Early Life Exposures, Gene-Environment Interactions, and Cognition in Old Age

Jason M. Fletcher (University of Wisconsin)
Atheendar S. Venkataramani (University of Pennsylvania)

December 27, 2018

Abstract:
Although there is a large literature linking early childhood exposures to childhood and adult outcomes, the causal evidence on how these exposures affect outcomes in late adulthood and the elderly years is limited. Moreover, the extent to which genetic factors modify the long-run consequences of these early exposures is not well understood. We examined the effects of early life exposure to pneumonia – a leading cause of infant death in the early 20th century – on cognitive outcomes among elderly adults. Leveraging the introduction of sulfonamide antibiotics in 1937 – which led to dramatic reductions in pneumonia morbidity and mortality – along with state-level differences in baseline disease rate – we find that infant exposure led to faster cognitive decline in adulthood. These effects were largest for individuals with higher genetic endowments (as measured by polygenetic scores (PGS) for cognition), and null for those with lower endowments. One interpretation of our finding is that, as environments are improved, those with genetic advantages are more fully able to leverage these improvements, which may increase inequality in cognitive performance over time.

Acknowledgements: The authors thank Amy Finkelstein, Adriana Lleras-Muney, Jonathan Skinner, Lauren Schmitz, and seminar participants at the NBER Summer Institute and NIA Biomarker Meeting for helpful comments and suggestions. Fletcher: jason.fletcher@wisc.edu. Venkataramani: atheenv@pennmedicine.upenn.edu.

1. Introduction

A large and robust literature in biology and economics has documented the long-run effects of early childhood experiences – including exposure to disease - on adult health and socioeconomic outcomes (Almond, Currie, and Duque, *Forthcoming*). The bulk of this literature has focused on health and socioeconomic outcomes in early childhood through the prime adult years. Less attention has been paid to the impacts of these experiences on outcomes in late adulthood and the elderly years (Case and Paxson, 2009) – with the bulk of this literature reporting descriptive associations.[1] In addition, while models of economic models of human capital formation posit that experiences and investments at different life stages may interact with each other in producing long-run health and well-being (Cunha and Heckman, 2007), the extent to which the impacts to early life investments vary by genetic endowments – both among the elderly and non-elderly - is not well understood. Providing insights into each of these areas is critical given the growing population of elderly worldwide, and emerging evidence of growing inequality in aging outcomes (Miller and Bairoliya 2017).

This study addresses these gaps. We first examine the effects of early life exposure to pneumonia – a leading cause of infant death in the early 20th century – on cognitive outcomes among elderly adults surveyed in the 2006-2010 U.S. Health and Retirement Survey. We focus on cognitive outcomes given its importance in the well-being of elderly, as well as emerging biomedical and epidemiological evidence of an association between early life experiences and age-related cognitive decline (Bale, 2015; Walhovd, et al 2016). To achieve causal identification, we leverage the introduction of sulfonamide antibiotics in 1937 – which led to dramatic reductions in pneumonia morbidity and mortality (Jayachandran, Lleras-Muney, and Smith, 2010) – along with state-level differences in baseline disease rates (Bhalotra and Venkataramani, 2015).

Focusing on cohorts born between 1920 and 1950, we find that exposure to pneumonia in the first year of life led to faster cognitive decline in late adulthood and the

[1] See Power, Kuh, and Morton (2013) for a review of the literature. Descriptive studies include Case and Paxson (2009) and Brandt, Deindl, and Hank (2012. Some notable quasi-experimental exceptions include Van der Berg, Lindeboom, Portrait (2006), McEniry and Palloni (2010), Brandt, Deindl, and Hank (2012), and Chang et al (2014).

post-retirement years. In particular, a 1 standard deviation increase in pneumonia mortality was associated with a 0.16 standard deviation decrease in the cognitive scores. The bulk of effects were driven by men, and robust to the inclusion of a number of state-year controls for disease environment, educational investments, and socioeconomic status; state-specific time trends; and use of narrower cohort windows. Moreover, they are likely a lower bound given the potential for mortality selection (Dominique et al 2017).

We then examine heterogeneity in impacts by using newly released genetic data in the HRS – specifically a polygenic score (PGS) strongly associated with cognitive performance (Rietveld et al 2014; Belsky and Israel, 2014; Davies et al, 2015; Ware et al, 2017) – to assess whether causal impacts vary with genetic predisposition. We find that effects were larger for individuals endowed with *higher* (above the median) PGS, while effect for individuals with low PGS were not significantly different from zero. The plausibility of these findings is demonstrated by null results for interactions using PGS for educational attainment, an attribute that is closely related to, but distinct from, cognition.

Our study makes several contributions to the literature. First, it provides rare causal evidence of the effects of early childhood health shocks on both adult cognition[2] and cognitive decline among elderly, specifically. The findings support Case and Paxson's (2009) seminal descriptive work which linked regional disease burdens at the time of birth on cognition in the same HRS data. Collectively, the findings demonstrate the importance of early childhood factors as a driver of cognitive performance in the final years of life.

Second, our study provides some of the first evidence on causal gene-environment interactions. Causal inference in this literature is challenged both by lack of exogenous variation (or selection into) exposure along with difficulties identifying appropriate candidate genes (which may reduce statistical power and challenges

[2] A number of studies link early life investments to short and medium-run cognitive outcomes, which are reviewed in Almond, Currie, and Duque (2017). Venkataramani (2012) is an exception in that it examines cognitive impacts of an early life health shock in adulthood. Another study that is important to highlight is, Adhvaryu et al (Forthcoming), who examine impacts of exposure to economic shocks early in life on a non-cognitive domain, adult mental health.

inference through multiple comparisons) (Fletcher and Conley, 2013). We address these issues by focusing on a well-known health policy shock to achieve exogenous variation in the broader disease environment and using newly available polygenic scores, which capture the contribution of many genes in explaining variation in the outcome of interest. Moreover, we conduct a highly challenging placebo check to rule out potential correlation between the cognitive PGS and other genetic and socioeconomic measures that may also modify the effects of early exposure to infant health shocks.

Third, the results inform models of human capital development and skill formation. One interpretation of our finding is that, as environments are improved, those with genetic advantages are more fully able to leverage these improvements, which increases inequalities over time. The empirical literature on the technology of skill formation has thus far has been focused on complementarity or substitutability of investments made in early life and thereafter on long-run outcomes – finding mixed results.[3] Our findings suggest that complementarity may exist between early life investments and genetic endowments. This has important implications for life cycle inequality – skill gaps that open up early in life due to differential investments may in fact be exacerbated by genetic predisposition.

The remainder of this paper is as follows. Section 2 describes the data. Section 3 discusses the empirical strategy and Section 4 presents the core findings and robustness checks. Section 5 concludes.

2. **Data**

We use data from the U.S. Health and Retirement Study (HRS), a nationally representative, longitudinal panel study of individuals over the age of 50 and their spouses. The HRS introduces a new cohort of participants every six years and interviews around 20,000 participants every two years. While the HRS has collected data on over 10,000 respondents beginning in 1992 (and refreshed in ongoing surveys),

[3] Recent work empirical examining complementarity and substitution across investments made at different life stages include Adhvaryu et al (2017), Duque, Rosales-Rueda, and Sanchez (2017), and Johnson and Jackson (2017). Almond, Currie, and Duque (2017) summarize this literature.

genetic data (from saliva samples) were first collected in 2006.[4] Together with additional collections in 2008 and 2010, the genetic subsample of HRS now has over 12,000 subjects (this number is soon expanding to over 20,000).

The genetic data is comprised of over 2.5 million genetic locations for each respondent, of the over 3 billion locations in the human genome[5]. The data reports, at each location, whether the respondent has an A, C, G, or T nucleotide. The 2.5 million locations were chosen to focus on places in the genome that differ in humans, "common variants", in at least 1% of the human populations and measure single nucleotide polymorphisms (SNP).[6] The 2.5 million locations in the HRS are then used to create a polygenic score for general cognition[7]. Following the literature (Ware et al. 2017), this score is created by weighing each nucleotide by the estimated beta coefficient linking each location with cognitive performance from a massive Genome Wide Association Study (GWAS) by Davies et al. (2015)[8]; the HRS respondents are removed so that the polygenic score is predicting out-of-sample.

Our outcome of interest are cognitive scores collected beginning in 1996. We use the RAND[9] summary measures (cogtot). The selected cognitive functioning measures include immediate and delayed word recall, the serial 7s test, counting backwards, naming tasks (e.g., date-naming), and vocabulary questions. In addition to the individual cognitive functioning measures, the HRS also derived three cognition indices, which summarizes the immediate and delayed word recall tasks. The mental status index sums scores from. The total cognition score we use sums the total recall

[4] See Domingue et al. (2017) for discussion of mortality selection in the genetic sample.
[5] These 2.5 million locations are expanded in a round of imputation to be over 21 million locations. http://hrsonline.isr.umich.edu/index.php?p=xxgen1&_ga=1.238849673.862524756.1380327234
[6] Recall, humans are estimated to be over 99.5% genetically identical to one another.
[7] See information about the polygenic scores available here: http://hrsonline.isr.umich.edu/index.php?p=shoavail&iyear=ZA&_ga=2.124445322.1863708380.1519752431-862524756.1380327234
[8] The PGSs for general cognition were created using results from a 2015 GWAS conducted across 31 cohorts by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. A total of 53,949 participants undertook multiple, diverse cognitive tests from which a general cognitive function phenotype was created within each cohort by principal component analysis. Thirteen genome-wide significant SNPs in three separate regions previously associated with neuropsychiatric phenotypes were reported
[9] The RAND HRS Data file is an easy to use longitudinal data set based on the HRS data. It was developed at RAND with funding from the National Institute on Aging and the Social Security Administration: https://www.rand.org/labor/aging/dataprod/hrs-data.html

and mental status indices (which includes counting, naming, and vocabulary tasks). In addition to the genetic data and cognitive data, we control for basic demographic variables, including age-at-survey, sex, and race.

Data for our key exposure – state-year pneumonia mortality rates – were obtained from Jayachandran, Lleras-Muney, and Smith (2010), who originally collected this data from U.S. vital statistics (Linder and Grove, 1947; Grove and Hetzel, 1968). We also obtained data for a number of state-year controls. From the above sources, we obtained data on mortality from other sulfa treatable conditions (maternal mortality, scarlet fever, meningitis) as well as other control diseases (tuberculosis, malaria, and diarrhea). Following Bhalotra and Venkataramani (2015), we used all-age pneumonia mortality given that it was predominantly driven by infant rates, but was less susceptible to inaccurate birth and death recording.[10] We obtained data on state-level physicians and pharmacists per capita, hospitals per capita, urbanization rates, illiteracy rates, per capita income, number of schools, and educational spending per capita from a variety of sources.[11] For all control variables, we calculate state means for the period 1930-1936 (pre sulfa drugs).

While the HRS has over 100,000 person-year observations for cognitive outcomes, we make a number of restrictions to create our analysis sample. First, we limit our analysis to respondents in the 2006, 2008, and 2010 waves because genetic information was first collected in 2006; this reduces the sample of person-years with cognitive measures to 37,000. We next limit our sample to birth years 1920-1950 to narrow our focus on the introduction and expansion of sulfa drugs, following previous work; this reduces our sample to approximately 30,000. Finally, we have some missing information in the historical data (pneumonia infant mortality rates between 1930-1936), which reduces the sample to approximately 25,000. Table 1 displays summary

[10] In their analysis of the long-run effects of early exposure to pneumonia, Bhalotra and Venkataramani (2015) show that their findings are unchanged – though somewhat noisier – when using infant pneumonia rates rather than all-age rates.

[11] See Appendix of Bhalotra and Venkataramani (2015) for details. These include: the Bureau of Economic Analysis (http://www.bea.gov/regional/spi/), the database maintained by Adriana Lleras-Muney (http://www.econ.ucla.edu/alleras/research/data.html), and a 5% sample of the 1930 U.S. Census (https://usa.ipums.org/usa/).

statistics for this analysis sample. The average age of the sample (at the time of the cognitive measurement) is 74 years old. We also note that some individuals did not provide DNA, so that our genetic sample is over 17,000. We show below (Table 2) that our pneumonia measures are not associated with whether an individual provides DNA.

3. Methods

To assess the causal effect of birth year exposure to pneumonia on old age cognition, we estimate versions of the following instrumental variable model:

$$Cognition_{i,j,c,t} = a_0 + a_1(PneumRate_{i,j,t}) + \boldsymbol{aX}_{i,j,c,t} + d_c + s_{i,j} + g_t + e_{i,j,t} \quad (1)$$

$$PneumRate_{i,j,c,t} = b_0 + b_1(Post_c)\left(BaselinePneumonia_{i,j}\right) + \boldsymbol{bX}_{i,j,c,t} + d_c + s_{i,j} + g_t + u_{i,j,t} \quad (2)$$

where *i* indexes the individual, *j* the state, *c* the birth cohort, and *t* the survey wave. *Cognition* represents the total cognition score described above; *PneumRate* represents all-age pneumonia mortality in the birth state and birth year; *Post* is an indicator = 1 if the individual was born after the arrival of sulfa drugs (1937 onwards), and *BaselinePneumonia* is the pre-sulfa average mortality rate (between 1930-1936) for the birth state. The terms $d_t,\ s_{i,j}, g_c$ represent birth cohort, birth state, and survey wave fixed effects.

Equation (1) and (2) leverage the sharp drop and convergence in pneumonia mortality rates with the arrival of sulfa drugs in 1937 to identify the causal effects of early exposure to pneumonia on cognition. That is, it assumes that states with higher burden of disease from pneumonia gained more from the arrival of sulfa drugs than states with lower burdens. This assumption is supported empirically. The sharp drop in pneumonia mortality nationwide starting in 1937 was demonstrated by Jayachandran, Lleras-Muney, and Miller (2010), and owes to the rapid uptake and diffusion of these agents.[12] The convergence across states by initial pneumonia mortality is illustrated by Bhalotra and Venkataramani (2015).

[12] This was in part enabled by the fact that sulfa drugs did not require prescriptions to obtain in the first few years of their existence. Hence, individuals could obtain these drugs from local pharmacists. There were also a number of different manufacturers of these agents. In 1939, because of highly publicized

Put differently, this empirical setup essentially leverages a "continuous" difference-in-differences strategy used widely in other work.[13] The exogenous variation we exploit is at the birth state (pre-existing disease burden) and cohort (birth relative to availability of sulfa drugs) level. The IV model allows us to address potential measurement error in pneumonia mortality. However, we also estimate reduced form (OLS), where the right-hand side of equation (2) is substituted into equation (1). Birth state and cohort fixed effects adjust for state-level time-invariant factors and secular trends in outcomes, respectively. For all models, we compute heteroscedasticity robust standard errors correcting for clustering at the state level.

The main threat to inference in our setup are birth state-birth year time varying unobserved factors. We address this in several ways. First, we include in the above models a rich set of baseline state attributes (interacted with post). These include measures of state socioeconomic status (per capita income, illiteracy, urbanization), which may both have affected the outcome and be correlated with the diffusion of sulfa drugs Jayachandran, Lleras-Muney, and Miller (2010); mortality from other sulfa and non-sulfa treatable disease (groups), including maternal mortality, meningitis, scarlet fever, malaria, tuberculosis, and diarrhea, to capture secular trends in the public health environment; availability of physicians, pharmacists, and hospitals per capita, to capture access to sulfa providers and medical care; and per capita spending on education. We show that the inclusion of these controls eliminates any pre-existing trends (as estimated in placebo specifications) in the outcome by baseline pneumonia mortality (Figure 1A).[14] Second, we test the sensitivity of our findings to the inclusion of state-specific linear time trends, which capture any differential pre-existing trends in the

deaths from some formulations, sulfa drugs became more tightly regulated. See Jayachandran, Lleras-Muney, and Miller (2008) for more details.

[13] This strategy was developed by Card (1992) in the labor economics literature and has since been used in a growing body of empirical work examining the consequences of large scale health interventions (e.g., Acemoglu and Johnson 2007, Bleakley 2007 and 2010, Cutler et al. 2010).

[14] Specifically, we estimate reduced form models using data for the 1925-1935 birth cohorts, all individuals who were over the age of 1 at the time of the policy. These individuals were less likely to benefit from sulfa drugs given the lower burden of disease from pneumonia among children above age 1. We estimate separate models assigning as exposed each birth year between 1931-1935 (placebo exposures). Without inclusion of the contextual controls, we find positive and significant estimates in several of the placebo tests. However, both the substantive and statistical significance of these estimates decline markedly with the inclusion of contextual controls.

outcome of interest that may covary with baseline pneumonia mortality. Third, we additionally assess the robustness of our models to narrower sample windows, to rule out the influence of other processes during the study era.

To estimate gene environment interactions, we first stratify HRS respondents by at the median of their cognitive PGS. We then estimate our core specifications for each of these samples – the high and low genetic endowment groups, respectively. Note that this is equivalent to estimating versions of equations (1) and (2) that fully interact all terms with the PGS – which we do, as well.

A key threat to inference here is that the PGS may be correlated with other genetic environmental factors that may influence the cognitive impacts of early exposure to pneumonia. For example, individuals with high PGS may be more likely to live in households where the parents are similarly endowed. These households may be better off with regard to other genetic endowment and socioeconomic status – and, consequently, estimates of heterogeneity in the impact of early life shocks by cognitive endowments may be picking up heterogeneity along other dimensions. To address this possibility, we conduct a placebo check where we assess heterogeneity with respect to a PGS score for another (related) outcome, education. This is a challenging check, as education is likely correlated with cognition. The absence of similar heterogeneity along this margin would speak to the specificity of the main interaction of interest.

A more general threat to inference is mortality selection – specifically, early exposure to pneumonia could lead to changes in cohort composition through any effect on life-cycle mortality risk. This would affect both the estimates on cognition as well estimates of interactions with genetic endowments. For the former, the bias is likely downwards; for the latter, the direction of bias is less clear.

4. **Results**

We begin our analysis be further exploring the composition of (i.e., selection into) our sample. We first test whether birth year pneumonia rates are associated with whether the respondent contributed DNA (saliva) for the PGS in 2006, 2008, or 2010. Although naïve results in Column 1 suggests such an association, the use of our strategy to focus on the introduction of sulfa drugs interacted with pre-drug levels of

pneumonia suggest no association (Column 2). Instrumenting birth year exposure with the sulfa drug X baseline interaction also suggests no detectable link with the provision of DNA data (Column 3 and first stage in Column 4). Similarly, we find no evidence of selection into the sample the IV models on key observable characteristics, including race, gender, and whether or not the respondent grew up in a poor household (Table 2A).

Panel 2 of Table 2 asks whether birth year exposures are linked with the genetic composition of our sample as measured by the cognitive polygenic score. Similar to our analysis of the provision on DNA, we find no statistically significant results. We interpret these results to suggest that, even if the environmental exposure has shaped the composition of the cohort through mortality selection, this culling is not related to the genetic predisposition of cognitive performance.

Panel 3 of Table 2 attempts to further examine the possibility of culling along another dimension. We examine whether adult height is associated with birth year pneumonia rates. Again, we find no statistically significant links, which suggests we can be less concerned about mortality selection processes before 2006.

Table 3 presents findings of associations between birth year exposure to pneumonia infant mortality rates and adult outcomes. Column 1 presents a baseline analysis with birth year fixed effects and shows that people born in places with higher pneumonia rates have lower old age cognition. Column 2 shows that the inclusion of state-of-birth fixed effects reduces the association by about 80%. Column 3 estimates the reduced form effects of the introduction of sulfa drugs interacted with baseline pneumonia rates and shows that people born in states that previously (before sulfa) had high pneumonia rates have higher adult cognition scores. A 1 per 1000 higher rate at baseline (about 5 SD increase) increases the cognitive index by 0.2 SD. Column 4 presents results from the instrumental variable approach. We find that lowering the birth year pneumonia rate by 1 per 1000 increases cognition by nearly 2 points (about 40% of SD).[15] The first stage F-statistic is strong (nearly 56).

[15] Appendix Table 3A shows that this result is similar if we add state-specific time trends as controls (Column 2), if we shorten the birth year window to include only 1930-1943 (Column 4), and if we limit the analysis to only those with genetic data (Column 6).

Prior work on the long-run effects of early exposure to sulfa drugs (Bhalotra and Venkataramani, 2015) examines heterogeneity by gender and income. Table 4 explores possible heterogeneity in this result. Results in Columns 1-4 suggest the impacts are larger for men. While we are unable to statistically distinguish the results for white vs. black respondents, the larger point estimates for black respondents is consistent with findings from prior work.

Table 5 stratifies our main results by genetic endowment in order to examine the possibility of gene-environment interaction between birth year exposure to pneumonia and polygenic scores predicting cognition. Comparing Columns 1 and 4, we see the reduced form effects of baseline (pre-sulfa) pneumonia for those born post-sulfa availability for individuals with "high" (above the median) polygenic scores versus those with low scores. Interestingly we find impacts only for those with advantageous genotypes. Similarly, comparison the IV estimates in Column 2 versus Column 5, we see effects for those with high polygenic scores but no effects for those with low scores. An interpretation of these findings is that, as environmental conditions improve, those individuals with genetic advantages appear more able to take advantage in terms of old age cognitive outcomes.[16] [17]

5. Discussion and Conclusion

Population aging worldwide – and particularly in high-income countries – has prompted new interest in factors influencing well-being and function in old age. While the role of early life factors in driving outcomes in old age has been posited and shown in descriptive work, causal evidence remains thin. Moreover, sources of growing inequality in well-being among older adults are not well understood, either.

Our results are among the first evidence of their kind in causally linking early life environments to an important determinant of welfare in old age – cognition. In addition, to demonstrating the long-reach of early childhood access to health technology, we also

[16] In unreported results, we find that the results are stronger if we control for maternal and paternal educational levels.

[17] Appendix Table 4A shows two additional robustness checks. Columns 1 and 2 show evidence that the results are specific to the cognitive PGS in that we find no reduced form differences if we instead split the sample by the education PGS. Columns 3 and 4 show that our results are robust to limiting the window of birth years to 1930-1943.

show that resulting improvements in the disease environment may actually increase inequality. Specifically, we find that individuals with better genetic endowments for life-time cognition were more sensitive to disease environments in infancy – and that secular improvements in the disease environments would have been more favorable for this group in the long-run. These novel results on gene-environment interactions also inform the literature on human capital production, specifically around complementarity between endowments and early life health investments.

References

Adhvaryu, Achyuta, James Fenske, and Anant Nyshadham. "Early life circumstance and adult mental health." Forthcoming, *Journal of Political Economy*.

Adhvaryu, Achyuta, Teresa Molina, Anant Nyshadham, and Jorge Tamayo. “Helping children catch up: early life shocks and the PROGRESA experiment.” *Mimeo,* 2017.

Almond, Douglas, Janet Currie, and Valentina Duque. “Childhood circumstance and adult outcomes: act II.” Forthcoming *Journal of Economic Literature*.

Bale, TL. Epigenetic and transgenerational reprogramming of brain development.” *Nature Reviews Neuroscience* 16, no 6 (2015): 332-344.

Belsky, Daniel W., and Salomon Israel. "Integrating genetics and social science: Genetic risk scores." *Biodemography and Social Biology* 60, no. 2 (2014): 137-155.

Bhalotra, Sonia R., and Atheendar Venkataramani. "Shadows of the captain of the men of death: Early life health interventions, human capital investments, and institutions." Mimeo, 2015.

Bleakley, Hoyt. “Disease and development: evidence from hookworm eradication in the American South.” *Quarterly Journal of Economics* 122, no 1 (2007): 73-117.

Bleakley, Hoyt. “Health, Human Capital, and Development.” *Annual Review of Economics* 2 (2010): 283-310.

Brandt, Martina, Christian Deindl, and Karsten Hank. “Tracing the origins of successful aging: the role of childhood conditions and social inequality in explaining later life health.” *Social Science and Medicine* 74, no. 9 (2012): 1418-1425.

Case, Anne and Christina Paxson. “Early life health and cognitive function in old age.” *American Economic Review: P&P* 99, no 2 (2009): 104-109.

Card, David. “Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage.” *Industrial and Labor Relations Review* 46, no 1 (1992): 22-3

Chang, Simon, Belton Fleisher, Seonghoon Kim, and Shinyung Liu. “Long-term health effects of malaria exposure around birth: evidence from colonial Taiwan.” *Economic Development and Cultural Change* 62, no. 3. (2014): 519-536.

Cunha, Flavio and James Heckman. “The technology of skill formation.” *American Economic Review*. 97, no 2. (2007): 31-47.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl. “Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India.” *American Economic Journal: Applied Economics* 2, no 2 (2010: 72-94.

Davies, Gail, N. Armstrong, Joshua C. Bis, Jan Bressler, Vincent Chouraki, Sudheer Giddaluru, Edith Hofer et al. "Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N= 53 949)." *Molecular Psychiatry* 20, no. 2 (2015): 183.

Domingue, Benjamin W., Daniel W. Belsky, Amal Harrati, Dalton Conley, David R. Weir, and Jason D. Boardman. "Mortality selection in a genetic sample and implications for association studies." *International Journal of Epidemiology* 46, no. 4 (2017): 1285-1294.

Duque, Valentina, Maria Rosales-Rueda, Fabio Sanchez. “Integrating early-life shocks and human-capital investments on children’s education.” *Mimeo*, 2017.

Fletcher, Jason and Dalton Conley. “The challenges of causal inference in gene-environment interaction research: leveraging research designs from the social sciences.” *American Journal of Public Health* 103, no S1 (2013): S42-S45.

Grove, Robert D and Hetzel, Alice M. *Vital Statistics in the United States 1940-1960.* Washington, DC: United States Government Printing Office (1968).

Jayachandran, Seema, Adriana Lleras-Muney, and Kimberly Smith. “Modern medicine and the twentieth century decline in mortality: evidence of the impact of sulfa drugs.” *American Economic Journal: Applied Economics* 2, no 2. (2010): 118-146.

Johnson, Rucker C., and C. Kirabo Jackson. “Reducing inequality through dynamic complementarity: evidence from Head Start and public school spending.” *NBER Working Paper No 23489.*

Linder, Forrest E and Robert D. Grove. *Vital Statistics in the United States 1900-1940.* Washington, DC: United States Government Printing Office (1947).

McEniry M and Alberto Palloni. “Early Life Exposures and the Occurrence and Timing of Heart Disease among the Older Adult Puerto Rican Population.” *Demography* 47, no 1 (2010): 23-43.

Miller, Ray and Neha Bairoliya. “Health, longevity, and welfare inequality of the elderly.” *Mimeo*, 2017.

Power C, D Kuh, and S. Morton. “From developmental origins of adult disease to life course research on adult disease and aging: insights from birth cohort studies.” *Annual Review of Public Health* 34, no 7. (203): 7-28.

Rietveld, Cornelius A., Tõnu Esko, Gail Davies, Tune H. Pers, Patrick Turley, Beben Benyamin, Christopher F. Chabris et al. "Common genetic variants associated with cognitive performance identified using the proxy-phenotype method." *Proceedings of the National Academy of Sciences* 111, no. 38 (2014): 13790-13794.

Venkataramani, Atheendar S. "Early life exposure to malaria and cognition in adulthood: evidence from Mexico." *Journal of Health Economics* 31, no. 5 (2012): 767-780.

Van den Berg, Gerard, Martin Lindeboom, France Portrait. “Economic Conditions Early in Life and Individual Mortality.” *American Economic Review* 96, no 1 (2006): 290-302.

Walhovd, Kristine B, Stine K. Krogsrud, Inge K. Amlien, Hauke Bartsch, Atle Bjornerud, Paulina Due-Tonnessen, Hakon Grydeland, Donald J. Hagler, Asta K Haberg, William S. Kremen, Lia Ferschmann, Lars Nyberg, Matthew Panizzon, Darius Rohani, Jon Skranes, Andreas B Storsve, Anne E. Solsnes, Christian K Tamnes, Wesley K Thomspon, Chase Reuter, Anders M. Dale, and Anders M. Fjell. “Neurodevelopmental origins of lifespan changes in brain and cognition.” *PNAS* 113, no 33 (2016): 9357-9362.

Ware, Erin B., Lauren L. Schmitz, Jessica D. Faul, Arianna Gard, Colter Mitchell, Jennifer A. Smith, Wei Zhao, David Weir, and Sharon LR Kardia. "Heterogeneity in polygenic scores for common human traits." *bioRxiv* (2017): 106062

Table 1
Summary Statistics
Health and Retirement Study, Birth Years 1920-1950
Survey Years 2006, 2008, 2010

Variable	Obs	Mean	Std Dev	Min	Max
Cognition	25,540	21.78	5.08	0	35
Female	25,540	0.58	0.49	0	1
Age at Survey	25,540	74.12	6.36	56	90
Black	25,540	0.13	0.34	0	1
White	25,540	0.85	0.36	0	1
State-Level Birth Year Pneumonia Infant Mortality Rate (1000s)	23,977	1.00	0.35	0.2	2.5
Baseline (1930-1936) Pneumonia IMR	25,540	1.07	0.19	0.8	1.5
Post Sulfa Indicator (Birth Year >1936)	25,540	0.40	0.49	0	1
Birth Year	25,540	1933.90	6.49	1920	1950
Birth State	25,540	30.48	14.06	4	56
Baseline Rate X Post Sulfa	25,540	0.43	0.54	0	1.5
Polygenic Score for Cognition	17,334	0.01	1.00	-3.8	3.7
Polygenic Score for Education	17,334	0.02	1.00	-3.7	3.4
Contextual Variables (Baseline (1930-1936), State Level):					
Complete Variable	25,540	0.42	0.41	0	1.0
Maternal Mortality Rate	25,540	2.67	2.69	0	8.2
Diarrhea Rate	25,540	3.29	3.65	0	15.5
Malaria Rate	25,540	49.83	90.72	0	423.1
Tuberculosis Rate	25,540	0.26	0.27	0	0.9
Scarlet Fever Rate	25,540	0.00	0.00	0	0.0
Meningitis Rate	25,540	0.00	0.00	0	0.0
Physicians Per 1,000	25,540	0.52	0.54	0	1.7
Pharmacists per capita	25,540	0.33	0.33	0	1.0
Hospitals per capita	25,540	0.02	0.02	0	0.1
Per Capita Income	25,540	192.67	205.22	0	681.6
Urbanization Rate	25,540	0.25	0.27	0	0.9
Illiteracy Rate	25,540	0.02	0.03	0	0.1
Number of Schools per capita	25,540	0.95	1.12	0	6.9
Missing Contextual Information	25,540	0.20	0.40	0	1

Table 2

Birth Year Pneumonia Exposure and Sample Composition
Associations with Genetic Information Availability, Genetic Scores and Adult Height

Outcome	Has DNA Info	Has DNA Info	Has DNA Info	Has DNA Info
Fixed Effects	BY, S	BY, S	BY, S	BY, S
Specification	OLS	RF	IV	First Stage
Birth Year Pneumonia Rate X 1000	-0.085**		-0.036	
	(0.034)		(0.133)	
Pre Pneumonia Rate X Post Sulfa interaction		0.020		-0.543***
		(0.073)		(0.073)
Age at Survey	0.009***	0.009***	0.009***	0.000
	(0.001)	(0.001)	(0.001)	(0.000)
Black	-0.030	-0.029	-0.029	-0.004
	(0.026)	(0.026)	(0.025)	(0.004)
Observations	23,977	23,977	23,977	23,977
R-squared	0.058	0.058	0.013	0.880
Number of birth_state			47	47
F			55.96	
Outcome	Cognitive PGS	Cognitive PGS	Cognitive PGS	Cognitive PGS
Fixed Effects	BY, S	BY, S	BY, S	BY, S
Birth Year Pneumonia Rate X 1000	-0.005		-0.149	
	(0.131)		(0.277)	
Pre Pneumonia Rate X Post Sulfa interaction		0.078		-0.524***
		(0.149)		(0.078)
Age at Survey	-0.001	-0.001	-0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.000)
Black	0.148***	0.148***	0.148***	-0.004
	(0.051)	(0.051)	(0.050)	(0.005)
Observations	16,311	16,311	16,311	16,311
R-squared	0.024	0.024	0.010	0.873
F			45.63	
Outcome	Height	Height	Height	Height
Fixed Effects	BY, S	BY, S	BY, S	BY, S
Birth Year Pneumonia Rate X 1000	0.011		0.005	
	(0.019)		(0.053)	
Pre Pneumonia Rate X Post Sulfa interaction		-0.003		-0.542***
		(0.029)		(0.072)
Age at Survey	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Black	0.003	0.003	0.003	-0.004
	(0.013)	(0.013)	(0.012)	(0.004)

Observations	23,967	23,967	23,967	23,967
R-squared	0.555	0.555	0.551	0.880
F			56.10	

Notes: BY: Birth Year Fixed Effects, S: State of Birth Fixed Effects, RF: Reduced Form Specification. Additional Controls: Constant, Female Indicator,
Robust standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3

Association between Birth Year Pneumonia IMR and Old Age Cognition

Outcome	Cognition	Cognition	Cognition	Cognition	Cognition
Fixed Effects	B	B, S FE	B, S FE	B, S FE	B, S FE
Specification	OLS	OLS	RF	IV	First Stage
Birth Year Pneumonia Rate X 1000	-1.546***	-0.217		-1.889**	
	(0.322)	(0.307)		(0.800)	
Pre Pneumonia Rate X Post Sulfa interaction			1.025**		-0.543***
			(0.471)		(0.073)
Female	0.777***	0.790***	0.789***	0.789***	0.000
	(0.090)	(0.089)	(0.089)	(0.088)	(0.002)
Age at Survey	-0.286***	-0.287***	-0.287***	-0.287***	0.000
	(0.021)	(0.021)	(0.021)	(0.021)	(0.000)
Black	-3.583***	-3.124***	-3.124***	-3.131***	-0.004
	(0.243)	(0.248)	(0.247)	(0.246)	(0.004)
Birth Year Post Sulfa (1937+)			-6.414***	-7.484*	-0.365
			(1.699)	(3.894)	(0.277)
Constant	41.183***	37.924***	43.605***		1.237***
	(3.404)	(2.054)	(1.486)		(0.034)
Observations	23,977	23,977	23,977	23,977	23,977
R-squared	0.148	0.164	0.164	0.125	0.880
Number of birth_state				47	47
F				55.96	

Notes: B: Birth Year Fixed Effects, S: State of Birth Fixed Effects, RF: Reduced Form Specification. Additional Controls: Contextual variables

Robust standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4

Association between Birth Year Pneumonia IMR and Old Age Cognition
Heterogeneity by Sex and Race

Outcome	Cognition	Cognition	Cognition	Cognition	Cognition	Cognition	Cognition	Cognition
Sample	Male	Male	Female	Female	White	White	Black	Black
Fixed Effects	B, S FE	B, S FE	B, S FE	B, S FE	B, S FE	B, S FE	B, S FE	B, S FE
Specification	RF	IV	RF	IV	RF	IV	RF	IV
Birth Year Pneumonia Rate X 1000		-2.868***		-0.994		-0.557		-4.688
		(1.096)		(1.129)		(1.016)		(4.213)
Pre Pneumonia Rate X Post Sulfa interaction	1.475**		0.563		0.299		2.906	
	(0.611)		(0.663)		(0.573)		(2.596)	
Female					0.837***	0.837***	0.692**	0.673***
					(0.092)	(0.091)	(0.263)	(0.258)
Age at Survey	-0.300***	-0.299***	-0.278***	-0.278***	-0.301***	-0.301***	-0.206***	-0.204***
	(0.027)	(0.027)	(0.023)	(0.023)	(0.022)	(0.022)	(0.047)	(0.046)
Black	-2.892***	-2.896***	-3.261***	-3.266***				
	(0.324)	(0.316)	(0.239)	(0.238)				
Birth Year Post Sulfa (1937+)	-0.318	6.180	-10.503***	-16.539***	-6.766***	-7.903*	2.179	2.368
	(2.348)	(4.915)	(2.192)	(4.773)	(2.160)	(4.439)	(11.409)	(28.055)
Constant	44.380***		43.846***		44.681***		34.259***	
	(1.910)		(1.652)		(1.574)		(3.513)	
Observations	10,229	10,229	13,748	13,748	20,339	20,339	3,131	3,130
R-squared	0.153	0.108	0.175	0.134	0.117	0.105	0.144	0.086
Number of birth_state		47		47		47		31
F		50.55		58.49		46.74		71.88

Notes: B: Birth Year Fixed Effects, S: State of Birth Fixed Effects, RF: Reduced Form Specification. Additional Controls: Contextual Variables

Robust standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5

Genetic Heterogeneity of the Impacts of Birth Year Pneumonia IMR on Old Age Cognition

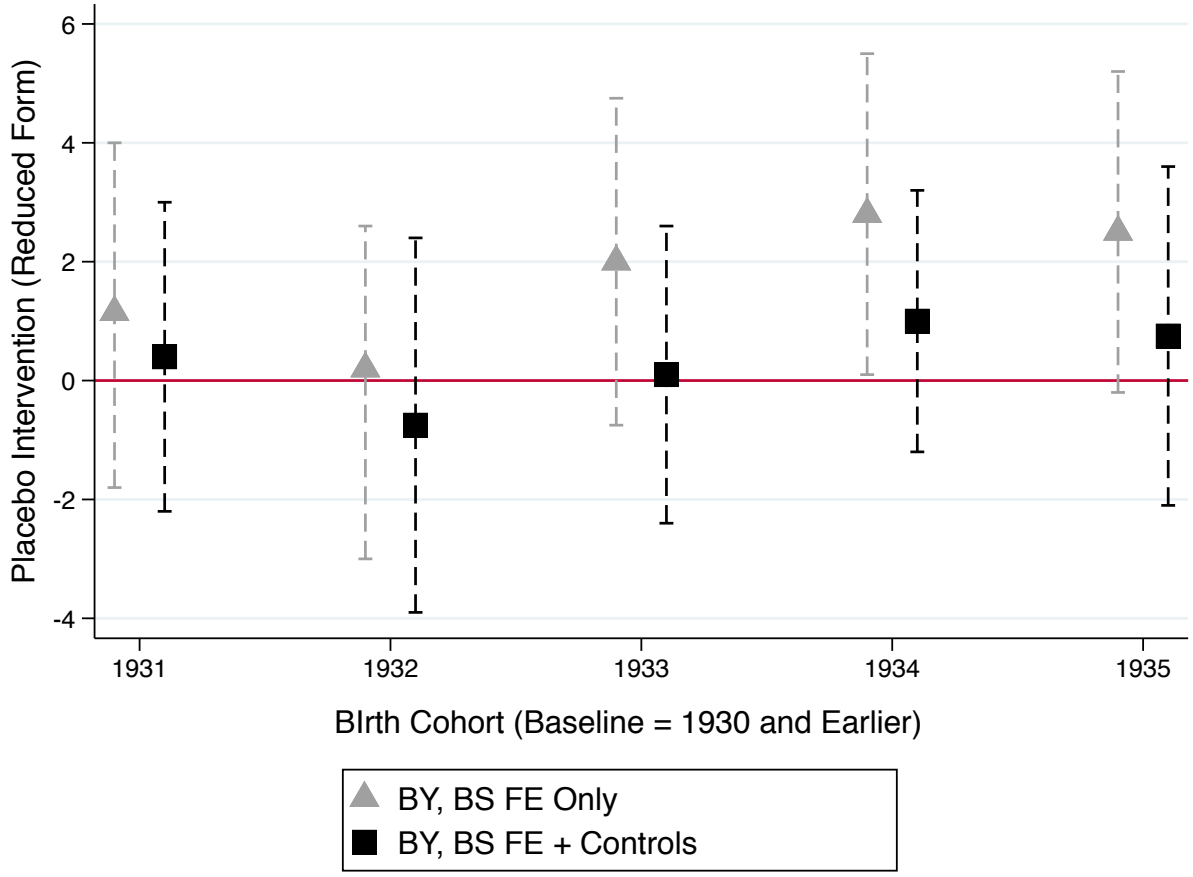
Outcome	Cognition	Cognition	Cognition	Cognition	Cognition	Cognition
Sample	High PGS	High PGS	High PGS	Low PGS	Low PGS	Low PGS
Fixed Effects	B, S FE	B, S FE	B, S FE	B, S FE	B, S FE	B, S FE
Specification	RF	IV	First Stage	RF	IV	First Stage
Birth Year Pneumonia Rate X 1000		-4.231***			0.378	
		(1.452)			(1.677)	
Pre Pneumonia Rate X Post Sulfa interaction	2.293***		-0.542***	-0.188		-0.496***
	(0.795)		(0.086)	(0.835)		(0.072)
female	0.992***	1.005***	0.003	0.903***	0.903***	-0.000
	(0.129)	(0.130)	(0.003)	(0.191)	(0.188)	(0.004)
Age at Survey	-0.330***	-0.330***	0.000	-0.333***	-0.333***	0.000
	(0.024)	(0.024)	(0.000)	(0.026)	(0.025)	(0.000)
black	-3.264***	-3.264***	-0.000	-4.058***	-4.055***	-0.008
	(0.269)	(0.278)	(0.008)	(0.321)	(0.318)	(0.007)
Birth Year Post Sulfa (1937+)	-5.369	-6.075	-0.349	-5.966	-10.625	-0.434
	(3.303)	(7.434)	(0.315)	(4.568)	(9.992)	(0.288)
Constant	46.802***		1.225***	47.164***		1.240***
	(1.678)		(0.041)	(1.770)		(0.038)
Observations	8,271	8,271	8,271	8,040	8,040	8,040
R-squared	0.197	0.137	0.871	0.197	0.155	0.878
Number of birth_state		47	47		47	47
F		39.70			47.94	

Notes: B: Birth Year Fixed Effects, S: State of Birth Fixed Effects, RF: Reduced Form Specification. Low PGS: Respondents with Cognitive Polygenic Scores<0. High PGS: Respondents with Cognitive Polygenic Scores>0. Additional Controls: Contextual Variables

Robust standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix

Figure 1A
Placebo Specifications and Pre-Trends



Notes: B: Birth Year Fixed Effects, S: State of Birth Fixed Effects, Controls: Contextual Variables. Estimates are from RF models for the 1925-1935 birth cohorts. Each point estimate is for a regression assigning Post to be a placebo treatment year (denoted on the x-axis). 95% CI computed based on standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1A

Summary Statistics for Genetic Subsample

Variable	Obs	Mean	Std Dev	Min	Max
Cognition	17,334	22.10	4.87	0	35
Female	17,334	0.57	0.49	0	1
Age at Survey	17,334	74.27	6.26	56	90
Black	17,334	0.13	0.33	0	1
White	17,334	0.87	0.34	0	1
State-Level Birth Year Pneumonia Infant Mortality Rate (1000s)	16,311	1.00	0.34	0.2	2.5
Baseline (1930-1936) Pneumonia IMR	17,334	1.06	0.18	0.8	1.5
Post Sulfa Indicator (Birth Year >1936)	17,334	0.39	0.49	0	1
Birth Year	17,334	1933.78	6.33	1920	1950
Birth State	17,334	30.59	13.94	4	56
Baseline Rate X Post Sulfa	17,334	0.42	0.53	0	1.5
Polygenic Score for Cognition	17,334	0.01	1.00	-3.8	3.7
Polygenic Score for Education	17,334	0.02	1.00	-3.7	3.4
Complete Variable	17,334	0.41	0.41	0	1.0
Maternal Mortality Rate	17,334	2.62	2.69	0	8.2
Diarrhea Rate	17,334	3.19	3.58	0	15.5
Malaria Rate	17,334	47.63	88.33	0	423.1
Tuberculosis Rate	17,334	0.26	0.27	0	0.9
Scarlet Fever Rate	17,334	0.00	0.00	0	0.0
Meningitis Rate	17,334	0.00	0.00	0	0.0
Physicians Per 1,000	17,334	0.51	0.54	0	1.7
Pharmacists per capita	17,334	0.32	0.33	0	1.0
Hospitals per capita	17,334	0.02	0.02	0	0.1
Per Capita Income	17,334	187.91	202.89	0	681.6
Urbanization Rate	17,334	0.25	0.26	0	0.9
Illiteracy Rate	17,334	0.02	0.03	0	0.1
Number of Schools per capita	17,334	0.96	1.16	0	6.9
Missing Contextual Information	17,334	0.19	0.39	0	1

Table 2A

Selection into the HRS Sample on Observables

	(1) P(Sample in 2010)	(2) P(White)	(3) P(Female)	(4) P(Poor Child)
Birth Year Pneumonia Rate X 1000	-0.036	0.082	-0.110	-0.147
	(0.133)	(0.099)	(0.118)	(0.122)
Observations	23,977	23,977	23,977	23,977
R-squared	0.013	0.245	0.011	0.033
F	55.96	55.76	55.76	55.76

Notes: All models include birth state, birth year, and survey year FE. Birth Year Pneumonia Rate instrumented by baseline pneumonia mortality X post. Additional Controls: Contextual Variables. Robust standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3A

Robustness of Associations between Birth Year Pneumonia IMR and Old Age Cognition
Adding State-Specific Time Trends and Sample Selection Considerations

Outcome	Cognition	Cognition	Cognition	Cognition	Cognition	Cognition	Cognition
Fixed Effects	B, S, Trends	B, S, Trends	B, S, Trends	B, S FE	B, S FE	B, S FE	B, S FE
Sample	Full	Full	Full	Short Window	Short Window	Genetic	Genetic
Specification	RF	IV	First Stage	IV	First Stage	IV	First Stage
Birth Year Pneumonia Rate X 1000		-1.544*		-1.508		-2.024*	
		(0.926)		(1.102)		(1.033)	
Pre Pneumonia Rate X Post Sulfa interaction	0.828		-0.536***		-0.449***		-0.524***
	(0.532)		(0.071)		(0.062)		(0.078)
Female	0.787***	0.787***	0.000	0.924***	-0.002	0.934***	0.002
	(0.091)	(0.090)	(0.002)	(0.093)	(0.002)	(0.101)	(0.003)
Age at Survey	-0.287***	-0.287***	0.000	-0.260***	-0.000	-0.331***	0.000
	(0.021)	(0.021)	(0.000)	(0.022)	(0.000)	(0.020)	(0.000)
Black	-3.160***	-3.165***	-0.003	-3.187***	-0.000	-3.674***	-0.004
	(0.242)	(0.241)	(0.004)	(0.270)	(0.003)	(0.268)	(0.005)
Birth Year Post Sulfa (1937+)		-6.927***	-0.293	-5.857	0.032	-7.054	-0.380
		(1.718)	(0.274)	(3.892)	(0.524)	(5.157)	(0.292)
Constant			1.332***		1.261***		1.235***
			(0.037)		(0.031)		(0.036)
Observations	23,977	23,977	23,977	17,607	17,607	16,311	16,311
R-squared	0.167	0.129	0.881	0.079	0.836	0.143	0.873
Number of birth_state		47	47	47	47	47	47
F		57.72		52.45		45.63	

Notes: B: Birth Year Fixed Effects, S: State of Birth Fixed Effects, RF: Reduced Form Specification. Short Window: Birth Years between 1930-1943. Additional Controls: Contextual Variables

Robust standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4A

Robustness of Genetic Heterogeneity Associations

Outcome	Cognition	Cognition	Cognition	Cognition
Sample	High Education PGS	Low Education PGS	High PGS Short Window	Low PGS Short Window
Fixed Effects	B, S FE	B, S FE	B, S FE	B, S FE
Specification	RF	RF	RF	RF
Pre Pneumonia Rate X Post Sulfa interaction	1.206	0.463	2.677***	-0.598
	(0.853)	(0.728)	(0.909)	(1.208)
Female	1.030***	0.906***	1.032***	1.111***
	(0.151)	(0.179)	(0.157)	(0.209)
Age at Survey	-0.340***	-0.322***	-0.291***	-0.295***
	(0.020)	(0.033)	(0.030)	(0.028)
Black	-3.525***	-3.884***	-3.402***	-4.064***
	(0.300)	(0.320)	(0.307)	(0.335)
Birth Year Post Sulfa (1937+)	-7.274***	-3.668	1.966	-8.339
	(2.569)	(4.357)	(9.846)	(9.215)
Constant	47.931***	45.894***	43.903***	44.258***
	(1.451)	(2.316)	(2.124)	(1.919)
Observations	8,367	7,944	5,975	6,064
R-squared	0.196	0.202	0.185	0.183
Number of birth_state				
F				

Notes: B: Birth Year Fixed Effects, S: State of Birth Fixed Effects, RF: Reduced Form Specification. Low PGS: Respondents with Cognitive Polygenic Scores<0. High PGS: Respondents with Cognitive Polygenic Scores>0. Low Education PGS: Respondents with Education Polygenic Scores<0. High Education PGS: Respondents with Education Polygenic Scores>0. Short Window: Birth Years between 1930-1943. Additional Controls: Contextual Variables

Robust standard errors clustered at the state-of-birth in parentheses. *** p<0.01, ** p<0.05, * p<0.1