On the Long Term Effects of the 1918 U.S. Influenza Pandemic

Ryan Brown
University of Colorado Denver

Duncan Thomas
Duke University

June 2018

Abstract

Using the 1918 Spanish influenza pandemic, Almond (2006) concludes that *in utero* exposure to maternal health insults has a large, negative impact on socio-economic status that reaches well into adulthood. A key assumption underlying this research is that birth cohorts exposed *in utero* to the influenza are statistically exchangeable with surrounding birth cohorts. The validity of that assumption is investigated using data from the 1920 and 1930 U.S. Censuses. We document that the exposed cohorts were born to families of lower socio-economic status relative to those who were not exposed. For example, fathers of the 1919 birth cohort were less likely to be literate, worked in lower-earning occupations, had lower socioeconomic status, were older, less likely to be white, had higher fertility and were less likely to be WWI veterans than the fathers of surrounding birth cohorts. Furthermore, after controlling for background characteristics, there is little evidence that individuals born in 1919 have worse socio-economic outcomes in adulthood relative to surrounding birth cohorts.

Comments from Douglas Almond, Mark Anderson, Michael Carter, Eileen Crimmins, Janet Currie, William Evans, Erica Field, James Heckman, Dan LaFave, Jeremy Lebow, Grant Miller, Andrew Noymer, John Parman, Daniel Rees, Seth Sanders, T. Paul Schultz, James P. Smith, and Alessandro Tarozzi have been very helpful.

Email addresses: Ryan Brown, ryan.p.brown@ucdenver.edu; Duncan Thomas, dthomas@econ.duke.edu

1. Introduction

Highly cited, influential work by Almond (2006) exploits the 1918 Spanish Influenza pandemic to estimate the causal effect of fetal health on human capital and economic outcomes in adulthood. Treating the influenza pandemic as a natural experiment, Almond uses U.S. census data to compare adult outcomes of the exposed birth cohort, those whose mothers had the highest probability of being exposed to influenza during the pregnancy, with comparable cohorts that were not exposed *in utero*. Two identification strategies are used. First, comparisons are drawn between births in 1919 with those born in surrounding cohorts, 1912-1922. Second, using the 1918-1920 birth cohorts, comparisons are drawn between those born in states with differential levels of maternal infection rates, which serve as proxies for influenza exposure intensity. Using either source of variation Almond finds that males in the exposed cohorts completed significantly less education and earned less as adults than those who were not exposed. The results are interpreted as evidence that fetal health has a long-lasting causal impact on economic prosperity in adulthood.

Identification of these causal effects relies on the assumption that the exposed and unexposed cohorts are statistically exchangeable: that is, there are no unobserved characteristics that could differentially affect adult outcomes of the exposed fetuses relative to other fetuses. We test this key assumption directly using data from the 1920 and 1930 U.S. Censuses. We document that the fathers of the 1919 birth cohort are negatively selected relative to fathers of children born in surrounding cohorts. Specifically, relative to other fathers, those of the 1919 birth cohort were less likely to be literate, worked in lower-earning occupations, had lower socio-economic status (SES), were older, had more children, were less likely to be white, and were significantly less likely to be World War I (WWI) veterans. The assumption that the exposed and unexposed birth cohorts are statistically exchangeable is rejected.

We proceed to investigate the extent to which background differences are able to account for the poorer adult socio-economic outcomes of the 1919 birth cohort relative to surrounding cohorts. For an array of markers of adults SES, there is no evidence that, after controlling background, the 1919 birth cohort have worse outcomes than the comparison cohorts. This is true for males, females and nonwhites. The findings cannot be explained by age heaping. We document the evidence that spatial variation in the intensity of in utero exposure to the influenza pandemic has a deleterious impact on adult SES outcomes is, at beast, weak. Adjusting for paternal background, there is no evidence of a negative dose-response. Our findings suggest the conclusion that *in utero* exposure in the United States to the 1918 influenza pandemic had a persistent adverse impact on the adult economic outcomes of the exposed cohorts is premature.

Finally, we discuss plausible reasons why the coincidence of the end of World War I (WWI) with the 1918 Spanish influenza pandemic substantially complicates isolating a causal impact of *in utero* exposure to the pandemic and may explain our findings. In addition, we discuss the implications of our work for interpretation of studies of populations outside the U.S. that compare adult SES of the 1919 birth cohort with other cohorts.

2. The long arm of the 1918 influenza pandemic

Fetal health and long-term economic outcomes

Recognizing the key roles of maternal and fetal health for birth outcomes and child health, maternal health during pregnancy has been a central plank in public health programs across the globe. Influential work by Barker (1994) posited that the long arm of *in utero* health insults likely reached well into adulthood. The combination of plausible biological mechanisms in humans and experimental evidence in animals, Barker and his colleagues traced out the impact of *in utero* health insults during specific periods of development of the fetus and health in later life including elevated risks of, for example, dyslipidemia, cardio-metabolic disease and premature mortality. A large literature provides considerable support for these predictions of the Fetal Origins Hypothesis, documenting links between fetal health and biological markers of health in adulthood.

It is, however, not obvious that the insights of the Barker hypothesis extend to adult socio-economic outcomes. First, a key strength of Barker's work lies in the careful documentation of biological mechanisms underlying the specific timing of *in utero* insults (such as when arteries are developed) and subsequent biological parameters that indicate health problems (hardened arteries in mid-life, for example). In contrast, while similar biological pathways may exist, there are no widely-accepted biological foundations that link *in utero* insults to socio-economic outcomes in adulthood. Second, behavioral responses, including post-natal interventions, have been shown to be effective in mitigating early life gaps in SES (Heckman, 2006). Whether the Fetal Origins Hypothesis extends to adult SES is, therefore, an empirical question.

The 1918 Influenza pandemic as a natural experiment

This question is taken up by Almond (2006) in a creative study that treats the 1918 influenza pandemic as a natural experiment to assess the long-term effects of *in utero* health on SES in adulthood using samples that are representative of the U.S. population. He argues that when the pandemic struck in the U.S. in October 1918[1], it was both unanticipated and its impact was immediate. He points out that 85% of all influenza deaths in the U.S. occurred between October 1918 and January 1919 and that the virus was almost completely inert by the end of January 1919 (Almond, 2006).[2] The speed, violence and unanticipated nature of the onset of the pandemic are key for identification. There was little scope for behavioral responses in anticipation of the influenza pandemic and births in early 1919 would have been conceived prior to the onset of the pandemic.

Almond reports intent-to-treat estimates of the effect of exposure to the influenza pandemic, citing three additional features of the pandemic (Jordan, 1927 as cited in Almond, 2006) that are important for his identification strategy. First, it is estimated that about 28% of the U.S. population was infected with the virus and that incidence was particularly high among pregnant women and women of child-bearing age. Second, noting that overall mortality attributed to the pandemic was not high (5 per thousand) Almond argues selective mortality (of mothers and fetuses) is unlikely to contaminate estimated effects. Third, the disease, which was transmitted through the air, was difficult to avoid and infection rates did not vary with SES.

There was also considerable spatial variation in the intensity of the pandemic across the United States that does not appear to be related to an area's wealth, climate or topology (Brainerd and Siegler, 2003) and can potentially be exploited to identify the impact of *in utero* health on adult SES.

3. Assessment of the evidence comparing cohorts

Adult SES of the 1919 birth cohort

Using the 1% sample of the 1960, a combined 3% sample of the 1970, and a 5%

[1]Historians have documented the first wave of influenza appeared in the U.S. in January 1918 in Haskell County, Kansas, spread to an army camp in Kansas in March where 522 men fell sick within two days and cases were documented in Queens, New York in March. However, the outbreak received minimal media coverage in the U.S. at the time. It was also detected in France, Spain and elsewhere in Europe in the spring of 1918. The only country in which there is documented evidence it received widespread media attention is Spain (Barry, 2004).
[2]There was a mild resurgence in cases in the spring of 1919, but the effect of the virus was relatively benign and drew little attention (1918.pandemic.gov). That resurgence is unlikely to contaminate estimated effects.

sample of the 1980 U.S. Censuses from IPUMS, Almond contrasts several indicators of human capital and economic well-being of the 1919 birth cohort with surrounding cohorts, 1912 to 1918 and 1920 to 1922. His primary specification measures the effect on a later life outcome, y_i, of being born in 1919, $I_i(YOB = 1919)$, relative to the comparison birth years, controlling for a yearly trend, YOB_i, and its quadratic, YOB_i^2:

$$y_i = \beta_0 + \beta_1 \cdot YOB_i + \beta_2 \cdot YOB_i^2 + \beta_3 \cdot I_i(YOB = 1919) + \varepsilon_i \qquad [1]$$

Table 1 presents estimates of $\hat{\beta}_3$ the coefficient on the 1919 year of birth indicator for males in 1960 from Almond (2006) in column 1, along with a replication using the IPUMS 1% sample of the 1960 U.S. Census in column 2.[3] Relative to those born in surrounding cohorts, males born in 1919 are significantly less likely to have graduated from high school, completed fewer years of education, have lower wage income, are more likely to be poor and have lower scores on the Duncan's Socioeconomic Index (SEI), an indicator or SES that is based on the occupation of the individual. The data for years of completed education is presented visually in Panel A of Figure 1, reproduced from Almond, which establishes the 1919 birth cohort, displayed in red, completed less education than predicted by the trend.

The 1920-1922 cohorts are excluded from the comparison cohorts in the third column of Table 1 for two reasons. First, we will exclude these comparison cohorts in some of the analyses reported below. Second, evidence from outside the U.S. indicates that experiencing the influenza pandemic affected subsequent fertility which altered the parental composition of the post 1919 cohorts (Boberg-Fazlic et al. 2016). The estimates based on the 1912-1919 cohorts are very similar to those based on the 1912-1922 cohort comparisons except they are estimated less precisely. Specifically, relative to the comparison cohorts, the1919 birth cohort has significantly less education and lower Duncan's SEI in adulthood.

The evidence in Table 1 can be interpreted as causal estimates of the impact of *in utero* health shocks and economic prosperity in adulthood if the 1919 and surrounding birth cohorts are statistically exchangeable. That is, no unobserved factors in the model, ε_i, have a different impact on the outcomes for the exposed birth cohort, $I_i(YOB = 1919)$, relative to the other cohorts. If, for example, parental SES of the 1919 birth cohort were lower than the parents of the comparison cohorts, then the credibility of the research design would be called into question. Almond dismissed this concern because there was no difference in the

[3] The difference between Almond's estimates and the replication estimates are likely due to changes in the public release versions of the IPUMS samples.

probability the parents were foreign born between the exposed and comparison birth cohorts using the 1960 and 1970 Census samples.

Paternal SES of males in the 1919 birth cohort

As our first step towards investigating whether this is, in fact, a legitimate concern, the other three panels of Figure 1 display paternal characteristics by year of birth of males born between 1912 and 1922 again with a linear trend. Data on paternal characteristics are drawn from the 1930 U.S. Census linking children with their fathers in the household.

The 1930 census does not record education and so, in panel B of the figure, the percentage of fathers who report themselves as literate is displayed for each birth cohort. The figure provides visual evidence that fathers of the 1919 cohort are substantially less likely to be literate relative to fathers of children in surrounding birth cohorts. The parallel with the figure for the child's education in panel A of the figure is striking. This pattern in paternal characteristic differences across cohorts is not restricted to literacy. Panel C displays the average Duncan's SEI and shows that fathers of the 1919 birth cohort have much lower SEI than surrounding cohorts. As shown in Panel D, fathers of the 1919 cohort are also older than predicted by the trend. These figures suggest the assumption of exchangeability of the birth cohorts is not obviously satisfied and, at least, warrants further inquiry (Thomas, 2010).

To this end, we compare a broad array of paternal characteristics of each birth cohort drawing on data from the full count versions of the 1920 and 1930 U.S. Census. The advantages of the 1920 Census are that it is proximate to the birth of the 1919 cohort, year of birth is collected from every respondent and, for those born in the previous 5 years, birth month is also collected. The drawback is that it precludes drawing comparisons with the post-1919 birth cohorts used in Almond (2006) although, as noted above for Table 1, that restriction does not affect conclusions about deficits of the 1919 birth cohort.

Drawing data from the 1930 Census permits inclusion of the full set of cohorts, 1912-1922 used in Almond (2006) but birth year has to be constructed. Specifically, the census was enumerated on April 1, 1930, and information on age was collected in whole years as of March 31, 1930. We have calculated birth year as 1930 less age less one so that each birth cohort covers the period from April 1 through December 31 of the birth year and from January 1 through March 31 of the following year. Thus, the 1919 birth cohort includes births in the second through fourth quarters of 1919 and the first quarter of 1920 which roughly translates into covering conceptions in the third and fourth quarters of 1918 as well as the first two quarters of 1919.

Relative to the 1920 Census, the longer hiatus between the birth of the child and the 1930 Census enumeration is a disadvantage. A potentially important issue is that analyses are restricted to children who coreside with their fathers at the time of the census. If paternal mortality or paternal co-residence patterns of the 1919 birth cohort are different from the surrounding cohorts, the estimates will be contaminated. This turns out to not be an important concern: for example, the fraction who do not coreside with their fathers is very similar for each birth cohort, paralleling patterns in the 1920 Census.

Model [1] is re-estimated replacing the adult outcomes of the child born in each cohort year with paternal characteristics of the males born in each cohort year. If the exchangeability assumption is correct, the coefficient on the 1919 birth cohort should not be statistically significant. Results are reported in Table 2.[4] Data from the 1920 full count Census are used in Panel A and from the 1930 full count Census in Panel B. Means for each paternal characteristic are reported in the first column of each panel and the coefficient from [1] on the indicator for the 1919 birth cohort of the child and associated standard error are in column 2 of each panel.

Since education is not reported in the censuses, we compare paternal literacy, a relatively blunt instrument. About 9% of fathers reported themselves as being illiterate in the 1920 Census and about 7% in the 1930 Census. The difference likely reflects the combination of literacy skill acquisition and a declining propensity to admit illiteracy over time. As shown in the first row of the table, consistent with Figure 1B, fathers of the 1919 birth cohort are significantly more likely to be illiterate than predicted by the trend in both censuses.

The 1920 and 1930 Censuses do not provide income data and so the next two rows of the table report indices of SES based on the occupation of the father as recorded in each Census. The Occupation Income Score assigns to each occupation the median total income (in hundreds of 1950 dollars) of all people in the 1950 census with that occupation. The Duncan SEI attributes an education and income level to each occupation as of 1950 to create the index. For both SES indicators, the fathers of the 1919 birth cohort are significantly lower SES in 1920 and in 1930. In addition, fathers of the 1919 cohort are significantly more likely to be non-white. All of this evidence points in one direction: the 1919 birth cohort had fathers with lower SES than surrounding cohorts.

[4] Due to the timing of data collection the trend is estimated with births from January 1, 1912 to December 31, 1919 when using the 1920 U.S. Census and from April 1, 1911 to March 31, 1923 when using the 1930 U.S. Census.

Moreover, as shown in the next three rows of the table, fathers of the 1919 cohort tend to be older when the child was born, have more children (as proxied by the number of children in the household), and marry earlier. This suggests that the fathers started fertility relatively early, continued longer, and produced larger families. Larger family sizes are associated with lower levels of human capital investments in each child and this may also contribute to the lower levels of adult SES of the 1919 birth cohort (Becker and Lewis, 1973). In addition, older fathers are likely to have less education given the secular increase in education across birth cohorts. These paternal behaviors are associated with SES and reinforce evidence that the fathers of the 1919 birth cohort were lower SES than the surrounding cohorts. The results are important because a large literature has established a positive link between parental income and SES, on one hand, and the outcomes of the child as an adult, on the other (Brooks-Gunn and Duncan 1997; Corcoran et al. 1992).

At this point, the key conclusion from Table 2 is that the exchangeability assumption necessary to draw causal inferences from comparisons of the 1919 birth cohort of males with surrounding cohorts is not supported by data from either the 1920 or the 1930 Census. The implications for adult SES outcomes of the 1919 birth cohort, relative to surrounding cohorts, are assessed in the next sub-section; we then discuss several potential reasons for the findings.

Adult SES of males in the 1919 birth cohort conditional on family background

Having established that paternal SES of the 1919 birth cohort of males is lower than paternal SES of surrounding cohorts of males, we assess the extent to which adjusting for these background differences accounts for the observed differences in adult SES of the 1919 birth cohort relative to surrounding birth cohorts. Unfortunately, this is not straightforward with census data because the parental background of adult respondents is not recorded. Therefore, we construct proxies for parental background for each birth cohort, b, by state of birth, s, and race, r, and extend models of adult SES, [1], by adding these proxies, P_{bsr}:

$$y_i = \gamma_0 + \gamma_1 \cdot YOB_i + \gamma_2 \cdot YOB_i^2 + \gamma_3 \cdot I_i(YOB = 1919) + \gamma_4 P_{bsr} + \upsilon_i \qquad [2]$$

Table 3 presents estimates of γ_3 for adult outcomes of males in the 1912-1922 birth cohorts adjusting for paternal characteristics that are constructed using the IPUMS 1930 full count Census. The characteristics are paternal age at marriage, paternal age at the birth of the index child, whether the father is white and whether the father was a WWI veteran.[5] For each of these characteristics, the paternal proxy is the average over all children born in each state and year of birth cohort calculated separately for whites and non-whites. In each panel in the table, the first column in the pair reports estimates of β_3 in [1] without paternal controls and, in the second column, estimates of γ_3 from [2] which adjust for paternal characteristics are reported. Each element in the table represents a separate regression with Panel A using adult SES outcomes measured in the 1960 Census, panel B using the 1970 Census and panel C using the 1980 Census.[6]

As shown in the first row of each panel, males born in 1919 were between 1.4 and 2.1 percentage points less likely to graduate from high school and this gap is statistically significant. Controlling paternal characteristics, however, the 1919 birth cohort are more likely to have graduated from high school and this positive difference is statistically significant in the 1970 and 1980 Censuses. Similarly, without background controls the 1919 birth cohort completed between 0.12 and 0.18 fewer years of schooling but, after taking into account paternal characteristics, they completed between 0.05 and 0.15 more years of schooling; all of these differences are statistically significant. The 1919 birth cohort earned less income, were more likely to be poor and scored lower on the Duncan SEI without controlling paternal characteristics but all of these effects are reversed in sign when paternal characteristics are controlled and, again, conditional on these background characteristics the 1919 birth cohort achieve better economic outcomes than their peers. Uncontrolled, disability is more likely to interfere with work for the 1919 cohort but, after adjusting for background, there are no differences when they are in their early fifties (in the 1970 Census) and disability is significantly less likely to be an impediment in their early sixties (in the 1980 Census). In their early fifties the 1919 cohort, whether including paternal controls or not, receive about $12 more per month in welfare income but this effect, while present for the 1919 cohort when they are in their early sixties when using the uncontrolled model, is smaller in magnitude and not statistically significant in the adjusted model. Social security income is reported in the final column: the 1919 birth cohort receive about $80 more per month in social security income when they are in their early sixties whether or not paternal characteristics are

[5]These paternal characteristics are selected because none should change substantially over time given the birth of the index child and so the specific timing of the measurement, in 1930, is likely immaterial. Including time-varying paternal characteristics based on the 1930 census would introduce potentially endogenous control variables and complicate interpretation.

[6]Indicator variables are included in the model when paternal characteristics are missing.

controlled. It is not entirely clear how to interpret this result: the fact that the 1919 cohort receives more social security income in 1980 than the other cohorts reflects the combination of earlier receipt and higher payouts potentially because of longer work histories or higher lifetime earnings or both.

Overall, the evidence in Table 3 completely changes the narrative in Almond (2006). In the unadjusted models, as documented by Almond, the 1919 birth cohort have less human capital, lower earnings, and more work limitations than the surrounding cohorts. Adjusting for paternal characteristics not only erases these deficits, but, for almost every outcome, the 1919 birth cohort has higher SES than the surrounding cohorts.[7]

Table 4 presents parallel estimates using the full count of the 1920 Census to measure paternal characteristics of males. Paternal characteristics in the 1920 Census are calculated for each year and state of birth, separately for whites and non-whites, and include paternal literacy, occupation income score, number of father's children in the household and paternal age at birth.

By necessity, the cohorts are restricted to the 1912 to 1919 birth years. Comparing the first column in each panel in Table 4 with the corresponding column in Table 3 documents that this restriction does not affect conclusions about males in the 1919 birth cohort having worse socioeconomic outcomes in adulthood than males in the comparison cohorts. However, adjusting for paternal characteristics in the second column of each panel of Table 4 reverses those conclusions: males in the 1919 cohort completed significantly more education and earn more than the comparison cohorts.[8]

Adult SES in the 1919 birth cohort: Summary of the evidence

Summarizing the results thus far we have established two facts regarding males in the 1919 birth cohort. First, the socio-economic status of the fathers of the cohort is lower than that of the surrounding birth cohorts. Second, adjusting for family background – with proxies for paternal characteristics – SES outcomes in adulthood of the 1919 birth cohort are no worse than those of the surrounding birth cohorts and there is some evidence that their SES outcomes are, in fact, better.

[7]Of the 26 differences between estimates with and without paternal controls, β_3-γ_3, in Table 3, 24 are statistically significantly different at the 1% level. The two exceptions are welfare income (in 1970) and social security income (in 1980).

[8]All differences between estimates with and without paternal controls in Table 4 are statistically significantly different at the 5% level with the exception of welfare income in 1970 and 1980 which is not statistically significant in any of the specifications.

Almond (2006) also describes results for females and nonwhites. We have repeated the analyses described above for males for these two demographic groups and results are summarized in Table 5. Paternal characteristics of the 1919 birth cohort, relative to the 1912-1918 cohorts are displayed in the first column of the table for females in the upper panel and for nonwhites in the lower panel. Three of the paternal characteristics, measured in the 1920 Census, are reported: literacy, occupation income score and Duncan's SEI. The fathers of the 1919 birth cohort are negatively selected on each of these characteristics. For example, fathers of females born in 1919 are 1.32% less likely to be literate and fathers of nonwhites are 0.95% less likely to be literate. For females, the deviations from trend for the 1919 cohort are negative and significant in both the 1920 and 1930 Censuses for all the paternal characteristics included in the analyses of males. For nonwhites, the deviations indicate negative selectivity for the same characteristics (except the occupation income score) and the deviations are significant for three of the five 1920 measures and five of the seven 1930 measures.

Outcomes in adulthood are presented in panel B of Table 5 for outcomes measured in the 1960, 1970 and 1980 Censuses. The even numbered columns replicate Almond (2006) and the odd numbered columns include measures of paternal background from the 1920 Census.

In models that do not adjust for background, the evidence that females in the 1919 birth cohort have worse adult outcomes relative to the comparison cohorts is not as clear as it is for males. Females born in 1919 are significantly less likely to have graduated from high school only in the 1970 Census and they have completed significantly fewer years of education in 1970 and 1980. Moreover, their total income is significantly higher in 1980. Adjusting for parental background in the odd columns in panel B, all the gaps turn positive and they are statistically significant indicating that adult outcomes of females in 1919 birth cohort are better than those of the comparison cohorts.

The evidence that nonwhites born in 1919 were negatively affected by the influenza pandemic is considerably weaker, as shown in the even numbered columns in the lower half of Table 5. While the education outcomes of the 1919 birth cohort are worse than the comparison cohorts, only one of the six estimates is statistically significant and none of the income differences is significant. However, adjusting for parental background, all but one of the estimates is positive and total income in 1980 is significantly higher than among the comparison cohorts.

Alternative explanations for cohort differences

We have documented that in the 1920 and 1930 Censuses, fathers of males, females and nonwhites born in 1919 are negatively selected relative to surrounding cohorts. Since parents reported the child's age and the less educated are more likely to heap on preferred ages, it is possible that heaping could explain these findings as the age of children assigned to the 1919 birth cohort in the 1920 and 1930 Census is 0 and 10, respectively. There are two issues that we explore with regard to this concern: the extent of age heaping among the children in the birth cohorts that we include in our analyses and the extent to which age heaping in these cohorts is related to paternal education and literacy.

First, to provide evidence on the extent of age heaping, for each year of birth, the number of living children reported in the full count 1920 and 1930 Censuses are compared with the number of births reported in the Vital Statistics, adjusting for mortality. Natality data are drawn from "Birth, Stillbirth, and Infant Mortality Statistics for the Birth Registration Area of the United States" (Roper and Austin, 1931) and mortality data are drawn from the annual "Mortality Statistics" reports published by the Census Bureau. Birth registration data were first collected in 1915 and covered 10 states and the District of Columbia; by 1922, registration data had been expanded to 31 states.[9] We draw comparisons between the natality and census data for all covered states in each year of birth.

In the 1920 Census, age is reported in years and months for births in the previous five years which is as far back as vital statistics reach. Column 2 of panel I of Table 6A displays the ratio of the number of births reported in the 1920 Census for each birth year to the number of births, adjusted for mortality, in vital statistics for children reported to have been born in the covered states. If there is age heaping on a specific birth cohort in the 1920 Census, the ratio for that cohort will be greater than unity. For the 1919 birth cohort, the ratio is 0.97 indicating a slight undercount. In fact, the ratio is very close to unity for all of the cohorts from 1915 through 1919. This evidence indicates that age heaping is not an important concern in the 1920 Census for the 1915 to 1919 cohorts.

[9] The 1915 registration area consisted of Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New York, Pennsylvania, Rhode Island, Vermont, and the District of Colombia. Maryland was added in 1916. Indiana, Kansas, Kentucky, and North Carolina were added in 1917. Ohio, Utah, Virginia, Washington, and Wisconsin were added in 1918. California, Oregon, and South Carolina were added and Rhode Island was dropped in 1919. Nebraska was added in 1920. Delaware, Mississippi, and New Jersey were added in 1921, and Rhode Island was readmitted as well. Illinois, Montana, and Wyoming were added in 1922 (Roper and Austin, 1931).

This is important because, as shown in the first column of panel I of Appendix Table 1, when attention is restricted to the 1915-1919 birth cohorts in the 1920 Census, fathers of the 1919 birth cohorts are significantly negatively selected. For example, fathers of the 1919 birth cohort are 1.20% (se=0.08) more likely to be illiterate than predicted by the trend estimated for the 1915-1919 cohorts. For comparison, using the 1912-19 cohorts to estimate the trend, these fathers are 1.21% (se=0.05) more likely to be illiterate. All but one of the comparisons of paternal characteristics indicate statistically significant negative paternal selection using the reduced set of comparison cohorts in the 1920 census for which we have shown age heaping is not a concern. Moreover, restricting the cohorts to those born in 1915 through 1919 in the analyses of adult socioeconomic outcomes in Table 4 does not affect any of our conclusions that, adjusting for paternal background, the outcomes of 1919 birth cohort are not worse than the earlier cohorts.[10] Thus, our conclusions based on the 1920 Census regarding negative selection of fathers of the 1919 birth cohort and the impact of adjusting for paternal background cannot be explained by age heaping.

Turning to the 1930 Census, the ratio of the number of children alive and reported as born in the covered states in each of the 1915-1922 birth cohorts to the number of births, adjusted for mortality, from vital statistics is displayed in column 2 of panel II of Table 6A. The ratio is 1.11 for the 1919 cohort. If this is because of age heaping, births that occurred in 1918 and 1920 would have been reported as having occurred in 1919 and the census to vital statistics ratios for 1918 and 1920 would be less than one. In fact, neither is less than one: the ratios are 1.08 and 1.07, respectively, and indicate, at most, modest heaping on the 1919 birth cohort.

One estimate of the extent of age heaping is the ratio of the number of births reported in 1919 to the average of those born one year earlier and one year later. The estimate for all children, reported in row 1 of column 1 of panel B of Table 6, is 1.034, confirming that heaping is modest. While heaping is substantially greater for children with illiterate fathers (1.113 in row 2) it is smaller for those with literate fathers (1.028 in row 3) even smaller among children of fathers who are literate and white and smaller still among children of

[10]For example restricting estimation of [1] to the 1915 to 1919 birth cohorts in the 1960 Census, the 1919 birth cohort is 3.5% less likely to have graduated from high school and completed 0.21 less years of education. These deficits are statistically significant and larger than the estimates in Table 3 which includes the 1912 to 1922 cohorts and in Table 4 which includes the 1912 to 1919 cohorts. As in Tables 3 and 4, adjusting for parental characteristics reverses these deficits. In comparison with the 1915 to 1918 cohorts, 1919 birth cohort is estimated to be 3.5% *more* likely to be a high school graduate and to have completed 0.34 *more* years of education. These effects are also statistically significant.

fathers who are literate, white and native born. The ratio for that group indicates that 1.1% of all births reported in 1919 are likely to be due to heaping on age 10.

For children of the latter three groups of fathers, age heaping on age 10 is arguably ignorable in the 1930 census. To evaluate whether heaping drives the 1930 Census results, columns 2 and 3 of Table 6B report the deviations from trend of two paternal indicators of socioeconomic status for the 1919 birth cohort, relative to the 1912 through 1922 birth cohorts as reported in the 1930 Census. The deviations for all children in the first row are from column 4 of Table 2. Deviations for the three sub-groups of children for whom we have shown heaping is modest are reported in rows 3 to 5: the fathers of the 1919 birth cohort are significantly negatively selected and the gaps are very similar to the gaps for all children in the first row.[11]

In sum, age heaping cannot explain the finding that the fathers of the 1919 birth cohort are negatively selected. Corroborating evidence is provided by Beach, Ferrie and Savedra (2018) who match male World War II enlistees to their records in the 1920 and 1930 Census and document that, in this sample, fathers of the 1919 birth cohort are negatively selected. Noting that ages of enlistees are well measured, they conclude age heaping does not explain this result.[12] Their evidence on enlistees also indicates that our proxies for paternal background perform well in this context.

4. Assessment of the evidence exploiting variation in exposure

The analyses thus far have exploited only cohort differences. Almond (2006) also investigated whether variation in virulence of the influenza explained variation in adult socio-economic outcomes. Unable to detect evidence of a dose-response, he concluded that "coverage and quality of data on timing and virulence prevent […] definitive conclusions" and turned to using the year- and state-specific maternal mortality rate, *MMR*, as a proxy for virulence. Restricting attention to the 1918 through 1920 birth cohorts in order to isolate an effect of fetal exposure, he investigated how each adult outcome, y_i, varies with the MMR measured in the year before the birth, *t-1*, in a model that included state fixed effects, μ_s and

[11] The same patterns describe the other paternal characteristics in Table 2.

[12] They assert our estimates are contaminated by age heaping which, as we have documented, is not the case. The reason for this discrepancy is that they investigate heaping after combining all children age 0 through 16 from the 1910, 1920 and 1930 Censuses. We do not use the 1910 Census and only include 0 to 8 year olds in 1920 and 8 to 18 year olds in 1930 in our study; restricting attention to these cohorts , we have used census and vital statistics data to show that age heaping is at most modest.

birth year fixed effects μ_t:

$$y_i = \alpha_0 + \alpha_1 MMR_{s_{t-1}} + \mu_s + \mu_t + \xi_{is_t} \qquad [3]$$

MMR is indicative of excess mortality largely due to influenza. Estimates of α_1 from Almond (2006) using adult socio-economic status of males in the 1960 census are displayed in the first column of Table 7. Our replication, in the second column, yields estimates that are very close. However, we checked the original sources (U.S. Public Health Service, 1947) for the MMRs and discovered two errors. First, for Virginia in 1919, Almond assigns an MMR of 6.3 while the rate recorded in the source was 8.3. Second, Almond includes 19 states although data on MMR are recorded for Washington D.C. in the same source. Estimates based on these amended data are displayed in the third column of the table. While the effect sizes are attenuated, the probability of high school graduation and completed years of schooling are significantly lower as MMR rises at a 5% size of test. However, using the corrected data, none of the other three of the five adult outcomes examined by Almond is significantly related to the maternal infection rate.[13]

Paralleling the cohort analyses in the previous section, model [3] is extended by adjusting for paternal background characteristics, P, measured in the 1930 Census for each state and year of birth, s_t, separately for whites and nonwhites:

$$y_i = \delta_0 + \delta_1 MMR_{s_{t-1}} + \delta_2 P_{s_t} + \mu_s + \mu_t + \zeta_{is_t} \qquad [4]$$

Estimates of δ_1 are displayed in column 4 along with p-values in column 5 for tests of the differences, $\delta_1 - \alpha_1$, between the adjusted and unadjusted estimates. Adjusting for paternal background controls cuts the estimated effects on education in half and neither is statistically significantly affected by MMR. The differences between the adjusted and unadjusted estimates, are statistically significant as shown by the p values.

Whereas none of the three income-related outcomes is significantly affected by the maternal mortality rate in either the adjusted or unadjusted models, it is worth noting that adjusting for paternal backgrounds reduces the effect sizes on the probability of being poor and the Duncan SEI by 100% and on income by 50%.

The fact that even the unadjusted estimates of α_1 are statistically significant for only two of the five adult outcomes is not strong evidence for a dose response effect of influenza

[13]The differences between the results in columns 2 and 3 are driven by the correction to the MMR for Virginia in 1919; whether Washington D.C. is included or excluded does not substantively affect the results.

exposure. Results from estimating the unadjusted model [3] for outcomes of males in the 1970 and 1980 Census are reported in columns 2 and 3 of Appendix Table 2. Whereas in 1960, males who were born in states with higher levels of excess maternal mortality are significantly less likely to have graduated from high school and completed significantly fewer years of education, by 1970 these same males are no less likely to have graduated from high school and report having completed significantly more years of education. The reason for this reversal when it is unlikely that these men, in their forties, completed more years of education is unclear. Moreover, in 1970, these males are also significantly less likely to be poor. By 1980, none of the estimated difference-in-differences is statistically significant.

Estimates from the same unadjusted models for females are displayed in panel B of Appendix Table 2. The only statistically significant effect indicates that females who were born in states with higher levels of excess maternal mortality are less likely to be poor – as was found for males in 1970. Panel C of the table presents results for non-whites: two of the estimated coefficients are statistically significant and, again, both indicate that higher exposure is associated with improved socioeconomic outcomes in adulthood.

Taking the evidence together, of 45 estimated coefficients only two indicate a statistically significant negative link between adult socioeconomic outcomes and excess maternal mortality and those estimates are not consistent over time. Moreover, as we have shown, neither is statistically significant after adjusting for paternal background. In contrast, four of the estimates indicate the link between maternal mortality and adult outcomes is significantly positive. The evidence there is a significant negative dose response is very weak. Indeed, adopting a testing procedure that takes into account the multiple comparisons in these analyses (Hommel, 1988), results in the conclusion that variation in the intensity of exposure to the 1918 influenza pandemic *in utero* has no statistically significant impacts on SES in adulthood.[14]

[14] Using World War II enlistees linked to the 1920 and 1930 Censuses, Beach, Ferrie and Savedra (2018) examine how outcomes in adulthood of males born between 1912 and 1922 vary with excess maternal mortality estimated at the city and year level. Putting aside the selectivity of this matched sample, they report that, relative to the reference group, males in the 1919 birth cohort born in high excess mortality states is significantly less likely to have graduated from high school and completed significantly fewer years of schooling, but they are significantly taller. However, since the reference group is all males born in 1912 through 1915, it is not clear that they have isolated the effect of fetal exposure. Their figures indicate no statistically significant differences in education outcomes among any of the males born in 1916 through 1922 who were exposed to elevated maternal mortality. To focus attention on the impact of *in utero* exposures, Almond compared the 1919 cohort with the 1918 and 1920 birth cohorts. Based on the data they present, there does not appear to be evidence in Beach, Ferrie and Savedra of an impact of *in utero* exposure to excess maternal mortality when comparisons are restricted to these three cohorts.

5. Discussion

Almond (2006) documented that adult socioeconomic outcomes of the 1919 birth cohort are worse than surrounding cohorts which he attributed to in utero exposure to the 1918 influenza pandemic. We document that the fathers of the 1919 birth cohort are negatively selected and, adjusting for background, accounts for the worse adult outcomes of this cohort. We turn next to a broader discussion of these findings as well as evidence drawn from other countries.

Paternal controls proxy for influenza virulence

The absence of evidence for a dose response effect notwithstanding, it is possible that our state- and year-specific measures of paternal socioeconomic status are proxies for higher virulence of the virus. To evaluate this potential issue, we have re-estimated the unadjusted models [1] and adjusted models [2] excluding the Northeast states[15] where virulence was highest. If these high virulence states drive the influenza effects that Almond reports, then excluding those states, the 1919 birth cohort outcomes should not be different from the surrounding cohorts. If paternal characteristics proxy for virulence then the inclusion of paternal characteristics should have an attenuated impact on the 1919 birth cohort differences for children born outside the Northeast.

Without paternal controls, children in the 1919 birth cohort born outside the Northeast states have significantly worse socioeconomic outcomes in 1960 relative to the surrounding cohorts and, for many outcomes, the gaps are actually larger in magnitude than the gaps for births from all states. For example, in 1960, males born outside the Northeast in 1919 completed 0.215 (s.e.=0.078) fewer years of education relative to those born outside the Northeast in 1912 through 1922. This gap is larger in magnitude than that for males born in all states (0.148 years, in Table 1, column 2). Adjusting for paternal background, the deficits of the 1919 birth cohort born outside the Northeast either disappear or are reversed. For example, the 1919 cohort completed 0.287 (s.e.=0.076) more years of education than the comparison cohorts.[16] Whereas this evidence is not consistent with paternal background serving as a proxy for virulence, it is consistent with absence of a dose response.

[15] The states are Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont.

[16] Results are very similar for analyses of outcomes in 1970 and 1980.

Effects of World War I

The fact that the 1918 influenza pandemic coincided with the end of World War I (WWI) suggests an alternative explanation for our findings. The United States declared war against Germany in April 1917 and started deploying troops to Europe in the summer of 1918. Thus during the conception period of the 1919 birth cohort, the United States was involved in a major global conflict.

There are multiple pathways through which WWI may have affected fertility decisions of couples and, therefore, the distribution of parental characteristics of the 1919 birth cohort. Over and above troop deployments and the threat of future deployments, there was greater uncertainty along with elevated levels of stress, reductions in income and food consumption as well as the potential for rationing. [17]

Whereas it is difficult to measure expectations and uncertainty at that time, there is good data on troop deployments. Figure 2, from Ayers (1919), documents that the number of men deployed in the American Army rose very dramatically in the last half of 1918, peaked at the end of the year and declined slowly during the first three months of 1919. This exactly coincides with the timing of the influenza pandemic.

Key for explaining the negative selection of parents in the 1919 cohort, those men who were actually deployed and those who were at risk of being deployed were unlikely to be drawn from the lower rungs of the SES ladder. There are several reasons for this fact. First, WWI was the first war in which a U.S. citizen was not allowed to hire a proxy to serve in his place. This ruled out the possibility of the upper class buying their way out of service. Second, due to the draft categories in use in 1917, men with greater levels of resources were more likely to be conscripted. While almost all draft eligible men were put in Class I, one of the main reasons for a deferment was the income dependency of the family of a potential draftee. A man was placed in a lower priority group if his family had little financial support apart from himself, because the family would have “insufficient” income to sustain itself if he were drafted (Nudd 2004). Third, draft eligibility was partly based on age with older men being less likely to be conscripted. Since educational attainment was rising substantially between these cohorts of men, an older father is likely to have less education. Fourth, deferments were awarded to men for health reasons and so the less healthy were less likely to be drafted.[18]

[17]For example, war-related famine in Allied countries triggered the U.S. to launch a major government food conservation campaign entitled “Food Will Win the War” and urged citizens to restrict their consumption of meat, wheat, fats, and sugars.

[18]In addition to the change in parental composition caused by WW1, the war may have impacted several other

Those couples who were most directly affected by WWI are likely to be better educated and have higher income and they are the couples who are most likely to defer fertility. Hence, the SES of parents of the 1919 birth cohort is likely to be lower than prior cohorts and, possibly, later cohorts.

The 1930 Census provides empirical evidence on the extent to which WWI veterans who survived to the Census and had a child between 1912 and 1922 were selected on SES. Results are summarized in Table 8. The overall mean for all fathers is displayed in the first column. The second column displays the difference in socioeconomic characteristics between the 7% who were veterans and the other fathers. Overall, 7.6% of the fathers were illiterate, but veterans were 1.22 percentage points less likely to be illiterate. They were also less likely to be non-white and had higher levels of SES in 1930 as indicated by both the Occupation Income score and Duncan SEI. All of these differences in paternal characteristics of veterans are statistically significant.

It turns out that a proxy for whether the father was a WWI veteran, the fraction of males who are veterans for each state, year of birth, and race cohort is a sufficient statistic for paternal background in the adjusted models, [2]. As shown in Appendix Table 3, only using that proxy yields the same conclusions as those drawn with the full set of paternal controls included in Table 3. Specifically, the disadvantage in adulthood SES on the 1919 birth cohort is no longer present after the inclusion of the WWI veteran control in analyses of male outcomes in the 1960, 1970 and 1980 censuses.

It is not possible to make predictions about the impacts of WWI, broadly construed, on the specific timing of fertility outcomes since behaviors of couples and their expectations are involved. This is evident in panel B of part I of Appendix Table 1 which displays the deviation from trend in the paternal characteristics of the 1919 birth cohort, quarter by quarter relative to the same birth quarters in the 1915 to 1918 cohorts. While negative selection is greatest in magnitude for those born in the first two quarters, the selection is negative and significant for 5 of the 6 markers among those born in the third and fourth quarters.

In sharp contrast, the fetal origins hypothesis makes very tight predictions about the timing of impacts. The deleterious effects on adult outcomes should be greatest on those born

---

aspects of life that could bias the estimates of the effect of prenatal exposure to the flu found in Almond (2006). The loss of enlisted fathers as breadwinners and changes in food prices (Rotwein, 1945), along with a national food conservation campaign, may have caused a restriction in nutrients consumed by pregnant mothers. Moreover, the mobilization effort may have caused elevated stress as some pregnant women were in a position where they needed to enter the workforce or make non-trivial lifestyle changes.

in the first two quarters of 1919 and there should be no effects on those born in the fourth quarter of 1919.

To provide evidence on this prediction that distinguishes the two hypotheses, Part II of Panel B of the table displays estimates of [1] again allowing the effects on adult SES to vary with quarter of birth in 1919, controlling birth quarter fixed effects, and using the 1915 to 1918 cohorts for comparison. Results for years of completed education of males in 1960, 1970 and 1980 are presented. Males born in the fourth quarter have the largest deficits in education and those are the only deficits that are significantly different from zero in all three censuses. This result is not consistent with the fetal origins hypothesis. [19] As before, when controls for paternal characteristics are included in the models, all the 1919 birth cohort differences turn positive and they are consistently significant for those born in the first three quarters of 1919.

While selective fertility appears to provide a better explanation of the deficits in the 1919 birth cohort than *in utero* exposure to influenza, it is important to underscore that there are likely to be other factors underlying variation in outcomes across the cohorts. Almond (2006) concludes selective mortality is unlikely to be an important factor. It is not clear that is correct. Young adults and, especially, pregnant women were at greatest risk of mortality and Barry (2005), for example, reports that, among hospitalized patients, between one- and three-quarters of pregnant women died and, among those who survived, one-quarter miscarried. Bloom-Feshbach et al (2011) use data from Europe and 10 states in the U.S. and estimate that the pandemic caused a cumulative decrease in the number of births of 1.8 per 1000 in the 1919 cohort which translates into a 7.6% decline in the birth rate. They interpret this as evidence of elevated rates of miscarriage. Drawing on evidence from countries in Europe that remained neutral during WW1, Mamelund (2012) concludes that troop deployments, elevated morbidity and anxiety about the future contributed to the lower birth rate. It is possible that the 1919 birth cohort that survived to 1920 and 1930 is also selected although Tewksbury (1926) concludes that the pandemic had at most a modest impact on infant mortality of the 1919 cohort, relative to surrounding cohorts. The extent to which effects on births in 1919 and survival to the 1920 and 1930 Censuses are selective on socioeconomic characteristics is not clear although it is plausible that the lower

[19] Males born in the first quarter of 1920 have completed 0.12 fewer years of education than the comparison cohorts in 1960 and 1970 which is larger than the deficit of the 1919 first quarter births. The deficit, which is statistically significant in the 1970 Census, cannot be explained by in utero influenza exposure.

socioeconomic status mothers and children were at greatest risk. In that case, our estimates, based on parents of children alive in the 1920 and 1930 censuses will tend to understate the extent of negative selection of conceptions and live births.

Further, the health and human capital of infants and children alive at the time of the influenza pandemic were potentially affected both directly, through being infected, and indirectly. Indirect effects may have included, for example, changes in parental time allocations if a parent or other child was sick or if a father was absent because of mobilization. There may have been consumption changes that affected nutritional status in response to fewer resources because of loss of income due to business closures, changes in food prices (Rotwein, 1945), or in response to a campaign that encouraged reduced consumption of meat and other foods to assure sufficient food for soldiers. These effects complicate interpretation of estimates of the impact of *in utero* exposure to the pandemic as the older cohorts may have been affected; some of the effects may be correlated with state-level virulence of the virus which complicates interpretation of estimates that exploit variation in maternal mortality rates. While these mechanisms potentially affected health and human capital outcomes, they do not affect our conclusion that the parents of the 1919 cohort born in the U.S. are negatively selected.[20]

Replacing proxies with own paternal characteristics

In general, under reasonable assumptions, the use of proxies for paternal background will affect precision of estimates but not result in biased estimates. Evidence from Parman (2015) speaks directly to this issue. He uses World War II male enlistees matched with their records in the 1930 Census and thereby links individual outcomes of the enlistees in adulthood with the characteristics of their parents recorded in the census. Parman investigates whether having a sibling exposed to the influenza *in utero* affects one's own human capital and thus focuses on the siblings of the 1919 birth cohort. However, he reports models that compare the 1919 birth cohort with surrounding cohorts and while he does not discuss the results, he establishes that, after adjusting for own parental characteristics, the 1919 birth is no worse off than surrounding cohorts in terms of years of education attained, high school graduation and ever attended high school. In fact, as in our analysis, the estimated coefficients on the 1919 birth cohort indicator are positive and, in some cases, statistically

[20] An advantage of drawing comparisons with the 1915 through 1918 birth cohorts is that it puts aside potential fertility selection after 1919.

significant. These results confirm our conclusions examining adult outcomes from a different data source and using individual- specific parental characteristics. Proxies for paternal background that we have constructed do not appear to be a source of bias in our analyses.

In utero exposure to the 1918 pandemic and adult SES: Evidence from outside the U.S.

Following Almond (2006), several published studies have investigated how exposure to the influenza pandemic has affected socioeconomic outcomes in adulthood in other countries. These studies conclude that the exposed birth cohorts are worse off than the cohorts around them. We provide a brief review of this work, highlight several methodological concerns over and above selective fertility which the published studies have not addressed, and present evidence that these additional methodological concerns are not ignorable. We then discuss recent work by Vollmer and Wójcik (2017) who use data from over 100 censuses and find little systematic evidence for deleterious long-term economic effects of the Spanish flu.

Using data from Switzerland that measured outcomes in 1970, Neelsen and Stratmann (2012) compare the 1919 birth cohort with surrounding birth cohorts. Their results are mixed. High school completion of the 1919 birth cohort is not statistically different from the other cohorts at a 5% size of test. Nor is there a statistically significant difference in a measure of occupational status. The 1919 cohort is 0.3 percentage points more likely to have never married and 0.5 percentage points less likely to have a vocational degree. While significant, the magnitude of each effect is very small, not substantively important, and the welfare implications are not obvious.

In contrast, findings from Sweden reverse those from the U.S. with the exposed cohorts being economically better off in adulthood. Bengtsson and Helgertz (2015) use both a cohort comparison approach and variation in mortality at the county and birth month level to isolate the impact of exposure to influenza. Males exposed to higher intensity of influenza in utero earn significantly more at age 55 and also at age 60 and they are more likely to be employed in a high SES occupation. To put these results in context, it is worth noting that both Sweden and Switzerland were neutral during World War I. Whereas Sweden was far from the epicenter of the conflict, Switzerland was not. Not only was it directly affected by the Allied blockade but Switzerland deployed troops to its border and, as a result, likely experienced greater uncertainty and economic hardship than Sweden because of the war.

Lin and Liu (2014) investigate the relationship between human capital outcomes and in utero exposure to influenza during the 1918 and 1920 outbreaks in Taiwan. They examine height of the 1917 through 1920 cohorts. When comparing height of children of different ages, it is standard practice to standardize heights using growth curves that taken into account age and gender-specific linear growth; this is particularly important in this study because all the children were measured in 1927 and so the treatment and comparison cohorts are different ages. As shown in their tables, there are no significant differences in height-for-age z scores of the 1919 birth cohort relative to the other birth cohorts. Using the 1980 Taiwan census, they report that relative to 1916-26 cohorts, educational attainment of the 1919 cohort is significantly lower as is education attainment of females in the 1921 cohort. The authors acknowledge two concerns. First, the study cohorts were around age 55 to 65 in 1980. Since life expectancy at birth of these birth cohorts was less than 37 years (Liu, 2004), these estimates are potentially contaminated by selective mortality. Second, as discussed for the U.S. case, the estimates are potentially contaminated by selective fertility. Even if the 1918 influenza pandemic was unexpected, it is unlikely that the 1920 pandemic was also unexpected and so selective fertility is more likely to contaminate estimates for that birth cohort. It turns out the estimates of education gaps are no different for the 1919 and 1921 birth cohorts which suggests selection effects may be similar for both cohorts.

The issue of mortality selection also arises in another study in a developing country context, Brazil. Using annual labor force surveys collected between 1986 and 1998, Nelson (2010) reports that relative to the 1912 through 1922 trend, the 1919 birth cohort completed fewer years of education, were less likely to complete college and less likely to be employed. The study cohorts were age 64 through 74 in the first survey and age 76 through 86 in the final survey. However, life expectancy at birth for these cohorts was less than 50 years.

To empirically assess the importance of this concern, we have re-estimated the models for the same cohorts using data from the 1960, 1970, 1980, 1991 and 2001 Brazilian censuses and report the deviation from trend for the 1919 birth cohort in Appendix Table 4. We examine three education outcomes: attained years of schooling, completion of more than primary education, and completion of college. There are three key results. First, in the 1960 and 1970 censuses, those in the 1919 birth cohort completed significantly more years of education and are more likely to have completed more than primary education than predicted by the trend. These estimates are the least likely to be contaminated by mortality selection. Second, by 1980, the 1919 birth cohort advantage has been cut in half and is significant only for attained years of education. In the absence of mortality selection, the education differences should not change across the censuses and so we conclude that mortality selection is a relevant concern even before 1986, the first wave used by Nelson. Third, in 1991, the estimated effects are negative but not statistically significant. There is no evidence of differences in the 2001 census. Relative to the 1980 census, mortality selection has an even larger effect in the 1991 (and 2001) censuses. Taken together, the evidence presented in Appendix Table 4 suggests that the negative effects reported for 1986 through 1998 by Nelson are likely driven by mortality selection.

Over and above the methodological concerns in these publications, Vollmer and Wójcik (2017) highlight an additional concern in this literature: publication bias. We have not found any published studies in this literature that report no differences between the exposed and surrounding birth cohorts. Vollmer and Wójcik (2017) provide a systematic analysis of this question by estimating a cohort comparison model in the same vein as Almond (2006) for education, employment and disability outcomes using 117 Censuses from 53 countries. Not only do they find that the vast majority of estimates are not significantly different from zero but those that are different are equally likely to be positive as negative. They conclude that publication bias is a legitimate concern in this literature.

6. Conclusion

Almond (2006) established that, relative to surrounding birth cohorts, the 1919 birth cohort in the U.S. attained lower levels of adult SES. Since this birth cohort was in utero during the 1918 influenza pandemic, this result has been interpreted as evidence of the long-term economic effects of in utero exposure to health insults. A key assumption underlying this inference is that the 1919 birth cohort is exchangeable with surrounding birth cohorts. This paper has investigated that assumption and found that it does not hold.

Using data from the 1920 and 1930 Censuses, we have shown that the fathers of the 1919 birth cohort have lower levels of SES than the fathers of surrounding cohorts. Specifically, fathers of the 1919 birth cohort are less likely to be literate, have lower occupation income scores and Duncan SEI, are less likely to be white and are less likely to be WWI veterans. These results cannot be explained by age heaping in the censuses. Moreover, there is no evidence of an adult SES disadvantage among the 1919 birth cohort in models that adjust for paternal characteristics using proxies constructed from the 1920 and 1930 Censuses.

There is, at best, weak evidence that the intensity of *in utero* exposure to the influenza pandemic among the 1918 to 1920 cohorts born in the U.S. predicts SES in adulthood among the 1918 to 1920 cohorts born in the U.S. There is no evidence of this significant dose response effect, though, after adjusting for paternal background. .

We conclude that drawing inferences about the deleterious impact of *in utero* exposure to the 1918 influenza on SES in adulthood in the U.S. is premature. The evidence we present is consistent with a long line of inquiry that has shown parental background is a key predictor of success. It is also consistent with evidence that post-natal interventions can mitigate early life disadvantage. It is important to underscore that our results speak only to impacts of fetal health on markers of socioeconomic success in adulthood; they do not speak to whether in utero health insults affect biological health risks.

References

Almond, D. (2006). “Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy*, 114 (4), 672-712.

Ayers, L. (1919). “The War with Germany: A Statistical Summary.” Washington D.C.. UNT Digital Library. http://digital.library.unt.edu/ark:/67531/metadc276266/.

Barker, D. (1994). *Mothers, Babies, and Disease in Later Life*, London: BMJ Publishing Group.

Barry, J. (2005). *The Great Influenza: The Story of the Deadliest Pandemic in History*. Penguin Books.

Barry, J. (2004). “The site of origin of the 1918 influenza pandemic and its public health origins”. *Journal of Translational Medicine*, 2.3.

Beach, B., Ferrie, J. P. and Savedra M. H. (2018). “Fetal Shock or Selection? The 1918 Influenza Pandemic and Human Capital Development” NBER WP 24725.

Becker, G. (1960). “An Economic Analysis of Fertility." in:*Demographic and Economic Change in Developed Countries*, Princeton: Princeton University Press for the National Bureau of Economic Research.

Becker, G. and Lewis, H.G. (1973). “On the Interaction between the Quantity and Quality of Children.” *Journal of Political Economy*, 81:S279-88.

Bengtsson, T. and Helgertz, J. (2015). “The Long Lasting Influenza: The Impact of Fetal Stress during the 1918 Influenza Pandemic on Socioeconomic Attainment and Health in Sweden 1968-2012.” IZA Discussion Paper No. 9327.

Bloom-Feshbach, K., Simonsen, L., Viboud, C., Molbak, K., Miller, M., Gottfredsson, M., and Andreasen, V. (2011). "Natality decline and miscarriages associated with the 1918 influenza pandemic: The Scandinavian and United States Experiences." *The Journal of Infectious Diseases*. 204 (8): 1157-1164.

Boberg-Fazlic, N. Ivets, M., Karlsson, M., and Nilsson, T. (2016). “Disease and fertility: Evidence from the 1918 Spanish flu epidemic in Sweden”. Mimeo

Brainerd, E. and Siegler M. (2003). “The Economic Effects of the 1918 Influenza Epidemic." *Discussion Paper no. 3791, Centre Econ. Policy Res.*, Paris. British Medical Journal , July, 13 1918.

Brooks-Gunn, J. and Duncan G. (1997). “The Effects of Poverty on Children." *The Future of Children*, 7 (2), 55-71.

Corcoran, M., Gordon, R., Laren, D., and Solon, G. (1992). “The Association Between Men’s Economic Status and Their Family and Community Origins".*J. Hum. Resour.* 27, 575-601.

Floris, J., Mayr H., Staub, K., and Woitek, U. "Survival of the weakest? Culling evidence from the 1918 Flu Pandemic." International Health Economics Association Annual Congress, July 2017. Boston, MA

Heckman, J. J. (2006). “Skill formation and the economics of investing in disadvantaged children.” *Science.* 312.5782:1900-2.

Hommel, G. (1988). “A stagewise rejective multiple test procedure based on a modified Bonferroni test.” *Biometrika* 75:383–386.

Lin, M. and Liu, E. (2014). “Does in utero exposure to Illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment.” *Journal of Health Economics*, 37 (September), 152-163.

Neelsen, S. and Stratmann, T. (2012). “Long-run effects of fetal influenza exposure: Evidence from Switzerland.” *Social Science and Medicine*, 74 (1), 58-66.

Nelson, R. (2010). “Testing the fetal origins hypothesis in a developing country: Evidence from the 1918 influenza pandemic.” *Health Economics*, 19, 1181-1192.

Nudd, J. (2004). “U.S. World War I Draft Registrations." *Yesterdays,* 24 (1), 34-41.

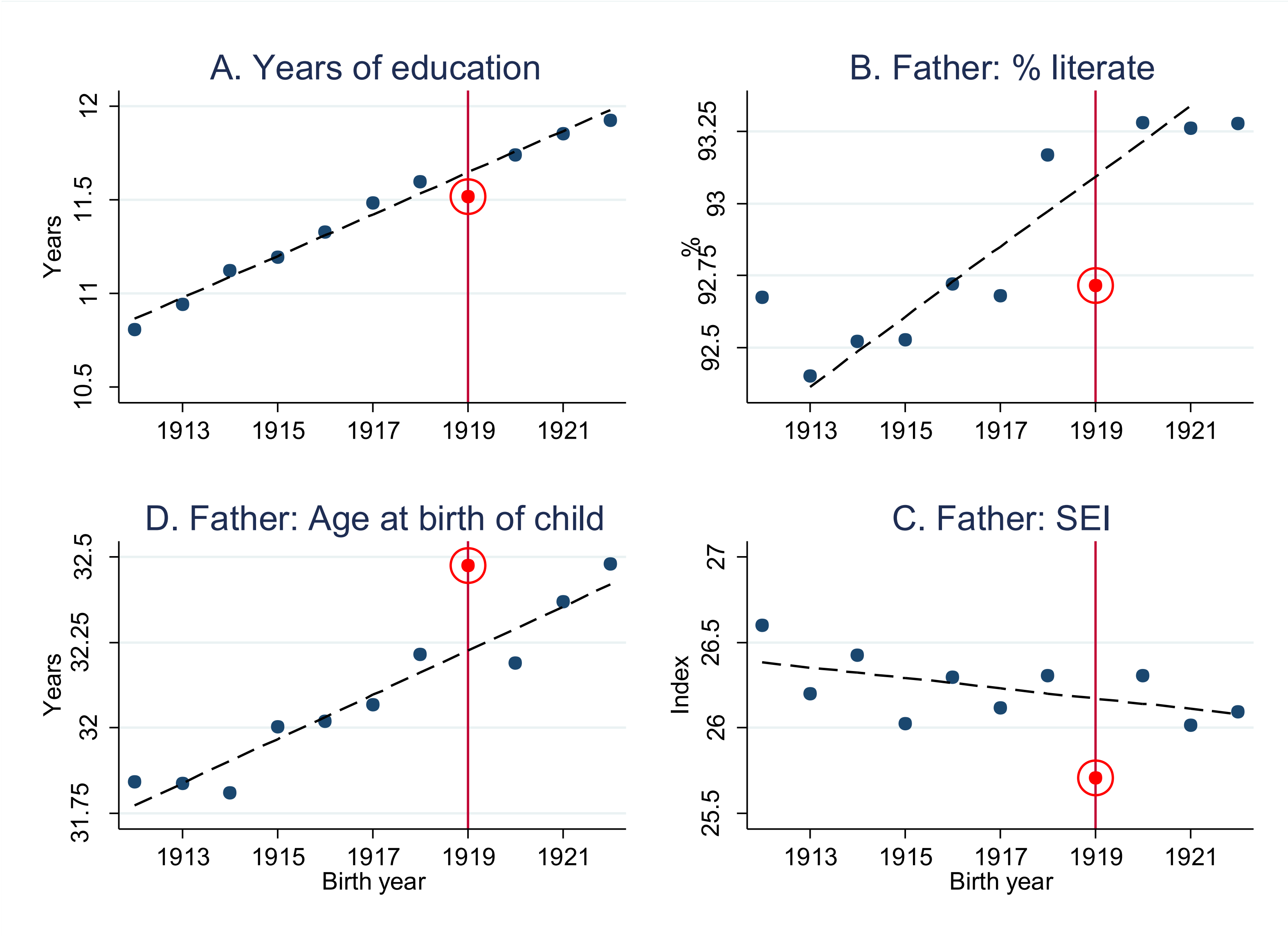
Parman, J. (2015) “Childhood health and sibling outcomes: Nurture reinforcing nature during the 1918 influenza pandemic.” *Explorations in Economic History* 58 (October), 22-43.

Rotwein, E. (1945). “Post-World War I Price Movements and Price Policy.” *Journal of Political Economy,* 53 (3), 234-257.

Thomas, D. (2010). “Health and Socioeconomic Status: The Importance of Causal Pathways." B. Pelskovic and J.Y. Lin (Eds.) *World Bank Annual Conference on Development Economics*.

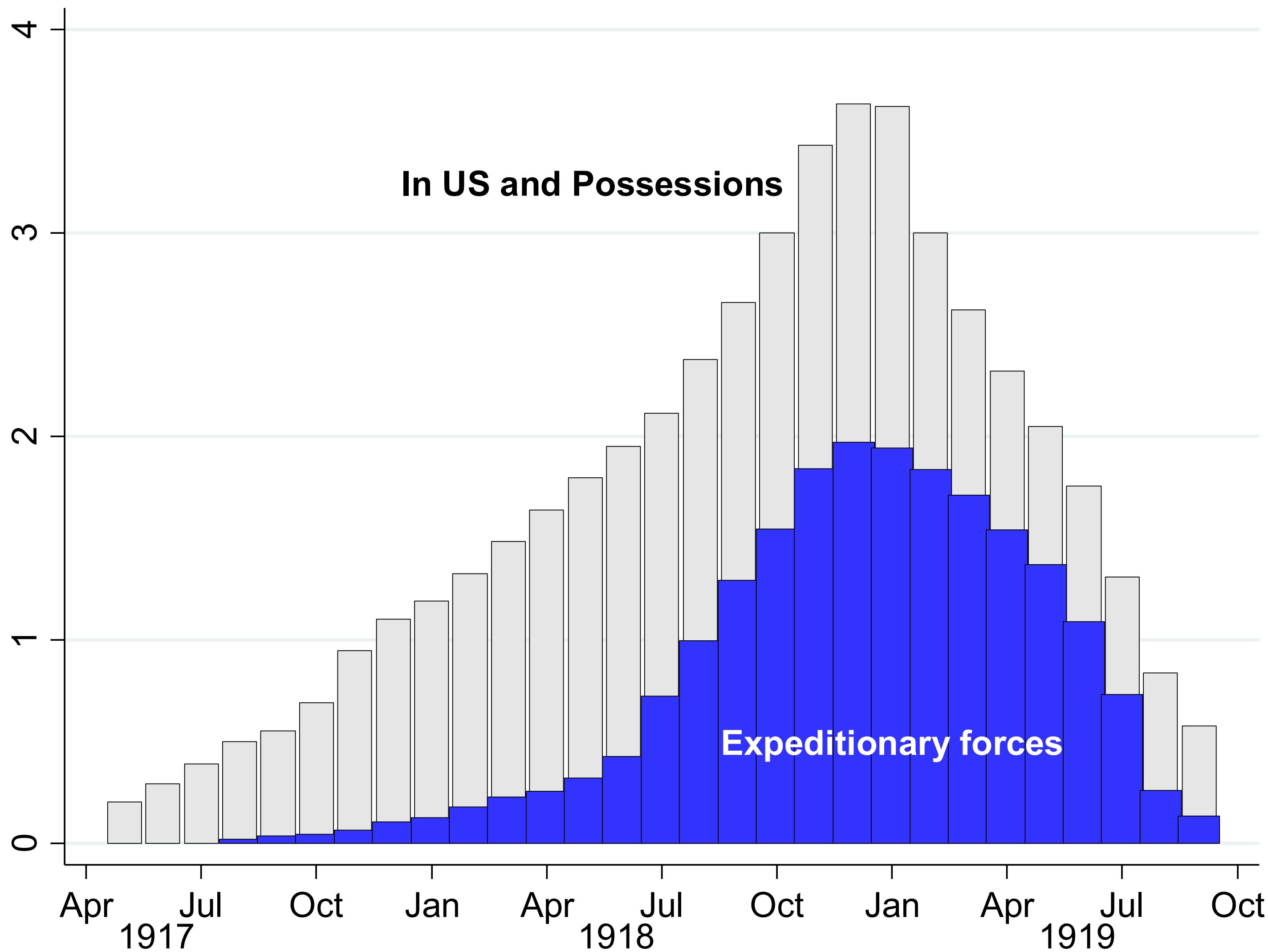
Vollmer, S. and Wójcik J. (2017). “The long-term consequences of the global 1918 influenza pandemic: A systematic analysis of 117 IPUMS international census data sets." Courant Research Centre: Discussion Paper No. 242.

Figure 1. Own education and paternal characteristics by own birth year



Sources: Public use samples of males from the 1960 Census (A) and 1930 Census (B-D).
A birth cohort is defined as from April 1 through March 31 of the following year in the 1930 Census.

Figure 2. Millions of soldiers in the American Army on the first of each month



Source: Ayers (1919) "*The War with Germany: A Statistical Analysis*"

Table 1
Differences in adult SES of 1919 birth cohort relative to surrounding cohorts Using males in the 1960 Census data

	Born in 1919		
	Relative to 1912-1922 cohorts		Relative to 1912-1918 cohorts
Socio-economic outcome in adulthood	Almond (2006)	Replication	
	(1)	(2)	(3)
1. High School Graduate	-0.021 **	-0.021 **	-0.022 *
	(0.005)	(0.005)	(0.009)
2. Years of Education (completed)	-0.150 **	-0.148 **	-0.188 **
	(0.038)	(0.039)	(0.064)
3. Total Income ($/month)	-573	-559	-539
	(295)	(292)	(498)
4. Wage Income ($/month)	-812 **	-802 **	-550
	(261)	(258)	(451)
5. Poor (<1.5 times the poverty level)	0.010 *	0.010 *	0.001
	(0.005)	(0.005)	(0.008)
6. Duncan's Socioeconomic Index	-0.640 *	-0.631 *	-0.884 *
	(0.259)	(0.260)	(0.436)
Observations	114,031	114,032	80,695

Notes: Estimates of β_3 from [1] and robust standard errors in parentheses reported for each dependent variable in column 1 and for each specification. Statistically significant at 5% (*) and 1% (**) size of test. All income values in 2005 dollars.

Table 2
Differences in paternal characteristics of 1919 birth cohort relative to surrounding cohorts
Using males in the 1920 and the 1930 Census data

	A. 1920 U.S. Census		B. 1930 U.S. Census	
		Born in 1919		Born in 1919
Paternal Characteristic	Mean	Relative to 1912-1918 cohorts	Mean	Relative to 1912-1922 cohorts
	(1)	(2)	(3)	(4)
1. Father is Illiterate (%)	9.21%	1.21% **	7.60%	0.29% **
		(0.05)		(0.03)
2. Father's Occupation Income Score	21.68	-0.23 **	22.74	-0.17 **
		(0.02)		(0.01)
3. Father's Duncan's Socioeconomic Index	22.60	-0.75 **	24.73	-0.35 **
		(0.04)		(0.02)
4. Father is Non-White (%)	15.91%	0.77% **	10.20%	0.81% **
		(0.06)		(0.03)
5. Father's Age at Birth	32.89	0.22 **	32.10	0.26 **
		(0.01)		(0.01)
6. Number of Father's Children in HH	3.67	0.32 **	4.12	0.07 **
		(0.00)		(0.00)
7. Father's Age at Marriage			24.59	-0.05 **
				(0.01)
8. Father is a WWI Veteran (%)			6.65%	-1.20% **
				(0.03)
Observations	9,335,388	9,335,388	12,175,857	12,175,857

Notes: Estimates of β_3 from [1] for each paternal characteristic. Robust standard errors in parentheses. Statistically significant at 5% (*) and 1% (**) size of test.
Trend estimates are from January 1, 1912 through December 31, 1919 in column A and from April 1, 1911 to March 31, 1923 in column B.
Birth cohorts are from January 1 through December 31 (in column A) and from April 1 through March 31 of the following year (in column B).

Table 3
Differences in adult SES of 1919 birth cohort relative to surrounding cohorts with and without paternal controls calculated using 1930 Census data

	A. 1960 Census		B. 1970 Census		C. 1980 Census	
Socio-economic outcome in adulthood	No paternal controls [1]	w/ paternal controls [2]	No paternal controls [1]	w/ paternal controls [2]	No paternal controls [1]	w/ paternal controls [2]
1. High School Graduate	-0.021 **	0.002	-0.020 **	0.008 *	-0.014 **	0.018 **
	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
2. Years of Education (completed)	-0.148 **	0.092 *	-0.178 **	0.053 *	-0.117 **	0.149 **
	(0.039)	(0.037)	(0.023)	(0.023)	(0.020)	(0.019)
3. Total Income ($/month)	-559	1,140 **	-1,218 **	989 **	-1,051 **	721 **
	(292)	(290)	(250)	(250)	(189)	(190)
4. Wage Income ($/month)	-802 **	572 *	-864 **	987 **	-679 **	720 **
	(258)	(256)	(230)	(229)	(177)	(178)
5. Poor (<1.5 times the poverty level)	0.010 *	-0.020 **	0.009 **	-0.008 **	0.006 **	-0.007 **
	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
6. Duncan's Socioeconomic Index	-0.631 *	0.627 *	-0.806 **	0.432 **	-0.813 **	0.470 **
	(0.260)	(0.259)	(0.157)	(0.158)	(0.137)	(0.138)
7. Disability Limits Work			0.005 **	0.001	0.005 *	-0.008 **
			(0.002)	(0.002)	(0.002)	(0.002)
8. Disability Prevents Work			0.004 **	0.000	0.001	-0.009 **
			(0.001)	(0.001)	(0.002)	(0.002)
9. Welfare Income ($/month)			12.281 *	11.902 *	16.936 *	8.733
			(5.844)	(6.053)	(7.039)	(7.267)
10. Social Security Income ($/month)			5.364	-9.393	81.687 **	75.180 **
			(9.228)	(9.383)	(18.658)	(19.350)
Observations	114,032	114,032	308,785	308,785	471,803	471,803

Notes: Robust standard errors are in parenthesis. Statistically significant at 5% (*) and 1% (**) size of test. All income values in 2005 dollars.

A birth cohort is defined as from April 1 through March 31 of the following year in the 1930 Census.

Table 4
Differences in adult SES of 1919 birth cohort relative to 1912-1918 cohorts with and without paternal controls calculated using 1920 Census data

Socio-economic outcome in adulthood	A. 1960 Census No paternal controls [1]	A. 1960 Census w/ paternal controls [2]	B. 1970 Census No paternal controls [1]	B. 1970 Census w/ paternal controls [2]	C. 1980 Census No paternal controls [1]	C. 1980 Census w/ paternal controls [2]
1. High School Graduate	-0.022 *	0.035 **	-0.018 **	0.041 **	-0.012 **	0.045 **
	(0.009)	(0.009)	(0.005)	(0.006)	(0.004)	(0.004)
2. Years of Education (completed)	-0.188 **	0.266 **	-0.169 **	0.242 **	-0.161 **	0.258 **
	(0.064)	(0.062)	(0.039)	(0.038)	(0.033)	(0.033)
3. Total Income ($/month)	-539	2,790 **	-644	3,060 **	-775	1,770 **
	(498)	(504)	(427)	(433)	(486)	(489)
4. Wage Income ($/month)	-550	2,160 **	-927 *	2,314 **	-1,085 *	-132
	(451)	(452)	(393)	(397)	(440)	(447)
5. Poor (<1.5 times the poverty level)	0.001	-0.037 **	0.004	-0.006	0.016 **	0.009
	(0.008)	(0.008)	(0.004)	(0.004)	(0.005)	(0.005)
6. Duncan's Socioeconomic Index	-0.884 *	1.181 **	-0.471	1.306 **	-0.218	1.389 **
	(0.436)	(0.441)	(0.267)	(0.271)	(0.234)	(0.240)
7. Disability Limits Work			0.005	0.003	0.006	-0.008 *
			(0.004)	(0.004)	(0.004)	(0.004)
8. Disability Prevents Work			0.001	-0.004	0.008 *	-0.005
			(0.003)	(0.003)	(0.003)	(0.004)
9. Welfare Income ($/month)			18.154	17.143	10.846	15.567
			(9.338)	(9.883)	(19.704)	(20.360)
10. Social Security Income ($/month)			-22.503	-48.289 **	770.069 **	699.787 **
			(16.532)	(17.334)	(59.434)	(61.671)
Observations	80,695	80,695	216,633	216,633	323,089	323,089

Notes: Robust standard errors are in parenthesis. Statistically significant at 5% (*) and 1% (**) size of test. All income values in 2005 dollars.

Table 5
Differences of 1919 birth cohort relative to surrounding cohorts: Females and nonwhite males
Paternal characteristics measured in 1920 Census and adult outcomes measured in 1960, 1970 and 1980 Censuses

	A. Paternal characteristics	B. Outcomes in adulthood					
		B1. 1960 Census		B2. 1970 Census		B3. 1980 Census	
	1919 cohort dev from trend	No paternal controls	w/ paternal controls	No paternal controls	w/ paternal controls	No paternal controls	w/ paternal controls
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
I. Females							
A. Paternal characteristics							
A.1. Father is Illiterate (%)	1.32% **						
	(0.05)						
A.2. Father's Occupation Income Score	-0.27 **						
	(0.02)						
A.3. Father's Duncan's SES Index	-0.76 **						
	(0.04)						
B. Outcomes in adulthood							
B.1. High School Graduate		-0.008	0.030 **	-0.021 **	0.023 **	-0.007	0.040 **
		(0.009)	(0.009)	(0.005)	(0.005)	(0.004)	(0.004)
B.2. Years of Education (completed)		-0.077	0.192 **	-0.176 **	0.094 **	-0.073 **	0.222 **
		(0.054)	(0.053)	(0.032)	(0.032)	(0.026)	(0.026)
B.3. Total Income (2005$/month)		211	849 **	-213	650 **	565 *	1,550 **
		(221)	(230)	(190)	(197)	(233)	(237)
Observations		83,730	83,730	233,482	233,482	383,531	383,531
II. Nonwhites							
A. Paternal characteristics							
A.1. Father is Illiterate (%)	0.95% **						
	(0.19)						
A.2. Father's Occupation Income Score	-0.05						
	(0.03)						
A.3. Father's Duncan's SES Index	-0.17 **						
	(0.04)						
B. Outcomes in adulthood							
B.1. High School Graduate		-0.015	0.006	-0.017	0.007	-0.011	0.014
		(0.017)	(0.016)	(0.011)	(0.010)	(0.008)	(0.008)
B.2. Years of Education (completed)		-0.039	0.132	-0.210 *	-0.081	-0.112	0.066
		(0.155)	(0.150)	(0.095)	(0.092)	(0.078)	(0.077)
B.3. Total Income (2005$/month)		698	929	-90	196	534	1,142 *
		(566)	(575)	(500)	(499)	(574)	(576)
Observations		15,995	15,995	41,726	41,726	71,227	71,227

Notes: Robust standard errors in parenthesis. Statistically significant at 5% (*) and 1% (**) size of test. Deviations from trend estimated using 1912-1919 birth cohorts.

Table 6. Age heaping in 1920 and 1930 Census

A. Comparison of census and vital statistics births by cohort

	1. 1920 Census		2. 1930 Census	
Reported year of birth	Reported age at Census	Census/ Vital Statistics	Calculated age at Census	Census/ Vital Statistics
	[1]	[2]	[1]	[2]
1922	-		7	1.05
1921	-		8	1.02
1920	-		9	1.07
1919	0	0.97	10	1.11
1918	1	0.99	11	1.08
1917	2	1.02	12	0.98
1916	3	1.04	13	1.03
1915	4	1.01	14	1.03

B. Heaping on 1919 birth cohort in 1930 Census and paternal characteristics

		1919 birth cohort deficits	
Population group	Ratio of (# of 10 yr olds)/ (avg. # of 9 & 11 yr olds) [1]	Paternal Occupation Income score [2]	Paternal Duncan's SEI [3]
1. All children	1.034	-0.17	-0.35
		(0.01)	(0.02)
2. Father: Illiterate	1.113	.	.
3. Father: Literate	1.028	-0.15	-0.32
		(0.01)	(0.02)
4. Father: Literate & white	1.020	-0.12	-0.26
		(0.01)	(0.03)
5. Father: Literate, white & US born	1.011	-0.14	-0.30
		(0.02)	(0.03)

Notes: 1920 Census birth cohorts in A based on month and year of birth. 1930 Census birth cohorts in A and B based on age in years taking into account enumeration was on April 1, 1930. Vital statistics for states covered in each year, adjusted for mortality. Defivits in paternal characteristics are deviation from trend for 1912-1922 cohorts in 1930 Census. Robust standard errors in parentheses.

Table 7

Impact of maternal infection rate in year before birth on adult socio-economic status in the 1960 Census for males born between 1918 and 1920 with and without backgrtound controls

Socio-economic outcome in adulthood	Almond (2006) (1)	Replication (2)	Corrected (3)	Adjusted w/ background characteristics (4)	P-value for test (3) = (4) (5)
1. High School Graduate	-0.101 **	-0.104 **	-0.086 *	-0.046	0.001
	(0.070)	(0.036)	(0.041)	(0.043)	
2. Years of Education (completed)	-0.756 **	-0.756 **	-0.692 *	-0.366	0.032
	(0.259)	(0.254)	(0.322)	(0.368)	
3. Log of Total Income	-0.165 **	-0.171 *	-0.166	-0.078	0.168
	(0.072)	(0.071)	(0.091)	(0.108)	
4. Poor (< 1.5 times the poverty level)	0.042	0.042	0.032	-0.003	0.036
	(0.026)	(0.026)	(0.033)	(0.036)	
5. Duncan's Socioeconomic Index	-2.711	-2.778	-2.393	0.179	0.080
	(1.735)	(1.660)	(2.035)	(2.682)	
Observations	16,566	16,566	16,659	16,659	

Notes: Standard errors clustered at the state and year of birth level are in parenthesis. Statistically significant at 5% (*) and 1% (**) size of test.

Table 8
Differences in characteristics of fathers who were WWI veterans relative to other fathers using the 1912 to 1922 male birth cohorts from the 1930 Census data

Paternal Characteristic	Mean [1]	WWI Veteran [2]
1. Father is Illiterate (%)	7.60%	-1.22% **
		(0.02)
2. Father's Occupation Income Score	22.74	3.29 **
		(0.02)
3. Father's Duncan's Socioeconomic Index	24.73	6.25 **
		(0.03)
4. Father is Non-White (%)	10.20%	-1.83% **
		(0.03)
Observations	12,175,857	12,175,857

Notes: Robust standard errors are in parenthesis. Statistically significant at 5% (*) and 1% (**) size of test.

Birth cohorts are from April 1 through March 31 of the following year.

Specification uses age of father, age of father squared, father's state of birth fixed effects, an indicator for being a white father, and an indicator for being a WWI veteran father as the independent variables.

Appendix Table 1

Departure of 1919 Birth Cohort from Trend[1] for Males Relative to the 1915-1918 Cohorts

	A. Year of birth difference	B. Quarter of birth differences				
	Born in 1919	1919Q1	1919Q2	1919Q3	1919Q4	Missing Birth Month in 1919
	(1)	(2)	(3)	(4)	(5)	(6)
I. Paternal characteristics in 1920 Census						
1. Father is Illiterate	1.20% **	2.10% **	1.28% **	1.89% **	0.53% **	-1.07% *
	(0.08)	(0.10)	(0.10)	(0.10)	(0.10)	(0.43)
2. Father's Occupation Income Score	-0.04	-0.25 **	-0.47 **	0.05	0.26 **	0.57 **
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.16)
3. Father's Duncan's Socioeconomic I	-0.60 **	-0.98 **	-1.07 **	-0.62 **	-0.14 *	1.00 **
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.31)
4. Father is Non-White (%)	1.06% **	2.34% **	1.97% **	1.42% **	-0.19%	4.11% **
	(0.10)	(0.11)	(0.11)	(0.12)	(0.12)	(0.56)
5. Father's Age at Birth	0.30 **	0.32 **	0.60 **	0.38 **	0.08 **	-0.37 *
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.15)
6. Number of Father's Children in HH	0.36 **	0.30 **	0.42 **	0.42 **	0.36 **	0.18 **
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
II. Completed years of education						
1. Measured in 1960 Census	-0.209 *	-0.159	-0.223	-0.138	-0.310 *	
	(0.101)	(0.122)	(0.122)	(0.123)	(0.121)	
2. Measured in 1970 Census	-0.182 **	-0.095	-0.152 *	-0.174 *	-0.320 **	
	(0.062)	(0.074)	(0.074)	(0.074)	(0.073)	
3. Measured in 1980 Census	-0.112 *	-0.108	-0.083	-0.055	-0.203 **	
	(0.053)	(0.063)	(0.064)	(0.063)	(0.062)	

Notes: ** indicates statistical significance at the 1% level, * indicates statistical significance at the 5% level. Robust standard errors in parentheses. 5,767,400 males in panel I and 51,462, 139,757 and 213, 481 in panels II rows 1 through 3, respectively. [1]Trend estimates are from January 1, 1915 through December 31, 1919.

Birth quarter regressions use model [1] replacing the 1919 birth cohort dummy variable with four 1919 birth quarter cohort indicator variables and birth quarter fixed effects.

Models of paternal characteristics include indicator variable for missing birth month in 1919 and a missing birth month indicator for all cohorts.

Appendix Table 2

Difference in difference estimates of maternal infection rates on adult socio-economic outcomes

Outcomes for males, females and nonwhites born between 1918 and 1920 in 1960, 1970 and 1980 censuses

	A. Males			B. Females			C. Non-whites		
Census:	1960	1970	1980	1960	1970	1980	1960	1970	1980
Socio-economic outcome in adulthood	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. High School Graduate	-0.086 *	0.031	0.006	0.015	0.020	-0.005	-0.106	0.020	0.031
	(0.041)	(0.020)	(0.014)	(0.033)	(0.019)	(0.017)	(0.084)	(0.050)	(0.032)
2. Years of Education (completed)	-0.692 *	0.344 **	-0.010	0.018	0.063	-0.032	-0.018	-0.086	-0.026
	(0.322)	(0.134)	(0.092)	(0.207)	(0.116)	(0.090)	(0.635)	(0.234)	(0.255)
3. Log of Total Income	-0.166	0.062	-0.078	-0.166	0.017	0.072	0.098	-0.155	0.298 **
	(0.091)	(0.041)	(0.042)	(0.102)	(0.037)	(0.041)	(0.139)	(0.112)	(0.093)
4. Poor (< 1.5 times the poverty level)	0.032	-0.053 **	0.020	-0.008	0.015	-0.033 *	-0.140	0.013	-0.037
	(0.033)	(0.013)	(0.011)	(0.029)	(0.015)	(0.015)	(0.094)	(0.049)	(0.032)
5. Duncan's Socioeconomic Index	-2.393	2.017	-1.584	-1.582	1.271	0.745	-1.105	4.348 *	0.495
	(2.035)	(1.287)	(0.823)	(1.482)	(0.900)	(0.728)	(2.979)	(1.979)	(1.600)
Observations	16,659	45,987	70,688	17,058	49,081	80,459	1,820	5,132	8,064

Notes: Standard errors clustered at the state and year of birth level are in parenthesis. Statistically significant at 5% (*) and 1% (**) size of test.

Appendix Table 3
Differences in adult SES of 1919 birth cohort relative to surrounding cohorts controlling state- and birth-year race-specific % of fathers that are WW1 veterans

Socio-economic outcome in adulthood	A. 1960 Census [1]	B. 1970 Census [2]	C. 1980 Census [3]
1. High School Graduate	0.000	0.004	0.012 **
	(0.005)	(0.003)	(0.003)
2. Years of Education (completed)	0.036	0.006	0.087 **
	(0.039)	(0.023)	(0.020)
3. Total Income ($/month)	577	274	921 **
	(295)	(253)	(192)
4. Wage Income ($/month)	149	370	619 **
	(262)	(232)	(180)
5. Poor (<1.5 times the poverty level)	-0.007	-0.002	-0.009 **
	(0.005)	(0.002)	(0.002)
6. Duncan's Socioeconomic Index	0.228	0.096	0.097
	(0.262)	(0.159)	(0.139)
7. Disability Limits Work		0.004 *	-0.002
		(0.002)	(0.002)
8. Disability Prevents Work		0.001	-0.005 **
		(0.001)	(0.002)
9. Welfare Income ($/month)		10.745	1.030
		(6.041)	(7.224)
10. Social Security Income ($/month)		-3.246	72.238 **
		(9.288)	(19.130)
Observations	114,032	308,785	471,803

Notes: Robust standard errors are in parenthesis. All income figures are given in 2005 dollars. Statistically significant at 5% (*) and 1% (**) size of test.
A birth cohort is defined as from April 1 through March 31 of the following year in the 1930 Census.

Appendix Table 4

Differences in adult SES of 1919 birth cohort relative to surrounding cohorts (1912-1922)

Using 1960, 1970, 1980, 1991, and 2000 Brazilian Census data samples from IPUMS

	1960 Census	1970 Census	1980 Census	1991 Census	2001 Census
Education outcome in adulthood	(1)	(2)	(3)	(4)	(5)
1. More than primary education	0.006 **	0.007 **	0.003	-0.003	-0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
2. College graduate	0.003	0.003	0.000	-0.002	-0.002
	(0.022)	(0.021)	(0.023)	(0.024)	(0.034)
3. Attained Years of Schooling	0.323 **	0.334 **	0.151 **	-0.032	0.036
	(0.022)	(0.021)	(0.023)	(0.024)	(0.034)
Observations	326,004	365,026	304,582	236,001	141,591

Notes: Estimates of β_3 from [1] and robust standard errors in parentheses. Statistically significant at 5% (*) and 1% (**) size of test.

The 1960, 1970, and 1980 data represents 5% samples, the 1991 data represents a 5.8% sample, and the 2000 data represents a 6% sample.

Birth cohorts are defined from September 1 of the previous year through August 31 of the stated birth cohort. Trend estimates are from September 1, 1911 to August 31, 1923.

Tthe 1919 birth cohort represents births from September 1st, 1918 to August 31, 1919.