constructed the life satisfaction variable based on ¹⁰³ by averaging the scores across all 5 items from the life satisfaction section. We set the final score to missing if there are three or more items with missing values, as indicated in the document.

We recoded 2006 variables since for that year the response scale was 6-point and changed to a 7-point scale thereafter.

C.6 Positive affect (positiveAffect)

Positive affect variables are not included in RAND files and were extracted directly from the leave-behind section from 2006 onwards. Variables extracted are **LB027A–LB027Y**.

We constructed the positive affect variable based on¹⁰³ by reverse-coding items c, d, f, g, h, k, p, q, t, u, v, x, and y, and averaging the scores across all 13 items. We set the final score to missing if there are more than six items with missing values.

C.7 Neuroticism (neuroticismScore)

Calculation of the neuroticism score is based on¹⁰⁴ and¹⁰³.

Neuroticism variables are not included in RAND files and were extracted directly from the leave-behind section from 2006 onwards. Variables extracted are **LB033A–LB033Z**.

We calculate neuroticism by reverse-coding questions d, h, l, q, and averaging the scores across the 4 items. We set the final score to missing if more than half of the items have missing values within each sub-dimension.

C.8 Depressive symptoms (depressScore)

For depressive symptoms we use RAND variable **RwCESD**.

The CESD score (**RwCESD**) is the sum of five “negative” indicators minus two “positive” indicators. The negative indicators measure whether the Respondent experienced the following sentiments all or most of the time: depression, everything is an effort, sleep is restless, felt alone, felt sad, and could not get going. The positive indicators measure whether the Respondent felt happy and enjoyed life all or most of the time. The corresponding RAND variable is **rcesd**.

C.9 Worry measurement, neuroticism factor (worryFeeling)

We constructed feeling worry from the question “Please indicate how well each of the following describes you: Worrying”, with answers “1. A LOT, 2. SOME, 3. A LITTLE, 4. NOT AT ALL”.

C.10 General risk-taking behavior (risk)

We calculated risk aversion based on¹⁰⁵, supplemental material section. Hypothetical gambles on lifetime income are available for waves 1992 and 1994, and 1998 to 2006. The questions change through rounds, therefore we created 3 measures that can be merged.

The HRS has fielded two versions, the first or “original” version and the revised or “status-quo bias free” version. The original questions begin:

- “Suppose that you are the only income earner in the family, and you have a good job guaranteed to give you your current (family) income every year for life. You are given the opportunity to take a new and equally good job, with a 50-50 chance it will double your (family) income and a 50-50 chance that it will cut your (family) income by a third. Would you take the new job?”

The revised questions (bias-free) begin:

- “Suppose that you are the only income earner in the family. Your doctor recommends that you move because of allergies, and you have to choose between two possible jobs. The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50-50 chance the second job would double your total lifetime income and a 50-50 chance that it would cut it by a third. Which job would you take—the first job or the second job?”

rc4_o1: Wave 1992 had only 2 gambles on the question, therefore only 4 categories are possible. Categories are defined from high risk aversion (rejects all gambles) to low risk aversion (accepts all gambles). This wave is susceptible to “status quo” bias.

rc6_f: Wave 1994 had 4 gambles on the question, therefore 6 categories are possible. Categories are defined from high risk aversion (rejects all gambles) to low risk aversion (accepts all gambles). This wave is susceptible to “status quo” bias.

rc6_f: From waves 1998 to 2006, the question had 4 gambles, therefore 6 categories are possible. Categories are defined from high risk aversion (rejects all gambles) to low risk aversion (accepts all gambles). This wave is free of “status quo” bias.

C.11 Loneliness (loneliness)

We measured loneliness based on ¹⁰³. HRS incorporates a 3-item and an 11-item scale of loneliness derived from the 20-item Revised UCLA Loneliness Scale.

We construct the loneliness measure by reverse-coding questions a, b, c, and e. Then we averaged the scores across all 11 items. We set the final score to missing if more than five items have missing values.

To create the original 3-item loneliness index, we reverse-code items a, b, c, and create an average of these three scores. We set the final score to missing if more than 1 item is missing.

C.12 Insomnia symptoms (insomniaFrequent)

Insomnia questions are present from 2002 to 2016. To create the variable for insomnia symptoms, we used the questions “How often have trouble with waking up during the night?” and “How often have trouble falling asleep?”. Possible answers were “1. Most of the time, 2. Sometimes, 3. Rarely or never”.

Variable **insomnia** is coded 1 if the person responded “Most of the time” or “Sometimes” to one of the two questions.

Variable **insomniaFrequent** is coded 1 if the person responded “Most of the time” to one of the two questions. Our selected variable is **insomniaFrequent**.

HRS variables 2002–2016:

(H, J, K, L, M, N, O, P)083: How often do you have trouble falling asleep—would you say most of the time, sometimes, or rarely or never?

(H, J, K, L, M, N, O, P)084: How often do you have trouble with waking up during the night—would you say most of the time, sometimes, or rarely or never?

C.13 Unipolar depression (**depress**)

We constructed unipolar depression from the question “Felt depressed” from the CES-D questionnaire. RAND variables are **rdepres** for wave 1 and **rdepres** for all other waves.

C.14 Anxiety (**anxietyScore**)

Five items were selected from the widely used Beck Anxiety Inventory (BAI). The Beck Inventory has been shown to distinguish symptoms of anxiety from depression and to be valid for use in older populations. This scale was not included after 2012.

Table C.1: 5 items (Q41a–Q41e)

Q41a	I had fear of the worst happening.
Q41b	I was nervous.
Q41c	I felt my hands trembling.
Q41d	I had a fear of dying.
Q41e	I felt faint.

Scaling: Responses to the 5 items are averaged to form an index of anxiety ranging from 1–4 (1: never feeling x to 4: most of the time feeling). Set the final score to missing if more than two of the items have missing values.

C.15 Alcohol consumption, drinks per week (**drinkWeek**)

RAND variables used to calculate the number of drinks per week are: drinks per day (**rdrinkn** and **rdrinkr**) and number of days per week drinking (**rdrinkd**).

C.16 Smoking initiation (**ageSmoke**)

Age at smoking initiation is constructed from the question “How old were you when you first started smoking cigarettes?”, taking the earliest reported age across all available waves.

C.17 Smoking cessation (`cesSmoke_GSCAN`)

Following the phenotype definition used in the GSCAN consortium¹⁰⁶, smoking cessation is coded as 1 for former smokers and 0 for current smokers, restricting the sample to ever-smokers. Constructed from RAND variable `rsmoken`, which indicates whether the Respondent currently smokes.

C.18 Number of cigarettes smoked per day (`cigsAll_Current`)

This variable was constructed from individual file data using the questions “NUM CIGARETTES SMOKED PER DAY” and “NUM CIGS PER DAY—WHEN SMOKED MOST”.

Table C.2: Variables from individual files HRS

Year	Num cig smoked per day	Num cig smoked per day, when smoked most
1992	V503	V505
1994	W453	
1996	E943	
1998	F1268	F1275
2000	F1401	G1408
2002	HC118	HC123
2004	JC118	JC123
2006	KC118	KC123
2008	LC118	LC123
2010	MC118	MC123
2012	NC118	NC123
2014	OC118	OC123

C.19 Physical activity measurement (`actModVig`)

RAND data on physical activity start from 2006 to 2016. We have created the variable `actModVig` that takes value 1 for those who do moderate or vigorous physical activity once a week or more often, and 0 otherwise.

Vigorous physical activity: “We would like to know the type and amount of physical activity involved in your daily life. How often do you take part in sports or activities that are vigorous, such as running or jogging, swimming, cycling, aerobics or gym workout, tennis, or digging with a spade or shovel: more than once a week, once a week, one to three times a month, or hardly ever or never?”

Moderate physical activity: “And how often do you take part in sports or activities that are moderately energetic, such as gardening, cleaning the car, walking at a moderate pace, dancing, floor or stretching exercises: more than once a week, once a week, one to three times a month, or hardly ever or never?”

C.20 Parental longevity (`ageParentsZ`)

Variables for parental longevity were created from `rmomage`, `rdadage`, `rmomliv`, `rdadliv`. We took the age of alive parents or at parent’s death.

The 1% oldest cut-off for HRS was 99 for women and 96 for men.

C.21 Osteoarthritis (arthritis)

Osteoarthritis is a variable from RAND files. RAND files contain the questions “R reports arthritis/rheumatism this wave” and “Whether or not a doctor has ever told the Respondent s/he had arthritis or rheumatism” (`rarthr`, `rarthre`).

C.22 Type II diabetes (diabetes)

Type II diabetes is a variable from RAND files. RAND files contain the questions “R reports diabetes this wave” and “Respondent has ever reported having diabetes” (`rdiab`, `rdiabe`).

C.23 Educational attainment (`educYears`)

Educational attainment in HRS is constructed from the question “R Years of Education” (`raedyrs`).

C.24 Age at first birth, AFB (`ageFirstBirth`)

Constructed using `HwAGEOKID` variables from RAND Family data files, renamed as `ageOldch`, the age of the Respondent’s oldest child. This variable is derived from the best-guess child’s age (`KwAGEBG`) in the Respondent-kid file. Some of the ages are over 80 years old; these ages are based on the reported birth year.

Some cleaning was required for this variable since there were cases where the age of the oldest child was greater than the age of the respondent. These cases were coded as missing values.

Table C.3: Inconsistent values in oldest child’s age

	Freq.	Percent	Cum.
Respondent is older	485,244	99.84	99.84
Child is older	780	0.16	100.00
Total	486,024	100.00	

We found cases where the difference between the respondent and the oldest child is less than 5 years or greater than 60. These inconsistencies depend on the reported age of the oldest child based on the

respondent’s best guess. In addition, we coded as missing those cases where the age at first birth is over 60 years old for women.

Table C.4: Age at first birth from RAND family data

Variable	Obs	Mean	Std. Dev.	Min	Max
ageFirstBirth	207,782	23.23	6.22	0	79

To account for these inconsistent values in the age at first birth (**ageFirstBirth**), the variable **FlagAFB** flags values under 5 for the full sample and over 60 for women, where childbearing is extremely unlikely. We replace these flagged values with missing.

C.25 Total cholesterol measurement (**totChol**)

The total cholesterol variable was extracted from the 2016 Venous Blood Study. The variable is measured in mg/dl.

C.26 Breast carcinoma (**cancerBreast**)

This variable is included only in years 1992 and 1994. The variable is coded as 1 if breast cancer was diagnosed since the last round (yes/no).

C.27 Prostate carcinoma (**cancerProstate**)

This variable is included only in years 1992 and 1994. The variable is coded as 1 if prostate cancer was diagnosed since the last round (yes/no).

D Data Cleaning: WLS

WLS private data contain 30 duplicated observations. Of these, 28 observations are pairs of identical twins. We delete one observation per twin pair. We merge genetic data by unique ID and keep a sample of White respondents.

D.1 Height (**height**)

Measures of height in WLS are self-reported for rounds 4 to 6. Round 6 also includes a variable for height measured in person, which we use. Height is measured in inches, and we convert it to centimeters to match the GWAS definition.

D.2 General intelligence latent factor (cognitionScore)

There are multiple IQ measures, but `gwiiq_bm` is the WLS recommended one if only one measure is to be used for WLS Graduates.

Cognitive ability (IQ) was measured in the freshman and junior year of high school (1956) using the Henmon–Nelson test of mental ability. The Henmon–Nelson test is a group-administered, 30-minute, multiple-choice assessment that consists of 90 verbal or quantitative items. It was administered in all Wisconsin high schools at various grade levels from the 1930s through the 1960s as part of a cooperative effort of high schools and colleges to identify youth who might succeed in college but did not originally plan to attend college.

Instead of using the best measure recommended by WLS, some studies also use the raw score. Raw Henmon–Nelson test scores were converted to IQ scores by standardizing to a mean of 100 based on centile rank. IQ scores ranged from 61 to 145; for an example see ¹⁰⁷.

There are other variables that measure some form of cognition. In waves 1992, 2004, and 2011, WLS included a subset of the fourteen items from the Wechsler Adult Intelligence Scale (WAIS). In wave 2004, cognition was measured with a fluency section and a word recall section.

Table D.1: Cognition variables: Wechsler Adult Intelligence Scale (WAIS)

<code>ri001re</code>	R4 total cognition score
<code>z_gi101re</code>	R5 6-item score for cognition
<code>z_gi106re</code>	R5 9-item score for cognition
<code>z_hi101re</code>	R6 6-item score summarizing R's performance on selected WAIS questions

D.3 Life satisfaction (lifeSatisfaction)

There is no direct measure of life satisfaction in WLS. The closest measure is psychological well-being, captured through eudaimonia. Questions are asked about six different aspects of life (autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance), with the answers combined to form a composite eudaimonic measure. Further explanation on how to construct this variable can be found in ¹⁰⁸. The continuous variable was constructed from 12 items in 1992–1993 and 18 items in 2003–2005.

D.4 Positive affect (positiveAffect)

There is no direct measure of positive affect in the WLS questionnaire. However, WLS measures emotion level based on the Health Utilities Index (HUI). This variable is constructed from 3 items:

D.5 Neuroticism (neuroticismScore)

We create neuroticism as described in the scales documentation of WLS ¹⁰⁸.

Table D.2: Psychological well-being: 12-item

z_rn016rec	I am influenced by people with strong opinions
z_rn017rec	I am in charge of how I live
z_rn018rec	It is hard to maintain close relations
z_rn019rec	I do not wander aimlessly through life
z_rn020rec	I am pleased with how things turned out
z_rn021rec	Demands of everyday life often get me down
z_rn022rec	Life is a continuous process of learning, changing, growing
z_rn023rec	Have not experienced many warm, trusting relationships
z_rn024rec	Live life one day at a time, do not think about the future
z_rn025rec	I judge myself by what I think is important
z_rn026rec	Gave up trying to make big improvements/changes in life
z_rn027rec	I like most aspects of my personality

Table D.3: Psychological well-being: 18-item

gn116re	I tend to be influenced by people with strong opinions.
gn125re	I judge myself by what I think is important, not by what others think.
gn128re	I have confidence in my own opinions even if they are contrary to general consensus.
gn117re	In general, I feel I am in charge of the situation in which I live.
gn121re	The demands of everyday life often get me down.
gn129re	I am quite good at managing the many responsibilities of my daily life.
gn122re	For me, life has been a continuous process of learning, changing, and growing.
gn126re	I gave up trying to make big improvements or changes in my life a long time ago.
gn130re	It is important to have new experiences that challenge how I think about myself/world.
gn118re	Maintaining close relationships has been difficult and frustrating for me.
gn123re	I have not experienced many warm and trusting relationships with others.
gn131re	People would describe me as a giving person, willing to share my time with others.
gn119re	Some people wander aimlessly through life, but I am not one of them.
gn124re	I live life one day at a time and do not really think about the future.
gn132re	I sometimes feel as if I have done all there is to do in life.
gn120re	When I look at the story of my life, I am pleased with how things have turned out.
gn127re	I like most aspects of my personality.
gn133re	In many ways, I feel disappointed about my achievements in life.

Table D.4: Emotion score questions

z_gx330re	During the past 4 weeks, have you been feeling happy or unhappy?
z_gx331re	Felt happy and interested in life, or somewhat happy?
z_gx332re	Felt somewhat unhappy, very unhappy, or so unhappy life not worthwhile?

Table D.5: Neuroticism variables

z_mh025rec	R4 measure of neuroticism
z_ih025rec	R5 summary score for neuroticism
z_jh025rec	R6 summary score for neuroticism

D.6 Depressive symptoms (depressScore)

The methodology to calculate the depression score with a 20-item scale is described in [109](#). For this scale we used variables `z_mu001rec`, `z_iu001rec`, and `z_ju001rec`.

Table D.6: 20-item depression variables

<code>z_mu001rec</code>	Summary score for psychological distress/depression—modified CES-D
<code>z_iu001rec</code>	Summary score for psychological distress/depression—modified CES-D
<code>z_ju001rec</code>	Summary score for psychological distress/depression—modified CES-D

To replicate the 8-item score used in other surveys, we used the following questions:

Table D.7: 8-item depression score

<code>z_ju013rer</code>	R6 On how many days during the past week did you feel depressed?
<code>z_ju012rer</code>	R6 On how many days during the past week did you feel sad?
<code>z_ju008rer</code>	R6 On how many days during the past week did you feel lonely?
<code>z_ju017rer</code>	R6 How many days in the past week felt that everything you did was an effort?
<code>z_ju020rer</code>	R6 On how many days during the past week did you sleep restlessly?
<code>z_ju022rer</code>	R6 On how many days during the past week did you feel you could not ‘get going’?
<code>z_ju006rer</code>	R6 On how many days during the past week did you feel happy?
<code>z_ju009rer</code>	R6 On how many days during the past week did you enjoy life?

The preferred measure is the 20-item score.

D.7 Worry measurement, neuroticism factor (worryFeeling)

Feeling worry was constructed from the question “On how many days in the past week did you worry over possible misfortune?”.

Table D.8: Feeling worry questions

<code>nu041rer</code>	R4 days past week worry over possible misfortune
<code>z_iu041rer</code>	R5 On how many days during the past week did you worry over possible misfortune?
<code>z_ju041rer</code>	R6 On how many days in the past week did you worry over possible misfortune?

D.8 General risk-taking behavior (risk)

To measure risk we used the question “Importance of having a low risk of losing your job vs. high pay” (`z_mg004rer`, `z_ig004rer`, `z_jfin05re`, `z_jfin10re`).

D.9 Loneliness (loneliness)

Loneliness is measured with the question “In a week how lonely are you (1–4)” (**z_ik016rer**). This is only asked in round 5.

D.10 Insomnia symptoms (insomniaFrequent)

We constructed **insomniaFrequent** from variables **mx020rer** and **z_ix020rer**, which ask: “How often have you had trouble sleeping in the past 6 months?”. **insomniaFrequent** is coded 1 if the person has sleeping problems more than once a week.

D.11 Unipolar depression (depress)

We constructed unipolar depression based on the question “Did R ever feel sad, blue, or depressed for two weeks or longer?” from variables **z_gu002re**, **z_ru002re**, and **z_au002re**.

D.12 Anxiety (anxietyScore)

As explained in [108](#), WLS uses the Spielberger State-Trait Anxiety Inventory (STAI) to assess anxiety. It can be used in clinical settings to diagnose anxiety and to distinguish it from depressive syndromes. It is also often used in research as an indicator of caregiver distress.

The index is constructed from survey questions:

Table D.9: Anxiety questions: On how many days during the past week did you...

IU035RER	“feel calm?”
IU037RER	“feel tense?”
IU039RER	“feel at ease?”
IU041RER	“worry over possible misfortune?”
IU043RER	“feel nervous?”
IU045RER	“feel jittery?”
IU047RER	“feel relaxed?”

The index was constructed by summing the valid values across the seven items if at least five items received a valid response. The variable was coded as -2 if fewer than five items received a valid response. Variables IU035RER, IU039RER, and IU047RER were reverse-coded in the creation of this variable.

D.13 Alcohol consumption, drinks per week (drinkWeek)

Drinks per week was constructed from the question “Total number of drinks in the last month” for rounds 4, 5, and 6 (variables **z_ru028re**, **z_gu028re**, **z_hu028re**). We divided the drinks in the last month by 30 and multiplied by 7. The final value is the maximum drinks per week across the three rounds.

D.14 Smoking initiation (ageSmoke)

Age at smoking initiation is constructed from the question “How old were you when you started smoking cigarettes regularly?”, asked in rounds 5 and 6. We take the minimum reported age across rounds.

D.15 Smoking cessation (cesSmoke_GSCAN)

Following the phenotype definition used in the GSCAN consortium¹⁰⁶, smoking cessation is coded as 1 for former smokers and 0 for current smokers, restricting the sample to ever-smokers. Constructed using current smoking status across rounds 4, 5, and 6 from variables `z_mx013rer`, `z_ix013rec`, and `z_jx013rec`.

D.16 Number of cigarettes smoked per day (cigsAll_Current)

Maximum cigarettes per day was constructed from the question “On average, how many packs of cigarettes do you smoke a day?” in rounds 4, 5, and 6 (variables `nx044red`, `z_ixt08rer`, `z_jxt08rer`). We use the maximum value across rounds.

D.17 Physical activity measurement (actModVig)

This survey asks only about light physical activity. The question is “Past year, how many hours per month did you do light physical activities with others?” for rounds 5 and 6 (variables `z_jz168rer`, `z_iz168rer`). We converted hours per month to minutes per week by multiplying by 60 and dividing by 4.3452 (the average number of weeks per month). We use the maximum value across rounds.

D.18 Parental longevity (ageParentsZ)

We calculated parental longevity from parent age (if alive) and parent’s age at death. We standardized parental age.

D.19 Osteoarthritis (arthritis)

We calculated this variable using the question “Has a medical professional ever said you have arthritis or rheumatism?”, present in rounds 4, 5, and 6 (variables `z_mx087rer`, `z_gx360re`, `z_jx211rer`). We code yes if the respondent has ever been told they have arthritis.

D.20 Type II diabetes (diabetes)

We calculated this variable using the question “Has a medical professional ever said you have diabetes?”, present in rounds 4, 5, and 6 (variables `z_mx095rer`, `z_gx342re`, `z_hx342re`). We code yes if the respondent has ever been told they have diabetes.

D.21 Educational attainment (`educYears`)

Years of education in WLS is constructed from `edhi`, indicating the highest level of education attained, converted into years using standard mappings (e.g., high school = 12, bachelor’s = 16, master’s = 18, doctoral = 21). For respondents with multiple reports across waves, we use the highest reported value.

D.22 Age at first birth, AFB (`ageFirstBirth`)

Age at first birth is asked only in round 3. If the age at first birth is missing, we replace it with a constructed age at first birth, calculated by subtracting the age of the oldest biological child from the age of the respondent.

D.23 Total cholesterol measurement (`totChol`)

There is no direct measure of cholesterol in WLS. We used the question “Has a medical professional ever said that you have high cholesterol?” to indicate high total cholesterol.

D.24 Breast carcinoma (`cancerBreast`)

Breast cancer is measured with the question “Has had breast cancer” (`z_gx350ere`). This question is present only in round 5.

D.25 Prostate carcinoma (`cancerProstate`)

Prostate cancer is measured with the question “Has had prostate cancer” (`z_gx351ere`). This question is present only in round 5.

E Data Cleaning: ELSA

E.1 Height (`height`)

Constructed from `htval`.

E.2 Obese vs. thin (`obesitySevere`)

We created the extreme obesity variable based on the definition from the UKB GWAS. Severe obesity is a dummy variable equal to 1 if BMI is greater than or equal to 40 kg/m^2 , and 0 otherwise.

E.3 Body Mass Index (bmi)

We extract the BMI measure directly from the survey (`bmival`). We replace extreme values (over 70 kg/m^2) with missing values.

Table E.1: Values over 70 BMI

over70	Freq.	Percent	Cum.
0	23,179	95.14	95.14
1	1,184	4.86	100.00
Total	24,363	100.00	

E.4 General intelligence latent factor (cognitionScore)

Based on ¹¹⁰, we construct general intelligence as the sum of 3 standardized cognition tests. These tests are verbal fluency, immediate and delayed verbal memory, and attention. Verbal (semantic) fluency was assessed by asking participants to name as many animals as they could think of in 1 minute. Immediate and delayed verbal memory was assessed by presenting a list of 10 nouns aurally on a computer, one every 2 seconds. Participants were asked to recall as many words as possible immediately and again after a short delay during which they carried out the other cognitive tests. Attention and mental speed were assessed using a letter cancellation task. Participants were given a clipboard to which was attached a page of 780 random letters of the alphabet set out in a grid of 26 rows and 30 columns, and were asked to cross out as many target letters (P and W) as possible in 1 minute.

The cognition score is the average of standardized verbal fluency, immediate and delayed verbal memory, and attention.

Table E.2: Questions on cognition

<code>cflisen</code>	Number of words recalled immediately
<code>cflisd</code>	Number of words recalled after delay
<code>cfani</code>	Number of animals mentioned (fluency)
<code>nrowcl</code>	Letter cancellation task: total number of letters searched ($30 \times (\text{nrow} - 1) + \text{nclm}$)

E.5 Life satisfaction (lifeSatisfaction)

Life satisfaction questions are included from wave 2 onwards.

We constructed the life satisfaction variable based on ¹¹¹ by averaging the scores across all 5 items from the life satisfaction section. We set the final score to missing if there are three or more items with missing values, as indicated in the document.

Table E.3: Satisfaction with life

sclifea	In most ways his/her life is close to his/her ideal
sclifeb	The conditions of his/her life are excellent
sclifec	Is satisfied with his/her life
sclifed	So far, he/she has got the important things he/she wants in life
sclifee	If could live his/her life again, would change almost nothing

E.6 Positive affect (positiveAffect)

Positive affect variables are included in the main questionnaire only for wave 5. Variables from the positive affect section are **scfede**, **scfeen**, **scfeac**, **scfepr**, **scfeint**, **scfeha**, **scfeat**, **scfeco**, **scfeins**, **scfeho**, **scfeal**, **scfeca**, and **scfeex**.

We constructed the positive affect variable based on ⁵⁵ by reverse-coding all questions and averaging the scores across all 13 items. We set the final score to missing if there are more than six items with missing values.

E.7 Neuroticism (neuroticismScore)

Calculation of the neuroticism score is based on ¹⁰⁴. Personality traits questions are:

Table E.4: Personality traits questions

How well does the following describe the respondent?	
scdewa	warm
scdewo	worrying
scdere	responsible
scdeli	lively
scdeca	caring
scdene	nervous
scdecr	creative
scdeha	hardworking
scdeim	imaginative
scdesof	soft-hearted
scdecal	calm
scdein	intelligent
scdecu	curious
scdeac	active
scdecar	careless
scdebr	broad-minded
scdesy	sympathetic
scdeta	talkative
scdeso	sophisticated
scdead	adventurous
scdeth	thorough

For the neuroticism variable we used **scdemo**, **scdewo**, **scdene**, **scdecal**.

E.8 Depressive symptoms (**depressScore**)

To construct the depression score we use¹¹² and¹¹³. We construct the variable based on the methodology presented for HRS, since the 8-item scale is the same as the one used in HRS by¹¹⁴.

The CESD score (**depressScore**) is the sum of five “negative” indicators minus two “positive” indicators. The negative indicators measure whether the Respondent experienced the following sentiments all or most of the time: depression, everything is an effort, sleep is restless, felt alone, felt sad, and could not get going. The positive indicators measure whether the Respondent felt happy and enjoyed life all or most of the time. If more than half of the questions are missing, the score is set to missing.

Table E.5: Variables used for depression

psceda	Whether felt depressed much of the time during past week
pscedb	Whether felt everything they did during past week was an effort
pscedc	Whether felt their sleep was restless during past week
pscedd	Whether was happy much of the time during past week
pscede	Whether felt lonely much of the time during past week
pscedf	Whether enjoyed life much of the time during past week
pscedg	Whether felt sad much of the time during past week
pscedh	Whether could not get going much of the time during past week

E.9 Worry measurement, neuroticism factor (**worryFeeling**)

We constructed feeling worry based on the question **scdewo** from the personality trait questionnaire.

E.10 General risk-taking behavior (**risk**)

The risk module is only included in wave 5. Question **ririsk** asks “Whether prepared to take risks on a scale of 0 to 10”.

E.11 Loneliness (**loneliness**)

ELSA measures loneliness with a four-item scale¹¹⁵, which is based on the widely-used 20-item Revised UCLA Loneliness Scale. A fifth item was added in round 3.

We created both a 4-item and a 5-item loneliness scale. We construct the loneliness measure by reverse-coding questions a, b, c, and e, then averaging the scores across all 4 or 5 items. We set the final score to missing if more than 2 items are missing.

In addition, we added a variable called **lonely** that asks: “How often respondent feels lonely”.

E.12 Insomnia symptoms (**insomniaFrequent**)

We created insomnia symptoms from the questions “Sleep: how often respondent has difficulty falling asleep” (**heslpa**) and “Sleep: frequency wake up several times at night” (**heslpb**). Frequent insomnia is a dummy variable equal to 1 if the person had difficulty falling asleep and woke up several times a night more than once a week, and 0 otherwise.

E.13 Unipolar depression (**depress**)

Unipolar depression is constructed from the question “Psychiatric problem has: depression” (**hepsyde**).

E.14 Anxiety (**anxietyScore**)

We constructed anxiety from the question **hepsy**an for round 3 onward and from questions **hepsy1**, **hepsy2**, **hepsy3**, **hepsy4**, **hepsy5**, **hepsy6** for rounds 1 and 2.

E.15 Parental longevity (**ageParentsZ**)

Variables for parental longevity were created from **dianm**, **dianf**, **dinma**, **dinfa**, **mthagd**, **ftthagd**, **dimad**, **difad**. We took the age of alive parents or at parent’s death. Ages under 10 years old were set to missing.

E.16 Osteoarthritis (**arthritis**)

Osteoarthritis is created from variables **heart1**, **heart2**, **heart3**, **heartoa**.

Table E.6: Arthritis questions

heart1	Which types of arthritis do you have? 1st
heart2	Which types of arthritis do you have? 2nd
heart3	Which types of arthritis do you have? 3rd
heartoa	Whether has osteoarthritis

E.17 Type II diabetes (**diabetes**)

Diabetes is calculated from variables **dheacd**, **heacd**, “Ever been told had diabetes (from feed or doctor)”.

E.18 Educational attainment (**educYears**)

Years of education is calculated from the variable “age finished continuous full-time education, merged from current and previous waves” (**edend**).

E.19 Age at first birth, AFB (`ageFirstBirth`)

We calculated age at first birth by subtracting the age of the oldest natural son or daughter from the reported age of the respondent at the moment of the survey.

Age of the oldest natural child: We calculated the age of the oldest child by taking the maximum age across reported ages (`chage1–chage16`) for natural children only (`chtype1–chtype16`). However, the age of the oldest child appears inconsistent in some cases.

Age at first birth: Age at first birth is calculated as the difference in age between the respondent and the oldest child in each round. We found some inconsistencies in the data, including negative ages. To account for these inconsistent values, the variable **FlagAFB** flags ages under 5 for the full sample and over 60 for women, where childbearing is extremely unlikely. We replace these flagged values with missing.

Table E.7: Age at first birth (`ageFirstBirth`)

Variable	Obs	Mean	Std. Dev.	Min	Max
<code>ageFirstBirth</code>	58,893	26.64	5.55	5	66

E.20 Number of children ever born, NEB (`childrenEverBorn`)

Number of children ever born is calculated by counting all natural children of the respondent (`chtype1–chtype16`) in each round and taking the maximum value across rounds.

Table E.8: Children ever born

Variable	Obs	Mean	Std. Dev.	Min	Max
<code>childrenEverBorn</code>	7,569	2.07	1.40	0	13

E.21 Total cholesterol measurement (`totChol`)

Only available for years 2004, 2008, and 2012.

E.22 Prostate carcinoma (`cancerProstate`)

“Ever had prostate cancer” is available from 2006 onward.

E.23 Smoking variables

All the smoking variables used in the analysis are self-reported:

1. Max cigarettes per day (*maxCPD*): Maximum weighted average of cigarettes per day across all waves. Observations below the 1st percentile are set to the 1st percentile and observations above the 99th percentile are set to the 99th percentile.
2. Ever smoker (*everSmoker*): If the respondent smoked cigarettes at one point in life.
3. Former smoker (*formerSmoker*): The last non-missing answer the respondent gave was used and categorized either into former smoker or current/never smoker.
4. Age started smoking (*ageStartSmoke*): Year the respondent started smoking daily.

F Quality Control

We performed extensive quality control procedures on each of our genetic datasets, closely following the steps and thresholds recommended by Marees et al. (2018)⁷⁶.

F.1 SNP filtering

We begin by removing variants with low imputation-quality metrics. When available, we remove all variants with an INFO score (generated by the imputation software IMPUTE2¹¹⁶) below 0.7, using the threshold recommended in¹¹⁷. We then remove all variants which are missing for more than 20% of samples. After doing the equivalent for samples (described later in the section), we proceed with a more stringent threshold, removing all variants which are missing for more than 2% of samples. We remove all variants with a minor allele frequency (MAF) < 0.01 to avoid the inclusion of variants lacking sufficient power to identify phenotypic correlation. In order to eliminate potential genotyping errors, we exclude variants which deviate from Hardy–Weinberg Equilibrium (HWE) with a p -value $< 1 \times 10^{-6}$.

F.2 Sample filtering

After enforcing the first variant missingness threshold described earlier, we remove all samples missing more than 20% of remaining variants. We then remove all samples for which we find a mismatch between the sex assigned by the dataset and the sex indicated solely by the genotypic data. To do this, we calculate X-chromosome heterozygosity/homozygosity rates and remove individuals outlying from the cluster of female and male samples. In executing this procedure for HRS and WLS, we raised the threshold for X-chromosome homozygosity for females because the female cluster extended well beyond the recommended threshold of 0.2. After the variant filter process is complete, we remove all samples missing more than 2% of remaining variants.

We proceed with a heterozygosity rate filter to ensure sample quality and remove examples of possible inbreeding. Considering only a pruned subset of variants, we remove samples for which the heterozygosity rate is more than 3 standard deviations from the mean. Because this paper assumes that each dataset used

is composed of unrelated samples, we must also investigate the relatedness of subjects in each dataset. In the case of HRS¹¹⁸ and WLS¹⁰⁸, associated quality control reports provide a list of samples to remove in order to achieve a fully “unrelated” sample. In the case of ELSA, we follow a procedure roughly similar to that used in the HRS and WLS reports. We calculate the identity by descent (IBD) of all sample pairs and flag all pairs indicated to be second-degree relatives or closer, removing the individual in each pair with the lowest call rate (i.e. the highest rate of variant missingness).

The final step of the QC process is to ensure we include only those samples with genotypic data indicating European ancestry by following a detailed population stratification procedure. This process is necessary in order to strengthen the associations we are able to find in our data due to the wide variance of allele frequencies across subpopulations. For our population stratification procedure, we conduct a multidimensional scaling (MDS) analysis on a pruned set of SNPs. Next, we follow the same procedure with data from the 1000 Genomes (1KG) project¹¹⁹ on a set of SNPs overlapping with our dataset. With these two results, we can visualize the clusters of genetic similarity for each of our datasets, overlaid with that of the 1KG data, for which we know the ethnicity of the samples. We are thus able to isolate the individuals with European ancestry in each of our datasets by including only the individuals overlapping with the European ancestry cluster in the 1KG data.