

The production of within-family inequality: Insights and implications of integrating genetic data¹

Jason M. Fletcher
Yuchang Wu
Zijie Zhao
Qiongshi Lu
University of Wisconsin-Madison

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 strong evidence of 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.

¹ The authors gratefully acknowledge use of the facilities of the Center for Demography of Health and Aging at the University of Wisconsin-Madison, funded by NIA Center Grant P30 AG017266. 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. Corresponding Author: Jason Fletcher, jason.fletcher@wisc.edu

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 as well as failing to examine parental responses to endowments that are not associated with birthweight. 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.

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 birthweight. 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.

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 socioeconomic status and estimate the relative ranking effects outlined above.

Results

We analyzed educational attainment in the full sample and a sibling sub-sample from respondents with European ancestry in the UK Biobank. Siblings were matched based on kinship estimates of genetic relatedness (see Methods). SI Appendix Table A1 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 sub-sample (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. SI Table 2 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 ~ 0.3 . On average siblings differ in PGS by ~ 0.75 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 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 0.5 years of schooling more than the younger sibling of the pair (28-30).² The interaction between the indicator for higher ranked EA-PGS and the EA-PGS score 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 Figure 1). This result suggests, on average, compensatory processes within families and a concomitant reduction in within-family inequalities in educational attainment.

Table 2 Column 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 gender gap is smaller in high SES areas, and the compensatory effects are much larger in high SES areas (p -value <0.05) (See Figures 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 penetrance for the case of height. The results suggest no effect and can rule out even modest size effects.

² We would have preferred to control for birth order, but the UKB only asked birth order information for a subset of the sample.

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, non-sibling pairs.

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 datasets.

Our results show lower levels of genetic penetrance for educational attainment for children who have higher PGS than their sibling. This phenomenon is larger 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.

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, 31, 32). 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 (33) 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 (34) and EA3(23) has been a bit more mixed, with the within-family analysis often reducing the estimates of genetic penetrance 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, 35, 36). New work (36) 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)

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(37), 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 continue to reflect family processes rather than pure “genetic effects”. Future work will need to incorporate additional strategies to separate “genetic” and “family” causal processes.

Methods and Materials

Data. We used data from the UK Biobank project(38). 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/.

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 $\sim 22,000$ sibling dyads. We then chose to keep only one dyad from any family with more one dyad, leaving $\sim 17,600$ 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 (39) and Educational Attainment(40). 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 UK Biobank 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 score. Second, we created an indicator for the sibling with the higher Height-PGS score. 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 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: SI Table 1 presents descriptive statistics for the full sample and the analysis sub-sample 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 SI Table 3A and Dyadic Characteristics are shown in SI Table 4A.

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 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.(45)

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, sex), a polygenic score and controls for 20 principle components and a family-clustered error term:

$$\text{education}_{if} = \beta_0 + \beta_1 X_{if} + \beta_2 PGS_{if} + PC_{if} + \varepsilon_{if}$$

We next add sibling fixed effects to the model.

$$\text{education}_{if} = \beta_0 + \beta_1 X_{if} + \beta_2 PGS_{if} + F_f + PC_{if} + \varepsilon_{if}$$

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:

$$\begin{aligned} \text{education}_{if} = & \beta_0 + \beta_1 X_{if} + \beta_2 PGS_{if} \\ & + \beta_3 \text{LargerPGS}_{if} + \beta_4 \text{LargerPGS}_{if} * PGS_{if} + F_f + PC_{if} + \varepsilon_{if} \end{aligned}$$

Where the key coefficient is β_4 , which reflects differences in genetic penetrance 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.³ Therefore, we stratify the model between families who were born in high socioeconomic status areas/years vs. low socioeconomic status areas/years.

³ 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).

Tables

Table 1
Associations between PGS-Education and Educational Attainment
Comparing Between Family and Within Family Results

Outcome	Education	Education	Education
Sample	Full	Sibling	Sibling
Fixed Effects?	None	None	Sibling
PGS (std)	0.908*** (0.008)	0.959*** (0.026)	0.444*** (0.048)
Female	-0.746*** (0.016)	-0.777*** (0.055)	-0.860*** (0.070)
Age Fixed Effects	X	X	X
PC Controls (1-20)	X	X	X
Observations	404,409	34,209	33,044
R-squared	0.068	0.063	0.669

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2
Within Family Associations between PGS-Education and Educational Attainment
Including Sibling Dyad Measures of Relative Position

Outcome Sample Fixed Effects?	Education	Education	Education
	Sibling	High SES Place	Low SES Place
	Sibling	Sibling	Sibling
PGS (std)	0.389*** (0.082)	0.433*** (0.113)	0.353*** (0.119)
Female	-0.857*** (0.070)	-0.626*** (0.096)	-1.083*** (0.103)
Older Sib of Pair	0.443*** (0.090)	0.323*** (0.122)	0.576*** (0.133)
Sib with larger PGS	0.124 (0.076)	0.183* (0.106)	0.078 (0.108)
Larger X PGS	-0.086* (0.050)	-0.179** (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-squared	0.669	0.669	0.641

Notes: Robust standard errors 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. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Falsification Exercise
Within Family Associations between PGS-Height and Height

Outcome	Height Full	Height Sibling None	Height Sibling Sibling	Height Sibling Sibling
Sample				
Fixed Effects?	None			
Height PGS (std)	2.516*** (0.010)	2.557*** (0.034)	2.151*** (0.050)	2.154*** (0.085)
Female	-13.319*** (0.019)	-13.295*** (0.065)	-13.468*** (0.070)	-13.467*** (0.070)
Older Sib of Pair				0.019 (0.089)
Sib with Larger PGS				-0.034 (0.075)
Larger X 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-squared	0.600	0.598	0.901	0.901

Robust standard errors in parentheses (clustered by family), *** p<0.01, ** p<0.05, * p<0.1

Table 4: Falsification Exercise
Between-Relative Associations of PGS-Education, Educational Attainment, and Relative Rankings of PGS

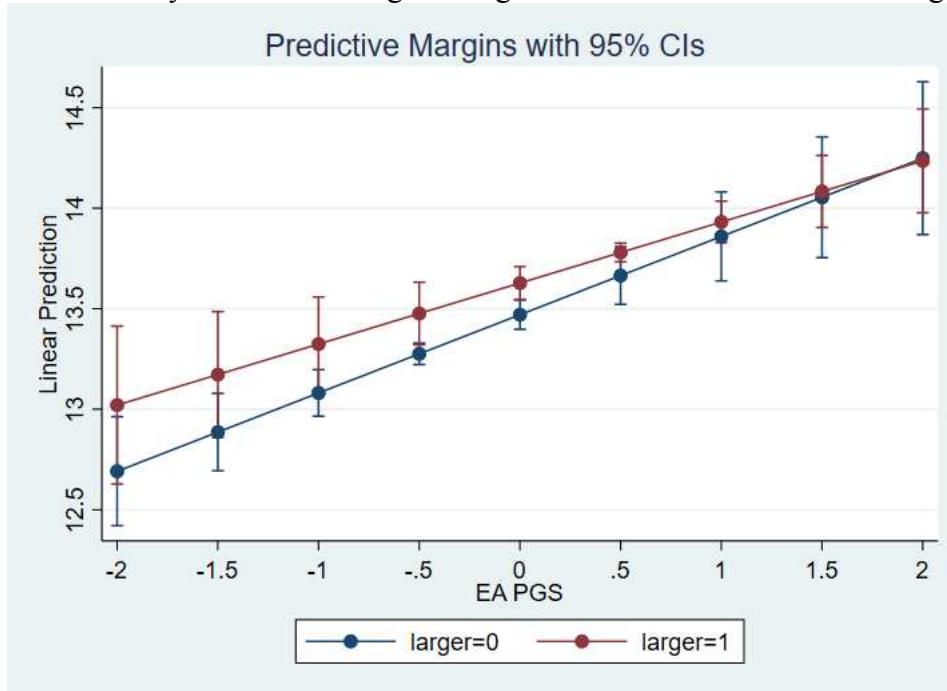
Outcome Sample Fixed Effects?	Education Non Sibling None	Education Non Sibling Family	Education Non-Sibling Family
PGS (std)	0.912*** (0.018)	0.513*** (0.048)	0.493*** (0.053)
Female	-0.638*** (0.037)	-0.709*** (0.052)	-0.709*** (0.052)
Older Relative of Pair		0.119* (0.063)	0.120* (0.063)
Relative with larger PGS		0.089 (0.060)	0.090 (0.061)
Larger X 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-squared	0.070	0.606	0.606

Robust standard errors in parentheses (clustered by “family”), *** p<0.01, ** p<0.05, * p<0.1

Figures

Figure 1

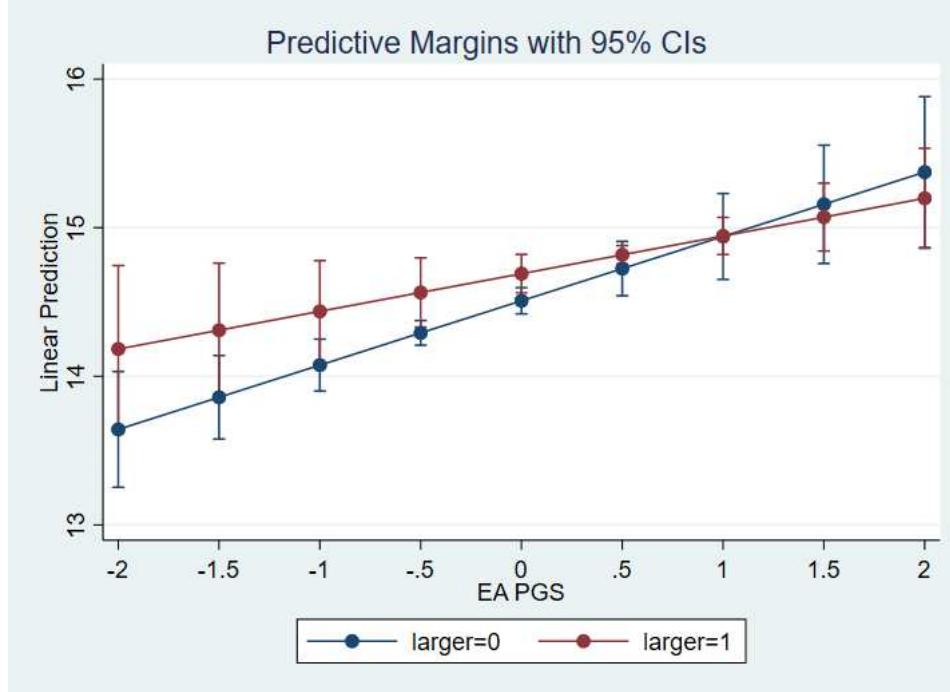
Plot of Associations between EA-PGS and Educational Attainment
Stratified by Whether Sibling has larger or smaller PGS than co-Sibling



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

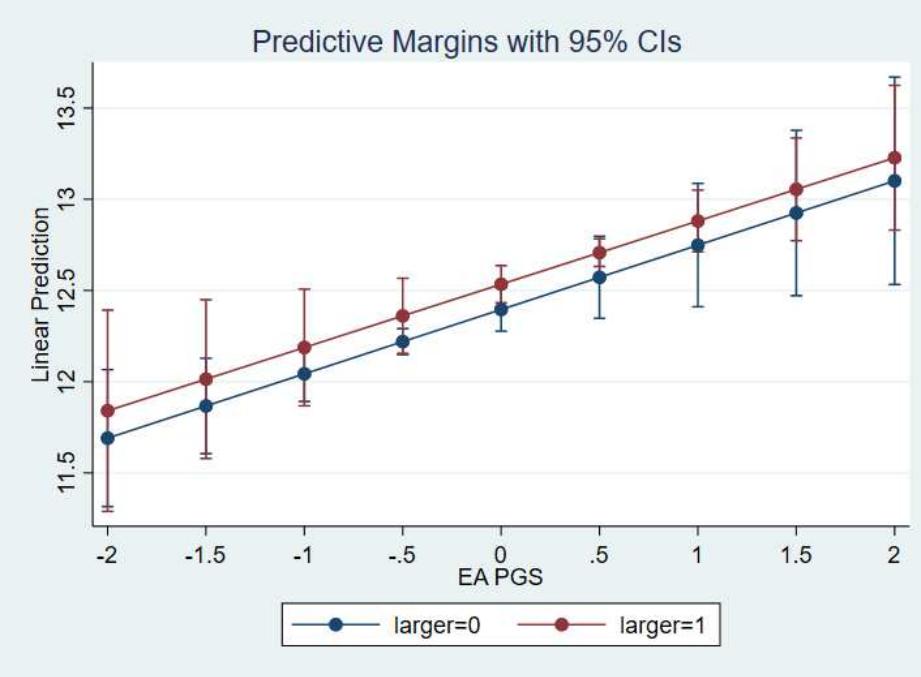
Figure 2
Plot of Associations between EA-PGS and Educational Attainment
Stratified by Whether Sibling has larger or smaller PGS than co-Sibling
High SES Places



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

Figure 1
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

Supplemental Information
Appendix Tables

Table 1A
Comparison of Full UKB Sample and Sibling Analysis Sample

Variable	Full Sample					Sibling Sample				
	Obs	Mean	Std Dev	Min	Max	Obs	Mean	Std Dev	Min	Max
Education	399,442	13.8	5.1	7.0	20.0	33,815	13.6	5.1	7.0	20.0
Female	403,183	0.54	0.5	0.0	1.0	34,095	0.58	0.5	0.0	1.0
Age	403,183	56.8	8.0	39.0	73.0	34,095	57.0	7.3	40.0	70.0
EA3 PGS (std)	403,183	0.00	1.0	-4.7	5.0	34,095	0.05	1.0	-3.7	4.0
Number of Siblings	397,344	2.0	1.8	0.0	32.0	34,019	2.8	1.9	1.0	22.0
PC 1	403,183	-12.4	1.6	-18.5	-6.1	34,095	-12.4	1.6	-18.3	-6.1
PC 2	403,183	3.8	1.5	-2.0	9.5	34,095	3.8	1.5	-1.7	9.4
PC 3	403,183	-1.6	1.6	-7.7	4.4	34,095	-1.6	1.6	-7.4	4.3
PC 4	403,183	1.3	2.9	-10.1	12.8	34,095	1.5	2.9	-9.8	12.5
PC 5	403,183	-0.7	6.7	-17.7	24.8	34,095	-0.5	6.9	-16.3	24.7
PC 6	403,183	-0.4	1.6	-6.7	6.0	34,095	-0.4	1.6	-6.6	5.9
PC 7	403,183	0.3	1.9	-17.7	11.5	34,095	0.3	1.8	-13.8	7.2
PC 8	403,183	-0.5	2.0	-9.0	16.7	34,095	-0.5	1.9	-8.7	16.7
PC 9	403,183	0.0	4.6	-36.6	12.6	34,095	0.3	4.4	-32.4	11.4
PC 10	403,183	0.2	2.3	-11.3	11.4	34,095	0.2	2.2	-9.8	9.8
PC 11	403,183	0.0	3.6	-11.0	20.3	34,095	0.0	3.5	-10.0	16.6
PC 12	403,183	0.2	2.3	-11.3	10.9	34,095	0.2	2.3	-9.2	9.0
PC 13	403,183	0.0	1.6	-7.2	7.2	34,095	0.0	1.6	-7.2	7.0
PC 14	403,183	0.2	3.3	-14.1	18.2	34,095	0.2	3.2	-11.5	17.3
PC 15	403,183	0.0	1.9	-13.8	9.6	34,095	0.0	1.8	-10.8	9.0
PC 16	403,183	0.0	3.1	-27.8	20.5	34,095	0.1	3.0	-22.7	16.7
PC 17	403,183	0.0	2.0	-9.9	8.7	34,095	0.0	1.9	-9.1	7.9
PC 18	403,183	0.0	2.9	-16.6	21.2	34,095	0.0	2.8	-12.2	15.9
PC 19	403,183	0.0	2.9	-11.7	14.6	34,095	0.0	2.7	-9.7	11.6
PC 20	403,183	0.0	2.9	-12.6	14.9	34,095	0.1	2.7	-10.9	12.9

Table 2A
Dyadic (Siblings) Measures
Differences and Correlations

Variable	Obs	Differences				Correlations
		Mean	Std Dev	Min	Max	
EA PGS (Std)	16,792	0.77	0.59	0.0	4.3	0.54
Height PGS (Std)	16,792	0.75	0.57	0.0	4.1	0.57
PC1	16,792	1.21	0.93	0.0	6.0	0.54
Education	16,522	4.14	4.24	0.0	13.0	0.32
Height	16,734	8.68	6.62	0.0	60.0	0.29
Age	16,792	4.25	2.75	0.0	12.0	0.76

Table 3A
Descriptive Statistics for Relatives (Non-Full-Siblings) Sample

Variable	Obs	Mean	Std Dev	Min	Max
Education	62,360	13.19	5.12	7.00	20.00
Female	62,360	0.54	0.50	0.00	1.00
Age	62,360	56.96	7.94	39.00	70.00
EA3 PGS (std)	62,360	-0.02	1.00	-4.57	4.08
Height PGS (std)	62,360	-0.02	1.00	-4.40	4.19
Height (cm)	62,229	168.37	9.22	100.00	204.00
Larger EA3 PGS than Relative	62,360	0.50	0.50	0.00	1.00
Older Than Relative	62,360	0.50	0.50	0.00	1.00
Number of Siblings	61,312	2.55	1.86	1.00	23.00
PC 1	62,360	-12.40	1.59	-18.43	-6.10
PC 2	62,360	3.79	1.50	-1.91	9.50
PC 3	62,360	-1.62	1.56	-7.39	4.30
PC 4	62,360	1.53	2.90	-10.14	12.46
PC 5	62,360	-0.37	6.78	-16.02	24.79
PC 6	62,360	-0.38	1.61	-6.72	5.94
PC 7	62,360	0.27	1.82	-17.67	8.27
PC 8	62,360	-0.46	1.93	-8.44	16.07
PC 9	62,360	0.18	4.37	-33.88	11.36
PC 10	62,360	0.19	2.19	-9.10	10.72
PC 11	62,360	0.08	3.43	-10.62	17.37
PC 12	62,360	0.14	2.28	-10.88	9.56
PC 13	62,360	0.03	1.59	-6.88	6.78
PC 14	62,360	0.21	3.15	-11.91	18.16
PC 15	62,360	0.00	1.82	-12.26	8.53
PC 16	62,360	-0.02	2.97	-22.79	18.16
PC 17	62,360	0.00	1.87	-8.24	8.43
PC 18	62,360	0.10	2.80	-12.22	17.61
PC 19	62,360	-0.05	2.71	-11.39	11.43
PC 20	62,360	0.13	2.70	-12.45	13.17

Table 4A
Dyadic (Relatives, Non Full Siblings) Measures
Differences and Correlations

Variable	Obs	Differences			Correlations	
		Mean	Std Dev	Min	Max	
EA3 (Std)	31180	1.0	0.77	0.0	5.2	0.20
Height PGS (Std)	31180	1.0	0.8	0.0	5.4	0.19
PC1	31180	1.6	1.22	0.0	9.9	0.20
Education	31180	5.0	4.4	0.0	13.0	0.16
Height	31049	10.0	7.32	0.0	73.5	0.10
Age	31180	8.4	6.67	0.0	30.0	0.09

References

1. Becker GS, Tomes N. Human capital and the rise and fall of families. *Journal of labor economics*. 1986;4(3, Part 2):S1-S39.
2. Becker GS, Lewis HG. On the Interaction between the Quantity and Quality of Children. *Journal of political Economy*. 1973;81(2, Part 2):S279-S88.
3. Behrman JR, Rosenzweig MR, Taubman P. Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *Journal of political economy*. 1994;102(6):1131-74.
4. Grätz M, Torche F. Compensation or reinforcement? The stratification of parental responses to children's early ability. *Demography*. 2016;53(6):1883-904.
5. Behrman JR, Pollak RA, Taubman P. Parental preferences and provision for progeny. *Journal of Political Economy*. 1982;90(1):52-73.
6. Cunha F, Heckman J. The technology of skill formation. *American Economic Review*. 2007;97(2):31-47.
7. Cunha F, Heckman JJ. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*. 2008;43(4):738-82.
8. Aizer A, Cunha F. The production of human capital: Endowments, investments and fertility. *National Bureau of Economic Research*, 2012 0898-2937.
9. Datar A, Kilburn MR, Loughran DS. Endowments and parental investments in infancy and early childhood. *Demography*. 2010;47(1):145-62.
10. Restrepo BJ. Parental investment responses to a low birth weight outcome: who compensates and who reinforces? *Journal of Population Economics*. 2016;29(4):969-89.
11. Hsin A. Is biology destiny? Birth weight and differential parental treatment. *Demography*. 2012;49(4):1385-405.
12. Rosenzweig MR, Schultz TP. The behavior of mothers as inputs to child health: the determinants of birth weight, gestation, and rate of fetal growth. *Economic aspects of health*: University of Chicago Press; 1982. p. 53-92.
13. Rosenzweig MR, Schultz TP. Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight. *Journal of political economy*. 1983;91(5):723-46.
14. Conway KS, Deb P. Is prenatal care really ineffective? Or, is the 'devil' in the distribution? *Journal of Health Economics*. 2005;24(3):489-513.
15. Lhila A, Simon KI. Prenatal health investment decisions: Does the child's sex matter? *Demography*. 2008;45(4):885-905.
16. Torche F, Villarreal A. Prenatal exposure to violence and birth weight in Mexico: Selectivity, exposure, and behavioral responses. *American sociological review*. 2014;79(5):966-92.
17. Belsky DW, Moffitt TE, Corcoran DL, Domingue B, Harrington H, Hogan S, Houts R, Ramrakha S, Sugden K, Williams BS. The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*. 2016;27(7):957-72.
18. Wertz J, Moffitt TE, Agnew-Blais J, Arseneault L, Belsky DW, Corcoran DL, Houts R, Matthews T, Prinz JA, Richmond-Rakerd LS. Using DNA from mothers and children to study parental investment in children's educational attainment. *bioRxiv*. 2018:489781.

19. Wertz J, Belsky J, Moffitt TE, Belsky DW, Harrington H, Avinun R, Poulton R, Ramrakha S, Caspi A. Genetics of nurture: A test of the hypothesis that parents' genetics predict their observed caregiving. *Developmental psychology*. 2019.
20. Haveman R, Wolfe B. The determinants of children's attainments: A review of methods and findings. *Journal of economic literature*. 1995;33(4):1829-78.
21. Li H, Zhang J, Zhu Y. The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography*. 2008;45(1):223-43.
22. Conley D. Bringing sibling differences in: Enlarging our understanding of the transmission of advantage in families. *Social class: How does it work*. 2008:179-200.
23. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linnér R, Fontana MA, Kundu T, Lee C, Li H, Li R, Royer R, Timshel PN, Walters RK, Willoughby EA, Yengo L, Alver M, Bao Y, Clark DW, Day FR, Furlotte NA, Joshi PK, Kemper KE, Kleinman A, Langenberg C, Mägi R, Trampush JW, Verma SS, Wu Y, Lam M, Zhao JH, Zheng Z, Boardman JD, Campbell H, Freese J, Harris KM, Hayward C, Herd P, Kumari M, Lencz T, Luan Ja, Malhotra AK, Metspalu A, Milani L, Ong KK, Perry JRB, Porteous DJ, Ritchie MD, Smart MC, Smith BH, Tung JY, Wareham NJ, Wilson JF, Beauchamp JP, Conley DC, Esko T, Lehrer SF, Magnusson PKE, Oskarsson S, Pers TH, Robinson MR, Thom K, Watson C, Chabris CF, Meyer MN, Laibson DI, Yang J, Johannesson M, Koellinger PD, Turley P, Visscher PM, Benjamin DJ, Cesarini D. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*. 2018. doi: 10.1038/s41588-018-0147-3.
24. Barcellos SH, Carvalho LS, Turley P. Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*. 2018;115(42):E9765-E72.
25. Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, Conley D, Fletcher JM, Freese J, Herd P. Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences*. 2018;115(31):E7275-E84.
26. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, Benonisdottir S, Oddsson A, Halldorsson BV, Masson G. The nature of nurture: Effects of parental genotypes. *Science*. 2018;359(6374):424-8.
27. Trejo S, Domingue BW. Genetic nature or genetic nurture? quantifying bias in analyses using polygenic scores. *bioRxiv*. 2019:524850.
28. Black SE, Devereux PJ, Salvanes KG. The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*. 2005;120(2):669-700.
29. Price J. Parent-child quality time does birth order matter? *Journal of human resources*. 2008;43(1):240-65.
30. Booth AL, Kee HJ. Birth order matters: the effect of family size and birth order on educational attainment. *Journal of Population Economics*. 2009;22(2):367-97.
31. Domingue BW, Fletcher J. Separating Measured Genetic and Environmental Effects: Evidence Linking Parental Genotype and Adopted Child Outcomes. *Behav Genet*. 2020. Epub 2020/05/01. doi: 10.1007/s10519-020-10000-4. PubMed PMID: 32350631.
32. Cheesman R, Hunjan A, Coleman JR, Ahmadzadeh Y, Plomin R, McAdams TA, Eley TC, Breen G. Comparison of adopted and non-adopted individuals reveals gene-environment interplay for education in the UK Biobank. *BioRxiv*. 2019:707695.

33. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra H-J, Shakhbazov K, Abdellaoui A, Agrawal A. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *science*. 2013;340(6139):1467-71.

34. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen G-B, Emilsson V, Meddents SFW. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016;533(7604):539.

35. Salvatore JE, Barr PB, Stephenson M, Aliev F, Kuo SIC, Su J, Agrawal A, Almasy L, Bierut L, Bucholz K. Sibling comparisons elucidate the associations between educational attainment polygenic scores and alcohol, nicotine and cannabis. *Addiction*. 2020;115(2):337-46.

36. Morris TT, Davies NM, Hemani G, Smith GD. Population phenomena inflate genetic associations of complex social traits. *Science Advances*. 2020;6(16):eaay0328.

37. Boardman J, Fletcher J. To cause or not to cause? That is the question, but identical twins might not have all of the answers. *Social Science & Medicine*. 2015;127:198-200. doi: 10.1016/j.socscimed.2014.10.013. Epub 2014 Oct 8; PMCID: PMC5830117.

38. Collins R. What makes UK Biobank special. *Lancet*. 2012;379(9822):1173-4.

39. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Ja, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Leach IM, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Ärnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson Å, Jolley J, Juliusdottir T, Junntila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundström J, Swertz MA, Syvänen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein

SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin M-R, Jousilahti P, Jula AM, Kaprio J, Kastelein JJP, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lupoli S, Madden PAF. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*. 2014;46(11):1173-86. doi: 10.1038/ng.3097.

40. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Karlsson Linner R, Fontana MA, Kundu T, Lee C, Li H, Li R, Royer R, Timshel PN, Walters RK, Willoughby EA, Yengo L, andMe Research T, Cogent, Social Science Genetic Association C, Alver M, Bao Y, Clark DW, Day FR, Furlotte NA, Joshi PK, Kemper KE, Kleinman A, Langenberg C, Magi R, Trampush JW, Verma SS, Wu Y, Lam M, Zhao JH, Zheng Z, Boardman JD, Campbell H, Freese J, Harris KM, Hayward C, Herd P, Kumari M, Lencz T, Luan J, Malhotra AK, Metspalu A, Milani L, Ong KK, Perry JRB, Porteous DJ, Ritchie MD, Smart MC, Smith BH, Tung JY, Wareham NJ, Wilson JF, Beauchamp JP, Conley DC, Esko T, Lehrer SF, Magnusson PKE, Oskarsson S, Pers TH, Robinson MR, Thom K, Watson C, Chabris CF, Meyer MN, Laibson DI, Yang J, Johannesson M, Koellinger PD, Turley P, Visscher PM, Benjamin DJ, Cesarini D. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50(8):1112-21. Epub 2018/07/25. doi: 10.1038/s41588-018-0147-3. PubMed PMID: 30038396; PMCID: PMC6393768.

41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-75. Epub 2007/07/25. doi: 10.1086/519795. PubMed PMID: 17701901; PMCID: PMC1950838.

42. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 2019;8(7). doi: 10.1093/gigascience/giz082. PubMed PMID: 31307061; PMCID: PMC6629542.

43. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Johannesson M, Koellinger P, Turley P, Visscher P. Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat Genet*, in press. 2018.

44. Fletcher JM. Adolescent Depression and Educational Attainment: Results Using Sibling Fixed Effects. *Health Economics*. 2010;19(7):855-71. doi: 10.1002/hec.1526.

45. Davies NM, Dickson M, Smith GD, Van Den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. *Nature human behaviour*. 2018;2(2):117-25.