Separating Measured Genetic and Environmental Effects: Evidence Linking Parental Genotype and Adopted Child Outcomes

Benjamin W. Domingue[1]

Jason Fletcher[2]

Acknowledgements

This research uses data from the Wisconsin Longitudinal Study (WLS) of the University of Wisconsin-Madison. Since 1991, the WLS has been supported principally by the National Institute on Aging (AG-9775, AG-21079, AG-033285, and AG-041868), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin-Madison. Since 1992, data have been collected by the University of Wisconsin Survey Center. A public use file of data from the Wisconsin Longitudinal Study is available from the Wisconsin Longitudinal Study, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 and at http://www.ssc.wisc.edu/wlsresearch/data/. The opinions expressed herein are those of the authors. This research was supported by core grants to the Center for Demography and Ecology at the University of Wisconsin-Madison (P2C HD047873) and to the Center for Demography of Health and Aging at the University of Wisconsin-Madison (P30 AG017266) and by award 96-17-04 from the Russell Sage Foundation and the Ford Foundation. The authors thank Ron Lee and participants at the Population Association of America 2018 conference for helpful suggestions.

[1] Graduate school of Education, Stanford University; bdomingue@stanford.edu

[2] La Follette School of Public Affairs, Department of Sociology, and Center for Demography of Health and Aging, University of Wisconsin-Madison

Abstract

There has been widespread adoption of genome wide summary scores (polygenic scores) as tools for studying the importance of genetics and associated lifecourse mechanisms across a range of demographic and socioeconomic outcomes. However, an often unacknowledged issue with these studies is that parental genetics impact both child environments and child genetics, leaving the effects of polygenic scores difficult to interpret. This paper uses multi-generational data containing polygenic scores for parents (n=7,193) and educational outcomes for adopted (n=855) and biological (n=20,939) children, many raised in the same families, which allows us to separate the influence of parental polygenic scores on children outcomes between environmental (adopted children) and environmental and genetic (biological children) effects. Our results complement recent work on "genetic nurture" by showing associations of parental polygenic scores with adopted children's schooling, providing additional evidence that polygenic scores combine genetic and environmental influences and that research designs are needed to separate these estimated impacts.

Key Words

Indirect genetic effects, Polygenic Score, Educational Attainment, Wisconsin Longitudinal Study, Adoption Study

1. Introduction

This paper focuses on the issue of separating genetic and environmental influences in important adult socioeconomic outcomes. The rapid advancement in the collection, measurement, and understanding of genetic data has led to new genetic discoveries linking specific genetic variants with health outcomes as well as schooling outcomes (1–3). While, individually, these variants do not have large effects (4), the development of genome wide summary scores (polygenic scores, PGS) offers the possibility of a genetically-based analytic tool that is predictive (5) of important adult outcomes, including health (6) and disease outcomes (7) but also behavioral outcomes such as schooling (8).

Many studies have begun to examine the characteristics of a PGS constructed to predict years of schooling (i.e., educational attainment). Studies have examined associations between the educational attainment PGS and outcomes of interest at different stages of the lifecourse including childhood and adolescence (9–11), middle age (9,12,13), and later-life (14,15). Other studies have examined the impact of the educational attainment PGS on economic attainments such as earnings and wealth (16–18). Still others have used this PGS to examine gene-environment interactions (19–22). While these studies have begun to offer novel insights into human behavior, they also raise questions about how to interpret associations between this polygenic score and the various outcomes in question.

The issue of how to interpret polygenic score associations with phenotypes has been repeatedly raised by scholars, especially when it comes to educational attainment (23). PGS are created by summing up results of GWAS effect estimates across the genome; however, the acknowledgement that GWAS, as a research design, is not adequate to separate causal and non-

causal genetic variants when estimating effects leads to the implication that PGS measures capture both causal and non-causal genetic effects on the phenotype.

As a first exploration of this issue, researchers have examined the associations between PGS and phenotype with a sibling design—leveraging differences in PGS between biological siblings as a means to attempt to separate causal and non-causal components of the PGS. Specifically, analysis of siblings is powerful as it minimizes the possibility for confounding due to factors that do not vary within-family. Early evidence (19,23,24) suggested that the PGS for educational attainment is indeed capturing a crucial signal of individual difference in the sense that the sibling with the higher PGS tended to also have more schooling. However, more recent work based on siblings suggests that associations between PGS and educational attainment are substantially smaller when drawn from family-based samples as compared to association estimates drawn from unrelated individuals (25). This suggest the possibility of bias—in particular, upwards bias in PGS estimates—when non-family samples are analyzed (26).

Relatedly, newer work has begun to produce evidence of “social genetic effects”, where the genetics of a focal individual’s social connections could play a role in the focal individual’s outcomes (27–29). Social genetic effects may, in many cases, be phenotypic effects in action—they may be mediated by observable phenotypes of an individual’s social peers—but, to the extent that they exist, they may lead to conceptual and methodological difficulties by adding a new layer of confounding to the PGS. The effects of polygenic scores would then need to be interpreted accordingly (e.g., they capture processes that are not within-person) and even sibling difference models may suffer from stable unit treatment value assumption (SUTVA) violations that can produce results that are biased in either direction (30).

A particularly important potential source of confounding with respect to the educational attainment polygenic score is the observation that variation in the parental genotype associated with educational attainment may also be associated with differences in the environments that parents create for their children. Were this hypothesis true, parental genetics would have implications for both child environments and child genetics, offering an additional reason the estimated effects of polygenic scores would be difficult to interpret as purely within-person genetic effects. Indeed, emerging work is beginning to examine the impacts of parental genotype (either net of child genotype or focusing on the non-transmitted portion) on children's outcomes, termed "genetic nurture" (31–35). Note that while our focus on educational attainment is motivated primarily by its centrality to many social science questions, the findings of Kong et al. (31) suggest that educational attainment may be a phenotype that is especially impacted by genetic nurture.

Existing estimates of genetic nurture may understate the total contribution of family background to child's phenotype. Kong et al. (31) partitions the parental genetic effects on offspring phenotype into transmitted and non-transmitted alleles. A PGS created from the non-transmitted alleles are the "indirect genetic effects" (or genetic nurture) and a PGS created from the transmitted alleles are the "direct genetic effects" in this framework. However, this partitioning exercise does not account for the fact that the transmitted alleles can affect children both directly and indirectly through parentally provided environments. In this paper, we introduce an alternative assessment that does not partition between transmitted and non-transmitted allele and instead estimates the "genetic nurture" effect via a different design.

This paper offers novel evidence on the magnitude of genetic nurture effects by using multi-generational data that collected polygenic scores from 7,200 parents and outcomes for

22,000 of their adopted and biological children, which allows us to separate the influence of parental polygenic scores on children outcomes between environmental (i.e. genetic nurture) (adopted children) and environmental and genetic (biological children) effects. We do this by focusing on a test of whether the polygenic score of a parent predicts the educational attainment of adopted offspring (to whom they have not transmitted their genes).

2. Methods

2A. Data

The Wisconsin Longitudinal Study (WLS) is a long-term study of over 10,000 men and women based on a random 1/3 sample of the graduating high school class of 1957 in Wisconsin (36). These individuals continue to be followed through the present. Alongside the respondent who graduated in 1957, siblings were also impaneled (we refer to “graduate” and “sibling” respondents throughout to distinguish between the two types). The WLS provides an opportunity to study the life course, relationships, family functioning, physical and mental health and well-being, and morbidity and mortality from late adolescence through 2008. WLS data also cover social background, youthful aspirations, schooling, military service, labor market experiences, family characteristics and events, social participation, psychological characteristics, and retirement. Further, in the latest complete round of survey collection, the respondents have provided biological specimens for DNA analysis. The WLS now contains genome-wide DNA information for many of their respondents.

The WLS also asks respondents about their offspring. In addition to the graduate and sibling respondents reporting information on their own lives, they also report information on all their surviving children, including both the relationships with each child and information about

that child's education. For each child, the respondent was asked their relationship with that child (possible choices: biological, adopted, step, foster, legal ward, niece/nephew, other non-relative, child of lover/partner). Given complexities associated with these different relationships (e.g., assortative mating may induce correlations between polygenic scores of parents and children of their partners, environments may vary across these different familial relationships), we focus on adopted (n=855) and biological children (n=20,939) in remaining analyses.

In total, we observe 21,794 parent/child pairs for 7,193 WLS parents who have been genotyped. We note two points about this sample. First, the WLS parents are clustered into 5,695 families (1493 families have two WLS parents and two families contain three and four WLS parents respectively). This clustering has implications for statistical inference as parents within the same family are not independent (e.g., they grew up in a similar environment); we cluster standard errors accordingly as discussed below. Second, our analytic sample (N=7,193 WLS parents) shows small differences on key demographics from the full WLS sample (N=19,046). Focusing on the graduate respondents in each sample, graduates in the analytic sample were slightly younger (they were born about a week later; the difference was significant, p=0.009) and slightly less likely to be male (47% of our graduates were male as compared to 48% in the full sample; the difference was not significant). However, our analytic sample had a higher concentration of graduates; in the full WLS sample, 54% of the respondents are graduates while in our analytic sample 65% of the respondents are graduates.

2B. Measures

Parent Polygenic Score

We use the educational attainment polygenic score constructed by WLS and the SSGAC following the completion of the third generation educational attainment GWAS (8). Complete details regarding score construction are available elsewhere (37); we offer a summary of key points here. As WLS respondents were used in the main educational attainment GWAS, GWAS estimates were re-computed without the WLS respondents. These estimates were then used in combination with LDpred (38) to create a polygenic score. Given issues associated with polygenic prediction in diverse samples (39,40), scores were constructed only for respondents of European ancestry. We residualize the score based on the first ten principal components and then standardized to have mean=0 and SD=1 in this ancestrally homogeneous subsample (N=8,509).

Parent Education

Parental education is compiled via the collection of responses given by WLS respondents to education-related questions in the various waves of WLS. In particular, as in both the original GWAS (8) and in other work with this data (18,20), we use the maximum number of years of education reported by a WLS respondent over the different waves of data collection.

Child education

Offspring educational attainment was collected in 2004 when WLS respondents were asked versions of the question "What is the highest grade of regular school your child ever attended?" for all of their children. We used the coding of the WLS variable (e.g., https://www.ssc.wisc.edu/wlsresearch/documentation/waves/?wave=grad2k&module=gkid) which was meant to capture the years of schooling (e.g., a high school graduate is coded as having 12 years of education; note that we coded those with GEDs as having 11 years of education).

Table 1 presents descriptive statistics for the analytic sample and stratifies these descriptive statistics by child status (biological or adopted). Nearly two thirds of the children have parents who are graduate respondents (as opposed to siblings of WLS graduates). The graduate respondents were born in a narrow window around 1939 (recall that they are the graduating class of 1957); there is substantially more variation in the birthyears of the sibling respondents (SD for graduates=0.52, SD for siblings=6.8). By construction, the education polygenic score of parents is centered at 0; however, note that the polygenic scores of parents who have adopted children are higher than those who have biological children. Parents of adopted children also have higher educational attainments. The birth years of the children are centered in the mid-1960s although, as expected, the birth years of adopted children are later. Children's educational attainment is over 14 years on average, although this figure is lower for adopted children.

2C. Analysis

So as to evaluate differential prediction of the PGS between biological and adopted children, we focus on linear models that contain interactions between the parental PGS and the child status (biological or adopted). For child j of parent i, we focus on linear models of the form

$$\text{Child Education}_{ij}=b_0+b_1\text{PGS}_i+b_2\text{Biological}_{ij}+b_3\text{Biological}_{ij}\text{PGS}_i+\text{BX}_{ij}+e_{ij}.$$

Our focal coefficient is b_1 which captures the association of parent PGS with adopted offspring educational attainment. Our key hypothesis is that $b_1>0$, indicating that the parental PGS predicts the educational attainment of their adopted children. Coefficients b_2 and b_3 describe differences between biological and adopted children. The b_2 coefficient describes differences in schooling between biological and adopted children that may occur if there are systematic differences in

environments associated with adoption status (e.g., parents construct different rearing environments for biological versus adopted children, the challenges faced by adoptees predating placement for adoption may have effects on their later attainments, etc.). Given the fact that biological children tend to have more years of education than adopted children, we anticipate $b_2>0$. The b_3 coefficient, which we expect to be positive given that it will also capture part of the effect of the genetics of biological children directly on their phenotype, will inform us about differences in the association between parental PGS and offspring educational attainment for biological versus adopted siblings. X is a matrix of controls including child sex and age (birth year); previous work (41) has shown the need for interactions in such models and we include those terms as needed. Standard errors are clustered—in particular, we use sandwich techniques (42) to obtain cluster-robust standard errors (43)—based on the family structuring of WLS participants (graduates plus siblings).

3. Results

Amongst biological children, the parental PGS is a robust predictor of child attainment. A one SD increase in the parental PGS predicts nearly an extra half year of schooling for a biological child (Table 2; M3). This is consistent with a large amount of previous work using a similar polygenic score (18,20,28). We next considered a test of the basic question: does the parental PGS predict educational attainment for non-biological offspring in the form of adopted children? In a baseline analysis, we considered a model (M0) containing main effects and interactions between parent PGS and child status. While the parental PGS is more predictive of the educational attainment of biological offspring, it also predicts variation in the variation of educational attainment in adopted children. A one standard deviation increase in PGS is

associated with roughly an extra quarter year of schooling in adopted children (b=0.253, p=0.025). Results are summarized in Figure 1; the fact that the parent PGS is more predictive of educational attainments for biological children (i.e., $b_3>0$) but that it also predicts variation in outcomes for adopted children (i.e., $b_1>0$) is readily observed. Results were similar when we added controls for birthyear. We obtained a slightly attenuated result when we consider a model saturated with main effects and interactions of child gender (M2). We also obtained comparable results when we focus on models stratified by child status (M3 & M4).

One issue with these comparisons between families who adopt and those who do not is our finding that parental PGS is higher among adoptive parents (Table 1), which suggests the possibility of other family-level differences between families with adopted versus biological children. As a sensitivity test, we also conducted analysis using only those families who have both adopted and biological children (44). Table 3 reports results for families with both adoptive and biological children. Our baseline results continue to support the idea that parental PGS predicts adoptee educational attainment (although, we note the result becomes insignificant in the saturated model). Note that even in these families with both biological and adopted children, we also find that biological children complete nearly three quarters of a year of schooling more on average compared with adopted children.

4. Discussion

A large and growing literature leverages the rapid and important advances in genotyping and prediction using genetics and the associated creation of polygenic scores to test novel hypotheses about the links between genetics and phenotypes of interest. A key component of this literature focuses on estimating statistical associations between individual genotype and

phenotype with a parsimonious set of controls based on the reasoning that genetic variation precedes all developmental processes and is thus independent of environmental feedback.

A limitation of this approach is that, due to the nature of meiosis, a child's genotype is highly correlated with the genotype of their biological parent. By extension, the child's genotype may also be correlated with environmental inputs during childhood (to the extent that the parental genotype is correlated with such inputs). Children with a high PGS for education are likely to have grown up in households with highly educated parents and both influences—that of the child's genetics and that of their environments—likely contribute to the children's educational attainments. Children with high PGS for education will experience differences in their surrounding environments—neighborhoods (23,45), for example—compared to children with a lower PGS.

In an attempt to separate children's genetic effects from parental effects, earlier studies (31,32) used the fact that half of each parent's alleles are not transmitted to children. If GWAS findings related to educational attainment were due to entirely direct (i.e., within-person) genetic effects, then the non transmitted alleles should contain no predictive information about the biological offspring's educational attainment, net of the directly transmitted alleles. Yet this was not the case; a sizeable portion of the "genetic effect" on children's schooling was estimated to be due to indirect genetic effects. In contrast, parallel findings for other phenotypes suggested a much smaller role for indirect genetic effects (31).

Our work is motivated by a different study design meant to capture a broader measure of parental effects. Building on a tradition of earlier work in both behavior genetics (46) and economics (47–49) that used adopted children to separate child genotype and parental genotype/environment, we test whether the polygenic score of parents predicts the educational

attainment of their adopted offspring. We utilize the fact that, given the absence of direct genetic inheritance, there is no mechanical (i.e. transmitted) correlation between the polygenic score of a parent and adopted child. We find that a one-SD increase in the PGS of a parent predicts roughly 0.25 years of schooling for an adopted offspring as compared to roughly 0.5 years of schooling for a biological offspring. Our findings are consistent with those that utilized the design based on non-transmitted alleles (31,32) in suggesting that the educational attainment polygenic score is capturing both direct and indirect genetic (i.e. parental) effects.

While our results are the first to use our proposed research design, we acknowledge several limitations. The WLS is representative of the high school graduating class of 1957 in Wisconsin and a selected sibling. While Wisconsin is broadly representative of the US during this period (36), we acknowledge that our results may not be generalizable to other populations or cohorts. This limitation would be larger to the extent that individuals in Wisconsin during this period had different adoption patterns than individuals not represented in this sample. This limitation is countered by the recognition that very few datasets contain the information we require for our analysis, including two generations of family members, information on adoptive status and completed schooling in the children's generation, and genetic information. A second limitation is that we have incomplete information about the adoption process for the families in our data, including a lack of information about selectivity in the adoption process and at what age the adoptee joined the adoptive family. Results from Table 3, wherein we focus only on families that choose to adopt, are meant to ensure that our results are not explained by selection along these lines. To the extent that adopted children begin living with their adoptive parents at later ages in our sample (compared with at birth), this would likely lead our estimates to be

relatively conservative estimates of the association between parental PGS on the adopted children's outcomes.

Our findings suggest the standard approach to linking genotype-phenotype using polygenic scores may inadvertently include the influence of both indirect and direct genetic effects. An implication of our results, that parental genotype is associated with *adopted* children's phenotypes, is that estimates of the associations between own-genotype on own-phenotypes are not strictly capturing direct genetic effects but instead are confounded by family background. The magnitude of this confounding—labeled "genetic nurture" in recent studies—may well vary by phenotype. Educational attainment is a phenotype with a relatively large amount of confounding by genetic nurture (31).

While other studies have produced evidence of the existence of confounding of polygenic scores, the data requirements are restrictive. For example, Kong et al. (31) required trio data from families in the Icelandic population to generate evidence of genetic nurture. Our results add a new test of this confounding by making use of adoptive children and parental PGS data. While each approach has limitations, we think that the consistent findings from both approaches are instructive. Our results thus add to the accumulating evidence that PGS effects should be interpreted with caution and are not pure "genetic" effects for educational attainment and likely other phenotypes. This suggests that since polygenic scores, as currently constructed, combine genetic and environmental influences, new research designs—perhaps including the use of within family studies at the GWAS stage (50) or attempts to estimate structural parameters associated with genetic nurture (26)—will need to be used to separate these estimated impacts.

Compliance with Ethical Standards

-Conflict of Interest: The authors declare that they have no conflict of interest.

-Informed consent: Informed consent was obtained from all individual participants included in the study.

-Ethical Approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Stanford IRB #39751 & University of Wisconsin Madison Minimal Risk IRB 2013-1477) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards

References

1. Conley D, Fletcher J. The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future. Princeton: Princeton University Press; 2017.

2. Cesarini D, Visscher PM. Genetics and educational attainment. npj Science of Learning [Internet]. 2017 Dec [cited 2018 Sep 25];2(1). Available from: http://www.nature.com/articles/s41539-017-0005-6

3. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics. 2017 Jul 6;101(1):5–22.

4. Chabris CF, Lee JJ, Cesarini D, Benjamin DJ, Laibson DI. The Fourth Law of Behavior Genetics. Curr Dir Psychol Sci. 2015 Aug 1;24(4):304–12.

5. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9(3):e1003348.

6. Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019 Apr;177(3):587-596.e9.

7. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics. 2018;1.

8. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018 Aug;50(8):1112–21.

9. Belsky DW, Moffitt T, Corcoran D, Domingue B, Harrington H, Hogan S, et al. The genetics of success: How SNPs associated with educational attainment relate to life course development. Psychological Science. 2016;

10. de Zeeuw EL, van Beijsterveldt CE, Glasner TJ, Bartels M, Ehli EA, Davies GE, et al. Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2014;165(6):510–520.

11. Selzam S, Krapohl E, von Stumm S, O'Reilly PF, Rimfeld K, Kovas Y, et al. Predicting educational achievement from DNA. Mol Psychiatry [Internet]. 2016 Jul 19; Available from: http://dx.doi.org/10.1038/mp.2016.107

12. Wedow R, Zacher M, Huibregtse B, Harris KM, Domingue B, Boardman J. Education, Smoking, and Cohort Change: Forwarding a Multidimensional Theory of the Environmental Moderation of Genetic Effects. American Sociological Review. In press;

13. Mõttus R, Realo A, Vainik U, Allik J, Esko T. Educational attainment and personality are genetically intertwined. Psychological science. 2017;28(11):1631–1639.

14. Ding X, Barban N, Tropf F, Mills M. The relationship between cognitive decline and a genetic predictor of educational attainment. Social Science & Medicine. 2019;112549.

15. Marioni RE, Ritchie SJ, Joshi PK, Hagenaars SP, Okbay A, Fischer K, et al. Genetic variants linked to education predict longevity. PNAS. 2016 Oct 31;201605334.

16. Barth D, Papageorge NW, Thom K. Genetic Endowments and Wealth Inequality. National Bureau of Economic Research; 2018.

17. Papageorge NW, Thom K. Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study. 2016;

18. Belsky D, Domingue B, Wedow R, Arseneault L, Boardman J, Caspi A, et al. Genetic analysis of social-class mobility in five longitudinal studies. Proceedings of the National Academy of Sciences. 2018;

19. Conley D, Domingue B, Cesarini D, Dawes C, Rietveld C, Boardman J. Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype? Sociological Science. 2015;2:82–105.

20. Trejo S, Belsky D, Boardman J, Freese J, Harris K, Herd P, et al. Schools as Moderators of Genetic Associations with Life Course Attainments: Evidence from the WLS and Add Health. Sociological Science. 2018;5:513–40.

21. Rimfeld K, Krapohl E, Trzaskowski M, Coleman JR, Selzam S, Dale PS, et al. Genetic influence on social outcomes during and after the Soviet era in Estonia. Nature Human Behaviour. 2018;1.

22. Schmitz LL, Conley D. The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes. Economics of education review. 2017;61:85–97.

23. Domingue BW, Belsky DW, Conley D, Harris KM, Boardman JD. Polygenic Influence on Educational Attainment. AERA Open. 2015;1(3).

24. Rietveld CA, Conley D, Eriksson N, Esko T, Medland SE, Vinkhuyzen AA, et al. Replicability and robustness of genome-wide-association studies for behavioral traits. Psychological science. 2014;25(11):1975–1986.

25. Selzam S, Ritchie SJ, Pingault J-B, Reynolds CA, O'Reilly PF, Plomin R. Comparing Within- and Between-Family Polygenic Score Prediction. The American Journal of Human Genetics. 2019;105(2):351–63.

26. Trejo S, Domingue BW. Genetic nature or genetic nurture? quantifying bias in analyses using polygenic scores. Biodemography and Social Biology. 2019;

27. Domingue BW, Belsky DW. The social genome: Current findings and implications for the study of human genetics. PLoS genetics. 2017;13(3):e1006615.

28. Domingue BW, Belsky DW, Fletcher JM, Conley D, Boardman JD, Harris KM. The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. PNAS. 2018 Jan 9;201711803.

29. Sotoudeh R, Harris KM, Conley D. Effects of the peer metagenomic environment on smoking behavior. Proceedings of the National Academy of Sciences. 2019;116(33):16302–16307.

30. Boardman JD, Fletcher JM. To cause or not to cause? That is the question, but identical twins might not have all of the answers. Social science & medicine (1982). 2015;127:198.

31. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. Science. 2018;359(6374):424–428.

32. Bates TC, Maher BS, Medland SE, McAloney K, Wright MJ, Hansell NK, et al. The Nature of Nurture: Using a Virtual-Parent Design to Test Parenting Effects on Children's Educational Attainment in Genotyped Families. Twin Res Hum Genet. 2018 Apr;21(2):73–83.

33. Wertz J, Belsky J, Moffitt TE, Belsky DW, Harrington H, Avinun R, et al. Genetics of nurture: A test of the hypothesis that parents' genetics predict their observed caregiving. Developmental Psychology. 2019 Jul;55(7):1461–72.

34. Wertz J, Moffitt TE, Agnew-Blais J, Arseneault L, Belsky DW, Corcoran DL, et al. Using DNA from mothers and children to study parental investment in children's educational attainment [Internet]. Genetics; 2018 Dec [cited 2019 Oct 9]. Available from: http://biorxiv.org/lookup/doi/10.1101/489781

35. Armstrong-Carter E, Trejo S, Hill L, Crossley K, Mason D, Domingue B. The Earliest Origins of Genetic Nurture: Prenatal Environment Mediates the Association Between Maternal Genetics and Child Development [Internet]. PsyArXiv; 2019 Jul [cited 2019 Oct 9]. Available from: https://osf.io/2usk8

36. Herd P, Carr D, Roan C. Cohort profile: Wisconsin longitudinal study (WLS). International journal of epidemiology. 2014;43(1):34–41.

37. Okbay A, Benjamin D, Visscher P. Documentation (Construction of WLS Polygenic Scores) [Internet]. 2018. Available from: https://www.ssc.wisc.edu/wlsresearch/documentation/GWAS/Lee_et_al_(2018)_PGS_WLS.pdf

38. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. The American Journal of Human Genetics. 2015;97(4):576–592.

39. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. The American Journal of Human Genetics. 2017;100(4):635–649.

40. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nature communications. 2019;10(1):3328.

41. Keller MC. Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. Biological Psychiatry. 2014 Jan 1;75(1):18–24.

42. Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. Journal of Statistical Software. 2004;11(10):1–17.

43. Arai M. Cluster-robust standard errors using R. Note available http://people su se. 2011;

44. Fletcher J, Wolfe B, Song J, Mailick MR, Greenberg J. My sibling is adopted. 2016; Available from: https://www.lifecoursecentre.org.au/wp-content/uploads/2018/06/2016-05-LCC-Working-Paper-Fletcher-et-al..pdf

45. Belsky DW, Caspi A, Arseneault L, Corcoran DL, Domingue BW, Harris KM, et al. Genetics and the geography of health, behaviour and attainment. Nature Human Behaviour [Internet]. 2019 Apr 8 [cited 2019 Apr 22]; Available from: http://www.nature.com/articles/s41562-019-0562-1

46. Scarr S, Weinberg RA. The Minnesota adoption studies: Genetic differences and malleability. Child Development. 1983;260–267.

47. Silles MA. The intergenerational transmission of education: New evidence from adoptions in the usa. Economica. 2017;84(336):748–778.

48. Sacerdote B. How large are the effects from changes in family environment? A study of Korean American adoptees. The Quarterly Journal of Economics. 2007;122(1):119–157.

49. Björklund A, Lindahl M, Plug E. The origins of intergenerational associations: Lessons from Swedish adoption data. The Quarterly Journal of Economics. 2006;121(3):999–1028.

50. Morris TT, Davies NM, Hemani G, Smith GD. Population phenomena inflate genetic associations of complex social traits. bioRxiv. 2019;630426.

Table 1. Summary Statistics: Wisconsin Longitudinal Study; Comparing Biological and Adopted Children

A. Child-Level Descriptives (Means and Standard Deviations, SD)

	All	Biological	Adopted
% Graduates	0.66	0.66	0.67
SD	0.47	0.47	0.47
Parent Birthyear	1939.03	1939.01	1939.55
SD	3.94	3.95	3.59
Parent PGS	-0.01	-0.01	0.13
SD	0.99	0.99	0.95
Parent Edu	13.61	13.57	14.39
SD	2.31	2.29	2.58
Child Birthyear	66.86	66.71	70.62
SD	6.46	6.40	6.87
Child Edu	14.40	14.43	13.74
SD	2.42	2.41	2.65
N	21794	20939	855

B. Parent-level Descriptives (Means and Standard Deviations, SD)

	All	No Adopted	Adopted
% Graduates	0.65	0.65	0.67
SD	0.48	0.48	0.47
Parent Birthyear	1939.41	1939.4	1939.6
SD	4.11	4.14	3.71
Parent PGS	0	-0.01	0.11
SD	1	1	0.96
Parent Education	13.75	13.7	14.35
SD	2.36	2.33	2.59
N	7193	6648	545

Table 2. Associations between child educational attainment and parental polygenic score (PGS) for educational attainment.

	M0	M1	M2	M3-bio only	M4-adopted only
Intercept	13.708	14.097	14.147	14.407	14.140
SE	0.105	0.110	0.154	0.032	0.154
p-value	<1E-5	<1E-5	<1E-5	<1E-5	<1E-5
PGS	0.253	0.263	0.227	0.479	0.273
SE	0.113	0.106	0.109	0.022	0.108
p-value	2.53E-02	1.36E-02	3.73E-02	<1E-5	1.16E-02
Bio Child	0.728	0.341	0.258		
SE	0.107	0.112	0.157		
p-value	<1E-5	2.37E-03	9.94E-02		
PGS:Bio Child	0.227	0.216	0.224		
SE	0.115	0.109	0.109		
p-value	4.72E-02	4.73E-02	4.00E-02		
Child birthyear		-0.098	-0.098	0.006	-0.097
SE		0.023	0.023	0.005	0.023
p-value		1.99E-05	2.64E-05	2.23E-01	3.35E-05
Bio Child:Child birthyear		0.105	0.104		
SE		0.023	0.024		
p-value		<1E-5	1.23E-05		
PGS:Child birthyear		0.003	0.002		
SE		0.004	0.004		
p-value		4.90E-01	6.18E-01		
Male Child			-0.086	0.070	-0.075
SE			0.207	0.047	0.207
p-value			6.76E-01	1.32E-01	7.15E-01
Bio Child:Male Child			0.157		
SE			0.210		
p-value			4.55E-01		
PGS:Male Child			0.064		
SE			0.044		
p-value			1.45E-01		
N	21021	21010	21010	20189	821

Table 3. Associations between child education and parental polygenic score in families with both adoptive and biological children.

	Baseline	Saturated
Intercept	13.708	14.140
SE	0.105	0.156
p-value	<1E-5	<1E-5
PGS	0.253	0.229
SE	0.113	0.134
p-value	2.57E-02	8.90E-02
Bio Child	0.733	0.700
SE	0.140	0.207
p-value	<1E-5	7.49E-04
PGS:Bio Child	0.296	0.292
SE	0.154	0.151
p-value	5.41E-02	5.39E-02
Male Child		-0.085
SE		0.209
p-value		6.84E-01
Child Birthyear		-0.097
SE		0.023
p-value		2.34E-05
Bio Child:Male Child		-0.541
SE		0.270
p-value		4.50E-02
PGS:Male Child		-0.013
SE		0.173
p-value		9.42E-01
Bio Child:Child Birthyear		0.087
SE		0.035
p-value		1.36E-02
PGS:Child Birthyear		0.014
SE		0.018
p-value		4.27E-01
N	1491	1486
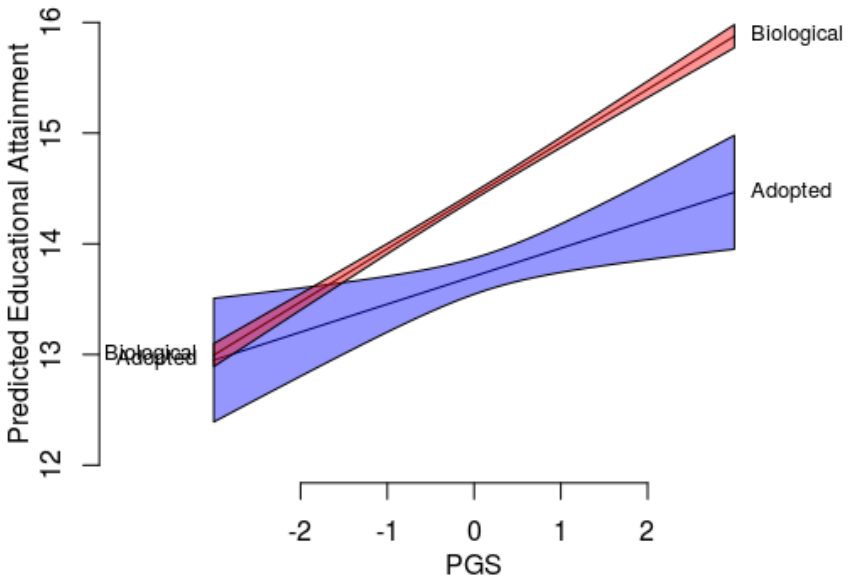
Figure 1. A comparison of predicted educational attainment as a function of parental PGS and child status (biological or adopted).