

Walking with Feet Tied to the Past: Childhood Exposure to the 1918 Influenza Pandemic and Later-Life Male Mortality*

Hamid Noghanibehambari[†]

Jason Fletcher[‡]

Abstract

Many previous studies have used cross-cohort comparisons and a focus on in utero exposure to document the lingering effects of the 1918 influenza pandemic on survivors' outcomes later in life. A subset has examined impacts on old-age health and mortality, finding mixed results. This paper contributes to the literature by extending the analysis to evaluate early-life and childhood exposure to the pandemic on old-age longevity along with a set of candidate mechanisms. We employ Social Security Administration death records linked to the full-count 1940 census and employ event-study and difference-in-difference regressions to compare longevity of cohorts with different age-at-exposure who were born in cities with high versus low exposure based on 1918 influenza-related mortality. Our treatment-on-treated calculations suggest that cohorts with age-at-exposure of 0-5 and 6-11 experience 7.2 and 10.9 months reductions in old age longevity; importantly, we note that these cohorts are often used as "controls" in the broader literature. While our main analysis focuses on males due to data constraints, supplementary results suggest similar effects for females. An analysis of census data suggests reductions in health, education and socioeconomic scores as potential mechanisms. The findings of this paper have important policy implications for interventions and identifying the populations at higher risks of adverse long-term outcomes during major pandemics.

Keywords: Mortality, Longevity, Pandemic, Childhood Exposures

* The authors claim that they have no conflict of interest to report. The authors would like to acknowledge financial support from NIA grants (R01AG060109, R01AG076830) as well as useful comments from participants at the Frontiers in Mortality, Risk and Insurance seminar series at the University of Wisconsin Madison and the Data-Intensive Research Conference.

[†] College of Business, Austin Peay State University, Marion St, Clarksville, TN 37040, USA

Email: noghanih@apsu.edu, Phone: +1-806-620-1812

[‡] La Follette School of Public Affairs, University of Wisconsin-Madison, 1225 Observatory Drive, Madison, WI 53706-1211, USA

Email: jason.fletcher@wisc.edu

JEL Codes: I18, J13, N31, N32

1. Introduction

The years 1918-1919 witnessed the largest pandemic in recorded history, the infamous Spanish flu. The estimates suggest that about one third of the world population contracted the disease and death tolls are estimated to be about 50 million, with approximately 650K deaths in the US (Gagnon et al., 2013; Patterson & Pyle, 1991). While the effects on the US economy was relatively modest⁴ and local areas that were hit harder recovered faster and even experienced larger wage growth during the 1920s (Brainerd & Siegler, 2003), the effects on individuals lasted for decades.

A strand of research examines the long-term effects on individuals' later-life health and socioeconomic measures with a specific focus on early-life exposure to the pandemic. For instance, Almond (2006) documents reductions in education and income and increases in disability for in-utero exposed cohorts. Cook et al. (2019) show that these negative effects spill over to the next generations. Fletcher (2018b) shows that women exposed during the first trimester marry lower educated men. Children and grandchildren of cohorts exposed to the pandemic during early-life reveal reductions in education and income. A narrow strand of this literature evaluates the effects on old-age health and mortality outcomes. Almond & Mazumder (2005) find deteriorations in subjective and physical health of exposed cohorts compared with the adjacent cohorts. Although there are evidence of the effects on education and socioeconomic measures, the impacts on mortality and longevity is mixed and inconclusive.⁵ Fletcher (2018b) employs data from the National Longitudinal Mortality Study and fails to find significant impacts on mortality for in-utero exposure. However, Fletcher (2018a) uses similar data and find small increases in mortality for those born in 1918 versus those born in 1917 and 1919. Mazumder et al. (2010) employ data from National Health Interview Surveys and document increases in cardiovascular diseases for 1918-born cohorts compared with adjacent cohorts. Myrskylä et al. (2013) document mixed evidence on mortality by cause of death. They show that in-utero exposed cohorts reveal reductions in cancer mortality but increases in non-cancer deaths.

⁴ Barro et al. (2020) estimate a roughly 1.5 percent drop in GDP and about 2 percent reduction in consumption for the US economy during the 1918-1919 pandemic.

⁵ Studies of other countries also find similar pattern of effects for a wide array of later-life education, socioeconomic status, and health outcomes (Beach, Clay, et al., 2022; Helgertz & Bengtsson, 2019; Karlsson et al., 2014; Lin & Liu, 2014; Nelson, 2010).

This paper employs newly released data and explores the effects of early-life and childhood exposure to the 1918 pandemic on later-life longevity. Our data allows us to infer individuals' city-of-childhood and implement event-study difference-in-difference regressions to examine the effects across different ages-at-exposure in high/low affected cities. We find significant reductions in longevity of those with age-at-exposure of less than 12 and smaller and insignificant effects for those exposed between 12-17. We find slightly larger effects among those with lower maternal education and from lower socioeconomic status families. Our analysis from census data reveals a very similar pattern of negative effects on education and measures of socioeconomic status.

Our paper adds to the existing literature in three ways. First, previous studies rely on cohort differences in mortality using adjacent cohorts as the control group (Cook et al., 2019; Fletcher, 2018a, 2018b, 2018c; Mazumder et al., 2010; Myrskylä et al., 2013). Therefore, they fail to account for the dosage of exposure across different geographic regions. We take advantage of a city-level measure of influenza mortality and employ a difference-in-difference method which enables us to absorb unobserved cohort heterogeneity and place heterogeneity in longevity. Second, the focus of these studies is almost exclusively on in-utero exposures relying on the relevance of insults during prenatal development for life-cycle outcomes (Almond & Currie, 2011; Barker, 1994, 1995, 1997). However, our focus spans several years of childhood and argues that childhood experiences of disease environment is also an important contributor for long-run outcomes, including old-age mortality (Almond et al., 2018; Hayward & Gorman, 2004). Third, as noted by Beach, Brown, et al. (2022), it is important to control for family characteristics as the 1918 pandemic affected cohorts come from lower socioeconomic status families.⁶ Our data is linked to historical censuses and allows us to observe a rich set of family characteristics, which we use as additional controls in our regressions. This aspect of our analysis is also unique as other studies of long-run effects of 1918 pandemic on mortality fail to control for family background characteristics.

The results of this paper have important policy implications. They point to the long-term costs of pandemics for individuals' outcomes. The magnitude of the effects could be used in cost-benefit analyses to justify policy interventions during and following pandemics. The results also reveal subpopulations at risk which help policymakers in identifying those who could most benefit

⁶ They show that the results of the seminal paper by Almond (2006) become small and insignificant when they control for family background characteristics.

from pandemic-related policy initiatives. Finally, our analysis shows that the 1918 pandemic had long lasting health and economic effects on cohorts who are often used as “controls” in the literature.

The rest of the paper is organized as follows. Section 2 discusses the empirical method. Section 3 introduces data sources. Section 4 reviews the results. Section 5 discusses the economic meaning of the estimated effects. Finally, section 6 concludes the paper.

2. Econometric Method

Our primary analysis leverages an event-study-type difference-in-difference method that compares the longevity of individuals at various ages of exposure in cities with higher versus lower 1918-1919 influenza mortality rates. Specifically, we implement the following regressions:

$$y_{icrb} = \alpha + \sum_{i=-9}^{i=-2} \xi_i I(1918 - b = i) \times Exp_{cr} + \sum_{j=0}^{j=19} \zeta_j I(1918 - b = j) \times Exp_{cr} \quad (1)$$

$$+ \gamma X_i + \eta Z_{crb} + \lambda_{rb} + \theta_c + \varepsilon_{icrb}$$

The outcome y is age-at-death (in months) for individual i in city c in census-region r born in year b . The function $I(.)$ takes a value of one if the inside argument is true and zero otherwise. The parameter Exp indicates whether the city of birth/childhood is at the top three-quartiles of influenza mortality rate.⁷ Therefore, the coefficients ξ represent the effects of exposure for individuals born after 1919 in cities with higher pandemic-related mortality. Since these cohorts are not directly affected by the influenza outbreak, their respective coefficients can be interpreted as placebo tests. The coefficients ζ represent the effects of cohorts who were exposed in high-mortality cities at various ages (0 to 19). In X , we include individual race, maternal education dummies, and paternal socioeconomic status dummies. In Z , we include city-year covariates extracted from full-count decennial censuses and interpolated for inter-decennial years, including the share of married people, share of literate people, and average socioeconomic score. Birth-

⁷ The use of a dummy variable to indicate high versus low exposure eases the interpretation of the results. However, one may be concerned with differential effects of exposure across different quartiles of influenza rate and also the extent of nonlinearities in the effects. In Appendix C, we extensively discuss this by replicating the main results across different age of exposure and different quartiles of influenza rate. We find that for several age-at-exposures, the effects of top-quartile cities become considerably smaller than those second and third quartiles, implying the role of mortality and selection. Therefore, it is not surprising to observe insignificant and smaller effects when we look at exposure using linear variables.

cohort by region-of-birth fixed effects (λ) control for secular trends in longevity with heterogeneous evolutionary paths across different regions in the country. City fixed effects (θ) absorb time-invariant unobserved heterogeneity of city influences in health and longevity. Finally, ε is a disturbance term. We cluster standard errors at the city and cohort level to account for both serial and spatial correlations in error terms.

We complement the event studies with a series of difference-in-difference regressions in which we group several event-studies and focus on three age groups, as follows:

$$y_{icrb} = \alpha + \zeta_1 Age_b^{0-5} Exp_{cr} + \zeta_2 Age_b^{6-11} Exp_{cr} + \zeta_3 Age_b^{12-17} Exp_{cr} + \gamma X_i + \eta Z_{crb} + \lambda_{rb} + \theta_c + \varepsilon_{icrb} \quad (2)$$

Each *Age* variable is a dummy that equals one if the individual belongs to age-at-exposure range shown in superscripts. All other parameters are as in equation 1.

3. Data and Sample Selection

The primary source of data is Death Master Files (DMF) of Social Security Administration records linked to the full-count 1940 census extracted from the CenSoc Project (Goldstein et al., 2021). The data reports deaths to male individuals that occurred between years 1975-2005.⁸ We use cross-census linking rules provided by the Census Linking Project to locate individuals in censuses 1900-1930 (Abramitzky et al., 2020). The cross-census data allows us to observe city of birth-childhood of individuals.⁹ This is an important identifier to measure the intensity of treatment and enables us to implement cross-city cross-cohort comparisons in a difference-in-difference setting.

We extract city-level influenza mortality data from Beach, Clay, et al. (2022). We use the mortality counts due to outbreaks of influenza for about 489 cities over the years 1918-1919. We calculate quartiles of influenza mortality rate for these cities as the dose-response measure in our regressions. Figure 1 shows the geographic distribution of influenza rate across counties (top panel) and location of cities within these counties in the final sample (bottom panel).

⁸ In Appendix B, we employ an alternative data that also covers females but includes a narrower death window. We argue that the effects are similar among females compared with males. Moreover, the effects are smaller for narrower death windows, suggesting that the inclusion of earlier deaths increase the magnitude of coefficients.

⁹ We use the city in the first census of observation as the city of birth/childhood. For instance, if an individual is born in 1901, the first census of observation is 1910. We assign the city-of-residence in 1910 as the city of birth/childhood.

We limit the sample to pre-1926 cohorts to have several cohorts born post-pandemic. We also remove cohorts aged above 20 in 1918. Therefore, our sample covers birth cohorts of 1898-1926 who died between 1975-2005. Summary statistics of the final sample are reported in the top-panel of Table 1. The final sample covers 1,471,443 observations. Average age-at-death is 76.4 years (916.4 months). Figure 2 shows the density distribution of age-at-death for those above and below-median of the 1918 influenza rate. About 19.5, 15.2, and 9.6 percent of individuals live in high exposure cities and are aged 0-5, 6-11, and 12-17 years old, respectively. Although the sample underrepresents blacks and overrepresents whites (about 3 percent blacks and 96.7 percent whites), both subpopulations have similar characteristics as the original cohorts in the original 1940 census (Breen & Osborne, 2022). In section 3.1, we further delve into this issue and argue that these differential representations due to data linking process is not likely to confound our estimates.

The employ alternative data sources to examine mechanism channels. We use the information reported in the 1940 census to examine the effects on educational achievements and other socioeconomic measures. Further, we also use census data for the years 1980-2000 combined with the American Community Survey data for the year 2005 extracted from (Ruggles et al., 2020). The reason for this selection of years is to have a sample consistent with death years of the DMF data (1975-2005). We also use information on World War II enlistment records that are linked to the 1940 census and our final sample. This data is extracted from (Goldstein et al., 2023). The bottom panels of Table 1 report summary statistics of these data sources. We will further discuss these data and their relevance in section 4.5.

4. Results

4.1. Concerns over Endogeneity

The Spanish flu resulted in influenza mortality rates that varied substantially by individuals' age and sociodemographic backgrounds (Clay et al., 2019; Gagnon et al., 2013; Noymer & Garenne, 2000). The differential mortality may confound the estimates of equations 1 and 2 if the exposure results in excess mortality of people with specific characteristics that are related to their longevity. For instance, higher exposure may result in overrepresentation of people from higher socioeconomic status families due to their higher survival during the pandemic. Since these people have, on average, higher longevity for other usually unobserved reasons, our results may underestimate the true (negative) effects of exposure in longevity. We empirically test this

using balancing tests in which we regress a series of pre-determined observable characteristics on exposure measures of equation 2, conditional on city and region-year fixed effects. The results are reported in Table 2. We do not find significant associations between exposure and being white, black, father schooling, mother schooling, and father socioeconomic score. The point estimates are economically small relative to the mean of outcomes reported in the last row. For instance, exposure at ages 0-5 and 6-11 in high-mortality cities is associated with 0.04- and 0.16-units lower father socioeconomic score, off a mean of 32.7, respectively. However, we observe several significant coefficients for missing value of paternal and maternal characteristics (columns 4, 6, and 8). We should note that the reference group in equation 2 are those born after the pandemic (cohorts of 1919-1925) as well as those aged 18-19 at the time of the pandemic. Since parental characteristics are observed in 1940 census and the 1919-1925 cohorts are less likely to have moved from their original family by 1940, it is not unexpected to observe positive coefficients for indicators of missing values. The overall picture of this table reduces the concern that differential mortality and survival into adulthood changes the sociodemographic composition of the final sample. Since we do not find associations between exposure measures and observables, assume that there are also no association with unobservable characteristics (Altonji et al., 2005; Fletcher et al., 2021).

4.2. Endogenous Merging

Individuals in the final sample come from death records that are linkable to the 1940 census as well as cross-census linking between 1940 and 1900-1930. This sample selection may cause concerns about generalizability of the results. Further, the selection may also confound the estimates if those who are merged have some characteristics that are correlated with their exposure and longevity. In this section, we empirically explore whether there is any association between merging and our exposure measure. The idea behind this test is that if more/less exposed individuals are more/less likely to be in the final sample, then the estimates might pick up on characteristics of linked individuals. We start our analysis by using the full-count 1940 census and limit the sample to cohorts of 1897-1925 who reside in cities that are covered in our final sample. We then merge this with the final sample and generate a variable indicating successful merging. We then regress this successful merging indicator on the right-hand side variables of equation 2. The results are reported in Table 3. We find quite small and statistically insignificant associations. For instance, the those in high exposure cities with age-at-exposure of 0-5 and 6-11 reveal 3.8 and

10.4 basis-points change, off a mean of 0.088. Since these results do not support an association with our main independent variables, there is little concern about endogenous merging.¹⁰

4.3. Main Results

The main results of equation 1 are depicted in the top-panel of Figure 3. We remove the coefficient of 1919 cohorts (age-at-exposure of -1) so that all coefficients are in comparison with longevity of those born in 1919.¹¹ The effects on individuals born after the pandemic (age-at-exposure of -2 to -9) reveal no difference compared with those born in 1919. Moreover, the results do not reveal a significant impact for those with age-at-exposure of 0-1. This is in contrast with several studies that show excess mortality for in-utero exposure to the 1918 influenza (Fletcher, 2018a; Myrskylä et al., 2013) but in line with Fletcher (2018b) that reports null results. However, we find significant impacts for age-at-exposure of 2-11. For age-at-exposures of 12-19, we observe reductions in longevity compared with the reference point, but the differences are relatively small and insignificant. To account for potential heterogeneity in the treatment effects that are ignored by ordinary least square estimations, we implement the method developed by Sun & Abraham (2021) and replicate the main event-study results in the bottom-panel of Figure 3. We observe a very similar pattern as those produced by OLS.

One plausible explanation of small effects for ages 0-1 is the very large mortality rates for this age group compared with other childhood ages. Thomas (1951) documents roughly 112 under-1 deaths per 1,000 infants in the year 1918. This is in sharp contrast with 15.7 deaths per 1,000 for age group 1-4 and 4.1 deaths per 1,000 for age group 5-14. This difference is even more evident when we focus on deaths related to influenza and pneumonia. Rogers (1920) tabulates deaths by cause and age for the year 1918 for Philadelphia and for Kansas and Indiana. He documents a large difference between death related from influenza and pneumonia among ages 0-1 and later childhood ages in all three samples. For instance, in Indiana, influenza-related death rate (per

¹⁰ We should also note that both DMF-census and cross-census linking rules are based on name commonalities rather than specific individuals' characteristics that are correlated with their health. Therefore, it is unlikely that linking rules confound the exposure-longevity relationships.

¹¹ Based on the results of Figure 3, we observe no discernible and statistically significant difference between longevity of cohorts born in 1918, 1919, and several years afterward. One notable outlier is for cohorts of 1920 which depicts positive effect versus those of 1919 cohort. Moreover, we should note that the 1919 cohort is also treated as pandemic continued toward the early months of 1919 and many individuals were born in this year had been in utero during 1918 or the early months of 1919. Therefore, this cohort may also be considered in utero exposed. If there are indeed some influences for in utero exposure for this cohort, our results produce conservative estimates of the true effects as a portion of the control group has also received the treatment.

100,000) for males aged 0 and 1 was 3,483 and 2,683 while for males aged 5-9 was merely 409.¹² This extremely high mortality difference may impose a selection on survivors. As noted by Bozzoli et al. (2009), at high levels of infant mortality rates, selection effects may dominate scarring effects and leave the fittest infants to survive into adulthood. Therefore, it is not surprising to observe no longevity reductions among these individuals. On the other hand, at lower rates of mortality (i.e., higher survival rates into adulthood), especially for those aged 2-11, the scarring effects dominate and we observe the negative effects of exposure on old-age outcomes. Another fact to note is that the 1918 influenza mortality rate revealed an unexpected “W” shape, where the effects were large for infants, old people, and surprisingly for young adults, specifically ages 15-44 (Gagnon et al., 2013; Noymer & Garenne, 2000). This pattern is consistent with the smaller effects on older age groups as higher mortality could force selection.

We group the event-study point estimates in the manner discussed in equation 2 and report the results in Table 4. Column 1 reports the results of regressions with only city and birth-year fixed effects. We add more covariates across consecutive columns. We find relatively robust coefficients. Based on the full specification of column 4, individuals with age-at-exposure of 0-5 and 6-11 in high exposure cities reveal 1.8 and 2.7 months reductions in longevity, respectively. We find an insignificant effect of -0.89 for age-at-exposure of 12-17.

Several studies show that during the 1918 influenza pandemic, there were varying rates of infection and mortality based on socioeconomic indicators (Beach, Clay, et al., 2022; Mamelund et al., 2021). Therefore, one expects to observe similar heterogeneity for long-run outcomes. In Table 5, we explore the heterogeneity by race, maternal education, and paternal socioeconomic score. While the effects appear to be larger among whites of age-at-exposure 0-5, we observe larger impacts among blacks for age-at-exposure of 6-11 and 12-17 (compared to the mean of outcome). For age-at-exposure of 0-5 and 6-11, we observe comparable effects and change with respect to the mean for individuals with low/high maternal education and low/high paternal socioeconomic index. However, we find substantially larger effects for age-at-exposure of 12-17 who belong to lower parental education and socioeconomic group.¹³

¹² In Appendix B, we implement a state-level analysis and show relatively large increases in infant mortality rates in 1918 compared with adjacent years.

¹³ Another potential source of heterogeneity comes from the historical fact that World War I veterans were more likely to contract the disease, carry the virus upon their return, and expose their family members. In Appendix E, we examine

4.4. Robustness Checks

In Table 6, we explore the robustness of results to alternative specifications. In column 1, we add birth-state by birth-year fixed effects to compare cross-cohort outcomes across cities within a state. This model will absorb all time-varying policy and economic shocks at the birth-state level that could influence the penetration of pandemic and appear in longevity equations. The effects only slightly drop in magnitude compared with those of column 4 of Table 4.

In column 2, we allow for city fixed effects to have heterogeneous impacts by race and parental characteristic dummies. We observe increases in the magnitudes. In this model, the effect of age-at-exposure of 12-17 also becomes significant. In column 3, we control for unobserved city characteristics that evolve linearly over cohorts by adding a city-specific trend in birth year. The coefficients become slightly smaller than the main model of Table 4.

To account for seasonality in birth and seasonal factors in mortality that may confound longevity equations, we include birth-month and death-month fixed effects. The results, reported in column 4, are very similar to the main results. In column 5, we add a battery of additional city controls, including population, labor force participation rate, average occupational income score, share of different race groups, age groups, and occupation groups. The coefficients drop by about 16-27 percent.

In columns 6-8, we examine the sensitivity of the functional form. In column 6, we replace the outcome with log of age-at-death. The coefficients of age-at-exposure 0-5 and 6-11 suggest 0.19 and 0.29 percent change, almost identical to the implied percentage change with respect to the mean of the outcome in Table 4. In columns 7-8, we replace the outcome with binary variables indicating longevity beyond 75 and 80 years, respectively. We observe larger effects for longevity at older ages, suggesting that the negative effects are delayed until much older ages. This is true for both point estimates and their implied changes with respect to the outcomes' mean.

Since our death window is limited, we do not observe the full history of longevity of each cohort. Goldstein et al. (2021) use the Human Mortality Database to estimate gender-specific cohort-specific longevity. Based on these estimates, they provide a weighting variable for a subset of observations that take into account differential longevity of cohorts based on the observation

the source of heterogeneity and find mixed evidence with insignificant results. However, the coefficients suggest slightly larger impacts for age of exposures of 0-5 and 12-17.

window. In column 9, we weight regressions by their provided “weight” variable. We observe quite similar effects to those reported in Table 4. Finally, in column 10, we show that standard errors are quite robust when we cluster at the city level.

4.5. Mechanisms

There are several channels through which the exposure to influenza during childhood may affect later life health outcomes. One possible channel is the interruption of schooling due to business closures, city lockdowns, and general Non-Pharmaceutical Interventions (NPI). Although studies show the potential benefits of NPIs during the pandemics on mortality rates, they are also associated with adverse effects on mental health outcomes and interruptions in schooling and other social interactions all of which may have long-lasting consequences (Garrett, 2007; Lai et al., 2020). For instance, city lockdowns are documented to interrupt children’s socioemotional and cognitive developments. Therefore, these NPIs may lead to lower human capital accumulation and reductions in schooling which in turn affects later-life health (Adams-Prassl et al., 2022; Goodman et al., 2011; Spinelli et al., 2020). Another pathway is that the pandemic may result in parental death. Studies document that the experience of parental death during early life and childhood have long-lasting legacies, specifically for mental health outcomes and suicide (Rostila & Saarela, 2011; Smith et al., 2014).¹⁴ Further, childhood exposure to pandemic might be accompanied by significant psychosocial stress. Prolonged exposure to this stress during the developmental period may have long-lasting impacts on health outcomes (Brew et al., 2022). Furthermore, infection during the pandemic among children may trigger inflammatory responses that accumulates and contributes to chronic inflammation, which is documented to be associated with later life cardiovascular disease, diabetes, and other autoimmune diseases (Burgner et al., 2015). However, the changes in immune system programming can also be beneficial specifically if infected during childhood. Exposure to certain pathogens may contribute to changes in immune response system and make individuals less susceptible to future infectious diseases, bringing benefits for exposed children in the long-run (Jain, 2020).

Although the scarcity of data limits our ability to test many of these pathways, we can employ the information available in censuses to provide suggestive channels. For instance,

¹⁴ In Appendix D, we use parental absence in the 1920 full count census relative to the 1910 full count census as a proxy for parental death. We observe significant and large increases in parental absence in 1920 versus 1910 in cities with higher exposure versus those with lower exposure.

changes in education and socioeconomic measures may operate as mediatory channels of impact as studies document the associations between these outcomes and longevity (Chetty et al., 2016; Fletcher, 2015; Lleras-Muney, 2005; Miller & Bairoliya, 2021). Since the final sample is linked to the 1940 census, we can examine the effects of exposure on these outcomes, as observed in early adulthood. To avoid those with uncompleted education, we limit this sample to those born before 1921. We report these results in columns 1-4 of Table 7. We do not find significant effects on schooling (column 1). Point estimates are quite small in magnitude representing less than 0.4 percent change with respect to the mean. This is in contrast with findings of Almond (2006) and Fletcher (2018b) but in line with the more recent study of Beach, Clay, et al. (2022) that find small and insignificant effects on educational outcomes after adjusting for proxies of parental characteristics. However, we find reductions in the occupational income score and socioeconomic index (columns 2-3). Consistent with the reduced-form results in Figure 3 and Table 4, we observe larger effects for the age-at-exposure of 6-11. Relative to the mean of the outcomes, age-at-exposure of 0-5, 6-11, and 12-17 for high-exposure cities are associated with 1.9, 3.6, and 1.5 percent drop in socioeconomic score. Furthermore, we observe reductions in the probability of having a white-collar occupation. However, the stronger effects appear to be concentrated for age-at-exposures of 6-11 and 12-17.

To complement this analysis, we also use decennial censuses 1960-1980, implement similar sample selection and empirical method, and replicate columns 1-4 in columns 5-8 of Table 7. We find reductions in all three measures, including schooling, with larger effects on age-at-exposure of 6-11. This pattern is consistent with both columns 1-4 of this table and the reduced form results of longevity. A noticeable difference in the results of columns 1-4 versus 5-8 is the relatively higher schooling and socioeconomic score of observations in 1960-1980 censuses versus those observed in 1940 censuses. The larger socioeconomic score could be attributable to the influence of time and accrued wealth and experience in the labor market. Higher schooling could also be attributable to acquired education and training during post-1940 years. This could be those who re-attend to complete high school or those who select to attend college, specifically community colleges in which age profile tends to be much higher than four-year colleges (Kane

& Rouse, 1999; Phillippe et al., 2005).¹⁵ If the observed difference is driven by later educational attainments, the negative effects of column 4 indicates that exposed cohorts were less successful in post-1940 educational achievements. Another explanation is that less educated individuals with higher exposure revealed lower mortality prior to 1960 compared with better educated individuals and column 4 reflect this mortality differential. This is less plausible given the literature that suggests benefits of education for mortality and longevity (Halpern-Manners et al., 2020; Lleras-Muney, 2005).

To complement the analysis of this section and further explore the mechanisms, we turn to World War II enlistment data. There are two advantages of using this data source. First, the data is linked to the 1940 full census and therefore we can link our final sample to the enlistment data at the individual level. Second, the data reports several important mediatory outcomes related to anthropometric measures and cognitive ability. Specifically, the data reports the Army General Classification Test (AGCT) scores. The AGCT was the comprehensive test developed by Army technicians in order to evaluate intelligence and other cognitive abilities of individuals in order to assign them to specific tasks and occupations within the Army. This measure is useful because intelligence and cognitive abilities are linked to early life exposures and also, they are documented to influence later life health and longevity (Deeg et al., 1990; Kremen et al., 2019; Lee et al., 2022). In addition, the data also reports height and weight of individuals. Earlier research points to the association between several anthropometric outcomes on schooling, socioemotional, and health outcomes (Deaton, 2007; Kim & Fletcher, 2021). Therefore, these measures provide us with an opportunity to explore several alternative pathways.

We use enlistment-based outcomes in regressions similar to the main results and report them in Table 8. In column 1, we explored the effects on the probability that an individual in the final sample is observed in the enlistment data. We find that higher exposure to the 1918 influenza results in a lower probability of being in the enlistment data with the largest effects for ages 0-12. Although the coefficients are relatively large with respect to the mean of the outcome, they are noisy and limit further interpretation. In column 2, we explored the results on the standardized height-for-age variable. We observe significant reductions for ages 0-5 and 6-11. Both coefficients

¹⁵ This is one possible explanation given the sharp increase in the number of high school expansions prior to 1940 and new college openings that occurred between 1940 and 1970 (Currie & Moretti, 2003; Goldin, 1998; Goldin & Katz, 1999).

suggest a reduction of about 6.5% of the standard deviation of height-for-age. In column 3, we examined the effects on AGCT score. We observe reductions in the score for ages 0-5 and 6-11 but small and insignificant increase for age-at-exposure of 12-17. The coefficient of age at exposure of 6-11 is significant and suggests a reduction of roughly 4.6 units, off a mean of 81.1.

Next, we use information on height and weight to construct the Body Mass Index (BMI). We then generate a dummy variable that indicates the BMI being in the normal range (i.e., above 18.5 and below 24.5). The results are reported in column 4. We observe significant reductions for age-at-exposure of 6-11 suggesting a decrease of 1.2 percentage points, equivalent to a change of 1.7% with respect to the mean of the outcome. We also observed larger effects for age at exposure of 12-17 but the coefficient is a statistically insignificant.

5. Discussion

The results suggest reductions in longevity of children of ages 0-11 who were exposed to the 1918 influenza in high mortality cities. We should note that the documented estimates are intent-to-treat effects across the whole population of children. Although obtaining the number of contracted cases is difficult, CDC (2018) estimates that about 25 percent of the US population were inflicted with the disease. Assuming the same rate for all age groups, we can use this number as the first stage to inflate the effects of reduced-form of Table 4. A back-of-the-envelope calculation to convert intent-to-treat effects into treatment-on-treated effects suggests that those with age-at-exposure of 0-5 and 6-11 in high exposure cities experienced 7.2 and 10.9 months reductions in longevity, respectively. The original population of these cohorts in high exposure cities add up to about 2.76 and 2.24 million individuals. Based on the reduced-form treatment-on-treated estimates, we calculate roughly 3.7 million life-years lost to these cohorts due to childhood exposure to the pandemic.

One way to put the magnitudes into perspective is to employ Value of Statistical Life (VSL) estimates. The average age-at-death of exposed individuals (age-at-exposure of 0-11) in the final sample is roughly 77 years. The average life expectancy of males survived to 77 is estimated at roughly 10 years (Arias, 2014). Further, studies suggest an average VSL of about \$10.9 million for US-born individuals (Colmer, 2020; Kniesner & Viscusi, 2019; Viscusi, 2018). Using the additional expected life-years of 10 and an arbitrary discount rate of 3 percent, we reach a Value

of Statistical Life year of about \$1.27 million for exposed cohorts in the final sample.¹⁶ Using the life-years lost calculations, we estimate a total loss of \$4.69 trillion due to longevity reductions of cohorts exposed to the pandemic at ages 0-11.

We can also understand the magnitude of the effects by comparing these effects with other determinants of mortality. For instance, Halpern-Manners et al. (2020) documents that each additional year of schooling is associated with about 4 months additional age at death. Therefore, the calculated treatment-on-treated effects are equivalent to a drop of about 1.8-2.7 years of schooling. Aizer et al. (2016) show that children of mothers who received a 3-year cash transfer under the Mother's Pension program during early decades of the 20th century live, on average, 11.6 months additional life. These cash transfers accounted for about 30-40 percent of pre-transfer maternal income. Therefore, the effects of childhood exposure to the pandemic are about 62-94 percent of a relatively large and long-lasting cash transfer program.

6. Conclusion

The 1918 influenza pandemic, as the largest pandemic in recorded history, claimed hundreds of thousands of lives in the US. However, its massive death toll and welfare impacts were not limited to only short-run outcomes. Its impacts were also detectable in a wide array of outcomes for the survivors decades later. Studies suggest that early-life exposure to the pandemic is associated with lower education, socioeconomic measures, and deteriorated health outcomes (Beach, Clay, et al., 2022). In this paper, we provided empirical evidence that childhood exposure to the 1918 influenza resulted in significant reductions in longevity.

We used social security administration death records linked with the full-count 1940 census. We implemented event-study and difference-in-difference method to compare longevity across different age-at-exposure of individuals living in cities with high influenza-related mortality versus low exposure cities. Our results suggest reductions in longevity of about 1.8 and 2.7 months for age-at-exposure of 0-5 and 6-11, respectively. We calculated treatment-on-treated effects of about 7.2 and 10.9 months, respectively. Further, we used information in the 1940 census and

¹⁶ This is calculated using the formula provided by Colmer (2020), as follows: $VSLY = \frac{rVSL}{1-(1+r)^{-L}}$ where r is a discount rate and L is the estimated remaining life years.

1960-1980 censuses and document reductions in education and socioeconomic scores as potential mechanism channels.

Our results can be used in cost-benefit calculations to justify government interventions during pandemics. Further, since the results reveal childhood ages as the primary age of impact, the findings may also help identifying the population at higher risks who would benefit most from interventions during major pandemics.

Reference

- Abramitzky, R., Boustan, L., & Rashid, M. (2020). *Census Linking Project: Version 1.0 [dataset]*. <https://doi.org/https://censuslinkingproject.org>
- Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2022). The impact of the coronavirus lockdown on mental health: evidence from the United States. *Economic Policy*, 37(109), 139–155. <https://doi.org/10.1093/EPOLIC/EIAC002>
- Aizer, A., Eli, S., Ferrie, J., & Muney, A. L. (2016). The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review*, 106(4), 935–971. <https://doi.org/10.1257/AER.20140529>
- Almond, D. (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. <https://doi.org/10.1086/507154>
- Almond, D., & Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, 25(3), 153–172. <https://doi.org/10.1257/JEP.25.3.153>
- Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, 56(4), 1360–1446.
- Almond, D., & Mazumder, B. (2005). The 1918 influenza pandemic and subsequent health outcomes: {An} analysis of {SIPP} data. *American Economic Review*, 95(2), 258–262.
- Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1), 151–184. <https://doi.org/10.1086/426036>
- Arias, E. (2014). United States Life Tables, 2009. *Public Health Resources*. <https://digitalcommons.unl.edu/publichealthresources/322>
- Bailey, M., Clay, K., Fishback, P., Haines, M., Kantor, S., Severnini, E., & Wentz, A. (2016). U.S. County-Level Natality and Mortality Data, 1915-2007. *Inter-University Consortium for Political and Social Research*. <https://doi.org/https://doi.org/10.3886/E100229V4>
- Barker, D. J. P. (1994). *Mothers, babies, and disease in later life*. BMJ publishing group London.
- Barker, D. J. P. (1995). Fetal origins of coronary heart disease. *BMJ*, 311(6998), 171–174.

- <https://doi.org/10.1136/BMJ.311.6998.171>
- Barker, D. J. P. (1997). Maternal nutrition, fetal nutrition, and disease in later life. *Nutrition*, 13(9), 807–813. [https://doi.org/10.1016/S0899-9007\(97\)00193-7](https://doi.org/10.1016/S0899-9007(97)00193-7)
- Barro, R. J., Ursúa, J. F., Weng, J., Ursúa Dodge, J. F., & Weng EverLife, J. (2020). *The Coronavirus and the Great Influenza Pandemic: Lessons from the “Spanish Flu” for the Coronavirus’s Potential Effects on Mortality and Economic Activity*. <https://doi.org/10.3386/W26866>
- Beach, B., Brown, R., Ferrie, J., Saavedra, M., & Thomas, D. (2022). Reevaluating the Long-Term Impact of In Utero Exposure to the 1918 Influenza Pandemic. *Journal of Political Economy*, 130(7), 1963–1990. https://doi.org/10.1086/719757/SUPPL_FILE/2016445DATA.ZIP
- Beach, B., Clay, K., & Saavedra, M. (2022). The 1918 Influenza Pandemic and Its Lessons for COVID-19. *Journal of Economic Literature*, 60(1), 41–84. <https://doi.org/10.1257/JEL.20201641>
- Bozzoli, C., Deaton, A., & Quintana-Domeque, C. (2009). Adult height and childhood disease. *Demography*, 46(4), 647–669. <https://doi.org/10.1353/DEM.0.0079>
- Brainerd, E., & Siegler, M. V. (2003). The economic effects of the 1918 influenza epidemic. *Available at SSRN 394606*.
- Breen, C. F., & Osborne, M. (2022). *An Assessment of CenSoc Match Quality*. <https://doi.org/10.31235/OSF.IO/BJ5MD>
- Brew, B. K., Lundholm, C., Caffrey Osvald, E., Chambers, G., Öberg, S., Fang, F., & Almqvist, C. (2022). Early-Life Adversity Due to Bereavement and Inflammatory Diseases in the Next Generation: A Population Study in Transgenerational Stress Exposure. *American Journal of Epidemiology*, 191(1), 38–48. <https://doi.org/10.1093/AJE/KWAB236>
- Burgner, D. P., Cooper, M. N., Moore, H. C., Stanley, F. J., Thompson, P. L., De Klerk, N. H., & Carter, K. W. (2015). Childhood Hospitalisation with Infection and Cardiovascular Disease in Early-Mid Adulthood: A Longitudinal Population-Based Study. *PLOS ONE*, 10(5), e0125342. <https://doi.org/10.1371/JOURNAL.PONE.0125342>
- CDC. (2018). *History of 1918 Flu Pandemic*. <https://www.cdc.gov/flu/pandemic-resources/1918->

commemoration/1918-pandemic-history.htm

- Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001–2014. *JAMA*, 315(16), 1750–1766. <https://doi.org/10.1001/JAMA.2016.4226>
- Clay, K., Lewis, J., & Severnini, E. (2019). What explains cross-city variation in mortality during the 1918 influenza pandemic? Evidence from 438 U.S. cities. *Economics & Human Biology*, 35, 42–50. <https://doi.org/10.1016/J.EHB.2019.03.010>
- Colmer, J. (2020). What is the meaning of (statistical) life? Benefit–cost analysis in the time of COVID-19. *Oxford Review of Economic Policy*, 36(Supplement_1), S56–S63. <https://doi.org/10.1093/OXREP/GRAA022>
- Cook, C. J., Fletcher, J. M., & Forgues, A. (2019). Multigenerational Effects of Early-Life Health Shocks. *Demography*, 56(5), 1855–1874. <https://doi.org/10.1007/S13524-019-00804-3>
- Currie, J., & Moretti, E. (2003). Mother’s Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings. *The Quarterly Journal of Economics*, 118(4), 1495–1532. <https://doi.org/10.1162/003355303322552856>
- Deaton, A. (2007). Height, health, and development. *Proceedings of the National Academy of Sciences*, 104(33), 13232–13237. <https://doi.org/10.1073/PNAS.0611500104>
- Deeg, D. J. H., Hofman, A., & Van Zonneveld, R. J. (1990). THE ASSOCIATION BETWEEN CHANGE IN COGNITIVE FUNCTION AND LONGEVITY IN DUTCH ELDERLY. *American Journal of Epidemiology*, 132(5), 973–982. <https://doi.org/10.1093/OXFORDJOURNALS.AJE.A115740>
- Fletcher, J., Kim, J., Nobles, J., Ross, S., & Shaorshadze, I. (2021). The effects of foreign-born peers in us high schools and middle schools. *Journal of Human Capital*, 15(3), 432–468. https://doi.org/10.1086/715019/SUPPL_FILE/200111APPENDIX.PDF
- Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: Compulsory schooling laws revisited. *Social Science & Medicine*, 127, 101–107. <https://doi.org/10.1016/J.SOCSCIMED.2014.09.052>
- Fletcher, J. M. (2018a). New evidence on the impacts of early exposure to the 1918 influenza

- pandemic on old-age mortality. *Biodemography and Social Biology*, 64(2), 123–126.
<https://doi.org/10.1080/19485565.2018.1501267>
- Fletcher, J. M. (2018b). Examining the long-term mortality effects of early health shocks. *Applied Economics Letters*, 26(11), 902–908.
<https://doi.org/10.1080/13504851.2018.1520960>
- Fletcher, J. M. (2018c). The effects of in utero exposure to the 1918 influenza pandemic on family formation. *Economics & Human Biology*, 30, 59–68.
<https://doi.org/10.1016/J.EHB.2018.06.004>
- Gagnon, A., Miller, M. S., Hallman, S. A., Bourbeau, R., Herring, D. A., Earn, D. J. D., & Madrenas, J. (2013). Age-Specific Mortality During the 1918 Influenza Pandemic: Unravelling the Mystery of High Young Adult Mortality. *PLOS ONE*, 8(8), e69586.
<https://doi.org/10.1371/JOURNAL.PONE.0069586>
- Garrett, T. A. (2007). Economic effects of the 1918 influenza pandemic. *Federal Reserve Bank of St. Louis*, 26, 74–94.
- Goldin, C. (1998). America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century. *The Journal of Economic History*, 58(2), 345–374. <https://doi.org/10.1017/S0022050700020544>
- Goldin, C., & Katz, L. F. (1999). The Shaping of Higher Education: The Formative Years in the United States, 1890 to 1940. *Journal of Economic Perspectives*, 13(1), 37–62.
<https://doi.org/10.1257/JEP.13.1.37>
- Goldstein, J. R., Alexander, M., Breen, C., Miranda González, A., Menares, F., Osborne, M., Snyder, M., & Yildirim, U. (2021). Censoc Project. In *CenSoc Mortality File: Version 2.0*. Berkeley: University of California. <https://censoc.berkeley.edu/data/>
- Goldstein, J. R., Breen, C., Alexander, M., Miranda González, A., Menares, F., Osborne, M., Snyder, M., Yildirim, U., & Wickle, A. (2023). *CenSoc Army Enlistment Records*.
<https://doi.org/10.7910/DVN/ZFVVNA>
- Goodman, A., Joyce, R., & Smith, J. P. (2011). The long shadow cast by childhood physical and mental problems on adult life. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), 6032–6037. <https://doi.org/10.1073/PNAS.1016970108>

- Halpern-Manners, A., Helgertz, J., Warren, J. R., & Roberts, E. (2020). The Effects of Education on Mortality: Evidence From Linked U.S. Census and Administrative Mortality Data. *Demography*, 57(4), 1513–1541. <https://doi.org/10.1007/S13524-020-00892-6>
- Hayward, M. D., & Gorman, B. K. (2004). The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography* 2004 41:1, 41(1), 87–107. <https://doi.org/10.1353/DEM.2004.0005>
- Helgertz, J., & Bengtsson, T. (2019). The Long-Lasting Influenza: The Impact of Fetal Stress During the 1918 Influenza Pandemic on Socioeconomic Attainment and Health in Sweden, 1968–2012. *Demography*, 56(4), 1389–1425. <https://doi.org/10.1007/S13524-019-00799-X>
- Jain, N. (2020). The early life education of the immune system: Moms, microbes and (missed) opportunities. *Gut Microbes*, 12(1). <https://doi.org/10.1080/19490976.2020.1824564>
- Kane, T. J., & Rouse, C. E. (1999). The Community College: Educating Students at the Margin between College and Work. *Journal of Economic Perspectives*, 13(1), 63–84. <https://doi.org/10.1257/JEP.13.1.63>
- Karlsson, M., Nilsson, T., & Pichler, S. (2014). The impact of the 1918 Spanish flu epidemic on economic performance in Sweden: An investigation into the consequences of an extraordinary mortality shock. *Journal of Health Economics*, 36, 1–19.
- Kim, J., & Fletcher, J. M. (2021). The effects of relative body weight on socioemotional and schooling outcomes among female adolescents in the United States. *Social Science & Medicine*, 289, 114418. <https://doi.org/10.1016/J.SOCSCIMED.2021.114418>
- Kniesner, T. J., & Viscusi, W. K. (2019). The Value of a Statistical Life. *Oxford Research Encyclopedia of Economics and Finance*. <https://doi.org/10.1093/ACREFORE/9780190625979.013.138>
- Kremen, W. S., Beck, A., Elman, J. A., Gustavson, D. E., Reynolds, C. A., Tu, X. M., Sanderson-Cimino, M. E., Panizzon, M. S., Vuoksima, E., Toomey, R., Fennema-Notestine, C., Hagler, D. J., Fang, B., Dale, A. M., Lyons, M. J., & Franz, C. E. (2019). Influence of young adult cognitive ability and additional education on later-life cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 116(6), 2021–2026.

- https://doi.org/10.1073/PNAS.1811537116/SUPPL_FILE/PNAS.1811537116.SAPP.PDF
- Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., Wesolowski, A., Santillana, M., Zhang, C., Du, X., Yu, H., & Tatem, A. J. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* 2020 585:7825, 585(7825), 410–413. <https://doi.org/10.1038/s41586-020-2293-x>
- Lee, H., Lee, M. W., Warren, J. R., & Ferrie, J. (2022). Childhood lead exposure is associated with lower cognitive functioning at older ages. *Science Advances*, 8(45), 5164. <https://doi.org/10.1126/SCIADV.ABN5164>
- Lin, M.-J., & Liu, E. M. (2014). Does in utero exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment. *Journal of Health Economics*, 37, 152–163.
- Lleras-Muney, A. (2005). The Relationship Between Education and Adult Mortality in the United States. *The Review of Economic Studies*, 72(1), 189–221. <https://doi.org/10.1111/0034-6527.00329>
- Mamelund, S. E., Shelley-Egan, C., & Rogeberg, O. (2021). The association between socioeconomic status and pandemic influenza: Systematic review and meta-analysis. *PLOS ONE*, 16(9), e0244346. <https://doi.org/10.1371/JOURNAL.PONE.0244346>
- Mazumder, B., Almond, D., Park, K., Crimmins, E. M., & Finch, C. 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. <https://doi.org/10.1017/S2040174409990031>
- Miller, R., & Bairoliya, N. (2021). Health, Longevity, and Welfare Inequality of Older Americans. *The Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/REST_A_01103
- Myrskylä, M., Mehta, N. K., & Chang, V. W. (2013). Early life exposure to the 1918 influenza pandemic and old-age mortality by cause of death. *American Journal of Public Health*, 103(7). <https://doi.org/10.2105/AJPH.2012.301060>
- Nelson, R. E. (2010). Testing the Fetal Origins Hypothesis in a developing country: evidence from the 1918 Influenza Pandemic. *Health Economics*, 19(10), 1181–1192. <https://doi.org/10.1002/HEC.1544>

- Noymer, A., & Garenne, M. (2000). The 1918 Influenza Epidemic's Effects on Sex Differentials in Mortality in the United States. *Population and Development Review*, 26(3), 565–581. <https://doi.org/10.1111/J.1728-4457.2000.00565.X>
- Patterson, K. D., & Pyle, G. F. (1991). The geography and mortality of the 1918 influenza pandemic. *Bulletin of the History of Medicine*, 65(1), 4–21.
- Phillippe, K. A., Sullivan, L. G., & American Association of Community Colleges, W. (2005). *National Profile of Community Colleges Trends and Statistics, 4th Edition*. 201.
- Rogers, S. L. (1920). Special tables of mortality from influenza and pneumonia. *Indiana, Kansas, and Philadelphia, PA*.
- Rostila, M., & Saarela, J. M. (2011). Time Does Not Heal All Wounds: Mortality Following the Death of a Parent. *Journal of Marriage and Family*, 73(1), 236–249. <https://doi.org/10.1111/J.1741-3737.2010.00801.X>
- Ruggles, S., Flood, S., Goeken, R., Grover, J., & Meyer, E. (2020). IPUMS USA: Version 10.0 [dataset]. *Minneapolis, MN: IPUMS*. <https://doi.org/10.18128/D010.V10.0>
- Smith, K. R., Hanson, H. A., Norton, M. C., Hollingshaus, M. S., & Mineau, G. P. (2014). Survival of offspring who experience early parental death: Early life conditions and later-life mortality. *Social Science & Medicine*, 119, 180–190. <https://doi.org/10.1016/J.SOCSCIMED.2013.11.054>
- Spinelli, M., Lionetti, F., Pastore, M., & Fasolo, M. (2020). Parents' Stress and Children's Psychological Problems in Families Facing the COVID-19 Outbreak in Italy. *Frontiers in Psychology*, 11, 556582. <https://doi.org/10.3389/FPSYG.2020.01713/BIBTEX>
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199. <https://doi.org/10.1016/J.JECONOM.2020.09.006>
- Thomas, B. (1951). Historical Statistics of the United States, 1789–1945. *Journal of the Royal Statistical Society: Series A (General)*, 114(1), 110–110. <https://doi.org/10.2307/2980918>
- Viscusi, W. K. (2018). Best Estimate Selection Bias in the Value of a Statistical Life. *Journal of Benefit-Cost Analysis*, 9(2), 205–246. <https://doi.org/10.1017/BCA.2017.21>

Tables

Table 1 - Summary Statistics

	Mean	SD	Min	Max
<i>DMF-1940-Census Data:</i>				
Death Age (Months)	916.38379	108.66594	589	1286
Birth Year	1914.2996	7.46569	1898	1926
Death Year	1989.6736	8.56601	1975	2005
Age 0-5	.27084	.4444	0	1
Age 6-11	.2137	.40992	0	1
Age 12-17	.13253	.33906	0	1
1918 Influenza Quartile 2-4	.72096	.44853	0	1
Age 0-5 × Q 2-4	.19427	.39564	0	1
Age 6-11 × Q 2-4	.1517	.35873	0	1
Age 12-17 × Q 2-4	.0956	.29404	0	1
White	.96704	.17854	0	1
Black	.03082	.17284	0	1
Father SEI 1 st Quartile	.57895	.49373	0	1
Father SEI 2 nd Quartile	.1096	.31239	0	1
Father SEI 3 rd Quartile	.11596	.32017	0	1
Father SEI 4 th Quartile	.10195	.30258	0	1
Father SEI Missing	.09354	.29119	0	1
Mother Education < High School	.37587	.48435	0	1
Mother Education = High School	.09244	.28964	0	1
Mother Education > High School	.01631	.12668	0	1
Mother Education Missing	.51538	.49976	0	1
City Literacy Rate	.95179	.02857	.70326	1
City Share of Married	.5595	.03529	.18262	.67497
City Socioeconomic Index	31.48381	3.09182	9.5442	47.26709
Observations		1,471,443		
<i>Mechanism Data (1940 Census):</i>				
Schooling	10.33	2.88	0	20
Occupational Income Score	27.16	9.06	3	80
Duncan Socioeconomic Index	34.12	21.65	3	96
White-Collar Occupation	.15	.36	0	1
Observations		1013221		
<i>Mechanism Data (1980-2000 Census and 2005 ACS):</i>				
Schooling	12.89	2.59	0	20
Occupational Income Score	26.59	9.54	3	80
Duncan Socioeconomic Index	42.65	22.83	4	96
White-Collar Occupation	.19	.39	0	1
Observations		1096712		
<i>Enlistment Data:</i>				
Height-for-Age (STD)	0	1	-2.68	3.39
AGCT Score	81.11	43.37	1	150
Normal BMI	.72	.45	0	1
Observations		219380		

Table 2 - Balancing Tests

	<i>Outcomes:</i>							
	White	Black	Father SEI	Father SEI Missing	Father Schooling	Father Schooling Missing	Mother Schooling	Mother Schooling Missing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age 0-5 × Q 2-4	-.02733 (.01692)	.02529 (.01648)	-.04102 (.20704)	.03325*** (.00941)	.04039 (.0328)	.03407*** (.00913)	.0492 (.04101)	.0316*** (.00978)
Age 6-11 × Q 2-4	-.02542 (.01695)	.02359 (.01644)	-.16166 (.41799)	.00116 (.00814)	.12602 (.08771)	.00664 (.00659)	.11065 (.0792)	.00588 (.00536)
Age 12-17 × Q 2-4	-.02384 (.0169)	.02196 (.01633)	.05251 (.67801)	-.00479 (.00888)	-.01636 (.16365)	.0021 (.00688)	.04409 (.1273)	.00181 (.00519)
Observations	1471443	1471443	540648	1471443	604840	1471443	713093	1471443
R-squared	.12379	.13152	.02825	.35884	.06887	.34533	.08769	.37341
Mean DV	0.967	0.031	32.689	0.579	6.998	0.589	6.919	0.515
City FE	✓	✓	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 3 - The Association between Exposure to 1918 Influenza and Successful DMF-Census Merging

	<i>Outcome: Successful Merging between Final Sample and Original Cohort of 1940 Census</i>	
	(1)	(2)
Age 0-5 × Q 2-4	.00055 (.00152)	.00038 (.00074)
Age 6-11 × Q 2-4	-.0011 (.00215)	-.00104 (.00139)
Age 12-17 × Q 2-4	-.0004 (.00212)	-.00061 (.00146)
Observations	16224287	16224287
R-squared	.00717	.11102
Mean DV	0.088	0.088
City FE	✓	✓
Region-Year FE	✓	✓
Controls		✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Table 4 - Childhood Exposure to 1918 Influenza and Later-Life Male Mortality

	<i>Outcome: Age at Death (Months)</i>			
	(1)	(2)	(3)	(4)
Age 0-5 × Q 2-4	-1.49665 (.92493)	-1.95447** (.80084)	-1.8679** (.82137)	-1.81434** (.8086)
Age 6-11 × Q 2-4	-1.90054* (.99022)	-2.74827*** (.92542)	-2.79489*** (.99368)	-2.74531*** (.98771)
Age 12-17 × Q 2-4	.13115 (.26367)	-.7638 (.49956)	-.82524 (.56076)	-.8852 (.56303)
Observations	1471443	1471443	1471443	1471443
R-squared	.23119	.23138	.23188	.23188
Mean DV	916.384	916.384	916.384	916.384
City FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Region-Year FE		✓	✓	✓
Family Controls			✓	✓
City Controls				✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Table 5 - Heterogeneity Analysis of Childhood Exposure to 1918 Influenza and Later-Life Male Mortality

	<i>Outcome: Age at Death (Months), Subsamples:</i>					
	Whites	Blacks	High Educated Mother	Low Educated Mother	Father SEI High	Father SEI Low
	(1)	(2)	(3)	(4)	(5)	(6)
Age 0-5 × Q 2-4	-1.93318** (.80472)	-.5447 (4.29906)	-1.91514** (.70918)	-1.86044* (1.01075)	-1.86647** (.79864)	-1.6141 (1.2647)
Age 6-11 × Q 2-4	-2.84417*** (1.01543)	-2.83478 (4.63234)	-2.99102*** (.92804)	-2.54786* (1.43138)	-2.67193** (1.00285)	-2.94193 (1.85041)
Age 12-17 × Q 2-4	-.92264 (.62592)	-4.63085 (4.10812)	-.68216 (.50517)	-3.31517*** (.95864)	-.49512 (.55205)	-5.73369** (2.23788)
Observations	1422937	45318	918370	553073	1139551	331892
R-squared	.2292	.30202	.21696	.16673	.22619	.16473
Mean DV	916.929	899.633	933.997	887.138	925.802	884.048
City FE	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Table 6 - Robustness Checks

	<i>Outcome: Age at Death (Months)</i>				
	Birth-State by Birth-Year FE	Interacting City FE by Covariate Dummies	Adding City Trend	Adding Death-Month and Birth-Month FE	Adding More City Controls
	(1)	(2)	(3)	(4)	(5)
Age 0-5 × Q 2-4	-1.53382* (.79592)	-2.06448** (.8527)	-1.63376* (.79985)	-1.78772** (.81042)	-1.48164* (.74071)
Age 6-11 × Q 2-4	-2.32527** (1.00084)	-3.1339*** (.97885)	-2.3751** (.96479)	-2.73177*** (.98692)	-1.9577** (.86105)
Age 12-17 × Q 2-4	-.33479 (.67542)	-1.2627** (.54117)	-.42167 (.41653)	-.92617 (.5694)	-.01489 (.6855)
Observations	1471443	1471409	1471443	1471443	1471443
R-squared	.23233	.23327	.23214	.23253	.23192
Mean DV	916.384	916.384	916.384	916.384	916.384
	Outcome: Log Age at Death	Outcome: Age at Death > 75	Outcome: Age at Death > 80	Using CenSoc Weights	Clustering on City
	(6)	(7)	(8)	(9)	(10)
	(6)	(7)	(8)	(9)	(10)
Age 0-5 × Q 2-4	-.00193** (.00086)	-.00571 (.00364)	-.00837** (.00382)	-1.52191** (.68079)	-1.81434*** (.70155)
Age 6-11 × Q 2-4	-.00285*** (.00096)	-.00979* (.00517)	-.01338** (.00651)	-2.59639** (1.26548)	-2.74531*** (1.01036)
Age 12-17 × Q 2-4	-.00086* (.00048)	-.00059 (.0022)	-.00462 (.00315)	-.69015 (.77445)	-.8852 (.61494)
Observations	1471443	1471443	1471443	1295978	1471443
R-squared	.2315	.13079	.12514	.13296	.23188
Mean DV	6.813	0.542	0.322	937.868	916.384

Notes. Standard errors, clustered on city and year, are in parentheses. Regressions include birth-city and birth-year fixed effects. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate. City controls in column 5 include population, share of immigrants, average occupational income score, labor force participation rates, share of people in difference race groups, age groups, and occupation groups.

*** p<0.01, ** p<0.05, * p<0.1

Table 7 - Mechanism Analysis: Childhood Exposure to 1918 Influenza and Later-Life Education and Socioeconomic Measures

	<i>Outcomes and Subsamples:</i>							
	Final Sample in 1940				Decennial Censuses 1960-1980			
	Schooling	Occupational Income Score	Socioeconomic Index	White-Collar Occupation	Schooling	Occupational Income Score	Socioeconomic Index	White- Collar Occupation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age 0-5 × Q	.05254	-.13911	-.48163*	-.00621	-.1799***	-.44495***	-.974***	-.1799***
2-4	(.03544)	(.10459)	(.24908)	(.00385)	(.04437)	(.09855)	(.22082)	(.04437)
Age 6-11 × Q	-.00093	-.49707***	-1.02853***	-.01803***	-.30887***	-.8944***	-1.46526***	-.30887***
2-4	(.0358)	(.09811)	(.21994)	(.0037)	(.04957)	(.10521)	(.25132)	(.04957)
Age 12-17 ×	.04622	-.16377	-.34496	-.01161***	-.24212***	-.71881***	-1.27743***	-.24212***
Q 2-4	(.03812)	(.10186)	(.22666)	(.00392)	(.06074)	(.11752)	(.2875)	(.06074)
Observations	998989	1013221	1013221	1013221	349133	1096712	1096712	349133
R-squared	.08346	.07135	.07829	.03287	.15518	.11144	.09501	.15518
Mean DV	10.333	27.160	34.116	0.155	12.892	26.587	42.650	12.892
City FE	✓	✓	✓	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies and county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Table 8 - Exploring the Effects on Mediatory Outcomes Using WWII Enlistment Data

	<i>Outcomes:</i>			
	Being in the Enlistment Data	AGCT Score	Height-for-Age (Standardized)	Normal BMI
	(1)	(2)	(3)	(4)
Age 0-5 × Q 2-4	-.14381 (.16416)	-.06413** (.03019)	-4.13686 (2.78773)	-.00244 (.00509)
Age 6-11 × Q 2-4	-.12427 (.10407)	-.06556** (.02651)	-4.64079* (2.55934)	-.01228** (.00536)
Age 12-17 × Q 2-4	-.094 (.08986)	-.0499 (.03102)	1.59868 (3.70105)	-.02927 (.02119)
Observations	1471443	219377	37118	218959
R-squared	.1509	.94952	.29884	.02744
Mean DV	0.146	0.000	81.135	0.724
City FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓
City Controls	✓	✓	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Figures

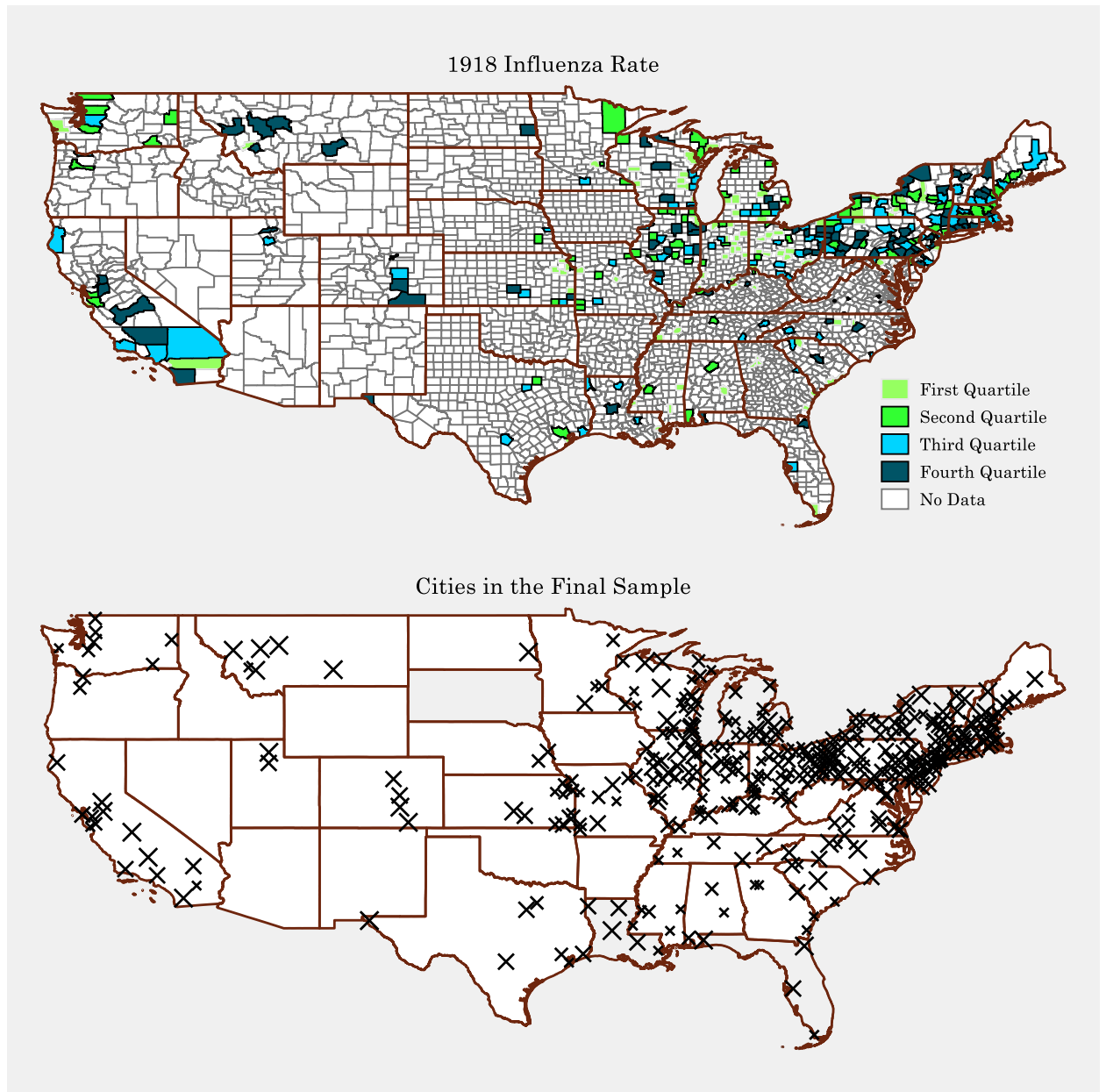


Figure 1 - Geographic Distribution of Cities in the Final Sample

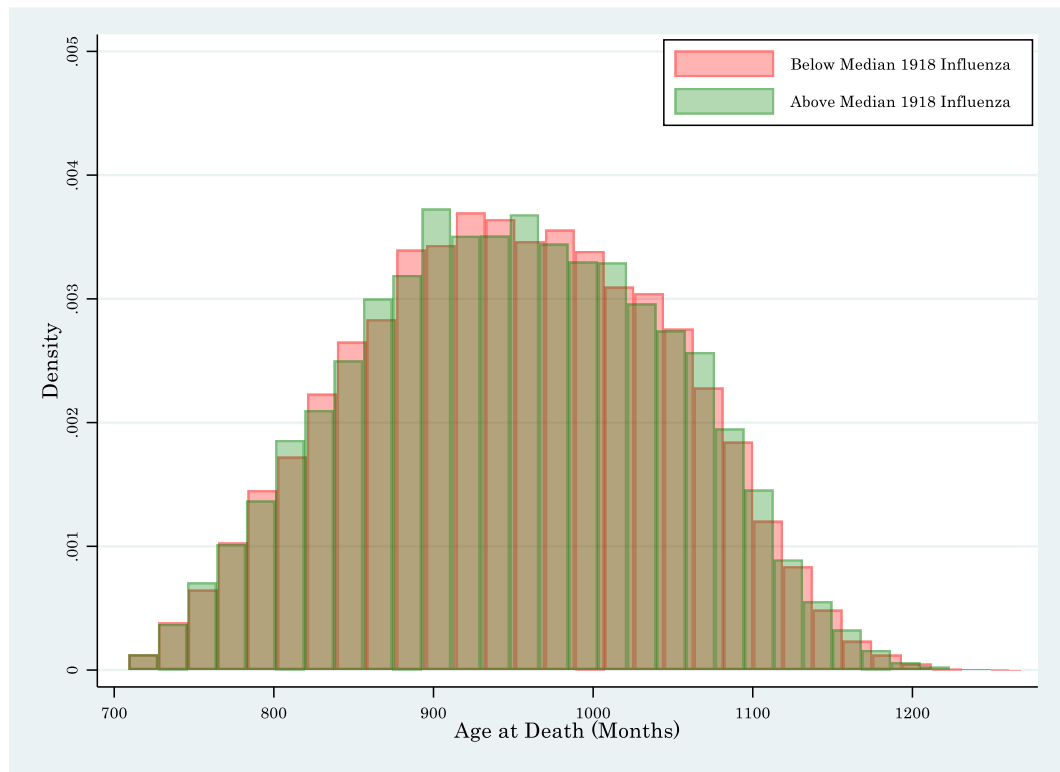


Figure 2 - Density Distribution of Age at Death based on Cumulative Childhood Exposure to the 1918 Influenza Pandemic

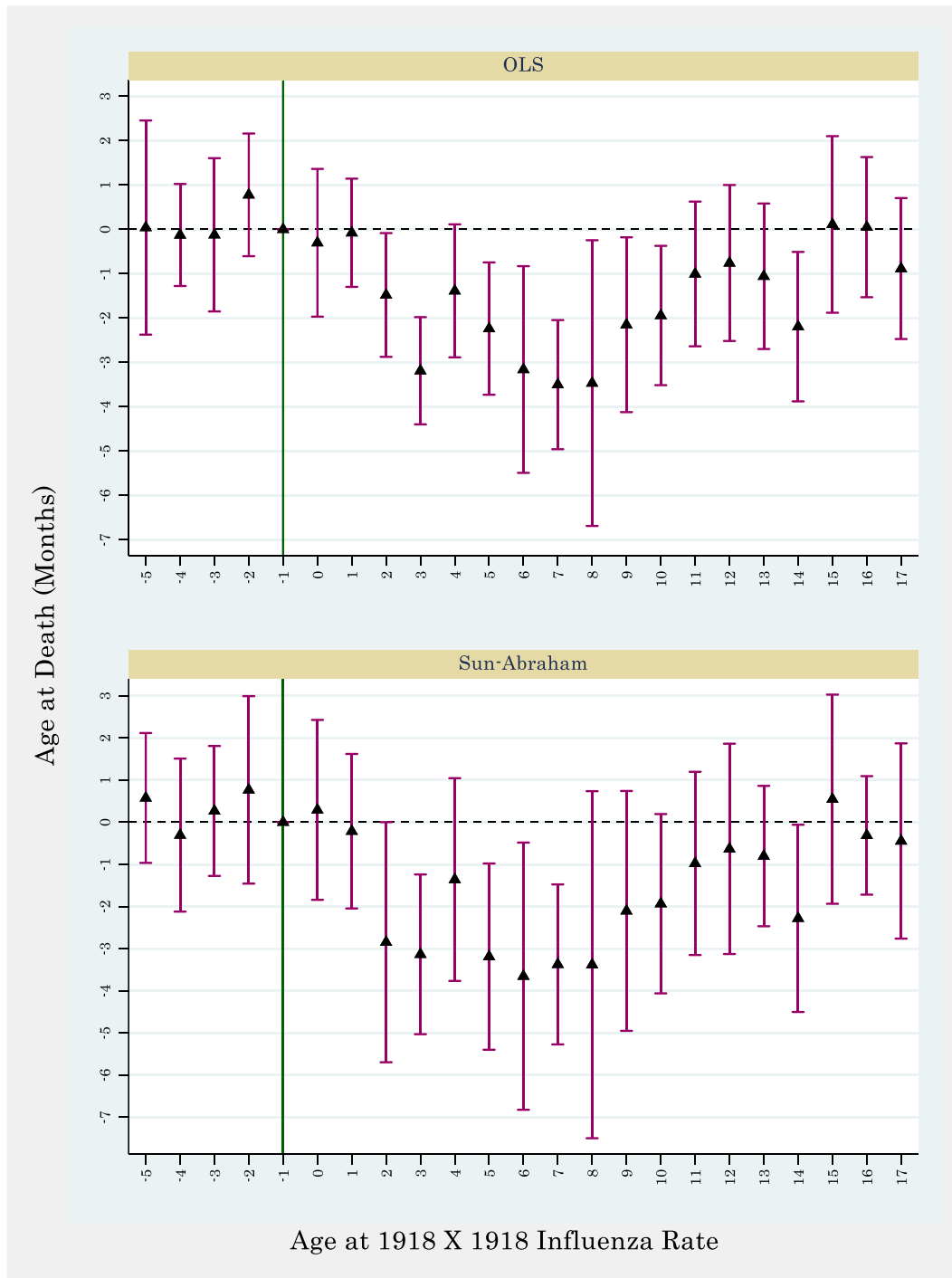


Figure 3 – Event Study of Childhood Exposure to 1918 Influenza and Later-Life Longevity

Notes. Point estimates and 90 percent confidence intervals are depicted. Standard errors are clustered on city and year. Regressions include birth-city and birth-region by birth-year fixed effects. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

Appendix A

The DMF data used in the paper covers deaths only to male individuals. To examine the effects on females, we use an alternative data source. We employ Numerical Identification (Numident) data extracted from Goldstein et al. (2021). The Numident data covers deaths that occur to both genders over the years 1988-2005. Similar to DMF, Numident is linked to the full-count 1940 census. We employ similar sample selections as in the DMF and implement equation 2 to compare the effects with those of DMF. The results are reported in Appendix Table A-1. In column 1, we replicate column 4 of Table 4. In column 2, we restrict DMF data to the Numident death years, i.e., years 1988-2005. We observe reductions in magnitude of the effects suggesting that the effects are larger as we expand death window and cover earlier death years. In column 3, we replicate the regressions using Numident data (death years 1988-2005) and restrict the sample to males to be able to compare the results with those of column 2. We observe almost identical effects for age 0-5 but substantially larger coefficient for age 6-11. In column 4, we restrict the Numident data to females. The effects are only slightly smaller than those of column 3 suggesting little heterogeneity in the effects by gender. Therefore, we argue that if we had earlier death data for females, we would have observed comparable effects for females as those in column 1. Finally, we report the results of full Numident data in column 5. We observe comparable effects compared with those in columns 3-4.

Appendix Table A-1 - Comparing the Results with the Numident Data Analysis for Death Years 1988-2005

	<i>Outcome: Age at Death (Months)</i>				
	DMF (1975-2005)	DMF (1988-2005)	Numident (1988-2005), Males Only	Numident (1988-2005), Females Only	Numident (1988-2005), Both Gender
	(1)	(2)	(3)	(4)	(5)
Age 0-5 × Q 2-4	-1.81503** (.80958)	-.76964 (.4702)	-.78577 (.49808)	-.59811** (.26576)	-.68366** (.33126)
Age 6-11 × Q 2-4	-2.74976*** (.99022)	-.40864 (.42274)	-1.54117* (.83855)	-1.14445 (.79825)	-1.08625 (.70115)
Age 12-17 × Q 2-4	-.89202 (.5648)	.06205 (.44342)	-.74311 (.81285)	-.73776 (.84221)	-.55275 (.65324)
Observations	1471443	839974	831405	1002701	1834106
R-squared	.23188	.53325	.39119	.53747	.50046
Mean DV	916.384	968.186	941.246	975.260	959.841
City FE	✓	✓	✓	✓	✓
Region-Year FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Regressions include birth-city and birth-year fixed effects. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Appendix B

In this appendix, we employ a state-level analysis to examine the effects on infant mortality rates and birth rates. We use data from Bailey et al. (2016) that reports infant deaths and births by race for a subset of states for the years 1915-onwards. We focus on several years before and after the pandemic for a balanced panel of states. We implement regressions of the following forms:

$$y_{st} = \alpha + \beta_1 I(t = 1918) \times Exp_s + \beta_2 I(t = 1919) \times Exp_s + \gamma X_{st} + \lambda_s + \eta_t + \varepsilon_{st} \quad (B-1)$$

Where y is log of infant mortality rate and birth rate for state s and year t . $I(.)$ is an indicator function. Exp is a dummy indicating the state is at the above-median of 1918 influenza mortality and zero otherwise. In X , we include state-level share of married, average socioeconomic score, and literacy rate. State and year fixed effects are included in λ and η , respectively. The sample covers the years 1916-1925. Standard errors are clustered on state.

The results are shown in Appendix Table B-1. Compared to adjacent years, we observe about 8.4 percent rise in infant mortality rate, 7.9 percent increase in white infant mortality rate, and 21 percent rise in black infant mortality rate (columns 1-3). However, we do not observe significant changes in birth rates and birth rate by race groups.

Appendix Table B-1 - State-Level Analysis of Exposure to the 1918 Influenza on Infant Mortality Rate and Birth Rate

	<i>Outcomes:</i>					
	Log Infant Mortality Rate	Log Infant Mortality Rate, Whites	Log Infant Mortality Rate, Blacks	Log Birth Rate	Log Birth Rate, Whites	Log Birth Rate, Blacks
	(1)	(2)	(3)	(4)	(5)	(6)
Year 1918 × Q 2-4	.08477** (.03087)	.07953** (.03274)	.21003** (.07847)	.01198 (.01041)	.02211 (.01472)	-.02487 (.06919)
Year 1919 × Q 2-4	-.00582 (.01904)	-.0034 (.01829)	.04165 (.10629)	-.01077 (.02219)	-.00771 (.02313)	-.03992 (.0488)
Observations	235	235	235	235	235	235
R-squared	.95681	.95779	.68481	.96246	.95688	.99341
Mean DV	4.356	4.308	4.843	3.375	3.299	-0.263
State FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include state-level share of married, average socioeconomic score, and literacy rate. The sample covers the years 1916-1925.

*** p<0.01, ** p<0.05, * p<0.1

Appendix C

In the main results, we focused on cities at the top three quartiles of influenza rate during 1918-1919. We chose this threshold for ease of interpretation and comparison of highly affected areas with lower affected areas. In Appendix Figure C-1, We disaggregate the exposure measure across different quartiles of influenza rate and across different ages from 0 to 17. This figure reports the results of one regression that includes all covariates and fixed effects as the main results. For most ages, the effects across different quartiles of influenza rate are quite close and comparable. However, we also observe slight discrepancies. For instance, for age at exposure of 9, 11, 12, and 13, we observe virtually zero effect for the top quartile cities while for quartiles two and three we observe negative effects. One possible explanation is the influence of mortality and selection in high exposed cities while the scarring effects for medium exposure cities. Nonetheless, the overall picture is closely aligned with the results of Figure 3.

To further explore the influence of linearity in the effects, in Appendix Table C-1, we replicate the main results for specifications that replace the top three quartiles dummy with log of influenza rate (column 1) and standardized influenza rate (column 2). Although we observe a very similar pattern as the main results, we do not find statistically significant coefficients. The difference between bottom quartile and top-three quartile cities and influenza rate is roughly 12.5 (per 1,000). This is also approximately equivalent to a change of 100 percent from the mean of influenza rate in the final sample. Based on the results of Table 4, this increase is equivalent to 1.8 and 2.7 months for age at exposure of 0-5 and 6-12, respectively. Using column 1 of Appendix Table C-1, a 100 percent increase in influenza rate is associated with 0.6 and 1 months reductions in age at death. These numbers are considerably smaller than those implied by the main results. One reason to explain this discrepancy is the observed discrepancies in the effects on different quartiles of exposure, as depicted in Appendix Figure C-1. One possible scenario is that top quartile of exposure cities have substantially higher mortality rates and higher selection, leading to much smaller effects on later life longevity, bringing down the overall effects when we examine the linear exposure measure as in Appendix Table C-1.



Appendix Figure C-1 – Early-Life and Childhood Exposure to 1918 Influenza and Old Age Longevity by Age and by Quartile of City-Level Influenza Rate

Notes. Point estimates and 90 percent confidence intervals are depicted. Standard errors are clustered on city and year. Regressions include birth-city and birth-region by birth-year fixed effects. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

Appendix Table C-1 - Replicating the Main Results Using Influenza Rate as a Linear Exposure Measure

	<i>Outcome: Age-at-Death (Months)</i>	
	(1)	(2)
Age 0-5 × Log Influenza Rate	-.58492 (.51451)	
Age 6-11 × Log Influenza Rate	-1.04767 (.7017)	
Age 12-17 × Log Influenza Rate	.02835 (.48108)	
Age 0-5 × Influenza Rate (STD)		-.10659 (.23994)
Age 6-11 × Influenza Rate (STD)		-.21471 (.1478)
Age 12-17 × Influenza Rate (STD)		-.13345 (.27858)
Observations	1471417	1471443
R-squared	.23187	.23187
Mean DV	916.384	916.384
City FE	✓	✓
Region-Year FE	✓	✓
Controls	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Appendix D

As discussed in section 4.5, one possible channel of impact is parental death. A narrow strand of research suggests the long-run impacts of exposure to parental death during infancy and childhood for later life health outcomes, specifically mental health and suicide. We do not have any information on parental death in the census. However, we can observe the change in parental absence in census records of children in the same age range from 1910 to 1920. In so doing, we use the full count censuses 1910 and 1920 and limit the sample to male individuals aged 0-17. We then merge this data with the influenza rate data to cover similar cities as in the final sample of the paper. We use “person number” of parents in the household to detect presence of parents. Thus, a missing value of this variable may represent, among alternative explanations, parental death. We then regress parental death dummy on covariates and fixed effects similar to the main regressions of the paper. We also add your fixed effects as the comparison is derived from cross city and overtime variation. The results are reported in Appendix Table D-1. We observed positive coefficients across different ages at exposure. The coefficients suggest larger effects for mother absence. Since fathers (relative to mothers) are more likely to be absent for non-death reasons, the fact that we observed larger effects among mothers may point to possible influence of excess death observe in 1920 versus 1910. Moreover, these coefficients suggest relatively large impacts. For instance, we observe 37-93 basis points increases in the likelihood of maternal absence, off a mean of 0.06. This is equivalent to 5.8-14.7% increases with respect to the mean of maternal absence.

To examine whether there is heterogeneity by parental absence, we link the final sample of the paper and the sample that we constructed for Appendix Table D-1. We then examined the effects on longevity based on subsamples by parental absence. The results are reported in Appendix Table D-2. In column 1, we showed the effects for the full-linked sample. In columns 2-5, we show the effects by subsamples of absent mothers, presents mothers, absent fathers, and present fathers, respectively. For age at exposure of 0-5, we observe larger coefficients in the subsamples of absent fathers and absent mothers. However, for other age-of-exposures, we find mixed evidence. For age of exposure of 6-11, we observe similar coefficients based on maternal absence but larger coefficient in case of paternal presence. For age at exposure of 12 to 17, we observe larger coefficients for paternal absence but not for maternal absence. The overall consistent picture arises only for very early life exposure to parental death (i.e., ages 0-5).

Appendix Table D-1 - Exploring the Effects of the Pandemic on the Probability of Absent Parents (Proxy for Parental Death) Using Censuses 1920-1910

	<i>Outcomes:</i>	
	Father Absent	Mother Absent
	(1)	(2)
Age 0-5 × Q 2-4	.00395* (.0022)	.00376*** (.00124)
Age 6-11 × Q 2-4	.00278 (.00274)	.00416*** (.00132)
Age 12-17 × Q 2-4	.0049 (.00375)	.00936*** (.00331)
Observations	20428847	20428847
R-squared	.03708	.01881
Mean DV	0.111	0.063
City FE	✓	✓
Region-Year FE	✓	✓
Year FE	✓	✓
Controls	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table D-2 - Exploring the Heterogeneity by Parental Absence in 1920

	<i>Outcome: Age at Death (Months), Sample:</i>				
	Full-Linked Sample	Mother Absent	Mother Present	Father Absent	Father Present
	(1)	(2)	(3)	(4)	(5)
Age 0-5 × Q 2-4	-1.59608 (1.01078)	-2.31518 (5.92124)	-1.6126 (1.10693)	-2.34989 (4.23365)	-1.60223 (.94195)
Age 6-11 × Q 2-4	-3.66084*** (1.14057)	-3.72631 (4.31799)	-3.71824*** (1.20614)	-1.38857 (4.41642)	-3.98477*** (1.10359)
Age 12-17 × Q 2-4	-1.20153 (.84628)	.2424 (4.35382)	-1.34301 (.92791)	-2.05357 (3.74306)	-1.26195* (.7197)
Observations	378333	15217	363079	27475	350839
R-squared	.09572	.12889	.09555	.11599	.09502
Mean DV	929.608	940.669	929.143	940.534	928.750

Notes. Standard errors, clustered on city and year, are in parentheses. Regressions include birth-city and birth-year fixed effects. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate. City controls in column 5 include population, share of immigrants, average occupational income score, labor force participation rates, share of people in difference race groups, age groups, and occupation groups.

*** p<0.01, ** p<0.05, * p<0.1

Appendix E

The 1918 influenza pandemic started from Spain, hence so-called Spanish flu. World War I veterans were among the first victims and also the main channel for the spread of the flu. Therefore, it is possible that returning veterans carried the virus back to their home, resulting in higher probability of infection among their families. In the 1930 census, the veteran status variable reports whether individuals served the military during World War I. We link these records to parental records in 1940 census and use this final sample to implement some heterogeneity analyses based on father veteran status. The results are reported in Appendix Table E-1. In column 1, report the results for the full linked sample. In column 2, we restrict the sample to veteran fathers and in column 3, we focus on non-veteran subsample. We observed larger though insignificant coefficients for age at exposure of 0-5 and 12-17 for veterans' subsample but smaller coefficients for age at exposure of 6-11. However, among nonveterans, we observed significant and large effects for age of exposure of 6-11. Although the numbers of column 2 may suggest higher dosage of exposure to the disease on those aged 0-5 and 12-17, we do not observe the overall pattern of the full sample of column 1 or the main results.

Appendix Table E-1 - Heterogeneity in the Results based on Father's Veteran Status during World War I

	<i>Outcome: Age at Death (Months)</i>		
	Linked with 1930 Sample	Father WWI Veteran	Non-WWI Veteran
	(1)	(2)	(3)
Age 0-5 × Q 2-4	-.90865 (.90346)	-1.21134 (4.78761)	-.79064 (.92643)
Age 6-11 × Q 2-4	-2.5931** (1.02763)	-1.05699 (4.55616)	-2.50197** (.91206)
Age 12-17 × Q 2-4	-.67942 (.81326)	-2.58078 (8.97212)	-.53531 (.7635)
Observations	824922	49787	775122
R-squared	.23242	.12742	.22778
Mean DV	915.519	863.926	918.831
City FE	✓	✓	✓
Region-Year FE	✓	✓	✓
Controls	✓	✓	✓

Notes. Standard errors, clustered on city and year, are in parentheses. Controls include race dummies, father's socioeconomic status dummies, maternal education dummies, county-level share of married, average socioeconomic score, and literacy rate.

*** p<0.01, ** p<0.05, * p<0.1

