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

DISCUSSION PAPER SERIES

IZA DP No. 11544

Environmental Bottlenecks on Children's Genetic Potential for Adult Socioeconomic Attainments: Evidence from a Health Shock

Jason M. Fletcher

MAY 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11544

Environmental Bottlenecks on Children's Genetic Potential for Adult Socioeconomic Attainments: Evidence from a Health Shock

Jason M. Fletcher
University of Wisconsin-Madison and IZA

MAY 2018

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

Environmental Bottlenecks on Children's Genetic Potential for Adult Socioeconomic Attainments: Evidence from a Health Shock[1]

This paper explores gene-environmental interactions between family environments and children's genetic scores in determining educational attainment. The central question is whether poor childhood family environments reduce the ability for children to leverage their genetic gifts to achieve high levels of educational attainments. The multigenerational information and genetic data contained in the Health and Retirement Study is used to separate two mechanisms of intergenerational transmission of socioeconomic status – genetic endowments and family environments. Using parental in utero exposure to the 1918/1919 influenza pandemic as a source of quasi-experimental variation to family environments (but not affecting children's genetic endowments), this paper estimates interactions between parental investments and children's genetic potential. The main finding suggests that girls with high genetic potential whose fathers were exposed to influenza face reduced educational attainments – a gene-environment interaction – but there is no similar effect for boys.

JEL Classification: J62, J1, J24

Keywords: in utero exposure, gene-environment interactions, polygenic score, intergenerational effects

Corresponding author:
Jason M. Fletcher
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706
USA

E-mail: jason.fletcher@wisc.edu

[1] The Health and Retirement Study (HRS accession number 0925-0670) is sponsored by the National Institute on Aging (grant numbers NIA U01AG009740, RC2AG036495, and RC4AG039029) and is conducted by the University of Michigan. Additional funding support for genotyping and analysis were provided by NIH/NICHD R01 HD060726. I thank Dan Belsky, Jason Boardman, Dalton Conley, Ben Domingue, Florencia Torche and participants at the 2017 PAA meeting for helpful comments. I thank Ben Domingue for help with the polygenic score data.

Large literatures across many social science disciplines have sought to understand key determinants of children's socioeconomic status attainments during the transition to adulthood. Behavioral geneticists consistently report heritability estimates of educational attainment of approximately 40%, suggesting the substantial importance of genetic endowments for adult socioeconomic status.[2] However, a host of papers in sociology, economics, and associated disciplines have shown that environmental factors and policies also predict schooling outcomes (See Haveman and Wolfe (1995) for a review). In the past, scholars have debated whether genes or environments are more important in predicting socioeconomic status, but current research has shifted attention to the interplay between the two forces (e.g. Heckman 2007). This shift was anticipated by results from behavioral genetics that showed that estimates of heritability vary by the socioeconomic status of the family (e.g. Turkheimer et al. 2003), suggesting gene-environment interplay. Indeed, an important recent meta-analysis of twin studies showed evidence that genetic influences on educational attainment differed across countries and birth cohorts (Branigan et al., 2013).

While the importance of gene-environment interaction in determining socioeconomic status attainments is conceptually attractive, providing direct empirical evidence of interaction presents many challenges. Most studies are unable to leverage research designs that can separate gene-environment interaction effects from other processes. In particular, because parents contribute to both the genetic and environmental advantages/disadvantages of their children (labeled gene-environment correlation in the literature), it is difficult to separately estimate the main effects of genetics, environments, and their interactions. This is an important issue, as Belsky et al. (2016) have shown that advantageous genetic and environment factors are positively correlated. As discussed above, much of this work has not been able to measure genotype explicitly, relying instead on family based studies and comparisons of heritability estimates (genes) across measures of environment, such as parental socioeconomic status.

With newly available measures of genotype available in large social science surveys, a small number of studies have examined gene-environmental effects related to determinants of socioeconomic status, and very few have been able to use research designs aimed at uncovering

[2] Heritability estimates rely on comparisons using twin and family based samples rather than measured genetic data to provide evidence of genetic influence.

causal estimates[3]. Conley et al. (2015) estimate associations between parental education and child genotype using the Framingham Heart Study and Health and Retirement Study and find no evidence of interaction, though this analysis was not able to separate gene-environment interaction from gene-environment correlation[4].

This paper contributes to the literature exploring gene-environmental interaction determinants of socioeconomic status by leveraging quasi-experimental variation in parental environment (but unrelated to parental genotype) to estimate causal effects. Using the multigenerational framework of the Health and Retirement Study, this paper creates measures of in utero exposure to the 1918/1919 influenza pandemic by relying on respondent reports of parental birth years.

Focusing on this exposure follows from a large literature that has shown that in utero exposure created long term negative impacts on educational attainment, family income, and disability (Almond 2006, Almond and Mazumder 2005). Thus, parental exposure is used to measure a cascade of poor family environmental characteristics throughout childhood that are presumed to reduce their educational attainments. This paper extends this larger literature relating parental inputs in children's educational success to examine whether children whose parents experienced the shock are differentially affected based on the children's genetic scores related to educational attainment. The main finding is that parental in utero exposure reduces the ability of females who grew up in the 1940's to live up to their genetic potential—I find an important negative interaction between parental exposure and the child's genetic score in predicting the child's educational attainment. There is no evidence for gene-environmental interaction for male children in the data.

Data

This paper uses the Health and Retirement Study (HRS), a nationally representative, longitudinal panel study of individuals over the age of 50 and their spouses. Although the study

[3] Schmitz and Conley (2016) find that (birth date based) eligibility for the Vietnam draft interacts with a genetic score in predicting education attainments of men in the data.
[4] Interestingly, the authors did find evidence of interaction between the genetic score of the mother and genetic score of the child in determining the child's eventual educational attainment. However, without controls for the father's genotype, it is unclear how to interpret this interaction.

began in 1992, genetic data was first collected in 2006[5]. The HRS introduces a new cohort of participants every six years and interviews around 20,000 participants every two years. This paper uses a variety of information available across the various data waves. Respondents report their parents' ages (if living) and year of death/age of death of each parent in each survey, which I use to construct parental birth years. Additionally, respondents report parental education and own education as well as demographic measures (race, sex, etc).

For the purposes of this study, I link the Genotype Data Version 1 (2006 and 2008 samples) with the main HRS files. Since this paper uses findings from a GWAS conducted on individuals of European ancestry to construct the genetic scores, the analysis sample excludes respondents who report being of Hispanic, African, American Indian, Alaskan Native, Asian, or Pacific Islander ethnic origin (Carlson et al., 2013).[6] There are approximately 8200 individuals with genetic data and non-missing parental birth year information. This paper focuses the analysis on parental birth years in a window around the 1918/19 influenza pandemic (birth years 1908-1928) to eliminate longer term secular trends in education, which eliminates half the sample, leaving 4,100 observations. In the main analysis, I focus on respondents with data for parental birth years that are consistently reported across survey years, which eliminates about 450 observations. The final sample consists of over 2200 female and 1500 male respondents linked to their genotype and parental data. Table 1 presents summary statistics for the analysis sample. Appendix Table 1 presents summary statistics for samples across various data restrictions and Appendix Table 2 shows no evidence that the 1918/1919 paternal exposure indicator is related to the composition of the analysis sample.

Following the literature (Schmitz and Conley 2016) the key genetic score is calculated based on results from a genome wide association study (GWAS) meta-analysis of educational attainment by the Social Science Genetics Association Consortium (SSGAC) that aggregates

[5]See Domingue et al. (2016) for evidence of limited impacts of mortality selection in the HRS genetic sample. A worry for the current analysis is that individuals with higher PGS for education are more likely to survive and be included in the sample and thus genotype would be correlated with parental birth years (i.e. age). Results in Table 2 suggest limited evidence of gene-environment correlation in that parental birth year in 1918/19 is not statistically significantly linked with child PGS.

[6] There are several reasons to exclude non-white respondents. The current PGS are less predictive in non-white samples and the samples sizes of non-white respondents for these birth cohorts (1908-1928) are small, suggesting a limited ability to detect effects.

thousands of genetic markers in the form of single nucleotide polymorphisms (SNPs) that capture variation at single positions (loci) across the genome (Okbay et al., 2016). Each SNP value is weighted by the association between the SNP and educational attainment in the larger GWAS (of over 350,000 observations) and summed across over 1 million loci to create a single scalar measure that predicts education attainment for each person in the data; the measure is then standardized within the HRS sample to aid interpretation.

While the scalar measure captures only genetic information that has been shown to be predictive of educational attainment in the GWAS meta-analysis and the sample is limited to individuals who self-report white race, the genetic score might still capture subtle population differences across respondents who report white race (e.g. those with primarily Italian vs. German ancestry) and these population differences might also be related to educational attainment through social processes, such as patterns of migration and settlement. This source of confounding is called population stratification. This paper follows the literature and corrects for this source of confounding through adding controls for 20 genome-wide principal components of the genetic data.

Empirical Specification

The primary question of interest for the analysis is whether genotype interacts with environment to determine children's educational outcomes. The key quasi-experimental variation comes from respondents' parents in utero exposure to the 1918 influenza pandemic. Following much of the literature, I code birth years 1918 and 1919 as "exposed" because the pandemic occurred during the later quarter of 1918 and first half of 1919 so that a large proportion of individuals born in these years were at least partially exposed in utero. I estimate regressions of the following form (Almond 2006):

$$y_i = \beta_0 + \beta_1(Birthyear = 1918/1919) + \beta_2 Birthyear + \beta_3 Birthyear^2 + \beta_4 X_i + \varepsilon_i$$

where $y_i$is the outcome for individual i, (parental) birth year is entered as linear and quadratic terms, demographic controls (gender) are used, and there is an idiosyncratic shock(ε_i). The coefficient of interest is β_1 which estimates the "jump" in the outcome attributed to in utero exposure to the 1918/1919 influenza pandemic. Appendix Table 3 replicates results in the literature that used other samples (Almond 2006) showing reductions of a few months of schooling for the first generation (i.e. those directly effected).

Extending the literature, this paper examines multigenerational effects of exposure of the first generation on the second generation. The first analysis examines whether the genetic score of the second generation is affected by paternal exposure:

$$genetic\ score_i = \beta_0 + \beta_1(Birthyear = 1918/1919) + \beta_2 Birthyear + \beta_3 Birthyear^2 + \beta_4 X_i + \varepsilon_i$$

The final specification focuses on potential gene-environment interactions between first generation exposure and the second-generation genetic score in determining the second generation's educational attainment.

$$y_i = \beta_0 + \beta_1(Birthyear = 1918/1919) * GeneticScore + \beta_2 Birthyear + \beta_3 Birthyear^2 + \beta_4 X_i + \beta_5 GeneticScore + \beta_6(Birthyear = 1918/1919) * +\varepsilon_i$$

Results

Table 2 presents results for predicting the genetic score of the second generation (the HRS respondents). While the effect is positive, suggesting higher genetic scores for children of affected fathers, the result is imprecisely estimated. Column 2 adds a control for paternal education, which reduces the estimate on influenza exposure and is highly predictive of the child's genetic score (see also Appendix Table 2).

Table 3 presents the estimates of gene-environment interactions. Column 1 shows that paternal exposure has a small and imprecisely estimated effect on the second generation's educational attainment for the full sample. Column 2 examines gene-environment interaction in the full sample. While the genetic score is highly predictive of educational attainment, the interaction is negative but not statistically significant in the full sample. Columns 3 and 4 split the sample by the sex of the second generation. Column 3 suggests large gene-environment interactions for females in the second generation, while the effects of a one standard deviation increase in the genetic score is to increase educational attainment by nearly 2/3rds of a year, this effect is reduced by half for females with fathers who were exposed to influenza (See also Figure 1 and 2). Column 4 suggests no interaction for males in the sample but a similar effect of the genetic score. Column 5 shows no interactions with maternal exposure (in results not shown, there is no interaction for sex-stratified estimates).

Conclusion

This research note makes several contributions to the literature examining environmental shocks on long term outcomes. First, the use of the HRS data to examine the 1918 influenza pandemic is novel to the literature; in doing so, I replicate findings from Almond (2006) that show education reductions for exposed cohorts. This replication supports the use of adult children's reports of parental birth years and could be extended in future work on the mechanisms of the impact of the pandemic. Second, this paper is one of the first to examine the intergenerational impacts of the 1918 influenza pandemic on the second generation[7]. Finally, the key contribution of the paper is to leverage quasi-experimental variation in environmental exposure to poor conditions to examine gene-environment interactions predicting educational attainments in a national sample.

The results suggest that females who have high genetic scores but whose father was environmentally exposed attain less schooling than those who did not have affected fathers. These findings show how environmental factors can constrain genetic potential, particularly in an era (birth cohorts around 1940) when women faced many barriers to educational attainment. Future work should seek to explore the potential mechanisms behind these interaction effects. One possibility is that affected fathers had children with women with lower schooling than unaffected fathers, which lowered the genetic endowment of the next generation. The findings in Fletcher (2017) do not support this mechanism. A broader possibility is that affected fathers faced a series of health challenges over their life courses, documented in part by Almond (2006) and Almond and Mazumder (2005), that constrained the educational attainments of their daughters who would otherwise have achieved higher levels of schooling, potentially shifting daughters into caregiver and work roles that were not shifted to sons.

[7] Cook et al. (2016) uses the Wisconsin Longitudinal Study data to explore effects of the influenza pandemic across three generations.

References

Almond, Douglas. "Is the 1918 Influenza pandemic over? Long-term effects of in utero Influenza exposure in the post-1940 US population." *Journal of political Economy* 114, no. 4 (2006): 672-712.

Almond, Douglas, and Bhashkar Mazumder. "The 1918 influenza pandemic and subsequent health outcomes: an analysis of SIPP data." *American Economic Review* (2005): 258-262.

Belsky, D.W., Moffitt, T.E., Corcoran, D.L., Domingue, B., Harrington, H., Hogan, S., Houts, R., Ramrakha, S., Sugden, K., Williams, B.S. and Poulton, R., 2016. The Genetics of Success How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. Psychological science, p.0956797616643070.

Belsky, Jay, and Michael Pluess. "Beyond diathesis stress: differential susceptibility to environmental influences." Psychological bulletin 135, no. 6 (2009): 885.

Bettinger, Eric P. "Paying to learn: The effect of financial incentives on elementary school test scores." *Review of Economics and Statistics* 94, no. 3 (2012): 686-698.

Boardman, Jason D., and Jason M. Fletcher. "To cause or not to cause? That is the question, but identical twins might not have all of the answers." *Soc Sci Med* 127 (2015): 198-200.

Branigan, Amelia R., Kenneth J. McCallum, and Jeremy Freese. "Variation in the heritability of educational attainment: An international meta-analysis." *Social forces* 92, no. 1 (2013): 109-140.

Conley, Dalton, Benjamin W. Domingue, David Cesarini, Christopher Dawes, Cornelius A. Rietveld, and Jason D. Boardman. "Is the effect of parental education on offspring biased or moderated by genotype?." *Sociological Science* 2 (2015): 82-105

Cook, C. Justin, and Jason M. Fletcher. "Understanding heterogeneity in the effects of birth weight on adult cognition and wages." *Journal of health economics* 41 (2015): 107-116.

Cook, C. Justin, Jason M. Fletcher, and Angela Forgues. (2016) “Multigenerational Effects of Early Life Health Shocks.” Working Paper

Domingue, Benjamin W, Belsky, Daniel W, Harrati, Amal, Conley, Dalton, Weir, David, & Boardman, Jason. (2016). Mortality selection in a genetic sample and implications for association studies. *BioRxiv*, 049635.

Fletcher, Jason M. "Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts." *PLoS One* 7, no. 12 (2012): e50576.

Fletcher, Jason M. (2017) “The Effects of In Utero Exposure to the 1918 Influenza Pandemic on Family Formation" Working Paper

Fletcher, Jason M, & Conley, Dalton. (2013). The challenge of causal inference in gene–environment interaction research: Leveraging research designs from the social sciences. *American Journal of Public Health, 103*(S1), S42-S45

Fletcher, Jason M., and Steven F. Lehrer. "Genetic lotteries within families." *Journal of health economics* 30, no. 4 (2011): 647-659.

Haveman, Robert, and Barbara Wolfe. "The determinants of children's attainments: A review of methods and findings." *Journal of economic literature* 33, no. 4 (1995): 1829-1878

Heckman, James J. "The economics, technology, and neuroscience of human capability formation." *Proceedings of the national Academy of Sciences* 104, no. 33 (2007): 13250-13255

Okbay, Aysu, Beauchamp, Jonathan P, Fontana, Mark Alan, Lee, James J, Pers, Tune H, Rietveld, Cornelius A, . . . Meddens, S Fleur W. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539-542.

Papageorge, Nicholas W., and Kevin Thom. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." (2016). SSRN Working Paper: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2801921

Plug, Erik, and Wim Vijverberg. "Schooling, family background, and adoption: Is it nature or is it nurture?." *Journal of political Economy* 111, no. 3 (2003): 611-641

Price, Alkes L, Patterson, Nick J, Plenge, Robert M, Weinblatt, Michael E, Shadick, Nancy A, & Reich, David. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904-909.

Richter, André, and Per Olof Robling. "Multigenerational effects of the 1918-19 influenza pandemic in Sweden." *Swedish Institute for Social Research* 5 (2013).

Sacerdote, Bruce. "How large are the effects from changes in family environment? A study of Korean American adoptees." *The Quarterly Journal of Economics* (2007): 119-157.

Scarr-Salapatek S. (1971). Race, social class, and IQ. *Science*, 4, 1285– 1295.

Schmitz, Lauren L., and Dalton Conley. *The Effect of Vietnam-Era Conscription and Genetic Potential for Educational Attainment on Schooling Outcomes*. No. w22393. National Bureau of Economic Research, 2016.

Thompson, Owen. "Economic Background and Educational Attainment The Role of Gene-Environment Interactions." *Journal of Human Resources* 49, no. 2 (2014): 263-294.

Turkheimer, Eric, Andreana Haley, Mary Waldron, Brian D'Onofrio, and Irving I. Gottesman. "Socioeconomic status modifies heritability of IQ in young children." Psychological science 14, no. 6 (2003): 623-628.

Tables

Table 1
Descriptive Statistics for HRS Analysis Sample
Full Sample and Stratified by Gender

		Analysis Sample			
Variable	Obs	Mean	Std Dev	Min	Max
Education	3734	13.52	2.47	0	17
Genetic Score	3741	-0.05	0.99	-3.71	3.28
Female	3741	0.60	0.49	0	1
Age in 2006	3738	62.02	6.83	39	83
Height	3741	5.58	0.32	4.58	6.83
Paternal Birth Year	3741	1916	5.70	1908	1928
Paternal Flu Indicator	3741	0.09	0.29	0	1
Paternal Education	3460	10.73	3.40	0	17
Maternal Birth Year	3698	1919	7.37	1827	1946
Maternal Flu Indicator	3698	0.10	0.31	0	1
Maternal Education	3575	11.06	2.85	0	17
		Females			
Variable	Obs	Mean	Std Dev	Min	Max
Education	2226	13.35	2.38	0	17
Genetic Score	2229	-0.04	0.99	-3.32	3.05
Female	2229	1.00	0.00	1	1
Age in 2006	2227	61.66	6.79	39	82
Height	2229	5.38	0.21	4.58	6.25
Paternal Birth Year	2229	1916	5.75	1908	1928
Paternal Flu Indicator	2229	0.09	0.29	0	1
Paternal Education	2057	10.65	3.42	0	17
Maternal Birth Year	2210	1919	7.30	1827	1946
Maternal Flu Indicator	2210	0.11	0.31	0	1
Maternal Education	2141	10.99	2.85	0	17
		Males			
Variable	Obs	Mean	Std Dev	Min	Max
Education	1508	13.76	2.57	0	17
Genetic Score	1512	-0.06	1.00	-3.71	3.28
Female	1512	0.00	0.00	0	0
Age in 2006	1511	62.55	6.86	41	83
Height	1512	5.87	0.23	5.17	6.83
Paternal Birth Year	1512	1916	5.62	1908	1928
Paternal Flu Indicator	1512	0.09	0.29	0	1
Paternal Education	1403	10.86	3.37	0	17
Maternal Birth Year	1488	1918	7.47	1857	1938
Maternal Flu Indicator	1488	0.10	0.30	0	1
Maternal Education	1434	11.18	2.83	0	17

Table 2
Effects of Parental In Utero Exposure on Genetic Composition of Next Generation

VARIABLES	Genetic Score 20 year sample dad clean birth year genetic sample	Genetic Score 20 year sample dad clean birth year genetic sample	Genetic Score 20 year sample mom clean birth year genetic sample
Father Born in 1918/1919	0.074	0.063	
	(0.046)	(0.058)	
Father Birth Year (Centered)	0.001	-0.004	
	(0.003)	(0.003)	
Father Birth Year Squared	0.000	0.000	
	(0.001)	(0.001)	
Mother Born in 1918/1919			-0.015
			(0.075)
Mother Birth Year (Centered)			-0.005***
			(0.002)
Mother Birth Year Squared			-0.000
			(0.000)
Father Education		0.059***	
		(0.003)	
Constant	-0.054**	-0.678***	-0.032
	(0.024)	(0.047)	(0.024)
Observations	3,741	3,460	4,315
R-squared	0.000	0.041	0.001

Notes: 20 Year Sample Refers to Paternal Birth Cohorts between 1908-1928. Clean Birth Year refers to observations with consistent paternal birth year information. Genetic Sample refers to sample with genetic data available. PC controls not shown
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

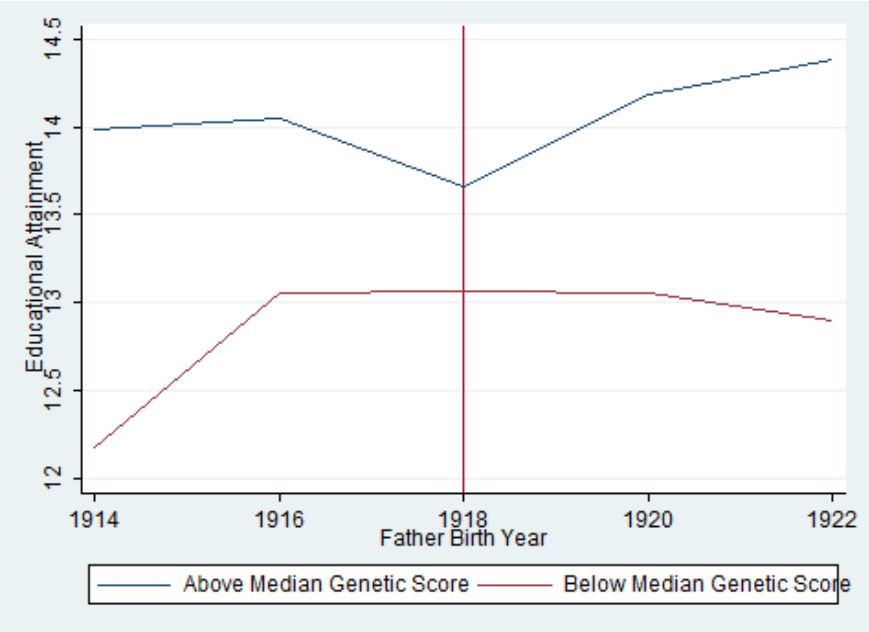
Table 3
Gene-Environment Interactions Determining Children's Educational Attainment
Genetic Scores X Paternal In Utero Exposure to 1918/19 Influenza Pandemic

VARIABLES	Education 20 year sample clean birth year genetic sample	Education 20 year sample clean birth year genetic sample	Education 20 year sample clean birth year female genetic sample	Education 20 year sample clean birth year male genetic sample	Education 20 year sample clean birth year genetic sample
Father Born in 1918/1919	-0.038	-0.083	-0.163	0.074	
	(0.088)	(0.090)	(0.106)	(0.086)	
Genetic Score (Standardized)		0.595***	0.636***	0.539***	0.620***
		(0.036)	(0.055)	(0.055)	(0.041)
Genetic Score X Father Born in 1918/1919		-0.156	-0.332**	0.104	
		(0.149)	(0.122)	(0.190)	
Father Birth Year (Centered)	-0.026***	-0.019***	-0.022**	-0.013	
	(0.006)	(0.006)	(0.008)	(0.016)	
Father Birth Year Squared	-0.001	-0.001*	-0.001	-0.002	
	(0.001)	(0.001)	(0.001)	(0.001)	
Female	0.128	0.083			0.114
	(0.106)	(0.099)			(0.127)
Age in 2006	-0.085***	-0.076***	-0.078***	-0.071***	-0.079***
	(0.006)	(0.005)	(0.007)	(0.013)	(0.007)
Height	1.252***	1.158***	1.162***	1.189***	1.274***
	(0.216)	(0.213)	(0.276)	(0.318)	(0.201)
Mother Born in 1918/1919					0.110
					(0.073)
Genetic Score X Mother Born in 1918/1919					0.007
					(0.041)
Mother Birth Year (Centered)					-0.012
					(0.008)
Mother Birth Year Squared					0.000
					(0.001)
Constant	11.756***	11.424***	9.666***	11.022***	10.783***
	(1.332)	(1.321)	(1.730)	(2.030)	(1.504)
Observations	3,731	3,731	2,224	1,507	4,304
R-squared	0.065	0.130	0.145	0.116	0.147

Notes: 20 Year Sample Refers to Paternal Birth Cohorts between 1908-1928. Clean Birth Year refers to observations with consistent paternal birth year information. Genetic Sample refers to sample with genetic data available. The first 20 principal components of genome wide data are controlled by not shown.
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

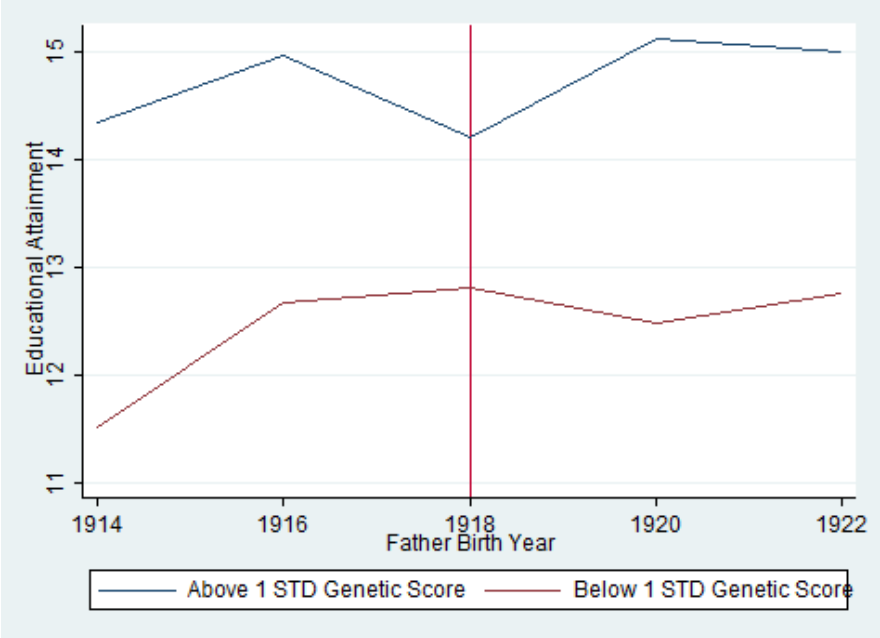
Figures

Figure 1
Effects of Paternal In Utero Influenza Exposure on Child's Educational Attainment
Results stratified by Median Split of Genetic Score



Notes: Birth Year Collapsed into two-year cells. Plots show unadjusted means.

Figure 2
Effects of Paternal In Utero Influenza Exposure on Child's Educational Attainment
Results stratified by +/- 1 Standard Deviation of Genetic Score



Appendix Tables

Appendix Table 1
Descriptive Statistics by Sample Selection Steps

		Non Missing Parent Birth Day			Paternal Birth Year Window	
Variable	Obs	Mean	Std Dev	Obs	Mean	Std Dev
Education	27716	12.46	3.25	12917	12.80	3.14
Genetic Score	8192	-0.02	1.00	4145	-0.05	1.00
Female	27823	0.55	0.50	12966	0.55	0.50
Age in 2006	27768	62.85	10.90	12943	60.31	8.25
Height	27823	5.57	0.33	12966	5.58	0.33
Missing Height Indicator	27823	0.01	0.09	12966	0.00	0.07
Paternal Birth Year	26101	1912	14	12966	1917	6
Paternal Flu Indicator	26101	0.05	0.21	12966	0.09	0.29
Paternal Education	22928	9.44	4.26	11384	9.85	4.15
Maternal Birth Year	27301	1917	13	12741	1920	8
Maternal Flu Indicator	27301	0.05	0.22	12741	0.09	0.28
Maternal Education	24812	9.70	3.92	11952	10.10	3.78
		Consistent Paternal Birth Day			Analysis Sample	
Variable	Obs	Mean	Std Dev	Obs	Mean	Std Dev
Education	11884	12.84	3.12	3734	13.52	2.47
Genetic Score	3741	-0.05	0.99	3741	-0.05	0.99
Female	11933	0.56	0.50	3741	0.60	0.49
Age in 2006	11910	60.07	8.32	3738	62.02	6.83
Height	11933	5.58	0.33	3741	5.58	0.32
Missing Height Indicator	11933	0.01	0.07	3741	0.00	0.00
Paternal Birth Year	11933	1917	6	3741	1916	6
Paternal Flu Indicator	11933	0.09	0.29	3741	0.09	0.29
Paternal Education	10443	9.91	4.13	3460	10.73	3.40
Maternal Birth Year	11726	1920	8	3698	1919	7
Maternal Flu Indicator	11726	0.09	0.28	3698	0.10	0.31
Maternal Education	10997	10.16	3.76	3575	11.06	2.85

Notes: Birth Year Window is 1908-1928.

Appendix Table 2
Associations between Paternal In Utero Exposure to 1918/1918 Influenza Pandemic and Children's Demographic Outcomes

Child's Outcome	Height	Height	Age	Female
	20 year sample clean birth year	20 year sample clean birth year	20 year sample clean birth year	20 year sample clean birth year
VARIABLES	genetic sample	full sample	genetic sample	genetic sample
Father Born in 1918/1919	-0.006	-0.003	-0.167	0.003
	(0.008)	(0.009)	(0.410)	(0.012)
Father Birth Year (Centered)	0.001	-0.001***	-0.718***	0.001
	(0.001)	(0.000)	(0.010)	(0.002)
Father Birth Year Squared	-0.000	0.000	0.005*	0.000
	(0.000)	(0.000)	(0.002)	(0.000)
Constant	5.587***	5.575***	60.738***	0.586***
	(0.008)	(0.006)	(0.155)	(0.012)
Observations	3,741	11,872	3,738	3,741
R-squared	0.000	0.000	0.368	0.001

Notes: 20 Year Sample Refers to Paternal Birth Cohorts between 1908-1928. Clean Birth Year refers to observations with consistent paternal birth year information. Genetic Sample refers to sample with genetic data available. This table shows a lack of association between the in utero exposure indicator and the child's adult height, age in 2006, and sex.
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 3
Replication of 1st Generation Impacts of In Utero Exposure
Examination of Differences by Analysis Sample

VARIABLES	Father Education 20 year sample clean birth year full sample	Father Education 20 year sample clean birth year genetic sample	Father Education 20 year sample 2SD birth year genetic sample	Father Education 20 year sample all birth year genetic sample
Father Born in 1918/1919	-0.074	-0.268**	-0.258*	-0.132
	(0.075)	(0.127)	(0.124)	(0.127)
Father Birth Year (Centered)	0.081***	0.077***	0.073***	0.073***
	(0.007)	(0.008)	(0.007)	(0.007)
Father Birth Year Squared	-0.001	-0.005**	-0.005***	-0.004***
	(0.001)	(0.002)	(0.002)	(0.001)
Constant	10.010***	11.049***	11.055***	10.979***
	(0.038)	(0.085)	(0.069)	(0.054)
Observations	10,443	3,460	3,641	3,839
R-squared	0.014	0.021	0.020	0.018

Notes: 20 Year Sample Refers to Paternal Birth Cohorts between 1908-1928. Clean Birth Year refers to observations with consistent paternal birth year information (birth year reports have variation less than 1 standard deviation). 2SD Birth Year refers to observations where the birth year reports have variation less than two standard deviations. Genetic Sample refers to sample with genetic data available.
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.