NBER WORKING PAPER SERIES

SIBLING CORRELATION IN EDUCATIONAL ATTAINMENT: A TEST OF GENETIC NURTURE

John Cawley
Euna Han
Jiyoon Kim
Edward C. Norton

Working Paper 27336
http://www.nber.org/papers/w27336

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w27336.ack

Sibling Correlation in Educational Attainment: A Test of Genetic Nurture
John Cawley, Euna Han, Jiyoon Kim, and Edward C. Norton
NBER Working Paper No. 27336
June 2020
JEL No. D1,D13,I2,I21,I23,I24,J13

ABSTRACT

The educational attainment of siblings is highly correlated. We test for a specific type of peer effect between siblings in educational attainment: genetic nurture. Specifically, we test whether a person's educational attainment is correlated with their sibling's polygenic score (PGS) for educational attainment, controlling for their own PGS for educational attainment. Models estimated using genetic data in the National Longitudinal Survey of Adolescent to Adult Health (Add Health) yield strong evidence of such genetic nurture between siblings, and this result is robust to alternative measures of educational attainment and different measures of polygenic score.

John Cawley
2312 MVR Hall
Department of Policy Analysis
and Management
and Department of Economics
Cornell University
Ithaca, NY 14853
and NBER
JHC38@cornell.edu

Euna Han
College of Pharmacy
Yonsei University
162-1 Songdo-Dong
Yeosu-Gu
Incheon, South Korea
eunahan@yonsei.ac.kr

Jiyoon Kim
Department of Economics
Swarthmore College
500 College Avenue
Swarthmore, PA 19081
junekim@umich.edu

Edward C. Norton
Department of Health Management and Policy
Department of Economics
University of Michigan
School of Public Health
1415 Washington Heights, M3108 SPHII Ann
Arbor, MI 48109-2029
and NBER
ecnorton@umich.edu

I. Introduction

Educational attainment is an important determinant of current and future socioeconomic status (SES) (Bandura 2001; Hastings et al. 2013; Hanushcek and Woessman 2020; Ashby 2010; Wickrama 2013; Blau and Duncan 1967; Nielsen and Roos 2015), and has also been described as the most important correlate of good health (Grossman and Kaestner, 1997). Given the importance of educational attainment for SES and health, there is great interest in factors that determine it, such as within-family influences (Black et al., 2011).

Many studies have noted a high correlation of educational attainment between siblings, which varies by the age and country of the sample, and ranges from 0.4 to 0.6 (Lindahl, 2011; Raaum et al., 2006; Solon, 2000; Nicoletti and Rabe, 2013). Siblings might influence each other's educational attainment in a number of ways (Griliches, 1979; Conley et al., 2007; Yi, Heckman, et al. 2015). Youths might tutor their siblings or help them with their homework. Youths may serve as role models or examples for their siblings (Benin and Johnson 1984; Nicoletti and Rabe 2019). Youths may update their beliefs about their own academic potential based on the achievements of their siblings (Altmejd et al., 2020). In addition, parents might reallocate family resources based on each child's academic performance (Griliches, 1979; Yi, Heckman et al., 2015).

However, estimating peer effects is notoriously difficult (Manski, 1993, Manski, 2000; Angrist, 2014). One challenge is the "reflection problem" — it is hard to distinguish the effect of sibling A on sibling B from vice-versa. Another challenge is correlated environments or shocks that are unobserved to the econometrician. In other contexts there is also the challenge of self-selection, that people may choose peers similar to themselves, although this is not an issue when the peers in question are full siblings and thus they did not self-select.

There have been several creative studies with credible identification strategies to address these challenges. Exploiting admissions thresholds for colleges, Altmejd et al. (2020) find evidence that an older sibling's enrollment in a more competitive college increases the probability that the younger sibling enrolls in college at all. Landersø et al. (2019) exploit Danish school entry policies and estimate that a later school start for the focal child improves the academic performance of older siblings. Karbownik and Ozek (2019) estimate regression discontinuity models around school starting ages using data from a school district in Florida. They find evidence of positive spillovers from older to younger siblings in lower-income households — younger siblings earned higher test scores if their older siblings were born right after (as opposed to right before) the school-entry cutoff, and thus tend to be the old for their grade. Interestingly, they find the opposite effect for higher-income households; there, students perform worse if their older sibling is old for their grade. Joensen et al. (2018) exploit an unexpected pilot scheme in Denmark which lowered the cost of advanced math and science at high school. They estimate that students were 2-3 percentage points more likely to take advanced math or science courses in high school if their older siblings did. Qureshi (2017) found that younger siblings benefit if their older sibling had a more experienced teacher; there is no such benefit from younger siblings having more experienced teachers.

In this study, we contribute to the literature on sibling effects in education and, more broadly, to that on peer effects in education (see the review in Paloyo (2020)). Specifically, we test for a particular form of sibling influence on education: *genetic nurture*, which is defined as one individual's genes affecting another individual's outcomes (Wolf, Brodie, Cheverud, Moore, & Wade, 1998). Empirically, we test whether the focal person's educational attainment is associated with their sibling's polygenic score (PGS) for educational attainment, controlling for the focal person's own PGS for educational attainment. A *polygenic score* (PGS) is an index of the number of alleles — or specific forms of a gene —that are associated with a specific trait; in this case, educational attainment. In the absence of genetic nurture, the focal person's educational attainment should not be associated with their sibling's PGS after controlling for the focal person's own PGS for that outcome.

This empirical approach addresses the problems that confound empirical research on peer effects. There is no issue of self-selection because the peers in question are full siblings (people who have the same two biological parents) so they do not self-select into the peer group. The reflection problem is addressed because we examine the effect of the PGS, which is determined before birth (at conception).

This paper also contributes to the literature that tests for genetic nurture. Kong et al. (2018) test for genetic nurture from parents to children; specifically, they test whether parental alleles that were not transmitted to (i.e., not inherited by) the child are still associated with the child's outcomes. In their Icelandic sample, they find that parental alleles associated with education that were *not* transmitted to the child have 29.9 percent effect of the alleles that *were* transmitted to the child. Domingue et al. (2015) test for genetic nurture between friends in the National Longitudinal Survey of Adolescent to Adult Health (Add Health); they find that educational attainment is correlated with the average of friends' PGS for educational attainment, even after controlling for own PGS. However, because friends are selected, this correlation could be the result of assortative matching or omitted variable bias from similar environments. In an earlier study (Cawley, Han, Kim, Norton, 2019), we found no evidence of genetic nurture among siblings with respect to body mass index or obesity.

Testing for genetic nurture is only possible if genes in fact influence the trait in question. Numerous studies have confirmed that educational attainment has a strong genetic component. A meta-analysis by Branigan, McCallum, and Freese (2013) estimates that the heritability of educational attainment is roughly 40%. Of course, there is variation in these estimates. The estimated heritability of educational achievement has declined over decades in the United States (Nielsen and Roos 2015; Reardon 2011; Bailey and Dynarski 2011; Branigan, McCallum, and Freese, 2013). A recent study using Add Health estimates that the heritability of highest grade completed is roughly 23% (Nielsen and Roos 2015) in the US. Studies using data from Denmark, Sweden, and Finland estimate that the heritability of educational attainment is 35% to 50% (Bingley, Christensen, and Walker 2005; Isacsson 1999; Silventoinen et al. 2004). Okbay (2016) identifies 74 genome-wide significant loci associated with the number of years of schooling, which are mostly found in genomic regions regulating gene expression in fetal brain. The high explanatory power found in these studies makes educational attainment well-suited for a test of genetic nurture.

II. Methods

Our empirical test for genetic nurture in educational attainment exploits the natural experiment of Mendelian randomization. The variation across full siblings in the set of genes received at conception is the result of each separately drawing genes from the same two parents. To appeal to a classic classroom example in statistics, full siblings each draw balls (genes) from the same two urns (the set of genes of their parents). Although they draw from the same urns, there will be random variation in the genes they draw, which is the concept of Mendelian randomization.

We measure each individual's genetic endowment using the polygenic score (PGS) for educational attainment, which reflects the combined additive influence of single-nucleotide polymorphisms (SNPs) associated with educational attainment across the entire genome. The specific PGS we use is based on a Genome-Wide Association Study (GWAS) of educational attainment with a sample size of 1,131,881 (Okbay et al. 2018, Lee et al. 2018). This PGS represents the associations of more than 1,271 SNPs across entire genome with educational attainment, and explains 11–13% of the variance in that outcome (Lee et al. 2018)

The PGS has several advantages in this context. It has high explanatory power for a single variable and has good out-of-sample reliability. Also, alleles are determined prior to birth (at conception) and are largely immutable. Therefore, alleles are not affected by the actions or characteristics of one's siblings. To clarify, gene *expression* can be affected by the environment (this is the subject of the field of epigenetics), but generally not one's alleles, and thus one's PGS is exogenously determined.

We follow the nomenclature of previous studies (e.g., Cohen-Cole and Fletcher, 2008) and refer to the individual whose educational attainment may be affected as the *ego*, and the individual who may be affecting them as the *alter*.

Our basic reduced-form model regresses the ego i's educational attainment (Y_i) on the alter's PGS for educational attainment ($AlterPGS_i$), the ego's own PGS for educational attainment ($EgoPGS_i$), and various other ego-level exogenous explanatory variables X_i:

$$Y_i = \beta_0 + \beta_1 AlterPGS_i + \beta_2 EgoPGS_i + \beta_3 X_i + \beta_4 PC_i + \epsilon_i \qquad (1)$$

In our main results, the dependent variable Y is an indicator variable for whether the respondent has earned a four-year college degree (BA/BS) or higher. In subsequent robustness checks, we examine alternative measures of educational attainment: (i) years of education; (ii) at least high school completion; and (iii) at least graduate school completion. Given the binary nature of the dependent variable and the goal of estimating an average treatment effect, we estimate linear probability models.

The PGS is equal to the number of *risky* alleles associated with *lower* educational attainment, each weighted by its relative contribution to educational attainment. Therefore, a *higher* PGS score is associated with *lower* educational attainment. Thus, the sign of β_2 (i.e. the coefficient on an individual's own PGS for educational attainment) should be negative.

The test for genetic nurture is whether β_1, the coefficient on the alter's PGS for educational attainment, equals zero. Because the model controls for the ego's own PGS for educational attainment, we are testing whether the *variance in genetic predisposition of the alter that is uncorrelated with that of the ego* is correlated with the ego's educational attainment. If β_1<0, that represents evidence of genetic nurture; specifically, of a form that generates a positive correlation in education among siblings (i.e., if the alter has more risky alleles associated with a lower education, the lower the education for the ego as well). If β_1>0, that is consistent with genetic nurture that generates a negative correlation in education among siblings (i.e., if the alter has more risky alleles associated with a lower education, the higher the education for the ego). Based on the previous literature on sibling peer effects in education, we predict that genetic nurture would create a positive correlation in educational attainment between siblings; in other words, that β_1<0.

To clarify, we do not examine whether specific SNPs inherited by the alter (but not the ego) affect the ego's educational attainment, but instead whether the aggregate measured genetic predisposition of the alter that is independent of that of the ego is correlated with the ego's educational attainment.

The distribution of allele frequency is not random across race and ethnic groups (Price et al., 2006; Conley et al., 2016). If one did not address this, it could cause omitted variable bias. A standard approach in the literature is to control for the principal components of the GWAS; by doing so one controls for ethnic and racial variation across the genome (Price et al., 2006). We follow that approach; the regression model controls for 10 principal components from the GWAS (denoted PC_i in equation 1).

In the regression model (1), the vector X_i includes the following set of control variables: the respondent's age in years; sex; race (African-American, Hispanic, other, with white as the omitted reference category); an indicator for whether the ego currently lives with the parents; total household income (categorical variables for \$20k-\$40k; \$40k-\$50k; \$50k-\$75k; \$75k-\$99; > \$100k, with < \$20k as the reference); parental educational attainment (high school graduate, some college, college graduate, with less than high school as the reference); the respondent's marital status (married, with not married as the reference); and the respondent's employment status (work full time, part time, with unemployed or out-of-labor force as the reference).

One strength of the reduced-form approach is that it does not restrict the possible mechanisms for genetic nurture. They may operate through role model effects, tutoring, budget or time constraints of the parents, or any other mechanism. Moreover, our test may pick up genetic nurture from other family members. Genetic nurture may operate through the observed sibling (the alter), another sibling that we do not observe whose genes are correlated with those of the alter, or the parents (from whom the alter drew genes). Any and all such effects will be reflected in the reduced-form estimate and are of interest to us because they all represent genetic nurture within families affecting educational attainment.

To avoid the selection bias that would result from studying a self-selected peer group, our peer group of interest is full siblings (i.e., siblings with the same two biologic parents). We exclude twins from our sample because they may exhibit different patterns of genetic nurture than other types of siblings. Specifically, twins may actively seek to establish their own identities, or more enthusiastically mimic each other. In addition, the model cannot be estimated for monozygotic (i.e., identical) twins because for them ego and alter genes are perfectly collinear, so there is no independent variation in the alter PGS controlling for ego PGS.

All individuals in the sample appear in two observations per wave of the data: one in which they are the ego (i.e., their educational attainment is the dependent variable) and one in which they are the alter (i.e., their PGS is the key independent variable). To account for the correlation in error terms among siblings, we cluster the standard errors by family.

III. Data

We examine data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), which is a longitudinal study of a nationally representative sample of adolescents in the U.S. The Add Health is an ideal data set to test for genetic nurture because it contains genetic data on a large number of full siblings and has followed the sample for enough years that nearly everyone will have completed their formal education.

We use the most recent wave — Wave IV — which was conducted in 2008–2009 when the subjects were aged 25–34 years old because by that age most people will have completed their formal education. We identify sibling pairs from the Wave I sibling pair data and restrict our sample to full siblings (i.e., siblings with the same two biological parents) who are not twins.

Our main regressor of interest is the polygenic score (PGS) for educational attainment. The PGS was constructed by the Social Science Genetic Association Consortium (SSGAC) and is available for respondents who, during Wave IV, had saliva samples collected by trained field interviewers for genetic testing. Roughly 96% consented to data collection and 78% to archiving of their data. There are two types of polygenic scores available based on different sets of summary statistics: (i) a score based on standard GWAS (Genome Wide Association Study) summary statistics, which are the coefficient estimates from univariate GWAS of educational attainment; and (ii) a score based on MTAG (Multi Trait analysis of GWAS) summary statistics, which are obtained from a multivariate analysis. The correlation between these two polygenic scores for education attainment is 0.95. Because it is not *a priori* obvious which measure is preferred, we run all analyses twice, once with each measure of PGS.

There are 1,251 full sibling pairs in Wave I sibling data. Out of them 557 pairs are matched with SSGAC PGS variable. We drop 241 sibling pairs because either of them has missing or invalid PGS for education. When data on control variables (other than PGS) are missing, the variable with the missing value is set to 0 and an indicator for missing values of that variable is set to 1 and included in the model. The main sample has 632 observations for 316 sibling pairs.

IV. Results

A. Summary Statistics

Table 1 presents the summary statistics for the sample. Our primary dependent variable is whether the respondent has completed a four-year college degree; this is true for 34.2% of the sample. The mean educational attainment is just over 14 years. High school graduates constitute 92.1% of the sample, and 12.0% have a graduate school degree. The PGS (based on MTAG) has a mean of 0.294 with a standard deviation of 0.16 and a minimum value of -0.165 and maximum value of 0.669. The PGS (based on GWAS) has a similar distribution with slightly smaller mean (0.202) and standard deviation (0.146). The sample is overwhelmingly white (98%), with a slight majority female (52.2%) and mean age of 28.5.

The correlation in years of education between siblings is 0.44 and the correlation in college completion is 0.42. The correlation between siblings in PGS based on MTAG is 0.52 and that in PGS based on GWAS is 0.49.

B. Results of Reduced-Form Regressions

Table 2 contains the results of our reduced-form regressions in which the dependent variable is an indicator variable for having a college degree or higher education. Columns 1 and 2 report results from models using the PGS based on MTAG, whereas column 3 and 4 report results from models using the PGS based on GWAS. For each measure of the PGS (MTAG and GWAS), we estimate two models: with and without controlling for parental educational attainment.

Looking across the first row of Table 2, the *ego's own* PGS is consistently negatively and significantly correlated with the probability that the ego has a college degree. This makes sense; the PGS was constructed to predict the person's own educational attainment. The negative sign is to be expected, given that the PGS is a count of alleles that predispose one to lower educational attainment. Therefore, the higher the PGS, the lower the probability of a college degree.

Recall that our test of genetic nurture is whether the *alter's* PGS is correlated with the ego's educational attainment, after controlling for the ego's PGS for educational attainment. The results, shown in in the second row of Table 2, are consistent with genetic nurture between siblings in educational attainment. Across all four models, the alter's PGS is negatively and significantly correlated with the probability that the ego has a college degree. These reduced-form results cannot tell us the mechanism for these effects, but indicate that the effects do exist. The effects are also substantial in magnitude: the correlation of alter's PGS with the ego's educational attainment is roughly one-third to one-half as large as the correlation of that outcome with the ego's own PGS.

To better explain the magnitudes of the coefficients, we describe the change in the probability that the ego has a college degree, associated with a two-standard-deviation change in alter's PGS. This corresponds to a change big enough to move the alter from the 50^{th} percentile of PGS to the 98^{th} percentile of PGS. In models that control for parental education, a two-standard deviation decrease in alter's PGS (associated with a better-educated alter) is associated with a 7.8% increase in the probability that the ego has a college degree. That is based on the PGS based on MTAG.   If the PGS based on GWAS is used, the corresponding change is 7.1%. For both measures of the PGS, the magnitude of the estimated genetic nurture is substantial.

The finding of genetic nurture is robust to alternative measures of PGS and of including parental education. Adding a control for parental educational attainment should reduce the coefficient on alter PGS because both are correlated with the unobserved parental genes associated with education. It is debatable whether the model should include parental education. An argument against including it is that the alter's genes were drawn from the parents' genes, so controlling for parental education may be an over-control (Angrist and

Pischke, 2009). On the other hand, controlling for parental education may give us a better sense of whether the genetic nurture operates through the sibling or the parent. We find, in columns 2 and 4 of Table 2, that controlling for parental education reduces the coefficient on alter's PGS by about one-third (and reduces the magnitude of the coefficient on alter PGS relative to that on ego's PGS from roughly 50% to roughly 35-40%), and reduces its significance level from 1% to either 5% or 10%, but it remains statistically significant. This suggests that some of the genetic nurture is operating through the sibling, as opposed to exclusively through the parents.

For additional robustness checks, we also re-estimated the models using different dependent variables that represent alternate measures of educational attainment. These appear in Table 3A (highest grade completed), 3B (binary variable for high school graduate), and 3C (binary variable for graduate school completion). For both highest grade completed (Table 3A) and high school completion (Table 3B), the pattern is robust that alter's PGS is significantly and negatively correlated with ego's educational attainment, even controlling for ego's own PGS. This is true using the PGS based on MTAG or the PGS based on GWAS, and whether or not one controls for parental education. In Table 3B, which concerns high school completion, the coefficients on alter PGS are actually larger than those on ego PGS, although the difference is not statistically significant.

The pattern is slightly different for graduate school completion (Table 3C). For that outcome, the alter's PGS predicts ego's educational attainment but only when parental education is excluded from the model. When the model controls for parental education, the coefficient on alter's PGS is smaller and not statistically significant. On the whole, however, the pattern that alter's PGS is significantly negatively correlated with ego's educational attainment is strongly robust for three measures of educational attainment (college completion, high school completion, and years of schooling). For subsequent models, we consistently use the dependent variable of college completion.

Several past studies of sibling peer effects in education have found evidence of heterogeneity (Paloyo, 2020), such as a greater peer effect from older to younger siblings than vice-versa (e.g. Qureshi, 2017). It is possible that the degree of genetic nurture depends on the age difference between the siblings and whether they are the same as opposed to opposite sex. To test this, we add to the model controls for the age difference in years between the siblings, and an indicator for whether the siblings are the same sex. We also add interactions of the alter's PGS with those two variables. The results, shown in Table 4, are somewhat mixed. The interaction term between age difference and alter PGS is positive and statistically significant, but only when the model includes a control for parental education. There is no detectable heterogeneity by whether the siblings are the same versus opposite sex; the interaction between alter PGS and whether the siblings are the same sex is not statistically significant in either case (i.e. whether or not one controls for parental education).

We next test for heterogeneity in genetic nurture in a different way: by estimating our models separately based on whether the ego is older or younger than the alter, whether the ego is the same versus opposite sex as the alter, whether the ego is female or male, and whether the ego's PGS for educational attainment is above or below average.

The results are consistent with heterogeneity in genetic nurture. Table 5A indicates that there is only detectable genetic nurture when the ego is younger (not older) than the alter. Table 5B indicates that genetic nurture is only detectable when the siblings are of different sex, not the same sex. Table 5D indicates that genetic nurture is only detectable when the ego has a PGS that is below average (i.e. predisposed to greater education).

Models estimated separately by ego's sex (Table 5C) suggest that genetic nurture is stronger for female than male egos, although it is only significant when parents' education is not controlled for. All of the aforementioned results used the PGS based on MTAG; we also estimated models using the PGS based on GWAS. Those estimates (shown in Tables 6A-6D) are very similar, with the exception that there is stronger evidence that genetic nurture exists only for female egos. Overall, the results suggest that genetic nurture is stronger when: the ego is younger than the alter, the ego and the alter are of different sex, the ego has a below-average PGS (implying higher education), and the ego is female.

V. Discussion

We test for evidence of genetic nurture in educational attainment by regressing an individual's educational attainment on both their own polygenic score (PGS) for educational attainment, and that of their full sibling (the alter). We consistently find that the alter's PGS is significantly correlated with the ego's educational attainment; this is robust to using various measures of educational attainment and two different calculations of PGS, and whether or not one controls for parental education.

Moreover, the magnitude is large; a 2-standard-deviation decrease in alter PGS (implying a predisposition to higher education in the alter) is associated with a 7.1% to 7.8% increase in the probability that the ego has a college degree. Subgroup analyses suggest that genetic nurture is stronger when the ego is younger than the alter, the ego and the alter are of a different sex, and when the ego has a below-average PGS (implying higher education).

These findings contribute to the literature on peer effects in education (e.g. Paloyo, 2020) and more specifically on family effects on education (Altmejd et al., 2020; Landersø et al., 2019; Karbownik and Ozek, 2019; Joensen et al., 2018; Qureshi, 2017). In particular, the results of this paper add to the evidence of within-family spillovers in education.

The following factors should be kept in mind when generalizing from these results. We study genetic nurture for one type of peer: full siblings. This is an important strength, because it is an exogenously determined peer group; any similarity cannot be due to self-selection. However, other peer groups may exhibit different degrees of genetic nurture.

The fact that we find strong and consistent evidence of genetic nurture in education does not mean that there is genetic nurture among full siblings in all outcomes. In fact, our prior study (Cawley, Han, Kim, and Norton, 2019) using the same methods consistently failed to detect any evidence of genetic nurture in body mass index (BMI) or waist circumference. The non-significance of those results was not driven by imprecision; the point estimates were quite small in magnitude. Thus, genetic nurture does not seem to be a general rule among siblings for all heritable traits.

Other limitations of the study include that the PGS explains a modest amount of variance in educational attainment: 11–13% (Lee et al. 2018). As more powerful PGS become available, further studies should be conducted. Another issue for generalizability is that we examine siblings during the ages of 25 to 34 years; it is possible that genetic nurture in education differs with age. Despite these limitations, this study provides important evidence of genetic nurture — and, more generally, family effects — in educational attainment.

Works Cited

Altmejd, Adam, Andres Barrios-Fernandez, Marin Drlje, Joshua Goodman, , … , and Jonathan Smith. 2020. “O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries.” Ed Working Paper: 20-230.

Angrist, Joshua. 2014. “The Perils of Peer Effects.” *Labour Economics* 30(C): 98–108.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist’s Companion. Princeton, NJ: Princeton University Press

Ashby, Julie, and Ingrid Schoon. 2010. “Career Success: The Role of Teenage Career Aspirations, Ambition Value, and Gender in Predicting Adult Social Status and Earnings.” *Journal of Vocational Behavior* 77 (3): 350-360.

Bailey, Martha, and Susan Dynarski. 2011. “Gains and Gaps: Changing Inequality in U.S. College Entry and Completion.” NBER Working Paper No. 17633.

Bandura, Albert. 2001. “Social Cognitive Theory: An Agentic Perspective.” *Annual Review of Psychology* 52: 1-26.

Benin, Mary, and David Johnson. 1984. “Sibling Similarities in Educational Attainment: A Comparison of Like-sex and Cross-sex Sibling Pairs.” *Sociology of Education* 57 (1):11-21.

Black, Sandra, Paul J. Devereaux, and Kjell Salvanes. 2011. “Older and Wiser? Birth Order and IQ of Young Men.” *CESifo Economic Studies* 57(1): 103-120.

Bingley, Paul, Kaare Christensen, and Ian Walker. 2005. “Twin-Based Estimates of the Returns to Education: Evidence from the Population of Danish Twins.” Unpublished Manuscript.

Blau, Peter, and Otis Duncan. 1967. *The American Occupational Structure*. New York: John Wiley & Sons

Branigan, Amelia, Kenneth McCallum, and Jeremy Freese. 2013. “Variation in the Heritability of Educational Attainment: An International Meta-Analysis.” *Social Forces* 92(1): 109-140

Cawley, John, Euna Han, Jiyoon Kim, and Edward C. Norton. 2019. “Testing for Family Influences on Obesity: The Role of Genetic Nurture.” *Health Economics* 28: 937-952.

Cohen-Cole, Ethan, and Jason Fletcher. 2008. “Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic.” *Journal of Health Economics* 27(5): 1382–1387.

Conley, Dalton, Kathryn Pfeiffer, and Melissa Velez. 2007. “Explaining Sibling Differences in Achievement and Behavioral Outcomes: The Importance of Within- and Between-Family Factors.” *Social Science Research* 36 (3): 1087-1104.

Conley, Dalton, Thomas Laidley, Jason Boardman, and Benjamin Domingue. 2016. “Changing Polygenic Penetrance on Phenotypes in the 20th Century Among Adults in the US Population.” *Scientific Reports*, 6, 30348.

Domingue, Benjamin. Daniel Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason Boardman. 2015. “Polygenic Influence on Educational Attainment: New Evidence from The National Longitudinal Study of Adolescent to Adult Health.” *AERA Open* 1(3), 1–13.

Griliches, Zvi. 1979. “Sibling Models and Data in Economics—Beginnings of a Survey.” *Journal of Political Economy* 87(5), S37–S64.

Grossman, Michael, and Robert Kaestner. 1997. *Effects of Education on Health. In the Social Benefits of Education.* Ann Arbor, Michigan: University of Michigan Press.

Hastings, Justine, Christopher Neilson, and Seth Zimmerman. 2013. "Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile." NBER Working Paper No. 19241.

Hanushcek, Eric, and Ludger Woessmann. 2020. "Education, Knowledge Capital, and Economic Growth." In *Economics of Education,* ed. Steve Bradley and Colin Green, 171–182. London: Academic Press.

Isacsson, Gunnar. 1999. "Estimates of the Return to Schooling in Sweden from a Large Sample of Twins." *Labour Economics* 6(4):471—89

Joensen, Juanna, and Helena Nielsen. 2018. "Spillovers in Education Choice." *Journal of Public Economics* 157: 158–183.

Karbownik, Krzysztof, and Umut Ozek. 2019. "Setting a Good Example? Examining Sibling Spillovers in Education Achievement Using a Regression Discontinuity Design." *CESifo Working Paper Series* 7531

Kong, Augustine, Gudmar Thorleifsson, Michael Frigge, Bjarni Vilhjalmsson, Alexander Young, … , and Kari Stefansson. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359(6374): 424–428.

Landersø, Rasmus, Helena Nielsen, and Marianne Simonsen. 2019. "Effects of School Starting Age on the Family." *The Journal of Human Resources* 1117-9174R1

Lee, James, Robbee Wedow, … , and David Cesarini. 2018. "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." *Nature Genetics* 50: 1112-1121

Lindahl, Lena. 2011. "A Comparison of Family and Neighborhood Effects on Grades, Test Scores, Educational Attainment and Income—Evidence from Sweden." *The Journal of Economic Inequality* 9(2): 207-226.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60(3): 531-42.

________. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14 (3): 115-136.

Nicoletti, Cheti, and Birgitta Rabe. 2013. “Inequality in Pupils’ Test Scores: How Much Do Family, Sibling Type and Neighbourhood Matter?” *Economica* 80(3):197–218.

________. 2019. “Sibling Spillover Effects in School Achievement.” *Journal of Applied Econometrics* 34:482–501

Nielsen, François, and J. Micha Roos. 2015. “Genetics of Educational Attainment and the Persistence of Privilege at the Turn of the 21st Century.” *Social Forces* 94(2): 535–561.

Okbay, Aysu, Jonathan Beauchamp, …., and Daniel Benjamin. 2016. “Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533, 539–542.

Okbay Aysu, Patrick Turley, Daniel Benjamin, Peter Visscher, David Braudt, and Kathleen Mullan Harris. 2018. “SSGAC Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health).” doi:10.17615/C6166F

Paloyo, Alfredo. 2020. “Peer Effects in Education: Recent Empirical Evidence.” In *Economics of Education: A Comprehensive Review*, ed. Steve Bradley and Colin Green, 291–305. Elsevier Academic Press.

Price, Alkes, Nick Patterson, Robert Plenge, Michael Weinblatt, Nancy Shadick, and David Reich. 2006. “Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies.” *Nature Genetic* 38: 904-909

Qureshi, Javaeria, 2017. “Siblings, Teachers and Spillovers on Academic Achievement.” *Journal of Human Resources* 53(1): 272-297.

Raaum, Oddbjørn. Kjell G. Salvanes, and Erik Ø. Sørensen. 2006. “The Neighbourhood is Not What it used to be.” *The Economic Journal*, 116(508): 200-222.

Reardon, Sean. 2011. “The Widening Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. In Whither Opportunity?”, ed. Greg Duncan and Richard Murnane, 91-115. New York: Russell Sage Foundation.

Silventoinen, Karri, Robert Krueger, Thomas Bouchard Jr., Jaakko Kaprio, and Matt McGue. 2004. “Heritability of Body Height and Educational Attainment in an International Context: Comparison of Adult Twins in Minnesota and Finland.” *American Journal of Human Biology* 16(5):544-5

Solon, Gary, Marianne Page, and Greg Duncan. 2000. “Correlation between Neighboring Children in their Subsequent Educational Attainment.” *The Review of Economics and Statistics* 82(3): 383-392.

Wickrama , K. A. S., and Catherine Walker O'Neal. 2013. “Family of Origin, Race/Ethnicity, and Socioeconomic Attainment: Genotype and Intraindividual Processes.” *Journal of Marriage and Family* 75 (1): 75-90.

Wolf, Jason, Edmund Brodie, James Cheverud, Allen Moore, and Michael Wade. 1998. “Evolutionary Consequences of Indirect Genetic Effects.” *Trends in Ecology & Evolution* 13(2): 64-69.

Yi, Junjian, James Heckman, Junsen Zhang, and Gabriella Conti. 2015. “Early Health Shocks, Intra-household Resource Allocation and Child Outcomes.” *The Economic Journal* 125(588): F347-F371.

TABLES

Table 1: Summary statistics

	Mean	S.D.	Min.	Max.
Outcome of Interest				
At least High School Completion	0.921	0.270	0	1
At least BA Completion	0.342	0.475	0	1
Graduate School Completion	0.120	0.325	0	1
Highest Education Completed	14.041	2.334	8	21
Ego Characteristics				
Polygenic Score (PGS) for Education (MTAG)	0.294	0.160	-0.165	0.669
Polygenic Score (PGS) for Education (GWAS)	0.202	0.146	-0.243	0.607
Male	0.478	0.499	0	1
Age	28.48	1.694	25	32
African-American	0.002	0.039	0	1
Hispanic	0.003	0.056	0	1
Other Race	0.013	0.112	0	1
White	0.979	0.142	0	1
Married	0.582	0.494	0	1
Working Part-time	0.226	0.419	0	1
Working Full-time	0.774	0.419	0	1
Live with Parents	0.106	0.308	0	1
Alter Characteristics				
Polygenic Score (PGS) for Education (MTAG)	0.295	0.161	-0.165	0.669
Polygenic Score (PGS) for Education (GWAS)	0.203	0.146	-0.243	0.607
Age Difference (months)	27.378	10.993	11	63
Same-sex Sibling	0.571	0.495	0	1
Household Characteristics				
20K<Income<39K	0.180	0.385	0	1
40K<Income<49K	0.147	0.355	0	1
50K<Income<75K	0.278	0.449	0	1
75K<Income<99K	0.144	0.351	0	1
100K<Income	0.139	0.346	0	1
Parents: less than HS	0.073	0.260	0	1
Parents: HS grad	0.318	0.466	0	1
Parents: Less than College	0.304	0.460	0	1
Parents: College Grad	0.214	0.410	0	1

Data: Add Health Wave IV. There are categories for “missing” for income and parents’ education, which are included in the models but are not shown here.

Table 2. The Effects of Sibling's PGS for Education on Ego's Education Attainment

	(1)	(2)	(3)	(4)
VARIABLES	Outcome: BA or more (Mean: 34.2%)			
	PGS based on MTAG		PGS based on GWAS	
Ego's PGS for Education	-0.715***	-0.602***	-0.808***	-0.684***
	[0.112]	[0.113]	[0.125]	[0.127]
Alter's PGS for Education	-0.367***	-0.244**	-0.379***	-0.244*
	[0.119]	[0.119]	[0.131]	[0.132]
Constant	0.016	-0.215	-0.033	-0.250
	[0.313]	[0.313]	[0.310]	[0.311]
Parents' Education	N	Y	N	Y
Observations	632	632	632	632
R-squared	0.256	0.311	0.257	0.311

Data: Add Health Wave IV, Full Siblings only

Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

Table 3. Regression Results using Different Education Measures

A. Outcome: Highest Education Completed

	(1)	(2)	(3)	(4)
	Outcome: Highest Educ completed (Mean: 14.04 years)			
	PGS based on MTAG		PGS based on GWAS	
Ego’s PGS for Education	-3.730***	-3.090***	-4.178***	-3.482***
	[0.590]	[0.585]	[0.636]	[0.6410]
Alter’s PGS for Education	-2.311***	-1.600***	-2.511***	-1.740***
	[0.582]	[0.574]	[0.633]	[0.625]
Constant	11.797***	9.847***	11.525***	9.664***
	[1.446]	[1.464]	[1.431]	[1.447]
Parents' Education	N	Y	N	Y
Observations	632	632	632	632
R-squared	0.299	0.371	0.301	0.373

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego’s age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for Highest Education completed outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

B. Outcome: At Least High School Completion

	(1)	(2)	(3)	(4)
	Outcome: At least high school completed (Mean: 92.09%)			
	PGS based on MTAG		PGS based on GWAS	
Ego's PGS for Education	-0.098	-0.053	-0.105	-0.059
	[0.099]	[0.099]	[0.109]	[0.110]
Alter's PGS for Education	-0.207**	-0.153*	-0.252**	-0.196*
	[0.093]	[0.091]	[0.105]	[0.103]
Constant	0.752***	0.475**	0.739***	0.471**
	[0.189]	[0.209]	[0.188]	[0.207]
Parents' Education	N	Y	N	Y
Observations	632	632	632	632
R-squared	0.092	0.145	0.096	0.148

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for at least high school completion outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

C. Outcome: Graduate School Completion

	(1)	(2)	(3)	(4)
	Outcome: At least graduate school completed (Mean: 12.03%)			
	PGS based on MTAG		PGS based on GWAS	
Ego's PGS for Education	-0.365***	-0.298***	-0.415***	-0.338***
	[0.097]	[0.093]	[0.104]	[0.099]
Alter's PGS for Education	-0.210**	-0.135	-0.199**	-0.114
	[0.086]	[0.087]	[0.096]	[0.098]
Constant	-0.12	-0.236	-0.151	-0.261
	[0.236]	[0.242]	[0.236]	[0.241]
Parents' Education	N	Y	N	Y
Observations	632	632	632	632
R-squared	0.182	0.227	0.179	0.225

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for at least high school completion outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

Table 4. The Effects of Sibling's PGS for Education on Ego's Education Attainment – Controlling for Age/Sex Difference and Interaction Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Outcome: BA or more (Mean: 34.2%)							
	PGS based on MTAG				PGS based on GWAS			
Ego's PGS for Education	-0.715***	-0.701***	-0.602***	-0.591***	-0.808***	-0.803***	-0.684***	-0.681***
	[0.112]	[0.114]	[0.113]	[0.113]	[0.125]	[0.126]	[0.127]	[0.128]
Alter's PGS for Education	-0.367***	-0.827***	-0.244**	-0.810***	-0.379***	-0.988***	-0.244*	-0.943***
	[0.119]	[0.313]	[0.120]	[0.291]	[0.131]	[0.347]	[0.132]	[0.324]
Age difference		-0.002		-0.004		-0.001		-0.002
		[0.003]		[0.003]		[0.003]		[0.003]
Same sex sibling		-0.052		-0.066		-0.064		-0.073
		[0.0812]		[0.076]		[0.067]		[0.063]
Alter's PGS * Same Sex		0.198		0.185		0.347		0.305
		[0.227]		[0.218]		[0.246]		[0.240]
Alter's PGS * Age Diff		0.014		0.018*		0.016		0.020*
		[0.010]		[0.010]		[0.011]		[0.011]
Constant	0.0162	0.1135	-0.2145	-0.0695	-0.0332	0.0709	-0.2502	-0.112
	[0.3127]	[0.3328]	[0.3134]	[0.3234]	[0.3099]	[0.3210]	[0.3105]	[0.314]
Parents' Educ	N	N	Y	Y	N	N	Y	Y
Observations	632	632	632	632	632	632	632	632
R-squared	0.256	0.263	0.311	0.317	0.257	0.266	0.311	0.320

Data: Add Health Wave IV, Full Siblings only

Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

Table 5. Regression Results for Subgroup Analysis, using MTAG based PGS

A. When Ego is old vs. When Ego is young

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	ego=Older		ego=Younger	
Mean of DV	36.59%		31.75%	
Ego's PGS for Education	-0.7792***	-0.6940***	-0.6539***	-0.5076***
	[0.1782]	[0.1738]	[0.1637]	[0.1542]
Alter's PGS for Education	-0.2834	-0.1659	-0.4069**	-0.2982*
	[0.1859]	[0.1803]	[0.1755]	[0.1659]
Age Difference (in months)	0.0017	0.0012	0.0020	0.0016
	[0.0026]	[0.0025]	[0.0028]	[0.0027]
Same Sex Sibling	0.0099	-0.0072	0.0248	0.0107
	[0.0537]	[0.0518]	[0.0508]	[0.0496]
Parent's Education	N	Y	N	Y
Observations	317	317	315	315
R-squared	0.284	0.327	0.293	0.359

Data: Add Health Wave IV, Full Siblings only

Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

B. Same-sex Sibling vs. Different-sex Sibling

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	Same-sex Siblings		Different-sex Siblings	
Mean of DV	33.24%		35.42%	
Ego's PGS for Education	-0.7150***	-0.6385***	-0.7019***	-0.5502***
	[0.1429]	[0.1428]	[0.1832]	[0.1811]
Alter's PGS for Education	-0.2059	-0.1122	-0.5582***	-0.3846*
	[0.1648]	[0.1640]	[0.1840]	[0.1955]
Age Difference (in months)	0.0069***	0.0060**	-0.0035	-0.0041
	[0.0025]	[0.0024]	[0.0028]	[0.0026]
Parent's Education	N	Y	N	Y
Observations	361	361	271	271
R-squared	0.284	0.331	0.293	0.364

Data: Add Health Wave IV, Full Siblings only

Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

C. When Ego is Female vs. When Ego is Male

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	ego=Female		ego=Male	
Mean of DV	36.97%		31.12%	
Ego's PGS for Education	-0.6911***	-0.5444***	-0.7837***	-0.7178***
	[0.1589]	[0.1682]	[0.1750]	[0.1571]
Alter's PGS for Education	-0.4215**	-0.2681	-0.2580	-0.1623
	[0.1779]	[0.1816]	[0.1750]	[0.1637]
Age Difference (in months)	0.0049*	0.0053**	-0.0007	-0.0024
	[0.0028]	[0.0026]	[0.0027]	[0.0026]
Same Sex Sibling	-0.0169	-0.0332	0.0437	0.0314
	[0.0554]	[0.0534]	[0.0544]	[0.0527]
Parent's Education	N	Y	N	Y
Observations	330	330	302	302
R-squared	0.300	0.348	0.287	0.351

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

D. When Ego's PGS for Education>Avg vs. When Ego's PGS for Education<Avg

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	Ego's PGS>Average		Ego's PGS<Average	
Mean of DV	21.77%		46.67%	
Ego's PGS for Education	-0.8299***	-0.7558***	-0.8193***	-0.5520*
	[0.2264]	[0.2238]	[0.2842]	[0.2880]
Alter's PGS for Education	-0.1085	-0.0415	-0.5875***	-0.3752*
	[0.1623]	[0.1592]	[0.2064]	[0.2076]
Age Difference (in months)	0.0061**	0.0061**	-0.0014	-0.0037
	[0.0027]	[0.0026]	[0.0026]	[0.0025]
Same Sex Sibling	-0.0247	-0.0308	0.0164	-0.0180
	[0.0503]	[0.0482]	[0.0589]	[0.0539]
Parent's Education	N	Y	N	Y
Observations	317	317	315	315
R-squared	0.225	0.259	0.252	0.327

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

Table 6. Regression Results for Subgroup Analysis, using GWAS based PGS

A. When Ego is old vs. When Ego is young

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	ego=Older		ego=Younger	
Mean of DV	36.59%		31.75%	
Ego's PGS for Education	-0.8654***	-0.7928***	-0.7407***	-0.5560***
	[0.1914]	[0.1860]	[0.1888]	[0.1790]
Alter's PGS for Education	-0.2729	-0.1187	-0.4204**	-0.3250*
	[0.2097]	[0.1982]	[0.1857]	[0.1751]
Age Difference (in months)	0.0019	0.0013	0.0021	0.0016
	[0.0026]	[0.0025]	[0.0028]	[0.0027]
Same Sex Sibling	0.0078	-0.0098	0.0231	0.0101
	[0.0538]	[0.0517]	[0.0505]	[0.0495]
Parent's Education	N	Y	N	Y
Observations	317	317	315	315
R-squared	0.281	0.327	0.293	0.359

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

B. Same-sex Sibling vs. Different-sex Sibling

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	Same-sex Siblings		Different-sex Siblings	
Mean of DV	33.24%		35.42%	
Ego's PGS for Education	-0.7885***	-0.7065***	-0.8359***	-0.6538***
	[0.1619]	[0.1626]	[0.1919]	[0.2005]
Alter's PGS for Education	-0.1452	-0.0539	-0.6404***	-0.4444**
	[0.1790]	[0.1774]	[0.1981]	[0.2144]
Age Difference (in months)	0.0070***	0.0060**	-0.0035	-0.0039
	[0.0025]	[0.0024]	[0.0028]	[0.0026]
Parent's Education	N	Y	N	Y
Observations	361	361	271	271
R-squared	0.280	0.330	0.304	0.369

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

C. When Ego is Female vs. When Ego is Male

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	ego=Female		ego=Male	
Mean of DV	36.97%		31.12%	
Ego's PGS for Education	-0.7394***	-0.5621***	-0.8982***	-0.8405***
	[0.1818]	[0.1944]	[0.1849]	[0.1682]
Alter's PGS for Education	-0.4921**	-0.3466*	-0.2058	-0.0620
	[0.1914]	[0.1943]	[0.1897]	[0.1729]
Age Difference (in months)	0.0050*	0.0054**	-0.0005	-0.0022
	[0.0029]	[0.0026]	[0.0027]	[0.0026]
Same Sex Sibling	-0.0165	-0.0324	0.0373	0.0255
	[0.0551]	[0.0531]	[0.0546]	[0.0531]
Parent's Education	N	Y	N	Y
Observations	330	330	302	302
R-squared	0.299	0.348	0.288	0.352

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1

D. When Ego's PGS for Education>Avg vs. When Ego's PGS for Education<Avg

	(1)	(2)	(3)	(4)
	Outcome: BA or More			
	Ego's PGS>Average		Ego's PGS<Average	
Mean of DV	21.77%		46.67%	
Ego's PGS for Education	-0.6826***	-0.6199***	-0.9520***	-0.6865**
	[0.2241]	[0.2256]	[0.2756]	[0.2790]
Alter's PGS for Education	-0.0870	-0.0068	-0.6012***	-0.3859*
	[0.1764]	[0.1725]	[0.2227]	[0.2228]
Age Difference (in months)	0.0065**	0.0064**	-0.0014	-0.0035
	[0.0028]	[0.0027]	[0.0027]	[0.0026]
Same Sex Sibling	-0.0209	-0.0280	0.0057	-0.0242
	[0.0504]	[0.0484]	[0.0585]	[0.0536]
Parent's Education	N	Y	N	Y
Observations	317	317	315	315
R-squared	0.214	0.250	0.261	0.333

Data: Add Health Wave IV, Full Siblings only
Notes: Robust standard errors in brackets, clustered at the family level. First 10 principal components are included in the model. Demographic variables controlled for are ego's age, sex, race, marital status, whether work full time/part time, whether live with parents, and household income categories. The reported number for BA or more outcome is OLS coefficient. *** p<0.01, ** p<0.05, * p<0.1