Published in final edited form as:

Popul Stud (Camb). 2019 March ; 73(1): 139–148. doi:10.1080/00324728.2018.1498533.

Environmental bottlenecks in children's genetic potential for adult socio-economic attainments: Evidence from a health shock

Jason M. Fletcher

University of Wisconsin–Madison

Abstract

This paper explores gene–environment interactions—interactions between family environments and children's genetic predispositions—in determining educational attainment. The central question is whether poor childhood family environments reduce children's ability to leverage their genetic gifts to achieve high levels of educational attainment—are there important 'bottlenecks' for poor children? The multigenerational information and genetic data contained in the United States' Health and Retirement Study are used to separate two mechanisms for intergenerational transmission of socio-economic status: genetic endowments and family environments. Using parental in utero exposure to the 1918–19 influenza pandemic as a source of quasi-experimental variation in family environments (that did not affect children's genetic endowments), I estimate 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.

Keywords

gene–environment interactions; multigenerational analysis; educational attainment; foetal origins; social mobility

Introduction

Large literatures across many social science disciplines have sought to understand key determinants of children's socio-economic status attainments during the transition to adulthood. Behavioural geneticists consistently estimate the heritability of educational attainment at approximately 40 per cent, suggesting that genetic endowments make a substantial contribution to adult socio-economic status. 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 socio-economic 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

¹Please direct all correspondence to Jason Fletcher, La Follette School of Public Affairs, Department of Sociology, Center for Demography and Ecology, Center for Demography of Health and Aging, University of Wisconsin–Madison, USA; or by jason.fletcher@wisconsin.edu.

behavioural genetics that showed estimates of heritability to vary by the socio-economic 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 differ across countries and birth cohorts (Branigan et al. 2013), and similar findings of heterogeneity have been found using molecular genetic measures (Tropf et al. 2017).

While the importance of gene–environment interaction in determining socio-economic status attainments is conceptually attractive, providing direct empirical evidence of interaction presents many challenges (Fletcher 2012; Fletcher and Conley 2013; Cook and Fletcher 2015). Most studies are unable to leverage research designs that can separate gene–environment interaction effects from other processes. This issue is, of course, not particular to gene–environment interactions, but stems from the difficulty of attempting to interpret empirical models with statistical interactions when the two interacting forces are not independent. In many examples of gene–environment interactions, the ‘environment’ is chosen purposely (high-quality schools, safe neighbourhoods, etc.) and thus these environments are likely to be correlated with attributes (i.e., genetic factors) of the parents (and children). This particular issue is labelled ‘gene–environment correlation’ in the literature and reflects a confounding process. More particular to the case of gene–environment interactions is that, because parents contribute to both the genetic and environmental advantages and disadvantages of their children, it is difficult to estimate the main effects of genetics, environments, and their interactions separately. This is an important issue, because Belsky et al. (2016) have shown broad evidence that advantageous genetic and environmental factors are positively correlated. Finally, 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 socio-economic status.

With newly available measures of genotype in large social surveys, a small number of studies have examined gene–environment effects related to determinants of socio-economic status, but very few have been able to use research designs aimed at uncovering causal estimates. One exception is Schmitz and Conley (2016), who found that (birthdate-based) eligibility for the Vietnam draft interacted with a genetic score in predicting educational attainments of men in the data. Conley et al. (2015) estimated associations between parental education and child genotype using the Framingham Heart Study and the Health and Retirement Study, and found no evidence of interaction, though this analysis was not able to separate gene–environment interaction from gene–environment correlation. Interestingly, the authors did find evidence of interaction between the genetic score of the mother and that 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.

This paper contributes to the literature exploring gene–environment interaction determinants of socio-economic status by leveraging quasi-experimental variation in the parental environment (that is unrelated to parental genotype) to estimate causal effects. Using the multigenerational framework of the Health and Retirement Study, I create measures of *in*

utero exposure to a health shock—the 1918–19 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 to the influenza pandemic created long-term negative impacts relating to educational attainment, family income, and disability (Almond and Mazumder 2005; Almond 2006). Thus, parental exposure is used to measure a cascade throughout childhood of poor family environmental characteristics that are presumed to reduce children's educational attainments. This paper extends the larger literature relating parental inputs in children's educational success by examining whether children whose parents experienced the shock are differentially affected, based on the children's genetic predispositions relating to educational attainment. The main finding is that parental *in utero* exposure reduced the ability of females who grew up in the 1940s to live up to their genetic potential—suggesting an environmental 'bottleneck'. I find an important negative interaction between parental exposure and the child's genetic predisposition in predicting girls' educational attainment. However, there is no evidence in the data for gene–environment interaction for male children.

Data

This paper uses the Health and Retirement Study (HRS), a nationally representative, longitudinal panel study in the United States (US), covering individuals aged 50 and over and their spouses. Although the study began in 1992, genetic data were first collected in 2006. 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. At every survey wave, respondents report their parents' ages (if living) and (if not) year of death and age at death of each parent; information that I use to construct parental birth years. Additionally, respondents report parental education and own education, as well as demographic measures (race, sex, etc.).

Because the HRS first collected genetic data in 2006, a concern for the current analysis is that individuals with higher polygenic scores (PGSs) 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., ages). Results shown later in Table 2 suggest limited evidence for gene–environment correlation, as links between parental birth year of 1918 or 1919 and child PGS are not statistically significant. On a related note, Domingue et al. (2016) found limited impacts of mortality selection in the HRS genetic sample.

For the purposes of this study, I link the Genotype Data Version 1 (an extra module of the HRS; 2006 and 2008 samples) with the main HRS files. Since this paper uses findings from a genome-wide association study (GWAS) conducted on individuals of European ancestry to construct the genetic scores (i.e., 'genetic predispositions'), the analysis sample excludes respondents who report being of Hispanic, African, American Indian, Alaskan Native, Asian, or Pacific Islander ethnic origin. There are several reasons for excluding non-white respondents. The current PGSs are less predictive of educational attainment in non-white

samples and the sample sizes of non-white respondents for these birth cohorts (1908–28) are small, suggesting a limited ability to detect effects.

There are approximately 8,200 individuals in the HRS with both genetic data and non-missing parental birth year information. This paper focuses the analysis on parental birth years in a 20-year window around the 1918–19 influenza pandemic (birth years 1908–28) to eliminate longer-term secular trends in education; this eliminates half the sample, leaving approximately 4,100 observations. In the main analysis, I focus on respondents whose data for parental birth years are consistently reported across survey years; this selection step eliminates about 400 observations. The final genetic sample consists of over 2,200 female and 1,500 male respondents linked to their genotype and parental data. Table 1 presents summary statistics for the analysis sample. Table A1 in the Appendix presents summary statistics for samples across the various selection steps described and Table A2 shows no evidence that the 1918–19 paternal exposure indicator is related to the composition of the analysis sample in terms of adult height, age, or sex.

Following the literature (Schmitz and Conley 2016), I calculated the key genetic score based on results from a GWAS meta-analysis of educational attainment by the Social Science Genetic Association Consortium (SSGAC), which aggregates 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 educational attainment for each person in the data; the measure is then standardized within the HRS sample to aid interpretation. A potential issue with using genetic scores that originate from GWAS findings is that, arguably, these scores will have lower power to detect gene–environment interactions because they represent the genetic variants across the genome that are most robust across environments (and therefore less sensitive to environmental interactions).

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). 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 1918–19 influenza pandemic. Following much of the literature, I code birth years 1918 and 1919 as ‘exposed’

because the pandemic occurred during the last quarter of 1918 and first half of 1919, so 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):

$$y_i = \beta_0 + \beta_1(\text{Birthyear} = 1918 \text{ or } 1919) + \beta_2\text{Birthyear} + \beta_3\text{Birthyear}^2 + \beta_4X_i + \varepsilon_i \quad (1)$$

where y_i is the educational attainment for individual i , (parental) birth year is entered as both 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 1918–19 influenza pandemic. Table A3 in the Appendix 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 affected).

Extending the literature, this paper examines the 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:

$$\begin{aligned} \text{genetic score}_i &= \alpha_0 + \alpha_1(\text{Birthyear} = 1918 \text{ or } 1919) \\ &\quad + \alpha_2\text{Birthyear} + \alpha_3\text{Birthyear}^2 \\ &\quad + \alpha_4X_i + u_i \dots \end{aligned} \quad (2)$$

This analysis is conducted in order to examine whether there is evidence of gene–environment correlation. Specifically, parental exposure to the influenza pandemic could have shaped marriage patterns. Taken together, Almond’s (2006) results showing reductions in years of schooling from influenza exposure and Mare and Schwartz’s (2006) findings on assortative mating patterns in education suggest that exposed individuals might have partners with lower education than unexposed individuals. This mechanism would suggest lower PGSs in the children from these partnerships.

The second specification focuses on potential gene–environment interactions between the first-generation exposure and the second-generation genetic score in determining the second generation’s educational attainment:

$$\begin{aligned} y_i &= \delta_0 + \delta_1(\text{Birthyear} = 1918 \text{ or } 1919) \\ &\quad \times \text{GeneticScore} + \delta_2\text{Birthyear} \\ &\quad + \delta_3\text{Birthyear}^2 + \delta_4X_i \\ &\quad + \delta_5\text{GeneticScore} \\ &\quad + \delta_6(\text{Birthyear} = 1918 \text{ or } 1919) + \zeta_i. \end{aligned} \quad (3)$$

Results

Table 2 presents results for predicting the genetic score of the second generation (the HRS respondents). While the effect in column (1) 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 for influenza exposure and is highly predictive of the child's genetic score. Column (3) finds no statistically significant relationship between mother's exposure and child's genetic score.

Table 3 presents the estimates of main effects of parental exposure as well as 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 pooled sample of male and female children combined. The set of controls is extended to include respondents' heights as a control for 'hardiness' in order to reduce the possibility of mortality selection further. Table A2 in the Appendix shows that height is not associated with exposure. Column (2) examines gene–environment interaction in the pooled sample. While the genetic score is highly predictive of educational attainment, the interaction is negative but not statistically significant.

Columns (3) and (4) of Table 3 split the sample by the sex of the second generation, based on evidence of differential intergenerational mobility of income for sons and daughters (Mazumder 2005). Torche (2015) also found differences in the strength of inter-generational persistence in socio-economic outcomes in Mexico, with boys having higher levels of persistence, while Fairbrother and Mahadevan (2016) found different patterns of mobility for sons and daughters in Australia. 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 two-thirds of a year, this effect is reduced by half for females with fathers who were exposed to influenza. This is illustrated further in Figures 1 and 2. Column (4) suggests no interaction for males in the sample but a similar effect of the genetic score as for females. 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 the impacts of environmental shocks on long-term outcomes. First, the use of the HRS data to examine the 1918–19 influenza pandemic is novel to the literature; in doing so, I replicate findings from Almond (2006) that showed reductions in educational attainment 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 1918–19 influenza pandemic on the second generation (see also Cook et al. 2016). Finally, the key contribution of the paper is to lever-age 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 with high genetic scores, but whose fathers were environmentally exposed, attain less schooling than those whose fathers were not affected. These findings show how environmental factors can constrain genetic potential, particularly in an era when women faced many barriers to educational attainment, as was the case for cohorts born around 1940. 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 (2018) do not support this mechanism. A broader possibility is that affected fathers faced a series of health challenges over their life courses, as documented in part by Almond and Mazumder (2005) and Almond (2006). These challenges may have constrained the educational attainments of their daughters—who would otherwise have achieved higher levels of schooling—potentially by shifting daughters into caregiving and work roles that were not shifted to sons.

acknowledgements

2 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 conducted by the University of Michigan. Additional funding support for genotyping and analysis was provided by NIH/ NICHD R01 HD060726.

3 I thank Dan Belsky, Jason Boardman, Dalton Conley, Ben Domingue, Florencia Torche, and participants at the 2017 Population Association of America meeting for helpful comments. I thank Ben Domingue for help with the polygenic score data and the Institute for Research on Poverty for editorial assistance. This work was also 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).

Appendix

Table A1

Descriptive statistics for HRS sample in the US, by sample selection steps

Variable	Observations	Mean	Standard deviation	Observations	Mean	Standard deviation
	<i>Sample with non-missing parental birth year</i>			<i>Sample included within paternal birth year window (1908–28)</i>		
Education (years)	27,716	12.46	3.25	12,917	12.80	3.14
Genetic score (standardized)	8,192	−0.02	1.00	4,145	−0.05	1.00
Sex (proportion female)	27,823	0.55	0.50	12,966	0.55	0.50
Age in 2006	27,768	62.85	10.90	12,943	60.31	8.25
Adult height (feet)	27,823	5.57	0.33	12,966	5.58	0.33
Missing height indicator (proportion)	27,823	0.01	0.09	12,966	0.00	0.07
Paternal birth year	26,101	1912	14	12,966	1917	6
Paternal influenza indicator (proportion)	26,101	0.05	0.21	12,966	0.09	0.29
Paternal education (years)	22,928	9.44	4.26	11,384	9.85	4.15
Maternal birth year	27,301	1917	13	12,741	1920	8
Maternal influenza indicator (proportion)	27,301	0.05	0.22	12,741	0.09	0.28
Maternal education (years)	24,812	9.70	3.92	11,952	10.10	3.78

Variable	Observations	Mean	Standard deviation	Observations	Mean	Standard deviation
	<i>Sample with consistent paternal birth year across surveys</i>			<i>Final analysis sample (with genetic score information)</i>		
Education (years)	11,884	12.84	3.12	3,734	13.52	2.47
Genetic score (standardized)	3,741	−0.05	0.99	3,741	−0.05	0.99
Sex (proportion female)	11,933	0.56	0.50	3,741	0.60	0.49
Age in 2006	11,910	60.07	8.32	3,738	62.02	6.83
Adult height (feet)	11,933	5.58	0.33	3,741	5.58	0.32
Missing height indicator (proportion)	11,933	0.01	0.07	3,741	0.00	0.00
Paternal birth year	11,933	1917	6	3,741	1916	6
Paternal influenza indicator (proportion)	11,933	0.09	0.29	3,741	0.09	0.29
Paternal education (years)	10,443	9.91	4.13	3,460	10.73	3.40
Maternal birth year	11,726	1920	8	3,698	1919	7
Maternal influenza indicator (proportion)	11,726	0.09	0.28	3,698	0.10	0.31
Maternal education (years)	10,997	10.16	3.76	3,575	11.06	2.85

Source: As for Table 1.

Table A2

Associations between paternal *in utero* exposure to the 1918–19 influenza pandemic and children's demographic outcomes in the US

<i>In utero</i> exposure variables	Child's outcome			
	Adult height (genetic sample)	Adult height (full sample)	Age in 2006 (genetic sample)	Sex (proportion female) (genetic sample)
Father born in 1918 or 1919	−0.006 (0.008)	−0.003 (0.009)	−0.167 (0.410)	0.003 (0.012)
Father's birth year (centred)	0.001 (0.001)	−0.001 *** (0.000)	−0.718 *** (0.010)	0.001 (0.002)
Father's birth year squared	−0.000 (0.000)	0.000 (0.000)	0.005 * (0.002)	0.000 (0.000)
Constant	5.587 *** (0.008)	5.575 *** (0.006)	60.738 *** (0.155)	0.586 *** (0.012)
Observations	3,741	11,872	3,738	3,741
R-squared	0.000	0.000	0.368	0.001

 $p < 0.01$

**
 $p < 0.05$

*
 $p < 0.1$.

Notes: All four samples shown refer to the 20-year sample (paternal birth cohorts born 1908–28) with consistent paternal birth year information. Genetic sample refers to sample with genetic data available, while the full sample includes those without genetic data available (but excludes those with missing height). This table shows a lack of association between the *in utero* exposure indicator and the child's adult height, age in 2006, or sex. Robust standard errors in parentheses.

Source: As for Table 1.

Table A3

Replication of first-generation impacts (father's education) of *in utero* exposure to the 1918–19 influenza pandemic in the US: examination of differences by analysis sample

<i>In utero</i> exposure variables	Father's education			
	Clean birth year (full sample)	Clean birth year (genetic sample)	2SD birth year (genetic sample)	All birth years (genetic sample)
Father born in 1918 or 1919	−0.074 (0.075)	−0.268 ** (0.127)	−0.258 * (0.124)	−0.132 (0.127)
Father's birth year (centred)	0.081 *** (0.007)	0.077 *** (0.008)	0.073 *** (0.007)	0.073 *** (0.007)
Father's birth year squared	−0.001 (0.001)	−0.005 ** (0.002)	−0.005 *** (0.002)	−0.004 *** (0.001)
Constant	10.010 *** (0.038)	11.049 *** (0.085)	11.055 *** (0.069)	10.979 *** (0.054)
Observations	10,443	3,460	3,641	3,839
R-squared	0.014	0.021	0.020	0.018

 $p < 0.01$

**

 $p < 0.05$

*

 $p < 0.1$.

Notes: All four samples shown refer to the 20-year sample (paternal birth cohorts born 1908–28). Clean birth year refers to observations with consistent paternal birth year information (birth year reports having variation of less than one standard deviation), while 2SD birth year refers to observations where the birth year reports vary by less than two standard deviations. Genetic sample refers to sample with genetic data available, while the full sample includes those without genetic data available. Robust standard errors in parentheses.

Source: As for Table 2.

References

- Almond Douglas. 2006 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(4): 672–712.
- Almond Douglas and Mazumder Bhashkar. 2005 The 1918 influenza pandemic and subsequent health outcomes: An analysis of SIPP data, *American Economic Review* 95(2): 258–262. [PubMed: 29125265]
- Belsky Daniel W., Moffitt Terrie E., Corcoran David L., Domingue Benjamin, Harrington HonaLee, Hogan Sean, ... Caspi Avshalom. 2016 The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development, *Psychological Science* 27(7): 957–972. [PubMed: 27251486]
- Branigan Amelia R., McCallum Kenneth J., and Freese Jeremy. 2013 Variation in the heritability of educational attainment: An international meta-analysis, *Social Forces* 92(1): 109–140.
- Conley Dalton, Domingue Benjamin W., Cesarini David, Dawes Christopher, Rietveld Cornelius A., and Boardman Jason D.. 2015 Is the effect of parental education on offspring biased or moderated by genotype?, *Sociological Science* 2: 82–105 [PubMed: 29051911]
- Cook C. Justin and Fletcher Jason M.. 2015 Understanding heterogeneity in the effects of birth weight on adult cognition and wages, *Journal of Health Economics* 41: 107–116. [PubMed: 25770970]
- Cook C. Justin, Fletcher Jason M., and Forgues Angela. 2016 Multigenerational effects of early life health shocks. Working Paper
- Domingue Benjamin W., Belsky Daniel W., Harrati Amal, Conley Dalton, Weir David, and Boardman Jason. 2016 Mortality selection in a genetic sample and implications for association studies, *International Journal of Epidemiology* 46(4): 1285–1294.

- Fairbrother David and Mahadevan Renuka. 2016 Do education and sex matter for intergenerational earnings mobility? Some evidence from Australia, *Australian Economic Papers* 55(3): 212–226.
- Fletcher Jason M. 2012 Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts, *PLoS ONE* 7(12): e50576. [PubMed: 23227187]
- Fletcher Jason M. 2018 The effects of in utero exposure to the 1918 influenza pandemic on family formation, *Economics & Human Biology* 30: 59–68. [PubMed: 29957334]
- Fletcher Jason M. and 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. [PubMed: 23927518]
- Haveman Robert and Wolfe Barbara. 1995 The determinants of children’s attainments: A review of methods and findings, *Journal of Economic Literature* 33(4): 1829–1878.
- Heckman James J. 2007 The economics, technology, and neuroscience of human capability formation, *Proceedings of the National Academy of Sciences* 104 (33): 13250–13255.
- Mare Robert D. and Schwartz Christine R.. 2006 Educational assortative mating and the family background of the next generation, *Sociological Theory and Methods* 21(2): 253–278.
- Mazumder Bhashkar. 2005 Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data, *Review of Economics and Statistics* 87(2): 235–255.
- Okbay Aysu, Beauchamp Jonathan P., Mark Alan Fontana James J. Lee, Pers Tune H., Rietveld Cornelius A., ... Benjamin Daniel J.. 2016 Genome-wide association study identifies 74 loci associated with educational attainment, *Nature* 533(7604): 539–542. [PubMed: 27225129]
- Schmitz Lauren L. and Conley Dalton. 2016 The Effect of Vietnam-Era Conscription and Genetic Potential for Educational Attainment on Schooling Outcomes Cambridge, MA: National Bureau of Economic Research Working Paper 22393.
- Torche Florencia. 2015 Intergenerational mobility and gender in Mexico, *Social Forces* 94(2): 563–587.
- Tropf Felix C., Lee S. Hong, Verweij Renske M., Stulp Gert, van der Most Peter J., de Vlaming Ronald, ... Mills Melinda C.. 2017 Hidden heritability due to heterogeneity across seven populations, *Nature Human Behaviour* 1(10): 757–765.
- Turkheimer Eric, Haley Andreana, Waldron Mary, D’Onofrio Brian, and Gottesman Irving L.. 2003 Socioeconomic status modifies heritability of IQ in young children, *Psychological Science* 14(6): 623–628. [PubMed: 14629696]

Figure 1.

Effects of paternal *in utero* exposure to the 1918–19 influenza pandemic on female children's educational attainment in the US: results stratified by median split of genetic score *Note:* Birth year collapsed into two-year cells. Plots show unadjusted means of each cell. *Source:* Author's calculations based on HRS data.

Figure 2.

Effects of paternal *in utero* exposure to the 1918–19 influenza pandemic on female children's educational attainment in the US: results stratified by ± 1 standard deviation (SD) of genetic score *Note:* Birth year collapsed into two-year cells. Plots show unadjusted means of each cell. *Source:* As for Figure 1.

Table 1
Descriptive statistics for HRS analysis sample: pooled sample of females and males in the US

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
<i>Pooled sample</i>					
Education (years)	3,734	13.52	2.47	0	17
Genetic score (standardized)	3,741	-0.05	0.99	-3.71	3.28
Sex (proportion female)	3,741	0.60	0.49	0	1
Age in 2006	3,738	62.02	6.83	39	83
Adult height (feet)	3,741	5.58	0.32	4.58	6.83
Paternal birth year	3,741	1916	5.70	1908	1928
Paternal influenza indicator (proportion)	3,741	0.09	0.29	0	1
Paternal education	3,460	10.73	3.40	0	17
Maternal birth year	3,698	1919	7.37	1827	1946
Maternal influenza indicator (proportion)	3,698	0.10	0.31	0	1
Maternal education	3,575	11.06	2.85	0	17
<i>Females</i>					
Education (years)	2,226	13.35	2.38	0	17
Genetic score (standardized)	2,229	-0.04	0.99	-3.32	3.05
Sex (proportion female)	2,229	1.00	0.00	1	1
Age in 2006	2,227	61.66	6.79	39	82
Adult height (feet)	2,229	5.38	0.21	4.58	6.25
Paternal birth year	2,229	1916	5.75	1908	1928
Paternal influenza indicator (proportion)	2,229	0.09	0.29	0	1
Paternal education (years)	2,057	10.65	3.42	0	17
Maternal birth year	2,210	1919	7.30	1827	1946
Maternal influenza indicator (proportion)	2,210	0.11	0.31	0	1
Maternal education (years)	2,141	10.99	2.85	0	17
<i>Males</i>					
Education (years)	1,508	13.76	2.57	0	17
Genetic score (standardized)	1,512	-0.06	1.00	-3.71	3.28
Sex (female)	1,512	0.00	0.00	0	0

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
Age in 2006	1,511	62.55	6.86	41	83
Adult height (feet)	1,512	5.87	0.23	5.17	6.83
Paternal birth year	1,512	1916	5.62	1908	1928
Paternal influenza indicator (proportion)	1,512	0.09	0.29	0	1
Paternal education (years)	1,403	10.86	3.37	0	17
Maternal birth year	1,488	1918	7.47	1857	1938
Maternal influenza indicator (proportion)	1,488	0.10	0.30	0	1
Maternal education (years)	1,434	11.18	2.83	0	17

Source: Author's calculations based on HRS.

Table 2

Effects of parental *in utero* exposure to the 1918–19 influenza pandemic on genetic score of second generation in the US

Variables	Genetic score		
	(1) Father's exposure	(2) Father's exposure and education	(3) Mother's exposure
Father born in 1918 or 1919	0.074 (0.046)	0.063 (0.058)	–
Father's birth year (centred)	0.001 (0.003)	–0.004 (0.003)	–
Father's birth year squared	0.000 (0.001)	0.000 (0.001)	–
Mother born in 1918 or 1919	–	–	–0.015 (0.075)
Mother's birth year (centred)	–	–	–0.005*** (0.002)
Mother's birth year squared	–	–	–0.000 (0.000)
Father's education	–	0.059*** (0.003)	–
Constant	–0.054** (0.024)	–0.678*** (0.047)	–0.032 (0.024)
Observations	3,741	3,460	4,315
R-squared	0.000	0.041	0.001

 $p < 0.01$ **
 $p < 0.05$ *
 $p < 0.1$.

Notes: All three samples shown refer to the 20-year sample (paternal birth cohorts born 1908–28) with consistent paternal birth year information and genetic data available. Principal component controls not shown. Robust standard errors in parentheses.

Source: As for Table 1.

Gene-environment interactions determining children's educational attainment in the US, using genetic scores $\times$ parental *in utero* exposure to 1918–19 influenza pandemic

Table 3

Variables	Children's educational attainment (years)				
	(1) All children (father's exposure)	(2) All children (father's exposure and gene-environment interaction)	(3) Female children (father's exposure and gene-environment interaction)	(4) Male children (father's exposure and gene-environment interaction)	(5) All children (mother's exposure and gene-environment interaction)
Father born in 1918 or 1919	-0.038 (0.088)	-0.083 (0.090)	-0.163 (0.106)	0.074 (0.086)	-
Genetic score (standardized)	-	0.595 *** (0.036)	0.656 *** (0.055)	0.539 *** (0.055)	0.620 *** (0.041)
Genetic score $\times$ father born in 1918 or 1919	-	-0.156 (0.149)	-0.332 ** (0.122)	0.104 (0.190)	-
Father's birth year (centred)	-0.026 *** (0.006)	-0.019 *** (0.006)	-0.022 ** (0.008)	-0.013 (0.016)	-
Father's birth year squared	-0.001 (0.001)	-0.001 * (0.001)	-0.001 (0.001)	-0.002 (0.001)	-
Female	0.128 (0.106)	0.083 (0.099)	-	-	0.114 (0.127)
Age in 2006	-0.085 *** (0.006)	-0.076 *** (0.005)	-0.078 *** (0.007)	-0.071 *** (0.013)	-0.079 *** (0.007)
Adult height	1.252 *** (0.216)	1.158 *** (0.213)	1.162 *** (0.276)	1.189 *** (0.318)	1.274 *** (0.201)
Mother born in 1918 or 1919	-	-	-	-	0.110 (0.073)
Genetic score $\times$ mother born in 1918 or 1919	-	-	-	-	0.007 (0.041)
Mother's birth year (centred)	-	-	-	-	-0.012 (0.008)
Mother's birth year squared	-	-	-	-	0.000 (0.001)
Constant	11.756 *** (1.332)	11.424 *** (1.321)	9.666 *** (1.730)	11.022 *** (2.030)	10.783 *** (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

 $p < 0.01$

Notes: All five samples shown refer to the 20-year sample (paternal birth cohorts born 1908–28) with consistent paternal birth year information and genetic data available. The first 20 principal components of genome-wide data are controlled but not shown. Robust standard errors in parentheses.

Source: As for Table 1.

**
*
 $p < 0.05$
 $p < 0.1$