I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15862

Understanding Sibling Correlations in Education: Molecular Genetics and Family Background

Jason Fletcher
Qiongshi Lu
Bhashkar Mazumder
Jie Song

JANUARY 2023

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15862

Understanding Sibling Correlations in Education: Molecular Genetics and Family Background

Jason Fletcher
University of Wisconsin-Madison and IZA

Qiongshi Lu
University of Wisconsin-Madison

Bhashkar Mazumder
Federal Reserve Bank of Chicago

Jie Song
University of Wisconsin-Madison

JANUARY 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Understanding Sibling Correlations in Education: Molecular Genetics and Family Background*

Sibling correlations in socioeconomic status are one of the key measures of equality of opportunity and social mobility, providing an omnibus examination of the importance of family background. Typically, these correlations are interpreted as the combined effects of shared sibling background and experiences, including genetics and family environments. The UK Biobank allows us to specifically control for sibling and parental genetics (polygenic scores, PGS) in order to gauge their relative importance compared with broader family background effects. We use >17,000 sibling pairs from the UK Biobank in order to further decompose standard sibling correlations of educational attainment found in the literature. In general, we find modest (up to 20%) contributions of molecular genetics to the similarity of sibling outcomes, suggesting a large amount of the observed similarity in sibling educational outcomes are due to parents and environments of children.

JEL Classification: J62, J12, J24
Keywords: sibling correlations, educational mobility, genetics

Corresponding author:
Jason Fletcher
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706
USA
E-mail: jason.fletcher@wisc.edu

---

* Authorship is alphabetized. The authors gratefully acknowledge use of the facilities of the Center for Demography of Health and Aging at the University of Wisconsin-Madison, funded by NIA Center Grant P30 AG017266. We thank Patrick Turley, Joe Ferrie, members of the Social Genomics Working Group at University of Wisconsin, and participants at IGSS and Genes, Social Mobility, and Inequalities across the Life-Course, and Bridges conferences for helpful comments. This research has been conducted using the UK Biobank Resource under Application 57284. The views expressed here do not reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Introduction

The correlation in socioeconomic status among siblings provides an omnibus measure of the importance of family background in determining children's outcomes by capturing factors shared in common by siblings such as parental involvement, schools, and neighborhoods (e.g. (Corcoran, Jencks, and Olneck 1976; Mazumder 2008; Solon, Page, and Duncan 2000; Bingley and Cappellari 2019; Björklund, Jäntti, and Lindquist 2009). By highlighting the overall role of family background, and not just one particular dimension of status (e.g. parent income), the sibling correlation is a particularly useful broad measure of relative intergenerational mobility and equality of opportunity (Deutscher and Mazumder). The sibling correlation in several economic outcomes in the US is estimated to be around 0.5 (Mazumder, 2008), suggesting that half of economic inequality could be attributed to family background and that, therefore, intergenerational economic mobility in the US is fairly modest. These figures are lower in many Nordic countries(Aaberge et al. 2002; Björklund et al. 2002). In the UK, many studies estimate intergenerational persistence of educational of approximately 0.35-0.4 (Dearden, Machin, and Reed 1997; Björklund and Jäntti 2020).

While sibling correlations provide valuable information about social mobility patterns, taken on their own, they provide limited insight into mechanisms. Broadly speaking, potential mechanisms are often indirectly explored by comparing differences in mobility across time and place[2]. A few studies have also examined the extent to which sibling correlation estimates change as additional variables are added to the model—for example, adolescent test scores and related human capital measures have been shown to explain about 50% of the sibling correlation in earnings and wages (Mazumder 2008).

Some observers have cited evidence that suggests that genetic inheritance explains much of the variation in economic outcomes in the US and that therefore genes can largely explain the observed degree of inequality in opportunity (Mankiw 2013). This perspective implies that there may be a "ceiling" on how much environmental changes could impact intergenerational mobility.

---

[2] A related literature is from genetics examines variation in heritability (i.e. the genetic contribution to an outcome, like education). Branigan et al. (2013) showed both a wide variation in heritability estimates of education across time and place and an average value around 0.4—the datasets from the UK included in the meta-analysis have estimates of 0.24 and 0.39. More recent analysis has been able to use measured molecular data to estimate a subset of heritability, called SNP-heritability. SNP-heritability can capture the subset of heritability that is produced by common (not rare) genetic variation in the population. Lee et al. estimated a SNP-heritability for educational attainment for the UK Biobank sample of 0.23.

Some work has used comparisons of different genetic "types" of siblings and relatives to suggest a large component of these correlations are related to genetics (Bjorklund, Jantti, and Solon 2005). Indeed, the rapid increase in genetic measurements available in large survey data now allows for consideration of this and related questions using more direct measures.

Thus far, there have only been a limited number of studies that use genetic measures to unpack patterns of social mobility. A major question is how to measure "genetics". The convention in the literature is to use a genome-wide summary measure, called a polygenic score (PGS), that aggregates the effects of many (often > 100,000) genetic variants on the outcome of interest[3]. One previous study found that adding parental and child polygenic scores did not change estimates of intergenerational education mobility, though the sample was small and from a single city in the US, and the specific polygenic scores were constructed from older genetics findings with high levels of measurement error (Conley et al. 2015).

We extend the literature that explores the mechanisms underlying sibling correlations in educational attainment by using a large sample of siblings (>17,000 pairs) from the UK Biobank (UKB). We also use new methods that impute parental genetics from sibling genetics inputs for additional controls. We follow typical practice in social science by examining random-effect models of educational attainment, adding polygenic scores for educational attainment of the siblings and the parents to explore changes to sibling correlation estimates. We find moderate sibling correlations in educational attainment (~0.32) that are reduced by only ~10% with polygenic scores. We additionally correct for potential measurement error in the polygenic scores and find reductions of ~20%, or a residual sibling correlation of 0.25. All our results point to modest impacts of molecular genetics on sibling correlations and thus suggest other aspects of the family and shared environments are more important in explaining the levels of sibling similarity in educational attainment. While this is a useful contribution to the literature, we note that, from the perspective of policy makers, it may not matter so much whether genes or environment play a larger role in explaining sibling correlations, as policies can potentially be crafted to improve opportunity irrespective of the cause (Manski 2011; Goldberger 1979).

---

[3] These effect estimates are taken from large scale (often >1M) genome wide association studies (GWAS) that link each genetic location to the outcome using regression analysis, controlling for overall genetic effects (principal components) and limiting the sample to those of the same ancestral group.

Data

This section repeats some text from our other work using the same data(Fletcher et al. 2020). We used data from the UKB project(Collins 2012). The participants, aged between 37 and 74 years, were originally recruited between 2006 and 2010. These data are restricted, but one can gain access by following the procedures described in www.ukbiobank.ac.uk/register-apply/.

Although siblings are not identified in the survey, respondents' genetics can be used to measure genetic relatedness among all pairs of respondents. We first use the UKB-provided kinship file, listing all pairwise kinships among 100,000 pairs in the sample of nearly 500,000 individuals. We chose pairs with kinship >0.2, which reflects first degree biological relatives (parents/siblings). We then chose remaining pairs who are <13 years apart in age, leaving ~22,000 sibling dyads. We then chose to keep only one dyad from any family with more one dyad, leaving ~17,600 dyads. We include only respondents of European ancestry in our analysis due to limited predictive power of PGS in non-European samples (Martin et al. 2017).

We constructed PGS for two traits for which large genome wide association studies (GWAS) are publicly available and do not contain UKB samples: height (Wood et al. 2014) and educational attainment(J. J. Lee, Wedow, Okbay, Kong, Maghzian, Zacher, Nguyen-Viet, et al. 2018) with UKB samples removed. We also constructed a second PGS within the UKB using four-fold cross-validation for some analysis (Song, Zheng, et al. 2022). For each fold, we conducted GWAS on 75% of UKB samples while adjusting for sex, age, and genetic principal components, and produced PGS on the 25% holdout samples using the same protocol described below[4]. For parental genetics, we used full sibling data and SNIPAR to impute "average" parental PGS (Young et al. 2020).

We followed standard protocols in creating our PGS. We removed single-nucleotide polymorphisms (SNPs) in strong linkage disequilibrium (LD). We LD-clumped the GWAS summary data by PLINK(Purcell et al. 2007), using 1000 Genomes Project Phase III European genotype data as reference. We used a LD window size of 1Mb and a pairwise r^2 threshold of 0.1. We did not apply any p-value thresholding to select SNPs. Final scores were produced using PRSice-2(Choi and O'Reilly 2019). The PGS were normalized to have mean zero and SD one and oriented so that each PGS was positively correlated with its corresponding outcome.

---

[4] We have also constructed a PGS using the non-siblings in the UKB for the GWAS inputs. Results are quite similar to the cross-validation PGS.

Empirical Model and Results

Our analysis begins with a typical sibling correlation/association examination for education attainment using a random-effect regression model. We then add genetic measurements as control variables to explore changes in the rho (sibling similarity) parameter. Table 1 shows descriptive statistics for the UKB and the sibling sub-sample, both only for those with European ancestry. The average years of schooling is 13.55 and age at interview is 57. The sample is 58% female. The PGS are standardized on the full sample. Appendix Table 1A shows that the sibling sample is quite similar to the full sample, though slightly older and less educated.

Table 2 presents our first results of sibling correlations. Column 1 controls for age, age-squared, and gender and estimates rho as 0.318[5]. We first add controls for parental polygenic scores for education in Column 2, which reduces rho by 7% to 0.296; Column 3, instead of parental PGS, controls for child PGS and shows a similar reduction of 8% to 0.294 from the baseline[6]. Column 4 adds 10 genetic principal components to explore "general" genetic effects that are not targeted to educational attainment and shows a minimal reduction of rho. Column 5 then adds all these genetic measures into a single analysis and finds an overall 12% reduction of rho to 0.279 and an overall incremental r^2 of 5.2%.

The results suggest a modest amount of sibling correlation is explained by including a broad set (i.e. both parent and child) of genetic measures that are trained to predict educational attainment. However, there are several measurement and conceptual issues to note. One key conceptual issue is that it is unclear whether to apportion parental PGS as "genetics" or "non-genetics" because they include genetics not passed to the child (i.e. non-transmitted alleles). A portion of parental PGS could just as well be added to "non-genetic" family factors. Leaving this issue aside for now, we also need to consider measurement error in the polygenic scores. While these polygenic scores are the most predictive measures available, the "child" PGS only explains 3.5% of the overall variation. We can benchmark this against what is termed the SNP-heritability of educational attainment in the UKB, which is the total variation explained by all common genetic

---

[5] To gauge the representativeness of the UKB data, we used representative UKHLS data to estimate sibling correlations in educational attainment to be 0.35.

[6] Recall that parental EA-PGS and child EA-PGS are correlated at 0.5.

variants (single nucleotide polymorphisms, or SNPs), which is estimated to be 23% (James J Lee, Wedow, Okbay, Kong, Maghzian, Zacher, Johannesson, et al. 2018).

To further pursue this issue, we use methods developed in other work to create a second PGS to be used as an instrument for the first PGS (DiPrete, Burik, and Koellinger 2018) and, separately, to include as a second measure of child genetics. Table 3 reports results using two PGS for educational attainment. As expected, once we correct for measurement error in the child's PGS, our estimate or rho falls further, by 17%, from Column 1 (baseline, repeated from Table 2) to Column 2. We achieve a larger reduction adding both PGS rather than using one to instrument the other—Column 3 reduces rho by 20%. Adding parental PGS increases this by 3% in each case, so that the largest reduction we see, in Column 5, includes 2 children's PGS, parental PGS, and PCs, reducing rho by 23% with an incremental r^2 of 6.8%.[7]

Table 4 presents estimates of sibling correlation in family income. This measure is reported in five categories, so it is not ideal for a full parallel exploration with educational attainment but can be suggestive if it is consistent or inconsistent with our earlier results. Recent work has estimated the SNP-heritability of income in the UKB sample to be 11%(Hill et al. 2019). Our rho estimates start at 0.2 and fall by only ~13% when including molecular genetic controls as above. We note that while our genetic measurements are targeted at predicting educational attainment rather than income, research has shown that the genetic correlation between income and education is 0.94 (Hill et al. 2019).

Table 5 presents estimates for sibling correlations in height that adds child/parent polygenic scores for height to benchmark the results for education against an outcome that we expect will have much stronger genetic contributions. The main takeaway is that shared genetic factors contribute a modest (~10%) amount of the sibling correlation in height.

Discussion

This paper presents new evidence on the sources of sibling correlation in educational attainment using national data on ~17K sibling pairs from the UK. While some researchers suggest that genetic transmission may place a ceiling on mobility patterns, or "lock in" advantages, our

---

[7] New work in biostatistics has used a method of moments approach to decompose variance in educational attainment into genetic and (residual) environmental correlation. Using the sibling data in the UKB, (Song, Zou, et al. 2022) estimated sibling correlations (net of genetics) of 0.21.

results suggest the ceiling is quite high. Indeed, we find that adding measured genetic predictors of educational attainment at both the parent and child levels explain a modest amount of the overall sibling correlation in educational attainment in our UK sample. These results add to an emerging literature showing an often-surprising *lack* of measured genetic effects on key social processes that are thought to have moderate or higher genetic heritability based on earlier work without the benefit of measured genetics. The results indirectly point to the importance of environmental effects on educational attainment that are shared by siblings, such as schools, neighborhoods, etc. Future research should expand these results in additional cohorts and for additional domains.

Tables

Table 1
Descriptive Statistics
UKB Sibling Sample

Variable	Obs	Mean	Std Dev	Min	Max
Education (years)	34,232	13.55	5.10	7	20
Height (cm)	34,456	168.06	9.18	126	204
Age	34,516	57.13	7.29	40	70
Year of Birth	34,516	1950.89	7.29	1937	1970
Female	34,516	0.58	0.49	0	1
Parent EA-PGS (std)	33,073	0.01	0.99	-3.80	4.73
EA-PGS, CV (std)	34,232	-0.01	1.03	-4.10	3.13
EA-PGS (std)	34,516	0.05	1.01	-4.52	3.99
EA-PGS, 2 (std)	34,232	0.00	0.20	-0.78	0.60
Height PGS	34,516	0.00	1.00	-3.98	4.18

PGS: Polygenic Score; CV: Cross-Validated; EA: Educational Attainment

Table 2
Sibling Correlations in Education Attainment in the UK Biobank
Contribution from Polygenic Scores for Educational Attainment

Outcome	Education	Education	Education	Education	Education
Age.	0.426***	0.461***	0.411***	0.412***	0.448***
	(0.055)	(0.057)	(0.054)	(0.054)	(0.056)
Age-sq	-0.005***	-0.005***	-0.004***	-0.004***	-0.005***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Female	-0.802***	-0.812***	-0.794***	-0.794***	-0.806***
	(0.053)	(0.054)	(0.053)	(0.053)	(0.053)
EA PGS (std)			0.865***	0.888***	0.765***
			(0.028)	(0.028)	(0.029)
Parent EA PGS (std)		0.835***			0.667***
		(0.031)			(0.031)
Constant	4.670***	3.737**	5.100***	5.195***	4.213***
	(1.529)	(1.579)	(1.505)	(1.519)	(1.577)
Observations	34,232	32,802	34,232	34,232	32,802
Number of famid	17,687	16,557	17,687	17,687	16,557
withinr2	0.0123	0.0124	0.0163	0.0163	0.0171
betweenr2	0.0257	0.0668	0.0781	0.0820	0.104
overallr2	0.0215	0.0481	0.0563	0.0586	0.0736
sigma_u	2.843	2.704	2.681	2.669	2.589
sigma_e	4.164	4.173	4.153	4.153	4.162
rho	0.318	0.296	0.294	0.292	0.279
% Reduction		-7%	-8%	-8%	-12%
Incremental R2		0.0266	0.0348	0.0371	0.0521

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

Table 3
Sibling Correlations in Education Attainment in the UK Biobank
Contribution from Polygenic Scores for Educational Attainment
Measurement Error Adjustments

Outcome	Education	Education	Education	Education	Education
Model		2Step PGS PCs	2 Kid PGS PCs	2Step PGS Parent PGS PCs	2 Kid PGS Parent PGS PCs
Age.	0.426***	0.408***	0.399***	0.445***	0.436***
	(0.055)	(0.054)	(0.054)	(0.056)	(0.056)
Age-sq	-0.005***	-0.004***	-0.004***	-0.005***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Female	-0.802***	-0.798***	-0.791***	-0.808***	-0.803***
	(0.053)	(0.053)	(0.052)	(0.054)	(0.053)
EA PGS (std)			0.775***		0.697***
			(0.028)		(0.029)
Parent EA PGS (std)				0.648***	0.526***
				(0.032)	(0.032)
Two Step EA-PGS		3.789***		3.021***	
		(0.136)		(0.144)	
EA-PGS CV (std)			0.588***		0.484***
			(0.026)		(0.028)
Constant	4.670***	5.196***	5.588***	4.171***	4.560***
	(1.529)	(1.522)	(1.504)	(1.581)	(1.567)
Observations	34,232	34,232	34,232	32,802	32,802
Number of famid	17,687	17,687	17,687	16,557	16,557
withinr2	0.0123	0.000597	0.00482	0.00164	0.00640
betweenr2	0.0257	0.118	0.140	0.122	0.145
overallr2	0.0215	0.0607	0.0832	0.0716	0.0899
sigma_u	2.843	2.487	2.415	2.435	2.374
sigma_e	4.164	4.156	4.143	4.165	4.152
rho	0.318	0.264	0.254	0.255	0.246
% Reduction		-17%	-20%	-20%	-23%
Incremental R2		0.0392	0.0617	0.0501	0.0684

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

Table 4
Sibling Correlations in Income
Contribution from Polygenic Scores for Educational Attainment

Outcome	Income	Income	Income	Income	Income	Income	Income	Income	Income
					PGS			2Step PGS	2 Kid PGS
				PGS	Parent PGS	2Step PGS	2 Kid PGS	Parent PGS	Parent PGS
Model		Parent PGS	PGS	PCs	PCs	PCs	PCs	PCs	PCs
Age.	4,419.296***	4,536.627***	4,385.263***	4,393.059***	4,507.135***	4,384.647***	4,363.741***	4,485.991***	4,464.833***
	(304.750)	(316.665)	(302.810)	(302.277)	(314.837)	(301.923)	(300.949)	(315.246)	(314.322)
Age-sq	-49.746***	-50.902***	-49.518***	-49.606***	-50.693***	-49.556***	-49.428***	-50.501***	-50.362***
	(2.731)	(2.829)	(2.714)	(2.709)	(2.813)	(2.706)	(2.698)	(2.817)	(2.809)
Female	-5,648.960***	-5,820.271***	-5,640.510***	-5,638.337***	-5,810.324***	-5,697.072***	-5,684.747***	-5,868.545***	-5,854.406***
	(300.618)	(305.984)	(299.132)	(298.749)	(304.607)	(298.618)	(297.848)	(304.997)	(304.287)
EA PGS (std)			2,787.771***	2,921.422***	2,528.553***		2,085.066***		2,009.424***
			(152.433)	(154.210)	(160.927)		(160.265)		(164.889)
Parent EA PGS (std)		2,668.688***			2,079.879***			1,395.234***	1,166.945***
		(165.290)			(167.692)			(179.393)	(179.590)
Two Step EA-PGS						12,314.920***		10,264.283***	2,417.121***
						(541.334)		(605.878)	(175.301)
EA-PGS CV (std)							2,884.052***		
							(159.259)		
Constant	-43,109.087**	-45,854.663**	-42,098.959**	-39,791.551**	-42,502.558**	-40,203.664**	-39,187.122**	-42,460.399**	-41,465.759***
	(8,399.430)	(8,758.850)	(8,345.160)	(8,417.112)	(8,800.539)	(8,409.838)	(8,382.506)	(8,815.333)	(8,789.284)
Observations	29,811	28,548	29,811	29,811	28,548	29,735	29,735	28,473	28,473
Number of famid	17,099	16,087	17,099	17,099	16,087	17,084	17,084	16,073	16,073
withinr2	0.0432	0.0442	0.0436	0.0436	0.0452	0.0428	0.0435	0.0449	0.0455
betweenr2	0.132	0.147	0.148	0.153	0.162	0.162	0.170	0.165	0.172
overallr2	0.106	0.116	0.118	0.121	0.127	0.127	0.132	0.128	0.133
sigma_u	11339	11070	10961	10845	10723	10575	10397	10610	10437
sigma_e	23177	23199	23160	23162	23185	23186	23176	23209	23199
rho	0.193	0.185	0.183	0.180	0.176	0.172	0.168	0.173	0.168
Standard errors in pa	% Change	4%	5%	7%	9%	11%	13%	10%	13%

Notes: Income categories imputed to pounds as 1=15000 2=25000 3=45000 4=75000 5=125000

Table 5
Sibling Correlations in Height (CM) in the UK Biobank
Contribution from Polygenic Scores for Height

VARIABLES	Height	Height Parent PGS	Height PGS PCs	Height PGS Parent PGS PCs
Age	0.041	0.021	0.036	0.029
	(0.065)	(0.066)	(0.061)	(0.062)
Age-sq	-0.002***	-0.001**	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Female	-13.405***	-13.400***	-13.379***	-13.382***
	(0.060)	(0.060)	(0.056)	(0.057)
Parent Height PGS		2.015***		1.230***
		(0.041)		(0.041)
Height PGS (std)			2.433***	2.115***
			(0.033)	(0.035)
Constant	178.935***	179.559***	181.283***	181.287***
	(1.809)	(1.826)	(1.700)	(1.750)
Observations	34,456	33,015	34,456	33,015
Number of famid	17,699	16,560	17,699	16,560
withinr2	0.686	0.686	0.718	0.719
betweenr2	0.439	0.506	0.536	0.555
overallr2	0.522	0.570	0.598	0.613
sigma_u	4.646	4.178	4.137	3.969
sigma_e	4.325	4.340	4.093	4.108
rho	0.536	0.481	0.505	0.483
% Reduction		-10%	-6%	-10%
Incremental R2		0.048	0.076	0.091

Notes: Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix Tables

Appendix Table 1
Comparison of Full UK Biobank Sample with Sibling Sub-Sample
European Ancestry Respondents

Variable	Obs	Mean	Std Dev	Obs	Mean	Std Dev
Education (years)	404,436	13.77	5.12	34,232	13.55	5.10
Height (cm)	407,384	168.72	9.25	34,456	168.06	9.18
Age	408,248	56.91	8.00	34,516	57.13	7.29
Year of Birth	408,248	1951.14	8.01	34,516	1950.89	7.29
Female	408,248	0.54	0.50	34,516	0.58	0.49
EA-PGS, CV (std)	404,436	0.00	1.00	34,232	-0.01	1.03
EA-PGS (std)	408,248	0.00	1.00	34,516	0.05	1.01
EA-PGS, 2 (std)	404,436	0.00	0.19	34,232	0.00	0.20
Height PGS	408,248	0.00	1.00	34,516	0.00	1.00

References

Aaberge, Rolf, Anders Björklund, Markus Jäntti, Mårten Palme, Peder J Pedersen, Nina Smith, and Tom Wennemo. 2002. "Income inequality and income mobility in the Scandinavian countries compared to the United States." *Review of Income and Wealth* 48 (4): 443-469.

Bingley, Paul, and Lorenzo Cappellari. 2019. "Correlation of brothers' earnings and intergenerational transmission." *Review of Economics and Statistics* 101 (2): 370-383.

Bjorklund, Anders, Markus Jantti, and Gary Solon. 2005. "Influences of nature and nurture on earnings variation." *Unequal chances: Family background and economic success*: 145-164.

Björklund, Anders, Tor Eriksson, Markus Jäntti, Oddbjörn Raaum, and Eva Österbacka. 2002. "Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States." *Journal of Population Economics* 15 (4): 757-772.

Björklund, Anders, and Markus Jäntti. 2020. "Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: a comparison of four approaches." *Research in Social Stratification and Mobility*: 100455.

Björklund, Anders, Markus Jäntti, and Matthew J Lindquist. 2009. "Family background and income during the rise of the welfare state: brother correlations in income for Swedish men born 1932–1968." *Journal of Public Economics* 93 (5-6): 671-680.

Choi, S. W., and P. F. O'Reilly. 2019. "PRSice-2: Polygenic Risk Score software for biobank-scale data." *Gigascience* 8 (7). https://doi.org/10.1093/gigascience/giz082. https://www.ncbi.nlm.nih.gov/pubmed/31307061.

Collins, Rory. 2012. "What makes UK Biobank special." *Lancet* 379 (9822): 1173-1174.

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.

Corcoran, Mary, Christopher Jencks, and Michael Olneck. 1976. "The effects of family background on earnings." *The American Economic Review* 66 (2): 430-435.

Dearden, Lorraine, Stephen Machin, and Howard Reed. 1997. "Intergenerational mobility in Britain." *The Economic Journal* 107 (440): 47-66.

Deutscher, Nathan, and Bhashkar Mazumder. "Measuring Intergenerational Income Mobility: A Synthesis of Approaches." *Journal of Economic Literature*.

DiPrete, Thomas A, Casper AP Burik, and Philipp D Koellinger. 2018. "Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data." *Proceedings of the National Academy of Sciences*: 201707388.

Fletcher, Jason, Yuchang Wu, Zijie Zhao, and Qiongshi Lu. 2020. "The production of within-family inequality: Insights and implications of integrating genetic data." *BioRxiv*.

Goldberger, Arthur S. 1979. "Heritability." *Economica* 46 (184): 327-347.

Hill, W David, Neil M Davies, Stuart J Ritchie, Nathan G Skene, Julien Bryois, Steven Bell, Emanuele Di Angelantonio, David J Roberts, Shen Xueyi, and Gail Davies. 2019. "Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income." *Nature communications* 10 (1): 1-16.

Lee, J. J., R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. Karlsson Linner, M. A. Fontana, T. Kundu, C. Lee, H. Li, R. Li, R. Royer, P. N.

Timshel, R. K. Walters, E. A. Willoughby, L. Yengo, Team andMe Research, Cogent, Consortium Social Science Genetic Association, M. Alver, Y. Bao, D. W. Clark, F. R. Day, N. A. Furlotte, P. K. Joshi, K. E. Kemper, A. Kleinman, C. Langenberg, R. Magi, J. W. Trampush, S. S. Verma, Y. Wu, M. Lam, J. H. Zhao, Z. Zheng, J. D. Boardman, H. Campbell, J. Freese, K. M. Harris, C. Hayward, P. Herd, M. Kumari, T. Lencz, J. Luan, A. K. Malhotra, A. Metspalu, L. Milani, K. K. Ong, J. R. B. Perry, D. J. Porteous, M. D. Ritchie, M. C. Smart, B. H. Smith, J. Y. Tung, N. J. Wareham, J. F. Wilson, J. P. Beauchamp, D. C. Conley, T. Esko, S. F. Lehrer, P. K. E. Magnusson, S. Oskarsson, T. H. Pers, M. R. Robinson, K. Thom, C. Watson, C. F. Chabris, M. N. Meyer, D. I. Laibson, J. Yang, M. Johannesson, P. D. Koellinger, P. Turley, P. M. Visscher, D. J. Benjamin, and D. Cesarini. 2018. "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals." *Nat Genet* 50 (8): 1112-1121. https://doi.org/10.1038/s41588-018-0147-3. https://www.ncbi.nlm.nih.gov/pubmed/30038396.

Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, M Johannesson, PD Koellinger, P Turley, and PM Visscher. 2018. "Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment." *Nat Genet, in press*.

Mankiw, N Gregory. 2013. "Defending the one percent." *Journal of economic perspectives* 27 (3): 21-34.

Manski, Charles F. 2011. "Genes, eyeglasses, and social policy." *Journal of Economic Perspectives* 25 (4): 83-94.

Martin, Alicia R, Christopher R Gignoux, Raymond K Walters, Genevieve L Wojcik, Benjamin M Neale, Simon Gravel, Mark J Daly, Carlos D Bustamante, and Eimear E Kenny. 2017. "Human demographic history impacts genetic risk prediction across diverse populations." *The American Journal of Human Genetics* 100 (4): 635-649.

Mazumder, Bhashkar. 2008. "Sibling similarities and economic inequality in the US." *Journal of Population Economics* 21 (3): 685-701.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. "PLINK: a tool set for whole-genome association and population-based linkage analyses." *Am J Hum Genet* 81 (3): 559-75. https://doi.org/10.1086/519795. https://www.ncbi.nlm.nih.gov/pubmed/17701901.

Solon, Gary, Marianne E. Page, and Greg J. Duncan. 2000. "Correlations between Neighboring Children in Their Subsequent Educational Attainment." *The Review of Economics and Statistics* 82 (3): 383-392. https://doi.org/10.2307/2646799. http://www.jstor.org/stable/2646799.

Song, Jie, Fengyi Zheng, Yuchang Wu, Jason Fletcher, and Qiongshi Lu. 2022. Generating polygenic risk scores in large biobanks through cross-validation. Working Paper.

Song, Jie, Clara Zou, Yuchang Wu, Jiacheng Miao, Ze Yu, Jason Fletcher, and Qiongshi Lu. 2022. **Decomposing heritability and genetic covariance by direct and indirect effect paths** In *Working Paper*.

Wood, Andrew R., Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H. Pers, Stefan Gustafsson, Audrey Y. Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, Najaf Amin, Martin L. Buchkovich, Damien C. Croteau-Chonka, Felix R. Day, Yanan Duan, Tove Fall, Rudolf

Fehrmann, Teresa Ferreira, Anne U. Jackson, Juha Karjalainen, Ken Sin Lo, Adam E. Locke, Reedik Mägi, Evelin Mihailov, Eleonora Porcu, Joshua C. Randall, André Scherag, Anna A. E. Vinkhuyzen, Harm-Jan Westra, Thomas W. Winkler, Tsegaselassie Workalemahu, Jing Hua Zhao, Devin Absher, Eva Albrecht, Denise Anderson, Jeffrey Baron, Marian Beekman, Ayse Demirkan, Georg B. Ehret, Bjarke Feenstra, Mary F. Feitosa, Krista Fischer, Ross M. Fraser, Anuj Goel, Jian Gong, Anne E. Justice, Stavroula Kanoni, Marcus E. Kleber, Kati Kristiansson, Unhee Lim, Vaneet Lotay, Julian C. Lui, Massimo Mangino, Irene Mateo Leach, Carolina Medina-Gomez, Michael A. Nalls, Dale R. Nyholt, Cameron D. Palmer, Dorota Pasko, Sonali Pechlivanis, Inga Prokopenko, Janina S. Ried, Stephan Ripke, Dmitry Shungin, Alena Stancáková, Rona J. Strawbridge, Yun Ju Sung, Toshiko Tanaka, Alexander Teumer, Stella Trompet, Sander W. van der Laan, Jessica van Setten, Jana V. Van Vliet-Ostaptchouk, Zhaoming Wang, Loïc Yengo, Weihua Zhang, Uzma Afzal, Johan Ärnlöv, Gillian M. Arscott, Stefania Bandinelli, Amy Barrett, Claire Bellis, Amanda J. Bennett, Christian Berne, Matthias Blüher, Jennifer L. Bolton, Yvonne Böttcher, Heather A. Boyd, Marcel Bruinenberg, Brendan M. Buckley, Steven Buyske, Ida H. Caspersen, Peter S. Chines, Robert Clarke, Simone Claudi-Boehm, Matthew Cooper, E. Warwick Daw, Pim A. De Jong, Joris Deelen, Graciela Delgado, Josh C. Denny, Rosalie Dhonukshe-Rutten, Maria Dimitriou, Alex S. F. Doney, Marcus Dörr, Niina Eklund, Elodie Eury, Lasse Folkersen, Melissa E. Garcia, Frank Geller, Vilmantas Giedraitis, Alan S. Go, Harald Grallert, Tanja B. Grammer, Jürgen Gräßler, Henrik Grönberg, Lisette C. P. G. M. de Groot, Christopher J. Groves, Jeffrey Haessler, Per Hall, Toomas Haller, Goran Hallmans, Anke Hannemann, Catharina A. Hartman, Maija Hassinen, Caroline Hayward, Nancy L. Heard-Costa, Quinta Helmer, Gibran Hemani, Anjali K. Henders, Hans L. Hillege, Mark A. Hlatky, Wolfgang Hoffmann, Per Hoffmann, Oddgeir Holmen, Jeanine J. Houwing-Duistermaat, Thomas Illig, Aaron Isaacs, Alan L. James, Janina Jeff, Berit Johansen, Åsa Johansson, Jennifer Jolley, Thorhildur Juliusdottir, Juhani Junttila, Abel N. Kho, Leena Kinnunen, Norman Klopp, Thomas Kocher, Wolfgang Kratzer, Peter Lichtner, Lars Lind, Jaana Lindström, Stéphane Lobbens, Mattias Lorentzon, Yingchang Lu, Valeriya Lyssenko, Patrik K. E. Magnusson, Anubha Mahajan, Marc Maillard, Wendy L. McArdle, Colin A. McKenzie, Stela McLachlan, Paul J. McLaren, Cristina Menni, Sigrun Merger, Lili Milani, Alireza Moayyeri, Keri L. Monda, Mario A. Morken, Gabriele Müller, Martina Müller-Nurasyid, Arthur W. Musk, Narisu Narisu, Matthias Nauck, Ilja M. Nolte, Markus M. Nöthen, Laticia Oozageer, Stefan Pilz, Nigel W. Rayner, Frida Renstrom, Neil R. Robertson, Lynda M. Rose, Ronan Roussel, Serena Sanna, Hubert Scharnagl, Salome Scholtens, Fredrick R. Schumacher, Heribert Schunkert, Robert A. Scott, Joban Sehmi, Thomas Seufferlein, Jianxin Shi, Karri Silventoinen, Johannes H. Smit, Albert Vernon Smith, Joanna Smolonska, Alice V. Stanton, Kathleen Stirrups, David J. Stott, Heather M. Stringham, Johan Sundström, Morris A. Swertz, Ann-Christine Syvänen, Bamidele O. Tayo, Gudmar Thorleifsson, Jonathan P. Tyrer, Suzanne van Dijk, Natasja M. van Schoor, Nathalie van der Velde, Diana van Heemst, Floor V. A. van Oort, Sita H. Vermeulen, Niek Verweij, Judith M. Vonk, Lindsay L. Waite, Melanie Waldenberger, Roman Wennauer, Lynne R. Wilkens, Christina Willenborg, Tom Wilsgaard, Mary K. Wojczynski, Andrew Wong, Alan F. Wright, Qunyuan Zhang, Dominique Arveiler, Stephan J. L. Bakker, John Beilby, Richard

N. Bergman, Sven Bergmann, Reiner Biffar, John Blangero, Dorret I. Boomsma, Stefan R. Bornstein, Pascal Bovet, Paolo Brambilla, Morris J. Brown, Harry Campbell, Mark J. Caulfield, Aravinda Chakravarti, Rory Collins, Francis S. Collins, Dana C. Crawford, L. Adrienne Cupples, John Danesh, Ulf de Faire, Hester M. den Ruijter, Raimund Erbel, Jeanette Erdmann, Johan G. Eriksson, Martin Farrall, Ele Ferrannini, Jean Ferrières, Ian Ford, Nita G. Forouhi, Terrence Forrester, Ron T. Gansevoort, Pablo V. Gejman, Christian Gieger, Alain Golay, Omri Gottesman, Vilmundur Gudnason, Ulf Gyllensten, David W. Haas, Alistair S. Hall, Tamara B. Harris, Andrew T. Hattersley, Andrew C. Heath, Christian Hengstenberg, Andrew A. Hicks, Lucia A. Hindorff, Aroon D. Hingorani, Albert Hofman, G. Kees Hovingh, Steve E. Humphries, Steven C. Hunt, Elina Hypponen, Kevin B. Jacobs, Marjo-Riitta Jarvelin, Pekka Jousilahti, Antti M. Jula, Jaakko Kaprio, John J. P. Kastelein, Manfred Kayser, Frank Kee, Sirkka M. Keinanen-Kiukaanniemi, Lambertus A. Kiemeney, Jaspal S. Kooner, Charles Kooperberg, Seppo Koskinen, Peter Kovacs, Aldi T. Kraja, Meena Kumari, Johanna Kuusisto, Timo A. Lakka, Claudia Langenberg, Loic Le Marchand, Terho Lehtimäki, Sara Lupoli, and Pamela A. F. Madden. 2014. "Defining the role of common variation in the genomic and biological architecture of adult human height." *Nature Genetics* 46 (11): 1173-1186. https://doi.org/10.1038/ng.3097. https://doi.org/10.1038/ng.3097.

Young, Alexander I, Seyed Moeen Nehzati, Chanwook Lee, Stefania Benonisdottir, David Cesarini, Daniel J Benjamin, Patrick Turley, and Augustine Kong. 2020. "Mendelian imputation of parental genotypes for genome-wide estimation of direct and indirect genetic effects." *BioRxiv*.