Separating Measured Genetic and Environmental Effects: Evidence Linking Parental Genotype and Adopted Child Outcomes[1]

Ben Domingue
Stanford University

Jason Fletcher
University of Wisconsin-Madison

Abstract
The bio-demographic literature has begun to use genome wide summary scores (polygenic scores) to predict a broad set of demographic and socioeconomic outcomes to understand the importance of genetics as well as potential life course mechanisms. However, a largely 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 that collected parental polygenic scores and child outcomes for 30,000 adopted and biological children, 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 show large effects of parental polygenic scores on adopted children's schooling, suggesting that polygenic scores combine genetic and environmental influences and that additional research designs will need to be used to separate these estimated impacts.

[1] 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)
The authors thank Ron Lee and participants at the Population Association of America 2018 conference for helpful suggestions.

Introduction and Background

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, the so-called genomics revolution, has led to a set of new genetic discovering linking specific genetic variants with health outcomes as well as schooling outcomes.(Conley and Fletcher 2017, Okbay et al. 2016) While, individually, these variants do not have large effects on schooling outcomes, the development of genome wide summary scores (polygenic scores, PGS) offers the possibility of a genetically-based analytic tool that is highly predictive of important adult outcomes, such as schooling, IQ, and health.(Ware et al. 2017)

Indeed, several studies have examined the impacts of PGS on a series of developmental outcomes during childhood to further understand potential mechanisms that link genes to distal outcomes.(Belsky et al. 2016) Other studies have examined the impact of these PGS on later life outcomes, such as earnings and wealth.(Barth, Papageorge and Thom 2017, Papageorge and Thom 2017) Still others have used these measures to examine gene-environment interactions.(Conley et al. 2015, Schmitz and Conley 2016)

Often, research in these areas has used an individual's polygenic score to assess effects of "genetic factors" without a full acknowledgment that these measures are confounded by parentally provided environments. In particular, parental genetics impact both child environments and child genetics, leaving the estimated effects of polygenic scores difficult to interpret as "purely genetic" effects.[2]

Perhaps some of the strongest evidence that polygenic scores associations with education phenotypes are not purely due to confounding by parental characteristics (genotype or phenotype) are results based on sibling differences. For example, results using the Framingham Heart Study suggest that associations between educational attainment and within-family (between sibling) variation in polygenic scores is actually stronger than associations that use variation between families.(Conley et al. 2015) Similar models that use Add Health data report

[2] Indeed, emerging work is beginning to examine the impacts of parental genotype that are not passed on to children on children's outcomes, which they term "genetic nurture". See Kong, Augustine, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V Halldorsson and Gisli Masson. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359(6374):424-28..

nearly identical associations between polygenic scores and educational attainment when comparing between family and within family (between sibling) models.(Domingue et al. 2015) Both these results could suggest that confounding bias from parental characteristics are small or nonexistent in practice. Alternatively, these models could be producing an alternative source of bias even as they are eliminating bias due to the shared environments of siblings: the possibility of spillover effects between siblings, either phenotypically or genotypically. Indeed, newer work has begun to produce evidence of so-called "social genomic effects", where the genetics of social connected individuals (e.g. classmates) could play a role in own-outcomes.(Domingue and Belsky 2017, Domingue et al. 2018). To the extent that these spillovers exist and are important, sibling difference models suffer from SUTVA violations and can indeed produce results that are biased in either direction. Thus, alternative research designs that do not suffer from this bias could produce new evidence of whether parental genotype confounds the relationship between child genotype and child outcome.

This paper overcomes these issues by using multi-generational data that collected parental polygenic scores and child outcomes for 30,000 adopted and biological children, 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.

Data

The Wisconsin Longitudinal Study (WLS) is a long-term study of over 10,000 men and women, a random 1/3 sample of the graduating high school class of 1957 in Wisconsin (Herd, Carr and Roan 2014). These individuals continue to be followed through the present. 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 over 8,000 respondents. In addition to the graduate and sibling respondents reporting information on their own lives, they also report educational and other information on all their surviving children, including the relationships with each child (biological, step, adopted)[3] so that our sample size of parent/child pairs is over 30,000.

Table 1 presents descriptive statistics of the sample and stratifies these descriptive statistics by child status of adopted, biological, and both. Nearly two thirds of the children have parents who are graduate respondents (compared to their siblings). The graduate respondents were born in a narrow window around 1939 (recall that they are the graduating class of 1957), although the sibling respondents have a larger range. By construction, the education polygenic score of parents is centered at 0, although an important finding is that the polygenic scores of parents who have adopted children are higher than those who have biological children. We will return to this issue in the analysis. 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.

Results

Table 2 presents basic regression results predicting children's educational attainments. Key variables include parental polygenic score (PGS) and biological status of the child; we also control for child sex and age (birth year). Our key result is that we find that parental PGS predicts child education for adopted children—a one standard deviation increase in PGS is associated with 0.14 years of schooling in adopted children, though the result is marginally statistically significant. Because, for biological children, the correlation between parental and child PGS is approximately 0.5, our result suggests that prior research using child PGS to predict own outcomes, such as schooling, are likely biased by parental genotype.

Table 2 also contains findings consistent with past research. As expected, we find that parental PGS predicts child education for biological children—a standard deviation increase in parental PGS predicts 0.36 (0.14+0.22) years of additional schooling for the child. This is a reasonable magnitude, as the effects of child PGS on child educational attainment are 0.37 in Add Health (Domingue et al. 2015), where the analysis used the previous version of the PGS that had lower predictive accuracy. We also find that biological children have nearly 2/3rds a year

[3] For each child, the respondent was asked the relationship, with possible choices: biological, adopted, step, foster, legal ward, niece/nephew, other non-relative, child of lover/partner

more schooling, compared to adopted children. In Table 3, we find evidence that parental PGS predicts schooling differentially for male versus female children—the effect on adopted children is strong for males but non-existent for females. This difference deserves additional analysis for possible explanations and/or mechanisms.

Table 4 extends the model in Table 2 to control for parental education attainment. We estimate this model for two reasons. First, we want to examine whether there are both environmental and genetic effects of parental PGS, and we interpret the parental phenotype (conditional on the PGS) as a reasonable "environmental" measure.[4] We also want to provide evidence of whether controlling for parental educational is "enough" to delink the possible confounding of child genotype with parental genotype on child educational attainment. Including a control for maternal education does reduce the link between maternal PGS and child education for adopted children by a small amount. We do not find differences in the associations between parental education and child education by adoptive/biological status.

As we noted above, 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, which suggests the possibility of other family-level differences between families with adopted versus biological children. In order to more fully investigate this issue, we pursue two strategies. First we focus our analysis on families who have both adopted and biological children. Second we deploy a "cousin fixed effects" strategy that makes use of the fact that the parents in the sample have siblings who are also in the sample. In this way, we can compare the transmission of parental PGS to children outcomes among parents who are more similar (i.e. siblings) than are a random parent in the population.

Table 5 reports results for families with both adoptive and biological children. The results are now stronger when we remove some of the selectivity of having adopted children. The estimate suggests that parental PGS is associated with an increase in adopted children's education by 0.23 years, although again the result is not statistically significant in our modest sample. In these families, we also find that biological children complete over a year of additional schooling on average compared with adopted children.

[4] An issue with this interpretation is that the PGS is measured with error, so that some of the effects of parental genetics may still "load onto" the phenotypic measure of education.

Finally, to more fully control for selection, we take advantage of a feature of the WLS sample, where respondents (parents in our case) and their siblings have been surveyed longitudinally. This allows a comparison of the outcomes of cousins (children of the siblings), holding constant larger family level factors. Table 6 reports these results. We find that parental PGS is associated with a 0.31 increase in child's years of schooling, and importantly we find no difference for adopted children. Like earlier results, this finding suggests that parental genotype is associated with the outcome of children, both adopted (who have no biological relationship with the parent) and also biological children.

Conclusion

A large and growing literature is leveraging the rapid and important advances in genotyping and prediction in genetics and the associated creation of polygenic scores to create new understandings in 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 because of the logic that genetic variation "happens first" in developmental processes and is immune to environmental feedback.

A limitation of this logic is that children's genotype is highly correlated with parental genotype, and, by extension, is correlated with some environmental inputs during childhood. Children with high PGS for education are likely to have grown up in households with highly educated parents and both influences (genetics and environments) likely contribute to the children's educational attainment. Indeed, this issue is part of the logic in earlier work that used adopted children to break the link between child genotype and parental genotype/environment.

The results in our paper show that the limitation in the logic in the literature linking genotype-phenotype using polygenic scores is consequential—an implication of our results, that parental genotype is associated with *adopted* children's phenotypes, is that estimates of the associations between child genotype on child phenotypes are not 'purely genetic' effects. Results in the literature tying own genotype to outcomes, especially socioeconomic outcomes, are a combination of genetic and environmental factors and are not as "pure" as typically interpreted in the literature. This suggests that since polygenic scores combine genetic and environmental influences, additional research designs will need to be used to separate these estimated impacts. One such design is a return to adoption studies, which we have pursued in this work.

Tables

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

	Total	biological	adopted
Graduate Respondent Family	0.65	0.66	0.67
Parent Birth Year	1939	1939	1939
SD	3.98	3.97	3.75
Education Polygenic Score	-0.02	-0.01	0.09
SD	1.00	1.00	0.90
Parent Educational Attainment	13.61	13.57	14.35
SD	2.31	2.29	2.57
Child Birth Year	1966	1967	1969
SD	9.40	6.79	11.92
Child Educational Attainment	14.33	14.42	13.72
SD	2.43	2.41	2.65
N	24467	21637	905

Table 2
Associations between Child Educational Attainment and Parent Genetics

	Child Educational Attainment
Constant	13.93
SE	127.79
Parent Polygenic Score	0.14
	1.46
Biological Child Indicator	0.66
	6.00
Male	-0.31
	-2.24
Child Birth Year	-0.04
	-5.41
Polygenic Score X Biological Child	0.22
	2.38
Polygenic Score X Male	0.05
	1.71
Biological Child X Male	0.09
	0.67
Polygenic Score X Child Birth Year	-0.01
	-2.26
Biological Child X Child Birth Year	0.03
	4.25
N.groups	7545.00
N.obs	21733.00

Table 3
Associations between Child Education and Parent Genetics
Stratified Results by Sex-of-Child

	Females	Males
	Child Educational Attainment	Child Educational Attainment
Constant	13.91	13.67
SE	122.68	112.44
Parent Polygenic Score	-0.02	0.23
	-0.14	1.72
Biological Child Indicator	0.65	0.69
	5.60	5.58
Child Birth Year	-0.02	-0.03
	-2.81	-3.06
Polygenic Score X Biological Child	0.38	0.17
	3.03	1.30
Polygenic Score X Male	NA	NA
	NA	NA
Biological Child X Male	NA	NA
	NA	NA
Polygenic Score X Child Birth Year	-0.01	0.00
	-1.59	-1.26
Biological Child X Child Birth Year	0.03	0.02
	3.26	1.93
N.groups	6195.00	6280.00
N.obs	10761.00	10972.00

Table 4
Associations between Child Education, Parental Genetics, and Parental Education

	Child Education	Child Education
Constant	12.79	12.79
SE	93.03	37.98
Parent Polygenic Score	0.11	0.09
	1.14	0.94
Biological Child Indicator	0.67	0.67
	6.10	1.97
Male	-0.32	-0.32
	-2.34	-2.35
Child Birth Year	-0.04	-0.01
	-5.54	-4.86
Parental Education	0.11	0.10
	13.62	3.44
Polygenic Score X Biological Child	0.23	0.25
	2.42	2.63
Polygenic Score X Male	0.05	0.05
	1.60	1.60
Biological Child X Male	0.10	0.10
	0.73	0.74
Polygenic Score X Child Birth Year	-0.01	-0.01
	-2.53	-2.47
Biological Child X Child Birth Year	0.03	NA
	4.05	NA
Biological Child X Parental Education		0.01
		0.21
N.groups	7486.00	7486.00
N.obs	21573.00	21573.00

Table 5
Associations between Child Education and Parental Genetics
Sample of Families with Both Adoptive and Biological Children

	Estimate	Std Error	T-stat	Pr(>|t|)
Parental Polygenic Score	0.23	0.18	1.24	0.22
Biological Child Indicator	1.05	0.16	6.74	0.00
Grad/sib sample	-0.04	0.23	-0.16	0.87
Male Child	-0.27	0.10	-2.69	0.01
Child Birth Year	-0.02	0.01	-3.18	0.00
Child Birth Order	-0.07	0.04	-1.68	0.09
Polygenic Score X Biological Child	0.31	0.16	1.89	0.06
Polygenic Score X Grad Sample	0.37	0.25	1.47	0.14
Biological X Grad Sample	-0.09	0.27	-0.34	0.73
PGS x Bio X Grad	-0.50	0.29	-1.73	0.08

Table 6
Associations between Child Education and Parental Genetics
Cousin Fixed Effects

	Estimate	Std. Error	t value	Pr(>|t|)
Parental Polygenic Score	0.31	0.06	5.59	0.00
Adopted Child Indicator	-0.80	0.20	-3.99	0.00
Male Child	-0.24	0.03	-7.51	0.00
Child Birth Year	-0.02	0.00	-5.16	0.00
Child Birth Order	-0.03	0.01	-1.83	0.07
Polygenic Score X Adopted Child	-0.05	0.11	-0.46	0.65
Polygenic Score X Male Sample	0.04	0.03	1.44	0.15
Polygenic Score X Child Birth Year	-0.01	0.00	-3.42	0.00
Polygenic Score X Birth Order	0.02	0.01	1.54	0.12
Adopted X Male	-0.07	0.16	-0.45	0.66
Adopted X Child Birth Year	-0.03	0.01	-3.18	0.00
Adopted X Birth Year	-0.04	0.07	-0.63	0.53

References

Barth, Daniel J, Nicholas W Papageorge and Kevin Thom. 2017. "Genetic Ability, Wealth, and Financial Decision-Making."

Belsky, Daniel W, Terrie E Moffitt, David L Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, Sandhya Ramrakha, Karen Sugden and Benjamin S Williams. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-Course Development." *Psychological Science* 27(7):957-72.

Conley, Dalton, Benjamin W Domingue, David Cesarini, Christopher Dawes, Cornelius A Rietveld and Jason D Boardman. 2015. "Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype?". *Sociological Science* 2:82-105.

Conley, Dalton and Jason Fletcher. 2017. *The Genome Factor: What the Social Genomics Revolution Reveals About Ourselves, Our History, and the Future*: Princeton University Press.

Domingue, Benjamin W, Daniel W Belsky, Dalton Conley, Kathleen Mullan Harris and Jason D Boardman. 2015. "Polygenic Influence on Educational Attainment: New Evidence from the National Longitudinal Study of Adolescent to Adult Health." *AERA open* 1(3):2332858415599972.

Domingue, Benjamin W and Daniel W Belsky. 2017. "The Social Genome: Current Findings and Implications for the Study of Human Genetics." *PLoS genetics* 13(3):e1006615.

Domingue, Benjamin W, Daniel W Belsky, Jason M Fletcher, Dalton Conley, Jason D Boardman and Kathleen Mullan Harris. 2018. "The Social Genome of Friends and Schoolmates in the National Longitudinal Study of Adolescent to Adult Health." *Proceedings of the National Academy of Sciences*:201711803.

Herd, Pamela, Deborah Carr and Carol Roan. 2014. "Cohort Profile: Wisconsin Longitudinal Study (Wls)." *International journal of epidemiology* 43(1):34-41.

Kong, Augustine, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V Halldorsson and Gisli Masson. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359(6374):424-28.

Okbay, Aysu, Jonathan P Beauchamp, Mark A Fontana, James J Lee, Tune H Pers, Cornelius A Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson and S Fleur W Meddens. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature* 533(7604):539.

Papageorge, Nicholas and Kevin Thom. 2017. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study."

Plomin, Robert and Sophie von Stumm. 2018. "The New Genetics of Intelligence." *Nature Reviews Genetics*.

Schmitz, Lauren L and Dalton Conley. 2016. "The Effect of Vietnam-Era Conscription and Genetic Potential for Educational Attainment on Schooling Outcomes." Vol.: National Bureau of Economic Research.

Ware, Erin B, Lauren L Schmitz, Jessica D Faul, Arianna Gard, Colter Mitchell, Jennifer A Smith, Wei Zhao, David Weir and Sharon LR Kardia. 2017. "Heterogeneity in Polygenic Scores for Common Human Traits." *bioRxiv*:106062.