Nature, nurture, and socioeconomic traits: New evidence from sib pairs and molecular genetic data*†

Gareth Markel[1*] Jonathan Pierre Beauchamp[2*] Rafael Ahlskog[3]
Joakim Coleman Ebeltoft[4] René Mõttus[5] Sven Oskarsson[6]
Uku Vainik[7] Eivind Ystrøm[8]

March 21, 2025

Abstract

A consequence of Mendel's First Law is that siblings' genetic relatedness varies randomly (with a mean of 50% and a standard deviation of $\sim$ 4%). We use molecular genetic data to compute the genetic relatedness of $\sim$80,000 sib pairs. We then compare the pairs' genetic relatedness to their similarity on 15 outcomes in the cognitive and educational, labor market, risk taking, health, and anthropometric domains, to estimate the relative importance of genetic (i.e., heritability) and family environmental influences on each outcome. We find evidence of sizeable genetic influences on risk tolerance, subjective wellbeing, cognitive performance, height, and BMI, and robust evidence of family environmental influences on educational attainment and labor market outcomes.

Keywords: nature, nurture, heritability, family environment, economic outcomes, molecular genetics

JEL classifications: A12, I12, J24, J300

*For helpful comments and suggestions, we thank Garett Jones, Richard Karlsson Linnér, Jonathan Schulz, Patrick Turley, and Alexander Young, as well as seminar participants at George Mason University. Acknowledgments specific to the six datasets used in this study are located in the Appendix.

†Markel and Beauchamp are the leading authors; other authors are listed alphabetically. Author contributions: Markel led the analyses and the writing of the manuscript; Beauchamp conceived the study, supervised the analyses, and led the writing of the manuscript; Rafael Ahlskog and Sven Oskarsson conducted the analyses in the Swedish Twin Registry (STR) data; Joakim Coleman Ebeltoft and Eivind Ystrøm conducted the analyses in the Norwegian Mother, Father and Child Cohort Study (MoBa) data. René Mõttus and Uku Vainik shared data and assisted with the analyses in the Estonian Biobank (EstBB) data; all authors provided critical feedback on the manuscript.

[1] Interdisciplinary Center for Economic Science and Department of Economics, George Mason University.
[2] Interdisciplinary Center for Economic Science and Department of Economics, George Mason University.
[3] Department of Government, Uppsala University.
[4] Department of Psychology, University of Oslo.
[5] School of Philosophy, Psychology and Language Sciences, University of Edinburgh.
[6] Department of Government, Uppsala University
[7] Institute of Psychology Institute of Genomics, University of Tartu
[8] Department of Psychology, University of Oslo.

*Correspondence to: garethmarkel@gmail.com, jonathan.pierre.beauchamp@gmail.com

The theory is confirmed that the pea hybrids form egg and pollen cells which, in their constitution, represent in equal numbers all constant forms which result from the combination of the characters united in fertilization.

—Gregor Mendel, 1865

1 Introduction

Mendel's First Law, also called the Law of Segregation, is a fundamental principle in genetics. It states that during the formation of a gamete (sperm or egg cell), at each location in the genome, a parent transmits to the gamete one strand of their DNA which is randomly selected among their two strands.[1,2]

A consequence of Mendel's First Law is that the genetic relatedness of any two siblings—defined here as the share of their DNA they commonly inherited from their parents—varies randomly.[3] At any location in their genomes, two sibs either inherit, with probability $\frac{1}{2}$, the same DNA strand from their mother, or they do not—and likewise from their father. As a result, their expected (genome-wide) genetic relatedness is 0.50, with a standard deviation of $\sim$ 0.04 (Visscher, Hill, and Wray, 2008). Figure 2 shows a histogram of the genetic relatedness for 19,142 sib pairs in the UK Biobank (UKB) dataset (one of the six datasets we analyze).

Here, we leverage the natural experiment implied by Mendel's First Law. We assume the standard "ACE" model from behavior genetics (Plomin et al., 2003), but instead of comparing dizygotic (DZ; also known as fraternal) vs. monozygotic (MZ; also known as identical) twins or adoptive vs. biological kin, for the first time (to our knowledge) in the economics literature we compare sib pairs' quasi-random genetic relatedness computed with their molecular genetic data. We link the sibs' genetic relatedness to their resemblance on a suite of 15 outcomes of relevance to economists in the labor market, cognitive and educational, risk taking, health, and anthropometric domains. This allows us to estimate the share of the variation in each outcome that is attributable to genetic factors—i.e., the heritability—and thus to test for genetic influences on each outcome. This also allows us to

[1] At every location in the genome, humans have two stands of DNA—one inherited from their father and one from their mother. These are organized across 23 pairs of chromosomes.

[2] Mendel's original statement of what is now known as his First Law appeared in his famous 1865 paper (Mendel, 1865) and differs from its modern interpretation. The above epigraph from Mendel's 1865 paper represents a central part of his original formulation.

[3] Throughout, unless otherwise noted, we use the terms "sibs" or "siblings" to refer to full biological siblings who have the same biological mother and father (but excluding monozygotic twins). As we further discuss below, the technical expression for what we call "genetic relatedness in the genetics literature is "identity by descent" (IBD).

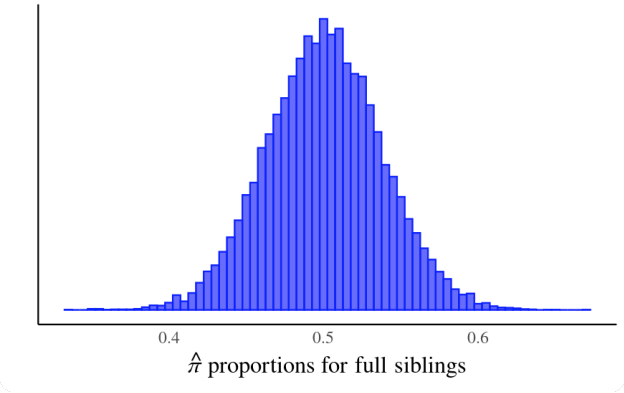


Figure 1: Histogram of the genetic relatedness of sib pairs in the UKB dataset.
Notes: The histogram shows the genetic relatedness (IBD, denoted $\hat{\pi}$ below) for the 19,142 sib pairs in the UKB dataset that passed our data quality control filters (see the Online Appendix). The mean genetic relatedness across the pairs is 0.501, and the standard deviation is 0.037. Genetic relatedness was computed using the SNIPAR software (Young, Nehzati, et al., 2022).

estimate the share of the variation that is attributable to the common family environment shared by the sibs.

Our method of comparing sibs' genetic similarity computed with molecular genetic data to their outcomes similarity was pioneered by geneticists nearly two decades ago (Visscher et al., 2006), and applied to estimate the heritability of height. Because height is a highly heritable trait, Visscher et al. (2006) were sufficiently well powered in a sample of only 3,375 sib pairs. For social-scientific outcomes like those we study, heritability tends to be lower and, as our power calculations suggest (see Online Appendix), much larger samples are necessary. To have sufficient statistical power, we therefore assembled a large dataset comprising $\sim 80,000$ sib pairs from six cohorts with rich outcomes and molecular genetic data.

We find statistically significant evidence (at the 5% level of significance) of genetic influences on several of the outcomes we study. We estimate the heritability ("h^2") of cognitive performance, risk tolerance, and subjective well-being to be 0.75 ($S.E. = 0.20$), 0.44 ($S.E. = 0.18$), and 0.34 ($S.E. = 0.17$), respectively; and that of BMI and height to be 0.58 ($S.E. = 0.11$), and 0.64 ($S.E. = 0.08$), respectively. We also find suggestive evidence (at the 10% level of significance) of genetic influences on fertility (number of children) ($h^2 = 0.24, S.E. = 0.16$). Surprisingly given our sample size of nearly 80,000 sib pairs for that outcome, power calculations, and prior estimates in the literature, our heritability estimate for educational attainment (years of education) is small and insignificant ($h^2 = 0.08, S.E. = 0.10$). Our heritability estimates for the remaining outcomes are not statistically significant, but this could be due to lower statistical power due to smaller samples and smaller true effects. For the labor market outcomes, our estimates are also likely

attenuated by the fact that male-female sib correlations for the period our data cover tend to be lower than the same-sex ones.

We also find statistically significant evidence of common family environmental influences on a number of outcomes, including educational attainment, family income, occupational income, occupational status, smoking and drinking behavior, and (more puzzlingly, at first glance) height.

The baseline ACE model we adopt from behavior genetics assumes zero assortative mating (i.e., it assumes random mating), contrary to the empirical evidence for many traits (Border et al., 2021). As is well understood, when there is actually positive assortative mating—as has been documented for educational attainment, many socioeconomic outcomes, and height (Beauchamp et al., 2011; Clark, 2023; Eika, Mogstad, and Zafar, 2019; Stulp et al., 2017)—the zero-assortative-mating assumption leads to downward bias in the heritability estimates and upward bias in the estimates of the common family environment's contribution. Thus, our baseline estimates of the outcomes' heritabilities are likely to be on the low side, while those of the contribution of the common family environment may tend to be too high.

Nonetheless, even when we adjust them for high assumed levels of assortative mating, our estimates still indicate the common family environment accounts for a substantial share ($\sim 25 - 30\%$) of the variation in educational attainment and occupational status. For height, however, the common family environmental influences vanish after adjustment for moderate levels of assortative mating. Adjusting for high but plausible levels of assortative mating boosts the heritability of educational attainment, with the upper bound of the 95% confidence interval exceeding 0.50; thus, despite our low baseline estimate of the heritability of educational attainment, we cannot rule out that the true heritability is substantial.

Our study relates to a sizeable literature in economics that has sought to estimate the relative influences of genetic and family environmental factors on economically relevant traits (e.g., Taubman, 1976; Björklund, Lindahl, and Plug, 2006; Cesarini et al., 2009; Sacerdote, 2011). This literature has mainly relied on twin studies—which for a given trait compare the similarity of DZ twin pairs to that of DZ twin pairs—as well as on adoption studies—which compare the similarity of adoptive kin to that of biological kin. Consistent with our overall results, this literature has found both genetic and common family environmental influences on most traits, though reported genetic influences have typically been considerably larger than family environmental ones, and a good number of twin studies have found negligible common family environmental influences. While our results align broadly with the existing literature, our relatively low heritability estimates for education diverge from previous studies. Further, our moderately large estimates of common family environmental influences on education and labor market outcomes—some of which persist even after adjusting for assortative mating—

are among the largest reported in the literature.

Both twin and adoption study designs have been criticized. Twin studies rely on various assumptions, including in particular the equal environment assumption (EEA), which stipulates that pairs of DZ twins share their environment to the same extent as pairs of DZ twins. Though various studies have found support for the EAA, it remains a controversial assumption. As for adoption studies, they typically rely on the assumption that the assignment of adoptees to their adoptive families is random. With the exception of a few studies in which placement was known to be quasi-random (Beauchamp, Schmitz, et al., 2023; Fagereng, Mogstad, and Rønning, 2021; Sacerdote, 2007), this assumption is problematic.

Our study design circumvents these issues. The EEA is not relevant for our study because we utilize the quasi-randomly determined genetic relatedness of full sibs instead of comparing DZ vs. DZ twins. And selective placement is also irrelevant here since this is not an adoption study. Because of Mendel's First Law, the relatedness of the sibling pairs we study is truly randomly assigned, and our estimates are free of contamination by environmental factors.

The remainder of the paper is organized as follows. Section 2 introduces our model and empirical approach. Section 4 summarize our estimation approach. Section 3 discusses the data and summarizes the steps we followed to prepare the genetic and outcomes data. Section 5 presents our results and discusses possible limitations. Section 6 concludes

2 Model

We use the standard "ACE" model from behavior genetics (Plomin et al., 2003) to decompose the variation in a given outcome into shares attributable to additive genetic factors (A), the common environment (C), and other factors (E). Additive genetic factors A capture genetic effects that increase linearly with the number of variants present at a specific genomic location and do not involve interactions between variants at different locations; they explain the bulk of the genetic variation for most traits (Hill, Goddard, and Visscher, 2008; Hivert et al., 2021). C captures common family environmental influences that are shared by siblings raised in the same family. And E, which behavioral geneticists refer to as "individual environment", in practice captures everything that is unique to an individual (including any measurement error), and is therefore independent across individuals.

The ACE model assumes that the outcome Y is the sum of these three influences:

$$Y = A + C + E. \tag{1}$$

The model further assumes that A, C, and E are mutually independent. For simplicity, we

will also assume here that Y has been standardized to have mean zero and unit variance. This implies that

$$\sigma_Y^2 = 1 = \sigma_A^2 + \sigma_C^2 + \sigma_E^2.$$

It follows that the heritability of Y—defined as the share of the variance of Y that is attributable to genetic factors—is equal to σ_A^2. Similarly, the shares of the variance due to common family environmental factors and other factors are equal to σ_C^2 and σ_E^2, respectively.[4]

Consider now two siblings with genetic relatedness π. They share a fraction π of their DNA, so the expected product of their additive genetic factor is $\mathbb{E}[A_1 A_2] = \text{Cov}[A_1, A_2] = \pi\sigma_A^2$. If the sibs were raised in the same family, the expected product of their (standardized) outcomes is given by

$$\begin{aligned}\mathbb{E}[Y_1 Y_2] &= \text{Cov}[Y_1, Y_2] \\ &= \text{Cov}[A_1, A_2] + \text{Cov}[C_1, C_2] + \text{Cov}[E_1, E_2] \\ &= \pi\sigma_A^2 + \sigma_C^2. \end{aligned} \tag{2}$$

In a sample of siblings whose genetic relatednesses are known, one can thus estimate the heritability and the common environment share by estimating the following "sib-regression", which regresses the product of the sibs' (standardized) outcomes on their genetic relatedness:

$$Y_1 Y_2 = \pi\beta_1 + \beta_0 + \epsilon, \tag{3}$$

where β_1 estimates $h^2 = \sigma_A^2$ and β_0 estimates σ_C^2. From this, one can obtain $\sigma_E^2 = 1 - \sigma_A^2 - \sigma_C^2$.

Figure 2 illustrates this regression for the cognitive performance outcome among 8,044 pairs of full sibs (excluding MZ twins) of European ancestry in the Generation Scotland (GS) dataset, for whom we estimated genetic relatedness (as described below). The estimated slope $\hat{\beta}_1 = 0.543$ and intercept $\hat{\beta}_0 = 0.085$ give the estimated heritability and common environment share, respectively.

The sib-regression (Equation 3) also works with a sample of twins. For DZ twins, $\mathbb{E}[\pi] = 0.5$, and for MZ twins, $\pi = 1$.[5] Figure 2 illustrates this. This can help understand what is arguably the most contentious assumptions of twin studies: the equal environment assumption (EEA). As mentioned, the EEA stipulates that pairs of DZ twins share their common environment to the same extent as pairs of DZ twins do. When EEA fails, the common environmental similarity of DZ twins is lower than that of MZ twins. σ_C^2 is then

[4] When Y is a binary outcome, Equation 1 and the assumption that A, C, and E are independent cannot both hold, and the model must be seen as approximation.

[5] From this, one obtains the classical behavior genetics formula for the estimation of heritability from twin data: $\mathbb{E}_{DZ}[Y_1 Y_2] = 0.5\sigma_A^2 + \sigma_C^2$ and $\mathbb{E}_{MZ}[Y_1 Y_2] = \sigma_A^2 + \sigma_C^2$, so $h^2 = \sigma_A^2 = 2(\mathbb{E}_{MZ}[Y_1 Y_2] - \mathbb{E}_{DZ}[Y_1 Y_2])$.

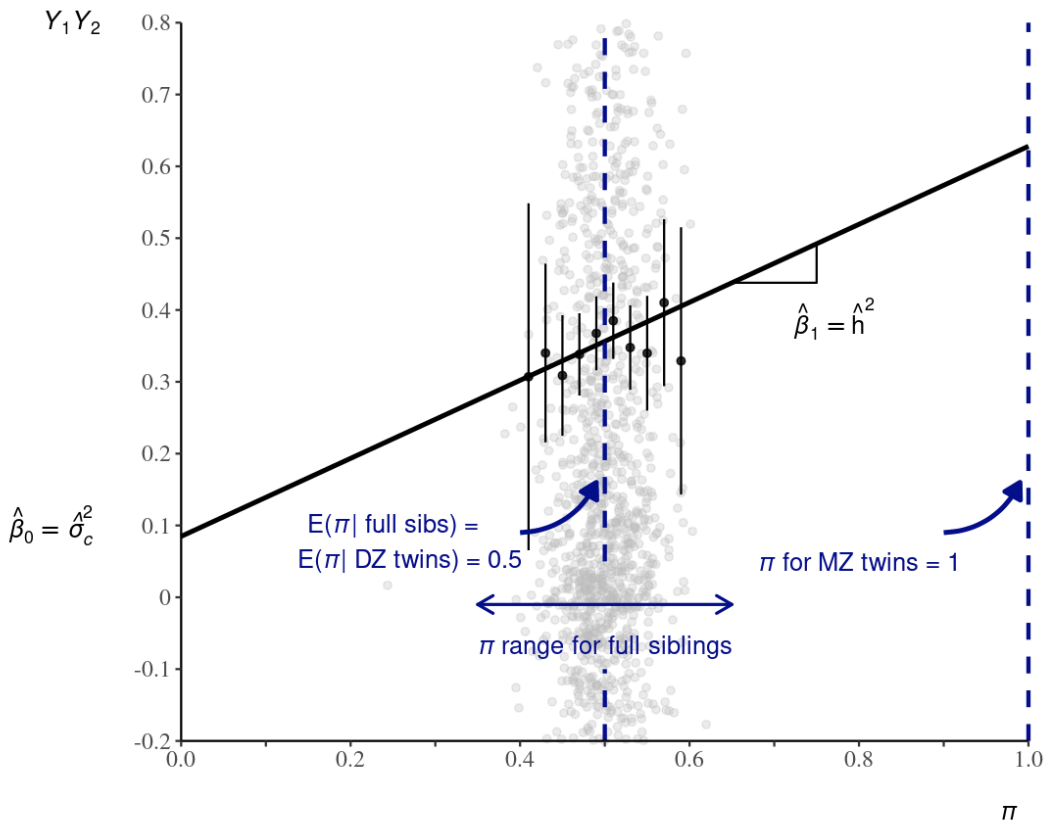


Figure 2: Illustrated sib-regression for cognitive performance in the GS dataset.

Notes: The figure illustrates the sib-regression (Equation 3) for cognitive performance among 8,044 sib pairs in the GS dataset. The y-axis gives the product Y_1Y_2 of the sibs' standardized cognitive performance residuals (see Section 3); the distribution of Y_1Y_2 has a much larger range than what is shown in the figure (mean=0.36, SD=1.07, min=-9.99, max=11.09). The x-axis gives their estimated genetic relatedness ($\hat{\pi}$). Each gray dot represents one sib pair; to reduce clutter, only 2,000 randomly sampled sib pairs are shown. The black dots and error bars represent the mean product and associated confidence interval for each of 10 equally sized bins defined based on the pairs' genetic relatednesses. The regression line illustrates the sib-regression, where the slope gives an estimate of the heritability ($\hat{h}^2 = \hat{\beta}_1 = 0.543$) and the intercept gives an estimate of the common environmental variance share ($\hat{\sigma}^2_C = \hat{\beta}_0 = 0.085$).

not a constant but correlates with π; its variable component is subsumed in the error term in Equation 3, which leads to omitted-variable bias in the estimation of the heritability β_1. By contrast, in the current study, the EEA is not a concern since we are not comparing the resemblance of DZ and MZ twins.

3 Data

Our data is drawn from the following six datasets from Estonia, Norway, Scotland, Sweden, and the United States. These datasets were selected because they have large numbers of

sibling pairs who have been genotyped and rich outcomes data:

The **Estonian Biobank (EstBB)** is a population-based dataset from Estonia, comprising over 200,000 participants—approximately one-fifth of the country's adult population. It integrates a wide range of health-related data, extensive lifestyle survey data, demographic information, and occupational histories for the participants. Outcome data are sourced from national education and health registries, as well as participant questionnaires (Milani et al., 2024; Arumäe et al., 2025). Notably, EstBB contains approximately 32,000 genotyped sibling pairs with data for this study.

Generation Scotland (GS) is a family-based cohort study from Scotland, comprising approximately 24,000 participants from around 7,000 families recruited between 2006 and 2011. It includes a broad range of health-related data, demographic and socioeconomic details, lifestyle factors, and cognitive function assessments. Outcome data are sourced from pre-clinic questionnaires, clinical assessments at intake, and national health records, including hospital, primary care, and prescription data (Smith et al., 2013). GS contains approximately 8,300 genotyped sibling pairs with data for this study.

Norwegian Mother, Father and Child Cohort Study (MoBa) is a population-based study from Norway involving over 280,000 participants. The study focuses on mothers, fathers (and their children), recruited before, during, and after pregnancy between 1999 and 2008. The study collects extensive data on health, lifestyle, diet, and environmental exposures through questionnaires administered during pregnancy and at various intervals after birth (Norwegian Institute of Public Health, n.d.), and supplements these by annual registry updates from Statistics Norway. Outcome data for this study come the Statistics Norway and questionnaires filled at week 15 of pregnancy and at 18 months after the child's birth (Norwegian Institute of Public Health, n.d.). MoBa includes approximately 11,400 genotyped sibling pairs with data for this study.

Swedish Twin Registry (STR) is a population-based registry with data on approximately 216,000 twins born in Sweden between 1886 and 2015. The registry collects extensive data on health, lifestyle, and environmental factors through self-administered questionnaires, physical measurements, and medical records (Zagai et al., 2019); we supplement these with military conscription data and data from Statistics Sweden. Excluding MZ twins, the STR contains approximately 5,600 pairs of genotyped DZ twins with outcome data for this study.

UK Biobank (UKB) is a biomedical database from the United Kingdom, comprising over 500,000 participants aged between 40 and 69 at recruitment (between 2006 and 2010). It integrates extensive health-related, lifestyle factors, cognitive function assessments, and imaging data (Bycroft, Freeman, Petkova, et al., 2018). Outcome data are sourced from initial assessments involving touchscreen questionnaires, verbal interviews, and physical

Table 1: The 15 outcomes and sample sizes

Outcome	**Notes**	**Sample size (N_{pairs})**						
		TOTAL	**EstBB**	**GS**	**MoBa**	**STR**	**UKB**	**WLS**
Panel A. Cognitive and educational								
Cognitive performance		17,332		8,044		687	6,908	1,693
EA	Years of education	79,714	35,124	7,968	11,378	5,672	18,949	623
Panel B. Labor market								
Employed	Observations at ages 30-60 only	37,290	11,342		11,367	4,659	9,462	460
Log family income	Observations at ages 30-60 only	26,976	3,241	4,639	11,354		7,742	
Log occupational income	Imputed using occupational codes	30,548	7,938		10,455	3,133	9,022	
Occupational Status	Imputed to SIOPS scale using occup. codes	25,072	2,183		10,615	3,351	8,923	
Panel C. Risk tolerance and risky behaviors								
Ever smoker	Indicator of whether one ever smoked	70,677	37,250	7,870	1,517	3,094	19,078	1,868
Cigarettes per day	Log(1 + cigarettes per day)	41,098	19,534	7,607	749	1,119	10,228	1,861
Drinks per week	Log(1 + drinks per week)	36,442	15,725	7,070			12,199	1,448
Risk tolerance	Extent to which one taskes risks; self-reported	28,827	9,249			1,425	18,153	
Panel D. Health-related & other								
Number of children	Women age 45+ and men age 50+ only	34,926	2,954		11,375	4,210	16,387	
General health	Self-rated	51,917	21,851	7,318	2,321	1,335	19,092	
Subjective well-being	Positive affect or life satisfaction; self-reported	29,949	9,076	7,323	9,494	1,440	2,616	
Panel E. Anthropometric								
BMI		78,456	37,300	8,312	9,502	2,930	19,112	1,300
Height		78,846	37,360	8,355	9,732	2,933	19,148	1,318

Notes: For each outcome, the total sample size is the sum of the sample sizes of the individual datasets.

measurements, with a subset of participants undergoing repeat assessments. The UKB includes approximately 20,000 genotyped sibling pairs for this study.

Wisconsin Longitudinal Study (WLS) is an ongoing cohort study that began in 1957, following a random sample of over 10,000 individuals who graduated from Wisconsin high schools that year, along with their siblings and spouses. It collects extensive data on education, employment, family, health, and social factors through periodic surveys conducted over six decades (Herd, Carr, and Roan, 2014). The WLS includes approximately 1,900 genotyped sibling pairs with data for this study.

The 15 outcomes we study are listed and briefly described in Table 1; Online Appendix Table E.1 provides additional details regarding the definition, construction, and source of each outcome variable for each dataset. For each outcome, Table 1 also the total number of genotyped sib pairs with nonmissing data across the datasets as well as as the number of sib pairs in each dataset.

We use the sib pairs' molecular genetic data to compute their genetic relatedness π. The technical expression for what we refer to as the pair's genetic relatedness is their "identity by descent" (IBD). When two sibs share a particular segment of DNA that was passed down from their mother or father, they are said to be "identical by descent" for that segment. A sib pair's IBD measures how much of their DNA they commonly inherited from their parents.[6] To compute the genetic relatedness (i.e., the IBD) of the sib pairs in each of our six datasets, we first apply some standard quality control filters to the genetic data. Among other such filters, we drop all sib pairs of non-European ancestry.[7] We then use the software tool SNIPAR (Young, Nehzati, et al., 2022) to compute the pairs' genetic relatedness. Online Appendix A provides additional details.

4 Estimation

To estimate the heritability (h^2) and the share of the variation that is due to the common environment (σ^2_C) for each outcome, we use the DeFries-Fulker (DF) regression (DeFries and Fulker, 1985). We use the DF regression instead of the "Haselman-Elston"-style regression from Equation 3 (Haseman and Elston, 1972) to maximize statistical power, as we discuss

[6]Technically, two individuals are identical by descent for a segment when they inherit the same segment from a common ancestor, which may also be, e.g., a great-grand-parent (if two sibs are the offspring of first-cousins) or a great-great-grand-parent. In Western countries, marriage among close kin is rare, so the IBD of sibs in our data is almost entirely due to shared segments inherited from their mothers or fathers.

[7]Different ancestral groups display different rates of genetic recombination, which can bias the IBD estimation (Kong et al., 2010). Our six datasets come from historically mostly homogeneous countries, so this filter required dropping few sib pairs.

in Online Appendix D, where we compare the efficiency of different estimation methods.

The DeFries-Fulker regression is

$$y_{1,j} = \beta_0 + \beta_1 y_{2,j} + \beta_2 y_{2,j}\hat{\pi}_j + \epsilon \tag{4}$$

where $y_{i,j}$ is the outcome of individual i in sib pair j and $\hat{\pi}_j$ is the estimated relatedness of sibling pair j. As we show in Appendix I (and as others have shown), $\hat{\beta}_2$ is an estimate of the heritability h^2 and $\hat{\beta}_1$ is an estimate of σ_C^2.

We estimate this regression in double-entry form, where each sib pair appears as an observation twice: once with sib 1 on the left-hand side and sib 2 on the right-hand side, and once with sib 2 on the left-hand side and sib 1 on the right-hand side. Standard errors are clustered at the family level (i.e., all the double entries of all the pairs of full siblings with the same biological mother and father form one cluster). Kohler and Rodgers, 2000 show that this double-entry form improves statistical power. To avoid introducing bias in our DF regression estimates, we do not constrain them to lie in the unit interval, even though they correspond to variance shares. And since we are statistically testing whether each variance component is larger than zero, we use one-sided tests (i.e., we test H_0: $h^2 = 0$ vs. H_1: $h^2 > 0$, and similarly for σ_C^2 and σ_E^2).

The baseline ACE model (Equation 1) and the DF estimation framework implicitly assume that variance of an outcome is constant with respect to sex and age (see Appendix I). In practice, however, for most of our outcomes the variance varies as a function of sex, birth year, and age at measurement. For that reason, before estimating the DF regression, we residualize each outcome on birth year dummies, sex, and their interactions, and then standardize the resulting residuals separately by sex or, in some cases, by sex-birth-year bins. For outcomes with multiple measurements, we also residualize out age at measurement and, after standardizing the resulting residuals, take the average across measurements. Each resulting residualized and standardized outcome thus has zero mean and a variance that is unity and constant across sexes, birth year, and measurement ages. Online Appendix B provides additional details.

Due to data-sharing constraints, the analysis was conducted separately for each dataset. To combine estimates from the different datasets, we use inverse variance weighted meta-analysis, whereby the meta-analyzed estimate of a given parameter θ is given by

$$\hat{\theta} = \frac{\sum_d \hat{\theta}_d / \sigma^2_{\hat{\theta}_d}}{\sum_d 1/\sigma^2_{\hat{\theta}_d}}, \tag{5}$$

where $\hat{\theta}_d$ is the estimate of θ in dataset d and $\sigma^2_{\hat{\theta}_d}$ is the variance of $\hat{\theta}_d$.

Online Appendix Figures D.5 and D.6 show the results of simulations to evaluate the statistical power to obtain significant estimates (at the 5% level of significance) of h^2 and σ_C^2 in our DF regressions, for various assumed true values of these parameters and as a function of sample size. If the true h^2 of a given outcome is 0.5 or higher, then power exceeds 90% even with a sample of only 20,000 sib pairs, which we have for all outcomes except cognitive performance; if the true h^2 is 0.3, then a little more than 40,000 sib pairs are needed to achieve 80% power; and with 80,000 sib pairs, as we nearly have for several traits, power to detect a true h^2 of 0.2 exceeds 70%. Power is considerably higher for σ_C^2: for a true σ_C^2 of only 0.2, nearly 80% power is achieved in a sample of only 20,000 pairs; and with a sample of 80,000 pairs, power to detect a true σ_C^2 of only 0.1 exceeds 70%. In sum, given the sample sizes for most of our outcomes, we are well powered to detect moderate true h^2's and even rather small true σ_C^2's.

5 Results

Table 2 reports the meta-analyzed estimates of h^2, σ_C^2, and σ_E^2 from the baseline ACE model without assortative mating. Four main findings emerge.

First, we find clear evidence of genetic influences on socioeconomic outcomes. For all outcomes, the heritability estimate is either positive or close to (and statistically not different from) zero. In other words, the slope $\hat{\beta}_1$ in Figure 2 (which is equal to the heritability) is either positive or nonnegative across the outcomes, thus implying that greater sib genetic relatedness tends to be associated with greater (and never lower) sib outcome similarity. Because genetic relatedness among sib pairs is quasi-random per Mendel's First Law, this association can be interpreted as causal and clearly points to causal genetic influences.

We obtain significant (at the 5% level) and, in some cases sizeable, heritability estimates for cognitive performance, risk tolerance, subjective well-being, BMI, and height. For cognitive performance, we estimate the heritability to be $h^2 = 0.746$, implying that 75% of the variance is attributable to genetic factors. However, the standard error of our estimate is large ($S.E. = 0.196$), so we cannot rule out more modest genetic influences. For self-reported risk tolerance and self-reported subjective well-being, we obtain heritability estimates of 0.442 ($S.E. = 0.176$) and 0.337 ($S.E. = 0.173$), respectively. And for BMI and height—for which we have nearly 80,000 sib pairs—we obtain sizeable and rather precise estimates of 0.575 ($S.E. = 0.108$) and 0.639 ($S.E. = 0.080$), respectively. We also obtain suggestive evidence (significant at the 10% level) that fertility (the number of children one has) is moderately heritable ($h^2 = 0.243$, $S.E. = 0.163$). These estimates are broadly in line with the existing behavioral genetics and social genomics literatures based on twin and

adoption studies, which have tended to report high heritabilities for BMI, height, and cognitive performance and moderate ones for preferences and personality measures (Beauchamp, Cesarini, and Johannesson, 2017; Polderman et al., 2015; Vukasović and Bratko, 2015).

Table 2: Baseline ACE model estimates

	N_{pairs}	$\hat{h^2}$	$\hat{\sigma}^2_C$	$\hat{\sigma}^2_E$	$\hat{\rho}_{sib}$
Panel A. Cognitive and educational					
Cognitive performance	17,332	0.746***	-0.040	0.294***	0.334***
		(0.196)	(0.099)	(0.098)	(0.007)
EA	79,714	0.076	0.323***	0.600***	0.362***
		(0.095)	(0.048)	(0.047)	(0.003)
Panel B. Labor market					
Employed	37,290	-0.083	0.127*	0.957***	0.081***
		(0.172)	(0.087)	(0.086)	(0.005)
Log family income	26,976	-0.106	0.247**	0.858***	0.202***
		(0.209)	(0.105)	(0.105)	(0.006)
Log occupational income	30,458	0.054	0.188***	0.758***	0.212***
		(0.157)	(0.079)	(0.079)	(0.006)
Occupational Status	25,072	-0.050	0.268***	0.782***	0.236***
		(0.165)	(0.083)	(0.083)	(0.006)
Panel C. Risk tolerance and risky behaviors					
Ever smoker	70,677	0.137	0.128***	0.735***	0.201***
		(0.109)	(0.055)	(0.055)	(0.004)
Cigarettes per day (logged)	41,098	0.135	0.152**	0.712***	0.227***
		(0.164)	(0.082)	(0.082)	(0.005)
Drinks per week (logged)	36,442	0.025	0.146**	0.829***	0.153***
		(0.162)	(0.081)	(0.081)	(0.005)
Risk tolerance	28,827	0.442***	-0.144	0.701***	0.081***
		(0.176)	(0.088)	(0.088)	(0.006)
Panel D. Health-related & other					
Number of children	34,926	0.243*	0.007	0.750***	0.132***
		(0.163)	(0.082)	(0.082)	(0.005)
Self-rated general health	51,917	0.057	0.105*	0.837***	0.133***
		(0.133)	(0.067)	(0.067)	(0.004)
Subjective wellbeing	29,949	0.337**	-0.060	0.724***	0.116***
		(0.173)	(0.087)	(0.086)	(0.006)
Panel E. Anthropometric					
BMI	78,456	0.575***	-0.009	0.443***	0.279***
		(0.108)	(0.054)	(0.046)	(0.003)
Height	78,846	0.639***	0.183***	0.178***	0.500***
		(0.080)	(0.040)	(0.040)	(0.003)

Notes: The table reports the estimates of h^2, σ^2_C, and σ^2_E from the baseline ACE model (without assortative mating). The estimates were obtained by meta-analyzing the dataset-level estimates, as described in the text. N_{pairs} is the total number of sib pairs in the meta-analysis. ρ_{sib} is the correlation in the outcome across sib pairs (from Equation 3, $\rho_{sib} = \sigma^2_C + \pi h^2$). Stars indicate the significance of the estimates on one-sided tests, as described in the text: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Our second main finding is that the common family environment appears to play an im-

portant role for many of our outcomes, including education attainment ($\sigma^2_C = 0.323, S.E. = 0.048$), the labor market outcomes (with $\sigma^2_C \sim 0.20 - 0.25$ for log family income, log occupational income, and occupational status), smoking and drinking behavior ($\sigma^2_C \sim 0.15$), and height ($\sigma^2_C = 0.183, S.E. = 0.040$).

As is well known, the baseline ACE model assumes no assortative mating, and this can bias estimates of σ^2_C upwards and those of h^2 downwards when assortative mating is in fact present. Online Appendix C provides further details on this and derives formula for assortative-mating-adjusted estimators for the heritability and the common environmental share of a given outcome. The key parameter for these adjusted estimators is the correlation r between mothers' and fathers' additive genetic factors for the outcome. r cannot be observed directly with current technologies, but evidence from both observable spousal outcome correlations and from molecular genetic data suggests that it is positive and non-negligible for most of our outcomes, including educational attainment, labor market outcomes, substance use, cognitive performance, and height (Border et al., 2021; Beauchamp et al., 2011; Clark, 2023; Eika, Mogstad, and Zafar, 2019; Okbay et al., 2022; Stulp et al., 2017).

To test the robustness of our baseline σ^2_C estimates to the zero-assortative-mating assumption, we re-estimated σ^2_C under various assumed levels of assortative mating ($r = 0.1$, $r = 0.3$, and $r = 0.5$). Online Appendix Table E.4 shows the results. While most significant baseline σ^2_C estimates remain so under weak assortative mating ($r = 0.1$), those for height and for drinking and smoking behavior become small and lose their significance under moderate or high assortative mating ($r = 0.3, 0.5$). By contrast, the σ^2_C estimate for educational attainment, log family income, and occupational status remain sizeable ($\sim 0.25 - 0.35$) even with high levels of assortative mating (though that for log family income loses statistical significance as its standard errors becomes larger).

Sizeable σ^2_C estimates for educational attainment have previously been reported in the literature (e.g., Branigan, McCallum, and Freese, 2013), but Wolfram and Morris, 2023 and Silventoinen et al., 2020 find that correcting for assortative mating renders these small or nil. Sizeable σ^2_C estimates for smoking and drinking behavior have also been reported (Beauchamp, Schmitz, et al., 2023; Sacerdote, 2007), but those for labor market outcomes such as income have often been lower (see, e.g., Table 1 in Hyytinen et al., 2019). Thus, our σ^2_C estimates for educational attainment and labor market outcomes and with their robustness to adjusting for assortative mating stand out vs. comparable estimates in the literature, and point to strong influences of the family environment on these outcomes.

Our third main finding is that much of the variation in the 15 outcomes is due to what behavioral geneticists call the "individual environment". Indeed, σ^2_E is significant for all outcomes, and exceeds 0.60 for all but three outcomes (cognitive performance, BMI, and

height). This finding of large individual environmental influences is consistent with what has repeatedly been found for most outcomes in the behavioral genetics and social genomics literatures, and has been dubbed the "Third Law of behavior genetics" (Turkheimer, 2000).

In addition to these three findings, we note that our heritability estimates for many of our outcomes are small and not statistically different from zero. In particular, for educational attainment, we estimate a small and insignificant heritability ($h^2 = 0.076 S.E. = 0.095$). This result stands out because our sample of nearly 80,000 sib pairs translates into over 70% and $\sim 95\%$ power to detect true heritabilities of 0.2 and 0.3, while a recent meta-analysis and a large-scale twin study both estimated heritabilities around 0.40 (Branigan, McCallum, and Freese, 2013; Silventoinen et al., 2020). The heritability estimates for the four labor market outcomes, for drinking and smoking behavior, and for self-rated general health are also all small and statistically indistinguishable for zero.

However, caution is warranted in interpreting these results. For educational attainment, high levels of assortative mating have been documented (e.g., Eika, Mogstad, and Zafar, 2019; Okbay et al., 2022); adjusting for high levels of assortative mating ($r = 0.5$) yields an adjusted heritability estimate of 0.153 (Table E.3), with a 95% confidence interval ranging up to 0.52. Thus, we cannot rule out substantial genetic influences on educational attainment. As for the other outcomes with small and insignificant heritability estimates, their sample sizes are lower ($N_{pairs} \sim 25,000 - 40,000$, except for general health), thus limiting our statistical power to detect small or moderate heritabilities. Further, assortative mating is likely present. For the labor market outcomes, outcome correlations tend to be lower for opposite-sex sib pairs than for same-sex pairs, thus also likely depressing estimates relative to those that would have been obtained using only same-sex pairs.

5.1 Limitations

Unlike twin studies, our empirical approach with sib pairs does not rely on the equal environment assumption (EEA, discussed in Section 2). Our baseline ACE model nonetheless makes several important structural assumptions that have been the subject of vigorous debates. In particular, it assumes only a specific type of genetic variance ("additive" variance) and it assumes no assortative mating and no gene-environment interactions or correlations. Additive genetic variance accounts for most of the genetic variance for most traits (Hill, Goddard, and Visscher, 2008), and recent research indicates that genetic dominance variance is negligible for most complex traits (Hivert et al., 2021). We have tested the robustness of our results to the zero-assortative mating assumption. As for gene-environment interactions and correlations, these can introduce bias in ACE model estimates (Eaves et al., 1977; Plomin

et al., 2003). Unfortunately, our data does not allow us to specify sufficiently many empirical moments to estimate more complex models that allow for these. Future research with data that include more pedigree relationships than just siblings or with polygenic indices should seek to address these issues.

6 Conclusion

We have leveraged a powerful natural experiment—Mendel's First Law and the quasi-random variation in genetic relatedness among full siblings—to estimate the relative contributions of genetic and common family environmental influences on a broad set of outcomes. To do so, we have used molecular genetic data to compute the genetic relatedness of ∼80,000 sibling pairs across six datasets, and compared that relatedness to the pairs' outcome similarity. Our approach sidesteps some of the main limitations of twin and adoption studies—namely, the equal environment assumption and selective placement.

Our findings provide compelling evidence for both genetic and common family environmental influences, with the relative importance varying significantly across traits. We document substantial heritability for cognitive performance, risk tolerance, subjective well-being, BMI, and height, reinforcing prior results from twin and genomic studies. At the same time, we uncover sizeable and robust shared environmental influences for educational attainment and labor market outcomes—results that persist even after adjusting for high levels of assortative mating.

Our findings invite further exploration into the mechanisms through which these influences operate. Future research should also aim to obtain more precise estimates by using even larger samples and incorporating a richer set of pedigree relationships than just sibs (Young, Frigge, et al., 2018), which could also allow a relaxation of the assumptions of no gene-environment correlations nor interactions.

References

Arumäe, Kadri, Mariliis Vaht, Anu Realo, Liisi Ausmees, Jüri Allik, et al. (2025). "Cohort Profiles: Personality Measurements at the Estonian Biobank of the Estonian Genome Center, University of Tartu". In.

Beauchamp, Jonathan, Lauren Schmitz, James J Lee, and Matt McGue (2023). "Nature-nurture interplay: Evidence from molecular genetic and pedigree data in Korean American adoptees". In.

Beauchamp, Jonathan P, David Cesarini, and Magnus Johannesson (2017). "The psychometric and empirical properties of measures of risk preferences". In: *Journal of Risk and Uncertainty* 54.3, pp. 203–237.

Beauchamp, Jonathan P, David Cesarini, Magnus Johannesson, Erik Lindqvist, and Coren Apicella (2011). "On the sources of the height intelligence correlation: New insights from a bivariate ACE model with assortative mating". In: *Behavior Genetics* 41.2, pp. 242–252.

Björklund, Anders, Mikael Lindahl, and Erik Plug (2006). "The origins of intergenerational associations: Lessons from Swedish adoption data". In: *The Quarterly Journal of Economics* 121.3, pp. 999–1028.

Border, Richard, Sean O'Rourke, Teresa de Candia, Michael E. Goddard, Peter M. Visscher, et al. (2021). "Assortative Mating Biases Marker-based Heritability Estimators". In: *bioRxiv*.

Branigan, Amelia R., Kenneth J. McCallum, and Jeremy Freese (June 2013). "Variation in the Heritability of Educational Attainment: An International Meta-Analysis". In: *Social Forces* 92.1, pp. 109–140.

Bycroft, C., C. Freeman, D. Petkova, et al. (2018). "The UK Biobank resource with deep phenotyping and genomic data". In: *Nature* 562.7726, pp. 203–209.

Cesarini, David, Christopher Dawes, Magnus Johannesson, Paul Lichtenstein, and Bjorn Wallace (2009). "Genetic Variation in Preferences for Giving and Risk Taking". In: *Quarterly Journal of Economics* 124.4, pp. 809–842.

Clark, Gregory (2023). "The inheritance of social status: England, 1600 to 2022". In: *Proceedings of the National Academy of Sciences* 120.27, e2300926120.

DeFries, J. C. and D. W. Fulker (Sept. 1985). "Multiple regression analysis of twin data". In: *Behavior Genetics* 15.5, pp. 467–473.

Eaves, Lindon J, Krystyna Last, Nicholas G Martin, and John L Jinks (1977). "A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences". In: *British journal of mathematical and statistical psychology* 30.1, pp. 1–42.

Eika, Lasse, Magne Mogstad, and Basit Zafar (2019). "Educational assortative mating and household income inequality". In: *Journal of Political Economy* 127.6, pp. 2795–2835.

Fagereng, Andreas, Magne Mogstad, and Marte Rønning (2021). "Why Do Wealthy Parents Have Wealthy Children?" In: *Journal of Political Economy* 129.3, pp. 703–756.

Haseman, Joseph K and Robert C Elston (1972). "The investigation of linkage between a quantitative trait and a marker locus". In: *Behavior genetics* 2.1, pp. 3–19.

Herd, Pamela, Deborah Carr, and Carol Roan (Feb. 2014). "Cohort profile: Wisconsin longitudinal study (WLS)". In: *International Journal of Epidemiology* 43.1, pp. 34–41.

Hill, William G, Michael E Goddard, and Peter M Visscher (2008). "Data and theory point to mainly additive genetic variance for complex traits". In: *PLoS genetics* 4.2, e1000008.

Hivert, Valentin, Julia Sidorenko, Florian Rohart, Michael E Goddard, Jian Yang, et al. (2021). "Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals". In: *The American Journal of Human Genetics* 108.5, pp. 786–798.

Hyytinen, Ari, Pekka Ilmakunnas, Edvard Johansson, and Otto Toivanen (2019). "Heritability of lifetime earnings". In: *The Journal of Economic Inequality* 17, pp. 319–335.

Kohler, Hans-Peter and Joseph Lee Rodgers (2000). "DF-Analyses of Heritability with Double-Entry Twin Data: Asymptotic Standard Errors and Efficient Estimation". In: *Behavior Genetics* 31, pp. 179–191.

Kong, Augustine, Gudmar Thorleifsson, Daniel F Gudbjartsson, Gisli Masson, Asgeir Sigurdsson, et al. (2010). "Fine-scale recombination rate differences between sexes, populations and individuals". In: *Nature* 467.7319, pp. 1099–1103.

LaBuda, M.C., J.C. DeFries, and D.W. Fulker (1986). "Multiple regression analysis of twin data obtained from selected samples". In: *Genetic Epidemiology* 3.6, pp. 425–433.

Mendel, Gregor (1865). *Experiments in Plant Hybridization*. Ed. by Roger Blumberg. Trans. by William Bateson. This version includes corrections and edits by Roger Blumberg as part of the MendelWeb project. Electronic Scholarly Publishing Project.

Milani, Lili, Maris Alver, Sven Laur, Sulev Reisberg, Toomas Haller, et al. (2024). "From Biobanking to Personalized Medicine: the journey of the Estonian Biobank". In: *medRxiv*.

Norwegian Institute of Public Health (n.d.). *Questionnaires from MoBa — fhi.no*. https://www.fhi.no/en/ch/studies/moba/for-forskere-artikler/questionnaires-from-moba/. [Accessed 11-02-2025].

Okbay, Aysu, Yeda Wu, Nancy Wang, Hariharan Jayashankar, Michael Bennett, et al. (2022). "Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals". In: *Nature genetics* 54.4, pp. 437–449.

Plomin, Robert, John C DeFries, Ian W Craig, and Peter McGuffin (2003). *Behavioral genetics*. American Psychological Association.

Polderman, Tinca JC, Beben Benyamin, Christiaan A De Leeuw, Patrick F Sullivan, Arjen Van Bochoven, et al. (2015). "Meta-analysis of the heritability of human traits based on fifty years of twin studies". In: *Nature genetics* 47.7, pp. 702–709.

Rodgers, J.L. and M. McGue (1994). "A simple algebraic demonstration of the validity of DeFries-Fulker analysis in unselected samples with multiple kinship levels". In: *Behavior Genetics* 24.3, pp. 259–262.

Sacerdote, Bruce (Feb. 2007). "How Large are the Effects from Changes in Family Environment? A Study of Korean American Adoptees*". In: *The Quarterly Journal of Economics* 122.1, pp. 119–157.

— (2011). "Nature and Nurture Effects on Children's Outcomes: What Have We Learned from Studies of Twins and Adoptees?" In: *Handbook of social economics*. Vol. 1. Elsevier, pp. 1–30.

Silventoinen, Karri, Aline Jelenkovic, Reijo Sund, Antti Latvala, Chika Honda, et al. (2020). "Genetic and environmental variation in educational attainment: an individual-based analysis of 28 twin cohorts". In: *Scientific reports* 10.1, p. 12681.

Smith, Blair H, Archie Campbell, Pamela Linksted, Bridie Fitzpatrick, Cathy Jackson, et al. (2013). "Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness". In: *International Journal of Epidemiology* 42.3, pp. 689–700.

Stulp, Gert, Mirre JP Simons, Sara Grasman, and Thomas V Pollet (2017). "Assortative mating for human height: A meta-analysis". In: *American journal of human biology* 29.1, e22917.

Taubman, Paul (1976). "The Determinants of Earnings: Genetics, Family, and Other Environments: A Study of White Male Twins". In: *American Economic Review* 66.5, pp. 858–870.

Turkheimer, Eric (2000). "Three laws of behavior genetics and what they mean". In: *Current directions in psychological science* 9.5, pp. 160–164.

Visscher, Peter, William Hill, and Naomi Wray (2008). "Heritability in the genomics era—concepts and misconceptions". In: *Nature reviews genetics* 9.4, pp. 255–266.

Visscher, Peter M, Sarah E Medland, Manuel A R Ferreira, Katherine I Morley, Gu Zhu, et al. (2006). "Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings". In: *PLoS genetics* 2.3, e41.

Vukasović, Tena and Denis Bratko (2015). "Heritability of personality: A meta-analysis of behavior genetic studies." In: *Psychological bulletin* 141.4, p. 769.

Wolfram, T. and D. Morris (2023). "Conventional twin studies overestimate the environmental differences between families relevant to educational attainment". In: *npj Science of Learning* 8.1, p. 24.

Young, Alexander, Michael Frigge, Daniel Gudbjartsson, Gudmar Thorleifsson, Gyda Bjornsdottir, et al. (2018). "Relatedness Disequilibrium Regression Estimates Heritability Without Environmental Bias". In: *Nature Genetics* 50.9, pp. 1204–1206.

Young, Alexander I, Seyed Moeen Nehzati, Stefania Benonisdottir, Aysu Okbay, Hariharan Jayashankar, et al. (2022). "Mendelian imputation of parental genotypes improves estimates of direct genetic effects". In: *Nature Genetics* 54, pp. 897–905.

Zagai, Ulrika, Paul Lichtenstein, Nancy L. Pedersen, and Patrik K. E. Magnusson (2019). "The Swedish Twin Registry: Content and Management as a Research Infrastructure". In: *Twin Research and Human Genetics* 22.6, pp. 672–680.

APPENDIX

I. Derivation of the DeFries-Fulker regression

From the additive genetic model, the correlation between two sibling's outcomes conditional on their genetic relatedness is given by the following equation:

$$corr(y_{2,j}y_{1,j}|\pi_j) = \sigma_C^2 + \sigma_A^2\pi_j. \tag{6}$$

Due to the double entry format—whereby each pair appears as an observation twice (once with sib 1 on the LHS and sib 2 on the RHS, and then vice versa)—$y_{2,j}$ and $y_{1,j}$ contain the same elements and have the same means and variances. Therefore, we can rewrite the conditional correlation as the coefficient of a regression of $y_{1,j}$ on $y_{2,j}$ for the set of sibling pairs with relatedness π_j:

$$corr(y_{2,j}y_{1,j}|\pi_j) = \frac{cov(y_{2,j}, y_{1,j}|\pi_j)}{\sqrt{var(y_{2,j})var(y_{1,j})}} = \frac{cov(y_{2,j}, y_{1,j}|\pi_j)}{var(y_{2,j})}. \tag{7}$$

This implies the following regression of $y_{1,j}$ on $y_{2,j}$ for the sibling pairs with relatedness π_j:

$$y_{1,j} = \beta_0 + \beta_{1,j}y_{2,j} + u, \tag{8}$$

where $\beta_{1,j} = \sigma_C^2 + \sigma_A^2\pi_j$.

This in turn implies the following regression of $y_{1,j}$ on $y_{2,j}$ and $y_{2,j}\pi_j$, for all the sibling pairs:

$$y_{1,j} = \beta_0 + \beta_1 y_{2,j} + \beta_2 y_{2,j}\pi_j + u, \tag{9}$$

where $\beta_1 = \sigma_C^2$ and $\beta_2 = \sigma_A^2$ (and $\beta_0 = 0$).

The full DeFries-Fulker regression also includes π_j as an additional covariate. As Rodgers and McGue, 1994 shows, the expectation of the coefficient on that additional covariate is $-\sigma_A^2 E[y]$. Here, all the outcomes are standardized, so $E[y] = 0$. Thus, and as the above shows, we do not need that additional covariate and so do not include it in our DeFries-Fulker regression. More details on this derivation are provided in LaBuda, DeFries, and Fulker, 1986 and Rodgers and McGue, 1994.

II. Dataset-specific acknowledgments

Estonian Biobank (EstBB): This work has been funded by Estonian Research Council's personal research funding start-up grants PSG656 and PSG759, and Estonian Research Council's team grants PRG2190 and PRG1291. The activities of the EstBB are regulated by

the Human Genes Research Act, which was adopted in 2000 specifically for the operations of the EstBB. Individual level data analysis in the EstBB was carried out under ethical approval 1.1-12/626, 13.04.2020 granted by the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application 3-10/GI/11571 from the Estonian Biobank.

Special thanks are owed to the EstBB research team (Andres Metspalu, Lili Milani, Tõnu Esko, Reedik Mägi, Mari Nelis and Georgi Hudjashov)—affiliated with the Estonian Genome Centre in the Institute of Genomics at University of Tartu—for performing genotyping, imputation, and quality control on samples from the EstBB.

Generation Scotland (GS): Ethical approval for the GS: Scottish Family Health Study was obtained from the Tayside Committee on Medical Research Ethics (on behalf of the National Health Service).

Moba [[[TBD]]]

STR This research uses data from the Swedish Twin Registry (STR), funded by the Swedish Research Council (Grants 2019-00244, 2023-01343 and 2024-06499).

UK Biobank (UKB): This research has been conducted using the UK Biobank Resource under application number 99086.

Wisconsin Longitudinal Study (WLS): This research uses data from the Wisconsin Longitudinal Study (WLS), funded by the National Institute on Aging (R01 AG009775; R01 AG033285; R01 AG060737, R01 AG041868)

ONLINE APPENDIX

A Methods: preparing the genetic data and computing genetic relatedness

We conduct the steps in this section separately for each dataset.

A.1 Genetic Data

For each dataset, we use imputed genetic data.

Cohort-specific details:

EstBB: Participants were genotyped with the Infinium Global Screening Array (GSA) with 309,258 SNPs. These data were then imputed to the Estonian Reference haplotype panels. The positions are in GRCh37 coordinates (Kuznetsov et al., 2023)

GS: Participants were genotyped with the Illumina HumanOmniExpressExome-8 v1.0 BeadChip array and the Illumina HumanOmniExpressExome-8 v1.2 BeadChip array. These data were then imputed to the Haplotype Reference Consortium panel (Nagy et al., 2017).

UKB: Participants were genotyped with the UK BiLEVE Axiom array and the UK Biobank Axiom array with over 800,000 SNPs (Bycroft, Freeman, Petkova, et al., 2017). These data were then imputed to the Haplotype Reference Consortium and UK10k haplotype panels. The positions are in GRCh37 coordinates (Bycroft, Freeman, Petkova, et al., 2018)

MoBa: Participants were genotyped in 26 batches using various versions of Illumina Global Screening Array, Illumina HumanOmniExpress array, and Illumina HumanCoreExome array Corfield et al., 2024. The MoBaPsychGen quality control pipeline combined these batches and provided 6,981,748 SNPs after post-imputation quality control Corfield et al., 2024.

STR: Participants were genotyped in three different batches: one on the Illumina OmniExpres array, one on the Illumina PsychChip array, and one on the Illumina GSA array. These batches were imputed to HRC using the Michigan imputation server (Das et al. 2016).

WLS: Participants were genotyped with the Illumina HumanOmniExpress array, with 713,014 SNPs (Herd, 2016). These data were then imputed to the Haplotype Reference Consortium (HRC) v.1.1 European reference panel (Haplotype Reference Consortium, n.d.).

A.2 Quality control (QC) of the genetic data

Our genetic QC follows the following steps, which are performed in Plink2 Chang et al., 2015:

1. We only include SNPs with an INFO score above 99%.
2. We only include HapMap3 SNPs.

3. We only include SNPs with genotyping rates > 99% (–geno 0.01), individuals with missingness rates <1% (–mind 0.01), SNPs with minor allele frequencies > 1% (–maf 0.01) and a Hardy-Weinberg p-value below 1e-6 (–hwe 1e-6).

Cohort-specific details:

EstBB: In the Estonian Biobank, all filters were applied with Plink and Plink2. This left 682,380 SNPs for the IBD calculation.

GS: In the Generation Scotland cohort, all filters were applied with Plink and Plink2. This left 837,433 SNPs for the IBD calculation.

MoBa: In MoBa, genotyping rates, missingness rates, minor allele frequencies, and Hardy-Winberg p-values filters were applied using genotype data from the MoBA quality control pipeline Corfield et al., 2024 and repeated using Plink. Filters on HapMap3 SNPs were applied using Plink. Only SNPs with INFO scores above 99% across all imputation batches were kept. This left 3,808,036 SNPs for the IBD calculation.

STR: We filter on imputation R-squared¿99% and otherwise apply all other filters as specified, using Plink and Plink2.

UKB: To filter on INFO score, we subset to variants with INFO > 99% by creating a list of rsIDs with failing INFO scores. We use Plink2 to apply the other filters. This left 760,326 SNPs for the IBD calculation.

WLS: In the Wisconsin Longitudinal Study, all filters were applied with Plink and Plink2. This left 980,097 SNPs for the IBD calculation.

A.3 Sibling Identification

Our general process to identify and code pairs of full siblings follows these steps:

1. Check if siblings are explicitly identified in the dataset.

2. If siblings are not explicitly identified in the dataset, use KING's –kingship flag (Manichaikul et al., 2010) to identify siblings. Pairs are coded as siblings if they had a kinship coefficient between $2^{-2.5}$ and $2^{-1.5}$ as well as $IBS_0 > 0.0012$.

3. If siblings are explicitly identified in the dataset, use KING's –kingship flag to verify their sibling status.

Cohort-specific details:

EstBB, *GS*, and *UKB*: Siblings are not explicitly identified in the UK Biobank, and were identified using KING Manichaikul et al., 2010. After this, we manually assign each sibship (i.e., the siblings with the same biological mother and father) an FID in order to provide a pedigree file to SNIPAR (Young, Nehzati, et al., 2022).

MoBA: Adult siblings were identified using KING and cross checked with kinship registry data.

STR: Siblings (in this case DZ twins; MZ twins were dropped) are explicitly identified in the STR since it is a twin register. Sibling status was verified using KING.

WLS: Siblings are explicitly identified by the WLS. The WLS also provides kinship coefficients for sibling pairs, which they calculated using King as described above Herd, 2016 and provided for all related pairs; we use these provided kinship coefficients to verify the sibling status of all sibling pairs identified by the WLS.

A.4 Sample Inclusion

Individuals are included in the sample (1) if they are a member of at least one sibling pair and (2) if both they and at least one of their siblings pass a standard set of checks for working with genetic data. These checks include the following:

1. Reported and genetic sex match.
2. No extreme levels of heterozygosity or missingness.
3. No sex chromosome aneuploidies.
4. European ancestry individuals only.

When there are more than one pair of siblings with the same mother and father, we include all the pairs separately (as described below, we cluster standard errors at the family level in the DeFries-Fulker regression).

Cohort-specific details:

EstBB: EstBB excludes individuals who failed these checks (with the exception of the ancestry check) from their provided data. As part of the pre-imputation sample QC, EstBB used bigsnpr to exclude all individuals without ancestry from Eastern Europe, Northwestern Europe, or Finland. For more details see Kuznetsov et al., 2023.

GS: GS excludes individuals who failed these checks (with the exception of the ancestry check) from their set or provided genotypes. As part of the pre-imputation sample QC, GS excludes all individuals who are more than 6 standard deviations away from the mean of the first two principal components of the genetic relatedness matrix. For additional details, see Amador et al., 2015.

MoBa: All checks are part of the quality control pipeline performed on MoBa genotype data used in this study Corfield et al., 2024.

STR: These checks were part of the pre-imputation QC procedure for all batches. Additionally, individuals with a missing genotype rate of more than 1% were dropped post-imputation.

UKB: All of these checks are performed using UKB-provided variables. (Reported and genetic sex match: Field 22001 for genetic sex and Field 31 for reported sex; no extreme levels of heterozygosity or missingness: Field 22027; no sex chromosome aneuploidies: Field 22019; and European ancestry individuals only: Field 22006). The UKB-provided ancestry variable was determined by first plotting the PCs of the UK Biobank sample. Within the subset who self-reported British ancestry, Bayesian outlier detection methods were used to identify the largest cluster of similar ancestries, which were then coded as having genetic British ancestry (Bycroft, Freeman, Petkova, et al., 2018).

WLS: The WLS excludes individuals who failed these checks (with the exception of the ancestry check) from their set or provided genotypes. As part of the pre-imputation sample QC, the WLS excludes all individuals who are not of European genetic ancestry. For additional details, see Herd, 2016.

A.5 IBD Calculation

IBD calculation is performed with the SNIPAR script ibd.py (Young, Nehzati, et al., 2022), after following the sample and genetic QC laid out above. We used a genotyping error probability of 0.0001. For the UKB and WLS, we used 20 threads and 5 batches.

This outputs an IBD file for each chromosome, which contains info on which ranges of base pairs are IBD0, IBD1, or IBD2 for each sib pair. For each sib pair, we use these to get the overall IBD for each chromosome using the following formula:

$$\hat{IBD}_c = \frac{\sum_i l^2_{s,c} + 0.5 * \sum_i l^1_{s,c}}{\sum_i l^2_{s,c} + \sum_i l^1_{s,c} + \sum_i l^0_{s,c}}, \tag{10}$$

where, $l^k_{s,c} \in \{0, 1, 2\}$ represents the length of segment s of chromosome c that is IBD k.

Then, we can take the weighted average of the chromosomal IBD proportions to get the genome-wide (GW) value for each sib pair. We weight by chromosome length, using chromosome lengths l_c given by Visscher et al., 2006:

$$\hat{\pi} = \hat{IBD}_{GW} = \frac{\sum_c l_c * \hat{IBD}_c}{\sum_c l_c} \tag{11}$$

B Methods: non-genetic data

B.1 Preparing the outcome variables

The Supplementary Table provides information on the phenotypes used in this analysis.

Since the mean and variance of each outcome typically vary as a function of sex, birth year, and age at measurement, and to deal with multiple measurements of the same outcome, we follow the following high-level steps, separately for each outcome and in each dataset (the details vary across datasets, as described below):

1. Residualize each measure of the outcome by regressing the outcome measures on sex, birth year, and age at measurements. For outcomes that are logged, the log is taken before the residualization process.

2. Standardize the resulting residualized measures separately by sex. Further, when there is substantial variation in the age at observation or in birth year within a cohort, the measures are standardized using the sex-specific standard deviation calculated within smaller age or birth-year buckets. This is to ensure that the variance of each observation is constant across sexes, birth years, and ages at measurement, since our DeFries-Fulker regression implicitly assumes a constant variance across observations.

3. For individuals with multiple measurements, we take the average across the measurements of the resulting standardized measures for each individual.

    - For the outcomes "ever smoker" and "number of children ever born", for the individual with multiple measurements, we instead take the maximal response (pre-residualization) and then follow steps 1 and 2.

Each resulting outcome variable thus has zero mean and a variance that is approximately constant across sexes, birth year, and measurement ages in the outcome's analysis sample in each dataset.

Cohort-specific details:

EstBB: Each phenotype is residualized on birth year dummies, age dummies, and their interactions with sex in the full EstBB sample. The residuals from that regression are standardized using the sex-specific standard deviation calculated within 6-year birth-year buckets within the full Estonian Biobank sample. For height and BMI, we take the median residual instead of the mean, after filtering out height measurements that differed by more than 5 cm from the average across measurements.

GS: Each phenotype is residualized on birth year dummies, age dummies, and their interactions with sex in the full GS sample. The residuals from that regression are standardized using the sex-specific standard deviation calculated within 6-year birth-year buckets within the full sample. These buckets start at birth year 1920 and end in 1992.

MoBa: Each phenotype is residualized on birth year dummies, age dummies, and sex. The residuals from that regression are standardized using the mean and standard deviation within

each sex and birth year for variables where we have information from the whole Norwegian population (population-wide registry data), and sex and 6-year bins that start at birth year 1920 and end in 1992 in the estimation sample for phenotypes with a restricted sample (i.e. MoBa questionnaire data).

STR: Each phenotype is residualized on birth year dummies, age dummies (when applicable), and sex. The residuals from that regression are standardized using the mean and standard deviation within each sex and birth decade in the estimation sample for each phenotype.

UKB: Each phenotype is residualized on birth year dummies, age dummies, and their interactions with sex in the full UKB sample. The residuals from that regression are standardized using the sex-specific standard deviation calculated within 6-year birth year buckets within the full UK Biobank sample. These buckets are as follows: 1935-1940, 1941-1946, 1947-1952,1953-1958, 1959-1964, and 1965-1970.

WLS: This residualization process was only performed with sex and birth year dummies, since age-at-survey was not collected, and standardized within 1 year age buckets.

B.2 Coding of occupational prestige

Occupational prestige was coded using the SIOPS scale Treiman, 1977; Ganzeboom and Treiman, 1996. To use the SIOPS scale, ISCO job codes are needed for each cohorts.

EstBB: In the Estonian Biobank, we have the major units (i.e., the highest-level category) of each participant in the ISCO-08 system, and this can be easily linked to SIOPS with Nicole Schwitter, n.d.

MoBa: The MoBa is linked to registry data with occupational codes procured and provided by Statistics Norway [ref. 14 in the spreadsheet]. Norwegian occupational codes (STYRK-98 and STYRK-08) were first converted to ISCO-88 codes and ISCO-88 codes were converted to SIOPS [Ref. 20 in the spreadsheet].

STR:

UKB: In the UKB, job codes are only readily available in the categorization of the UK SOC 2000 scale. Unfortunately, we could not find a crosswalk between that scale and SIOPS.[8] Instead, we link the UK SOC 2000 codes to the 1988 version of ISCO, using codes from CAMSIS, n.d. accessed using the Internet Archive. From there, we linked the ISCO-88 codes to SIOPS using Maarten Hermans, n.d.

[8]The UK SOC 2000 scale could be re-coded to the UK SOC 2010 scale, and there are mappings between the latter and SIOPS; however, using the 2010 scale would have required special permission to access the raw job data.

C Assortative mating

The results we present in Section 5 assume parents select mates randomly. The presence of assortative mating, when individuals with similar phenotypes mate, will bias estimates of h^2 and σ_C^2. Under assortative mating, h^2 is instead given by

$$\hat{h}^2 = \frac{\hat{\beta}_2}{1-r}, \tag{12}$$

with c given by

$$\hat{\sigma}_C^2 = \hat{\beta}_1 - \frac{r}{1-r}\hat{\beta}_2. \tag{13}$$

In these equations, r gives the degree of assortative mating. Tables E.3 and E.4 present results for $r \in \{0.1, 0.3, 0.5\}$.

D Power Calculations

We conducted power simulations to select the most powerful method to estimate heritability based on sibling IBD data. The methods under consideration are presented in Table D.1.

[OLD FROM MAIN TEXT] Several variants of the sib-regression given by Equation 4 have been proposed in the literature. These include two Haseman-Elston regressions, a variance component model, and the DeFries-Fulker regression (Haseman and Elston, 1972; Sham and Purcell, 2001; Xu et al., 2000; DeFries and Fulker, 1985). In Online Appendix Section D, we conduct simulations to compare the efficiency of the 4 different variants. We find that, for reasonable values of h^2, the DeFries-Fulker regression is more powerful than Haseman-Elston regression when calculating heritability from realized relatedness. Its power is roughly equal to VCA, but DeFries-Fulker is much faster than VCA and it is difficult to meta-analyze VCA estimates across disparate cohorts[9].

Name	**Model**	h^2 **Parameter**
Double-entry Defries-Fulker Regression	$y_{1,j} = \beta_0 + \beta_1 y_{2,j} + \beta_2 y_{2,j}\pi_j + \epsilon$	$h^2 = \beta_2$
Variance component analysis	$var(Y) = \sigma_a^2 + \sigma_c^2 + \sigma_e^2$	$h^2 = \sigma_a^2$
Squared Difference Haseman-Elston Regression	$(Y_1 - Y_2)^2 = \beta_0 + \beta_1 \pi_{1,2}$	$h^2 = -\frac{\beta_1}{2}$
Cross-product Haseman-Elston Regression	$Y_1 Y_2 = \beta_0 + \beta_1 \pi_{1,2}$	$h^2 = \beta_2$

Table D.1: Heritability Estimators

We simulated a trait from the following model (Equations 14-17), using real IBD relatedness from sibling pairs in the UK Biobank.

$$Y_i = A_i + C_i + E_i \tag{14}$$

$$cov(A_1, A_2) = \pi_{1,2} h^2 \tag{15}$$

$$cov(C_1, C_2) = I(\pi_{1,2} > 0) C^2 \tag{16}$$

$$A_i, C_i, E_i \sim N(0, 1) \tag{17}$$

We used a grid of "ground truth" heritability values, common environment effect values, and sample sizes. To estimate the power at sample sizes bigger than that of the UKB, we fit a normal distribution to the observed IBD relatedness values, and sampled from that distribution. We found that DeFries-Fulker regression and the variance component analysis were consistently higher powered than either Haseman-Elson specification. Results for $h^2 = 0.3$ and $\sigma_C^2 = 0.1$ are presented in Table D.2 and Figure D.1. Since DeFries-Fulker is easier to meta-analyze across samples, we picked that specification for our main analysis.

[9]It is difficult to retrieve standard errors for this model. See Visscher, 2006 for more discussion.

N	Df-F	HE-CP	HE-SD	VCA
10000	0.34	0.31	0.29	0.36
20000	0.53	0.46	0.44	0.55
40000	0.76	0.70	0.68	0.79
60000	0.90	0.84	0.82	0.92

Table D.2: This table presents estimated power for the 4 methods under consideration, for 4 different sample sizes. The assumed true heritability in this simulation is $h^2 = 0.3$, and the assumed environmental component is $\sigma_C^2 = 0.1$.

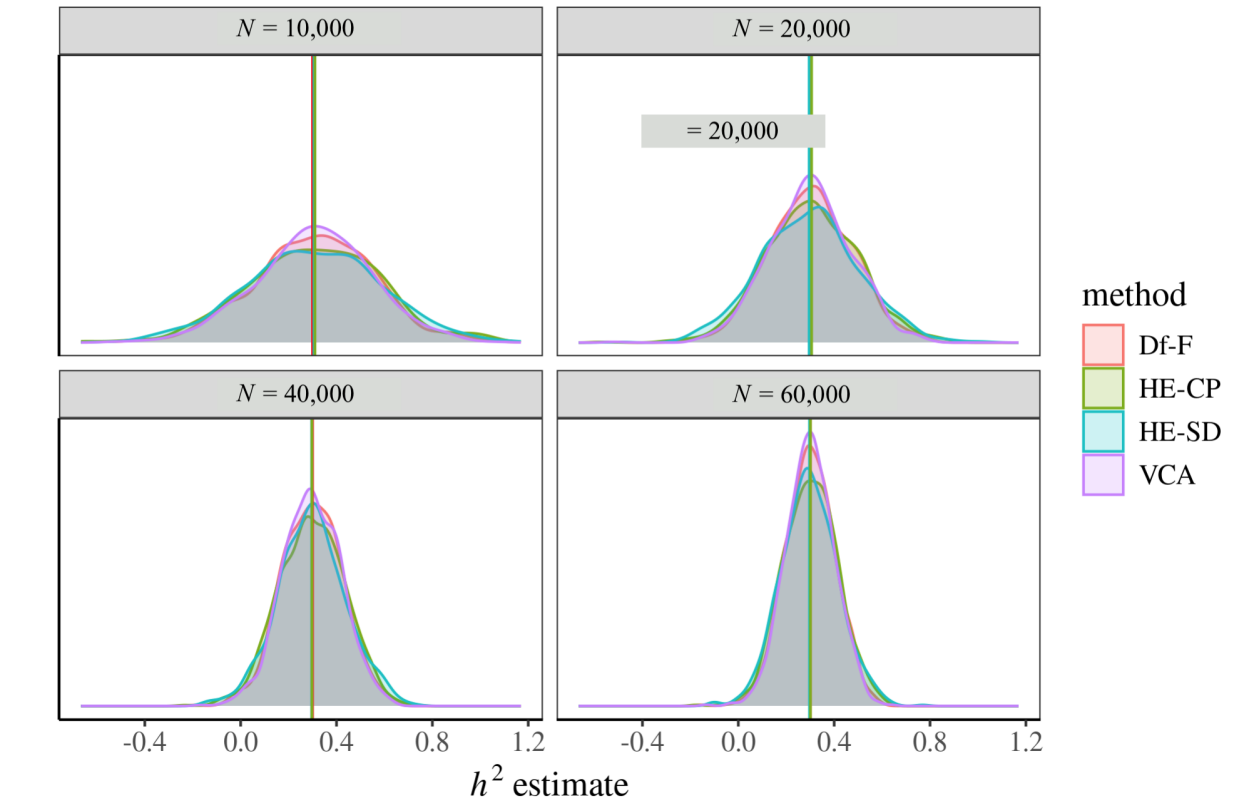


Figure D.1: Densities of h^2 estimates computed using 4 different methods on simulated data
Notes: This plot presents heritability estimates estimated using DeFries-Fulker regression, variance component analysis, Haseman-Elston cross product regression, and Haseman-Elston squared difference regression, for a grid of sample size values. The assumed true heritability in this simulation is $h^2 = 0.3$, and the assumed environmental component is $\sigma_C^2 = 0.1$. The true heritability from which the data is simulated is given by a red line, and the mean estimate is given by a vertical line in the relevant color. Where the red line is not visible, the mean estimate and the true value are identical.

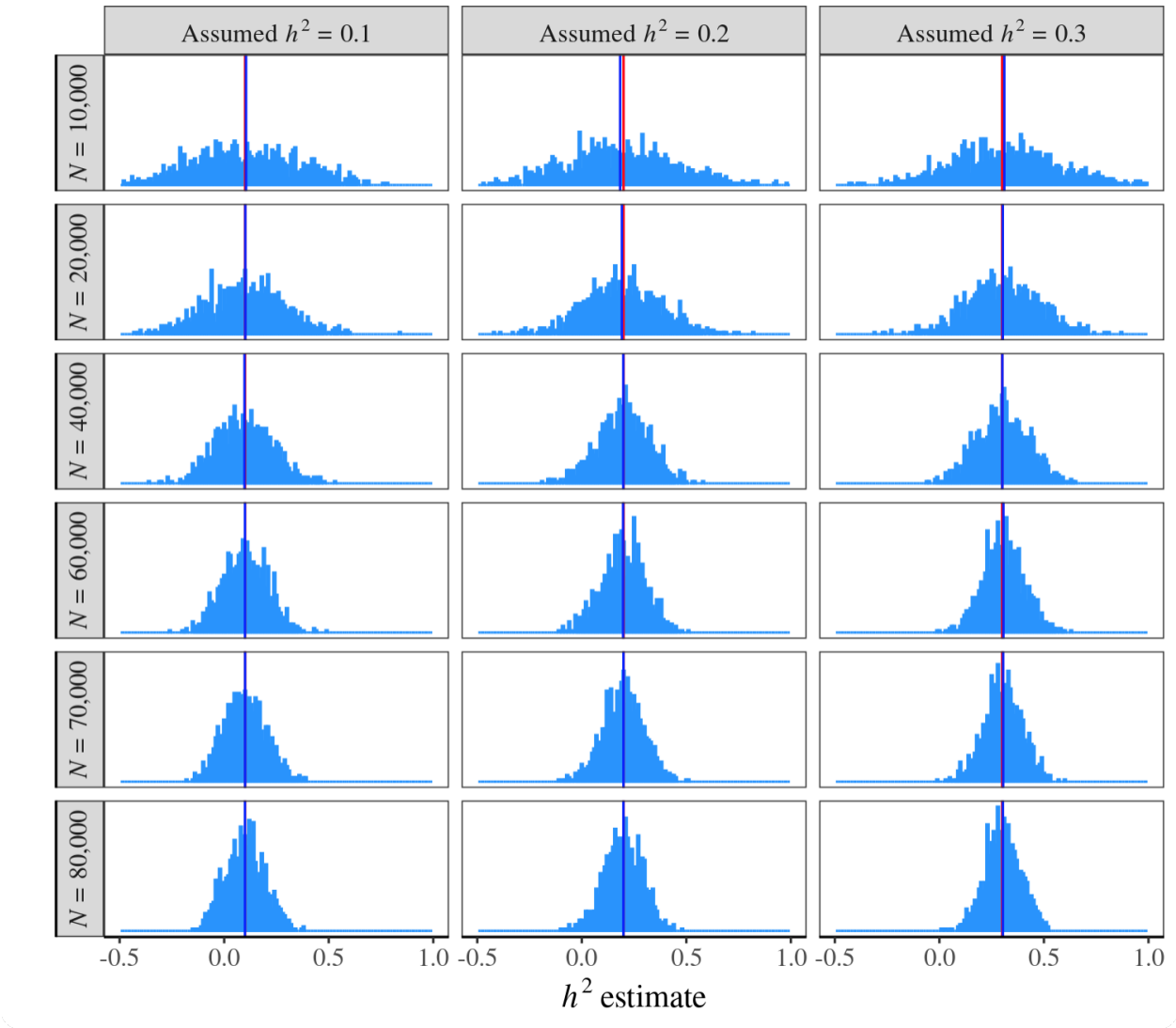


Figure D.2: Densities of h^2 estimates computed the Defris-Fulker method

Notes: This plot presents heritability estimates estimated using DeFries-Fulker regression for a grid of sample size and true heritability values. The true heritability from which the data is simulated is given by a red line, and the mean estimate is given in dark blue. Where the red line is not visible, the mean estimate and the true value are identical. All simulations assume $\sigma_C^2 = 0.1$

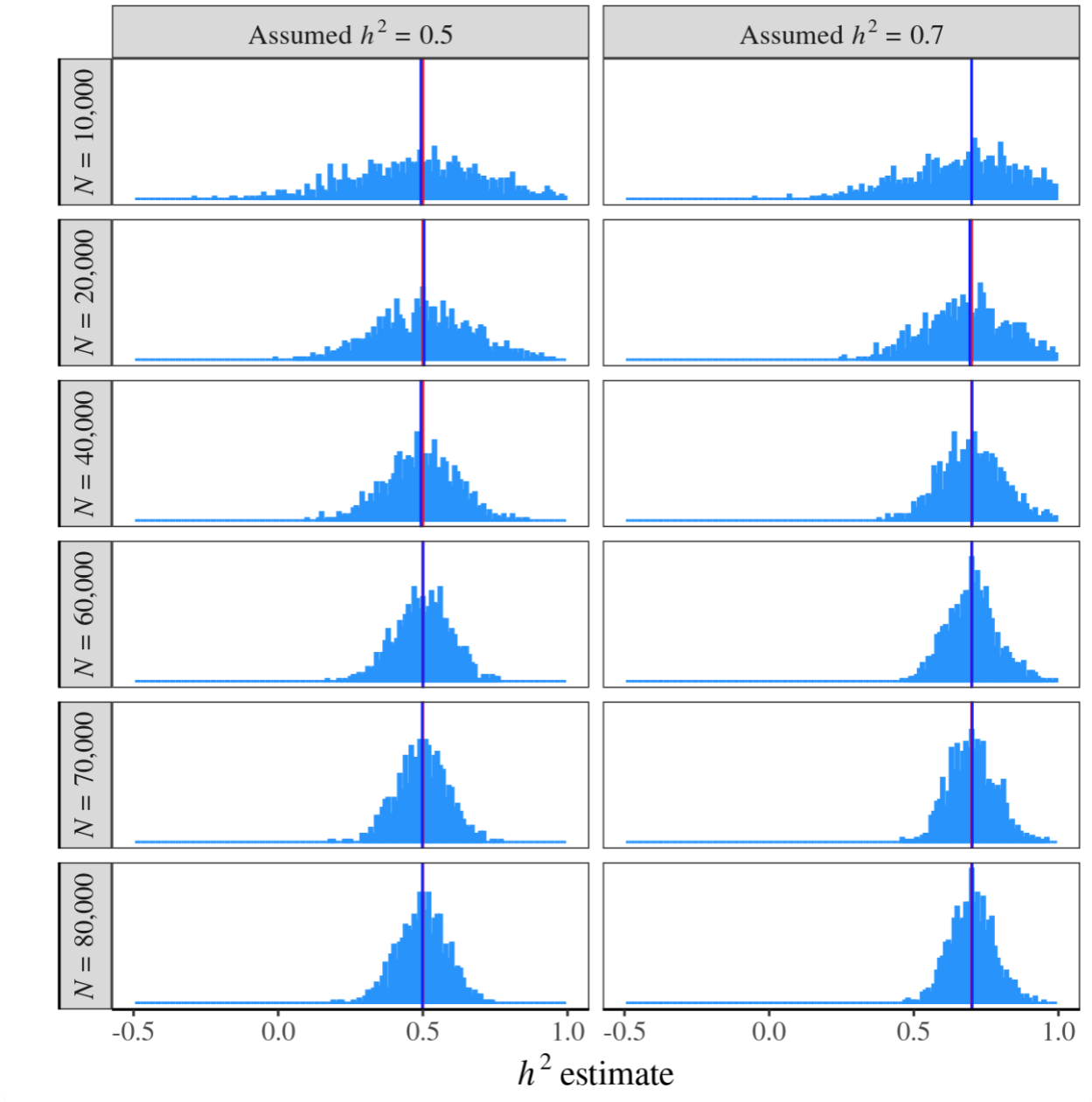


Figure D.3: Densities of h^2 estimates computed the Defris-Fulker method

Notes: This plot presents heritability estimates estimated using DeFries-Fulker regression for a grid of sample size and true heritability values. The true heritability from which the data is simulated is given by a red line, and the mean estimate is given in dark blue. Where the red line is not visible, the mean estimate and the true value are identical. All simulations assume $\sigma^2_C = 0.1$

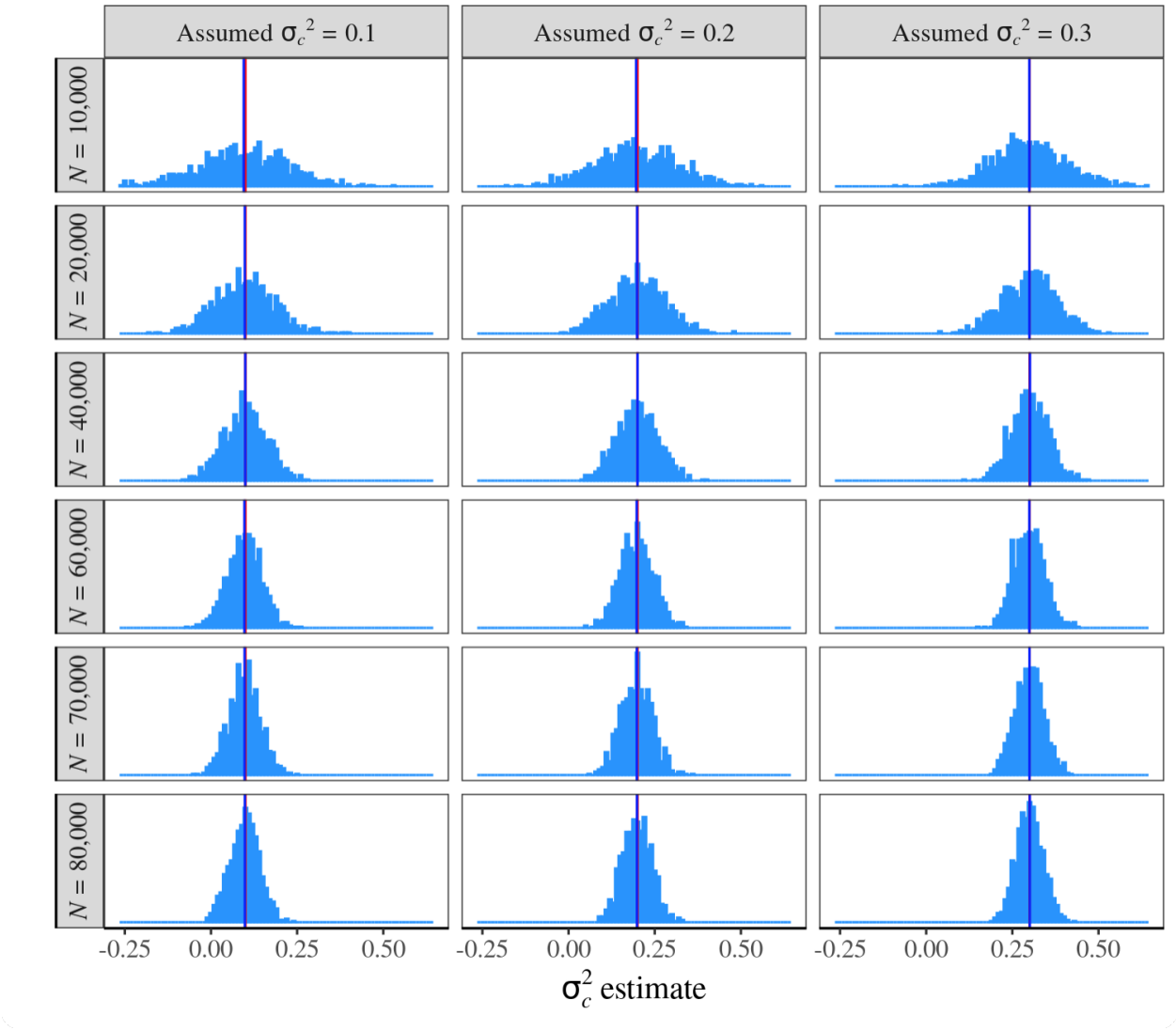


Figure D.4: Densities of σ^2_C estimates computed the Defris-Fulker method

Notes: This plot presents σ^2_C estimates estimated using DeFries-Fulker regression for a grid of sample size and true σ^2_C values. The true σ^2_C from which the data is simulated is given by a red line, and the mean estimate is given in dark blue. Where the red line is not visible, the mean estimate and the true value are identical. All simulations assume $h^2 = 0.3$

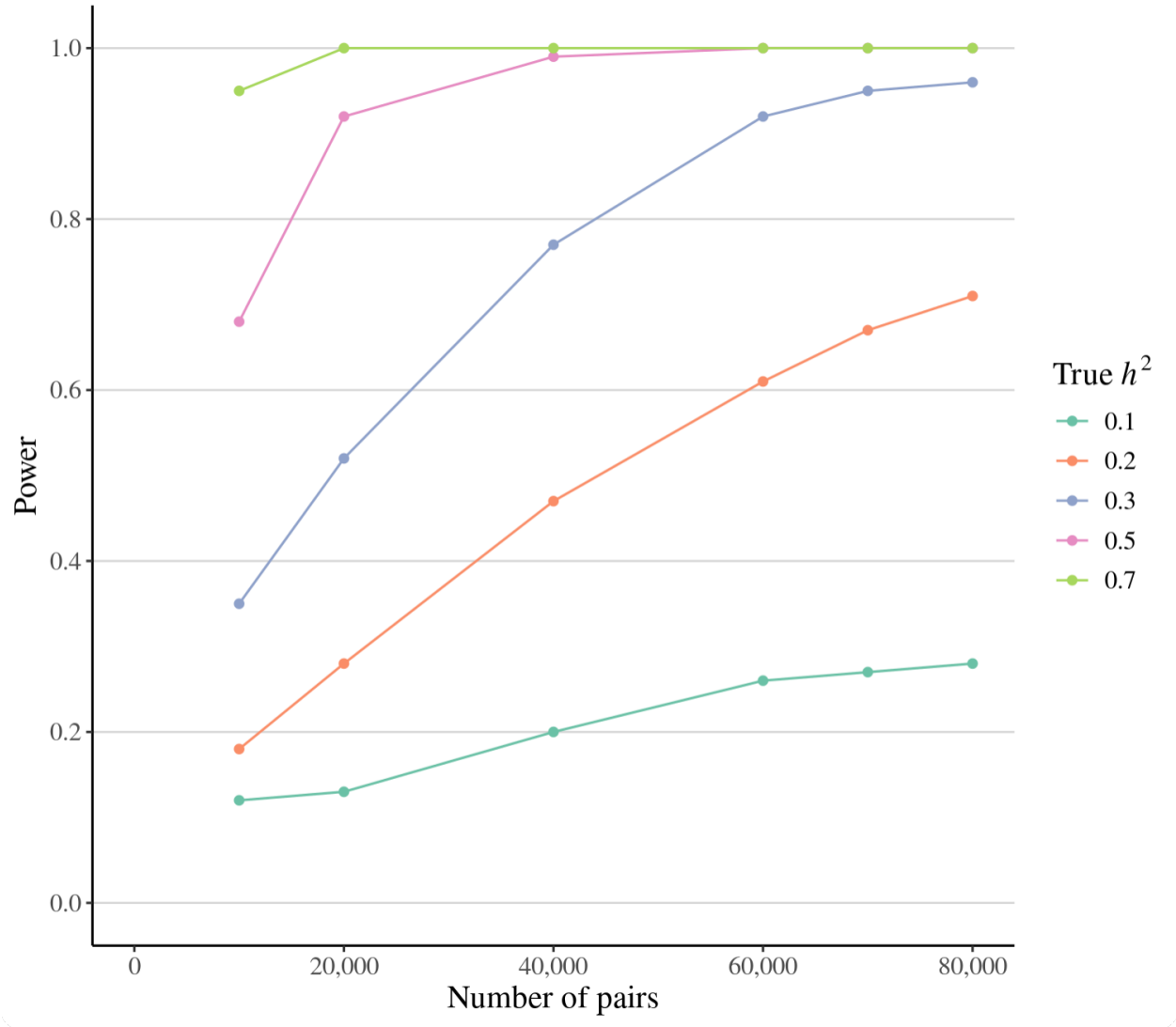


Figure D.5: DeFries-Fulker power at various h^2s and sample sizes.

Notes: This plot presents the power of estimates estimated using DeFries-Fulker regression for a grid of sample size and true h^2 values. All simulations assume $\sigma_C^2 = 0.1$

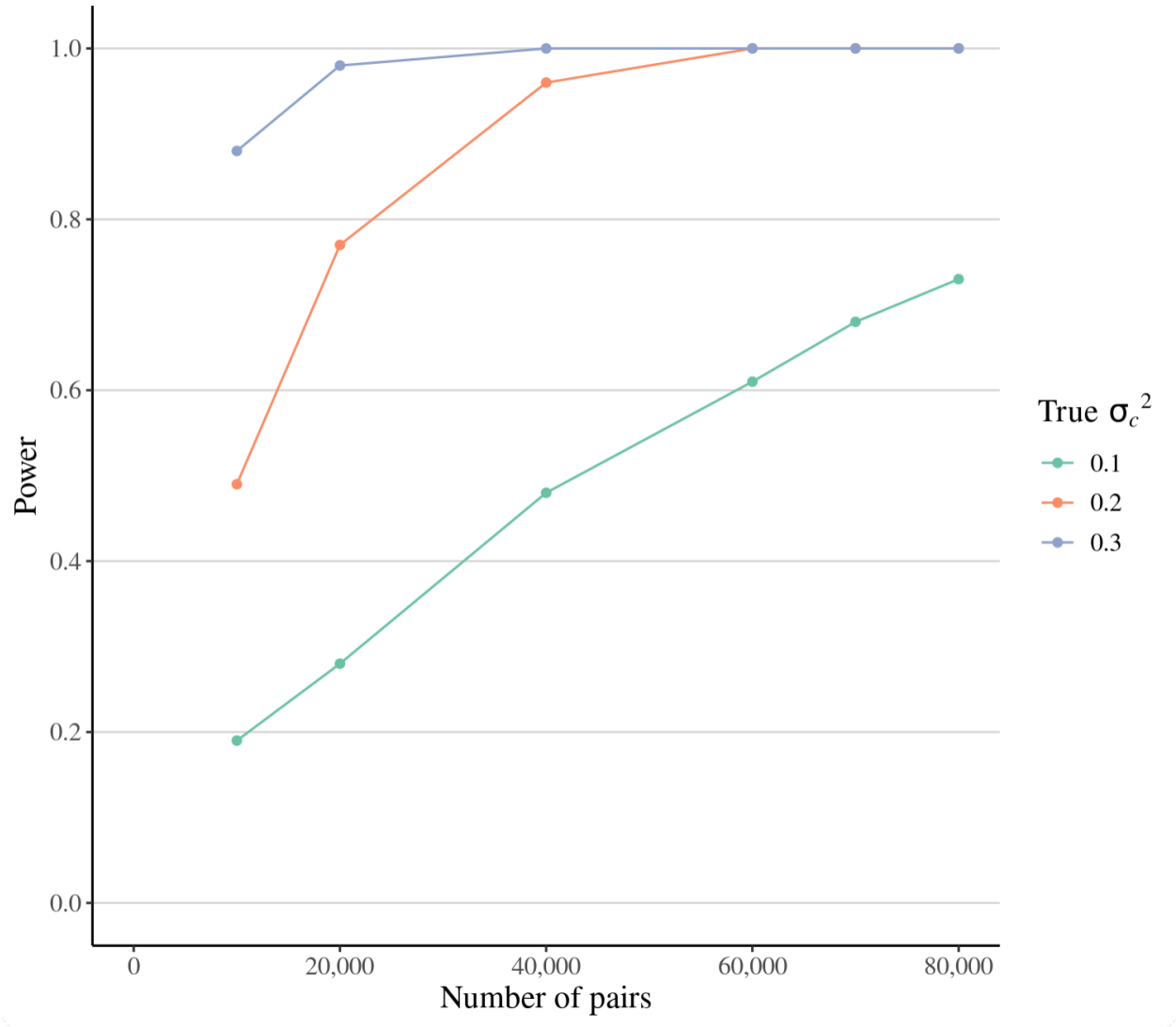


Figure D.6: DeFries-Fulker power at various σ_C^2s and sample sizes.

Notes: This plot presents the power of estimates estimated using DeFries-Fulker regression for a grid of sample size and true h^2 values. All simulations assume $h^2 = 0.3$

E Supplementary tables

Table E.1: Construction details for the 15 outcome variables

Outcome	Definition	Dataset	*Npairs*	Construction details	Treatment of repeated measures	Citations	Dataset codes
Panel A. Cognitive performance and education							
Cognitive performance	*Score on a test of cognitive performance*	EstBB	--	--	--	--	--
		GS	8,044	Following [2], cognitive performance was measured as the first PC of score on a logical memory test, (total of immediate and delayed), digit symbol test, and a verbal fluency test. Each subtest was variance normalized before performing principal component analysis.	NA	[10], [2]	g
		MOBA	--	--	--	--	--
		STR	687	Men in our sample were required by Swedish law to participate in military conscription around the age of 18; as part of the drafting procedure, they had to complete a test of cognitive ability, which consisted of four subtests: logical, verbal, spatial, and technical.	NA		The data comes from the Swedish National Service Administration.
		UKB	6,913	This cohort's cognitive performance score is the raw score on a 13-question fluid intelligence inventory.	Average of each standardized, residualized measure*	[5],[7]	20191, 20016
		WLS	1,693	Centile rank of the student's performance on the Henmon-Nelson test.	NA	[6]	ghncr_bm
Educational attainment	*Number of years of education*	EstBB	35,124	Data aggregated from national databases and self-reported questionnaires, preferring highest degree ever reported. The education levels were transformed into years of educational attainment following based on ISCED 2011 levels, similarly to [8].	NA	[15]	person_portrait_education_code,person_portrait_education_acode, person_portrait_education_group, education_pers_acode
		GS	7,968	Participants were asked "How many years altogether did you attend school or study full-time," and these answers were transformed into years of education following [2]. Participants were also asked to list the academic qualifications (e.g., "A levels/AS levels or equivalent"). These were transformed into years of educational attainment following [2; see SI Section 1] based on ISCED 1997 levels, for individuals who did not respond to the years of education question.	NA	[10],[2]	qualification
		MOBA	11,378	Latest registry data on ISCED category transformed to years of education in Norway.	NA	[20]	edu_years
		STR	5,668	Recoded from Swedish Educational Nomenclature standard to years of education. Available annually from 1990-2018; observations before age 25 were filtered out.	Highest measure used.	[13]	
		UKB	18,963	Participants were asked to list the academic qualifications (e.g., "A levels/AS levels or equivalent"). These were transformed into years of educational attainment following [2; see SI Section 1] based on ISCED 1997 levels.	Average of each standardized, residualized measure*	[2],[5]	6138
		WLS	623	The WLS directly provides years of education.	NA	[6]	edeqyr

(Continues)

Table E.1 (continued): Construction details for the 15 outcome variables

Outcome	Definition	Dataset	*Npairs*	Construction details	Treatment of repeated measures	Citations	Dataset codes
Panel B. Labor market							
Employed	*Binary variable indicating whether participant has a job*	EstBB	11,342	Participants are coded as "employed" ("1") if they reported being employed on the Estonian Biobank survey, and as unemployed ("0") otherwise. Only observations taken between ages 30 and 60 were used.	NA	[15]	work_is_working
		GS	--	--	--	--	--
		MOBA	11,367	Yearly registry data. Only observations taken between ages 30 and 60 were used.	Average of each standardized, residualized measure between ages 35-45.	[21]	
		STR	4,656	Participants were coded as unemployed in a given year if they earned income from wages or active business of less than 25% of the median income in that year. Only observations taken between ages 30 and 60 were used.	Average of each standardized, residualized measure*	[13]	
		UKB	9,473	Participants are coded as "employed" ("1") if they reported being "in paid employment or self-employed" on the UK Biobank survey, and as unemployed ("0") otherwise. Only observations taken between ages 30 and 60 were used.	Average of each standardized, residualized measure*	[5]	6142
		WLS	460	Participants are coded as employed ("1") if they reported being "Employed-- civilian" or "Employed--military" on the WLS survey, and unemployed ("0") otherwise. Only observations taken between ages 30 and 60 were used.	NA	[6]	z_lfstat
Log family income	*Log of total annual income earned by all household members*	EstBB	3,241	Participants were asked, "What is your personal average monthly income (after taxes, in euros)?" Only observations taken between ages 30 and 60 were used. For individuals who reported earning $0, we took log(1).	NA	[16]	pt_life_attitude_income_household
		GS	4,639	Participants were asked, "What is the average total income before tax of your entire household." The options were 1 = less than £10;000; 2 = between £10;000 and £30;000;3 = between £30;000 and £50;000; 4 = between £50;000 and £70;000 ;5 = more than £70;000 ; 6 = prefer not to answer; not known. For the <10,000 category, participants were coded as earning 10,000, and for the >70,000 category, participants were coded as earning 70,000. Only observations taken between ages 30 and 60 were used.	NA		income
		MOBA	11,354	Yearly registry data (population-wide income data from tax registers and other government registers cross-checked and produced by Statistics Norway) grouped by household (population registry based on administrative government data procured and provided by Statistics Norway). Only observations taken between ages 30 and 60 were used	Average of each standardized, residualized measure. Between 2005-2023.	[22] [23]	log_inc_hh
		STR	--	--	--	--	--
		UKB	7,748	Family income was coded as the midpoint of all categorical responses to a question asking what is your "average total household income before tax." The options were, "<18,000," "18,000-30,999," "31,000-51,999," "52,000-100,000," ">100,000," ""Do not know," and "prefer not to answer," all denominated in pounds. For the "<18,000" category, individuals were coded as earning 12,000 per year, while in the ">100,000" category, individuals were coded as earning "120,000" per year. Only observations taken between ages 30 and 60 were used.	Average of each standardized, residualized measure*	[5]	738
		WLS	--	--	--	--	--

(Continues)

Table E.1 (continued): Construction details for the 15 outcome variables

Outcome	Definition	Dataset	*Npairs*	Construction details	Treatment of repeated measures	Citations	Dataset codes
Panel B. Labor market (continued)							
Log occupational income	*Log of annual income imputed based based on occupational code*	EstBB	7,938	Imputed using data on hours worked, participant occupation codes, and external Estonian wage data, following the algorithm in and using code from [17]. Most recent job code, or most recent job code from before age 65 in the case of individuals over 65, were used.	NA	[15], [17]	curr_income
		GS	--	--	--	--	--
		MOBA	10,455	Occupational code at age 35-45. Mean log wages adjusted for CPI for each sex and ISCO-08 code in the whole population that year.	NA	[21] [22]	log_occ_inc
		STR	3,130	Imputed using full-population average income per ISCO-88 category and participant occupation code. Only years employed after age 30 used.	Average of each standardized, residualized measure*	[13]	--
		UKB	9,031	Imputed using data on hours worked, participant occupation codes, and external British wage data, following the algorithm in and using code from [4]. Most recent job code, or most recent job code from before age 65 in the case of individuals over 65, were used (as with the whole UK Biobank sample, all participants were 40 years old or older at the time of observation).	NA	[4],[5]	22617,22602,22603,20277
		WLS	--	--	--	--	--
Occupational status	*Occupational prestige on the SIOPS scale*	EstBB	2,183	Occupational status was coded by translating from the ISCO-08 system to SIOPS using 1 digit. Participants reported their major occupational category, which was then coded as ISCO-08 + 1 (so armed forces, category 0, was option 1, and so on).	NA	[15],[16]	work_current_occupation_code,work_current_occupation_name,work_main_occupation_code,work_main_occupation_name
		GS	--	--	--	--	--
		MOBA	10,615	Average SIOPS occupational prestige between ages 35-45. Scale from 0-100.	NA	[21] [27]	avg_occ_prestige
		STR	3,349	Occupational status was coded by translating four-digit ISCO-88 to SIOPS.	Average of each standardized, residualized measure*	[13]	
		UKB	8,932	Occupational status was coded by translating the UKB's SOC 2000 codes to the ISCO-88 system, and from that to SIOPS using up to 4 digits*.	NA	[4],[5]	22617,20277
		WLS	--	--	--	--	

(Continues)

Table E.1 (continued): Construction details for the 15 outcome variables

Outcome	Definition	Dataset	*Npairs*	Construction details	Treatment of repeated measures	Citations	Dataset codes
Panel C. Risk tolerance & risky behaviors							
Ever smoker	*Binary variable indicating whether participant ever smoked*	EstBB	37,250	Converted from aggregate smoking status estimated by the EstBB from self-report data and health registries. Former and current smoker are coded as 1, Never smoker as 0 (and "Unknown"'s are dropped).	NA	[15]	person_portrait_last_smoking_status_name
		GS	7,870	Participants were asked "Have you ever smoked tobacco?" Participants were coded as 1.0 if yes, 0.0 if no, and NA otherwise.	NA	[10]	ever_smoke
		MOBA	1,517	Self-report for both women and men at 15 weeks of pregnancy. Question 94: "Have you ever smoked?" 1 coded as yes.	NA	[24]	AA1355, F214
		STR	3,091	Participants of the 1973 STR survey and the SALT survey were asked to indicate if they are, or ever were, smokers.	NA	[12]	NYRA423,SMOK_SNUFF_NEJ_INTE_ENS_PROV,SMOK_SNUFF_JA_BARA_PROVAT,SMOK_SNUFF_ROKT_DA_OCH_DA,SMOK_SNUFF_ROKT_REGELBUNDET,SMOK_SNUFF_FESTROKT,SMOK_SNUFF_ROKER_DA_OCH_DA,SMOK_SNUFF_ROKER_REGELBUNDET,SMOK_SNUFF_FESTROKER
		UKB	19,092	Coding (following [1]): "Coded as 1 if a respondent reported that they were a current or previous smoker and 0 if they reported never smoking or only smoking once or twice."	Use maximum measure	[1], [5]	20160
		WLS	1,868	Participants were asked if they have ever smoked regularly, or if they smoke regularly now. They are asked this in 3 waves. If they ever report smoking regularly, they are coded as an ever smoker.	Use maximum measure	[6]	mx012rer,mx013rer,ix012rer,ix013rec,jx012rer,jx013rec
Log cigarettes per day	*Log of "1" plus number of cigarettes smoked per day when smoking*	EstBB	19,534	Estimated tobacco units from multiple tobacco questions from the health questionnaire ." The logarithm of "1" plus that measure was used.	Average taken before residualization.	[15]	smoking_cigarettes_per_day_usually;
		GS	7,607	For individuals who do not report ever smoking, we code CPD as 0. Else, following [1], we use the max number of cigarettes/pipes/cigars reported smoking in one day. The logarithm of "1" plus that measure was used.	NA	[10]	cigs_day, cigars_day, ever_smoke
		MOBA	749	Women: Self-report last three months before pregnancy. Men: Self-report last six months before pregnancy. Q96 from 15 weeks of pregnancy questionnaire: "If daily, how many cigarettes per day?" Number 0-99.	NA	[24]	AA1358, FF216
		STR	1,119	Participants who indicated that they were smokers were also asked to indicate cigarettes per day across several surveys. Two were used: the 1973 STR survey, and the SALT survey (later surveys are also available but were not used, as they surveys cohorts where other modes of tobacco consumption (snus) are the norm). The logarithm of "1" plus that measure was used.	Average of each standardized, residualized measure*	[12]	NYRA429, TOB_AM_SMOKE_ANT_CIG_DAG_V
		UKB	10,234	For individuals who do not report ever smoking, we code CPD as 0. Else, following [1], we use the max number of cigarettes/pipes/cigars reported smoking in one day. The logarithm of "1" plus that measure was used.	Average of each standardized, residualized measure*	[1], [5]	2887,3456,6183,20160
		WLS	1,861	Participants were asked how many packs of cigarettes they smoke per day when smoking regularly. We define CPD as (number of packs) x 20, 20 being the number of cigarettes in a standard pack. The logarithm of "1" plus that measure was used.	NA	[6]	jx015rer

(Continues)

Table E.1 (continued): Construction details for the 15 outcome variables

Outcome	Definition	Dataset	*Npairs*	Construction details	Treatment of repeated measures	Citations	Dataset codes
Panel C. Risk tolerance & risky behaviors (continued)							
Log drinks per week	*Log of "1" plus number of alcoholic beverages drunk per week*	EstBB	15,725	Estimated alcohol units from multiple alcohol questions from the health questionnaire. We multiplied this by 7 to calculate drinks per week. For individuals who report never drinking, we code this as 0. The logarithm of "1" plus the resulting measure was used.	NA	[15]	person_portrait_alcohol_unit_per_day
		GS	7,070	Participants were asked "During the past week, please record how many units of alcohol you have had." The logarithm of "1" plus that measure was used.	NA	[10]	units
		MOBA	--	--	--	--	--
		STR	--	--	--	--	--
		UKB	12,208	Coding (following [1]): "Constructed combining answers from multiple questions. First, respondents were asked how often they drink alcohol, and response options include 1) daily or almost daily; 2) three or four times per week; 3) once or twice per week; 4) one to three times per month; 5) special occasions only; and 6) never. Respondents who reported drinking once per week or more were asked how many glasses of various types of alcoholic beverages they consume per week. We use the sum of all alcoholic drinks per week in our drinks per week phenotype. Respondents who reported drinking less than once per week (one to three times per month or on special occasions only) were asked how many glasses of various types of alcoholic beverages they consume per *month*. For these respondents, we added the total number of drinks per month and divided by 4 to arrive at an approximated number of drinks per week. Respondents who reported never drinking were coded as 0." The logarithm of "1" plus that measure was used.	Average of each standardized, residualized measure*	[1], [5]	1558,1568,1578,1588,1598,1608,5364
		WLS	1,448	Participants were asked in up to 3 waves "how many drinks they had in the last month." This measure was divided by 4 to get drinks per week. The logarithm of "1" plus that measure was used.	Average of each standardized, residualized measure*	[6]	ru001re,ru028re,ru025re,ru026re,gua01re,gu028re,gu025re,gu026re,gu027re,hu101re,hu028re,hu025re,hu026re,hu027re
Risk tolerance	*The extent to which one is a risk taker (self-reported)*	EstBB	9,249	Participants were asked to evaluate the statement, "I take risks," with responses on a scale of 1-6. This result was coded as the z-score of the percentile rank of each respondent.	NA	[15]	extraversion20
		GS	--	--	--	--	--
		MOBA	--	--	--	--	--
		STR	1,422	Participants of the SALTY survey were asked 'How do you see yourself: are you a person who is prepared to take risks or do you try to avoid risks?' and 'Are you a person who is prepared to take financial risks or do you try to avoid financial risks?' with answer scales from 1-10 for both questions. The average of these two items was used.	NA	[12]	MORAL1, MORAL2
		UKB	18,166	Participants were asked: "Would you describe yourself as someone who takes risks? Yes / No." Coding (following [1]): "1 if response was "yes" and 0 if response was "no".	Average of each standardized, residualized measure*	[1], [5]	2040
		WLS	--	--	--	--	--

(Continues)

Table E.1 (continued): Construction details for the 15 outcome variables

Outcome	Definition	Dataset	*Npairs*	Construction details	Treatment of repeated measures	Citations	Dataset codes
Panel D. Health-related & other							
Number of children ever born	*Number of children ever born*	EstBB	2,954	Participants were asked: "How many children do you have (number)?" When individuals had data missing, then we added data from female health study question on number of live births. Male participants under age 50 and female participants under age 45 were dropped.	Maximum response at person level was taken from occasional double answers	[15],[16]	pt_life_attitude_childrens_count_name , missing data imputed from female_health_num_of_live_births
		GS	--	--	--	--	--
		MOBA	11,375	Number of children registered to the individual in the latest population registry data (based on government administrative data procured and provided by Statistics Norway). Male participants under age 50 and female participants under age 45 were dropped.	NA	[23]	n_barn
		STR	4,206	The multigeneration register was obtain number of children as of 2018. Male participants under age 50 and female participants under age 40 were dropped.	NA	[14]	
		UKB	16,394	Male participants were asked how many children they had ever fathered. Female participants were asked how many live births they had had. Male participants under age 50 and female participants under age 45 were dropped.	Use maximum measure		2734, 2405
		WLS	--	--	--	--	--
Self-rated general health	*Rating of general health (self-reported)*	EstBBB	21,851	Participants were asked a number of questions on their general health. We summed these responses together to form a level sum score, following [9]. We then took the negative z-score of the gender-specific percentage rank of each participant's LSS.	As we lacked dates for double answers, and therefore age at measurement, occasional double answers were averaged at person level.	[9], [15]	"health_movement_code", "health_selfcare_code", "health_common_activities_code", "health_pain_discomfort_code", "health_anxiety_depression_code"
		GS	7,318	Participants completed a 28 question battery on their general health (GHQ-28).	NA	[11]	ghq_total
		MOBA	2,321	Question from the World Health Organization's WHOQOL-BREF quality of life assessment: "How satisfied are you with health?" Responses are on a 5-point scale. Available only for women, 18 months after birth (Q102)	NA	[26]	EE672
		STR	1,332	Participants of the SALTY survey were asked to rate their state of health with a percentage from 0-100.	NA	[12]	SJALVUPPSKATTAD6
		UKB	19,106	Participants were asked "In general how would you rate your overall health?" Response choices included "excellent," "good," "fair," and "poor." We then normalized the responses for males and females separately	Average of each standardized, residualized measure*		2178
		WLS	--	--	--	--	--
Subjective wellbeing	*Subjective wellbeing (positive affect or life satisfaction; self-reported)*	EstBBB	9,076	The subjective wellbeing phenotype is coded as the first PCA of the following SWB questions: *I'm satisfied with my job* *I am satisfied with my choice of profession* *I'm satisfied with how I get along with my partner (spouse, partner)* *I am satisfied with my financial situation* *I'm satisfied with my place of residence* *I am satisfied with the way things are organized in our country*	NA	[15],[16],[18]	LS-
		GS	7,323	The subjective wellbeing phenotype is coded as the z-score associated with the percentile rank of the	NA	[10]	d3
		MOBA	9,494	Satisfaction with life scale (SWLS) at 15 weeks of pregnancy (Q133). 7-point likert scale to 5 statements. Z-score of average across the five statements was used.	NA	[24][25]	AA1527, AA1528, AA1529, AA1530, AA1531, FF269, FF270, FF271, FF272, FF273
		STR	1,437	Participants of the SALTY survey were asked 'Would you generally consider yourself to be: 1. Very happy, 2. Fairly happy, 3. Not very happy, 4. Not at all happy'.	NA	[12]	ATTITYD1
		UKB	2,618	Following [3], the subjective well being phenotype is coded as the z-score associated with the percentile rank of the participants response to the question, "In general how happy are you." Answer choices ranged from "[1] Extremely unhappy" to "[6] Extremely happy."	Average of each standardized, residualized measure*	[3], [5]	20458
		WLS	--	--	--	--	--

(Continues)

Table E.1 (continued): Construction details for the 15 outcome variables

Outcome	Definition	Dataset	*Npairs*	Construction details	Treatment of repeated measures	Citations	Dataset codes
Panel E. Anthropometric							
BMI	*BMI = weight / height^2 (metric units)*	EstBB	37,300	BMI computed with measured or self-reported height.	Height measurements that differed by more than 5cm from the average across measurements were filtered out. BMI was then computed for the remaining measurements, then standardized and residualized, and the median was taken.*	[15]	bmi_assembled_height
		GS	8,312	BMI computed with measured weight and height.	NA	[10]	bmi
		MOBA	9,502	BMI computed with self reported weight and height.	Median of each standardized, residualized measure*	[19]	bmi
		STR	2,929	BMI computed with measured weight and height.	NA	[12]	The data comes from several sources: military conscription data (age 18) for part of the male sample; health checkup data (varying ages) for some.
		UKB	19,126	BMI computed with measured weight and height.	NA	[5]	21001
		WLS	1,300	BMI computed with self reported weight and height.	NA	[6]	jx011rec
Height	*Participant's standing height*	EstBB	37,360	Height, measured or self-reported.	Height measurements that differed by more than 5cm from the average across measurements were filtered out. The remaining measurements were standardized and residualized, and the median was taken.*	[15]	bmi_assembled_bmi
		GS	8,355	Standing height (measured).	NA	[10]	height
		MOBA	9,732	Self-reported (or partner reported if self-report is missing) standing height.	Average of each standardized, residualized measure*	[19]	height
		STR	2,930	Standing height (measured).	NA	[12]	The data comes from several sources: military conscription data (age 18) for part of the male sample; health checkup data from TWINGENE (varying ages) for some.
		UKB	19,162	Standing height (measured).	Average of each standardized, residualized measure*	[5]	50
		WLS	1,318	Standing height (self reported).	NA	[6]	jx010rec

Table E.1 (continued): Construction details for the 15 outcome variables

Table E.1 references

[1] Karlsson Linnér, R., et al. (2019). 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*(2), 245–257.

[2] Okbay, A., et al. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics, 54*(4), 437–449.

[3] Okbay, A., et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics, 48*(6), 624–633.

[4] Kweon, H., et al. (2020). Genetic fortune: Winning or losing education, income, and health. Tinbergen Institute Discussion Paper, 2020-053/V. Available at: https://ssrn.com/abstract=3682041

[5] UK Biobank. (n.d.). UK Biobank touch-screen questionnaire: Final version. UK Biobank. Available at: http://biobank.ctsu.ox.ac.uk/crystal/docs/TouchscreenQuestionsMainFinal.pdf

[6] Herd, P., et al. (2014). Cohort Profile: Wisconsin Longitudinal Study (WLS). *International Journal of Epidemiology, 43*, 34–41.

[7] Davies, G., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications, 9*, 2098.

[8] Lee, J. J., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics, 50*(8), 1112–1121.

[9] Devlin, N., Parkin, D., & Janssen, B. (2020). *Methods for analysing and reporting EQ-5D data*. Cham: Springer.

[10] Smith, B. H., et al. (2013). Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). *International Journal of Epidemiology, 42*(3).

[11] Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological Medicine, 9*(1), 139–145.

[12] Zagai, U., et al. (2019). The Swedish Twin Registry: Content and management as a research infrastructure. *Twin Research and Human Genetics, 22*(6).

[13] Statistiska centralbyrån. (2019). Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier (LISA) – Bakgrundsfakta 1990–2017.

[14] Statistiska centralbyrån. (2017). Flergenerationsregistret 2016 – En beskrivning av innehåll och kvalitet.

[15] Milani, L., et al. (2024). From biobanking to personalized medicine: The journey of the Estonian Biobank. *medRxiv*.

[16] Vaht, M., et al. (2024). Cohort Profiles: Personality measurements at the Estonian Biobank of the Estonian Genome Center, University of Tartu. OSF Preprints. Available at: https://doi.org/10.31234/osf.io/2aey6

[17] Kweon, H., Burik, C. A. P., Ning, Y., et al. (2025). Associations between common genetic variants and income provide insights about the socio-economic health gradient. *Nature Human Behaviour*.

[18] Mõttus, R., et al. (2024). Most people's life satisfaction matches their personality traits: True correlations in multitrait, multirater, multisample data. *Journal of Personality and Social Psychology, 126*(4), 676–693.

[19] Norwegian Institute of Public Health. (2024). Questionnaires from MoBa. Norwegian Institute of Public Health. Available at: https://www.fhi.no/en/ch/studies/moba/for-forskere-artikler/questionnaires-from-moba/

[20] Statistics Norway. (2017). Norwegian standard classification of education. Statistics Norway. Available at: https://www.ssb.no/en/utdanning/norwegian-standard-classification-of-education

[21] Statistics Norway. (2024a). Employment, register-based. Statistics Norway. Available at: https://www.ssb.no/en/arbeid-og-lonn/sysselsetting/statistikk/sysselsetting-registerbasert

[22] Statistics Norway. (2025). Income and wealth statistics for households. Statistics Norway. Available at: https://www.ssb.no/en/inntekt-og-forbruk/inntekt-og-formue/statistikk/inntekts-og-formuesstatistikk-for-husholdninger

[23] Statistics Norway. (2024b). Families and households. Statistics Norway. Available at: https://www.ssb.no/en/befolkning/barn-familier-og-husholdninger/statistikk/familier-og-husholdninger

[24] Norwegian Institute of Public Health. (n.d.). Instrument documentation, 15 weeks women and men. *Norwegian Institute of Public Health*. Available at:
Women: https://www.fhi.no/globalassets/dokumenterfiler/studier/den-norske-mor-far-og-barn--undersokelsenmoba/instrumentdokumentasjon/instrument-documentation-q1.pdf
Men: https://www.fhi.no/globalassets/dokumenterfiler/studier/den-norske-mor-far-og-barn--undersokelsenmoba/instrumentdokumentasjon/instrument-documentation-q-father.pdf

[25] Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*, 71–75.

[26] Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. *Quality of Life Research, 13*, 299–310.

[27] Treiman, D. J. (1977). Developing the scale. In *Occupational prestige in comparative perspective* (pp. 159–189). Elsevier.

Table E.2: Results for each dataset

	Cohort-level results														
	EstBB					GS					MoBa				
Phenotype	N_{pairs}	h^2	σ^2_C	σ^2_E	ρ_{sib}	N_{pairs}	h^2	σ^2_C	σ^2_E	ρ_{sib}	N_{pairs}	h^2	σ^2_C	σ^2_E	ρ_{sib}
Panel A. Cognitive and educational															
Cognitive performance						8,044	0.543**	0.085	0.372***	0.357***					
							(0.287)	(0.144)	(0.144)	(0.010)					
EA	35,124	0.146	0.267***	0.587***	0.346***	7,968	-0.138	0.358**	0.780***	0.303***	11,378	-0.047	0.385***	0.662***	0.360***
		(0.154)	(0.077)	(0.077)	(0.005)		(0.326)	(0.164)	(0.163)	(0.011)		(0.236)	(0.119)	(0.118)	(0.009)
Panel B. Labor market															
Employed	11,342	0.176	-0.024	0.848***	0.057***						11,367	-0.055	0.112	0.943***	0.081***
		(0.362)	(0.182)	(0.181)	(0.009)							(0.287)	(0.145)	(0.142)	(0.009)
Log family income	3,241	-0.394	0.257	1.137***	0.061***	4,639	-0.346	0.406**	0.940***	0.224***	11,354	0.00	0.203	0.797**	0.205***
		(0.446)	(0.224)	(0.224)	(0.018)		(0.442)	(0.221)	(0.222)	(0.014)		(0.698)	(0.354)	(0.345)	(0.009)
Log occupational income	7,938	0.361	0.002	0.638***	0.184***						10,455	-0.124	0.264**	0.860***	0.203***
		(0.386)	(0.193)	(0.194)	(0.011)							(0.249)	(0.125)	(0.125)	(0.010)
Occupational Status	2,183	-0.242	0.211	1.030***	0.088***						10,615	0.125	0.203**	0.671***	0.265***
		(0.692)	(0.346)	(0.348)	(0.021)							(0.241)	(0.122)	(0.120)	(0.009)
Panel C. Risk tolerance and risky behaviors															
Ever smoker	37,250	0.160	0.130*	0.710***	0.218***	7,870	0.662**	-0.123	0.461***	0.199***	1,517	0.496	-0.052	0.556**	0.204***
		(0.158)	(0.080)	(0.079)	(0.005)		(0.329)	(0.165)	(0.165)	(0.011)		(0.656)	(0.330)	(0.328)	(0.025)
Cigarettes per day (logged)	19,534	0.154	0.136	0.710***	0.238***	7,607	0.610*	-0.107	0.497***	0.191***	749	1.638**	-0.667	0.029	0.162***
		(0.300)	(0.151)	(0.150)	(0.007)		(0.371)	(0.186)	(0.185)	(0.011)		(0.912)	(0.459)	(0.456)	(0.036)
Drinks per week (logged)	15,725	-0.222	0.225**	0.997***	0.111***	7,070	0.515*	-0.091	0.576***	0.159***					
		(0.271)	(0.136)	(0.135)	(0.008)		(0.356)	(0.178)	(0.178)	(0.012)					
Risk tolerance	9,249	0.030	0.078	0.892***	0.095***										
		(0.366)	(0.184)	(0.183)	(0.010)										
Panel D. Health-related & other															
Number of children	2,954	0.555	-0.142	0.587**	0.158***						11,375	-0.158	0.239*	0.919***	0.157***
		(0.543)	(0.274)	(0.271)	(0.018)							(0.327)	(0.164)	(0.164)	(0.009)
Self-rated general health	21,851	-0.125	0.186*	0.939***	0.130***	7,318	0.030	0.094	0.876***	0.100***	2,321	0.120	0.045	0.834***	0.105***
		(0.226)	(0.114)	(0.113)	(0.007)		(0.356)	(0.179)	(0.179)	(0.012)		(0.614)	(0.309)	(0.307)	(0.021)
Subjective wellbeing	9,076	0.806**	-0.271	0.465***	0.157***	7,323	-0.469	0.320*	1.149***	0.085***	9,494	0.209	-0.002	0.794***	0.099***
		(0.353)	(0.178)	(0.176)	(0.010)		(0.430)	(0.218)	(0.213)	(0.012)		(0.259)	(0.130)	(0.130)	(0.010)
Panel E. Anthropometric															
BMI	37,300	0.504***	0.028	0.468***	0.276***	8,312	0.245	0.176*	0.579***	0.307***	9,502	0.479*	0.021	0.500***	0.265***
		(0.174)	(0.088)	(0.062)	(0.005)		(0.251)	(0.127)	(0.126)	(0.010)		(0.301)	(0.150)	(0.151)	(0.010)
Height	37,360	0.731***	0.125**	0.143**	0.488***	8,355	0.809***	0.110	0.081	0.518***	9,732	0.572***	0.200**	0.228**	0.487***
		(0.125)	(0.063)	(0.062)	(0.005)		(0.261)	(0.132)	(0.130)	(0.009)		(0.203)	(0.103)	(0.101)	(0.009)

(Continues)

Table E.2 (continued): Results for each dataset

	Cohort-level results (continued)														
	STR					UKB					WLS				
Phenotype	N_{pairs}	h^2	σ^2_C	σ^2_E	ρ_{sib}	N_{pairs}	h^2	σ^2_C	σ^2_E	ρ_{sib}	N_{pairs}	h^2	σ^2_C	σ^2_E	ρ_{sib}
Panel A. Cognitive and educational															
Cognitive performance	687	0.452	0.240	0.308	0.468***	6,908	0.813***	-0.123	0.310**	0.287***	1,693	1.569***	-0.441	-0.128	0.342***
		(0.897)	(0.453)	(0.445)	(0.034)		(0.317)	(0.160)	(0.158)	(0.012)		(0.615)	(0.309)	(0.308)	(0.023)
EA	5,672	0.043	0.399***	0.558***	0.421***	18,949	0.121	0.332***	0.547***	0.394***	623	0.300	0.279	0.421	0.430***
		(0.321)	(0.161)	(0.160)	(0.012)		(0.179)	(0.090)	(0.089)	(0.007)		(1.102)	(0.545)	(0.560)	(0.036)
Panel B. Labor market															
Employed	4,659	-0.464	0.342*	1.122***	0.110***	9,462	-0.221	0.205	1.017***	0.094***	460	0.775	-0.303	0.528	0.081**
		(0.524)	(0.265)	(0.260)	(0.015)		(0.320)	(0.161)	(0.160)	(0.010)		(1.457)	(0.725)	(0.735)	(0.047)
Log family income						7,742	0.124	0.175	0.701***	0.239***					
							(0.306)	(0.153)	(0.153)	(0.011)					
Log occupational income	3,133	0.674*	-0.090	0.416**	0.250***	9,022	-0.120	0.294**	0.825***	0.233***					
		(0.452)	(0.229)	(0.224)	(0.017)		(0.279)	(0.140)	(0.140)	(0.010)					
Occupational Status	3,351	-0.040	0.289	0.750***	0.270***	8,923	-0.262	0.356***	0.906***	0.223***					
		(0.451)	(0.226)	(0.227)	(0.017)		(0.282)	(0.142)	(0.141)	(0.010)					
Panel C. Risk tolerance and risky behaviors															
Ever smoker	3,094	0.114	0.237	0.649***	0.294***	19,078	-0.101	0.208**	0.893***	0.156***	1,868	-0.098	0.205	0.893***	0.155***
		(0.475)	(0.239)	(0.237)	(0.017)		(0.200)	(0.100)	(0.100)	(0.007)		(0.635)	(0.319)	(0.318)	(0.023)
Cigarettes per day (logged)	1,119	0.808	-0.270	0.462	0.137***	10,228	-0.343	0.433***	0.910***	0.258***	1,861	0.346	-0.014	0.669**	0.155***
		(0.936)	(0.474)	(0.465)	(0.030)		(0.266)	(0.134)	(0.133)	(0.010)		(0.637)	(0.319)	(0.320)	(0.023)
Drinks per week (logged)						12,199	-0.038	0.212*	0.827***	0.197***	1,448	0.263	0.059	0.679**	0.190***
							(0.260)	(0.130)	(0.131)	(0.009)		(0.741)	(0.369)	(0.373)	(0.026)
Risk tolerance	1,425	0.213	-0.013	0.800**	0.095***	18,153	0.598***	-0.227	0.629***	0.072***					
		(0.710)	(0.360)	(0.352)	(0.026)		(0.210)	(0.105)	(0.105)	(0.007)					
Panel D. Health-related & other															
Number of children	4,210	-0.085	0.150	0.935***	0.107***	16,387	0.468**	-0.117	0.649***	0.117***					
		(0.438)	(0.222)	(0.218)	(0.015)		(0.226)	(0.114)	(0.113)	(0.008)					
Self-rated general health	1,335	-0.736	0.505	1.231***	0.136***	19,092	0.249	0.029	0.722***	0.151***					
		(0.836)	(0.422)	(0.416)	(0.027)		(0.201)	(0.101)	(0.101)	(0.007)					
Subjective wellbeing	1,440	1.485**	-0.631	0.146	0.113***	2,616	0.393	-0.080	0.687***	0.116***					
		(0.705)	(0.357)	(0.350)	(0.026)		(0.561)	(0.282)	(0.280)	(0.019)					
Panel E. Anthropometric															
BMI	2,930	-0.111	0.307	0.804***	0.252***	19,112	1.070***	-0.252	0.182**	0.284***	1,300	0.869	-0.191	0.322	0.245***
		(0.555)	(0.281)	(0.275)	(0.018)		(0.218)	(0.109)	(0.110)	(0.007)		(0.779)	(0.386)	(0.394)	(0.027)
Height	2,933	0.920**	0.015	0.064	0.478***	19,148	0.491***	0.283***	0.226***	0.528***	1,318	-0.241	0.518*	0.722**	0.398***
		(0.423)	(0.214)	(0.210)	(0.016)		(0.147)	(0.074)	(0.073)	(0.006)		(0.753)	(0.378)	(0.377)	(0.025)

Notes: This table mirrors Table 2, but provides the results separately for each dataset (instead of the meta-analyzed results). The table reports the estimates of h^2, σ_C^2, and σ_E^2 from the baseline ACE model (without assortative mating) for each of the six datasets. N_{pairs} is the number of sib pairs. ρ_{sib} is the correlation in the outcome across sib pairs (from Equation 3, $\rho_{sib} = \sigma_C^2 + \pi h^2$). Stars indicate the significance of the estimates on one-sided tests, as described in the text: $*** \ p < 0.01$, $** \ p < 0.05$, $* \ p < 0.1$.

Table E.3: $\hat{h}^2$ corrected for assortative mating

	N_{pairs}	$\hat{\rho}_{sib}$	$\hat{h}^2$	$\hat{h}^2_{r=0.1}$	$\hat{h}^2_{r=0.3}$	$\hat{h}^2_{r=0.5}$
Panel A. Cognitive and educational						
Cognitive performance	17,332	0.334***	0.746***	0.830***	1.066***	1.493***
		(0.007)	(0.196)	(0.218)	(0.281)	(0.393)
EA	79,714	0.362***	0.076	0.085	0.109	0.153
		(0.003)	(0.095)	(0.105)	(0.135)	(0.189)
Panel B. Labor market						
Employed	37,290	0.081***	-0.083	-0.093	-0.120	-0.168
		(0.005)	(0.172)	(0.192)	(0.246)	(0.345)
Log family income	26,976	0.202***	-0.106	-0.118	-0.151	-0.212
		(0.006)	(0.209)	(0.232)	(0.298)	(0.418)
Log occupational income	30,548	0.212***	0.054	0.060	0.077	0.108
		(0.006)	(0.157)	(0.174)	(0.224)	(0.314)
Occupational Status	25,072	0.236***	-0.050	-0.056	-0.071	-0.100
		(0.006)	(0.165)	(0.183)	(0.236)	(0.330)
Panel C. Risk tolerance and risky behaviors						
Ever smoker	70,677	0.201***	0.137	0.151	0.195	0.272
		(0.004)	(0.109)	(0.122)	(0.156)	(0.219)
Cigarettes per day (logged)	41,098	0.227***	0.135	0.150	0.193	0.271
		(0.005)	(0.164)	(0.182)	(0.234)	(0.328)
Drinks per week (logged)	36,442	0.153***	0.025	0.028	0.035	0.050
		(0.005)	(0.162)	(0.180)	(0.231)	(0.324)
Risk tolerance	28,827	0.081***	0.442***	0.491***	0.631***	0.884***
		(0.006)	(0.176)	(0.196)	(0.252)	(0.353)
Panel D. Health-related & other						
Number of children	34,926	0.132***	0.243*	0.270*	0.347*	0.486*
		(0.005)	(0.163)	(0.181)	(0.233)	(0.327)
Self-rated general health	51,917	0.133***	0.057	0.064	0.082	0.115
		(0.004)	(0.133)	(0.148)	(0.190)	(0.267)
Subjective wellbeing	29,949	0.116***	0.337**	0.373**	0.480**	0.672**
		(0.006)	(0.173)	(0.192)	(0.246)	(0.345)
Panel E. Anthropometric						
BMI	78,456	0.279***	0.575***	0.640***	0.822***	1.151***
		(0.003)	(0.108)	(0.120)	(0.154)	(0.216)
Height	78,846	0.500***	0.639***	0.709***	0.912***	1.277***
		(0.003)	(0.080)	(0.089)	(0.114)	(0.160)

Notes: The table reports the heritability estimates (h^2) adjusted for various assumed levels of assortative mating. The estimates were obtained by meta-analyzing the dataset-level estimates, as described in the text. N_{pairs} is the total number of sib pairs in the meta-analysis. ρ_{sib} is the correlation in the outcome across sib pairs. $\hat{h}^2$ is the baseline heritability estimate (also reported in Table 2). $h^2_{r=0.1}$, $h^2_{r=0.3}$, and $h^2_{r=0.5}$ are the heritability estimates adjusted for $r = 0.1$, $r = 0.3$, and $r = 0.5$, respectively, where r is the correlation r between mothers' and fathers' additive genetic factors for the outcome (see Online Appendix C for details). Stars indicate the significance of the estimates on one-sided tests, as described in the text: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E.4: $\hat{\sigma}^2_C$ corrected for assortative mating

	N_{pairs}	$\hat{\rho}_{sib}$	$\hat{\sigma}^2_C$	$\hat{\sigma}^2_{C,r=0.1}$	$\hat{\sigma}^2_{C,r=0.3}$	$\hat{\sigma}^2_{C,r=0.5}$
Panel A. Cognitive and educational						
Cognitive performance	17,332	0.334***	-0.040	-0.124	-0.361	-0.787
		(0.007)	(0.099)	(0.121)	(0.183)	(0.295)
EA	79,714	0.362***	0.323***	0.315***	0.291***	0.247**
		(0.003)	(0.048)	(0.058)	(0.088)	(0.142)
Panel B. Labor market						
Employed	37,290	0.081***	0.127*	0.136*	0.208*	0.211
		(0.005)	(0.087)	(0.106)	(0.161)	(0.259)
Log family income	26,976	0.202***	0.247**	0.259**	0.293*	0.353
		(0.006)	(0.105)	(0.128)	(0.194)	(0.314)
Log occupational income	30,548	0.212***	0.188*	0.182	0.165	0.134
		(0.006)	(0.079)	(0.096)	(0.146)	(0.236)
Occupational Status	25,072	0.236***	0.268***	0.273***	0.289**	0.318*
		(0.006)	(0.083)	(0.101)	(0.154)	(0.248)
Panel C. Risk tolerance and risky behaviors						
Ever smoker	70,677	0.201***	0.128***	0.113**	0.070	-0.008
		(0.004)	(0.055)	(0.067)	(0.102)	(0.164)
Cigarettes per day (logged)	41,098	0.227***	0.152**	0.138*	0.076	0.018
		(0.005)	(0.082)	(0.100)	(0.152)	(0.246)
Drinks per week (logged)	36,442	0.153***	0.146**	0.144*	0.135	0.122
		(0.005)	(0.081)	(0.099)	(0.150)	(0.243)
Risk tolerance	28,827	0.081***	-0.144	-0.192	-0.332	-0.585
		(0.006)	(0.088)	(0.108)	(0.164)	(0.265)
Panel D. Health-related & other						
Number of children	34,926	0.132***	0.007	-0.020	-0.098	-0.237
		(0.005)	(0.082)	(0.100)	(0.152)	(0.245)
Self-rated general health	51,917	0.133***	0.105*	0.099	0.080	0.048
		(0.004)	(0.067)	(0.082)	(0.124)	(0.200)
Subjective well-being	29,949	0.116***	-0.060	-0.097	-0.204	-0.546
		(0.006)	(0.087)	(0.106)	(0.161)	(0.259)
Panel E. Anthropometric						
BMI	78,456	0.279***	-0.009	-0.074	-0.256	-0.585
		(0.003)	(0.054)	(0.066)	(0.100)	(0.162)
Height	78,846	0.500***	0.183***	0.112**	-0.091	-0.456
		(0.003)	(0.040)	(0.049)	(0.075)	(0.120)

Notes: The table reports the estimates of outcome variance share attributable to the common family environment (σ^2_C), adjusted for various assumed levels of assortative mating. The estimates were obtained by meta-analyzing the dataset-level estimates, as described in the text. N_{pairs} is the total number of sib pairs in the meta-analysis. ρ_{sib} is the correlation in the outcome across sib pairs. $\hat{\sigma}^2_C$ is the baseline estimate (also reported in Table 2). $\hat{\sigma}^2_{C,r=0.1}$, $\hat{\sigma}^2_{C,r=0.3}$, and $\hat{\sigma}^2_{C,r=0.5}$ are the estimates adjusted for $r = 0.1$, $r = 0.3$, and $r = 0.5$, respectively, where r is the correlation r between mothers' and fathers' additive genetic factors for the outcome (see Online Appendix C for details). Stars indicate the significance of the estimates on one-sided tests, as described in the text: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

References

Amador, Carmen, Jennifer E Huffman, Holly Trochet, Veronique Vitart, Pau Navarro, et al. (2015). "Recent genomic heritage in Scotland". In: *BMC Genomics* 16.1, p. 437.

Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, et al. (2017). "Genome-wide genetic data on 500,000 UK Biobank participants". In: *bioRxiv*.

CAMSIS (n.d.). *SOC to ISCO crosswalk*. https://web.archive.org/web/20060304112311if_/http://www.cf.ac.uk:80/socsi/CAMSIS/occunits/OIUsoctoiscoxls.zip. Accessed: 2024-05-25.

Chang, Christopher C, Carson C Chow, Laurent C Tellier, Shashaank Vattikuti, Shaun M Purcell, et al. (Feb. 2015). "Second-generation PLINK: rising to the challenge of larger and richer datasets". In: *Gigascience* 4, p. 7.

Corfield, Elizabeth C., Alexey A. Shadrin, Oleksandr Frei, Zillur Rahman, Aihua Lin, et al. (2024). "The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1". In: *bioRxiv*.

Ganzeboom, Harry B. and Donald J. Treiman (1996). "Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations". In: *Social Science Research* 25.3, pp. 201–239.

Haplotype Reference Consortium (n.d.). *Haplotype Reference Consortium*. https://www.sanger.ac.uk/collaboration/haplotype-reference-consortium/. Accessed: 2024-05-25.

Herd, Pamela (2016). *Quality Control Report for Genotypic Data*. URL: https://www.ssc.wisc.edu/wlsresearch/documentation/GWAS/Herd_QC_report.pdf.

Kuznetsov, Ivan A., Mait Metspalu, Uku Vainik, Luca Pagani, et al. (2023). "Assessing the impact of 20th century internal migrations on the genetic structure of Estonia". In: *bioRxiv*.

Maarten Hermans (n.d.). *strat: Functions and data useful for social stratification research*. https://rdrr.io/rforge/strat/. Accessed: 2024-05-25.

Manichaikul, Ani, Josyf C Mychaleckyj, Stephen S Rich, Kathleen Daly, Michèle Sale, et al. (Nov. 2010). "Robust relationship inference in genome-wide association studies". In: *Bioinformatics* 26.22, pp. 2867–2873.

Nagy, Reka, Thibaud S Boutin, Jonathan Marten, Jennifer E Huffman, Shona M Kerr, et al. (2017). "Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants". In: *Genome Medicine* 9.1, p. 23.

Nicole Schwitter (n.d.). *ISCO08ConveRsions: Converts ISCO-08 to Job Prestige Scores, ISCO-88 and Job Name*. https://cran.r-project.org/package=ISCO08ConveRsions. Accessed: 2024-05-25.

Sham, P.C. and S. Purcell (2001). "Equivalence between Haseman-Elston and Variance-Components Linkage Analyses for Sib Pairs". In: *The American Journal of Human Genetics* 68.6, pp. 1527–1532.

Treiman, Donald J. (1977). "Occupational Prestige in Comparative Perspective". In: *Academic Press*.

Visscher, Peter M. (2006). "A Note on the Asymptotic Distribution of Likelihood Ratio Tests to Test Variance Components". In: *Twin Research and Human Genetics* 9.4, pp. 490–495.

Xu, Xin, Scott Weiss, Xiping Xu, and LJ Wei (2000). "A unified Haseman-Elston method for testing linkage with quantitative traits". In: *The American Journal of Human Genetics* 67.4, pp. 1025–1028.