




# The production of within-family inequality: Insights and implications of integrating genetic data

Jason M. Fletcher <sup>a,\*</sup>, Yuchang Wu<sup>b</sup>, Zijie Zhao <sup>b</sup> and Qiongshi Lu <sup>c</sup>

<sup>a</sup>La Follette School of Public Affairs, Center for Demography of Health and Aging, Department of Population Health Sciences, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706, USA

<sup>b</sup>Department of Biostatistics and Medical Informatics, Genetics-Biotech Center, 425 Henry Mall, Madison, WI 53706, USA

<sup>c</sup>Department of Biostatistics and Medical Informatics, Center for Demography of Health and Aging, Genetics-Biotech Center, University of Wisconsin-Madison, 425 Henry Mall, Madison, WI 53706, USA

\*To whom correspondence should be addressed: Email: [jason.fletcher@wisc.edu](mailto:jason.fletcher@wisc.edu)

Edited By: Ryan Webb

## Abstract

The integration of genetic data within large-scale social and health surveys provides new opportunities to test long-standing theories of parental investments in children and within-family inequality. Genetic predictors, called polygenic scores, allow novel assessments of young children's abilities that are uncontaminated by parental investments, and family-based samples allow indirect tests of whether children's abilities are reinforced or compensated. We use over 16,000 sibling pairs from the UK Biobank to test whether the relative ranking of siblings' polygenic scores for educational attainment is consequential for actual attainments. We find evidence consistent with compensatory processes, on average, where the association between genotype and phenotype of educational attainment is reduced by over 20% for the higher-ranked sibling compared to the lower-ranked sibling. These effects are most pronounced in high socioeconomic status areas. We find no evidence that similar processes hold in the case of height or for relatives who are not full biological siblings (e.g. cousins). Our results provide a new use of polygenic scores to understand processes that generate within-family inequalities and also suggest important caveats to causal interpretations the effects of polygenic scores using sibling difference designs. Future work should seek to replicate these findings in other data and contexts.

**Keywords:** polygenic scores, inequality, family environments, parental preferences

## Significance Statement

Parents choose to reinforce or compensate for differences in their children's abilities, which has implications for child outcomes as well as within-family inequalities. We show that the association between a child's genetics and their educational attainment is weaker if the child has a higher genetic propensity for educational attainment than their sibling, suggesting compensatory/equalizing parental investments. Our work also suggests caution in interpreting the effects of polygenic scores, even in the case of within-family designs, as causal.

## Introduction

A large social science literature has produced theoretical and empirical support that parental actions shape patterns of within-family inequalities over the life cycle (1–3). Theoretically, the key components of the question include parental attitudes about inequalities of outcomes of their children (4, 5) and the potential of differential returns to investment based on children's talents (i.e. technologies of skill formation) (6–8). Separating these mechanisms has proven difficult. Empirically, measurement limitations have been important bottlenecks in progress. Ideally, researchers could use measures that (a) occur early in life so they represent endowments and (b) capture endowments of children that are not related to parental behaviors and investments

(no feedback effects). Typically, birth weight has been used in analyses that examine whether parents reinforce or compensate for children's endowments (3, 9–11). However, focusing on birth weight is imperfect because it can be affected by parental behaviors and investments (12–16), and it limits the scope of analysis due to a focus on a single measurement. This scope limitation occurs both in terms of life course outcomes that can be tied to birth weight and failing to examine parental responses to endowments that are not associated with birth weight. Alternative measures, such as test scores (4) can be problematic because parents can shape these outcomes prior to research measurement and they may not be “early enough” to capture endowments—for example, before children acquire language skills.

**Competing Interest:** The authors declare no competing interest.

**Received:** December 20, 2022. **Revised:** March 14, 2023. **Accepted:** April 6, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

**Table 1.** Associations between PGS-education and educational attainment (comparing between family and within-family results).

Outcome	Education Full	Education Sibling	Education Sibling
Sample	None	None	Sibling
Fixed effects?			
PGS (std)	0.908 <sup>a</sup> (0.008)	0.959 <sup>a</sup> (0.026)	0.444 <sup>a</sup> (0.048)
Female	-0.746 <sup>a</sup> (0.016)	-0.777 <sup>a</sup> (0.055)	-0.860 <sup>a</sup> (0.070)
Age fixed effects	X	X	X
PC controls (1–20)	X	X	X
Observations	404,409	34,209	33,044
R <sup>2</sup>	0.068	0.063	0.669

Notes: Robust standard errors are in parentheses.

<sup>a</sup>P < 0.01.

Summary genetic assessments (i.e. polygenic scores, PGS) have the capacity to overcome these empirical limitations, as these measures are fixed at conception—thus, they have no feedback effects, and they can capture endowments tied to early outcomes. PGS also provide the possibility of expanding the domains of analysis outside of birth weight. Studies have begun to use these measures to show associations with early childhood outcomes (17), and some research has found evidence that PGS are associated with parental responses (18, 19). However, fewer studies have incorporated these measures into analyses of within-family inequalities. Might parents use observable phenotypic downstream outcomes tied to PGS in their efforts to increase or decrease differences in later outcomes of their children? We begin this direction of empirical analysis by linking two highly predictive PGS measures (education and height) to phenotypic outcomes in adulthood in a sample of over 16,000 sibling pairs from the UK Biobank (UKB).

We propose that theoretical models of parental responses to children's endowments can be assessed with PGS measures of siblings. Because our data do not include measures of early life, we use an indirect test of the accumulated parental responses to their children's abilities by comparing educational attainments of the siblings with their PGS measurements. In order to provide an omnibus test of compensation vs. reinforcement, we test whether the *relative ranking* of PGS within sibling pairs is consequential for predicting adult outcomes. We focus on the case of education as an exemplar where parents may have the means and desire to shape inequalities in their children's attainments (20) and the case of height, where they have neither, so that we have a negative test. We hypothesize that if parents prefer, on average, for equalizing their children's outcomes (i.e. compensatory behaviors), the PGS will be less predictive of attainment in the sibling with a higher relative rank. If parents prefer, on average, to reinforce their children's relative advantages, the PGS will be more predictive of attainment in the sibling with a higher relative rank. We then explore whether these patterns differ by the socioeconomic status of the families. A large literature has shown that parents in advantaged settings often show compensatory preferences and behaviors while parents in disadvantaged settings show reinforcing preferences and behaviors (11, 21, 22). Therefore, we divide the families in our data based on area-level ("neighborhood") socioeconomic status and estimate the relative ranking effects outlined above. We note here and below that our data cannot rule out some alternative hypotheses of differential sibling effects based on relative ranking, such as direct sibling effects and other family processes; thus we cautiously

**Table 2.** Within-family associations between PGS-education and educational attainment (including sibling dyad measures of relative position).

Outcome	Education Sibling	Education High SES place	Education Low SES place
Sample	Sibling	Sibling	Sibling
Fixed effects?			
PGS (std)	0.389 <sup>a</sup> (0.082)	0.433 <sup>a</sup> (0.113)	0.353 <sup>a</sup> (0.119)
Female	-0.857 <sup>a</sup> (0.070)	-0.626 <sup>a</sup> (0.096)	-1.083 <sup>a</sup> (0.103)
Older sib of pair	0.443 <sup>a</sup> (0.090)	0.323 <sup>a</sup> (0.122)	0.576 <sup>a</sup> (0.133)
Sib with larger PGS	0.124 (0.076)	0.183 <sup>b</sup> (0.106)	0.078 (0.108)
Larger × PGS	-0.086 <sup>b</sup> (0.050)	-0.179 <sup>c</sup> (0.070)	-0.006 (0.073)
Age fixed effects	X	X	X
PC controls (1–20)	X	X	X
Observations	33,044	16,810	16,234
R <sup>2</sup>	0.669	0.669	0.641

Notes: Robust standard errors are in parentheses (clustered by family). High SES place: families who grew up in places/years with above median predicted educational attainments. Low SES place: families who grew up in places/years with below median predicted educational attainments.

<sup>a</sup>P < 0.01, <sup>b</sup>P < 0.05, <sup>c</sup>P < 0.1.

interpret our findings in terms of their consistency with compensatory or reinforcing behaviors.

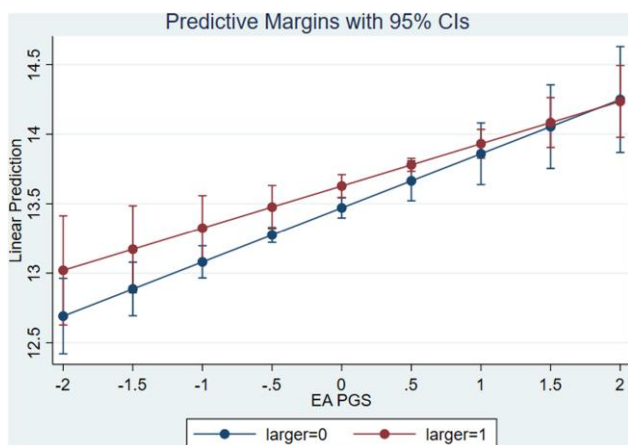
## Results

We analyzed educational attainment in the full sample and a sibling subsample from respondents with European ancestry in the UKB. Siblings were matched based on kinship estimates of genetic relatedness (see Materials and methods). Table S1 provides descriptive statistics of the full sample and our sibling analysis sample. Our baseline results conformed with other analyses (23–25) showing that the education attainment (EA)-PGS predicts educational attainment in the UKB sample (Table 1, column 1), that these associations are retained in the sibling subsample (Table 1, column 2), and that the association is reduced by >50% when family fixed effects are included (Table 1, column 3). Specifically, the association between a standard deviation increase in EA-PGS and EA falls from 0.91 years in column 1 to 0.44 years in column 3.

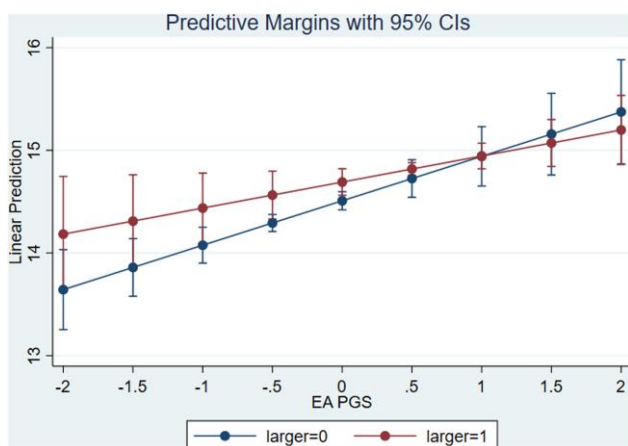
Sibling difference analysis controls for influences shared by siblings growing up in the same household, including the portion of the PGS associated with genetic nurture (26, 27) that is shared among siblings. However, unshared environmental factors, including parental actions to reinforce or compensate for sibling differences in PGS, are retained in the PGS associations. Table S2 presents average differences and correlations in sibling phenotype and genotype measures. As expected, sibling correlations in the genotype measures are  $r \sim 0.5$ . Correlations in EA and height are  $\sim 0.3$ . On average, siblings differ in PGS by  $\sim 0.75$  standard deviation (SD), 4 years of schooling, and 8 cm in height.

In order to conduct an omnibus test of reinforcing vs. compensating processes in reaction to sibling differences in PGS, we add dyadic measures of relative ranking of PGS and age (see Materials and methods) as well as an interaction between the indicator for higher relative ranking and PGS to predict EA.

Table 2 shows that inclusion of the dyadic measures is consequential. Conforming to the birth order literature, the older sibling [even accounting for age indicators (i.e. fixed effects)] attains over



**Fig. 1.** Plot of associations between EA-PGS and educational attainment. Stratified by whether sibling has larger or smaller PGS than co-sibling. Notes: coefficients plotted from Table 2, column 2. Larger = 1 refers to siblings who have larger values of PGS than their co-sibling. Larger = 0 refers to siblings who have smaller values of PGS than their co-sibling. The outcome is educational attainment. Same controls as Table 2. Command: Marginsplot from Stata.

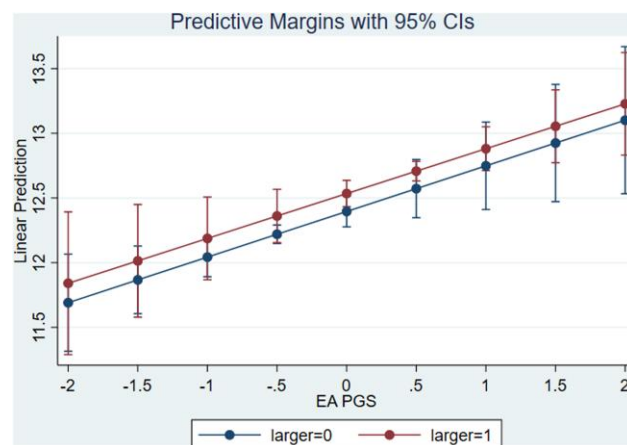


**Fig. 2.** Plot of associations between EA-PGS and educational attainment. Stratified by whether sibling has larger or smaller PGS than co-sibling. High SES places. Notes: coefficients plotted from Table 2, column 3. Larger = 1 refers to siblings who have larger values of PGS than their co-sibling. Larger = 0 refers to siblings who have smaller values of PGS than their co-sibling. The outcome is educational attainment. Same controls as Table 2. Command: Marginsplot from Stata.

0.5 years of schooling more than the younger sibling of the pair (28–30).<sup>1</sup> The interaction between the indicator for higher-ranked EA-PGS and the EA-PGS is negative ( $P$ -value  $< 0.09$ ), so that the EA-PGS has a smaller association with attainment for siblings who are higher-ranked within the sibship (see Fig. 1). This result suggests, on average, compensatory processes within families and a concomitant reduction in within-family inequalities in educational attainment. As we note further below, it is important to point out that we cannot rule out direct sibling and other family effects as a source of these associations.

Table 2 columns 3 and 4 stratify the sibling analysis based on area (place of birth)–estimates of socioeconomic status. Results show that birth order effects are smaller in high SES areas, the

<sup>1</sup> We would have preferred to control for birth order, but the UKB only asked birth order information for a subset of the sample.



**Fig. 3.** Plot of associations between EA-PGS and educational attainment. Stratified by whether sibling has larger or smaller PGS than co-sibling. Low SES places. Notes: coefficients plotted from Table 2, column 4. Larger = 1 refers to siblings who have larger values of PGS than their co-sibling. Larger = 0 refers to siblings who have smaller values of PGS than their co-sibling. The outcome is educational attainment. Same controls as Table 2. Command: Marginsplot from Stata.

gender gap is smaller in high SES areas, and the differences are much larger in high SES areas ( $P < 0.05$ ) (see Figs. 2 and 3). These results also show, in general, the importance of families for shaping links between genetic measures and social outcomes.

Table 3 performs a negative test. We hypothesized to find no effect of families and relative rank on genetic association for the case of height. The results suggest no effect and can rule out even modest size effects.

Table 4 performs a second negative test. We form dyads in the data who are related (e.g. cousins) but are not full siblings (kinship  $< 0.18$ ). We show that earlier results of impacts of relative rankings of PGS are absent in these related, nonsibling pairs. In order to explore the possibility of “ceiling effects” of EA that may mechanically induce an interaction between EA-PGS and other predictive measures of EA (sibling rank of EA-PGS), we include two additional tests of interaction between individual level predictors of EA and the EA-PGS between full siblings. Figures S1A (female sex) and S2A (indicator of older sibling in pair) show that, while each of these the main effects of these stratifying variables predicts EA, we see no evidence of interactions with EA-PGS in predicting EA. This further suggests that there is not a mechanical explanation for the presence of the interaction between EA-PGS relative ranking and EA-PGS in our main results.

## Discussion

These results contribute to several literatures in the social and genetic sciences. In the social sciences, the use of PGS to measure relative traits between siblings allows expanded analyses estimating the extent to which parents compensate or reinforce their children’s endowments. Future work could expand this direction by estimating a larger set of domains of PGS as well as focusing on early life measurements of children’s outcomes. A major limitation for this direction of inquiry is the modest sample of siblings in many data sets; future research should test PGS  $\times$  relative rank/status in other data sets to assess its wider importance in other contexts. A second limitation is the documented “healthy respondent” composition of the UKB sample, which limits the generalizability of the results (31).

**Table 3.** Falsification exercise (within-family associations between PGS-height and height).

Outcome Sample Fixed effects?	Height Full None	Height Sibling None	Height Sibling Sibling	Height Sibling Sibling
Height PGS (std)	2.516 <sup>a</sup> (0.010)	2.557 <sup>a</sup> (0.034)	2.151 <sup>a</sup> (0.050)	2.154 <sup>a</sup> (0.085)
Female	-13.319 <sup>a</sup> (0.019)	-13.295 <sup>a</sup> (0.065)	-13.468 <sup>a</sup> (0.070)	-13.467 <sup>a</sup> (0.070)
Older sib of pair				0.019 (0.089)
Sib with larger PGS				-0.034 (0.075)
Larger × PGS				0.052 (0.051)
Age fixed effects	X	X	X	X
PC controls (1–20)	X	X	X	X
Observations	402,320	34,036	32,700	32,700
R <sup>2</sup>	0.600	0.598	0.901	0.901

Robust standard errors are in parentheses (clustered by family).

<sup>a</sup>P < 0.01.

Our results show lower levels of genetic association for educational attainment for children who have higher PGS than their sibling. This phenomenon is stronger in places whose residents have higher socioeconomic status than in places with low socioeconomic status. Together, these results are consistent with parental preferences for equality among their children, which may be accomplished through compensating investments. We do acknowledge, though, that we are not able to rule out effects from siblings in our results; further distinguishing parental from sibling influence remains an important next step in this research area.

The findings are also consequential for genetic analyses more broadly. Some recent innovative analysis linking genetics to outcomes has taken seriously the confounding between children's genetics and family background when explaining children's later outcomes as adults, seeking to decompose PGS into direct (child) and indirect (parent) effects (26, 32, 33). This work has been motivated, in part, by previous research that used sibling comparisons in PGS to predict education. For example, some early use of educational attainment PGS (34) used sibling pairs to show that the PGS continued to predict education. The implication was then drawn that the PGS contained some causal effects of genetics on important outcomes such as schooling.

Later work has prioritized comparisons of between family estimates with within-family estimates to assure a lack of confounding by family background—using EA2 (35) and EA3 (23) has been a bit more mixed, with the within-family analysis often reducing the estimates of genetic association by 50% compared to between-family analysis. However, even with these smaller effects, researchers have relied on this framework to show the likelihood of causal PGS effects on education and related outcomes (25, 36, 37). New work (37) has discussed common issues in genetic analysis that can be overcome with family-based designs, including population stratification, assortative mating, and dynastic effects. One commonality of these issues is that they are presumed to reflect shared effects on all children in a household that can be eliminated through sibling comparisons or, alternatively, the use of parental genetic data. Indeed, researchers have stated that sibling analysis “rules out purely social transmission as an explanation for the associations between children's education-linked genetics and their attainment” (25).

**Table 4.** Falsification exercise (between-relative associations of PGS-education, educational attainment, and relative rankings of PGS).

Outcome Sample Fixed effects?	Education Nonsibling None	Education Nonsibling Family	Education Nonsibling Family
PGS (std)	0.912 <sup>a</sup> (0.018)	0.513 <sup>a</sup> (0.048)	0.493 <sup>a</sup> (0.053)
Female	-0.638 <sup>a</sup> (0.037)	-0.709 <sup>a</sup> (0.052)	-0.709 <sup>a</sup> (0.052)
Older relative of pair		0.119 <sup>b</sup> (0.063)	0.120 <sup>b</sup> (0.063)
Relative with larger PGS		0.089 (0.060)	0.090 (0.061)
Larger × PGS			0.040 (0.046)
Age fixed effects	X	X	X
PC controls (1–20)	X	X	X
Observations	62,360	62,360	62,360
R <sup>2</sup>	0.070	0.606	0.606

Robust standard errors in parentheses (clustered by “family”).

<sup>a</sup>P < 0.01, <sup>b</sup>P < 0.1.

However, while these may be the most common concerns for analysis linking genetics with outcomes, a rich set of theoretical models and empirical results from demography and adjacent social sciences suggest there are yet another set of complications that should be explored before claiming that genetic effects are causal. In particular, families shape outcomes, parents react to children's abilities and have preferences for their set of children's outcomes. Other family members (who also often share genetics with the child/children) also contribute. An important outcome of these processes is the possibility of statistical interference between the outcomes of siblings (38), so that siblings' designs both eliminate some empirical concerns but raise others.

The current results demonstrate that whatever causal effects exist between PGS and outcomes can be mediated through families, even with a sibling fixed effects design. This suggests caution in interpreting sibling fixed effects models as “causal genetic effects” and reemphasizes that the value of this strategy is to control for shared environments but not produce causal estimates. Indeed, our results suggest that within-sibling analysis of genetics and outcomes continues to reflect family processes rather than pure “genetic effects.” Future work will need to incorporate additional strategies to separate “genetic” and “family” causal processes as well as extend the analysis of “relative PGS” effects to other data sets in order to explore its reach in other contexts.

## Materials and methods

### Data

We used data from the UKB project (39). The participants, aged between 37 and 74 years, were originally recruited between 2006 and 2010. These data are restricted, but one can gain access by following the procedures described in [www.ukbiobank.ac.uk/register-apply/](http://www.ukbiobank.ac.uk/register-apply/).

Although siblings are not identified in the survey, respondents' genetics can be used to measure genetic relatedness among all pairs of respondents. We first use the UKB provided kinship file, listing all pairwise kinships among 100,000 pairs in the sample of nearly 500,000 individuals. We first choose pairs with kinship > 0.2, which reflects first-degree biological relatives (parents/siblings). We then choose remaining pairs who are < 13 years apart



in age, leaving ~22,000 sibling dyads. We then chose to keep only one dyad from any family with more one dyad, leaving ~17,600 dyads. We include only respondents of European ancestry in our analysis.

## Polygenic scores

We constructed PGS for two traits for which large genome-wide association studies (GWAS) are publicly available and do not contain UKB samples: height (40) and educational attainment (23). We removed single-nucleotide polymorphisms (SNPs) in strong linkage disequilibrium (LD). We LD-clumped the GWAS summary data by PLINK (41), using 1000 Genome Project Phase III European genotype data as reference. We used a LD window size of 1Mb and a pairwise  $r^2$  threshold of 0.1. We did not apply any  $P$ -value thresholding to select SNPs. Final weights were produced by using PRSice-2 (42). The PGSs were normalized to have mean zero and SD one and oriented so that each PGS was positively correlated with its corresponding outcome.

## Phenotypes

Educational levels of the UKB participants were measured by mapping each major educational qualification that can be identified from the survey measures to an International Standard Classification of Education (ISCED) category and imputing a years-of-education equivalent for each ISCED category (43). Height is measured standing height.

## Sibling variables

We created three variables indicating the relative status of the members of sibling pairs. First, we created an indicator for the sibling with the higher EA-PGS. Second, we created an indicator for the sibling with the higher height-PGS. Third, we created an indicator for the sibling who is older, due to well-known birth order effects on educational attainment (28–30). We note that self-reported birth order is only available for a subset of UKB respondents.

## Place-based socioeconomic status

The UKB does not contain information about childhood background socioeconomic status of the respondents. Therefore we created information based on place of birth (“neighborhoods”) and year of birth to predict which places/years of birth had high vs. low (based on median) predicted educational attainment. We split families based on whether neither sibling was born in a place/year with high predicted attainment vs. either sibling was born in a place/year with high predicted attainment. The median predicted schooling was 13.65 years in a regression that had place of birth fixed effects and year of birth fixed effects.

## Sample characteristics

Table S1 presents descriptive statistics for the full sample and the analysis subsample of siblings. Educational attainment is slightly higher in the full sample, and number of siblings (by definition) is higher in the sibling sample. EA3 is slightly higher on average in the sibling sample but demographic characteristics are quite similar.

## Sample characteristics for falsification exercise

We created a second sample of related individuals who were not full biological siblings in order to develop a negative (falsification) test of the analysis. To construct the sample, we used dyads with

kinship measurements below 0.18 and randomly chose one dyad in cases where a respondent was linked with multiple sample members. Sample characteristics are shown in Table S3A and dyadic characteristics are shown in Table S4A.

## Analysis

We tested associations using linear regression models. We clustered standard errors at the family level. We conducted sibling difference analysis using family fixed effects regression (44). We note that fixed effects and first-difference models are equivalent in our case, since we have two observations per unit (family) (45). We control for sex, age indicators, and 20 genetic principal components.

We controlled for age indicators (fixed effects) for the well-documented secular increases in schooling over this time period (46).

We use regression analysis to link PGS to educational attainments in our results. We compare results with and without sibling fixed effects and also with measures reflecting siblings’ relative position in the dyad based on EA3 and age. That is, we regress educational attainment for respondent  $i$  in family  $f$  on demographic characteristics (age and sex), a PGS and controls for 20 principal components, and a family-clustered error term:

$$\text{education}_{if} = \beta_0 + \beta_1 X_{if} + \beta_2 \text{PGS}_{if} + \beta_{\text{PC}} \text{PC}_{if} + \varepsilon_{if}$$

We next add sibling fixed effects to the model. As we note above, this is equivalent to taking differences of each measure between the two siblings in the dyad:

$$\text{education}_{if} = \beta_0 + \beta_1 X_{if} + \beta_2 \text{PGS}_{if} + \beta_F F_f + \beta_{\text{PC}} \text{PC}_{if} + \varepsilon_{if}$$

Finally, we enrich the model with dyadic measures of relative position, namely, an indicator for whether the respondent has a higher PGS than his/her sibling and the interaction between this indicator and PGS:

$$\text{education}_{if} = \beta_0 + \beta_1 X_{if} + \beta_2 \text{PGS}_{if} + \beta_3 \text{LargerPGS}_{if} + \beta_4 \text{LargerPGS}_{if} * \text{PGS}_{if} + \beta_F F_f + \beta_{\text{PC}} \text{PC}_{if} + \varepsilon_{if}$$

where the key coefficient is  $\beta_4$ , which reflects differences in genetic association based on relative ranking in the sibship on the PGS. A positive coefficient would suggest that parents reinforce early evidence of a child’s ability (relative to his/her sibling), while a negative coefficient suggests that parents compensate. It is possible that these effects vary by family socioeconomic status.<sup>2</sup> Therefore, we stratify the model between families who were born in high socioeconomic status areas/years vs. low socioeconomic status areas/years.

## Acknowledgments

The authors gratefully acknowledge use of the facilities of the Center for Demography of Health and Aging at the University of Wisconsin-Madison. We thank members of the Social Genomics Working Group at University of Wisconsin and Benjamin Domingue for helpful comments. This research has been conducted using the UK Biobank Resource under Application 57284. This manuscript was posted on as a preprint: <https://doi.org/10.1101/2020.06.06.137778>

<sup>2</sup> Unfortunately, the UKB is extremely limited in measurements of childhood experiences and has nearly no information about parents’ characteristics of the respondents (e.g. parental education, family socioeconomic status during childhood, etc.).

## Supplementary Material

Supplementary material is available at PNAS Nexus online.

## Funding

Funded by NIA Center Grant P30 AG017266.

## Author contributions

J.M.F. and Q.L. designed the study, analyzed, and interpreted the data and wrote the first draft of the paper. J.M.F., Y.W., Z.Z., and Q.L. assisted with the analysis and interpretation of the data and the writing of the paper. J.M.F. and Q.L. have had full access to all the data in the study and had complete freedom to direct its analysis and its reporting, without influence, editorial decision, or censorship from the sponsors. The authors declare no competing interest and have all approved manuscript submission.

## Data availability

Partial restrictions to the data and/or materials apply: data available from UK Biobank (email: [access@ukbiobank.ac.uk](mailto:access@ukbiobank.ac.uk)), statistical code available from authors and sample code included in supplementary file.

## References

- Becker GS, Tomes N. 1986. Human capital and the rise and fall of families. *J Labor Econ.* 4(3 Part 2):S1–S39.
- Becker GS, Lewis HG. 1973. On the interaction between the quantity and quality of children. *J Political Econ.* 81(2 Part 2):S279–SS88.
- Behrman JR, Rosenzweig MR, Taubman P. 1994. Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *J Political Econ.* 102(6):1131–1174.
- Grätz M, Torche F. 2016. Compensation or reinforcement? The stratification of parental responses to children's early ability. *Demography* 53(6):1883–1904.
- Behrman JR, Pollak RA, Taubman P. 1982. Parental preferences and provision for progeny. *J Political Econ.* 90(1):52–73.
- Cunha F, Heckman J. 2007. The technology of skill formation. *Am Econ Rev.* 97(2):31–47.
- Cunha F, Heckman JJ. 2008. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *J Hum Resour.* 43(4):738–782.
- Aizer A, Cunha F. The production of human capital: Endowments, investments and fertility. National Bureau of Economic Research, 2012 0898-2937.
- Datar A, Kilburn MR, Loughran DS. 2010. Endowments and parental investments in infancy and early childhood. *Demography* 47(1):145–162.
- Restrepo BJ. 2016. Parental investment responses to a low birth weight outcome: who compensates and who reinforces? *J Popul Econ.* 29(4):969–989.
- Hsin A. 2012. Is biology destiny? Birth weight and differential parental treatment. *Demography* 49(4):1385–1405.
- Rosenzweig MR, Schultz TP. 1982. *The behavior of mothers as inputs to child health: the determinants of birth weight, gestation, and rate of fetal growth.* Chicago (IL): University of Chicago Press. p. 53–92.
- Rosenzweig MR, Schultz TP. 1983. Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight. *J Political Econ.* 91(5):723–746.
- Conway KS, Deb P. 2005. Is prenatal care really ineffective? Or, is the 'devil' in the distribution? *J Health Econ.* 24(3):489–513.
- Lhila A, Simon KI. 2008. Prenatal health investment decisions: does the child's sex matter? *Demography* 45(4):885–905.
- Torche F, Villarreal A. 2014. Prenatal exposure to violence and birth weight in Mexico: selectivity, exposure, and behavioral responses. *Am Sociol Rev.* 79(5):966–992.
- Belsky DW, et al. 2016. The genetics of success: how single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychol Sci.* 27(7):957–972.
- Wertz J, et al. 2018. Using DNA from mothers and children to study parental investment in children's Educational attainment. *bioRxiv* 489781. <https://doi.org/10.1101/489781>, preprint: not peer reviewed.
- Wertz J, et al. 2019. Genetics of nurture: a test of the hypothesis that parents' genetics predict their observed caregiving. *Dev Psychol.* 55:1461–1472.
- Haveman R, Wolfe B. 1995. The determinants of children's attainments: a review of methods and findings. *J Econ Lit.* 33(4):1829–1878.
- Li H, Zhang J, Zhu Y. 2008. The quantity-quality trade-off of children in a developing country: identification using Chinese twins. *Demography* 45(1):223–243.
- Conley D. 2008. Bringing sibling differences in: enlarging our understanding of the transmission of advantage in families. In: *Social class: how does it work.* New York (NY): Russell Sage Foundation. p. 179–200.
- Lee JJ, et al. 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet.* 50:1112–1121.
- Barcellos SH, Carvalho LS, Turley P. 2018. Education can reduce health differences related to genetic risk of obesity. *Proc Natl Acad Sci USA.* 115(42):E9765–E9E72.
- Belsky DW, et al. 2018. Genetic analysis of social-class mobility in five longitudinal studies. *Proc Natl Acad Sci USA* 115(31):E7275–E7E84.
- Kong A, et al. 2018. The nature of nurture: effects of parental genotypes. *Science* 359(6374):424–428.
- Trejo S, Domingue BW. 2019. Genetic nature or genetic nurture? Quantifying bias in analyses using polygenic scores. *bioRxiv* 524850. <https://doi.org/10.1101/524850>, preprint: not peer reviewed.
- Black SE, Devereux PJ, Salvanes KG. 2005. The more the merrier? The effect of family size and birth order on children's Education. *Q J Econ.* 120(2):669–700.
- Price J. 2008. Parent-child quality time does birth order matter? *J Hum Resour.* 43(1):240–265.
- Booth AL, Kee HJ. 2009. Birth order matters: the effect of family size and birth order on educational attainment. *J Popul Econ.* 22(2):367–397.
- Fry A, et al. 2017. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol.* 186(9):1026–1034.
- Domingue BW, Fletcher J. 2020. Separating measured genetic and environmental effects: evidence linking parental genotype and adopted child outcomes. *Behav Genet.* 50(5):301–309.
- Cheesman R, et al. 2019. Comparison of adopted and non-adopted individuals reveals gene-environment interplay for education in the UK Biobank. *bioRxiv* 707695. <https://doi.org/10.1101/707695>, preprint: not peer reviewed.
- Rietveld CA, et al. 2013. GWAS Of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340(6139):1467–1471.

- 
- 35 Okbay A, et al. 2016. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533(7604):539.
- 36 Salvatore JE, et al. 2020. Sibling comparisons elucidate the associations between educational attainment polygenic scores and alcohol, nicotine and cannabis. *Addiction* 115(2):337–346.
- 37 Morris TT, Davies NM, Hemani G, Smith GD. 2020. Population phenomena inflate genetic associations of complex social traits. *Sci Adv.* 6(16):eaay0328.
- 38 Boardman J, Fletcher J. 2015. To cause or not to cause? That is the question, but identical twins might not have all of the answers. *Soc Sci Med.* 127:198–200.
- 39 Collins R. 2012. What makes UK Biobank special. *Lancet* 379(9822):1173–1174.
- 40 Wood AR, et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 46(11):1173–1186.
- 41 Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- 42 Choi SW, O'Reilly PF. 2019. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience* 8(7):giz082.
- 43 Lee JJ, et al. 2018. Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat Genet.* 50:1112–1121.
- 44 Fletcher JM. 2010. Adolescent depression and educational attainment: results using sibling fixed effects. *Health Econ.* 19(7): 855–871.
- 45 Wooldridge JM. 2010. *Econometric analysis of cross section and panel data*. Cambridge (MA): MIT press.
- 46 Davies NM, Dickson M, Smith GD, Van Den Berg GJ, Windmeijer F. 2018. The causal effects of education on health outcomes in the UK Biobank. *Nat Hum Behav.* 2(2):117–125.