PNAS RESEARCH ARTICLE | SOCIAL SCIENCES

Advantageous early-life environments cushion the genetic risk for ischemic heart disease

Samuel Baker[a], Pietro Biroli[b], Hans van Kippersluis[c], and Stephanie von Hinke[a,d,1]

Edited by Dalton Conley, Princeton University, Princeton, NJ; received August 15, 2023; accepted April 18, 2024

In one of the first papers on the impact of early-life conditions on individuals' health in older age, Barker and Osmond [*Lancet*, 327, 1077–1081 (1986)] show a strong positive relationship between infant mortality rates in the 1920s and ischemic heart disease in the 1970s. We merge historical data on infant mortality rates to 370,000 individual records in the UK Biobank using information on local area and year of birth. We replicate the association between the early-life infant mortality rate and later-life ischemic heart disease in our sample. We then go "beyond Barker," by showing considerable genetic heterogeneity in this association that is robust to within-area as well as within-family analyses. We find no association between the polygenic index and heart disease in areas with the lowest infant mortality rates, but a strong positive relationship in areas characterized by high infant mortality. These findings suggest that advantageous environments can cushion one's genetic disease risk.

Barker hypothesis | developmental origins | gene–environment interplay | ESSGN

Significance

We show evidence of substantial nature–nurture interplay in the development of later-life ischemic heart disease. We find that advantageous early-life environments, proxied by one's local infant mortality rate during the prenatal period, substantially cushion one's genetic risk for heart disease. Our findings highlight that early life conditions can mitigate arguably unfair genetic inequalities.

Ischemic heart disease is the most common cause of death in the developed world, accounting for more than 9 million deaths worldwide in 2016 (1). It is well known that adverse conditions during the prenatal and early childhood period affect cardiovascular health (2–5), as well as other health and economic outcomes in older age (for reviews, see, e.g., refs. 6–8). The vast literature on the so-called developmental origins of health and disease (DOHaD) hypothesis spans both the medical and social sciences (see, e.g., refs. 4, 9–14). Its best-known proponent is the British physician and epidemiologist David Barker. In one of the first of a set of papers, Barker and colleagues showed a strong positive geographical relationship between the infant mortality rate in the 1920s and ischemic heart disease mortality in the 1970s (15).

In addition to such "environmental" circumstances affecting the development of heart disease, genetic factors are known to play an important role. Twin studies have shown that heart disease is heritable (e.g., ref. 16), with more recent Genome-Wide Association Studies (GWAS) starting to unravel the specific genetic variants implicated in the disease (see, e.g., refs. 17–25). These gene-discovery studies have linked dozens of independent genetic loci to heart disease, facilitated a better understanding of the causal risk factors, and informed the development of new therapeutics (see, e.g., ref. 26).

Whereas the role of environmental and genetic main effects have been widely documented, Tiffany et al. (27) argue that the environmental and biological mechanisms (or $G \times E$) that lead to cardiovascular disease remain understudied and incompletely understood. Some earlier studies discuss the potential role of $G \times E$ interactions for cardiovascular disease (see, e.g., refs. 28–30), but none of these focus on the long-held DOHaD hypothesis. This is what we explore here, investigating the extent to which one's underlying genetic risk is moderated by the early life environment, or vice versa, examining to what extent the early life environments moderates one's genetic risk for heart disease.

There are multiple reasons why "nature" might interact with "nurture" to shape individuals' outcomes. First, there may be a biological channel: genetic variation may predispose individuals to certain health conditions or behaviors, but the phenotypic effect can depend on environmental circumstances (31, 32), e.g., through environmentally induced epigenetic regulation (see, e.g., refs. 27 and 33). The role of epigenetics driven by environmental circumstances in the development of heart disease has been studied extensively over the last 15 y (e.g., refs. 34–37). However, even if early life circumstances cause epigenetic changes (e.g., ref. 34), this does not necessarily imply that they affect later life outcomes. Indeed, there could be compensatory investments by the individual or her environment that offset the development of any epigenetic effects. For example, parental investments in children may differ in response to early life circumstances, potentially mitigating or reinforcing their effects (e.g., refs. 38–43). In other words, epigenetics in

Author affiliations: [a]School of Economics, University of Bristol, Bristol BS8 1TU, United Kingdom; [b]Department of Economic Sciences, University of Bologna, Bologna 40126, Italy; [c]Erasmus School of Economics, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands; and [d]Institute for Fiscal Studies, London WC1E 7AE, United Kingdom

Author contributions: P.B., H.v.K., and S.v.H. designed research; S.B., P.B., H.v.K., and S.v.H. performed research; S.v.H. analyzed data; S.B. digitized the historic data and merged it with individual-level data; and S.B., P.B., H.v.K., and S.v.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: s.vonhinke@bristol.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2314056121/-/DCSupplemental.

Published June 25, 2024.

response to environmental circumstances is not a sufficient condition for the existence of $G \times E$ interaction effects.

Furthermore, an interaction between nature and nurture need not be epigenetic; one's genetic predisposition for certain health conditions or behaviors may simply reflect the "type" of individual, with different "types" responding differently to different environments. For example, variation in the nicotinic acetylcholine receptor has been shown to moderate the influence of tobacco taxation on smoking (44, 45). A differential response to tobacco taxation is unlikely to occur due to epigenetic changes or compensatory effects and is more likely to indicate that those with the protective polymorphism can be seen as different types who respond differently to different environments. In other words, epigenetics is also not a necessary condition for the existence of $G \times E$ interaction effects.

Our analysis does not distinguish between the different mechanisms. Nevertheless, it allows us to advance our knowledge on the interplay between genetic and environmental factors that drive the world's major killer. We argue that this is not just a fundamental scientific advance, but may also inform governments on how environmental policies can reduce the—arguably unfair—inequalities in heart disease arising from one's genetic variation.

Results

We use the UK Biobank (46) to explore the role of $G \times E$ in heart disease, where we define G as one's polygenic index for heart disease and E denotes the local infant mortality rate that individuals were exposed to during the pregnancy period. We standardize both to have mean zero and SD one. We distinguish between the full sample and the sibling sample, where the latter allows us to estimate direct genetic effects and reduce confounding by family socioeconomic characteristics. *SI Appendix*, section A describes the data in more detail, and shows the spatial and time variation in infant mortality rates for the UK Biobank cohorts.

Between-Family Estimates. Column (1) in Table 1 presents the estimates from a linear regression of the binary indicator of ischemic heart disease on one's local infant mortality rate during the prenatal period as well as its square, controlling for gender and year × month of birth dummies. This confirms the results in ref. 15, showing a significant relationship between adverse early life conditions and later life cardiovascular health. Column (2) shows the predictive power of the polygenic index, corroborating the findings from existing GWAS and showing a significant positive relationship between the polygenic index for heart disease and the individual diagnosis. Including the infant mortality rate and the polygenic index simultaneously, as in columns (3) and (4), shows associations of similar magnitude. Column (4) further adds the interaction term between genetic variation and the prenatal environment, showing a positive and significant coefficient. In other words, an increase in the local infant mortality rate increases the probability of being diagnosed with ischemic heart disease by more for those with a high polygenic index for heart disease. Any gene-environment correlations (*rGE*) complicate the interpretation of the $G \times E$ coefficient; see (e.g., ref. 47). We explore this in *SI Appendix*, section B, showing some evidence of *rGE* in the full sample, but not within families. This suggests that the restriction to siblings allows us to identify true $G \times E$ interactions, rather than spurious $G \times G$ or $E \times E$ interactions.

To explore nonlinearities in the $G \times E$ interaction effect, we plot the relationship between the residualized outcome and the local infant mortality rate experienced during pregnancy using local polynomial plots, where the residual is obtained from a regression of ischemic heart disease on the set of covariates. Fig. 1 shows this for the quintiles of the polygenic index [*SI Appendix*, Fig. C.1 shows this using the regression estimates from column (4) of Table 1], illustrating a diverging pattern as the infant mortality rate increases. The figure suggests that one's polygenic index for heart disease makes little difference in explaining the probability of developing ischemic heart disease among those born in advantageous environments, yet plays a major role when exposed to adverse early life environments.

Table 1. Gene–environment interplay for ischemic heart disease

	Ischemic heart disease				
	(1)	(2)	(3)	(4)	(5)
IMR	0.0157***		0.0130***	0.0132***	0.0040***
	(0.0016)		(0.0013)	(0.0013)	(0.0011)
IMR^2	−0.0010*		−0.0006	−0.0007	0.0004
	(0.0005)		(0.0005)	(0.0005)	(0.0004)
PGI		0.0315***	0.0314***	0.0313***	0.0310***
		(0.0005)	(0.0005)	(0.0006)	(0.0006)
PGI^2		0.0042***	0.0042***	0.0042***	0.0042***
		(0.0004)	(0.0004)	(0.0004)	(0.0004)
IMR × PGI				0.0098***	0.0099***
				(0.0006)	(0.0006)
Covariates	Yes	Yes	Yes	Yes	Yes
District FEs	No	No	No	No	Yes
Mean	0.13	0.13	0.13	0.13	0.13
R^2	0.07	0.08	0.08	0.08	0.08
No. of obs.	378,785	378,785	378,785	378,785	378,785

Notes: Covariates include gender and all year × month of birth dummies. "Mean" is the mean of the dependent variable. IMR denotes the infant mortality rate during the prenatal period and PGI is the polygenic index for heart disease; both are standardized to have mean zero and SD one. Robust SE clustered by district in parentheses. *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

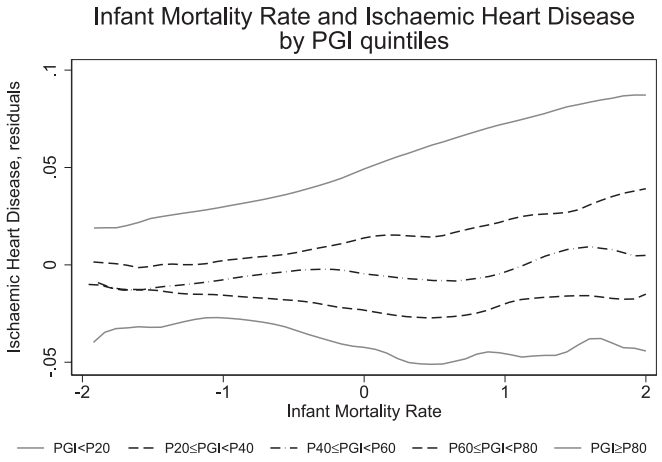

Fig. 1. $G \times E$ interplay for ischemic heart disease, local polynomial plots for the five quintiles of the PGI; full sample.

We next explore the extent to which the estimates in Table 1 are capturing unobserved (socioeconomic) differences across districts. *SI Appendix*, section A shows a correlation between infant mortality and the districts' socioeconomic composition, suggesting that part of the raw association may stem from confounding: lower social class families may experience both higher infant mortality rates as well as ischemic heart disease in later life producing a spurious association between infant mortality rates and later-life heart disease. Column (5) of Table 1 therefore adds district of birth fixed effects. The inclusion of both time and district fixed effects implies that the identification is driven by geographical as well as temporal variation. In other words, the district fixed effects take out average differences across districts in the IMR, and the year fixed effects take out average differences over the birth years. The remaining variation we use stems from variations in the actual experienced infant mortality rates in a certain district in a certain year that are not explained by these average differences across years and districts. The estimates that account for district fixed effects suggest that over half of initial association is due to time-invariant differences between districts. A one SD increase in the infant mortality rates is associated with a 0.4 percentage point (3%) increase in the probability of being diagnosed with ischemic heart disease (with its square insignificantly different from zero). The coefficients on the polygenic index and the interaction remain unchanged.

Within-Family Estimates. Table 2 provides the estimates for the reduced sibling sample without family fixed effects [columns (1)–(3)], and with family fixed effects [columns (4)–(6)]. Restricting to the sibling sample substantially increases all SE, yet only slightly reduces the coefficients of the infant mortality rate, with no changes in the polygenic index. Whereas the estimate on the interaction almost halves, suggesting that the sibling sample is different from the full UK Biobank, it remains statistically significant.

Moving to the estimates from the family fixed effects specifications [columns (4)–(6)], we find a very small reduction in the parameter estimate on the polygenic index. This suggests that most of the genetic effect stems from a direct genetic effect, with little influence of demography and indirect genetic effects for ischemic heart disease. The inclusion of family fixed effects attenuates the infant mortality rate coefficient, rendering it insignificantly different from zero. However, the SE increase even further, suggesting a strong reduction in the power of our family fixed effects analysis. Indeed, the confidence intervals on the infant mortality rate for the analysis with and without family fixed effects in Table 2 overlap, and we cannot claim a null-result with certainty. The $G \times E$ interaction effect, however, remains very similar, again suggesting that the infant mortality rate does not increase the probability of being diagnosed with ischemic heart disease for those with an average polygenic index, yet that it does increase this probability for those with a high polygenic index. Or vice versa, a one SD increase in the polygenic index increases the probability of being diagnosed with heart disease, and this effect is larger for those exposed to higher infant mortality rates during the intrauterine period. *SI Appendix*, Fig. C.2 confirms this with nonlinear local polynomial plots using the residualized outcome.

There are three key issues with the family fixed effects specification: 1) a linearity assumption when studying interaction effects for binary outcomes (48), 2) the problem of incidental parameters in a fixed effects regression that cannot be interacted with *IMR* and *PGI* (as in ref. 49), and 3) the fact that the

Table 2. Using the sibling sample and including family fixed effects

	Reduced (sibling) sample			With family fixed effects		
IMR	0.0087**	0.0077**	0.0080**	−0.0000	−0.0002	0.0002
	(0.0035)	(0.0035)	(0.0035)	(0.0061)	(0.0061)	(0.0061)
IMR^2	0.0008	0.0011	0.0010	0.0063***	0.0065***	0.0064***
	(0.0017)	(0.0016)	(0.0016)	(0.0023)	(0.0023)	(0.0023)
PGI		0.0289***	0.0285***		0.0201***	0.0197***
		(0.0017)	(0.0017)		(0.0034)	(0.0034)
PGI^2		0.0061***	0.0061***		0.0066***	0.0066***
		(0.0012)	(0.0012)		(0.0020)	(0.0020)
IMR × PGI			0.0056***			0.0057**
			(0.0019)			(0.0028)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE	No	No	No	No	No	No
R^2	0.07	0.08	0.08	0.57	0.57	0.57
N	33,054	33,054	33,054	33,054	33,054	33,054

Notes: Columns (1)–(3) show robust SE clustered by district on the sibling sample. Columns (4)–(6) use two-way clustering by family and district on the sibling sample. IMR denotes the infant mortality rate during the prenatal period and PGI is the polygenic index for heart disease; both are standardized to have mean zero and SD one. *$P < 0.1$, **$P < 0.05$, ***$P < 0.01$.

interaction effect in a within-family specification may be partially identified off of between-family variation (50). Since including fixed effects in a logistic specification is problematic (51–53) and does not deal with the incidental parameters problem, nor the identification issue, we explore these issues in two ways. First, we rerun our main specification, but rather than including family fixed effects, we control for the family mean infant mortality rate, the mid-parental (imputed) polygenic index, as well as the family mean of the $G \times E$ interaction. We impute the parental polygenic index following (54, SNIpar). Interacting these family means with *IMR* and *PGI* (as in ref. 49) deals with issue (2), and specifying a logistic regression addresses issue (1). An additional advantage of this approach may be one of increased precision in the PGI main effect as well as its interaction with the infant mortality rate. The findings from this specification, reported in *SI Appendix*, section D, are consistent with a positive interaction effect. Furthermore, including the family mean infant mortality rate instead of family fixed effects leads to a significant positive infant mortality rate coefficient. Finally, the inclusion of the mid-parental (imputed) polygenic index leads to more precision in the main PGI coefficient, but this does not carry over to the interaction term, which is estimated with less precision compared to that in Table 2. Apart from reduced precision in the coefficient of the interaction term, the approach also does not deal with issue (3): the interaction effect may still be partially identified off of between-family variation.

Second, therefore, we use so-called "segmented regressions" stratifying the sample by either the infant mortality rate or the polygenic index and estimating the coefficient of the other variable. In our setting, this approach has two main advantages. First, it allows us to use a logistic (or probit) regression to model the binary outcome without the need to include fixed effects and without having to estimate the $G \times E$ interaction term, solving issue (1). Second, using a within-family mean-deviation specification in the segmented regressions allows us to effectively estimate a within-family model, but does not require the inclusion of the interaction term, family fixed effects, nor interactions between the X's or *PGI* and family fixed effects, solving issues (2) and (3). The alternative of taking family deviations of all right-hand side variables does identify their main effects, but not the interaction term. Indeed, the interaction between the family deviations is not the same as the family deviations of the interaction, meaning that the former does not produce meaningful estimates (50). Finally, the within-family mean-deviation specification also avoids singleton observations in the subgroups being dropped from the estimation. Our results are consistent with the findings from the within-family specification. However, while the difference between the high and low polygenic index groups and between those exposed to a high and low infant mortality rate are salient and relatively large, so are the SE, implying that the differences across groups do not reach statistical significance at 5%. Nevertheless, the estimates do suggest that our results are robust to the nonlinear modeling of the binary outcome of heart disease, but also that the interaction term is primarily identified off of within-family variation.

In *SI Appendix*, section E, we explore potential mechanisms, investigating whether the main and interaction effects are associated with specific behaviors and other outcomes that also associate with heart disease, including BMI, blood pressure, height, drinking, and smoking. Although we find some significant effects in the between-family analysis, including family fixed effects renders most insignificantly different from zero. Only for diastolic blood pressure do we find a significant negative $G \times E$ effect, suggesting that individuals with a high polygenic index for heart disease are less sensitive to the effects of adverse prenatal environments; consistent with e.g., ref. 55. We find no further evidence that the potential mechanisms mediate the relationship between polygenic indices and infant mortality rates in the development of heart disease.

Finally, *SI Appendix*, section F reports the results of different heterogeneity analyses and robustness checks. We show that our estimates are slightly larger for males compared to females (*SI Appendix*, Table F.1) and that they are robust to controlling for additional district-level covariates, alternatively defined polygenic indices, and nonlinearities in G and E (*SI Appendix*, Table F.2). We also find that the timing of the infant mortality rate measure is important. Consistent with the DOHaD hypothesis, we find that the estimate for infant mortality rates during pregnancy and (to a lesser extent) in the first year of life are almost double the size compared to earlier or later years (*SI Appendix*, Figs. F.1 and F.2).

Discussion

This paper shows a previously hidden interplay between genes and early-life environments in the development of ischemic heart disease. For someone with an average polygenic index for heart disease, the association between the infant mortality rate and ischemic heart disease is quantitatively small and not robust to the inclusion of family fixed effects (Table 2). This suggests that the infant mortality rate partially proxies for unobserved characteristics that vary between families, such as socioeconomic status. The influence of the polygenic index itself is quantitatively meaningful, with a sibling born into an area with an average infant mortality rate and one SD higher polygenic risk than his/her sibling experiencing 2 percentage points (or 15% relative to the mean) higher risk for developing ischemic heart disease. Interestingly however, we find that in districts with the lowest infant mortality rates, the relationship between the polygenic index and heart disease turns insignificantly different from zero. This suggests that advantageous environments may cushion the influence of the polygenic index. It is worth highlighting, however, that this does not necessarily imply that one's genetic predisposition for heart disease has no impact in these environments. For one, the genetic variants that capture differences in average predisposition for a trait may not affect its environmental sensitivity (56). Second, the GWAS weights that are used to construct the polygenic index depend on the environmental and demographic context of the discovery sample (57–59). In other words, the polygenic index does not solely capture genetic predisposition.

We do not claim to estimate the causal impact of infant mortality rates. Even though our fixed effects specifications allow us to rule out that infant mortality captures district- or family-specific time-invariant conditions, we cannot open the black box of which factors are the driving causal mechanism. In other words, it is likely that the infant mortality rate proxies for other early life (socioeconomic) environments. Although the polygenic index is measured with error and only reflects common genetic variants, our results are at least suggestive that the infant mortality rate reflects early-life environmental conditions, since we find that variation in the infant mortality rate within families is independent of genetic variation. This is therefore consistent with the hypothesis that genes and early-life environments interact in the development of heart disease.

One important issue to take into account in the interpretation of the results is selection and scarring. First, it is well known that the UK Biobank is not representative of the UK population (60), affecting the generalizability of our findings. Second, individuals

born in districts with high infant mortality rates may not be observed in the data, as they may not have survived. If this is the case, our sample is a selected (healthier) sample of individuals, meaning that our estimates are likely to be a lower bound. Third, individuals born in districts with high infant mortality rates may have been "scarred" in early life because of their exposure to an adverse environment. In fact, this is one of the potential mechanisms suggested in the literature, and what we may be capturing in our analysis.

Our findings have at least two implications. First, they reject the notion of genetic or environmental "determinism," suggesting a more nuanced understanding of later life health, in which the effect of one's genetic predisposition depends on one's environment (see, e.g., ref. 61). This suggests that improving the early life environment not only reduces the mean prevalence in heart disease but also the variation for those with different genetic predispositions. This is consistent with findings of Barcellos et al. (62), who show that an early educational intervention reduces the obesity-gap between the top and bottom of the BMI polygenic index. It is an interesting contrast however, with findings for education and cognition outcomes, where genetic effects are typically weaker in deprived environments (63–65). This may suggest that deprived environments strengthen genetic effects on health, but weaken genetic effects on education and cognition. Exploring whether this holds more generally would be an interesting area for future research.

Second, despite the fact that our study explores infant mortality rates between the 1930s and 1970s, our findings do have policy relevance. Indeed, they highlight an alternative measure of inequality in the population, genetic inequality, and show that there is scope to reduce this. We show the salience of genetic inequality and highlight its downstream health effects: a two SD difference in the polygenic index for heart disease corresponds to an approximately six percentage point difference in the probability of being diagnosed with heart disease; this is nonnegligible. We show that this is mainly driven by direct (causal) genetic effects. Using a very large sample, we show, however, that these effects are strongly malleable. A substantial body of research has already shown that early life interventions mitigate phenotypic inequalities, such as those in education, income, health, and crime (see, e.g., refs. 66–69). Our results go further than that; they highlight that early life conditions are key in mitigating arguably unfair genetic inequalities. Hence, by focusing on interventions that target the early-life environment, policy makers can reduce not only phenotypic but also genotypic inequalities in the population.

Materials and Methods

Using individuals' eastings and northings of birth in the UK Biobank, we identify the Local Government District (henceforth: district; $n = 1{,}472$) in which individuals were born and merge in data on individuals' local environmental conditions in the year before birth (which we loosely refer to as the year of pregnancy). To this end, we take the Great Britain Historical Database (GBHD; 70) as a starting point, which contains district-level birth, death, and infant mortality counts, as well as population estimates for the years 1930–1957 and 1963–1973. We collect and digitize this information for the remaining years 1958–1962 (71), and systematically quality-control the entirety of the database. We then link the infant mortality rate during the prenatal period in one's district of birth (defined as the total number of local deaths in the first year of life per 1,000 live births) for all UK Biobank participants born in England or Wales. We use the genetic data to create a polygenic index (PGI) for heart disease, using GWAS summary statistics (22) that exclude the UK Biobank. The baseline empirical specification can be written as

$$\begin{aligned} Y_{id} = \alpha &+ \beta_1 IMR_{id,t=-1} + \beta_2 IMR^2_{id,t=-1} + \beta_3 PGI_i + \beta_4 PGI_i^2 \\ &+ \beta_5 PGI_i \times IMR_{id,t=-1} + \gamma \mathbf{X_i} + f(IMR, PGI, \mathbf{X}) \\ &+ \delta_d + u_{id}, \end{aligned} \quad [1]$$

where Y_{id} is equal to one if individual i, born in district d, has been diagnosed with ischemic heart disease (IHD, identified using ICD-10 codes (I20–I25) obtained from mortality records and primary/secondary diagnoses in individuals' hospital inpatient records), and zero otherwise. 13% of our sample has been diagnosed with IHD. The district-level infant mortality rate in the year before birth (i.e., $t = -1$) and the relevant polygenic index are given by $IMR_{d,t=-1}$ and PGI_i respectively, both standardized to have mean 0 and SD 1. The gene-environment interaction is denoted by $PGI_i \times IMR_{d,t=-1}$. We follow ref. 47 and include nonlinear terms of PGI_i as well as $IMR_{d,t=-1}$, though we show the robustness of our results to linear specifications in *SI Appendix*, section F.4.

The vector $\mathbf{X_i}$ includes gender and dummies for each year-month of birth (i.e., [(12 mo × 38 y) − 1] separate dummy variables). The latter will account for the fact that older individuals are more likely to be exposed to higher infant mortality rates, and are also more likely to have worse health, on average. We also include the first 10 principal components of the genetic relatedness matrix to control for any remaining genetic differences across ancestry groups (see, e.g., ref. 72). The function $f(IMR, PGI, \mathbf{X})$ denotes interactions between $IMR_{d,t=-1}$ and $\mathbf{X_i}$ and between PGI_i and $\mathbf{X_i}$ (as in ref. 49), and δ_d are district fixed effects (we estimate the model with and without). The error is denoted by u_{id}; we report heteroskedasticity-robust SE, clustered either by district (in the full sample) or by family and district (in the family fixed effects analysis).

Whereas district fixed effects go a long way in controlling for socioeconomic differences across districts, the composition of districts may change over time. As such, an increase in the infant mortality rate may reflect an increase in regional poverty. We therefore also exploit the fact that the UK Biobank includes a sample of siblings, exploring variation within sibling pairs. The family fixed effects specification is, given by

$$\begin{aligned} Y_{ijd} = \alpha &+ \beta_1 IMR_{ijd,t=-1} + \beta_2 IMR^2_{ijd,t=-1} \\ &+ \beta_3 PGI_{ij} + \beta_4 PGI_{ij}^2 \\ &+ \beta_5 PGI_{ij} \times IMR_{ijd,t=-1} + \gamma \mathbf{X_{ij}} \\ &+ f(IMR, PGI, \mathbf{X}) + \eta_j + u_{ijd}, \end{aligned} \quad [2]$$

where η_j are the family fixed effects; the other variables are defined above. As such, Eq. **2** exploits the fact that some individuals are exposed to low infant mortality rates during the prenatal period, but their siblings, who largely share the same family environment, may be exposed to higher or lower rates. Hence, this specification exploits variation in infant mortality, holding any other time-invariant (observed or unobserved) family characteristics fixed and as such accounts for any confounders that differ between households that may bias the estimates from Eq. **1**. An additional advantage of including family fixed effects is that estimation of the genetic effects is purged from concerns relating to genetic nurture (see, e.g., refs. 73 and 74).

In sum, given the random inheritance of genetic variants within families, β_3 and β_4 in Eq. **2** capture the direct (or "causal") genetic effect. Furthermore, since our environmental measure is uncorrelated with our polygenic index in the within-family analysis (*SI Appendix*, section B), we are able to estimate a genuine "$G \times E$" interaction, as opposed to spurious "$G \times G$" or "$E \times E$" due to, e.g., gene-environment correlations or genetic nurture. To support our interpretation as genuine $G \times E$ interactions, *SI Appendix*, Fig. G.1), shows that the infant mortality rate is uncorrelated to a wide range of other polygenic indices, and similarly, that the polygenic index for heart disease is uncorrelated with a range of alternative early life environments (*SI Appendix*, Fig. G.2). Indeed, finding evidence of systematic *rGE* between the infant mortality rate and alternative polygenic indices would suggest that our $G \times E$ estimate may in fact be picking up a gene–gene interaction effect. Vice versa, strong correlations between the polygenic index for heart disease and alternative early life environments would suggest that our $G \times E$ estimate may instead be capturing an environment–environment interaction effect. Our analysis shows no strong evidence of *rGE* in either case, with a wide range of polygenic indices and early life environments, reinforcing the argument that we are identifying genuine $G \times E$ interactions.

Finally, whereas the sibling sample is helpful to reduce concerns of residual confounding, a downside is that it is significantly smaller. This has two (related) implications, both leading to a loss of power. First, using the smaller sample directly inflates the SE. Second, including family fixed effects means we are only exploiting variation within families. As most siblings are born relatively close together and infant mortality rates do not change dramatically over the course of a few years within a given district, there is relatively little variation in infant mortality rates over time within the same family. Hence, for both reasons, the within-family analysis has much less power than the between-family analysis, as reflected in the SE.

Data, Materials, and Software Availability. District-level infant mortality rates for England and Wales between 1935 and 1970 will be made available through the UK Biobank website, in line with its policy. The UK Biobank data are only accessible upon payment of a fee. Researchers can apply for data access directly at https://www.ukbiobank.ac.uk/.

ACKNOWLEDGMENTS. We would like to thank Alice Carter, Dilnoza Muslimova, seminar participants at the Vrije Universiteit Amsterdam, Erasmus Medical Centre, the Centre for Health Economics, the Alpine Population Conference, and European Social Science Genetics Network (ESSGN) for valuable comments. We gratefully acknowledge employees and participants of the 23andMe, Inc cohort for sharing GWAS summary statistics for educational attainment. This research has been conducted using the UK Biobank Resource under Application No. 74002. It uses data provided by patients and collected by the National Health Service (NHS) as part of their care and support. The work is also based on data provided through www.VisionofBritain.org.uk and uses historical material which is copyright of the Great Britain Historical Geographic Information System (GIS) Project and the University of Portsmouth. We gratefully acknowledge financial support from New Opportunities for Research Funding Agency Cooperation in Europe (NORFACE) Dynamics of Inequality Across the Life-course (DIAL) (462-16-100), the Medical Research Council (MR/N0137941/1), and the European Research Council (ERC 851725).

1. WHO, *Top 10 Causes of Death* (World Health Organization, 2018).
2. P. D. Gluckman, M. A. Hanson, C. Cooper, K. L. Thornburg, Effect of in utero and early-life conditions on adult health and disease. *N. Engl. J. Med.* **359**, 61–73 (2008).
3. L. G. Andersen *et al.*, Birth weight, childhood body mass index and risk of coronary heart disease in adults: Combined historical cohort studies. *PLoS ONE* **5**, 1–7 (2010).
4. Y. Arima, H. Fukuoka, Developmental origins of health and disease theory in cardiology. *J. Cardiol.* **76**, 14–17 (2020).
5. L. Donghao *et al.*, Birth weight, gestational age and risk of cardiovascular disease in early adulthood: Influence of familial factors. *Am. J. Epidemiol.* **192**, 866–877 (2023).
6. P. D. Gluckman, T. Buklijas, M. A. Hanson, "Chapter 1: The developmental origins of health and disease (DOHaD) concept" in *The Epigenome and Developmental Origins of Health and Disease*, C. S. Rosenfeld, Ed. (Elsevier, 2016), pp. 1–15.
7. S. H. Gage, M. R. Munafò, G. D. Smith, Causal inference in developmental origins of health and disease (DOHaD) research. *Annu. Rev. Psychol.* **67**, 567–585 (2016).
8. G. Conti, G. Mason, S. Poupakis, "Developmental origins of health inequality" in *Oxford Research Encyclopedia of Health Economics*, A. M. Jones, Ed. (Oxford University Press, 2019).
9. M. W. Gillman, Developmental origins of health and disease. *N. Engl. J. Med.* **353**, 1848–1850 (2005).
10. P. Gluckman, M. A. Hanson, C. Cooper, K. L. Thornburg, Effect of in utero and early-life conditions on adult health and disease. *N. Engl. J. Med.* **359**, 61–73 (2008).
11. D. Almond, J. Currie, *Human Capital Development Before Age 5* (Elsevier, 2011), pp. 1315–1486.
12. D. Almond, J. Currie, Killing me softly: The fetal origins hypothesis. *J. Econ. Perspect.* **25**, 153–172 (2011).
13. D. Almond, J. Currie, V. Duque, Childhood circumstances and adult outcomes: Act II. *J. Econ. Lit.* **56**, 1360–1446 (2018).
14. J. J. Heindel, L. A. Skalla, B. R. Joubert, C. H. Dilworth, K. A. Gray, Review of developmental origins of health and disease publications in environmental epidemiology. *Reprod. Toxicol.* **68**, 34–48 (2017).
15. D. Barker, C. Osmond, Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet* **327**, 1077–1081 (1986).
16. Z. D. Drobni *et al.*, Heritability of coronary artery disease: Insights from a classical twin study. *Circ. Cardiovasc. Imaging* **15**, e013348 (2022).
17. N. J. Samani *et al.*, Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
18. A. Helgadottir *et al.*, A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
19. R. McPherson *et al.*, A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
20. H. Schunkert *et al.*, Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333 (2011).
21. P. Deloukas *et al.*, Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25 (2013).
22. M. Nikpay *et al.*, A comprehensive 1000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121 (2015).
23. J. M. M. Howson *et al.*, Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat. Genet.* **49**, 1113 (2017).
24. C. P. Nelson *et al.*, Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385 (2017).
25. P. van der Harst, N. Verweij, Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443 (2018).
26. A. V. Khera, S. Kathiresan, Genetics of coronary artery disease: Discovery, biology and clinical translation. *Nat. Rev. Genet.* **18**, 331 (2017).
27. T. M. Powell-Wiley *et al.*, Social determinants of cardiovascular disease. *Circ. Res.* **130**, 782–799 (2022).
28. S. Visvikis-Siest, G. Siest, The STANISLAS cohort: A 10-year follow-up of supposed healthy families. Gene–environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. *Clin. Chem. Lab. Med.* **46**, 733–747 (2008).
29. A. Hirvonen, Gene–environment interactions in chronic pulmonary diseases. *Mutat. Res.* **667**, 132–141 (2009).
30. E. Flowers, E. Froelicher, B. Aouizerat, Gene–environment interactions in cardiovascular disease. *Eur. J. Cardiovasc. Nurs.* **11**, 472–478 (2012).
31. P. Gluckman, M. Hanson, Adult disease: Echoes of the past. *Eur. J. Endocrinol.* **155**, S47–S50 (2006).
32. M. Rutter, *Genes and Behavior: Nature–Nurture Interplay Explained* (Blackwell Publishing, 2006).
33. P. Gluckman, M. Hanson, *The Developmental Origins of Health and Disease: The Breadth and Importance of the Concept, Chapter 1* (Eurekah.com/Springer Science+Business Media, 2006), pp. 1–7.
34. B. T. Heijmans *et al.*, Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17046–17049 (2008).
35. R. Feil, M. F. Fraga, Epigenetics and the environment: Emerging patterns and implications. *Nat. Rev. Genet.* **13**, 97–109 (2012).
36. A. P. Feinberg, M. D. Fallin, Epigenetics at the crossroads of genes and the environment. *JAMA* **314**, 1129–1130 (2015).
37. S. Costantino *et al.*, Epigenetics and precision medicine in cardiovascular patients: From basic concepts to the clinical arena. *Eur. Heart J.* **39**, 4150–4158 (2018).
38. J. Yi, J. J. Heckman, J. Zhang, G. Conti, Early health shocks, intra-household resource allocation and child outcomes. *Econ. J.* **125**, F347–F371 (2015).
39. A. R. Adhvaryu, A. Nyshadham, Endowments at birth and parents' investments in children. *Econ. J.* **126**, 781–820 (2016).
40. M. Grätz, F. Torche, Compensation or reinforcement? The stratification of parental responses to children's early ability. *Demography* **53**, 1883–1904 (2016).
41. T. Molina, Pollution, ability, and gender-specific investment responses to shocks. *J. Eur. Econ. Assoc.* **19**, 580–619 (2021).
42. H. Majid, Child endowments and parental investments: Intra-household allocation in oportunidades families in Mexico. *Rev. Dev. Econ.* **22**, 91–114 (2018).
43. D. Muslimova, H. van Kippersluis, C. Rietveld, S. von Hinke, F. Meddens, "Dynamic complementarity in skill production: Evidence from genetic endowments and birth order" in *Mimeo, Erasmus University Rotterdam* (2020), vol. 1, p. 1.
44. J. Fletcher, Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. *PLoS ONE* **7**, e50576 (2012).
45. E. A. W. Slob, C. A. Rietveld, Genetic predispositions moderate the effectiveness of tobacco excise taxes. *PLoS ONE* **16**, e0259210 (2021).
46. C. Sudlow *et al.*, UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
47. P. Biroli, *et al.*, The economics and econometrics of gene-environment interplay. arXiv [Preprint] (2022). https://arxiv.org/abs/2203.00729 (Accessed 23 May 2024).
48. B. W. Domingue, K. Kanopka, S. Trejo, M. Rhemtulla, E. M. Tucker-Drob, Ubiquitous bias and false discovery due to model misspecification in analysis of statistical interactions: The role of the outcome's distribution and metric properties. *Psychol. Methods*, 1–33 (2022), 10.1037/met0000532.
49. M. C. Keller, Gene x Environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biol. Psychiatr.* **75**, 18–24 (2014).
50. J. Myles Shaver, Interpreting interactions in linear fixed-effect regression models. *Strategy Sci.* **4**, 25–40 (2019).
51. G. Chamberlain, Analysis of covariance with qualitative data. *Rev. Econ. Stud.* **47**, 225–238 (1980).
52. C. Hsiao, "Logit and probit models" in *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, L. Mátyás, P. Sevestre, Eds. (Springer, 1996), pp. 410–428.
53. W. Greene, The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Economet. J.* **7**, 98–119 (2004).
54. A. I. Young *et al.*, Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nat. Genet.* **54**, 897–905 (2022).
55. G. J. Berg, S. von Hinke, R. A. H. Wang, Prenatal sugar consumption and late-life human capital and health: Analyses based on postwar rationing and polygenic scores. arXiv [Preprint] (2023). https://arxiv.org/abs/2301.09982 (Accessed 23 May 2024).
56. R. Johnson, R. Sotoudeh, D. Conley, Polygenic scores for plasticity: A new tool for studying gene-environment interplay. *Demography* **59**, 1045–1070 (2022).
57. H. Mostafavi *et al.*, Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* **9**, e48376 (2020).
58. N. Pirastu *et al.*, Genetic analyses identify widespread sex-differential participation bias. *Nat. Genet.* **53**, 663–671 (2021).
59. M. C. Mills, N. Barban, F. C. Tropf, *An Introduction to Statistical Genetic Data Analysis* (MIT Press, 2020).

60. A. Fry *et al*., Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
61. A. Caspi *et al*., Genetic sensitivity to the environment: The case of the serotonin transporter gene and its implications for studying complex diseases and traits. *Am. J. Psychiatry* **167**, 509–527 (2010).
62. S. H. Barcellos, L. S. Carvalho, P. Turleym, Education can reduce health differences related to genetic risk of obesity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E9765–E9772 (2018).
63. A. C. Heath *et al*., Education policy and the heritability of educational attainment. *Nature* **314**, 734–736 (1985).
64. E. M. Tucker-Drob, T. C. Bates, Large differences cross-national in gene × socioeconomic status interaction on intelligence. *Psychol. Sci.* **27**, 138–149 (2016).
65. K. P. Harden, *The Genetic Lottery: Why DNA Matters for Social Equality* (Princeton University Press, 2021).
66. J. Zhou, J. Heckman, B. Liu, M. Mai, "Treatment Effects and the Measurement of Skills in an Influential Home Visiting Program" (Tech. Rep. 1, 2020).
67. G. Conti, S. Cattan, C. Farquharson, R. Ginja, M. Pecher, "The Health Effects of Universal Early Childhood Interventions: Evidence from Sure Start" (CEPR Discussion Papers 16730, 2021).
68. J. L. García, F. H. Bennhoff, D. E. Leaf, J. J. Heckman, "The dynastic benefits of early childhood education" (Tech. Rep., National Bureau of Economic Research, 2021).
69. J. L. García, J. J. Heckman, V. Ronda, The lasting effects of early-childhood education on promoting the skills and social mobility of disadvantaged African Americans and their children. *J. Polit. Econ.* **131**, 1477–1506 (2023).
70. H. R. Southall, P. Ell, E. M. Garrett, G. Mooney, C. Galley, Great Britain Historical Database : Health and Health Care Data : Mortality Statistics, 1851-1973. UK Data Service. http://doi.org/10.5255/UKDA-SN-4570-1. Accessed 5 May 2020.
71. HMSO, *Registrar General Statistical Review of England and Wales* (The Registrar General Statistical Review of England and Wales, London, UK, 1934).
72. A. L. Price *et al*., Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
73. A. Kong *et al*., The nature of nurture: Effects of parental genotypes. *Science* **359**, 424–428 (2018).
74. J. Wertz *et al*., Genetics of nurture: A test of the hypothesis that parents' genetics predict their observed caregiving. *Dev. Psychol.* **55**, 1461–1472 (2019).