PNAS RESEARCH ARTICLE | DEMOGRAPHY GENETICS

Using the phenotype differences model to identify genetic effects in samples of partially genotyped sibling pairs

Sam Trejo[a,1] and Klint Kanopka[b]

Edited by Marcus Feldman, Stanford University, Stanford, CA; received March 19, 2024; accepted October 23, 2024

The identification of causal relationships between specific genes and social, behavioral, and health outcomes is challenging due to environmental confounding from population stratification and dynastic genetic effects. Existing methods to eliminate environmental confounding leverage random genetic variation resulting from recombination and require within-family dyadic genetic data (i.e., parent–child and/or sibling pairs), meaning they can only be applied in relatively small and selected samples. We introduce the *phenotype differences* model and provide derivations showing that it—under plausible assumptions—provides consistent (and, in certain cases, unbiased) estimates of genetic effects using just a single individual's genotype. Then, leveraging distinct samples of fully and partially genotyped sibling pairs in the Wisconsin Longitudinal Study, we use polygenic indices and phenotypic data for 24 different traits to empirically validate the phenotype differences model. Finally, we utilize the model to test the effects of 40 polygenic indices on lifespan. After a 10% false discovery rate correction, we find that polygenic indices for three traits—body mass index, self-rated health, chronic obstructive pulmonary disease—have a statistically significant effect on an individual's lifespan.

causal inference | biodemography | quantitative methods | genomics | premature mortality

Understanding whether and how variation in individual DNA sequence produces variation in life outcomes is a key goal of the field of human genomics. Over the last decade, researchers have been successful at assembling large genome-wide association study (GWAS) samples of unrelated individuals to precisely estimate genetic associations for a wide range of traits (1). However, the identification of causal relationships between specific genes and social, behavioral, and health outcomes—as well as the mechanisms that such effects operate through—is challenging due to between-family environmental confounding from population structure (2–4) and dynastic genetic effects (5, 6). Thus, it is difficult to know how well current between-family GWAS discoveries and existing polygenic indices capture the causal effects of genes (7).

While naive genetic associations are often environmentally confounded, there exist promising avenues for causal inference. In the case of DNA, we have the ultimate "natural" experiment; conditional on their parents' genes, a child's genes are randomly assigned via genetic recombination. Numerous strategies have been developed to leverage these random genetic differences between parents and their children in order to identify genetic effects.* These strategies include both i) trio methods (11–13), which explicitly condition on parental genotype, and ii) sibling methods (7, 14–16), which difference out all shared family-level factors, thereby indirectly conditioning on parental genotype. However, existing quantitative methods require the use of dyadic genetic data within families (i.e., parent–child pairs and/or sibling pairs) and, therefore, can only be applied in relatively small and selected samples. A dearth of such data exists; for instance, though the UK Biobank has roughly 500,000 genotyped individuals, it has only about 22,000 sibling pairs (and even fewer parent–child pairs) (17). At present, researchers are left with estimates of genetic effects that are either precise but environmentally biased or quite imprecise but environmentally unbiased.

*In research seeking to identify the effects of a specific genetic variant (or of a subset of genetic variants), genetic confounding may still be an issue. That is, estimates of the effects of a given genetic predictor g_{ij} on some outcome y_{ij} may be confounded by genetic variants not encompassed in g_{ij}. Nonetheless, the bias due to genetic confounding is expected to be smaller in within-family studies than in population studies (8). Moreover, this is generally not a concern in polygenic index analyses, where typically the only goal is to eliminate confounding due to nongenetic characteristics; importantly, the within-family regression coefficient of a polygenic index has been shown to reflect only the causal effects of genetic variants (9, 10).

Significance

Today, researchers have access to vast amounts of genomic data, but genetic influences and environmental influences on human traits can be difficult to disentangle. A key challenge is identifying whether specific genetic factors cause (rather than simply correlate with) various social, behavioral, and health outcomes. We introduce a statistical strategy for detecting genetic effects that can help increase the precision and generalizability of genetic discoveries, which we then validate using data on full sibling pairs from Wisconsin. We also use our strategy to test which genes affect premature mortality and identify genetic factors related to three different traits—body mass index, self-rated health, and chronic obstructive pulmonary disease—that have significant effects on which sibling lives longer than the other.

Author affiliations: [a]Department of Sociology and Office of Population Research, Princeton University, Princeton, NJ 08544; and [b]Steinhardt School of Culture, Education, and Human Development, Department of Applied Statistics, Social Science, and Humanities, New York University, New York, NY 10003

Author contributions: S.T. designed research; S.T. and K.K. performed research; K.K. analyzed data; and S.T. and K.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: samtrejo@princeton.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405725121/-/DCSupplemental.

Published November 26, 2024.

Typical sibling methods, such as the *fixed effects* model (Eq. **1**)[†] require four pieces of information: the genotype of both siblings and the phenotype of both siblings.

$$y_{1j} - y_{2j} = \beta^{\mathrm{FE}}(g_{1j} - g_{2j}) + \varepsilon_j^* \quad [1]$$

We introduce a within-family regression specification—which we call the *phenotype differences* model (Eq. **2**)—for comparing siblings and estimating direct genetic effects.

$$y_{1j} - y_{2j} = \alpha + \beta^{\mathrm{PD}}\big(g_{1j}(1-\rho)\big) + \varepsilon_{1j} \quad [2]$$

Here, y_{ij} is the outcome for individual i in family j, g_{ij} is the genetic predictor of individual i in family j, α is an intercept term, and ρ is the population correlation between g_{1j} and g_{2j}—that is, $\mathrm{corr}(g_{1j}, g_{2j})$. Importantly, the phenotype differences model provides, in expectation, the same estimates as fixed effects models but instead requires just a single individual's genotype (as well as the phenotype of that individual and one of their siblings).[‡] In doing so, the phenotype differences model can increase the statistical power (by increasing the size of analytic samples) and improve the external validity (by increasing the representativeness of samples) of within-family genetic analyses. While the phenotype differences model can hypothetically be applied when studying the effects of nongenetic variables, it is especially well-suited for genetic predictors because of our strong prior regarding the correlation of genes among full biological siblings (ρ).

We show that, when genetic effects are small, phenotype differences provide the same precision as fixed effects per genotype. The key assumptions of the phenotype differences model, as well as mathematical details on precision, can be found in the *Online Methods*. *SI Appendix*, section D provides unbiasedness and consistency proofs of the phenotype differences estimator. While phenotype differences are valid for both variant-level (e.g., in GWAS, where g_{ij} is a single nucleotide polymorphism) and genome-wide (e.g., when g_{ij} is a polygenic index or some other summary measure) analyses, in our empirical application, we focus on the genome-wide case. Note that when g_{ij} is standardized within-sample to have mean 0 and variance 1, as is common in the polygenic index literature, the phenotype differences estimator is, like the fixed effects estimator, biased in finite samples. However, the approximate expected bias in $\hat{\beta}$ is small; it is less than 1% when N—the number of (partially genotyped) sibling pairs used—is equal to just 30, and this bias disappears entirely as N grows large.

The mathematical intuition of the phenotype differences model is, perhaps, best understood by first considering key aspects of the fixed effects model. In general, if some family-level environment (e_j) is associated with *both* genotype (g_{ij}) and phenotype (y_{ij}), a naive between-family estimate of the relationship between genotype and phenotype will be confounded. The fixed effects model regresses the sibling difference in phenotype on the sibling difference in genotype which, under the standard assumption of random assignment of genotype within families, recovers genetic effect estimates that are free from environmental confounding. Notice that, by applying the 'within transformation' on genotype and phenotype—thereby breaking the link between e_j and *both* y_{ij} and g_{ij}—the fixed effects model *exceeds* the requirements for eliminating environmental confounding.

This fact forms the conceptual basis for the phenotype differences model, which breaks only the link between e_j and y_{ij}, and thereby requires the use of just a single sibling's genetic information. Because the environmental effect, e_j, does not vary within families, it is mechanically uncorrelated with the sibling difference in phenotypes, $y_{1j} - y_{2j}$. Thus, even though a correlation between e_j and g_{1j} may persist, it does not introduce environmental confounding. However, because the covariance of $y_{1j} - y_{2j}$ and g_{1j} becomes distorted, we must use the within-family correlation of genetic predictors, ρ, to reinflate our genetic effect estimates; in effect, multiplying g_{1j} by $1 - \rho$ allows one to regress the observed phenotypic difference on the *expected* (albeit unobserved) genotypic difference.

We also present an augmented version of the phenotype differences estimator that is robust to the existence of assortative mating. Rather than assuming a specific value for ρ (for instance, $\frac{1}{2}$, the expectation under random mating), $\hat{\rho}$ can be estimated using a distinct sample of M fully genotyped sibling pairs. *SI Appendix*, section D provides consistency proofs (as N and M both go infinity) for this version of the phenotype differences model; importantly, because the sample correlation is a biased estimator of the population correlation in finite samples, phenotype differences models using $\hat{\rho}$ produce $\hat{\beta}$ that are biased toward zero. However, the approximate magnitude of this bias is small, being less than 1% when $M = 50$.

As in all regression-based quasi-experimental designs, researchers must be careful not to include unnecessary covariates, which may produce collider bias; that is, when a variable of interest is (conditionally) ignorably assigned, the inclusion of additional control variables may reintroduce statistical dependence with potential confounders. We urge further caution regarding the use of covariates when applying the phenotype differences model, as control variables may lead the conditional expectation of the sibling genotypic difference to depart from $g_{1j}(1 - \rho)$ and thereby bias $\hat{\beta}$. Recall that, by virtue of leveraging the random assignment of genotype within families, the standard bivariate phenotype differences model provides estimates that are free of environmental confounding (and therefore the use of covariates is typically unnecessary and should be avoided).

If some covariate x_{ij} is statistically independent from g_{ij} and the genotyped and ungenotyped sibling are randomly sampled from all siblings within a family,[§] then $x_{1j} - x_{2j}$ will be independent from $g_{1j}(1 - \rho)$ and thus the sibling difference can be included as a covariate in the phenotype differences model without side effect. However, if x_{ij} and g_{ij} are statistically dependent, or x_{ij} may be related to selection into genotyping or phenotyping, we recommend instead residualizing $y_{1j} - y_{2j}$ on $x_{1j} - x_{2j}$ (which mechanically cannot influence the expected sibling genotypic difference) and then fitting phenotype differences on the residuals. Notably, while residualizing both the dependent and independent variable on a set of covariates before running a bivariate regression is statistically equivalent to combining all the relevant variables in a single multivariate regression (20, 21), residualizing only the dependent variable does not produce such an equivalency. As such, any resulting $\hat{\beta}$ must be interpreted as

[†]In general, fixed effects and first differences models are slightly different statistical approaches for making for within-group comparisons. However, in our case, where the number of observations i in each group j is equal to two (e.g., sibling pairs), the fixed effects and first differences specifications are algebraically identical [see chapter 5.1. of Angrist and Pischke (18)].

[‡]Both fixed effects and phenotypes difference models may suffer from bias in the presence of indirect effects between siblings (13); there currently exists mixed evidence regarding the prevalence of and magnitude of such sibling effects (5, 19).

[§]For instance, the sibling difference in a dichotomous variable indicating an individual's sex is likely orthogonal to the sibling difference in genetic predictor (as long as the genetic predictor is constructed using only autosomal information). Also note that many variables, such as mother's educational attainment, do not vary among full sibling pairs and therefore cannot be used as covariates in within-family models like phenotype differences.

Table 1. Wisconsin longitudinal study sibling data

A. Two Genotypes sample

	Graduate			Not graduate		
	Mean	SD	*N*	Mean	SD	*N*
Female	0.53	0.50	2,107	0.53	0.50	2,107
Birth Year	1,938.87	0.48	2,107	1,940.68	6.87	2,107
Deceased pre-2020	0.30	0.46	2,107	0.28	0.45	2,107
Survived to age 75	0.94	0.23	2,107	0.92	0.28	1,672
Lifespan	80.05	2.46	2,107	78.27	6.51	2,107

B. One Genotype sample

	Genotyped			Not genotyped		
	Mean	SD	*N*	Mean	SD	*N*
Graduate	0.74	0.44	3,494	0.26	0.44	3,494
Female	0.51	0.50	3,494	0.48	0.50	3,494
Birth year	1,939.31	3.50	3,494	1,939.22	6.67	3,494
Deceased pre-2020	0.31	0.46	3,494	0.49	0.50	3,494
Survived to age 75	0.93	0.25	3,309	0.60	0.49	3,101
Lifespan	79.52	3.99	3,494	72.93	10.82	3,494

This table uses data from the Wisconsin Longitudinal Study. WLS "graduates" are the original members of the study, who graduated from high school in 1957. Later, a randomly selected sibling of each graduate was empaneled into the study. Both panels include only self-reported full sibling pairs. Panel (*A*) contains the Two Genotypes sample, where both siblings were genotyped, while Panel (*B*) contains the One Genotype sample, where just a single sibling was genotyped. Genetic data indicate that 48 out of the 2,107 sibling pairs displayed in Panel (*A*) are actually half siblings. The lifespan variable is right censored; only deaths prior to 2020 are included.

the effect of the genetic predictor on the residualized outcome variable.[¶]

As an example application using a sample of sibling pairs drawn from the Wisconsin Longitudinal Study (WLS), we leverage polygenic index and phenotype data for 24 traits to empirically validate the phenotype differences model. The WLS is a longitudinal survey based on a $\frac{1}{3}$ sample of all 1957 Wisconsin high school graduates ($N = 10{,}317$) and a randomly selected sibling of these graduates (22). The graduates were originally empaneled with an in-person questionnaire at age 18; both WLS graduates and the randomly selected siblings were reinterviewed periodically across the life course, and in recent years genetic data was assayed from consenting respondents. All polygenic indices are drawn from the recent Social Science Genomics Association Consortium (SSGAC) Polygenic Index Repository (23).

Importantly, there exist two nonoverlapping samples of sibling pairs in the WLS: i) the One Genotype sample, where only a single sibling in each pair is genotyped, and ii) the Two Genotypes sample, where both siblings are genotyped. Table 1 displays summary statistics for these two distinct samples. To date, within-family GWAS and polygenic index analyses using the WLS data have focused on the 2,107 siblings pair that comprise the Two Genotypes sample (24–26). We show how the phenotype differences model allows us to extend within-family genomic analyses to the additional 3,494 sibling pairs in the One Genotype sample. By comparing results from fixed effects and phenotype differences models fit on the One Genotype and Two Genotypes samples, respectively, we demonstrate the accuracy and precision of the phenotype difference model. We also show how the phenotype differences model can help address mortality selection into genotyping when testing the effects of 40 polygenic indices on lifespan.[#]

Results

We begin by computing $\hat{\rho}$ estimates using the Two Genotypes sample. While we find evidence for assortative mating on a small number of traits, for most polygenic indices we observe $\hat{\rho}$ very close to 0.5. After implementing a 10% Benjamini, Krieger, and Yekutieli (BKY) two-step false discovery rate correction (27), polygenic indices for two traits remained statistically significant: height with a $\hat{\rho}$ of 0.6 (0.017) and cigarettes per day with a $\hat{\rho}$ of 0.44 (0.019).[‖] Dataset S1 contains $\hat{\rho}$ and its SE for each of the 47 polygenic indices considered; we utilize these estimates in the various phenotype differences models fit in this study, and SEs from these models are adjusted to account for the additional uncertainty resulting from $\widehat{\text{var}}(\hat{\rho})$ using the delta method (28, 29).

Then, we fit a series of regressions to empirically validate the performance of the phenotype differences model in the WLS. Fig. 1 compares the estimated $\hat{\beta}$ coefficients from fixed effects and phenotype differences regressions of 24 phenotypes on their respective polygenic index. These 24 traits are indicated in *SI Appendix*, Table S1 and represent all phenotypes in the SSGAC polygenic index repository for which at least 300 WLS siblings pairs in both the One Genotype and Two Genotypes samples have nonmissing phenotypic data.

All three panels of Fig. 1 display the same fixed effects estimates, which are derived from the full Two Genotypes sample. In Panels (*A* and *B*), the phenotype differences estimates come from a procedure using the Two Genotypes sample in which the genetic (but not phenotypic) data of a randomly selected sibling in each pair is discarded. For each phenotype, Panel (*A*) displays the

[¶] In addition, modeling decisions at the residualization stage can shape the interpretation of the resulting phenotype difference estimates; we recommend that researchers center continuous covariates at their mean prior to residualizing and select reference groups that are sensible for their specific application. Researchers will need to balance any potential precision gains with changes to interpretation of the estimand when deciding whether to residualize.

[#] Notably, we are forced to focus largely on *premature* mortality due to the fact that our lifespan variable is right censored in 2020. At that time, 64% of the members our analytic sample of sibling pairs were still alive.

[‖] SEs displayed in parentheses.

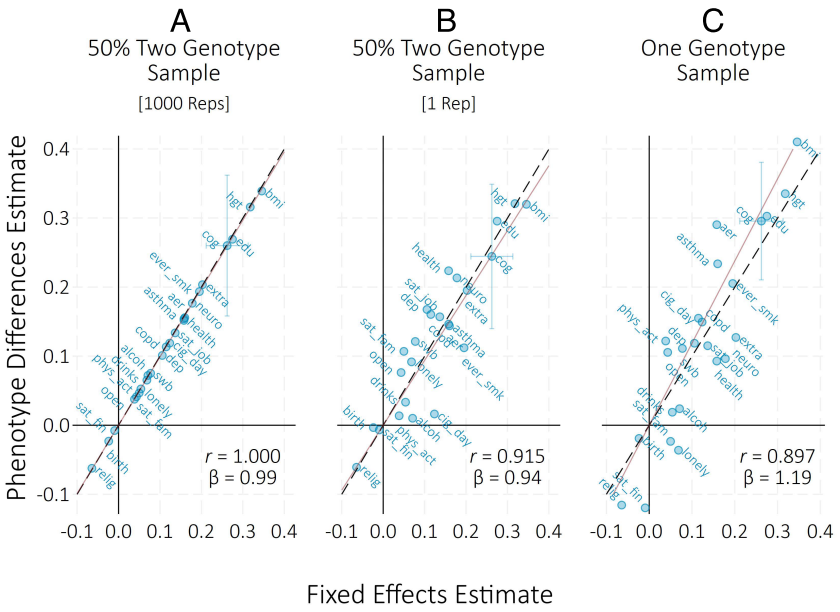

Fig. 1. Comparing Fixed Effects to Phenotype Differences Estimates. This figure compares the effects of various polygenic indices on their target phenotype estimated using fixed effects (FE) and phenotype differences (PD). All panels display the same FE estimates, which are derived from the full Two Genotypes sample. In Panels (*A* and *B*), the PD estimates come from a procedure using the Two Genotypes sample in which the genetic data of a randomly selected sibling in each pair is discarded; Panel (*A*) displays the mean from 1,000 iterations whereas Panel (*B*) displays the estimates from a single iteration. Panel (*C*) displays PD estimates derived from the One Genotype sample. Both polygenic indices and phenotypes are standardized over the full sample of genotyped WLS graduates. *SI Appendix*, Table S1 contains a list of the names and labels of all the phenotypes displayed in this figure. The 95% CIs of the fixed effects and phenotype differences estimates are displayed for a single phenotype (cognitive ability).

mean phenotype differences $\hat{\beta}$ coefficient from 1,000 iterations of this procedure, whereas Panel (*B*) displays the $\hat{\beta}$ coefficient from a single iteration. The phenotype differences $\hat{\beta}$ coefficients displayed in Panel (*C*) use data from the One Genotype sample, meaning the estimates plotted against each other in this panel are derived from entirely nonoverlapping samples. Because both the fixed effects and the phenotype differences models leverage only sibling comparisons, all $\hat{\beta}$ coefficients displayed in Fig. 1 are unbiased by environmental confounding.

As can be seen, there is a high correspondence (moving from left to right: $r = 0.99$, $r = 0.92$, and $r = 0.90$) between the fixed effects and phenotype differences estimates, even when fit on nonoverlapping samples. In Panel (*A*), the two estimates are virtually identical; this suggests that differences between the fixed effects and phenotype differences estimates in Panel (*B*) are simply the result of sampling variance. Importantly, the correspondence between fixed effects and phenotype differences estimates is similar across Panels (*B* and *C*), implying that the key phenotype differences assumptions are, indeed, largely met in the WLS One Genotype sample. To achieve a single estimate with the greatest precision, the two estimates displayed in Panel (*C*) can be pooled using inverse variance-weighted meta-analysis. All the $\hat{\beta}$ and SEs displayed in Fig. 1 can be found in Dataset S2.

Next, we use the phenotype differences model to estimate the effects of various polygenic indices on lifespan. While previous studies have shown that both specific genetic variants (30, 31) and polygenic indices (32, 33) are associated with mortality outcomes, the extent to which these associations represent genetic effects versus environmental confounding remains largely unknown. A key challenge for the identification of genetic effects on mortality, both in general and specifically in the WLS, is mortality selection. Panel (*A*) of Fig. 2 provides a binned scatter plot of the likelihood that a WLS respondent is genotyped as a function of their year of death; while no WLS respondents who died before 2006 were genotyped, roughly 80% of respondents who survived to 2015 were genotyped. Panel (*B*) of Fig. 2 displays overlaid histograms of the within-family difference in lifespan for both the One Genotype sample (in red) and the Two Genotypes sample (in blue). While both the One Genotype sample and the Two Genotypes sample may be influenced by mortality selection (as described in *SI Appendix*, section H), we observe that the amount of within-family variation in lifespan is more restricted in the Two Genotypes sample than in the One Genotype sample.[**]

Panel (*C*) of Fig. 2 displays the $\hat{\beta}$ coefficients from regressions of various polygenic indices on lifespan in years. We omit 4 sex-specific polygenic indices and 3 polygenic indices which failed to significantly ($P < 0.05$) predict their target phenotype in between-families regressions fit on the WLS sample, leaving us with polygenic indices for 40 traits (indicated in Dataset S3). Coefficient estimates are pooled via inverse variance-weighted meta-analysis across results from phenotype differences and fixed effects models fit on the One Genotype sample and Two Genotypes sample, respectively. Our analyses revealed polygenic indices for eight different phenotypes that have a $\hat{\beta}$ with a P-value of less than 0.05; however, after implementing a BKY 10% false discovery rate correction, polygenic indices for three phenotypes had statistically significant effects on lifespan: body mass index, self-rated health, and chronic obstructive pulmonary disease.

A 1 SD increase in the self-rated health polygenic index is associated with a 0.19 (0.071) year increase in lifespan; when corrected for measurement error in the polygenic index using the procedure described by Becker et al. (23), the self-rated health coefficient grows to 0.41.[††] A 1 SD increase in the

[**]In the One Genotype sample, the sample variance of the sibling difference in lifespan (in years) is 77.8, whereas in the Two Genotypes sample the sample variance of the sibling difference in lifespan is just 16.8.

[††]Note that the mean absolute difference between sibling pairs in our analytic sample for a given polygenic index is about 0.8 SD, whereas the mean absolute difference for two randomly selected unrelated individuals is approximately 1.1 SD.

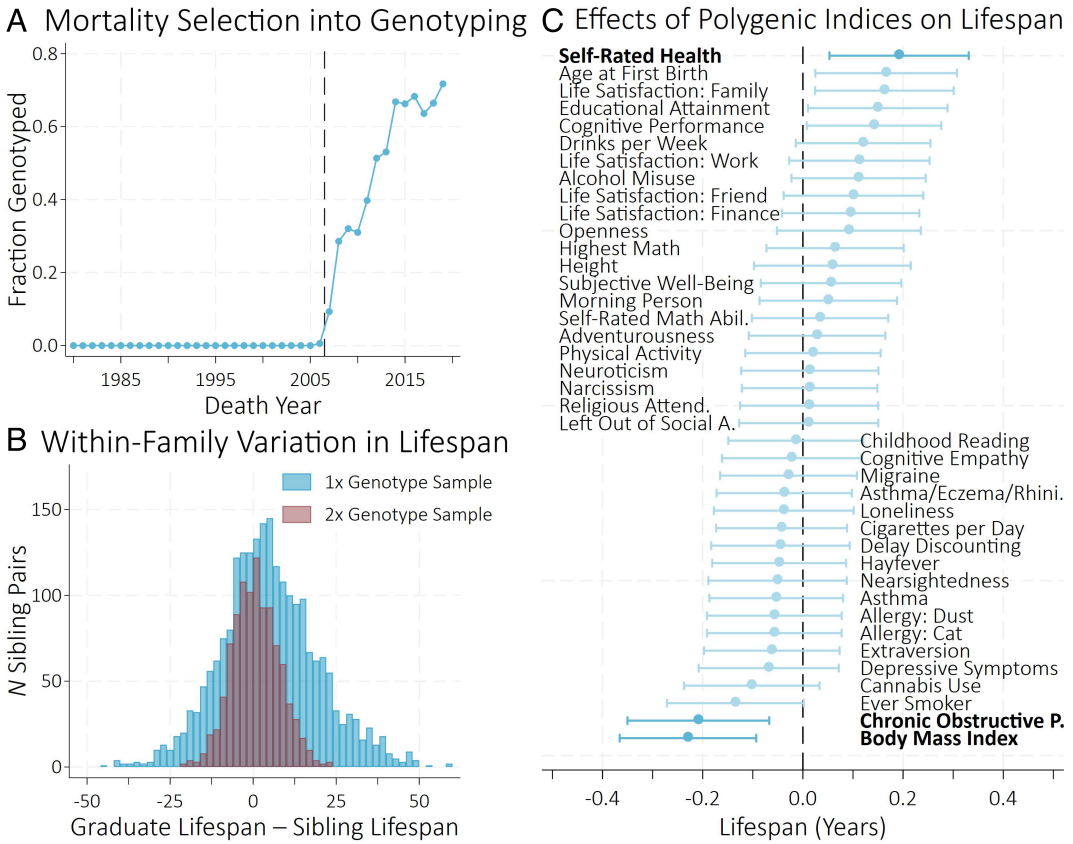

Fig. 2. Using Phenotype Differences to Study Premature Mortality. Panel (*A*) displays the fraction of WLS respondents with genotype data as a function of their year of death, with a vertical line drawn at 2006. Panel (*B*) displays a histogram comparing the within-family variation in lifespan for the One Genotype sample (blue bars) and Two Genotypes sample (red bars). Panel (*C*) displays $\hat{\beta}$ estimates and 95% CIs of the effect of 40 polygenic indices on lifespan in years; bolded polygenic indices are statistically significant after implementing a Benjamini, Krieger, and Yekutieli 10% false discovery rate correction. Sibling difference in lifespan is residualized on a second-degree polynomial of sibling difference in birth year, centered at its mean, prior to conducting phenotype differences regression.

body mass index and chronic obstructive pulmonary disease polygenic indices are associated with 0.23 (0.070) and 0.21 (0.072) year decreases in lifespan, respectively. When corrected for measurement error, these two coefficients grow to 0.31 and 0.62, respectively. In addition, *SI Appendix*, Fig. S1 and Dataset 4 replicate Panel (*C*) of Fig. 2 with the outcome variable instead being a dichotomous variable indicating whether an individual survived until at least age 75.[‡‡]

Discussion

Our analytical and empirical results demonstrate that the phenotype differences model is a robust estimator of genetic effects in the presence of environmental confounding. The estimator, which requires only a single genotype and therefore can be used on a wider range of data sources, has the potential to increase the precision and generalizability of within-family genomic studies. The potential applications of the phenotype differences model extend well beyond the two considered here. More broadly, our work highlights the value of collecting sibling phenotype data, even when sibling genotype data are unavailable.

Indeed, perhaps the most beneficial use cases of the phenotype differences model will come through new data collection efforts and/or creative uses of existing data resources. For phenotypes that are easily reported (like height and educational attainment), data could be collected by surveying unrelated individuals on the phenotypes of their siblings; or, when studying rare and sensitive phenotypes (such as severe mental disorders), sibling phenotypes could be collected by leveraging population registries and other administrative data bases. Table 2 describes these potential applications of the phenotype differences model in more detail.[§§] Notably, the phenotype differences model can also be applied in research using genetic characteristics as an instrumental variable, known as Mendelian randomization (*SI Appendix*, section G); conducting Mendelian randomization within families helps reduce bias that results from violations of the exclusion restriction (35).[¶¶]

The phenotype differences model is complementary to a recently introduced strategy known as Mendelian imputation which, like phenotype differences, leverages known biological features of genetic inheritance to make improvements to the fixed effects model specifically for applications in genomic research (13). Mendelian imputation leverages variant-level identity-by-descent information of genotyped sibling pairs to nonlinearly impute the parental genome; this, in turn, allows researchers to fit within-family models that have increased statistical power and identify additional parameters (specifically, indirect genetic

[‡‡] Of the 5,601 sibling pairs used in the lifespan analyses, a subsample of 4,773 pairs were born before 1945 (and therefore both siblings were able to reach age 75 by the time of mortality data collection in 2020); these 4,773 pairs become the analytic sample for our survival to 75 analyses.

[§§] For more information on the UK Biobank, see Bycroft, Freeman, Petkova et al. (17). For more details on the iPSYCH study, see Pedersen, Bybjerg-Grauholm et al. (34).

[¶¶] The phenotype differences model may be adapted for use in the study of polygenic index-by-environment interactions (36, 37), but such an extension is beyond the scope of the present paper.

Table 2. Potential applications of the phenotype differences model

Collecting sibling reported phenotypes	Imagine researchers decide to expand an existing data repository, like the UK Biobank, to increase the number of genotyped pairs of first-degree relatives available for within-family genomic analyses. Rather than empaneling, interviewing, and genotyping multiple members of the same family, for certain traits which are easily reported—such as height and educational attainment—it is likely more cost-effective to simply ask existing respondents (or a new sample of unrelated individuals) to provide the phenotypes of their siblings. One does not necessarily have to choose between the two strategies—both types of data can be simultaneously collected and utilized. Importantly, the phenotype differences model provides estimates that are robust to asymmetric bias and classical measurement error (e.g., reporting a sibling's height systemically lower or less accurately than one's own height), so long as such biases are not meaningfully genetically caused.
Leveraging existing administrative data	Assembling large samples of genotyped siblings can be challenging even when studying common and easily measured phenotypes, and it is only more difficult when studying rare and highly sensitive phenotypes (such as severe mental disorders). Case–control studies, such as iPSYCH, have achieved the impressive task of assembling a sufficient sample size for molecular genetic analyses, but it is currently difficult to integrate such data sources into a within-family framework. However, the same administrative data used to create iPSYCH could be leveraged to collect the phenotypes of siblings without the need to obtain consent for the collection of additional biological assays.
Accounting for sample selection	As we have shown using the WLS data, the ability to conduct within-families genomic analyses using a single genotype per family can be useful for dealing with issues related to selection into genotyping. In our case, such selection resulted largely from premature mortality, though other forms of selection into genotyping, such as an individual's concerns related to privacy or trust in research institutions, may exist.

effects from parents and siblings). The two methods are, in a sense, symmetric; whereas Mendelian imputation can be used to estimate direct effects in cases where phenotypic data is only available for a single sibling (but genotypic data is available for both siblings), the phenotype differences model allows researchers to estimate direct effects in cases where genotype data is only available for a single sibling (but phenotypic data is available for both siblings).

Our example application of the phenotype differences model to the field of biodemography provides a glimpse into the so-called genetic lottery (38, 39) for premature mortality in mid-century Wisconsin. After a 10% false discovery rate correction, we find that three polygenic indices—body mass index, self-rated health, chronic obstructive pulmonary disease[##]—have statistically significant effects on an individual's lifespan. That is, holding the other circumstances of one's birth constant, if a person were to have inherited a different DNA sequence (and, in turn, a different polygenic index), their expected lifespan would have also changed (40).

Nonetheless, the precise pathways through which the genetic effects observed in this study operate remain largely unknown. While some mistakenly conceptualize the effects of genes as operating strictly within the body, they instead often operate through long, complex causal chains meditated by social and environmental aspects of our world (41–43). Moreover, though the fixed effects and the phenotype differences models are both within-family estimators, the polygenic indices used in this study—like most existing polygenic indices—were constructed using allelic weights from between-family GWAS. So, even though the estimates of genetic effects we present are free of environmental confounding, processes like assortative mating and population stratification at the GWAS stage can influence the resulting polygenic index (and thereby alter which weighted set of genetic variants is tested for effects on a given outcome).

[##]With respect to our results regarding the chronic obstructive pulmonary disease polygenic index, it is worth noting that over 50% of WLS respondents were regular smokers at some point during their lifetime.

In our lifespan analyses, the estimated effects of polygenic indices may well serve as a lower bound of the true effects. Recall that our lifespan variable is right censored in 2020, when the average member of our analytic sample would be just 80 y old. Individuals with a lower genetic risk for premature mortality are more likely to live longer and, therefore, also more likely to be censored, producing attenuation bias toward zero.

However, there is also reason to remain cautious when interpreting our lifespan results. Notably, estimates from both fixed effects models fit on Two Genotypes samples and phenotype differences models fit on One Genotype samples are potentially impacted by selection into genotyping induced collider bias (44). In *SI Appendix*, section H, we provide a mathematical framework for considering selection-related collider bias in fixed effects and phenotype differences models. Unlike in the case of fixed effects, for phenotype differences the direction of such biases is difficult to ascertain. Reassuringly, however, because i) selection into genotyping must be influenced by our genetic predictor in order for collider bias to exist and ii) premature mortality is a key mediator of our genetic predictor's effects on selection, we believe that collider bias is relatively unlikely to introduce type I errors.

Broadly, our premature mortality results show that certain recent between-family genetic discoveries—as summarized by polygenic indices—predict sibling differences in lifespan, an outcome of substantive interest other than the traits that these scores were trained to predict. If, in fact, between-family GWAS were overwhelmingly capturing environmental confounding from population structure and dynastic genetic effects, we would be unlikely to see such a result. As genetic indices continue to become increasingly powerful causal predictors, policymakers may need to increase regulations of the uses of genomic data in order to protect citizens (45) and prevent adverse outcomes in insurance markets (33).

Materials and Methods

The Phenotype Differences Model.

Assumption of equal variance of genetic predictor. The phenotype differences model recovers estimates of genetic effects that are free of environmental

confounding when two key assumptions hold: random assignment of genotype within families, which is an assumption also required by fixed effects, and equal population variance of the genetic predictor in the genotyped and ungenotyped siblings, an assumption unique to phenotype differences.

Let g_{1j} have population variance σ_1^2, and let g_{2j} have population variance σ_2^2. Thus, the assumption of phenotype differences is that $\sigma_1^2 = \sigma_2^2$. Because we do not directly observe g_{2j}, this assumption is inherently untestable. However, if we know that we observe the genotype of a random sibling (i.e., random sampling of genotype within families), then this assumption is trivially met.

When g_{ij} is a normally distributed polygenic index, there exists no mechanical mean-variance dependence. In this case, average differences in genetic characteristics between the genotyped sibling and the ungenotyped siblings are not inherently a problem; this is because the existence of genetic characteristics that linearly increase or decrease an individual's likelihood of being the genotyped (versus ungenotyped) sibling would not distort the *variance* of g_{1j} (compared to g_{2j}) and therefore will not violate our assumption. However, nonlinear and/or nonmonotonic selection into genotyping may impact the variance of g_{1j} (relative to g_{2j}) and therefore induce violations.

When g_{ij} is the number of major alleles at a single-nucleotide polymorphism (i.e., taking the value of 0, 1, or 2), such as in GWAS, mean differences likely entail variance differences; therefore, systematic differences in the allele frequencies across the genotyped and ungenotyped sibling will most likely induce violations of the equal variance assumption. When this key assumption is not met, estimates of the true genetic effect become biased as a function of the degree of variance discordance. Beginning from *SI Appendix*, Eq. **S18** of the phenotype differences consistency proof, we observe that:

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{\mathrm{PD}} | \beta = \beta + \frac{\beta}{1-\rho} \times \left(\rho - \rho \times \frac{\sigma_1}{\sigma_2} \right). \quad [3]$$

When $\rho = \frac{1}{2}$, this reduces to:

$$\operatorname*{plim}_{N\to\infty} \hat{\beta}^{\mathrm{PD}} | \beta = \beta + \beta \times \left(1 - \frac{\sigma_1}{\sigma_2} \right). \quad [4]$$

We empirically test the equal variance assumption in two different ways. First, while it is not possible to test whether a difference in genetic predictor variance exists between individuals who survived until the year that genotyping began (2006) and those who did not, we can artificially "move back" the genotyping date. Specifically, we test whether the 47 polygenic indices have equal variance among genotyped individuals who survived until 2020 (N = 5,406) and among genotyped individuals who died prior to 2017 (N = 2,302) using a Brown–Forsythe test (46, 47). Across the 47 separate tests, we observe no $P < 0.05$, consistent with the lack of systematic differences in genetic predictor variance between genotyped individuals who did versus did not survive until 2020.

Second, while we definitionally do not observe the genetic predictor of both siblings in our One Genotype sample, we often do observe phenotype variables for both siblings. *SI Appendix*, Table S2 displays results from Brown–Forsythe equal variance tests using phenotypic data. We focus on 7 phenotypes that are largely continuously distributed to reduce the presence of mean-variance dependencies (which do not exist for normally distributed polygenic indices). Although one phenotype (neuroticism) has $P < 0.05$, the magnitude of any phenotypic variance discordance that we observe between the genotyped and ungenotyped siblings is generally small and insignificant.

Comparative precision. As we show in *SI Appendix*, section F, when the effects of g_{ij} are small, fixed effects estimates will have asymptotically identical SEs to phenotype differences estimates derived from the same number of genotypes (although, when using phenotype differences, one typically has half as many genotypes per family). However, as the fraction of within-family outcome variation explained by g_{ij} grows, phenotype difference provides less precise effect estimates than fixed effects per genotype. Specifically, for models fit on the same number of sibling pairs–that is, $N = M$, implying twice as many genotypes used for fixed effects than for phenotype differences–this decrease in comparative precision is governed by the following formula:

$$\operatorname*{plim}_{N\to\infty} \frac{\sqrt{\mathrm{var}(\hat{\beta}^{\mathrm{FE}})}}{\sqrt{\mathrm{var}(\hat{\beta}^{\mathrm{PD}})}} = \frac{\sqrt{1-\phi}}{\sqrt{4-\phi}}. \quad [5]$$

Here, ϕ is the fraction of within-family variation in the y_{ij} that is explained by within-family variation in g_{ij}:

$$\phi = \frac{\mathrm{cov}(g_{1j} - g_{2j}, y_{1j} - y_{2j})}{\mathrm{var}(y_{1j} - y_{2j})} \quad [6]$$

The within-R^2 of the fixed effects model can provide a useful estimate of ϕ. For currently available genetic predictors, this reduction in precision due to $\phi > 0$ is relatively modest. For example, when g_{ij} is the polygenic index for height–one of the most predictive scores in the SSGAC repository–$\hat{\phi} = 0.11$ in the WLS Two Genotypes sample. Therefore, the comparative precision *per genotype* is $2 \times \sqrt{\frac{1-0.11}{4-0.11}} = 0.96$. That is, when fit on the same number of genotypes, fixed effects estimates of the effect of the height polygenic index on phenotypic height will have expected SEs that are 96% as large as the SEs of phenotype difference estimates. See *SI Appendix*, Fig. S2 for a graphical display of the relationship between ϕ and the asymptotic ratio of SEs of the fixed effects and phenotype differences models. Note that, for ease of exposition, these derivations assume ρ is known and equal to $\frac{1}{2}$.

Assortative mating and estimating ρ. When meaningful assortative mating exists, maternal and paternal genetic predictors become correlated with one another and $\rho \neq \frac{1}{2}$. In such a case, $\hat{\rho}$ can be empirically estimated from a sample of fully genotyped sibling pairs, as we do with the WLS Two Genotypes Sample. It is important to note that while ρ is fixed, $\hat{\rho}$ is derived from a finite sample of M fully genotyped sibling pairs and is therefore random. Thus, when fitting phenotype differences models with an estimated within-family correlation of genetic predictor, the uncertainty from $\hat{\rho}$ propagates to $\hat{\beta}$ and must be accounted for when constructing SEs. In *SI Appendix*, section E, we use the multivariate delta method (28, 29, 48) to show that the variance of $\hat{\beta}^{PD}$ in this case is approximately:

$$\widehat{\mathrm{var}}(\hat{\beta}^{\mathrm{PD}}) \approx \underbrace{\frac{1}{N-2} \times \frac{\widehat{\mathrm{var}}(\hat{\varepsilon}_{1j})}{(1-\hat{\rho})^2}}_{\text{Typical OLS variance}} + \underbrace{\frac{(\hat{\beta}^{PD})^2 (1+\hat{\rho})^2}{M-3}}_{\text{Uncertainty due to } \hat{\rho}}, \quad [7]$$

where N is the number of sibling pairs in the phenotype differences sample and M is the number of sibling pairs used to estimate $\hat{\rho}$. As can be seen in *SI Appendix*, Fig. S3, when $M > 1{,}000$, this SE adjustment ends up being relatively small.

Nonpaternity events. Occasionally, sibling pairs believe themselves to be full biological siblings but are, in actuality, only half-siblings. Because half-siblings have an expected within-family correlation of genetic predictor of $\frac{1}{4}$ (under random mating), sufficient prevalence of nonpaternity events can produce $\rho < \frac{1}{2}$. In such a case, phenotype differences estimates become biased away from 0. In the WLS Two Genotypes sample, approximately one-in-fifty sibling pairs who self-report being full biological siblings are actually half-siblings. If the variance of the overall distribution of g_{ij} is identical for full and half-siblings, we would expect this frequency of nonpaternity events to induce $\rho \approx 0.495$. Such a small departure from $\rho = \frac{1}{2}$ will have only a trivial impact on phenotype differences estimates. Importantly, because misreporting individuals are unaware that they are half siblings, it is unlikely that there exist meaningful environmental differences between such half siblings that are correlated with paternal genotype, so confounding from population stratification and/or dynastic effects is unlikely.

Empirical Application.

The wisconsin longitudinal study. The WLS is a survey based on a $\frac{1}{3}$ sample of all 1957 Wisconsin high school graduates and a randomly selected sibling of these graduates (22). The graduate respondents were originally empaneled

with an in-person questionnaire at age 18 in 1957, which was followed by data collection at ages 25, 36, 54, 65, and finally 72 in 2012. The WLS includes a wide range of administrative and prospectively collected survey data from early life, adolescence, and adulthood. Genetic samples were assayed from saliva for a subsample of consenting WLS graduates and siblings. Genotyping was performed using the Illumina HumanOmniExpress 24 BeadChip arrays (Version 1/1.1; Illumina). We restrict our analytic sample to individuals of European ancestries because only these respondents have nonmissing polygenic index data in the SSGAC repository. However, due to the ancestral homogeneity of the WLS sample, this restriction is relatively inconsequential in practice; of the 9,012 genotyped WLS respondents, 8,927 were identified as European ancestries and have valid polygenic index data. While within-family models are robust to biases stemming from ancestry-related environmental confounding, including diverse ancestries together in a single phenotype differences model could produce inflated values of $\hat{\rho}$ [due to ancestral assortative mating (49, 50)]. For the sake of open science, our main text analyses utilize the public-use release of the WLS data in which it is impossible to identify the 48 half siblings in the Two Genotypes Sample who believe themselves to be full siblings; however, in *SI Appendix*, Tables S3–S6, we rerun our analyses using the restricted-use data with these 48 cases removes and our findings our substantively identical.

Polygenic indices and target phenotype variables. The polygenic indices used in this study are drawn from Version 1.1 of the Social Science Genomics Association Consortium Polygenic Index Repository (23). All polygenic indices are standardized over the full sample of genotyped WLS graduates. A list of which polygenic indices are utilized in each analysis presented in this study can be found in *SI Appendix*, Table S1.

We align our phenotypic variables with those used in as Becker et al. (23); see supplementary table 12 of that paper for a list of the specific WLS survey items used. Though the repository contains 47 distinct polygenic indices, in the WLS, there only exist phenotype data as a subset of 30 traits. When multiple measurements were available, for variables such as depression, we first standardize the variable within each wave and then take the average for an individual across waves. For variables like educational attainment with multiple measurements, we take the maximum value across waves. Finally, all phenotypes are standardized over the full sample of genotyped WLS graduates.

Mortality variables. The mortality data used in this study is derived from the National Death Index (51)–importantly, such data are available for all members of the WLS, regardless of how long they remained empaneled in the study. Mortality data are right censored in 2020; that is, we observe all deaths through December of 2019. Our key mortality variable is lifespan in years, which is calculated as the difference between death year and birth year (for individuals who are still alive, we use 2020 as their death year). Although this lifespan variable is right censored, because genotype is randomly assigned within families, we would not expect siblings differences in polygenic indices to be correlated with siblings differences in birth year. In *SI Appendix*, Fig. S1 and Dataset S4, we repeat our mortality analyses using a dichotomous indicator for survival to age 75. Sibling differences for these two mortality outcomes were residualized on a second-order polynomial of sibling difference in birth year.

Data, Materials, and Software Availability. Our main text analyses utilize the WLS long-form survey data (v14.03) and the WLS SSGAC polygenic index repository (v1.1). Both of these data sources are publicly available and can be downloaded from the WLS website) (citealpbib52). All syntax files needed to replicate our main text analyses are available at the following link: https://github.com/sam-trejo/phenotype_differences (citealpbib53). Supplemental analyses in *SI Appendix* use restricted-use WLS genomic data to conduct a set of robustness checks. These data can be accessed by qualified researchers via application to the WLS.

ACKNOWLEDGMENTS. We would like to thank Alan Aw, Daniel Benjamin, Dalton Conley, Benjamin Domingue, Jason Fletcher, Qiongshi Lu, Iain Mathieson, Brandon Stewart, Marissa Thompson, and Elliot Tucker-Drob for helpful comments on early versions of this manuscript. We are also grateful to Pamela Herd, Carol Roan, Kamil Sicinski, and the Wisconsin Longitudinal Study staff for access to the restricted-use data. Our research has also benefited from the use of the Social Science Genetic Association Consortium's Polygenic Index Repository. This work has been supported, in part, by the Institute of Education Sciences under Grant No. R305B140009. All opinions expressed are those of the authors alone and should not be construed as representing the opinions of any institution.

1. A. Abdellaoui, L. Yengo, K. J. Verweij, P. M. Visscher, 15 years of GWAS discovery: Realizing the promise. *Am. J. Hum. Genet.* **110**, 179–194 (2023).
2. J. Novembre *et al.*, Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
3. A. A. Zaidi, I. Mathieson, Demographic history mediates the effect of stratification on polygenic scores. *eLife* **9**, e61548 (2020).
4. A. Abdellaoui *et al.*, Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* **3**, 1332–1342 (2019).
5. A. Kong *et al.*, The nature of nurture: Effects of parental genotypes. *Science* **359**, 424–428 (2018).
6. S. Trejo, B. W. Domingue, Genetic nature or genetic nurture? Introducing social genetic parameters to quantify bias in polygenic score analyses. *Biodemography Soc. Biol.* **64**, 187–215 (2018).
7. L. J. Howe *et al.*, Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nat. Genet.* **54**, 581–592 (2022).
8. C. Veller, G. M. Coop, Interpreting population-and family-based genome-wide association studies in the presence of confounding. *PLoS Biol.* **22**, e3002511 (2024).
9. C. Veller, M. Przeworski, G. Coop, Causal interpretations of family GWAS in the presence of heterogeneous effects. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2401379121 (2024).
10. D. J. Benjamin, D. Cesarini, P. Turley, A. S. Young, *Social-Science Genomics: Progress, Challenges, and Future Directions* (Working Paper Series no. 32404, National Bureau of Economic Research, 2024). http://www.nber.org/papers/w32404. Accessed 1 July 2024.
11. A. I. Young *et al.*, Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* **50**, 1304–1310 (2018).
12. E. M. Eilertsen *et al.*, Direct and indirect effects of maternal, paternal, and offspring genotypes: Trio-GCTA. *Behav. Genet.* **51**, 154–161 (2021).
13. A. I. Young *et al.*, Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nat. Genet.* **54**, 897–905 (2022).
14. C. A. Rietveld *et al.*, Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychol. Sci.* **25**, 1975–1986 (2014).
15. B. W. Domingue, D. W. Belsky, D. Conley, K. M. Harris, J. D. Boardman, Polygenic influence on educational attainment: New evidence from the national longitudinal study of adolescent to adult health. *AERA Open* **1**, 2332858415599972 (2015).
16. A. Sjölander, T. Frisell, S. Öberg, Sibling comparison studies. *Annu. Rev. Stat. Appl.* **9**, 71–94 (2022).
17. C. Bycroft *et al.*, The UK biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
18. J. D. Angrist, J. S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, 2009).
19. A. Sanz-de Galdeano, A. Terskaya, Sibling differences in genetic propensity for education: How do parents react? *Rev. Econ. Stat.*, 1–44 (2023).
20. R. Frisch, F. V. Waugh, Partial time regressions as compared with individual trends. *Econom. J. Econom. Soc.* **1**, 387–401 (1933).
21. M. C. Lovell, Seasonal adjustment of economic time series and multiple regression analysis. *J. Am. Stat. Assoc.* **58**, 993–1010 (1963).
22. P. Herd, D. Carr, C. Roan, Cohort profile: Wisconsin longitudinal study (WLS). *Int. J. Epidemiol.* **43**, 34–41 (2014).
23. J. Becker *et al.*, Resource profile and user guide of the polygenic index repository. *Nat. Hum. Behav.* **5**, 1744–1758 (2021).
24. D. W. Belsky *et al.*, Genetic analysis of social-class mobility in five longitudinal studies. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7275–E7284 (2018).
25. J. J. Lee *et al.*, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
26. S. Trejo, Exploring the fetal origins hypothesis using genetic data. *Soc. Forces* **102**, 1555–1581 (2024).
27. Y. Benjamini, A. M. Krieger, D. Yekutieli, Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507 (2006).
28. J. L. Doob, The limiting distributions of certain statistics. *Ann. Math. Stat.* **6**, 160–169 (1935).
29. R. Dorfman, A note on the! d-method for finding variance formulae *Biom. Bull.* **1**, 129–138 (1938).
30. P. R. Timmers *et al.*, Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife* **8**, e39856 (2019).
31. J. Deelen *et al.*, A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **10**, 3669 (2019).
32. R. E. Marioni *et al.*, Genetic variants linked to education predict longevity. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13366–13371 (2016).
33. R. K. Linnér, P. D. Koellinger, Genetic risk scores in life insurance underwriting. *J. Health Econ.* **81**, 102556 (2022).
34. C. B. Pedersen *et al.*, The iPSYCH2012 case-cohort sample: New directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
35. B. Brumpton *et al.*, Avoiding dynastic, assortative mating, and population stratification biases in mendelian randomization through within-family analyses. *Nat. Commun.* **11**, 3519 (2020).
36. B. W. Domingue, S. Trejo, E. Armstrong-Carter, E. M. Tucker-Drob, Interactions between polygenic scores and environments: Methodological and conceptual challenges. *Sociol. Sci.* **7**, 465–486 (2020).
37. R. Johnson, R. Sotoudeh, D. Conley, Polygenic scores for plasticity: A new tool for studying gene-environment interplay. *Demography* **59**, 1045–1070 (2022).

38. J. M. Fletcher, S. F. Lehrer, Genetic lotteries within families. *J. Health Econ.* **30**, 647–659 (2011).
39. K. P. Harden, *The Genetic Lottery: Why DNA Matters for Social Equality* (Princeton University Press, 2021).
40. S. Trejo, D. O. Martschenko, Beware the phony horserace between genes and environments. *Behav. Brain Sci.* **46**, e228 (2023).
41. S. H. Barcellos, L. S. Carvalho, P. Turley, Education can reduce health differences related to genetic risk of obesity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9765–E9772 (2018).
42. K. Rimfeld *et al.*, Genetic influence on social outcomes during and after the soviet era in Estonia. *Nat. Hum. Behav.* **2**, 269–275 (2018).
43. P. Herd *et al.*, Genes, gender inequality, and educational attainment. *Am. Sociol. Rev.* **84**, 1069–1098 (2019).
44. F. Elwert, C. Winship, Endogenous selection bias: The problem of conditioning on a collider variable. *Annu. Rev. Sociol.* **40**, 31–53 (2014).
45. L. H. Seaver *et al.*, Points to consider to avoid unfair discrimination and the misuse of genetic information: A statement of the american college of medical genetics and genomics (ACMG). *Genet. Med.* **24**, 512–520 (2022).
46. M. B. Brown, A. B. Forsythe, Robust tests for the equality of variances. *J. Am. Stat. Assoc.* **69**, 364–367 (1974).
47. R. G. O'Brien, Robust techniques for testing heterogeneity of variance effects in factorial designs. *Psychometrika* **43**, 327–342 (1978).
48. E. S. Lee, R. N. Forthofer, *Analyzing Complex Survey Data* (Sage Publications, 2005).
49. J. Y. Zou *et al.*, Genetic and socioeconomic study of mate choice in latinos reveals novel assortment patterns. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13621–13626 (2015).
50. A. Mas-Sandoval, S. Mathieson, M. Fumagalli, The genomic footprint of social stratification in admixing American populations. *eLife* **12**, e84429 (2023).
51. National Center for Health Statistics, National Death Index. https://www.cdc.gov/nchs/ndi/index.html. Accessed 18 July 2024.
52. R. M. Hauser, W. H. Sewell, P. Herd, M. Engelman, Wisconsin Longitudinal Study (WLS) [graduates, siblings, and spouses]: 1957-2020 (Version 14.03, WLS, University of Wisconsin-Madison, Madison, WI, 2024). https://researchers.wls.wisc.edu/data/. Accessed 1 July 2024.
53. S. Trejo, K. Kanopka, Phenotype differences in the WLS replication code. Github. https://github.com/sam-trejo/phenotype_differences. Deposited 1 October 2024.