Northwestern | INSTITUTE FOR POLICY RESEARCH

Working Paper Series

WP-18-15

Fetal Shock or Selection? The 1918 Influenza Pandemic and Human Capital Development

Brian Beach
Assistant Professor of Economics
College of William & Mary

Joseph Ferrie
Professor of Economics
IPR Associate
Northwestern University

Martin Saavedra
Assistant Professor of Economics
Oberlin College

Version: June 20, 2018

DRAFT

ABSTRACT

Almond (2006) argues that in utero exposure to the 1918 influenza pandemic lowered socioeconomic status in adulthood, whereas Brown and Thomas (2016) find that the effect disappears after controlling for parental characteristics of the 1919 birth cohort. In this paper, the researchers link microdata from the 1920 and 1930 censuses to WWII enlistment records and city-level influenza data. The result is a data set with much more precisely measured influenza exposure and parental characteristics. Results indicate that in the absence of the pandemic, the 1919 birth cohort would have been more likely to graduate from high school and would have obtained more years of schooling. The impact on high school graduation is largely unaffected by including parental controls and city-specific time trends. Adding household fixed effects (and thus exploiting variation among brothers) yields similar but somewhat larger results.

1 Introduction

Following the influential work of Almond (2006), economists have become increasingly interested in understanding the extent to which early-life exposure to disease and deprivation affects adult outcomes. In that paper, Almond argued that the 1918 influenza pandemic offers a natural experiment for testing Barker's 1995 hypothesis that malnutrition while in utero increases susceptibility to heart disease. To this end, Almond compared adult outcomes of birth cohorts that were exposed to the pandemic while in utero to adjacent birth cohorts and found evidence that in utero flu exposure not only decreased health in adulthood but that it also lowered educational attainment, income, and socioeconomic status (SES) (Almond, 2006; Almond & Mazumder, 2005). Economists have since applied variations of this identification strategy to assess the implications of fetal shocks on individual outcomes and overall economic development. The majority of these studies yield consistent results: early-life exposure to disease and deprivation lowers SES and health in adulthood, which has important implications for development.[1]

The critical identifying assumption in this literature is that unobserved determinants of human capital vary continuously between birth cohorts. Recent work by Brown & Thomas (2016) casts doubt on this assumption, at least with respect to its application to the study of the 1918 influenza pandemic. Brown and Thomas draw on data from the 1920 and 1930 US censuses to test whether parents whose children were in utero during the pandemic were observably different from parents of children from adjacent birth cohorts. The authors find that parents of the exposed cohort were less likely to be literate, were less likely to be white, and were of lower SES. Brown and Thomas suggest this is due to the fact that the pandemic coincided with

[1]See Almond & Currie (2011) for an overview of research on the link between in utero conditions and individual outcomes. Notable studies examining the long-run economic impacts of early exposure to other infectious diseases include Bleakley (2007) on hookworm, Barreca (2010) on malaria, Beach *et al.* (2016) on typhoid fever, and Saavedra (2017) on yellow fever.

World War I (WWI), and WWI veterans were positively selected from the overall population. Importantly, Brown and Thomas show that many of Almond's findings are not robust to controlling for parental characteristics.

We make two contributions to this debate. First, we argue that measurement error may explain these diverging findings. Specifically, we show that while there was some selection amongst the 1919 birth cohort in general, the extent of that selection is overstated when one does not account for two important confounders: age heaping and the SES-child mortality gradient. In light of these issues, our second contribution is to adopt an identification strategy that allows us to more credibly recover a causal effect. Specifically, we draw on linked microdata which allows us to exploit geographic variation in the intensity of the pandemic as a source of identification. All of our results include birth year fixed effects, which account for the general rise in WWI deployment, and for a subset of our sample we can include household fixed effects. Results indicate that in utero exposure to the influenza pandemic did impair human capital development.

This paper begins by examining whether parental characteristics were systematically different in 1919, as suggested by Brown and Thomas (henceforth, BT). The evidence BT provide comes from the 1920 and 1930 censuses where birth cohort is inferred from reported age. This approach is complicated by two factors: age heaping and infant mortality. The issue of age heaping arises from the fact that birth year must be inferred from the census data. Given the timing of the pandemic and enumeration, this means that parents of the exposed birth cohort should report their child as age 0 in 1920 and age 10 in 1930. However, it is well documented that individuals of lower cognitive ability are more likely to report ages ending in a 5 or 0 (A'Hearn *et al.*, 2009), which offers an alternative explanation for the relationship documented by BT. As for the role of infant mortality, the SES infant mortality gradient was particularly pronounced during the early 20th century (Boustan & Margo, 2014; Collins & Thomasson, 2004). When BT observe that parents of 0 year olds are observably

different in the 1920 census, this could be because the children of low SES parents are less likely to survive past infancy. To the extent that those children die before appearing in later censuses, the aggregate parental characteristics obtained from the 1920 census will be systematically mismeasured.

To illustrate these points more broadly, we show that all children appearing in the 1910, 1920, and 1930 censuses with reported ages of zero or ten have lower SES parents. In other words, while parents that report having a child born in 1919 are of lower average SES, so are parents that report having a child born in 1899, 1909, and 1929. We estimate that approximately half of the differences documented by BT can be accounted for by age heaping. We also provide a simple model demonstrating why including these systematically mismeasured parental characteristics is likely to bias estimates towards finding that the 1919 cohort performed *better* than adjacent cohorts, which is exactly what BT find. Since not all of the parental differences can be explained by age heaping, the concern that the Almond (2006) estimates are confounded by parental selection cannot be ruled out without additional analysis.

After illustrating this issue, we then introduce an alternative way of examining whether in utero exposure to the pandemic impaired human capital development. Specifically, we construct a new data set linking young adult males from World War II (WWII) enlistment records to their corresponding record in the 1920 or 1930 censuses. This allows us to identify residence at the time of enumeration, which we interpret as their in utero environment. Because we observe individuals with their parents in the census, we collect (and subsequently control for) parental characteristics such as mother's and father's literacy, nativity, age, and occupational income score, as well as whether the family owned or rented their home and the number of siblings.

Next we digitize annual city-level disease data to construct a measure of pandemic intensity. This allows us to move away from the cohort analysis employed by Almond (2006) and Brown & Thomas (2016) and instead adopt a difference-in-differences identification strategy that includes birth cohort fixed effects. This means that our

empirical strategy first compares differences in long-run outcomes for those in utero during the pandemic relative to adjacent cohorts and then asks whether that difference varies with respect to the intensity of the city's exposure to the pandemic. In other words, our identifying assumption is now that unobserved determinants of human capital are not systematically related to pandemic intensity. Taking parental characteristics as outcome variables indicates that this assumption is likely valid.

Turning to long-run outcomes, we find robust evidence that in utero pandemic exposure decreased the likelihood of graduating high school. This is true before and after we control for parental characteristics and even when we control for city-specific time trends. For the set of families that had more than one child enlist in WWII, we are able to adopt a household fixed effects model, and we continue to find that the pandemic decreased educational attainment. As for biological outcomes, we find no consistent evidence that in utero exposure to the pandemic affected heights, weights, or BMI.

While these results clearly contribute to the debate between Almond (2006) and Brown & Thomas (2016), they may have broader implications for other early-life exposure papers. The majority of this literature adopts an empirical framework that pairs variation in cohort-specific exposure to epidemics or policy changes as a source of identification. As Brown & Thomas (2016) point out, it is important to understand whether other determinants of human capital are related to treatment. Linked microdata allow the researcher to more explicitly test this assumption. However, as we show, should this assumption fail it may still be possible to recover a causal effect by exploiting intensity of treatment as a source of identification. While we focus our attention on the 1918 influenza pandemic debate, this approach is likely to be useful for future scholars as well. For instance, this approach may help shed light on the debate over the long-run effects of eradicating hookworm in the American south (see Roodman (2018), Bleakley (2007), and Bleakley (2018)).

2 Age heaping and inference

In this section, we document age heaping in the 1920 and 1930 censuses and consider the implications for estimating the causal impacts of the 1918 influenza pandemic on later life outcomes. Subsection 2.1 documents that 1920 and 1930 census data likely overstate differences in parental characteristics across birth cohorts. Subsection 2.2 shows that age heaping largely ceases by the 1960-1980 censuses, implying cohorts are likely measured more accurately for those census years. Lastly, subsection 2.3 provides a simple model that demonstrates how age heaping in the 1920 and 1930 censuses can lead to biased estimates.

2.1 Were parents of the 1919 birth cohort different?

As noted previously, Almond (2006) compares the adult outcomes of the 1919 birth cohort, who would have been in utero during the 1918 influenza pandemic, to adjacent birth cohorts and attributes the differences in outcomes to the effect of the pandemic. Since Almond controls for birth cohort using a polynomial, the identifying assumption is that unobservable characteristics affecting human capital varied continuously across birth cohorts. BT cast doubt on this assumption, arguing that the 1919 birth cohort had systematically different parents since the height of WWI deployment corresponded with the pandemic. Thus, the pool of potential fathers for children born in 1919 was systematically different than the pool of potential fathers from adjacent birth cohorts. BT provide evidence for this story by examining the 1920 and 1930 censuses. Unfortunately, these censuses did not record an individual's birth year but instead recorded their age. BT infer birth year by taking the enumeration year minus the reported age minus 1.[2] Consequently, those who are inferred to have been from

[2]The 1920 census instructed enumerators to obtain an individual's age as of January 1, 1920. All children born after January 1, 1919 should thus be recorded as age 0, and so taking enumeration year minus age minus 1 allows us to recover an individual's age. The 1930 census instructed enumerators to obtain an individual's age as of April 1, 1930. Thus, measuring birth year as year minus age

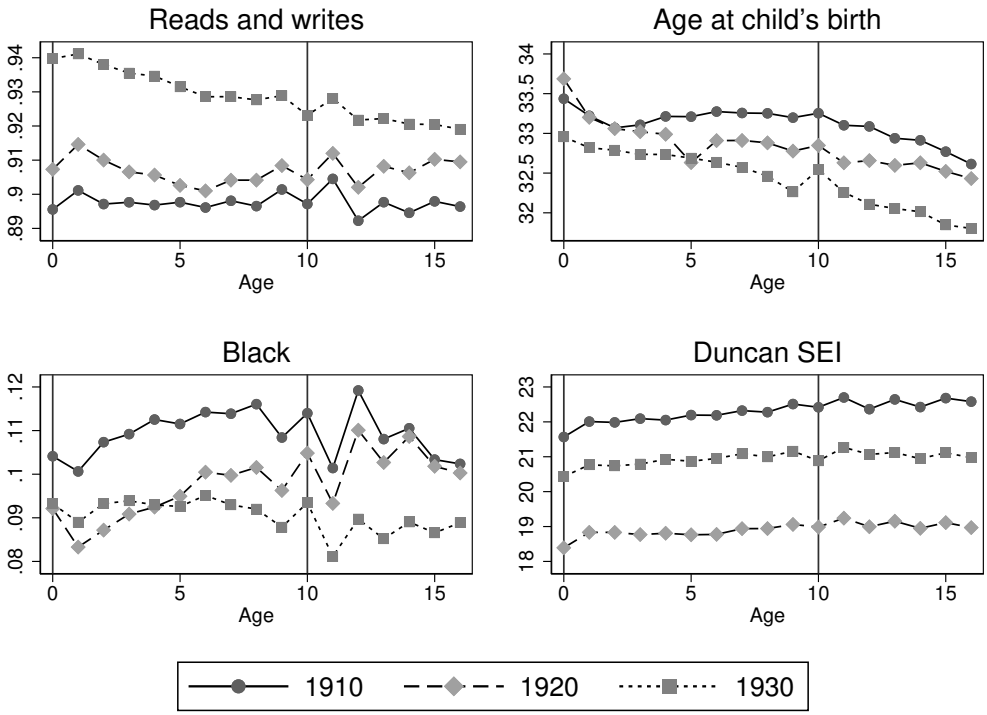
the 1919 birth cohort are those whose parents reported their child as 0 years old in the 1920 census or 10 years old in the 1930 census.

This fact complicates inference because previous studies have shown that lower SES individuals are more likely to report ages that are even numbers or numbers ending with a 0 or 5 (see A'Hearn *et al.* (2009), Parman (2017), and the citations therein). Because of this, we might expect parents who report having a 10 year old to systematically differ from parents of reported 9 and 11 year olds even if the parents of actual 9, 10, and 11 year olds had the same distribution of characteristics. Obtaining parental characteristics for children that are reported as age 0 is further complicated by the fact that infant mortality rates were more pronounced for lower SES families during this period. Non-white infants were about 75 percent more likely to die than white infants during the early 20th century (Collins & Thomasson, 2004). Consequently, parents of 0 year olds in the census appear to have lower SES than older children, regardless of the census year, because high SES children are over-represented among children who survive infancy.

To demonstrate the extent of these problems, we draw on the full count 1910, 1920, and 1930 census data from Ruggles *et al.* (2017). We then locate fathers who report having children between the ages of 0 and 16 and examine how fathers' literacy, age, race, and Duncan Socioeconomic Index (SEI) varied with the reported age of the child. Graphs of father characteristics by age and census year appear in Figure 1. Although fathers of reported 10 year olds in the 1930 census have SES below trend, this phenomenon is not unique to the 1930 census. Fathers of reported 10 year olds were less literate, older, had lower Duncan SEI, and were more likely to be black relative to 9 and 11 year olds in both the 1910 and 1920 censuses. Likewise, while reported newborns, or 0 year olds, in the 1920 census have lower SES fathers, the same is true for newborns in the 1910 and 1930 censuses.

minus 1 will be correct for those born after April 1, so long as their reported age is correct. For individuals with correct ages born before April 1, inferred birth year will be off by one year.

Figure 1: Father characteristics by child's age in the 1910, 1920, and 1930 censuses



Notes: Underlying data are 100% samples from the 1910, 1920, and 1930 censuses from IPUMS. Samples include any individual 16 or younger whose father resides in the same household at the time of enumeration. We then compute the age-specific mean for each census year. Father's age at birth is the simple difference between the father's reported age and the child's reported age. The Duncan Socioeconomic Index is a measure of occupational standing based on the income and education level associated with each occupation in 1950.

Figure 1 does not rule out that the 1919 birth cohort came from a select pool of fathers. The deviations from trend for father's age and Duncan SEI appear larger for 0-year-olds in the 1920 census and 10-year-olds in the 1930 census. To measure the level of selection, we combine data from all three censuses and regress father's characteristics on separate cubic polynomials for each census year and a dummy variable for the 1919 birth cohort.[3] These estimates quantify the degree of selection before accounting for age heaping. We then adjust for age heaping by repeating the regression with additional dummy variables for age 0 and age 10. We also include two additional proxies for socioeconomic status as dependent variables: the occupational income score and the lasso industry, demographic, and occupation (LIDO) score (Saavedra & Twinam, 2018).[4]

The results from this exercise appear in Table 1. Panel A displays the estimates without accounting for age heaping, and Panel B accounts for heaping at ages 0 and 10. Baseline estimates suggest that the 1919 birth cohort had fathers who were older, less literate, more likely to be black, and more likely to be employed in lower SES occupations (measured by either the Duncan SEI, occupational income scores, or LIDO scoress). In Panel B, we can see that those who report an age of 0 and 10, even if they were not part of the 1919 birth cohort, all report these differences in father characteristics. The coefficient on a dummy variable for being in the 1919 cohort are all of the same sign, remain statistically different from zero, but are approximately cut in half in magnitude. Thus, age heaping accounts for approximately half of the parental differences documented in Brown & Thomas (2016).

[3]For computational ease, we take a random 50% sample.

[4]The occupational income score assigns each occupation the median earnings of that occupation in 1950. The LIDO score uses lasso regression to predict 1950 income with industry, occupation, sex, race, age, and geography.

Table 1: Father's characteristics by birth cohort

	(1)	(2)	(3)	(4)	(5)	(6)
	DV: Father's characteristics					
	Age	Literate	Black	SEI	OCCSCORE	LIDO
Panel A: Baseline estimates						
1919 cohort	0.283***	-0.0056***	0.0076***	-0.314***	-0.108***	-0.178***
	(0.007)	(0.0002)	(0.0003)	(0.018)	(0.011)	(0.009)
Panel B: Accounting for age heaping and differential mortality patterns						
1919 cohort	0.157***	-0.0037***	0.0042***	-0.185***	-0.036***	-0.084***
	(0.009)	(0.0003)	(0.0003)	(0.022)	(0.014)	(0.011)
Age 0	0.218***	-0.0061***	0.0071***	-0.322***	-0.134***	-0.190***
	(0.009)	(0.0003)	(0.0003)	(0.021)	(0.012)	(0.010)
Age 10	0.093***	-0.0002	0.002***	-0.053***	-0.048***	-0.058***
	(0.006)	(0.0002)	(0.0002)	(0.015)	(0.009)	(0.008)
Observations	47,478,444	47,478,444	47,478,444	47,478,444	47,478,444	41,911,313

Notes: * p<0.1; ** p<0.05; *** p<0.01. Robust standard errors are in parenthesis. The data come from 50% samples of the 1910, 1920, and 1930 censuses and includes all children ages 0 to 16 with a father. The regression includes a separate cubic polynomial for each census year. The 1919 birth cohort corresponds to those with a reported age of 0 in 1920 or a reported age of 10 in 1930.

2.2 Was there age heaping in later censuses?

The previous subsection illustrated that the patterns documented in BT are not necessarily unique to the 1919 cohort. Specifically, because of the timing of the pandemic and the timing of census enumeration, we should expect parents of the 1919 birth cohort to appear as lower SES because of age heaping and the SES-child mortality gradient. Indeed, drawing on full count census data from 1910, 1920, and 1930, we showed that parents who reported their child as being born in 1899, 1909, and 1929 were also of lower average SES despite the absence of a World War and pandemic.

This raises a natural question: if BT are overstating the difference in parental characteristics might Almond's results be overstated for similar reasons? First, let us consider the issue of the SES-child mortality gradient. This should not affect Almond's results because if children of lower SES parents are less likely to survive

early infancy then they will not show up in the later censuses that Almond analyzes. If we assume that these children would have worse outcomes then if anything this differential mortality likely works against Almond and so Almond's results are understated.[5]

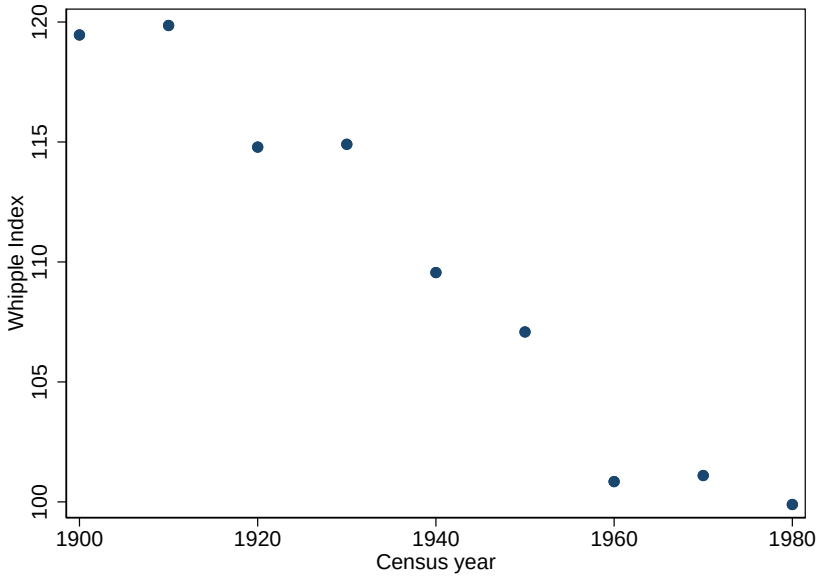
Age heaping, however, could be problematic because Almond draws on data from the 1960-1980 censuses and must infer year of birth from reported age. Thus, significant age heaping could threaten Almond's results. To gauge the level of age heaping in census years, we turn to the Whipple Index. The index was created by demographer George Chandler Whipple to quantify heaping at ages ending in 0 or 5 in census data, but does not measure heaping at other ages. Formally, the Whipple Index is the sum of individuals with an age ending in a 0 and 5 between the ages of 23 and 62 divided by the sum of all individuals between 23 and 65, multiplied by 500. Assuming the actual distribution of ages is continuous, we would expect the Whipple Index to be close to 100. In the sociology literature, a Whipple Index below 105 is typically considered accurate, while higher numbers indicate age heaping. We use data from the 1 percent samples of U.S. Censuses from 1900 to 1980. The results appear in Figure 2. The Whipple Index in 1920 and 1930 is approximately 115, suggesting there is significant age heaping. The Whipple Index for the 1960, 1970, and 1980 censuses are close to 100, suggesting no or minimal age heaping. Since Almond measures adult outcomes using the 1960-1980 censuses, it is unlikely that significant age heaping could be driving Almond's results.

2.3 Can measurement error explain the divergent findings?

BT replicate Almond's work and show that after controlling for parental characteristics, the 1919 birth cohort appears to be doing better than adjacent cohorts. Could

[5]One caveat is whether the outcomes of these individuals would have actually been affected by the pandemic in a measurable way. If we are measuring the effect on a discrete variable like high school graduation then for any individual that would not have graduated high school anyway, the pandemic may have had no effect.

Figure 2: Whipple index by census year



Notes: Data from the 1% census samples from IPUMS and are restricted to individuals between the ages of 23 and 62. The Whipple Index is the sum of individuals with an age ending in a 0 and 5 between divided by the sum of all individuals, multiplied by 500. A Whipple Index of 100 implies no age heaping.

these results be due to age heaping and selective infant mortality?

Let us consider a simple human capital model where son i's educational attainment (y_i^{son}) depends on in utero influenza exposure (flu_i), his father's educational attainment (y_i^{father}), and exogenous unobserved characteristics (ϵ_i), which are independent of the other explanatory variables, so that:

$$y_i^{\text{son}} = \beta \text{flu}_i + \gamma y_i^{\text{father}} + \epsilon_i. \tag{1}$$

Further assume that as in BT and Almond, flu_i is an indicator variable that affects all children from the 1919 birth cohort and otherwise equals zero. Estimation of equation 1 would yield consistent estimates of β if y_i^{son}, flu_i, and y_i^{father} were observable.

The advantage of the linked data approach we ultimately adopt is that we are able to observe all three of these variables. BT, however, do not observe y_i^{father}. Instead, BT observe average parental characteristics for a state-year cell. Assuming no age heaping or SES-child mortality gradient, we can express these parental characteristics as $\bar{y}_{s,t}^{\text{father}} = y_i^{\text{father}} + \nu_i$, where ν_i is the deviation from the mean for father i and is by assumption independent of $\bar{y}_{s,t}^{\text{father}}$.[6] Since flu_i is only a function of t, it follows that ν_i is also independent of flu_i. Then estimation of

$$y_i^{\text{son}} = \beta \text{flu}_i + \gamma \bar{y}_{s,t}^{\text{father}} + u_i \tag{2}$$

yields consistent estimates since $u_i = -\gamma\nu_i + \epsilon_i$, both of which are exogenous to the explanatory variables. Thus, in the absence of age-heaped data, the BT approach would successfully account for parental characteristics and return causal estimates of the effects of the 1918 influenza pandemic.

Now suppose that $\bar{y}_{s,t}^{\text{father}}$ is mismeasured because of age heaping. Specifically, suppose that relatively uneducated parents tend to round the ages of their children

[6]The two are by definition uncorrelated, but we strengthen the assumption to independence to simplify the discussion.

to heaped age (eg, ages 0, 10, and perhaps other even years), whereas relatively educated parents correctly report the ages of their children. Then average parental characteristics will be biased downwards for heaped ages (which, for example, include all actual 10-year-olds, plus 9- and 11-year-olds with relatively uneducated parents); whereas average parental characteristics will be biased upward for non-heaped ages (only 9 year olds with relatively educated parents report the age of 9).

Formally, let $\tilde{\bar{y}}^{\text{father}}_{s,t}$ be the mismeasured average parental characteristics, and $e_{s,t}$ be the error, so that $\tilde{\bar{y}}^{father}_{s,t} = \bar{y}^{\text{father}}_{s,t} + e_{s,t}$. For heaped ages, $e_{s,t} < 0$; for non-heaped ages, $e_{s,t} > 0$.

Consider estimation of

$$y^{\text{son}}_{i,s,t} = \beta \text{flu}_i + \gamma \tilde{\bar{y}}^{\text{father}}_{s,t} + \eta_i. \tag{3}$$

The error can be written as $\eta_i = \gamma(e_{s,t} - \nu_i) + \epsilon$, which is correlated with flu_i because those exposed to the pandemic in utero are also assigned lower SES parental characteristics due to of age heaping. This is precisely the bias we aim to overcome.

3 Measuring flu exposure with linked data

Thus far we have illustrated the problems associated with relying on aggregate census data to attach parental characteristics to the 1919 birth cohort. Having documented those issues, we now introduce an empirical approach that allows us to take the concerns of BT seriously while dramatically cutting down on measurement error. We start by constructing a novel dataset linking WWII enlistment records back to the 1920 and 1930 censuses. By linking individuals over time we are able to observe the same individual twice: first as a child with his parents at the time of enumeration and again as a young adult in the enlistment records. This means that instead of controlling average parental characteristics at the possibly mismeasured cohort level,

we can control directly for each child's own parental characteristics.

While drawing on linked data allows us to overcome issues associated with systematic mismeasurement of parental characteristics, the information observed in the census also allows us to move away from the cohort-based approach of Almond and BT and instead adopt a difference-in-differences framework. Specifically, by assuming that the individual's location at the time of enumeration was the same as their in utero environment we are able to exploit geographical variation in flu intensity as a source of identifying variation.

3.1 Constructing a linked dataset

To assess the extent to which in utero flu exposure affected long-run outcomes, we construct a dataset that links individuals across administrative records. We begin with the universe of WWII enlistment records.[7] These enlistment records span from 1938 to 1943 and cover over ten million men. Parman (2015b) estimates that about 65 percent of men born between 1913 and 1923 enlisted in WWII, and so the sample is more representative of the male population than one might initially think.[8] The exception to this is that registrants could be rejected for failing to meet the minimum education or physical standards. Failing to observe rejected applicants, however, would likely bias us against finding evidence that in utero flu exposure impaired human capital development.

The enlistment records include the following information: name, race, year of birth, state of birth, years of secondary and post-secondary education, height, and

[7]These records, which were digitized by the National Archives and Records Administration, are available at http://aad.archives.gov/aad/fielded-search.jsp?dt=893.

[8]While female records do exist, women were not required to register and so we might worry about whether the sample is representative. Because of this, we do not attempt to link female records. Failing to link women is in fact common practice in this literature, e.g. Aizer *et al.* (2016); Feigenbaum (2015); Long & Ferrie (2013). This is because women tend to change their name when they get married, and so without knowing their maiden name it is impossible to find their childhood administrative records.

weight.[9] To link these records to the 1920 and 1930 censuses, we draw primarily on the digitized name, year of birth, and state of birth variables. We focus our attention on individuals that would have been between the ages of 0 and 10 at the time of census enumeration in 1920 (i.e., the 1909-1919 birth cohorts) and we focus on those between the ages of 7 and 9 at the time of enumeration in 1930 (i.e., the 1920-1922 birth cohorts).

It is worth pointing out that our ultimate empirical specification includes birth cohort fixed effects, which will address differential changes in the linking introduced by the fact that our 1920-1922 cohorts come from a different census. We prefer to include these cohorts for our baseline analysis because these cohorts received no direct exposure to the pandemic, and so they are our cleanest set of control cohorts. Of course, including these cohorts comes at a cost, which is that the assumption that the place of enumeration is the same as their in utero environment is less likely to hold. Nevertheless, as a robustness check, we exclude these cohorts and we find nearly identical results.

Our linking procedure builds upon previous work (Long & Ferrie, 2013; Beach *et al.*, 2016).[10] In short, our algorithm can be summarized as follows: we take our set of enlistment records and cast a wide net to find a set of potential matches in the census. We then impose a number of requirements to ensure that the match is reasonable. If there is no match or more than one potential match, we classify the enlistment record as unable to be linked.

We begin by standardizing all given names (e.g., "Ed" and "Eddie" would be recoded as "Edward") in both the enlistment records and the census records. A potential census match is one where the birth state is the same, the race is the same, the birth year is similar (plus or minus three years), the first initial of the standardized

[9]Between March and June of 1943, the individual's AGCT score, a reasonable proxy for intelligence, was entered in place of weight (Ferrie *et al.*, 2012). Unfortunately we do not have enough observations in our sample to use this as an outcome.

[10]Alternative linking algorithms are available; however, recent work by Bailey *et al.* (2017) suggests that the method we employ is among the best in terms of reducing instances of false matches.

first names match, and the first initial of the last name matches. We then require that the raw first and last name strings are reasonably close. Specifically, we require that the Jaro-Winkler string distance for both the first names and the last names is greater than or equal to 0.8; a perfect match will have a string distance of 1. We do not require the names to match exactly for two reasons. First, during this time period census enumerators went door-to-door and recorded the information that was spoken to them; thus, the information recorded by the enumerator may represent a common spelling variant. Second, both the census records and the enlistment records were handwritten, and so relaxing the "exact name match" criteria relaxes any concerns about minor transcription errors.

A successful enlistment-to-census link is one where there is only one census record that survives this matching process.[11] Note however that the unique record may have an inconsistently reported birth year. The reason we allow year of birth to vary by up to three years is to accommodate the fact that the information comes from two different sources (the year of birth reported in the census likely comes from a parent, while the year of birth reported in the enlistment records comes from the individual). Thus, to the extent that numeracy differed, there may be disagreement between the two sets of records. However, because accurately identifying birth year is the point of this paper, we further restrict our final linked sample to include only those whose year of birth matches is consistently reported. This procedure allows us to link 853,141 enlistees to a record in the 1920 or 1930 censuses.

[11]Recent work by Bailey *et al.* (2017) has raised concerns about false matches in linking studies. Our modified linking algorithm (relative to some of our previous work, e.g. Beach *et al.* (2016)) responds to these concerns in two ways. First, we cast a very wide net to increase the chance that the true link is in our set of potential matches. Second, if a record has more than one potential match, rather than try to break that tie by choosing the record with the closest name or closest age (as was done in our previous work) we simply classify the record as unable to be linked. To assess the false positive rate with our algorithm we take the full census sample and create an modified version where we take the original record and modify both the names and ages to incorporate the types of spelling errors, transcription errors, and misreporting of birth years outlined above and in Goeken *et al.* (2017). We then try to link from the original census to the modified census using our linking algorithm. Our linking algorithm yields a successful match rate of 36.5 percent and a false positive rate of 1.8 percent. This highlights the conservativeness of our algorithm.

3.2 Measuring in utero flu exposure

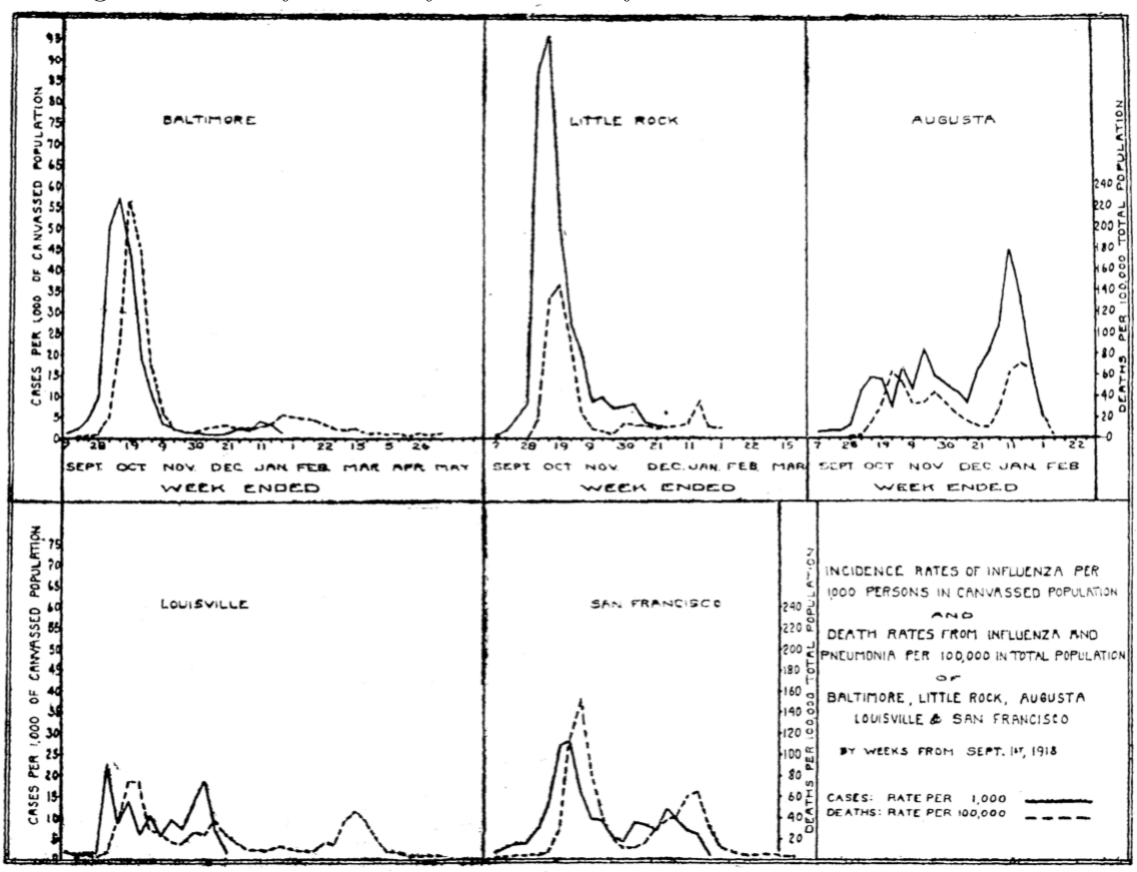
With our linked dataset in hand, we now turn our attention to measuring in utero flu exposure. Ideally we would have individual microdata identifying prenatal flu exposure (i.e., if and when exposure took place as well as a measure of severity). These data do not exist. Reliable data on maternal infection rates or morbidity are also unavailable as influenza morbidity data was not systematically collected prior to the pandemic. Beginning in 1900, however, annual influenza mortality data were collected and published in a systematic fashion for registration states and cities.[12] Thus, influenza mortality is a natural starting point for assessing in utero exposure.

The most important feature of the influenza mortality data series, which was published alongside other important causes of death on an annual basis by the US Census Bureau, is that the underlying data conform to a common reporting standard. This ensures that we are able to measure flu fatalities with a great degree of precision, but it does come with one drawback: the data are not comprehensive. Cities and states were only included in the published reports if the underlying data were deemed reliable. In 1900 there were 330 registration cities systematically collecting mortality data, but by 1920 the registration area included 662 cities spanning 41 states.

One natural question is whether influenza mortality trends track influenza morbidity trends reasonably well. While comprehensive reliable morbidity data do not exist, the best estimates of this relationship come from data collected by special canvases from the Public Health Service. During the fall of 1918, the Public Health Service went door-to-door collecting information about if and when a member of the

[12]Registration states and cities are those with laws requiring that mortality statistics be collected. In contrast to England, which standardized and mandated the reporting of deaths in 1846, the United States left this decision to state and local governments. Several large cities and states passed mandatory reporting laws by 1900, and in that year the Census Bureau worked with those registration areas to establish uniform reporting standards. The result of this was the adoption of a standardized death certificate and the international classification standard, as well as the distribution of "The Manual of International Classification of Causes of Death", which cross referenced terms appearing in causes of death from 1890 and 1900 reports with the new uniform classification standard.

Figure 3: Weekly morbidity and mortality for five cities in the fall of 1918



Notes: Figure reprinted from Frost (1920).

household had contracted the flu. They also collected information for members that eventually died. These canvases covered 12 cities. Our efforts to track down the underlying data were unsuccessful. However, the study was summarized by Frost (1920). Figure 3 comes from that report. That figure plots weekly morbidity and mortality rates for five cities during the fall of 1918. Here we can see that the fatality rate tracks the case rate reasonably well on a weekly basis but with a bit of a lag. Given that the data we ultimately draw on are reported at the annual level, this lag is unlikely to be much of a concern for us.

While influenza mortality data capture total influenza deaths in 1918, one concern is that influenza mortality will capture more than just the severity of the pandemic.

For instance, early public health scholars often noted that clean water interventions lowered mortality from waterborne causes as well as causes that are not typically thought of as waterborne (e.g., influenza, tuberculosis, pneumonia, kidney failure, and heart failure).[13] Relatedly, Clay *et al.* (2015), document that mortality rates in 1918 were higher in places with more coal pollution and places with high pre-existing typhoid fever rates, a reasonable proxy for water quality. These relationships are attributable to the fact that air pollution and typhoid fever compromise an individual's immune system, making them more susceptible to influenza. In light of this, observing high influenza mortality rates in 1918 could mean that a city was hit relatively hard by the pandemic, or that a city had relatively worse water and air quality, or both. Since air pollution and water quality have also been shown to impair human capital development, we may worry that influenza mortality rates do not cleanly identify the effect of in utero flu exposure.[14]

Our solution to this problem is as follows. First, we generate a counterfactual estimate of influenza mortality in 1918. To do so, we transcribe all city-level mortality statistics spanning 1900-1930 from the annual Mortality Statistics reports. We then run a series of city-level regressions where we restrict the sample to the 1900-1917 period and regress ln(influenza deaths) on a city-specific linear time trend.[15] Taking the exponential of the predicted values from this regression yields a prediction of influenza fatalities in the absence of the pandemic for post-1917 years.[16] This allows us to construct a measure of excess influenza deaths by simply subtracting predicted influenza deaths in 1918 from actual influenza deaths in 1918.

This measure of excess deaths gives us the unanticipated increase in influenza

[13]This phenomenon is often referred to as the Mills-Reincke Phenomenon. See Ferrie & Troesken (2008) for more discussion as well as an empirical test of this theory in Chicago.

[14]See Sanders (2012) and Isen *et al.* (2017) on early-life exposure to air pollution and Beach *et al.* (2016) on early-life exposure to typhoid fever.

[15]We only run these regressions for cities that appear in every report. This leaves us with 287 cities.

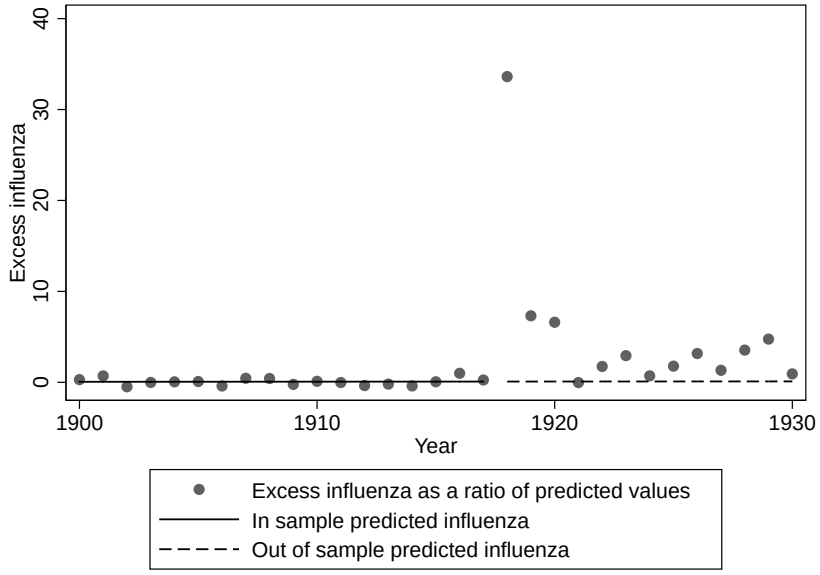
[16]The use of the natural logarithm ensures that the predicted number of influenza deaths is always greater than zero.

mortality due to the pandemic. Of course, we need to normalize this measure because excess deaths will be related to city size, and the goal is to have a measure of in utero exposure to influenza, not in utero exposure to large cities. Our options are to divide by population or to divide by predicted influenza deaths. Population and predicted influenza deaths are positively correlated: all else equal, a larger city should have more predicted influenza deaths. However, dividing by population ignores the fact that cities of similar sizes may have different underlying disease and pollution environments. Because of this, normalizing by population will likely overstate flu intensity for cities with particularly bad health environments. For this reason, and because accurate population data are only available in census years, we choose to use predicted influenza deaths as our denominator. Mechanically this measure is simply the ratio of the number of excess influenza deaths occurring in 1918 to the number of expected influenza deaths in 1918, where that expectation captures underlying trends in population growth and intrinsic differences in disease and pollution environments. Intuitively, this is a measure of flu severity that isolates the pandemic channel from other contributing factors.

After dealing with annexations, our final sample consists of 287 cities. For these cities, Figure 4 plots the average excess influenza ratio by year. The solid line is a line of best fit through the years of 1900-1917 (in the sample period).[17] Excess influenza remains close to zero until 1918, during which approximately 35 influenza deaths occurred for every expected influenza death. Influenza deaths are higher during than the 1920s than would have been predicted using the pre-pandemic data, however, even 13 years after 1917, excess influenza deaths are not far above zero. It appears influenza reached its new steady state in 1921.

[17]Alternatively, we could have used data from the 1900-1917 and 1920-1930 years, omitting the years during which the pandemic occurred. However, if the pandemic had lingering effects on influenza rates or city population counts, then data from the 1920-1930 period may be endogenous.

Figure 4: Excess influenza mortality ratio by year



Notes: Excess influenza ratio is calculated by taking actual influenza deaths minus predicted influenza deaths and then dividing by the predicted influenza deaths. City-specific trends in annual ln(influenza mortality) are estimated over the 1900-1917 period.

4 Empirical Approach

Combining our linked dataset with our measure of excess influenza yields a final sample of 218,662 linked records. The reduction in sample size largely stems from the fact that many of our linked records simply did not live in a city that we could calculate exposure to the pandemic.[18] Summary statistics are reported in Table 2. The average years of schooling for our sample is 11.4 with about 54 percent of our sample completing high school. In terms of family characteristics, the literacy rates among mothers and fathers were quite high at 94 and 92 percent, respectively. About

[18]Of our initial 853,141 links we are not able to use 597,486 of them because we do not have flu data for their birth city. We throw out an additional 26,556 links because the child is not observed with both parents at the time of enumeration and we throw out an additional 4,296 links because one or more of the relevant parental controls is missing. Next we discard an additional 6,141 links because the parental age at time of birth is implausible (e.g., mother or father's age at the time of birth is over the age of 50 or under the age of 18). We discard these observations because we are concerned that the indicated mother or father is not actually the birth mother or birth father.

36 percent of families owned their homes and only 4.5 percent of mothers were in the labor force at the time of enumeration.

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Excess flu in 1918	36.093	25.025	3.64	179.172	218,662
WWII variables					
Years of schooling	11.478	2.305	8	17	218,662
High school graduate	0.561	0.496	0	1	218,662
Height (inches)	68.284	2.679	50	80	168,317
Weight (pounds)	152.143	22.225	100	300	168,317
Birth year	1916.879	3.535	1909	1922	218,662
Census variables					
Father's age at child's birth	32.066	6.542	18	50	218,662
Father's OCCSCORE	21.712	14.83	0	80	218,662
Father reads and writes	0.949	0.22	0	1	218,662
Father is foreign	0.435	0.496	0	1	218,662
Mother's age at child's birth	28.334	5.899	18	50	218,662
Mother is in labor force	0.044	0.205	0	1	218,662
Mother reads and writes	0.928	0.259	0	1	218,662
Mother is foreign	0.383	0.486	0	1	218,662
Family owns home	0.361	0.48	0	1	218,662
Number of siblings	2.364	1.941	0	9	218,662

Notes: Excess flu in 1918 is calculated by taking actual influenza deaths minus predicted influenza deaths and then dividing by the predicted influenza deaths. See text for more details. WWII variables come from the enlistment records. Sample is restricted to linked males born between 1909 and 1922 in a city for which we have flu data. Census variables are observed in either 1920 for the 1909-1919 birth cohorts or 1930 for the 1920-1922 cohorts. OCCSCORE is the median income for a particular occupation as measured in 1950. This variable is measured in hundreds of 1950 dollars.

We adopt a difference-in-differences strategy to identify the impact of in utero flu exposure on adult outcomes. Specifically we estimate variations of two regressions. For the first, we regress:

$$y_{ibc} = \alpha_0 + \beta_b + \gamma_c + \sum_{b=1916}^{1922} \delta_b \mathbf{1}\left[\text{yob} = b\right] \times \text{High}_c + \Gamma X_i' + \epsilon_{ibc} \tag{4}$$

where y_{ibc} is outcome y of individual i from birth year b in birth city c. The parameters β_b and γ_c are birth year and birth city fixed effects, respectively. The variable High_c is equal to one if city c had above median excess influenza in 1918. The parameters δ_b are treatment effects that vary by birth cohort. If the fetal origins hypothesis is true, we would expect $\delta_{1916} = \delta_{1917} = \delta_{1918} = \delta_{1920} = \delta_{1921} = \delta_{1922} = 0$ and $\delta_{1919} < 0$ when the outcome is educational attainment. Of course, exposure to the pandemic in early life may matter as well (Currie & Almond (2011)). In this case we might expect δ_{1916} through $\delta_{1919} < 0$ with effects being more pronounced for δ_{1919}. Even in this setting, δ_{1920} through δ_{1922} should be equal to zero. The vector X_i includes additional controls such as race and parental characteristics. For some specifications, we only estimate a treatment effect for the 1919 birth cohort and include birth-city linear time trends. Standard errors are clustered at the birth city level.

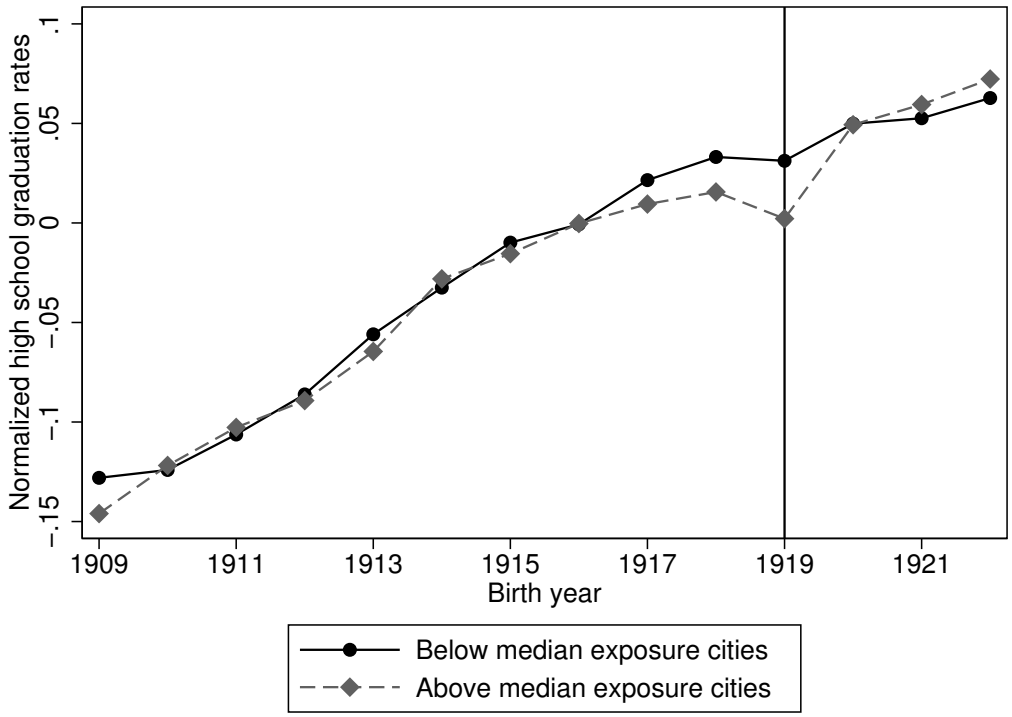
The previous specification is a simple difference-in-differences equation that assigns above median excess influenza cities as the treatment group and below median excess influenza cities as the control group. Additionally, we estimate a difference-in-differences model that uses continuous variation in treatment status:

$$y_{ibc} = \alpha_0 + \beta_b + \gamma_c + \sum_{b=1916}^{1922} \delta_b \mathbf{1}\left[\text{yob} = b\right] \times \frac{\text{flu}_{1918,c}}{E(\text{flu}_{1918,c}|\text{yob} = b)} + \Gamma X_i' + \epsilon_{ibc}. \quad (5)$$

The variable $\text{flu}_{1918,c}$ is excess influenza in 1918 for city c. This variable is divided by $E(\text{flu}_{1918,c}|\text{yob} = b)$ so that δ_b is the average effect of the pandemic for birth cohort b. All other parameters are defined as above.

Figure 5 depicts this identification strategy (and one of our main results) with raw data. Specifically, we plot the average high school graduate rate by birth cohort for cities that will ultimately have above and below median exposure to the pandemic. To ease interpretation, we normalize high school graduation rates at the city-level. The figure reveals that for the 1909 to 1916 birth cohorts, graduation rates in cities with above and below median exposure tended to follow a nearly identical trend.

Figure 5: Normalized graduation rates by birth cohort in above and below median pandemic exposure cities



Notes: Sample consists of 218,662 enlistees linked to either the 1920 or 1930 censuses. Graduation rates normalized by removing city-specific mean before collapsing to the cohort level. The median pandemic city had an excess flu ratio of 28.2.

Beginning with the 1917 cohort the average graduation rates start to fall relative to previous trends for both groups and the effects appear to be much more severe for those born in cities with above average exposure to the pandemic. While the 1919 cohort experiences the largest deviation from trend, by 1920 graduation rates appear to return to normal.

One key takeaway from Figure 5 is that a difference-in-difference strategy comparing outcomes in high versus low exposure cities is likely to satisfy the parallel trends assumption. However, it is worth noting that the slight deviations in trend for the 1917-1919 cohorts in low exposure cities may be due to early life exposure to the pandemic. In the context of our regressions, we treat those deviations as the counter-

factual trend that would have been observed in high exposure cities if the pandemic had never occurred. Thus, to the extent that the stagnation in low exposure (i.e., control) cities is due to the pandemic, our estimates can be thought of as a lower bound. However, if one believes that stagnation is due to other factors (e.g., selection into military service) then by treating low exposure cities as a control, our empirical approach will simply difference out that and any other selection mechanisms that are common across the entire 1919 birth cohort.

5 Results

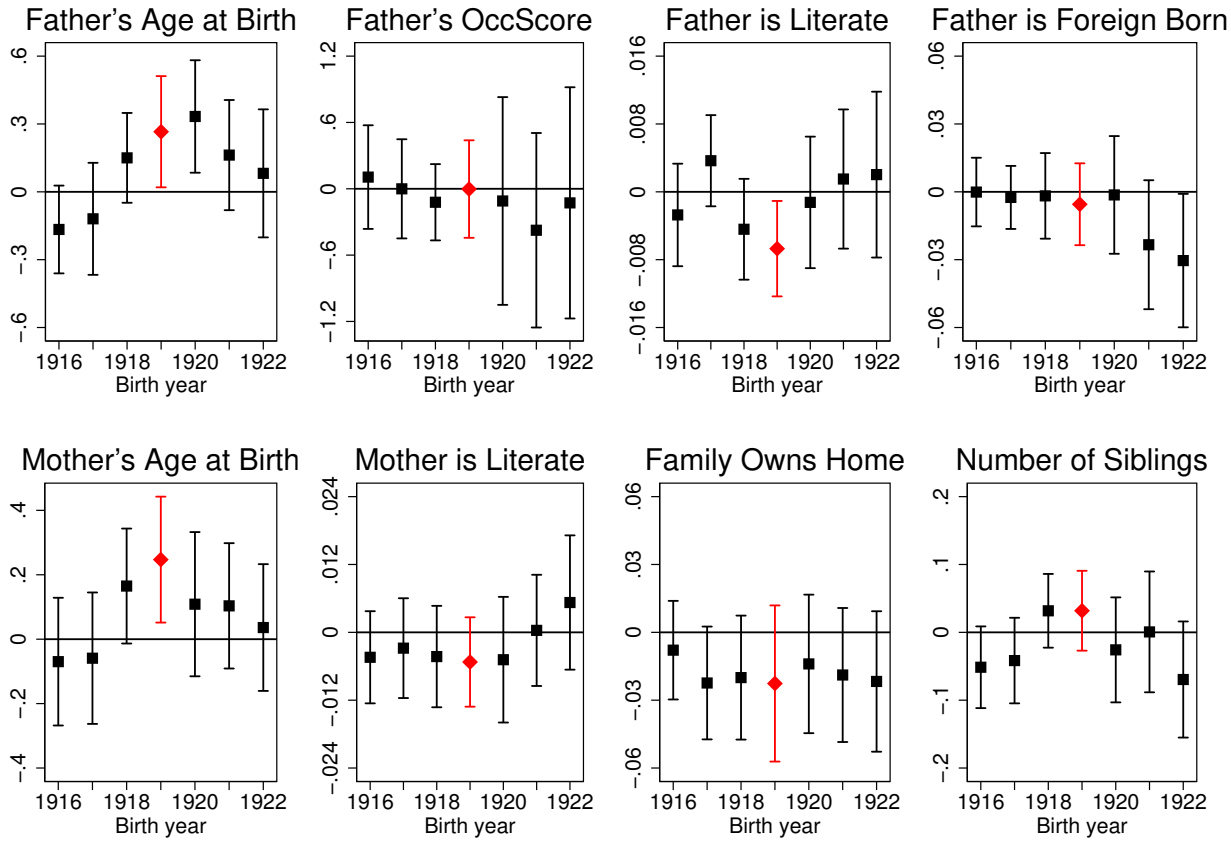
Before proceeding to our main results, Figure 6 provides additional support for our identification strategy. Each panel considers a different parental characteristic: mother and father's age at the time of birth, father's occupational income score, indicators for whether the mother and father are literate, an indicator for whether the father is a foreigner, an indicator of whether the family owns their home, and number of siblings. All of these characteristics are available from the census where we observe the child with his parents. Each panel plots four coefficients that represent the interaction between being born in a high exposure city and being born in either 1916, 1917, 1918, 1919, 1920, 1921, or 1922. These coefficients were obtained from running a variation of equation 4 that regresses each parental characteristic on a series of birth cohort fixed effects, city fixed effects, and the seven above interactions.

The purpose of this exercise is to take the concerns of Brown & Thomas (2016) seriously and show that our source of identifying variation is not systematically correlated with other determinants of human capital. The results of Figure 6 show little evidence that parents of children born in high exposure cities were systematically different than parents of children born in low exposure cities in 1919 or other adjacent years. While the 1919 coefficient is statistically significant when the dependent variables are mother's or father's age at birth, these deviations appears to be consistent

with a broader cycle. The other statistically significant deviation relates to father's literacy, however, relative to adjacent cohorts the deviation does not appear as a clear outlier. Further, the 0.8 percentage point decline in literacy is not meaningful when evaluated relative to the base literacy rate of nearly 95 percent.

For comparison purposes, we repeat this exercise taking the individual's high school completion, total years of schooling, and height (all measured at the time of enlistment) as our outcomes of interest. These results are presented in Figure 7. The top panels use the same estimating equation as Figure 6. The bottom panels add parental controls to the regressions. In both panels we see that high school graduation and years of schooling follow the same trends observed in Figure 5. That is, from 1917 to 1919 we see deviations from trend and the deviation is most pronounced for the 1919 cohort. By 1920, however, the impact of the pandemic appears to be close to zero. The results in Figure 7 indicate that individuals from high exposure cities that were in utero during the pandemic were roughly 2.5 percentage points less likely to complete high school and had about 0.1 fewer years of total schooling at the time of enlistment.

Figure 6: Flu intensity and parental characteristics



Notes: Sample consists of 218,662 enlistees linked back to either 1920 or 1930 censuses. Each coefficient corresponds to an indicator that equals 1 for individuals born in the specified year (1916 through 1922) and born in a high exposure city. High exposure cities had an excess flu ratio greater than 28.2. These coefficients come from a regression that also includes year fixed effects, city fixed effects, and race fixed effects. Implicitly, these coefficients are evaluated against parental characteristics from children born in high exposure cities between the years 1909 and 1915.

It is worth noting that the 1916-1918 cohorts can be thought of as treated to the extent that they also received exposure to influenza during early childhood, just not while in utero. Thus, the results of Figure 7 indicate that in utero exposure to the pandemic was much more detrimental to human capital development than exposure in the postnatal period.[19]

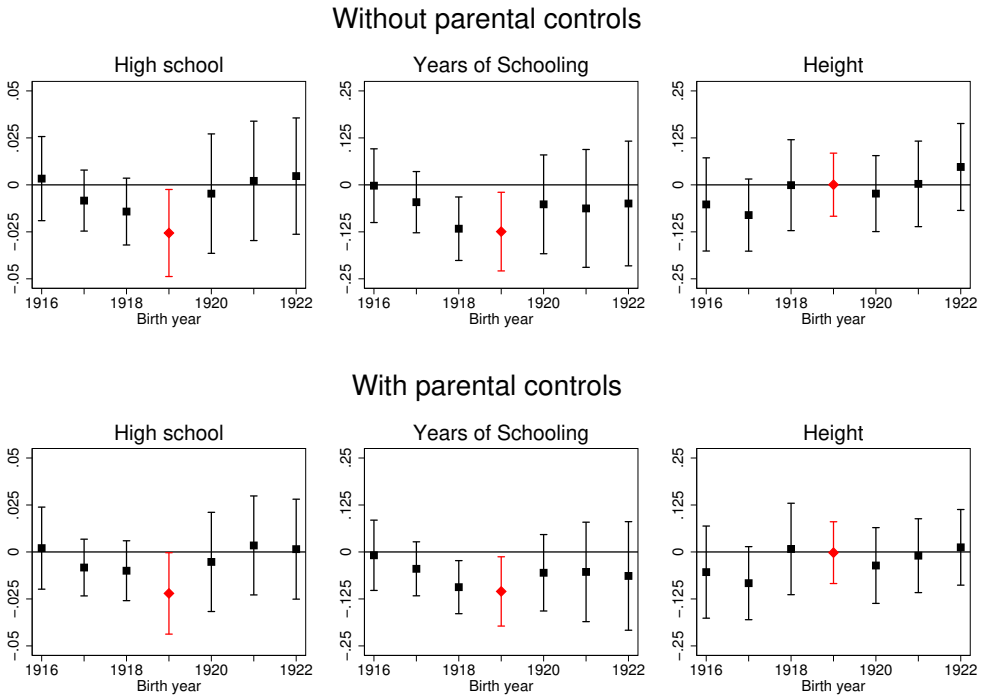
Interestingly, we do not find any systematic relationship between in utero (or other early life exposure) to the pandemic and height at the time of enlistment. Economic historians and demographers often use height as a proxy for exposure to disease and deprivation.[20] In their analysis of heights in enlistment records both Mazumder *et al.* (2010) and Parman (2015b) find the average heights of those born in 1919 to be lower than those in adjacent birth cohorts. There are two potential explanations for why our results differ from these two studies. First, our sample differs substantially from these studies in that we restrict our analysis to individuals that lived in cites. Second, those studies use the same identification strategy as Almond (2006) whereas we include birth cohort fixed effects and instead exploit flu intensity as a source of identifying variation. If both high and low flu exposure have similar effects on adult height, then our results would imply that in utero flu exposure did not affect height.

Having established that the effects of the pandemic were most pronounced for those that were in utero at the time of the pandemic, we now assess the robustness of this effect. Panel A of Table 3 presents results for high school graduation and Panel B presents results for total years of schooling. The first row in each panel models in utero flu exposure as a simple indicator for being born in a high exposure city in 1919. The second row models in utero exposure in a continuous fashion by interacting each city's excess influenza rate with an indicator for being born in 1919. To aid

[19]Since our identification strategy treats 1916-1919 cohorts in low exposure cities as a counterfactual, it is possible that post-natal exposure to flu impaired human capital development but did not have a differential effect for those in high exposure environments.

[20]Most relevant for our study is Parman (2015a). Parman analyzes the link between early childhood disease exposure and height using the same enlistment data as us and finds a negative relationship between disease exposure and height, even after including household fixed effects.

Figure 7: Flu intensity and long-run outcomes



Notes: Sample consists of 218,662 individuals when the outcomes are high school completion or years of schooling and 168,385 observations when the outcome is height. Each coefficient corresponds to an indicator that equals 1 for individuals born in the specified year (1916 through 1922) and born in a high exposure city. High exposure cities had an excess flu ratio greater than 28.2. These coefficients come from a regression that also includes year fixed effects, city fixed effects, and race fixed effects. Implicitly, these coefficients are evaluated against parental characteristics from children born in high exposure cities between the years 1909 and 1915. Parental controls include: father's age at time of birth, mother's age at time of birth, father's OCCSCORE, indicators for whether the father and mother can read and write, indicators for whether the father and mother are citizens, an indicator for whether the mother was in the labor force at the time of enumeration, an indicator for whether the family owns their own home, and number of siblings.

interpretation, we standardize this variable by dividing by the mean excess influenza rate for the 1919 birth cohort. This standardization means that the coefficient can be interpreted as the pandemic's average effect. Column 1 presents our baseline results. In column 2 we add controls for our parental characteristics (described above), while in column 3 we add city-specific linear trends. In column 4 we focus on brothers and add household fixed effects.[21]

The main results are as follows: we find robust evidence that the pandemic decreased high school graduation rates on the order of 2 percentage points. This result survives the inclusion of household fixed effects when we model in utero exposure continuously but the p-value falls just outside the 10-percent level of significance in our more conservative specification where we treat those with below median exposure to the pandemic as the control group. For schooling, we find consistent and negative effects, but we lose statistical significance in the household fixed effects specification. In the Appendix we show that omitting the 1930 links (the 1920 to 1922 birth cohorts) has little effect on our estimates.

Table 4 presents results for three biological outcomes: height (Panel A), weight (Panel B), and BMI (Panel C). The format of this table follows Table 3 in that the top row of each panel models in utero exposure discretely while the bottom rows model in utero exposure continuously. We find no evidence that in utero influenza exposure affected weights or BMI for any specifications. There is no effect on heights when influenza is measured discretely. When influenza is measured continuously, the pandemic appeared to increase heights but this result is not robust to the inclusion of household fixed effects. Even for the specification that is statistically significant, it is not economically significant. The estimate suggests that the pandemic may have

[21]Since the 1909 to 1919 cohorts are observed in the 1920 census and the 1920 to 1922 cohorts are observed in the 1930 census, we won't observe any of the 1920 to 1922 birth cohorts with their older brothers born before 1920. To remedy this we go back to census manuscripts and identify any older brothers that are still in the household in 1930. If any of those older brothers happen to be successfully linked between the enlistment records and 1920, then we are able to generate a household identifier that is consistent regardless of whether the brother was born in 1909-1919 (and thus enumerated in 1920) or born between 1920 and 1922 (and thus enumerated in 1930).

Table 3: In utero exposure and schooling

	(1)	(2)	(3)	(4)
Panel A: DV is High School Graduate				
Treatment is above median exposure	-0.023***	-0.020***	-0.020***	-0.041
	(0.007)	(0.007)	(0.006)	(0.030)
Continuous treatment (std.)	-0.016***	-0.014***	-0.013***	-0.043*
	(0.004)	(0.005)	(0.004)	(0.022)
Observations	218,662	218,662	218,662	23,395
Panel B: DV is Years of Schooling				
Treatment is above median exposure	-0.087**	-0.070**	-0.050*	-0.051
	(0.034)	(0.030)	(0.028)	(0.141)
Continuous treatment (std.)	-0.063***	-0.055***	-0.037**	-0.069
	(0.021)	(0.018)	(0.018)	(0.095)
Observations	218,662	218,662	218,662	23,395
Parental controls	N	Y	Y	Y
City-year trends	N	N	Y	Y
Household fixed effects	N	N	N	Y

Notes: * p<0.1; ** p<0.05; *** p<0.01. Standard errors (clustered at the city level) in parentheses. Parental controls include: father's age at time of birth, mother's age at time of birth, father's OCCSCORE, indicators for whether the father and mother can read and write, indicators for whether the father and mother are foreign born, an indicator for whether the mother was in the labor force at the time of enumeration, an indicator for whether the family owns their home, and number of siblings. The continuous treatment specifications interact each city's excess influenza rate with an indicator for being born in 1919. We standardize the excess influenza rate by dividing by the mean excess influenza rate, such that the coefficient can be interpreted as the pandemic's average impact.

Table 4: In utero exposure and adult health

	(1)	(2)	(3)	(4)
Panel A: DV is Height				
Treatment is above median exposure	0.013	0.015	0.027	-0.097
	(0.041)	(0.039)	(0.040)	(0.220)
Continuous treatment (std.)	0.044**	0.045**	0.047**	0.022
	(0.022)	(0.019)	(0.020)	(0.123)
Observations	168,317	168,317	168,317	18,051
Panel B: DV is Weight				
Treatment is above median exposure	0.157	0.227	0.472	-0.008
	(0.376)	(0.363)	(0.317)	(1.863)
Continuous treatment (std.)	0.159	0.174	0.213	-0.803
	(0.195)	(0.189)	(0.182)	(1.312)
Observations	168,317	168,317	168,317	18,051
Panel C: DV is BMI				
Treatment is above median exposure	0.012	0.021	0.049	0.079
	(0.053)	(0.050)	(0.045)	(0.240)
Continuous treatment (std.)	-0.004	-0.003	0.001	-0.140
	(0.028)	(0.027)	(0.025)	(0.159)
Observations	168,317	168,317	168,317	18,051
Parental controls	N	Y	Y	Y
City-year trends	N	N	Y	Y
Household fixed effects	N	N	N	Y

Notes: * p<0.1; ** p<0.05; *** p<0.01. Standard errors (clustered at the city level) in parentheses. Parental controls include: father's age at time of birth, mother's age at time of birth, father's OCCSCORE, indicators for whether the father and mother can read and write, indicators for whether the father and mother are citizens, an indicator for whether the mother was in the labor force at the time of enumeration, an indicator for whether the family owns their home, and number of siblings.

increased heights by less than one twentieth of one inch.[22]

6 Conclusion

Almond (2006) provides some of the first evidence in economics in favor of the fetal origins hypothesis by analyzing the 1918 influenza pandemic as a natural experiment. A key assumption of Almond's work is that the 1919 birth cohort would have had similar outcomes to adjacent birth cohorts if the pandemic had never occurred. This assumption is not obvious since other events coincided with the pandemic, most notably the height of WWI. Brown & Thomas (2016) argue that, because servicemen were positively selected from the pool of potential fathers, the 1919 birth cohort had systematically lower SES parents. After accounting for these differences in parental characteristics, they find that the long-run effects of the pandemic disappear.

We contribute to this debate in two ways. First, we show that early censuses suffered from age heaping at ages 0 and 10 and that this heaping is most consistent with (1) LSES parents rounding the ages of their children and (2) the children of LSES parents being less likely to survive past infancy. Accounting for this age heaping, we find that Brown & Thomas (2016) likely overstate the differences in parental characteristics by a factor of 2. However, age heaping alone cannot account for all of the parental differences in the 1919 birth cohort. For this reason, the BT critique should be taken seriously.

Second, we estimate the long-run effects of the 1918 influenza pandemic using linked data and city-level variation in influenza exposure. Using linked data allows us to accurately observe parental characteristics with minimal measurement error. Thus, we can directly control for parental differences in the 1919 birth cohort. Second, using city-level data allows us to more accurately measure local influenza environment and

[22]To put this in perspective, Parman (2015a) relates enlistee height to state and city-level disease environments and finds that a standard deviation increase in infant mortality rates decreases height by 0.3 inches, nearly 6 times greater than the effect we find.

provides us with the necessary variation to include birth cohort fixed effects. Using the linked data, we find that the 1918 influenza pandemic reduced educational attainment and that the estimates are of a similar order of magnitude to those in Almond (2006).

Our results should be interpreted with two important caveats in mind. First, our sample corresponds to males that enlisted in WWII and resided in a city at the time of enumeration in either 1920 or 1930. Thus, our sample is not nationally representative and may not generalize to females or those born in the countryside. For instance, female labor force opportunities were more compressed during this time period, and so our estimates may be less relevant for this population. Further, because influenza is a contagious disease it is likely that those residing in rural areas were less affected by the pandemic. The second caveat is that we do not observe post-WWII economic outcomes. It is unclear if our results would generalize to economic outcomes in the 1960s when the 1919 birth cohort would have reached peak earnings.

With these caveats in mind, it is worth pointing out that the WWII enlistment records have two notable advantages. The first is that they report year of birth instead of age and we find no evidence of age heaping in the enlistment data. The second advantage is that the enlistment records are the only publicly available dataset that include any measure of biological outcomes (height and weight) that can also be linked back to historical census records. While we do not find consistent evidence that influenza exposure affected height or weights, biological outcomes are an important part of Almond's analysis and so it is important that we are able to consider them in our analysis as well.

References

A'Hearn, Brian, Baten, Jörg, & Crayen, Dorothee. 2009. Quantifying quantitative literacy: Age heaping and the history of human capital. *The Journal of Economic History*, **69**(3), 783–808.

Aizer, Anna, Eli, Shari, Ferrie, Joseph, & Lleras-Muney, Adriana. 2016. The long-run impact of

cash transfers to poor families. *American Economic Review*, **106**(4), 935–71.

Almond, Douglas. 2006. Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population. *Journal of Political Economy*, **114**(4), 672–712.

Almond, Douglas, & Currie, Janet. 2011. Killing me softly: The fetal origins hypothesis. *Journal of economic perspectives*, **25**(3), 153–72.

Almond, Douglas, & Mazumder, Bhashkar. 2005. The 1918 influenza pandemic and subsequent health outcomes: an analysis of SIPP data. *American Economic Review*, **95**(2), 258–262.

Bailey, Martha, Cole, Connor, Henderson, Morgan, & Massey, Catherine. 2017. *How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth*. Tech. rept. National Bureau of Economic Research.

Barker, David J. 1995. Fetal origins of coronary heart disease. *BMJ: British Medical Journal*, **311**(6998), 171.

Barreca, Alan I. 2010. The long-term economic impact of in utero and postnatal exposure to malaria. *Journal of Human resources*, **45**(4), 865–892.

Beach, Brian, Ferrie, Joseph, Saavedra, Martin, & Troesken, Werner. 2016. Typhoid fever, water quality, and human capital formation. *Journal of Economic History*, **76**(01), 41–75.

Bleakley, Hoyt. 2007. Disease and development: evidence from hookworm eradication in the American South. *The quarterly journal of economics*, **122**(1), 73–117.

Bleakley, Hoyt. 2018. *The Hookworm Again*. Tech. rept. Working paper.

Boustan, Leah, & Margo, Robert A. 2014. *Racial Differences in Health in Long-Run Perspective: A Brief Introduction*. Tech. rept. National Bureau of Economic Research.

Brown, Ryan, & Thomas, Duncan. 2016. On the long term effects of the 1918 US influenza pandemic. *Unpublished Manuscript*.

Clay, Karen, Lewis, Joshua, & Severnini, Edson. 2015. *Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic*. Tech. rept. National Bureau of Economic Research.

Collins, William J, & Thomasson, Melissa A. 2004. The declining contribution of socioeconomic disparities to the racial gap in infant mortality rates, 1920-1970. *Southern Economic Journal*, 746–776.

Currie, Janet, & Almond, Douglas. 2011. Human capital development before age five. *Pages 1315–1486 of: Handbook of labor economics*, vol. 4. Elsevier.

Feigenbaum, James J. 2015. Intergenerational mobility during the great depression. *Unpublished working paper*.

Ferrie, Joseph P, & Troesken, Werner. 2008. Water and Chicagos mortality transition, 1850–1925.

Explorations in Economic History, **45**(1), 1–16.

Ferrie, Joseph P, Rolf, Karen, & Troesken, Werner. 2012. Cognitive disparities, lead plumbing, and water chemistry: Prior exposure to water-borne lead and intelligence test scores among World War Two US Army enlistees. *Economics & Human Biology*, **10**(1), 98–111.

Frost, Wade Hampton. 1920. Statistics of influenza morbidity: with special reference to certain factors in case incidence and case fatality. *Public Health Reports (1896-1970)*, 584–597.

Goeken, Ronald, Lee, Yu Na, Lynch, Tom, & Magnuson, Diana. 2017. Evaluating the Accuracy of Linked U. S. Census Data: A Household Linking Approach. *Unpublished Manuscript*.

Isen, Adam, Rossin-Slater, Maya, & Walker, W Reed. 2017. Every breath you takeevery dollar youll make: The long-term consequences of the clean air act of 1970. *Journal of Political Economy*, **125**(3), 848–902.

Long, Jason, & Ferrie, Joseph. 2013. Intergenerational occupational mobility in Great Britain and the United States since 1850. *American Economic Review*, **103**(4), 1109–37.

Mazumder, Bhashkar, Almond, Douglas, Park, Kyung, Crimmins, Eileen M, & Finch, Caleb E. 2010. Lingering prenatal effects of the 1918 influenza pandemic on cardiovascular disease. *Journal of developmental origins of health and disease*, **1**(1), 26–34.

Parman, John. 2015a. Childhood health and human capital: New evidence from genetic brothers in arms. *The Journal of Economic History*, **75**(1), 30–64.

Parman, John. 2015b. Childhood health and sibling outcomes: Nurture Reinforcing nature during the 1918 influenza pandemic. *Explorations in Economic History*, **58**, 22–43.

Parman, John. 2017. Health, gender and mobility: Intergenerational correlations in longevity over time. *Working paper*.

Roodman, David. 2018. *COMMENT: THE IMPACTS OF HOOKWORM ERADICATION IN THE AMERICAN SOUTH*. Tech. rept. Working paper.

Ruggles, Steven, Genadek, Katie, Goeken, Ronald, Grover, Josiah, & Sobek, Matthew. 2017. *Integrated Public Use Microdata Series: Version 7.0 [dataset]*. Minneapolis: University of Minnesota. http://doi.org/10.18128/D010.V6.0.

Saavedra, Martin. 2017. Early-life disease exposure and occupational status: The impact of yellow fever during the 19th century. *Explorations in Economic History*, **64**, 62–81.

Saavedra, Martin Hugo, & Twinam, Tate. 2018. A Machine Learning Approach to Improving Occupational Income Scores. *SSRN Working Paper*.

Sanders, Nicholas J. 2012. What doesnt kill you makes you weaker prenatal pollution exposure and educational outcomes. *Journal of Human Resources*, **47**(3), 826–850.

7 Appendix

Table A.1: In utero exposure and schooling – 1920 links only

	(1)	(2)	(3)	(4)
Panel A: DV is High School Graduate				
Treatment is above median exposure	-0.023***	-0.021**	-0.019***	-0.045
	(0.009)	(0.008)	(0.007)	(0.032)
Continuous treatment (std.)	-0.018***	-0.015***	-0.016***	-0.042**
	(0.004)	(0.005)	(0.004)	(0.020)
Observations	163,263	163,263	163,263	16,219
Panel B: DV is Years of Schooling				
Treatment is above median exposure	-0.091**	-0.077**	-0.031	-0.043
	(0.040)	(0.036)	(0.030)	(0.168)
Continuous treatment (std.)	-0.078***	-0.066***	-0.034	-0.066
	(0.021)	(0.019)	(0.027)	(0.088)
Observations	163,263	163,263	163,263	16,219
Parental controls	N	Y	Y	Y
City-year trends	N	N	Y	Y
Household fixed effects	N	N	N	Y

Notes: * p<0.1; ** p<0.05; *** p<0.01. Standard errors (clustered at the city level) in parentheses. Parental controls include: Father's age at time of birth, Mother's age at time of birth, Father's OCCSCORE, indicators for whether the father and mother can read and write, indicators for whether the father and mother are foreign born, an indicator for whether the mother was in the labor force at the time of enumeration, an indicator for whether the family owns their own home, and number of siblings. The continuous treatment specifications interact each city's excess influenza rate with an indicator for being born in 1919. We standardize the excess influenza rate by dividing by the mean excess influenza rate, such that the coefficient can be interpreted as the pandemic's average impact.