NBER WORKING PAPER SERIES

# POLLUTION, INFECTIOUS DISEASE, AND MORTALITY: EVIDENCE FROM THE 1918 SPANISH INFLUENZA PANDEMIC

Karen Clay
Joshua Lewis
Edson Severnini

Working Paper 21635
http://www.nber.org/papers/w21635

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2015, Revised May 2018

We thank Martha Bailey, Antonio Bento, Christian Dippel, Walker Hanlon, and seminar participants at the 2015 NBER DAE Summer Meetings, the 2015 CNEH Meetings, the 2015 AERE Summer Conference, the 2015 Annual Meeting of the Population Association of America, McGill, Carnegie Mellon, and USC for valuable comments and suggestions. Edson Severnini thanks the Energy Policy Institute at the University of Chicago for research support and its generous hospitality during his semester-long visit, when part of this project was executed. Clay and Severnini acknowledge financial support from Heinz College and the Berkman Fund at Carnegie Mellon. Lewis acknowledges financial support from the University of Montreal. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic
Karen Clay, Joshua Lewis, and Edson Severnini
NBER Working Paper No. 21635
October 2015, Revised May 2018
JEL No. I15,I18,N32,N52,Q40,Q53

## ABSTRACT

The 1918 Influenza Pandemic killed millions worldwide and hundreds of thousands in the United States. This paper studies the impact of air pollution on pandemic mortality. The analysis combines a panel dataset on infant and all-age mortality with a novel measure of air pollution based on the burning of coal in a large sample of U.S. cities. We estimate that air pollution contributed significantly to pandemic mortality. Cities that used more coal experienced tens of thousands of excess deaths in 1918 relative to cities that used less coal with similar pre-pandemic socioeconomic conditions and baseline health. Factors related to poverty, public health, and the timing of onset also affected pandemic mortality. The findings support recent medical evidence on the link between air pollution and influenza infection, and suggest that poor air quality was an important cause of mortality during the pandemic.

Karen Clay
Heinz College
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
and NBER
kclay@andrew.cmu.edu

Edson Severnini
Carnegie Mellon University
4800 Forbes Ave #2114B
Pittsburgh, PA 15213
erseverninni@gmail.com

Joshua Lewis
University of Montreal
Département de sciences économiques
3150 rue Jean-Brillant
Montréal, QC, H3T 1N8
joshua.lewis@umontreal.ca

The 1918 Influenza Pandemic was among the worst catastrophes in human history. The virus infected an estimated 500 million people worldwide, one-third of the population. It killed at least 50 million, more than all 20th century wars (Crosby 1989). In the U.S., more than 30 percent of the population was infected, and approximately 675,000 died (Crosby 1989). The 1918 Influenza Pandemic continues to be studied both as an extraordinary historical episode and because of its implications for current policy. As Taubenberger and Morens (2006, p.15) put it: "[u]nderstanding the 1918 pandemic and its implications for future pandemics requires careful experimentation and in-depth historical analysis."

In the U.S., the pandemic spread nationwide during September and October of 1918. There were large regional differences in pandemic mortality, but little consensus has emerged over the underlying causes of these mortality differences. Analysis of mortality rates in Chicago and Hartford shows that mortality rates were related to markers of poverty such as the percent foreign born, illiteracy, and homeownership (Tuckel et al. 2006; Grantz et al. 2016). Other scholars argue that pandemic timing and proximity to World War I military bases influenced severity (Sydenstricker 1918; Barry 2004; Byerly 2010). Bootsma and Ferguson (2007) and Markel et al. (2007) present evidence that public health measures such as school closings, cancelling of public meetings, and quarantines mitigated the effects. Still other researchers argue that pandemic-related mortality was unrelated to socioeconomic conditions or geography (Huntington 1923; Crosby 1989; Brainerd and Siegler 2003).

The possible relationship between air pollution and pandemic mortality has been largely overlooked, despite evidence from human and animal studies that air pollution can increase susceptibility to viral infection and heighten the risk of severe complications, post-infection (Jakab 1993; Jaspers et al 2005). This link could have been especially pronounced during 1918

outbreak, given the devastating impact of the H1N1 virus on lung function (Ireland 1928) and the high levels of air pollution in American cities (Appendix Tables A.1 and A.2).

This paper studies the impact of air pollution on mortality associated with the 1918 pandemic. The analysis draws on a panel of infant and all-age mortality for the period 1915 to 1925 in 180 American cities, representing 60 percent of the urban and 30 percent of the total population. Mortality is linked to a novel measure of city air pollution: coal-fired capacity for electricity generation. Information on electricity plants with at least 5 megawatts of capacity is available in 1915 including location, capacity and type of generation (coal or hydroelectric).

Coal-fired electricity generation was a major source of urban air pollution in the early 20$^{th}$ century. Given historical limitations in electricity transmission, coal-fired plants were typically located near urban areas, producing large volumes of unregulated emissions. A detailed study of Chicago found that in 1912 nearly one half of visible smoke was due to coal-fired electricity generation (Goss 1915). Unlike air pollution from residential coal use, which occurred primarily during the winter months (Barreca, Clay, and Tarr 2016), coal-fired plants produced emissions throughout the fall outbreak of 1918. Coal-fired capacity varied widely across cities, in part, because of differences in the availability of fuel. The empirical analysis is based on a difference-in-differences approach that compares changes in mortality in 1918 in high and medium coal-fired capacity cities to mortality changes in cities with low coal-fired capacity with similar baseline socioeconomic conditions and pre-pandemic mortality rates.

We find that air pollution exacerbated the impact of the 1918 Influenza Pandemic. Cities that used more coal for electricity generation experienced large relative increases in 1918 infant and all-age mortality: infant mortality increased by 11 percent in high coal-capacity cities and 8 percent in medium coal cities relative to low coal-capacity cities; meanwhile, the relative

increases for all-age mortality were 10 and 5 percent in high and medium coal-capacity cities. The estimates imply that pollution in high and medium coal cities was responsible for 30,000 to 42,000 additional deaths during the pandemic, or 19 to 26 percent of total pandemic mortality.

We evaluate alternative determinants of pandemic severity. Guided by the historical literature, we focus on factors related to city poverty, the timing of pandemic onset, and local public interventions. Pandemic mortality was somewhat more elevated in cities with high concentrations of immigrants and poor water quality, consistent with previous research on the relationship between poverty, baseline health, and pandemic severity. The timing of onset was also related to pandemic mortality. Cities hit by earlier outbreaks had particularly high mortality rates, consistent with the virus having weakened over time. We also find suggestive evidence that local interventions mitigated pandemic severity. The relationship between pollution and pandemic mortality is unaffected by the inclusion of controls for these alternative factors.

The 1918 Influenza Pandemic continues to be widely studied because of its relevance to preventing future outbreaks. A large medical literature has sought to understand the particular characteristics of the H1N1 strain responsible for the pandemic (see Taubenberger and Morens 2006, for a discussion). Beginning with Almond (2006), economists have also used the pandemic to examine the long-term outcomes of survivors, although there has been some debate about the size of the effects (see Brown and Thomas 2013; Beach et al. 2017). This paper contributes to this literature by providing evidence on another determinant of pandemic severity, air pollution. Drawing on a new panel dataset on mortality that covers a large sample of American cities, we are also able to evaluate the importance of a number of determinants of pandemic severity that have been previously identified by the historical literature. Given that the risks posed by a severe

influenza pandemic are substantial and unlikely to be met by the existing medical infrastructure, the findings may be relevant to the public health response to future outbreaks.

This paper also contributes to the literature on air pollution and mortality by providing evidence on the interaction between air pollution and infectious disease. A number of studies have shown a causal link between air pollution and mortality (e.g., Chay and Greenstone 2003a 2003b; Currie and Neidell 2005). These studies typically rely on short-term variation in air pollution to identify the health impact. There has been less research on the interaction. A number of epidemiological studies indicate associations between exposure to air pollutants and increased risk for respiratory virus infections (Ciencewicki and Jaspers 2007), although it is unclear whether those correlations have a causal interpretation. Our results demonstrate how exposure to air pollution can exacerbate the mortality effects of severe, if less frequent, health shocks.

## HISTORICAL BACKGROUND

### *The 1918-1919 Influenza Pandemic*

The influenza pandemic of 1918-1919 was brief, but severe. Estimates of worldwide fatalities range from 50-100 million (Crosby 1989; Johnson and Mueller 2002). In the United States, fatalities were between 675,000 and 850,000. In some victims, the virus triggered a 'cytokine storm', an overreaction of the body's immune system that typically led to a rapid deterioration in health. The majority of deaths, however, were caused by a secondary infection, such as bacterial pneumonia, which typically developed in the days and weeks after initial infection (Barry 2004). Figure 1 reports national influenza and pneumonia death rates by month for the 1918-1919 period and the average over the previous five years. Pandemic-related mortality was particularly elevated from October 1918 and January 1919. This four-month period accounted for over 90 percent of pandemic-related deaths.

The pandemic was caused by the H1N1 virus. Unlike the seasonal flu, which is typically caused by slight variations in pre-existing strains, the vast majority of individuals lacked immunity to the virus.[1] Approximately 30 percent of the U.S. population contracted the H1N1 virus in 1918-1919, and fatality rates among those who contracted the virus exceeded 2.5 percent, which is far higher than typical mortality of 0.1 percent (Collins 1931). The Spanish Flu was also characterized by an unusual 'W' age distribution of mortality (see Appendix Figure A.1). The high mortality rates among young adults have been linked to an overreaction of the immune system (Barry 2004). Infant mortality was caused by both postnatal exposure to the virus and higher rates of premature birth (Reid 2005).

The pandemic first appeared in the U.S. as part of a mild influenza outbreak during the spring of 1918. Historians have sought to identify the site of origin of the 1918 Influenza Pandemic. Some accounts suggest that the first human infection occurred in Haskell County, Kansas in late January and early February 1918, and then spread to Europe by American troops (Barry 2004). It is believed that a mutation in the strain during the summer led to a sharp increase in virulence.

The most serious wave of the pandemic originated in Camp Devens near Boston in the first week of September 1918, and then spread rapidly throughout the country. By mid-September, the pandemic had surfaced in most East Coast cities and then moved westward, diffusing nationwide by early October (Sydenstricker 1918).

*Determinants of Pandemic Mortality*

[1] There is some debate among medical historians over whether a previous strain of the virus was also responsible for the 1889-1890 pandemic (Barry 2004).

There were wide cross-city differences in pandemic severity. Pandemic mortality was more than 2.5 times higher in cities at the 90th percentile relative to cities at the 10th percentile.[2] Differences in pandemic mortality were large even among cities within the same state. For example, mortality rates in Gary, Indiana, were more than twice as high as those in Indianapolis. Although researchers have commented on the differences, there is little consensus on the underlying causes (Huntington 1923; Crosby 1989; Kolata 1999; Brainerd and Siegler 2003).

The medical and public health response to the pandemic appears to have been largely ineffective. Antibiotics had not yet been developed, and so could not be used to treat the bacterial pneumonia that often developed, and medicine, more generally, had little to offer beyond palliative care. Municipalities were often slow to adopt preventative measures, which included bans of public gatherings, regulations against spitting in public, and campaigns to the wearing of masks. Most researchers consider these public interventions to have had little effect on pandemic mortality (Brainerd and Siegler 2003; Crosby 1989).[3]

Historians have argued that the timing of local onset, driven in part by the movement of military personnel, influenced pandemic mortality (Barry 2004; Byerly 2010; Sydenstricker 1918; Crosby 1989). Some accounts suggest that the virus weakened substantially in mid-October of 1918, and that cities that experienced later outbreak were exposed to a less virulent strain. Poverty has also been linked to pandemic mortality, both because of higher transmission in poor neighborhoods and lower levels of baseline health capital, although notably no

[2] These calculations are based on the sample of 180 cities used in the analysis.

[3] In contrast, Bootsma and Ferguson (2007) and Markel et al. (2007), find evidence that public health interventions reduced pandemic mortality.

relationship has been found between crowding, measured by population density, and pandemic mortality (Grantz et al. 2016; Tuckel et al. 2006)

*Pandemic Severity, Air Pollution, and Coal-fired Electricity Generation*

Air pollution has received almost no attention from the historical literature on the pandemic, despite emerging evidence that air pollution exacerbates pandemics. In randomized control trials, mice exposed to higher levels of particulate matter (PM) experienced increased mortality when infected with a common strain of the influenza virus (Hahon et al. 1985; Harrod et al. 2003; Lee et al. 2014). Microbiology studies of respiratory cells also identify a link between pollution exposure and respiratory infection. Respiratory cells are the primary site for influenza virus infection and replication, and PM exposure increases the viral-load post-infection (Jaspers et al. 2005). Air pollution has also been shown to increase the severity of bacterial infections in the lungs (Jakab 1993).

The effects of air pollution may have been particularly acute during the 1918 pandemic given the pathology of the H1N1 virus. Pandemic mortality was often caused by a secondary infection, such as bacterial pneumonia. Contemporary researchers noted the impact of the pandemic virus on lungs. As reported in the *Journal of the American Medical Association*, doctors noted that "lung lesions, complex and variable, struck one as being quite different in character to anything one had met with at all commonly in thousands of autopsies one had performed during the last 20 years" (Ireland 1928, p.150).

Although systematic cross-city information on air quality was not available until the mid-1950s, intermittent monitor readings during the early 20th century suggest that air pollution was severe and varied widely across cities. Average levels of total suspended particulates (TSP) air pollution across 15 large American cities were seven times higher than the annual threshold and

twice the maximum daily threshold initially set under the Clean Air Act Amendments of 1970 (Appendix Table A.2). In 1912, the Bureau of Mines reported that 23 of 28 cities with populations over 200,000 were trying to combat smoke (Appendix Table A.1). Dozens of smaller cities also passed ordinances.

Electricity generation was a significant contributor to urban air pollution. In 1912, electricity generating plants accounted for 44 percent of visible smoke in Chicago, while residential coal consumption contributed just 4 percent (Goss 1915).[4] Moreover, coal-fired power plants operated continuously throughout the fall of 1918, whereas residential coal consumption was concentrated in the winter months (Barreca, Clay and Tarr 2016).

There were large differences in the sources of electricity generation based on local availability of inputs. For example, both Grand Rapids and Lansing had similar installed electricity capacity in 1915, although Grand Rapids, which had more abundant sources hydro potential, generated more than twice as much power from hydroelectricity. At the state level, there is a positive relationship between total coal consumption and coal-fired generating capacity, and a negative relationship between total coal consumption and hydro capacity (Appendix Figure A.2), reflecting the fact that coal-fired power was concentrated in the Midwestern states with abundant coal resources.

## DATA CONSTRUCTION AND CITY CHARACTERISTICS BY COAL-FIRED CAPACITY

[4] These estimates were based on Ringelmann Chart measurements, a method used to quantify emission according to the density of observed smoke. Ringelmann measurements were widely used throughout the first half of the 20th century to establish whether emissions exceeded the standards permissible under local ordinances (U.S. Department of the Interior 1967).

To study the impact of air pollution on pandemic severity, we combine information on city coal-fired capacity with a panel dataset on mortality. Infant and all-age deaths were collected from the *Mortality Statistics* for a panel of 180 registration cities for the period 1915-1925.[5] We begin with an initial sample of 283 cities with a population of at least 20,000 in 1921. From this sample, we drop 88 cities with missing information on covariates, and exclude an additional 15 cities located in states that did not use coal for electricity generation, leaving a final sample of 180 cities. Cities are linked to pre-pandemic county-level demographic and economic characteristics drawn from the census of population and census of manufacturing (Haines and ICPSR 2010).

We combine the data on infant and all-age deaths with information on city population and births in 1921 to calculate the infant mortality rates per 1,000 live births in 1921, and the all-age mortality rates per 10,000 city residents in 1921. Infant mortality is widely used in studies of air pollution, since infants are especially vulnerable to environmental exposure and current air pollution concentrations are a better reflection of lifetime exposure (Currie et al. 2014). Contemporary evidence suggests that both infant and all-age deaths were accurately recorded, although there was substantial underregistration of births, particularly among minority populations (Grove 1943). Because underreporting of births may bias estimates in panel regression analyses (Eriksson, Niemesh, and Thomasson 2017), we explore the sensitivity of the main results to several alternative measures of the infant mortality rate.

To construct a proxy for city-level pollution, we digitized information from a 1915 federal report on the location and capacity of coal-fired and hydroelectric power stations with

[5] Price Fishback generously provided these data.

installed capacity of at least 5 megawatts (U.S. Department of Agriculture 1916). These data account for 67 percent of installed coal-fired capacity and 83 percent of installed hydroelectric capacity in 1915.[6] We calculate total coal-fired capacity within a 30-mile radius of each city-centroid, and classify cities into terciles (low, medium, high) of coal-fired capacity. This radius was chosen to capture the fact that the majority of power plant emissions are dispersed locally (Seinfeld and Pandis 2012).[7] To assess whether local coal-fired capacity was related to urban air pollution, we estimate city-level regressions that compare the relationship between coal-fired capacity in 1915 and TSP concentrations in the mid-20th century, controlling for city population, the urban share, and total manufacturing employment.[8] Despite a lag of almost 50 years, there is a clear positive relationship between coal-fired capacity and measured air pollution, and a negative relationship between hydro capacity and TSP concentrations (Appendix Figure A.3).

Coal-fired generating capacity varied widely across cities. Cities with more coal capacity tended to have larger manufacturing sectors, perhaps reflecting higher electricity demand. They

[6] These calculations are derived from interpolated values of total installed generating capacity for electric utilities and industrial generating plants in 1912 and 1917 (Historical Statistics of the United States 1976, p.821).

[7]Recent evidence from Illinois found that over 40 percent of exposure occurred within 30 miles of a power plant (Levy et al. 2002). Historically, air pollution would have been substantially more localized, given the increase in power plant smoke stack heights that has occurred over the past 50 years (Hales 1976, p.10).

[8] Given the paucity of data on early 20th century city air quality, it is impossible to evaluate the contemporaneous link between local coal-fired capacity and air pollution.

also tended to use more coal for residential purposes, reflecting greater local availability (Appendix Table A.3). Despite these differences, almost half of the cross-city variation in coal capacity cannot be explained by standard socioeconomic measures. This idiosyncratic variation in coal capacity across cities will form the basis of our empirical strategy.

Table 1 (col. 1) reports mean characteristics for the sample of 180 cities. The infant mortality rate was 86 per 1,000 live births, and decreased over the sample period (Figure 2a). The all-age mortality rate was 138 per 10,000 residents, and remained roughly stable in non-pandemic years (Figure 2b). During the pandemic year, infant mortality exceeded its trend by 19 percent and all-age mortality exceeded its trend by 35 percent.

Table 1 reports estimated differences in city characteristics for medium coal capacity and high coal capacity relative to low coal capacity cities. We report unadjusted differences in outcomes (cols. 2 and 3) and propensity score adjusted differences (cols. 4 and 5).[9] Over the full period, there is no statistically significant difference in unadjusted or adjusted infant mortality rates by coal capacity. There is a slightly negative relationship between all-age mortality and coal capacity, indicating that healthier workers were somewhat more likely to reside in highly

[9] To construct propensity score adjusted differences we calculate a city-specific propensity score for tercile of coal use based on pre-pandemic socioeconomic conditions (log city population, fraction white, fraction foreign born, fraction urban, log manufacturing employment, log manufacturing payroll per worker, and tercile of residential coal consumption). Propensity scores are estimated using an ordered probit regression to allow for the three bin tercile specification. Adjusted differences in outcomes are then calculated using a linear regression model that controls for the propensity score.

polluted cities, potentially drawn to better labor market opportunities. These differences are eliminated after adjusting for baseline socioeconomic conditions in col. 4 and 5.[10] High and low coal cities are estimated to have had similar trends in mortality both prior to the pandemic and over the entire sample period. Despite these similarities in non-pandemic years, high coal cities experienced a differential rise in mortality in 1918. Cities in the top tercile had excess infant mortality rates that were 10.8 percentage points higher than cities in the bottom tercile, and the gap in excess all-age mortality rates was 6.3 percentage points. We also estimate large differences in coal-fired capacity across the three terciles that are not explained by baseline socioeconomic conditions (Panel B).

There were other differences in socioeconomic characteristics across the three terciles of coal-fired capacity. High coal cities were more populous, had larger manufacturing sectors, had a higher concentration of foreign-born residents, and burned more coal for residential use (Panels C and D). Adjusting for the propensity score largely eliminates these differences. To the extent that high and low coal cities were different in either pre-trends or levels, our empirical controls for both city and year fixed effects, and allows for differential non-pandemic trends in mortality and differential changes in mortality in 1918 according to each observable pre-pandemic characteristic and baseline dependent variables.

[10] These baseline differences suggest that the results are unlikely to be driven by the sorting of less healthy individuals into heavily polluted cities. Nevertheless, to the extent that less healthy individuals sorted into high pollution cities in ways that are not reflected by either pre-pandemic mortality rates or other socioeconomic variables, the estimates might overstate the influence of air pollution on pandemic severity.

## EMPIRICAL FRAMEWORK

To study the effects of air pollution on pandemic mortality, we adopt a difference-in-differences approach that combines the sharp timing of the pandemic with large cross-city differences in coal-fired capacity. The empirical analysis is based on a comparison of average changes in mortality during the pandemic across cities with higher levels of coal-fired capacity relative to changes in mortality in cities with lower levels of coal-fired capacity that had similar pre-pandemic observable characteristics and similar pre-pandemic mortality rates.[11] Formally, outcome $Y_{ct}$ in city *c* and year *t* is regressed on city and year fixed effects ($\mu_c$ and $\lambda_t$), indicators for high coal capacity ($H_c$) and medium coal capacity ($M_c$) that are each interacted with year fixed effects, separate controls for pre-pandemic mortality in 1915 and 1916 ($Y_{c,pre}$) that are each interacted with year fixed effects, pre-pandemic county characteristics ($X_c$) that are each interacted with a linear time trend and an indicator for 1918, and an error term ($\varepsilon_{ct}$):

$$Y_{ct} = \beta_{1t} H_c \cdot \lambda_t + \beta_{2t} M_c \cdot \lambda_t + \gamma_t Y_{c,pre} \cdot \lambda_t + \theta X_c \cdot t + \theta_{18} X_c \cdot I(18) + \mu_c + \lambda_t + \varepsilon_{ct}. \quad (1)$$

The coefficients for coal capacity ($\beta_{1t}$ and $\beta_{2t}$) are allowed to vary in each year. We set 1917 as the reference year. As a result, each coefficient $\beta_{1t}$ ($\beta_{2t}$) captures the differential change in mortality from 1917 to year t in high (medium) coal cities relative to the change in mortality in low coal cities over the same period.

Equation (1) includes controls for baseline mortality in 1915 and 1916 (separately) interacted with year fixed effects. These controls allow for differences in pandemic severity according to baseline population health. The vector $X_c$ includes the baseline demographic and

[11] The empirical strategy is similar to the approach used by Hornbeck (2012).

economic control reported in Table 1, panel C, along with city longitude and latitude. Each variable is interacted with a linear time trend and a dummy variable for 1918.[12] These controls allow for both differential trends in mortality and differential changes in mortality during the pandemic year according to city socioeconomic conditions and geography.

The identification assumption is that the increase in mortality in 1918 would have been similar across the three groups of cities in the absence of coal capacity differences. In practice, this assumption must hold after allowing for differential changes in mortality related to baseline city characteristics and pre-pandemic mortality rates. In the next section we demonstrate the validity of the empirical methodology and assess threats to identification.

Two other estimation details are worth noting. First, the regressions are unweighted. Standard errors are clustered at the city level to adjust for heteroskedasticity and within-city correlation over time.

## RESULTS

### *Infant and All-Age Mortality*

[12] Demographic controls include city population in 1921, and county-level variables for fraction urban, fraction foreign born, fraction nonwhite (all measured in 1910). Economic controls include manufacturing employment in 1910, manufacturing payroll per worker in 1900 (data is unavailable for 1910), and the tercile of residential coal use per capita in 1918 (Lesher 1918). Among other things, these manufacturing covariates control for alternative sources of city air pollution.

To illustrate the empirical approach, Figure 3 graphs estimated $\beta$ s with different sets of controls (see equation (1)).[13] This 'event-study' design compares changes in mortality in each year from 1915 to 1925 relative to the 1917 baseline year. The figure allows us to assess the identification assumption that absent the pandemic, mortality in high and low coal cities would have trended similarly in 1918. The left-hand figures report the coefficient estimates from regression models that include city fixed effects, year fixed effects, and geographic controls that capture the spread of the virus.[14] The right-hand figures report the coefficient estimates from the fully specified regression model reported in equation (1), with additional controls for 1915 and 1916 mortality interacted with year along with the full set of demographic and economic covariates. Panel A reports the results for infant mortality, and panel B for all-age mortality.

In 1918, infant mortality and all age mortality in high-capacity and medium-capacity cities increased relative to low-capacity cities. The rise in 1918 mortality was particularly large in high-capacity cities. The relative increases in mortality were temporary, and in the years following the pandemic, mortality changes were similar across the three groups of cities. In contrast, there is no statistically significant relationship between coal capacity and changes in mortality in non-pandemic years, supporting our identifying assumption that mortality would have trended similarly in 1918 in the absence of the pandemic.

[13] The coefficient estimates with the confidence intervals are reported in Appendix Figures A.5 and A.6.

[14] The geographic controls are city longitude and latitude, each interacted with a linear trend and a dummy variable for 1918.

Table 2, columns 1-3, reports results for infant mortality from estimating equation (1). In column 1, we include city and year fixed effects along with controls for baseline mortality and geography. In column 2, we add controls for baseline city demographic characteristics, and, in column 3, we include the full set of economic controls as described in equation (1). There is a strong relationship between coal capacity and pandemic-related infant mortality that is stable across the different specifications. In 1918, infant mortality increased by 11.0 percent more in high-capacity cities and 7.8 percent more in medium-capacity cities than in low-capacity cities (col. 3).

There were similarly large relative increases in pandemic all-age mortality in high coal cities. Table 2, cols. 4-6, reports the coefficient estimates for the 1918 interaction effect for high and medium coal cities. In 1918, all-age mortality increased by an additional 9.6 percent in high-capacity cities and 5.4 percent in medium-capacity cities as compared to changes in low-capacity cities (col. 6).

The differential increases in mortality in high- and medium-capacity cities during the pandemic year are consistent with the epidemiological and experimental evidence on the role of air pollution in increasing influenza morbidity and mortality. The observed relationships could reflect the effects of air pollution exposure in the months prior to the pandemic, exposure during the pandemic, or some combination of the two.

Because the regression models control flexibly for trends based on pre-pandemic mortality rates, the coefficient estimates capture the impact of coal capacity on pandemic mortality across cities with similar baseline health. In the fully specified model, we also include baseline demographic and economic covariates, each interacted with a time trend and a 1918 dummy to allow for differences in pandemic mortality according to each pre-pandemic factor.

The fact that these covariates have very little impact on the main coefficient estimates strongly suggests that there was an independent relationship between coal capacity and pandemic mortality that was not driven by differences in population characteristics or industrial composition.

To quantify the impact of air pollution on pandemic severity we calculate the number of deaths attributable to coal, based on the coefficient estimates from Table 2 and compare these to the total number of excess deaths in 1918 in the sample population. Table 3 reports the results. The top panel reports the estimates of the total number of excess deaths in 1918 for cities in each of the three terciles of coal capacity (see Appendix B). In total, we calculate that there were 158,000 excess deaths in 1918 in the sample.[15] Given that our sample comprises roughly 30 percent of the U.S. population, these calculations fall within the range of previous estimates of total U.S. pandemic mortality (Crosby 1989).

We evaluate the number of pandemic-related deaths in a counterfactual scenario in which coal capacity in high and medium is reduced to the low capacity level. The calculations are derived based on the coefficient estimates in column 6 of Table 2. We rely on two different approaches to calculate the number of deaths averted. In the first approach, we multiply the total exposed population by the change in mortality probability implied by the regression coefficients. In the second approach, we compare the observed excess 1918 mortality rate to the counterfactual excess mortality rate implied by the regression estimates (see Appendix B for calculations). Both approaches yield large counterfactual reductions in mortality. We calculate that 30,000 to 42,000 total deaths (5,600 to 6,500 infant deaths) would have been averted, a 19 to

[15] Infants accounted for less than 10 percent of total pandemic-related mortality.

26 percent reduction in pandemic mortality. The majority of the deaths averted would have occurred in high-capacity cities. These cities tended to be more populous, so the health impacts of coal were particularly severe.

The economic costs of air pollution during the pandemic were substantial. Applying a $1.1 million (2015 dollars) value of a statistical life in 1920 (Costa and Kahn 2004), we calculate that excess mortality in high and medium coal cities led to a loss of $45.9 billion, equivalent to almost 6 percent of total U.S. GDP in 1918. These losses do not account for the morbidity effects and the losses in worker output in 1918.

*Poverty, Timing of Onset, Local Interventions, and Pandemic Mortality*

Having established a link between coal capacity and pandemic mortality, we now explore other potential determinants of pandemic severity. Table 4 explores the importance of factors related to city poverty and the geographic spread of the pandemic throughout the country.

We assess the impact of various proxies for city poverty on pandemic-related mortality. We include measures of the percent white, percent foreign born, and the typhoid rate in 1900-1905, an indicator for poor quality of drinking water (Beach et al. 2016), all interacted with an indicator for 1918.[16] To separately identify the role of these poverty proxies, these regression models do not include baseline mortality controls. The coefficient estimates reflect the extent to which differences in various measures of socioeconomic conditions were related to pandemic severity.

We find some evidence that city poverty and baseline health conditions were related to pandemic mortality (Table 4, cols. 1 and 2). Higher concentrations of foreign born are associated

[16] Data on city typhoid mortality rates were compiled from Whipple (1908).

with excess all-age mortality, and the fraction white is negatively related to pandemic mortality although the coefficient estimates are not statistically significant. Poor water quality, as proxied by typhoid mortality, is also positively related to all-age pandemic mortality.

Historians have argued that the timing of pandemic onset was related to its severity (Crosby 1989; Sydenstricker 1918). Researchers have claimed that the virus weakened over the course of the fall of 1918, so that locations that experienced a delayed onset were exposed to a less virulent strain. The ability of public officials to respond to the outbreak may also have been related to the timing of local onset.

We assess whether factors related to the timing of onset were related to pandemic mortality. For this analysis, we omit controls for longitude and latitude to separately identify the role of geography. First, we use information on the week of pandemic onset from Sydenstricker (1918). The pandemic first surfaced along the East Coast in early September, and moved westward, diffusing nationwide by mid-October. We construct a dummy variable for 'late' arrival cities that experienced onset after September 27, and interact this variable with an indicator for 1918 to allow for differences in severity based on the time of onset.[17] The results

[17] This specification is chosen to reflect the views of historians that the evolution of weekly pandemic virulence was nonlinear. In particular, estimates of weekly virulence from Baltimore, a city thought to be representative of the nation as a whole, indicate that the virus gained strength through September and then weakened significantly in mid-October (Sydenstricker 1918). In practice, we find qualitatively similar results in regression that use the continuous measure of week of onset.

(reported in col. 3) show that both infant and all-age mortality were significantly lower in late arrival cities, consistent with previous claims about the evolution of the virus.

Next, we assess the role of World War I in influencing local pandemic severity. The movement of military personnel is believed to have been an important determinant of pandemic timing. Crosby (1989), Kolata (2001), Barry (2004), and Byerly (2010) provide detailed accounts of the pandemic in the military and the role of the Navy and Army in its spread. We digitized information on the location of major army training camps in 1918 (U.S. War Department 1919, p.1519), and construct a dummy variable for whether a city was below- or above-median distance from a base. We interact this variable with a 1918 indicator to allow for differences in pandemic severity according to exposure to WWI military bases. The results (col. 4) show that infant mortality was significantly higher in cities near a base. The coefficient estimates for all-age mortality are also positive, albeit smaller in magnitude and statistically insignificant.

Overall, the results in Table 4 support the historical narrative that both urban poverty and factors related to the timing of pandemic onset were related to local severity. Importantly, across all these alternative specifications and different samples, the impact of coal capacity remains stable, suggesting that the main results were not driven by one of these alternative mechanisms.

Some researchers have argued that other local public interventions, such as quarantines and bans on public gatherings, influenced severity (Markel et al. 2007). To assess the role of the local public health effort, we use data from Markel et al. (2007) on local interventions for a sub-sample of 32 cities and construct indicators for early and long-term interventions following their

classification. We interact these indicators with the 1918 dummy, and re-estimate a simplified version of equation (1) for the sub-sample of cities.[18]

The results are reported in Table 5. For comparison, we report the estimates from this modified specification in column 1. Restricting the sample to the 32 quarantine cities, the coefficient estimates for medium capacity are not statistically significant, although the coefficients for high coal-fired capacity remain statistically significant and similar in magnitude to the estimates in Table 2. The coefficient estimates for early and long-term intervention are negative although not statistically significant. Broadly, these findings support the conclusions of Markel et al. (2007) and Bootsma et al. (2007) that local public health initiatives may have played a role in mitigating the effects of the pandemic.

*Robustness Checks*

One potential concern with the previous results is misreporting of the mortality rate. Although infant and all-age deaths are generally thought to have been accurately recorded, underregistration of births could bias the estimates for infant mortality (Eriksson, Niemesh and Thomasson 2017). Moreover, because our measure of the infant mortality rate is constructed based on births in 1921 rather than contemporaneous births, annual fluctuations in fertility could bias the main estimates. We explore the sensitivity of the results to two alternate measures of the infant mortality rate: infant deaths per 1,000 annual births, and infant deaths per 10,000 city residents in 1921. Estimates based on the first measure will not be biased by annual fluctuations

[18] Given the limited sample size, we use a restricted set of covariates for city and year fixed effects, and longitude/latitude and city population (each interacted with a linear time trend and an indicator for 1918).

in fertility, although because a number of cities began collecting information on births midway through the sample period, we must omit roughly one quarter of the sample. Meanwhile, the second measure will not be affected by reporting error due to the underregistration of births.

Table 6 reports the results based on the alternate measures of infant mortality, which are both highly correlated with the original dependent variable. For reference, column 1 reports the baseline results. Column 2 reports the results based on infant deaths per 1,000 annual births. Despite the limited sample size, the estimated effects are similar in magnitude to the original findings. Column 3 reports the results based on infant death per city population in 1921. The coefficient estimates are also very similar to the baseline estimates, providing confidence that the main findings were not influenced by mismeasurement of births.

Table 7 provides a number of robustness checks. For reference, column 1 reports the baseline specification. Columns (2) to (4) explore sensitivity to alternate control variables. In column 2, we add controls for linear state-specific trends to allow for differential trends in mortality across states. The results are not affected by these covariates. In column 3, we replace the control for log population with log population density. This covariate allows for differences in pandemic transmission, for example, due to crowding. Because we lack information on contemporaneous population density, this variable is constructed by dividing city population in 1921 by city area reported in the 1944 City Books (Haines and ICPSR 2010).[19] The coefficient estimates are very similar to the main findings, consistent with epidemiological evidence

[19] The 1920 Census of Population only reported information on population density for cities with at least 100,000 residents. Among cities for which we observe both measures of population density, the correlation between the two variables is 0.89.

showing no association between population density and pandemic mortality (Grantz et al. 2016). In column (4), we explore the sensitivity of the main findings to controls for the population age structure. We add controls for fraction of the population age 18 to 44, who were particularly susceptible to the virus. The main findings are unaffected by this covariate.

Columns (5) to (8) examine the sensitivity of the results to alternate specifications and samples. In column 5, we re-estimate the model based on the mortality rate in levels rather than logs. The estimated effects are statistically significant and economically important. For infant mortality, the coefficient estimates imply increases in pandemic mortality of 15 (=13.0/85.5) percent and 9 (=7.6/85.5) percent in high coal and medium coal cities relative to low coal cities. For all-age mortality, the implied relative increases are 8 (=11.0/138.2) percent and 15 (=21.1/138.2) percent in high and medium coal cities. In column 6, we report the results, dropping cities for which more than one year of mortality data is missing. In column 7, we drop cities in the South. The coefficients on coal-fired capacity remain similar to the baseline values in sign, significance, and magnitude. In column 8, we re-estimate the regressions for cities with at least 50,000 residents in 1921 to examine the sensitivity of the results to outlier mortality rates in smaller cities. The results are not sensitive to this sample restriction.[20]

To conclude the empirical analysis, we provide two additional tests of the research design. First, we explore heterogeneity in the coal-pandemic relationship according to average city wind speed. Intuitively, the local impact of coal consumption should be mitigated by higher wind speeds, which disperse pollutants over a wider region (e.g., Wang and Ogawa 2015). We

[20] Coefficient estimates based on winsorized mortality rates are also similar in magnitude and significance.

assemble information on average annual speed at an elevation of 80 meters from the U.S. Department of Energy (2017), to identify cities above-median and below-median wind, and allow the effects of coal capacity to vary according to this variable.[21] The results show consistently larger mortality effects in low wind cities (col. 9), consistent with higher winds having mitigating the local health impacts of air pollution during the pandemic.

Second, we estimate a set of placebo regressions based on hydroelectric capacity, which generated electricity but was emissions free. In these regressions we interact indicators for medium and high hydroelectric capacity with 1918. The results show no significant relationship between hydro capacity and excess infant or all-age mortality in 1918 (col. 10).[22]

## CONCLUDING REMARKS

The 1918 Influenza Pandemic was an exceptional historical episode, with death rates 5 to 20 times higher than typical pandemics. A century later, the 'Spanish Flu' continues to be an active area of historical analysis, with researchers seeking to understand its origin, the sources of its virulence, and its epidemiological features. Despite ongoing research, basic questions remain about the spread of the virus, and the sources of the stark regional patterns in mortality.

---

[21] We focus on wind speed at an 80m elevation rather than at ground level to capture the dispersion of pollutants from power plant smoke stacks that typically exceeded 50m in the early 20th century (Hales 1976).

[22] In addition to these robustness tests, we have also explored the sensitivity of the main findings to a number of additional controls. The main findings cannot be attributed to differences in the size of the population eligible for military service, differences in city weather conditions during the pandemic, or access to the railway system.

This paper provides new evidence on role of air pollution in exacerbating the pandemic. The effects of air pollution on pandemic mortality were sizeable. Cities with high levels of coal capacity collectively experienced tens of thousands of excess deaths in 1918. Our analysis suggests that pre-pandemic socioeconomic and health conditions also contributed to pandemic severity as did the timing of its spread throughout the country.

Despite improvements in preventative practices and the development of modern antiviral drugs and vaccines, a moderately severe modern pandemic could lead to 2 million excess deaths worldwide (Fan, Jamison, and Summers 2016), and a pandemic virus with similar pathogenicity to the 1918 virus would likely kill more than 100 million (Taubenberger and Morens 2006). A better understanding of the factors that influenced mortality during 1918 Influenza Pandemic may offer critical insights for the mitigation of contemporary pandemics.

Although air quality has improved dramatically over the past 100 years in the United States, urban air pollution remains a major problem in many developing countries. In fact, pollution in cities in India and China is comparable to levels in the U.S. in the early 20$^{th}$ century (Appendix Table A.2). This study's findings thus have particular relevance to the developing world, where air pollution is often severe and where there is limited medical infrastructure. Further research on more recent outbreaks may help shed light on the potential for improved medical treatments and targeted pollution abatement strategies to mitigate the risks posed by a global pandemic.

## REFERENCES

Almond, Douglas. "Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy* 114, no. 4 (2006): 672-712.

Almond, Douglas, Yuyu Chen, Michael Greenstone, and Hongbin Li. “Unintended Consequences of China's Huai River Policy.” *American Economic Review: Papers and Proceedings* 99 (2009): 184-190.

Barreca, Alan, Karen Clay, and Joel Tarr. “Coal, Smoke, and Death: Bituminous Coal and American Home Heating.” NBER Working Paper No. 19881, Cambridge MA, 2016.

Barry, John M. *The Great Influenza: The Epic Story of the Deadliest Plague in History*. New York: Viking Press Books, 2004.

Beach, Brian. Joseph Ferrie, Martin Saavedra, and Werner Troesken. “Typhoid Fever, Water Quality, and Human Capital Formation.” *Journal of Economic History* 76 No. 1 (2016): 41-75.

Beach, Brian, Joseph Ferrie, and Martin Saavedra. “The 1918 Influenza Pandemic and the Fetal Origins Hypothesis: Evidence from Linked Data.” Working Paper, 2017.

Bootsma, Martin, and Neil Ferguson. “The effect of public health measures on the 1918 influenza pandemic in U.S. cities.” *PNAS* 104 no. 18 (2007): 7588-7593.

Brainerd, Elizabeth and Mark V. Siegler. “The Economic Effects of the 1918 Influenza Epidemic.” Centre Econ. Policy Research Discussion Paper No. 3791, Paris, 2003.

Brown, Ryan and Duncan Thomas. “On the Long Term Effects of the 1918 US Influenza Pandemic.” Mimeo, 2016.

Byerly, Carol. “The U.S. Military and the Influenza Pandemic of 1918-1919.” *Public Health Reports* 125 no. 3 (2010): 82-91.

Chay, Kenneth Y. and Michael Greenstone. “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession.” *Quarterly Journal of Economics* 118 (2003a): 1121-1167.

Chay, Kenneth Y. and Michael Greenstone. “Air Quality, Infant Mortality, and the Clean Air Act of 1970.” NBER Working Paper No. 10053, Cambridge MA, 2003b.

Ciencewicki, Jonathan, and Ilona Jaspers. “Air Pollution and Respiratory Viral Infection.” *Inhalation Toxicology: International Forum for Respiratory Research* 19 No. 14 (2007): 1135-1146.

Cohen, Aaron J., H. Ross Anderson, Bart Ostro, Kiran Dev Pandey, Michal Krzyzanowski, Nino Kullnzli, Kersten Gutschmidt, C. Arden Pope III, Isabelle Romieu, Jonathan M. Samet, and Kirk R. Smith. “Chapter 17: Urban Air Pollution” In *Comparative Quantification of Health Risks, Vol. 2*. Geneva: World Health Organization, 2004.

Collins, Selwyn D. “Age and Sex Incidence of Influenza and Pneumonia Morbidity and Mortality in the Epidemic of 1928-29 with Comparative Data for the Epidemic of 1918-19: Based on Surveys of Families in Certain Localities in the United States following the Epidemics.” *Public Health Reports* 46 No. 33 (1930): 1909-1937.

Costa, Dora L., and Matthew E. Kahn. “Changes in the Value of Life, 1940-1980.” *The Journal of Risk and Uncertainty* 29 No. 2 (2004): 159-180.

Crosby, Alfred D. *America's Forgotten Pandemic: The Influenza of 1918*. New York: Cambridge Univ. Press, 1989.

Currie, Janet, Joshua Graff-Zivin, Jamie Mullen, and Matthew Neidell. “What Do We Know About Short and Long Term Effects of Early Life Exposure to Pollution?” *Annual Review of Resource Economics* 6 (2014): 217-247.

Currie, Janet and Matthew Neidell. “Air Pollution and Infant Health: What Can We Learn From California's Recent Experience?” *Quarterly Journal of Economics* 120 (2005): 1003-1030.

Eisenbud, Merril. “Levels of Exposure to Sulfur Oxides and Particulates in New York City and their Sources.” *Bulletin of the New York Academy of Medicine* 54 (1978): 991-1011.

Eriksson, Katherine, Gregory T. Niemesh, and Melissa Thomasson. “Revising Infant Mortality Rates for the Early 20th Century United States.” NBER Working Paper No. 23263, Cambridge MA, 2017.

Fan, Victoria, Dean T. Jamison, and Lawrence H. Summers. “The Inclusive Cost of Pandemic Influenza Risk.” NBER Working Paper No. 22137, Cambridge MA, 2016.

Flagg, Samuel B. *City Smoke Ordinances and Smoke Abatement*. Washington D.C.: Government Printing Office, 1912.

Goss, William F. M. “Smoke Abatement and Electrification of Railway Terminals in Chicago. Report of the Chicago Association of Commerce.” Chicago: Chicago Association of Commerce, Committee of Investigation on Smoke Abatement Industry, 1915.

Grantz, Kyra H., Madhura S. Rane, Henrik Salje, Gregory E. Glass, Stephen E. Schachterle, and Derek A. T. Cummings. “Disparities in Influenza Mortality and Transmission Related to Sociodemographic Factors within Chicago in the Pandemic of 1918.” *PNAS* 113 No. 48 (2016): 13838-44.

Grove, Robert. “Studies in the Completeness of Birth Registration, Part 1. Completeness of Birth Registration in the United States, December 1, 1939 to March 31, 1940.” *Vital Statistics – Special Report Series, National Office of Vital Statistics, U.S. Public Health Service* 17 No. 18 (1943): 223-296.

Hahon, Nicholas, James A. Booth, Francis Green, and Trent R. Lewis. “Influenza virus infection in mice after exposure to coal dust and diesel engine emissions.” *Environmental Research* 37 No. 1 (1985): 44-60.

Haines, Michael R. and Inter-university Consortium for Political and Social Research (ICPSR). *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2010.

Hales, Jeremy M. “Tall Stacks and the Atmospheric Environment.” EPA Publication No. EPA-450/3-76-007, 1976.

Harrod, Kevin S., Richard J. Jaramillo, Cynthia L. Rosenberger, Shan-Ze Wang, Jennifer A. Berger, Jacob D. McDonald, and Matthew D. Reed. “Increased Susceptibility to RSV Infection by Exposure to Inhaled Diesel Engine Emissions.” *American Journal of Respiratory Cell and Molecular Biology* 28 No. 4 (2003): 451-463.

Hornbeck, Richard. “The Enduring Impact of the American Dust Bowl: Short- and Long-run Adjustments to Environmental Catastrophe.” *American Economic Review* 102 No. 4 (2012): 1477-1507.

Huntington, Ellsworth. “Causes of Geographical Variation in the Influenza Epidemic in the Cities of the United States.” *Bulletin Nat. Res. Council* 6 (1923): 1-36.

Ireland, Merritt W (Ed.). “Medical Department of the United States Army in the World War.” *Communicable Diseases* 9: Washington, D.C.: U.S. Army, 1928.

Ives, James E., Rollo H. Britten, David W. Armstrong, Wirt A. Gill, and Frederick H. Goldman. *Atmospheric Pollution of American Cities for the Years 1931 to 1933 with Special Reference to the Solid Constituents of the Pollution*. U.S. Treasury Department, Public Health Bulletin No. 224. Washington: Government Printing Office, 1936.

Jakab, George J. “The Toxicological Interactions Resulting from Inhalation of Carbon Black and Acrolein on Pulmonary Antibacterial and Antiviral Defenses.” *Toxicology and Applied Pharmacology* 121 (1993): 167-175.

Jaspers, Ilona, Jonathan M. Ciencewicki, Wenli Zhang, Luisa E. Brighton, Johnny L. Carson, Melinda A. Beck, and Michael C. Madden. “Diesel Exhaust Enhances Influenza Virus Infections in Respiratory Epithelial Cells.” *Toxicology Sciences* 85 No. 2 (2005): 990-1002.

Johnson, Niall. and Juergen Mueller. “Updating the Accounts: Global Mortality of the 1918-1920 “Spanish” Influenza Pandemic.” *Bulletin of Historical Medicine* 76 (2002): 105-115.

Kolata, Gina. *Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It.* New York: Touchstone, 1999.

Lee, Greg I., Jordy Saravia, Dahui You, Bishwas Shrestha, Sridhar Jaligama, Valerie Y. Hebert, Tammy R. Dugas, and Stephania A. Cormier. “Exposure to combustion generated environmentally persistent free radicals enhances severity of influenza virus infection.” *Particle and fibre toxicity* 11 No. 1 (2014): 57.

Lesher, C.E. “Coal in 1917.” *U.S. Geological Survey. Part B. Mineral Resources of the United States*. Part II: 1908-1956. Washington D.C.: U.S. Government Printing Office, 1918.

Levy, Jonathan I., John D. Spengler, Dennis Hlinka, David Sullivan, and Dennis Moon. “Using CALPUFF to evaluate the impacts of power plant emissions in Illinois: model sensitivity and implications.” *Atmospheric Environment* 36 (2002): 1063-1075.

Markel, Howard, Harvey B. Lipman, J. Alexander Navarro, Alexandra Sloan, Joseph R. Michalsen, Alexandra Minna Stern, Martin S. Cetron. “Nonpharmaceutical Interventions Implemented by US Cities During the 1918-1919 Influenza Pandemic.” *JAMA* 298 No. 6 (2007): 644-654.

Reid, Alice. “The Effects of the 1918-1919 Influenza Pandemic on Infant and Child Health in Derbyshire.” *Medical History* 49 (2005): 29-54.

Seinfeld, John H., and Spyros N. Pandis. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Hoboken, NJ: John Wiley & Sons, 2012.

Stern, Arthur C. “History of Air Pollution Legislation in the United States.” *Journal of the Air Pollution Control Association* 32 No. 1 (1982): 44-61.

Sydenstricker, Edgar. “Preliminary Statistics of the Influenza Epidemic.” *Public Health Reports* 33 (December 1918): 2305-21.

Taubenberger, Jeffery, and David M. Morens. 2006. “1918 Influenza: The Mother of All Pandemics.” *Emerging Infectious Diseases* 12 No. 1 (2006): 15-22.

Tuckel, Peter, Sharon L. Sassler, Richard Maisel, and Andrew Leykam. “The Diffusion of the Influenza Pandemic of 1918 in Hartford, Connecticut.” *Social Science History* 30 No. 2 (2006): 167-196.

U.S. Census Bureau. *Historical Statistics of the United States, Colonial Times to Present*. Washington D.C.: U.S. Census Bureau, 1976.

U.S. Census Bureau. *Fourteenth Census of the United States, Volume XI Mines and Quarries, General Report and Analytical Tables and Selected Industries*. Washington D.C.: U.S. Census Bureau, 1919.

U.S. Department of Agriculture. *Electric Power Development in the United States*. Washington D.C.: U.S. Government Printing Office, 1916.

U.S. Department of Energy. *WINDExchange: Annual Average Wind Speed at 80m [Map]*. Available at: https://windexchange.energy.gov/states. Accessed December 13, 2017.

U.S. Department of the Interior, Bureau of Mines. *Ringelmann Smoke Chart (Revision of IC 7718)*. Washington D.C.: U.S. Bureau of Mines, 1967.

U.S. Geological Survey. *Mineral Resources of the United States. Part II -- Nonmetals*. Washington D.C.: U.S. Government Printing Office, 1917.

U.S. National Office of Vital Statistics. *Vital Statistics of the United States*. Washington D.C.: U.S. Government Printing Office, various years.

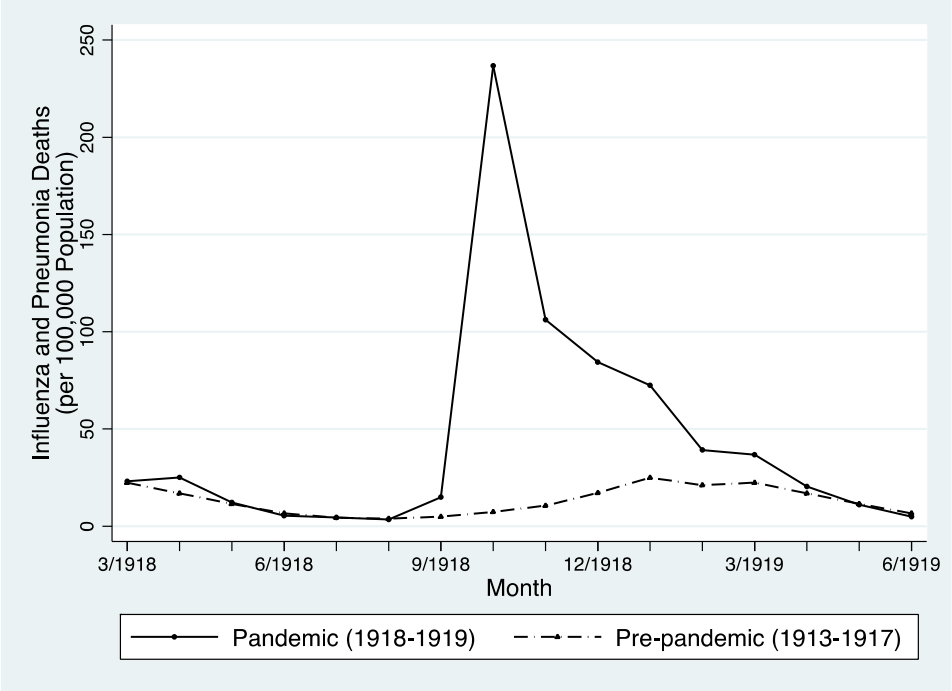
U.S. War Department. *Second report of the Provost Marshall General to the Secretary of War on the operations of the Selective Service System to December 20, 1918*. Washington, D.C.: U.S. Government Printing Office, 1919.

Wang, Jianhua and Susumu Ogawa. "Effects of Meteorological Conditions on PM2.5 Concentration in Nagasaki, Japan." *International Journal of Environmental Research and Public Health* 12 (2015): 9089-9101.

Whipple, George, C. *Typhoid Fever: Its Causation, Transmission, and Prevention*. New York: John Wiley and Sons, 1908.

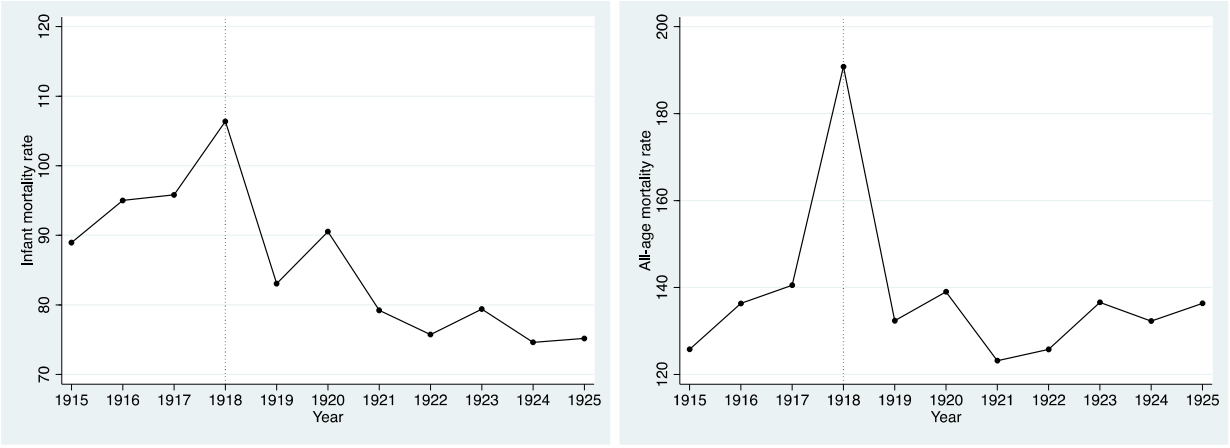
## FIGURES AND TABLES

Figure 1: Influenza and Pneumonia Death Rates, March 1918 - June 1919




*Source:* Mortality Statistics (1913-1919).

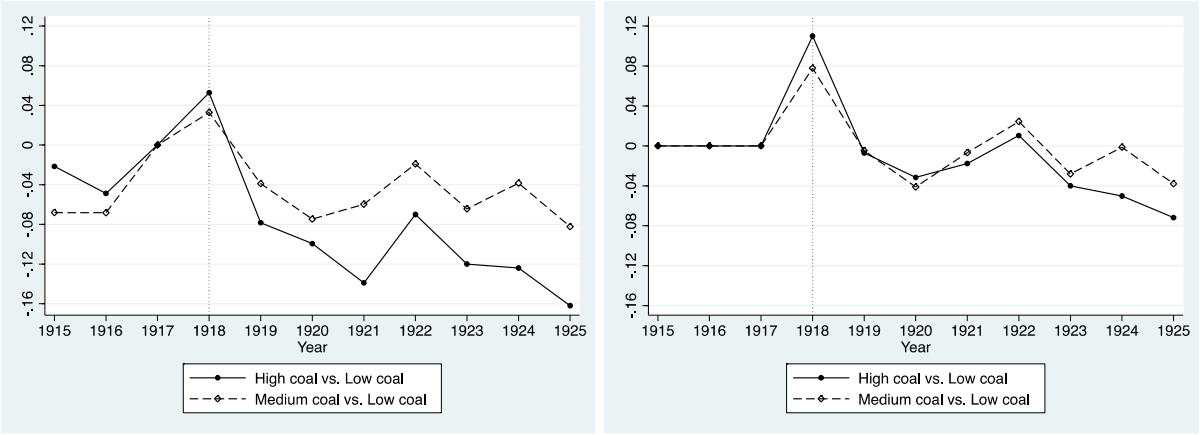
Figure 2: Infant and All-age Mortality, 1915-1925



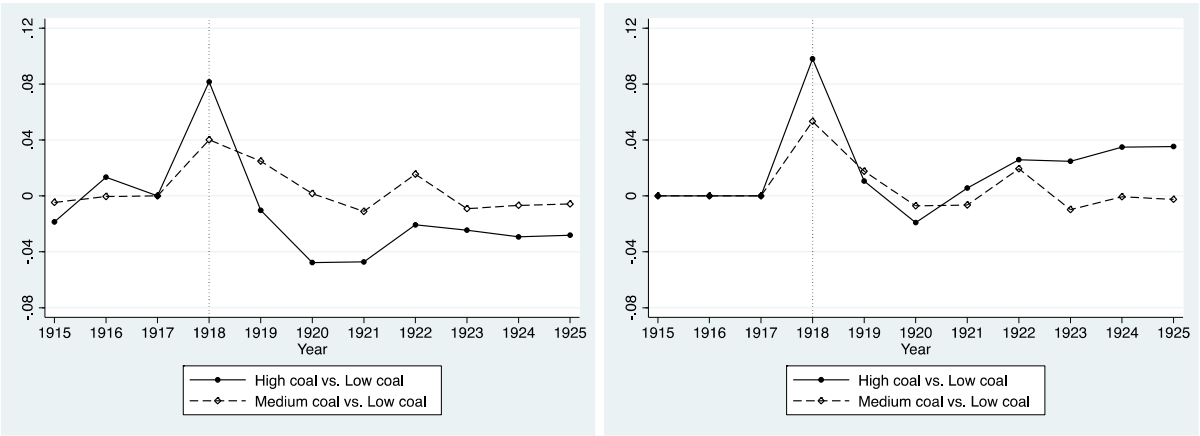

(a) Infant mortality (b) All-age mortality

*Notes:* This figure reports the annual mortality rate for the 180 cities used in the empirical analysis. The infant mortality rate is calculated per 1,000 live births in 1921. The all-age mortality rate is per 10,000 city residents in 1921.

Figure 3: Estimated Differences in Mortality by Coal-fired Capacity, 1915-1925




Panel A: Infant mortality




Panel B: All-age mortality

*Notes:* This figure reports the coefficient estimates from event study regressions based on different versions of equation (1) in the text. The coefficients estimates are relative to the 1917 baseline year. In each year, the solid line represents the differential change in log mortality in high coal relative to low coal cities ($\beta_{1t}$), and the dashed line represents the differential change in medium coal relative to low coal cities ($\beta_{2t}$). The lefthand figures report the coefficient estimates based on regression models that include only city fixed effects, year fixed effects, and geographic controls (longitude and latitude interacted with a linear time trend and an indicator for 1918). The righthand figures report the coefficient estimates based on the full set of controls from equation (1). The corresponding regression estimates with 95% confidence intervals are report in Appendix Figures A.5 and A.6.

# Table 1: Summary Statistics

| | | Differences relative to low coal | | | |
|---|---|---|---|---|---|
| | | Unadjusted | | Propensity Score Adjusted | |
| | Mean (all cities) | Medium coal | High coal | Medium coal | High coal |
| | (1) | (2) | (3) | (4) | (5) |
| *Panel A: Infant and all-age mortality, 1915-1925* | | | | | |
| Infant mortality rate, per 1,000 births | 85.6 | -1.53 | -1.76 | -3.65 | -5.81 |
| | | (2.43) | (3.46) | (2.58) | (3.57) |
| All-age mortality rate, per 10,000 population | 138.2 | -11.20** | -14.39*** | -3.26 | 0.80 |
| | | (4.99) | (4.97) | (4.92) | (5.47) |
| Infant mortality rate, Δ 1917-1915 | 7.3 | 5.19 | 2.61 | 1.60 | -4.31 |
| | | (3.17) | (3.53) | (3.18) | (4.39) |
| All-age mortality rate, Δ 1917-1915 | 14.7 | -0.27 | 0.99 | -3.88 | -5.95 |
| | | (3.14) | (3.69) | (3.23) | (4.32) |
| Infant mortality rate, Δ 1925-1915 | -15.6 | -4.04 | -11.44** | 1.08 | -1.49 |
| | | (4.25) | (5.13) | (4.17) | (5.68) |
| All-age mortality rate, Δ 1925-1915 | 12.1 | -2.69 | -4.72 | 0.51 | 1.49 |
| | | (3.55) | (4.63) | (3.71) | (4.78) |
| Excess infant mortality, in 1918 (%) | 18.5 | 7.45*** | 10.50*** | 7.64** | 10.80*** |
| | | (2.77) | (2.67) | (3.15) | (3.63) |
| Excess all-age mortality, in 1918 (%) | 35.1 | 3.45 | 9.29*** | 1.90 | 6.33** |
| | | (2.15) | (2.40) | (2.29) | (3.03) |
| *Panel B: Coal and hydro capacity in 1915* | | | | | |
| Coal capacity, < 30 miles (MWs) | 182.8 | 50.7*** | 473.3*** | 22.9** | 420.2*** |
| | | (6.0) | (56.3) | (11.6) | (54.7) |
| Hydro capacity, < 30 miles (MWs) | 11.7 | -8.7 | -25.0*** | -14.3 | -35.9** |
| | | (6.8) | (7.6) | (9.9) | (14.1) |
| *Panel C: Baseline population and socioeconomic characteristics in 1910* | | | | | |
| Log population density, in 1921 | 8.59 | 0.053 | 0.397*** | -0.067 | 0.168 |
| | | (0.094) | (0.125) | (0.107) | (0.147) |
| Fraction urban | 0.76 | 0.103*** | 0.199*** | 0.022 | 0.043 |
| | | (0.029) | (0.031) | (0.030) | (0.038) |
| Fraction white | 0.95 | 0.019 | 0.022* | 0.011 | 0.007 |
| | | (0.017) | (0.012) | (0.019) | (0.018) |
| Fraction foreign born | 0.21 | 0.028** | 0.121*** | -0.017 | 0.036** |
| | | (0.014) | (0.015) | (0.013) | (0.016) |
| Fraction workers in manufacturing | 0.34 | 0.054** | 0.087*** | -0.005 | -0.026 |
| | | (0.022) | (0.023) | (0.023) | (0.028) |
| Manufacturing payroll per worker, in 1900 | 5.25 | 0.23 | -0.09 | 0.30 | 0.05 |
| | | (0.46) | (0.46) | (0.47) | (0.494) |
| Residential coal per capita, by tercile 0-2, in 1918 | 1 | 0.58*** | 1.21*** | 0.05 | 0.19 |
| | | (0.13) | (0.13) | (0.11) | (0.14) |
| *Panel D: Other determinants of pandemic severity* | | | | | |
| Typhoid mortality rate per 100,000 population, 1900-1905 | 35.8 | -9.8* | -10.6* | -11.9* | -15.0* |
| | | (5.0) | (6.2) | (6.1) | (8.4) |
| Late pandemic onset | 0.29 | -0.086 | -0.18** | -0.03 | -0.06 |
| | | (0.083) | (0.077) | (0.092) | (0.105) |
| Distance to WWI base | 87.5 | -21.3 | -37.5** | -31.1* | -56.3*** |
| | | (16.6) | (15.1) | (17.3) | (19.7) |

*Notes:* Column 1 reports unweighted average values for the 180 sample cities. Columns 2 and 3 report the difference in each relevant characteristic for medium and high coal cities relative to low coal cities. These estimated differences are obtained from a single regression of the indicated characteristics on a dummy for medium and high coal (low coal is the omitted category) conditional on city longitude and latitude. Columns 4 and 5 report the estimated difference in each characteristic conditional on longitude and latitude and a city-specific propensity score. The propensity score is obtained from an order probit regression model of tercile of coal capacity on baseline city socioeconomic conditions (log population, fraction white, fraction foreign born, fraction urban, log manufacturing employment, log manufacturing payroll per worker, and tercile of residential coal consumption). Robust standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 2: The Effect of the Pandemic on Mortality, by Coal-Fired Capacity

| | Estimated effect | | |
|---|---|---|---|
| **Dependent variable: Log(infant mortality)** | | | |
| | (1) | (2) | (3) |
| I(Year=1918) × | | | |
| Medium vs. low coal capacity | 0.0649** | 0.0823** | 0.0784** |
| | (0.0317) | (0.0330) | (0.0340) |
| High vs. low coal capacity | 0.0827*** | 0.118*** | 0.109** |
| | (0.0317) | (0.0380) | (0.0424) |
| Observations | 1,771 | 1,771 | 1,771 |
| R-squared | 0.827 | 0.831 | 0.834 |
| Cities | 180 | 180 | 180 |
| **Dependent variable: Log(all-age mortality)** | | | |
| | (4) | (5) | (6) |
| I(Year=1918) × | | | |
| Medium vs. low coal capacity | 0.0301 | 0.0509** | 0.0544** |
| | (0.0232) | (0.0255) | (0.0258) |
| High vs. low coal capacity | 0.0671** | 0.0881*** | 0.0964*** |
| | (0.0270) | (0.0327) | (0.0350) |
| Observations | 1,770 | 1,770 | 1,770 |
| R-squared | 0.921 | 0.926 | 0.926 |
| Cities | 180 | 180 | 180 |
| City & year fixed effects | Y | Y | Y |
| 1915 & 1916 mortality × year | Y | Y | Y |
| Geographic controls | Y | Y | Y |
| Demographic controls | | Y | Y |
| Economic controls | | | Y |

*Notes:* Each column reports the coefficient estimates from a different regression from versions of equation (1) in the text. The coefficient estimates represent the interaction effects for medium vs. low coal ($\beta_{2,1918}$) and high vs. low coal ($\beta_{1,1918}$). Geographic controls include city longitude and latitude. Demographic controls include city population, and county-level controls for fraction urban, fraction foreign born, and fraction nonwhite. Economic controls include manufacturing employment, manufacturing payroll per worker, and tercile of residential coal use. Controls are interacted with a linear time and an indicator for 1918. Standard errors are clustered at the city-level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 3: Pandemic-Related Deaths Averted by Reducing Coal-Fired Capacity

| | High coal cities (1) | Medium coal cities (2) | Low coal cities (3) | All cities (4) |
|---|---|---|---|---|
| Excess deaths in 1918 | | | | |
| Total | 106,896 | 32,115 | 19,584 | 158,595 |
| Per 10,000 population | 56.6 | 46.2 | 45.5 | 52.2 |
| Deaths averted (approach 1) | | | | |
| Total | 34,844 | 6,885 | - | 41,729 |
| Per 10,000 population | 18.5 | 9.9 | | 13.7 |
| Deaths averted (approach 2) | | | | |
| Total | 25,195 | 5,138 | - | 30,333 |
| Per 10,000 population | 13.3 | 7.4 | | 10.0 |

*Notes:* Excess deaths in 1918 are calculated as the difference between observed mortality in 1918 and predicted mortality in 1918 based on a linear city-specific trend for the period 1915 to 1925. Estimates for approach 1 are calculated by multiplying the total population by the change in mortality probability implied by the estimated coefficients from Table 2, col. 6. Estimates for approach 2 are calculated by subtracting the coefficient estimates from Table 2, col. 6 from observed excess mortality in 1918 and multiplying by total population. See Appendix B for calculations.

Table 4: Other determinants of pandemic severity

| | Log(infant mortality) | | | | | Log(all-age mortality) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| I(Year=1918) × | | | | | | | | | | |
| Med vs. low capacity | 0.0910** | 0.0923** | 0.0746** | 0.0747** | 0.0825** | 0.0551** | 0.0617** | 0.0513** | 0.0528** | 0.0499* |
| | (0.0397) | (0.0390) | (0.0343) | (0.0338) | (0.0381) | (0.0261) | (0.0270) | (0.0249) | (0.0259) | (0.0258) |
| High vs. low capacity | 0.135*** | 0.128*** | 0.106** | 0.0939** | 0.0968** | 0.110*** | 0.104*** | 0.0933*** | 0.0872** | 0.0779** |
| | (0.0493) | (0.0442) | (0.0429) | (0.0416) | (0.0465) | (0.0352) | (0.0348) | (0.0339) | (0.0356) | (0.0371) |
| Log(baseline typhoid) | 0.0264 | | | | 0.0271 | 0.0495*** | | | | 0.0496*** |
| | (0.0228) | | | | (0.0244) | (0.0184) | | | | (0.0184) |
| Fraction white | | -0.198 | | | 0.152 | | -0.213 | | | -0.0911 |
| | | (0.181) | | | (0.169) | | (0.163) | | | (0.169) |
| Fraction foreign born | | -0.0393 | | | -0.0159 | | 0.352** | | | 0.204 |
| | | (0.167) | | | (0.147) | | (0.154) | | | (0.136) |
| Late pandemic arrival | | | -0.0577** | | -0.00220 | | | -0.0609** | | -0.0420 |
| | | | (0.0263) | | (0.0324) | | | (0.0242) | | (0.0270) |
| Near WWI military base | | | | 0.0595** | 0.0661** | | | | 0.0343 | 0.0438* |
| | | | | (0.0251) | (0.0293) | | | | (0.0240) | (0.0248) |
| Obs | 1,451 | 1,771 | 1,771 | 1,771 | 1,451 | 1,450 | 1,770 | 1,770 | 1,770 | 1,450 |
| R-squared | 0.773 | 0.777 | 0.833 | 0.834 | 0.780 | 0.893 | 0.903 | 0.925 | 0.924 | 0.893 |
| Cities | 142 | 180 | 180 | 180 | 142 | 142 | 180 | 180 | 180 | 142 |
| Full controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

*Notes:* Each column reports the coefficient estimates from a different regression. All models include the full set of controls reported in Table 2, col. 3. The coefficient estimates represent the interaction effects for medium vs. low coal ($\beta_{2,1918}$), high vs. low coal ($\beta_{1,1918}$), and other determinants of pandemic mortality. Baseline typhoid mortality is calculated as the average typhoid mortality rate between 1900 and 1905 from Whippel (1908). Late pandemic arrival is a dummy variable equal to one for cities whose initial recorded onset of the pandemic occurred after September 27, 1918 based on Sydenstricker (1918). Near WWI military base is a dummy variable equal to one for cities that were below median distance from a World War I army training camp based on the U.S. War Department (1919). Standard errors are clustered at the city-level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 5: Local interventions and pandemic severity

| | Log(infant mortality) | | | | Log(all-age mortality) | | | |
|---|---|---|---|---|---|---|---|---|
| | Full sample | Quarantine sample | | | Full sample | Quarantine sample | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| I(Year=1918) × | | | | | | | | |
| Med vs. low capacity | 0.0858*** | -0.0332 | -0.0252 | -0.00649 | 0.0463* | -0.00437 | -0.00176 | 0.0194 |
| | (0.0303) | (0.0447) | (0.0417) | (0.0519) | (0.0260) | (0.0434) | (0.0451) | (0.0455) |
| High vs. low capacity | 0.141*** | 0.0925* | 0.0934* | 0.111* | 0.109*** | 0.149*** | 0.149*** | 0.165*** |
| | (0.0309) | (0.0482) | (0.0473) | (0.0545) | (0.0290) | (0.0439) | (0.0436) | (0.0458) |
| Early intervention | | | -0.0234 | | | | -0.00761 | |
| | | | (0.0346) | | | | (0.0382) | |
| Long intervention | | | | -0.0557 | | | | -0.0626 |
| | | | | (0.0396) | | | | (0.0445) |
| Observations | 1,771 | 340 | 340 | 340 | 1,770 | 340 | 340 | 340 |
| R-squared | 0.755 | 0.874 | 0.874 | 0.882 | 0.893 | 0.926 | 0.926 | 0.926 |
| Cities | 180 | 32 | 32 | 32 | 180 | 32 | 32 | 32 |
| Restricted controls | Y | Y | Y | Y | Y | Y | Y | Y |

*Notes:* Each column reports the coefficient estimates from a different regression. All models are estimated for a restricted set of controls that include city and year fixed effects, and longitude/latitude and city population (each variable interacted with a linear time trend and an indicator for 1918). The coefficient estimates represent the interaction effects for medium vs. low coal ($\beta_{2,1918}$), high vs. low coal ($\beta_{1,1918}$), and characteristics of local nonpharmaceutical interventions for a sample of 32 cities from Market et al (2007). Early and long interventions are classified according to Market et al (2007), where 'early intervention' is a dummy variable for cities that implemented nonpharmaceutical interventions within one week of pandemic onset, and 'long intervention' is a dummy variable for cities that maintained nonpharmaceutical interventions for at least 65 days. Standard errors are clustered at the city-level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 6: Robustness Tests: Alternative Measures of Infant Mortality

| | **Dependent variable: Log(infant mortality)** | | |
|---|---|---|---|
| | Infant deaths per 1,000 births in 1921 (baseline) | Infant deaths per 1,000 annual births | Infant deaths per 10,000 city population in 1921 |
| | (1) | (2) | (3) |
| Correlation with baseline dep. var. | | 0.91 | 0.90 |
| I(Year=1918) × | | | |
| Medium vs. low coal capacity | 0.0784** | 0.0602* | 0.0751** |
| | (0.0340) | (0.0333) | (0.0328) |
| High vs. low coal capacity | 0.109** | 0.0874** | 0.0106** |
| | (0.0424) | (0.0441) | (0.0411) |
| Observations | 1,771 | 1,380 | 1,771 |
| Cities | 180 | 137 | 180 |
| Full controls | Y | Y | Y |

*Notes:* Each column reports the coefficient estimates from a different regression. All models include the full set of controls reported in Table 2, col. 3. The coefficient estimates represent the interaction effects for medium vs. low coal ($\beta_{2,1918}$) and high vs. low coal ($\beta_{1,1918}$). Standard errors are clustered at the city-level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

Table 7: Robustness tests

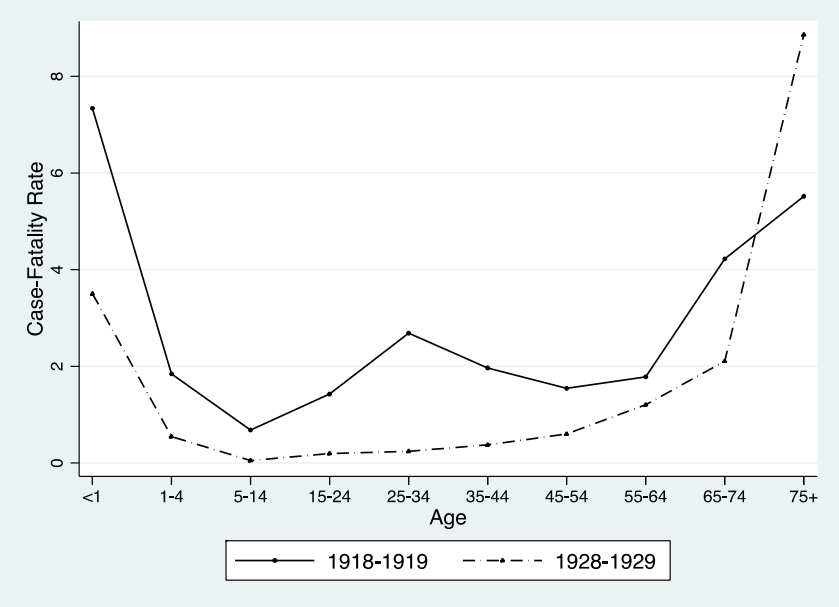
| | Baseline | Alternate controls | | | Alternate specifications and samples | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Linear state trend | City pop. density (1944 city area) | Fraction pop. age 18-44 | Mortality rate in levels | Balanced panel | Drop South | City pop. > 50,000 | Effects by average annual wind speed | Coal vs. hydro capacity |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| **Dep var: Log(infant mortality rate)** | | | | | | | | | | |
| I(Year=1918) × | | | | | | | | | | |
| Med vs. low coal | 0.0784** | 0.0839** | 0.0780** | 0.0733** | 7.62** | 0.0755** | 0.0712* | 0.0632* | | 0.0691** |
| | (0.0340) | (0.0359) | (0.0345) | (0.0331) | (3.40) | (0.0339) | (0.0370) | (0.0340) | | (0.0328) |
| ×I(<Median wind) | | | | | | | | | 0.0826** | |
| | | | | | | | | | (0.0358) | |
| ×I(≥Median wind) | | | | | | | | | 0.0732* | |
| | | | | | | | | | (0.0401) | |
| High vs. low coal | 0.109** | 0.121*** | 0.104** | 0.1102*** | 13.01*** | 0.104** | 0.101** | 0.121*** | | 0.108** |
| | (0.0424) | (0.0459) | (0.0428) | (0.0422) | (4.63) | (0.0423) | (0.0474) | (0.0425) | | (0.0421) |
| ×I(<Median wind) | | | | | | | | | 0.1465** | |
| | | | | | | | | | (0.0642) | |
| ×I(≥Median wind) | | | | | | | | | 0.1003** | |
| | | | | | | | | | (0.0426) | |
| Med. vs. low hydro | | | | | | | | | | -0.0362 |
| | | | | | | | | | | (0.0288) |
| High vs. low hydro | | | | | | | | | | 0.0084 |
| | | | | | | | | | | (0.0296) |
| Observations | 1,771 | 1,771 | 1,771 | 1,771 | 1,771 | 1,675 | 1,588 | 1,094 | 1,771 | 1,771 |
| Cities | 180 | 180 | 180 | 180 | 180 | 155 | 159 | 102 | 180 | 180 |
| **Dep var: Log(all-age mortality rate)** | | | | | | | | | | |
| I(Year=1918) × | | | | | | | | | | |
| Med vs. low coal | 0.0544** | 0.0605** | 0.0523** | 0.0543** | 10.96** | 0.0498* | 0.0384 | 0.0496* | | 0.0455* |
| | (0.0258) | (0.0269) | (0.0261) | (0.0257) | (4.58) | (0.0257) | (0.0273) | (0.0253) | | (0.0253) |
| ×I(<Median wind) | | | | | | | | | 0.0970*** | |
| | | | | | | | | | (0.0293) | |
| ×I(≥Median wind) | | | | | | | | | 0.0137 | |
| | | | | | | | | | (0.0295) | |
| High vs. low coal | 0.0964** | 0.104*** | 0.104*** | 0.0968*** | 21.14** | 0.0913** | 0.0771** | 0.102** | | 0.101*** |
| | (0.0350) | (0.0368) | (0.0347) | (0.0351) | (6.86) | (0.0361) | (0.0378) | (0.0402) | | (0.0343) |
| ×I(<Median wind) | | | | | | | | | 0.2305*** | |
| | | | | | | | | | (0.0328) | |
| ×I(≥Median wind) | | | | | | | | | 0.0688* | |
| | | | | | | | | | (0.0360) | |
| Med. vs. low hydro | | | | | | | | | | -0.0313 |
| | | | | | | | | | | (0.0294) |
| High vs. low hydro | | | | | | | | | | 0.0245 |
| | | | | | | | | | | (0.0249) |
| Observations | 1,770 | 1,770 | 1,770 | 1,770 | 1,770 | 1,674 | 1,587 | 1,093 | 1,770 | 1,770 |
| Cities | 180 | 180 | 180 | 180 | 180 | 155 | 159 | 102 | 180 | 180 |

*Notes:* Each column reports the coefficient estimates from a different regression. All models include the full set of controls reported in Table 2, col. 3. Standard errors are clustered at the city-level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

# APPENDIX

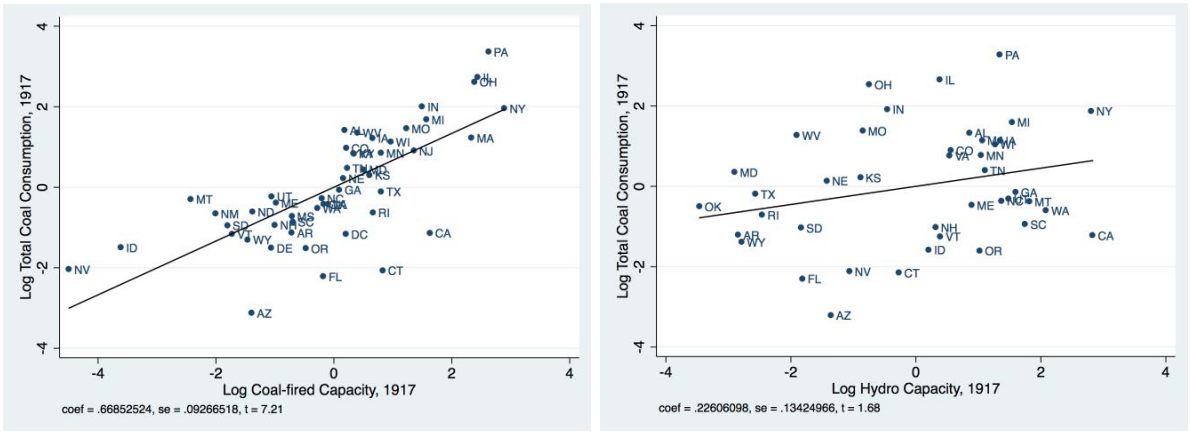
## *A. Figures and Tables*

Figure A.1: Case-fatality Rate from Influenza and Pneumonia by Age




*Notes:* Based on surveys conducted by the United States Public Health Service in 12 localities in 1918-1919 and 14 localities in 1928-1929.
*Source:* Collins (1931).

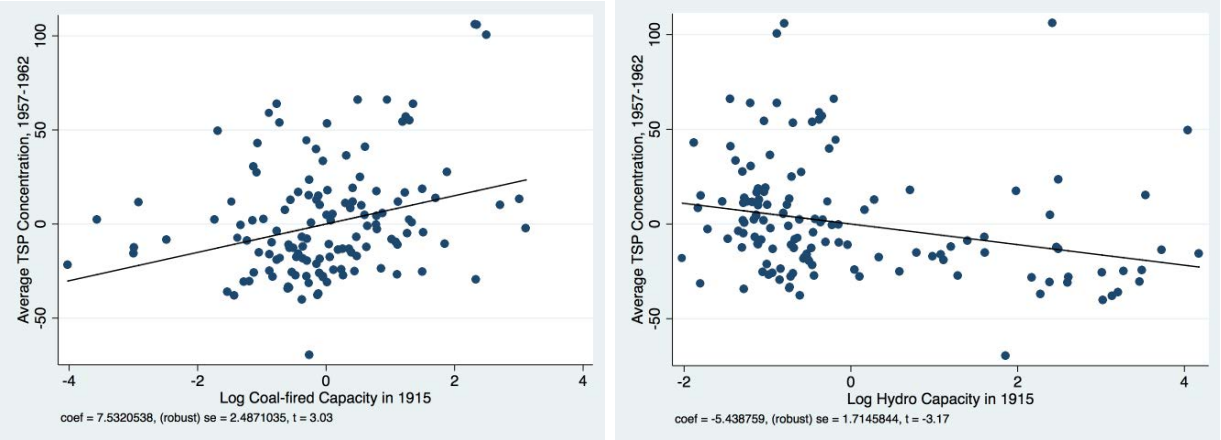
Figure A.2: Electricity Capacity and Total State Coal Consumption in 1917




(a) Coal-fired capacity (b) Hydro capacity

*Notes:* These figures report the relationship between the logarithm of coal-fired capacity, hydro capacity, and total coal consumption at the state level in 1917.
*Source:* Mineral Resources of the United States (1917, p.1254).

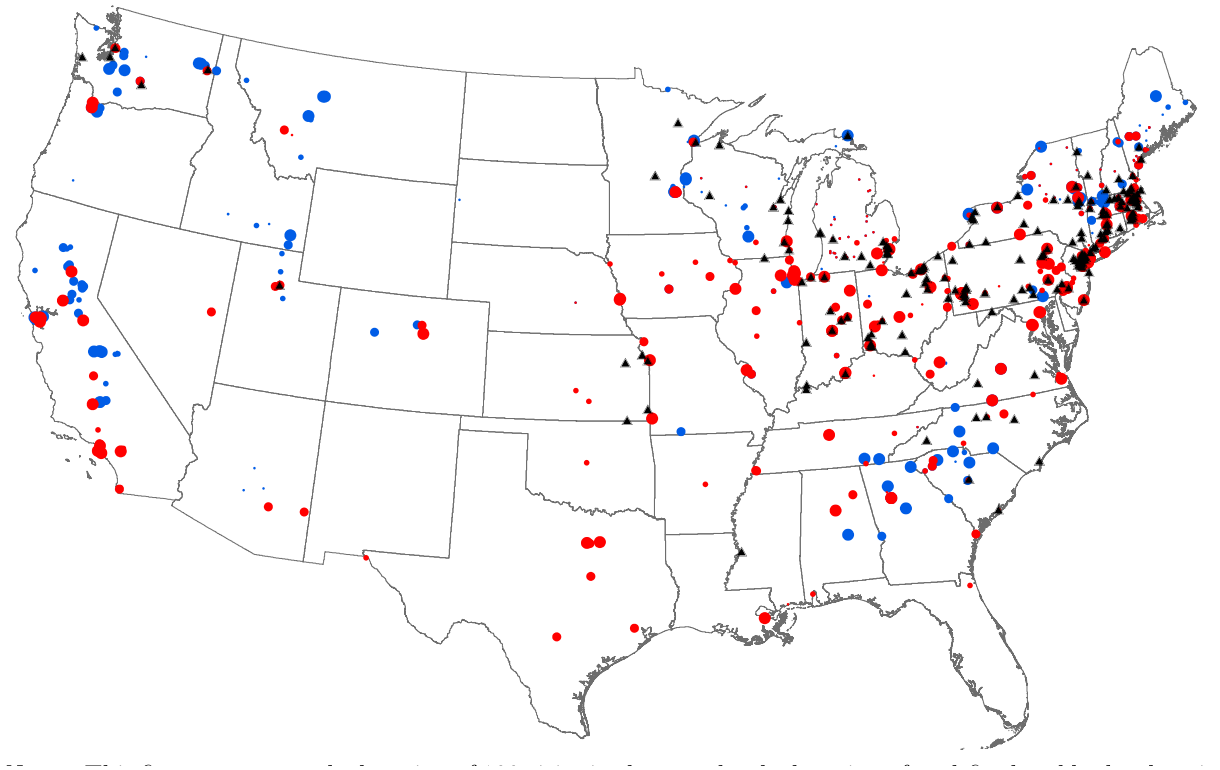
Figure A.3: Electricity Capacity in 1915 and Mid-20th Century Air Pollution




(a) Coal-fired capacity (b) Hydro capacity

*Notes:* These figures report the residual relationship from city-level regressions that relate coal-fired and hydro capacity in 1915 to average TSP concentrations from 1957 to 1962. The regressions control for the logarithm of city population in 1921, manufacturing employment in 1910, and urban share in 1910. Capacity is calculated within a 30-mile radius of each city-centroid. TSP concentrations are calculated as an unbalanced mean of annual concentrations for the years 1957 to 1962.

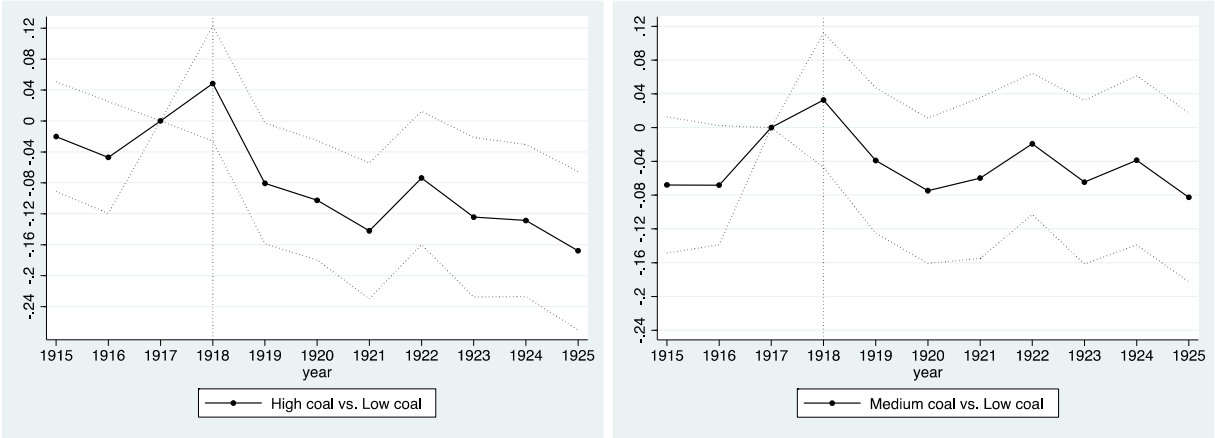
Figure A.4: Cities in Sample and the Location of Coal-Fired and Hydroelectric Power Plants



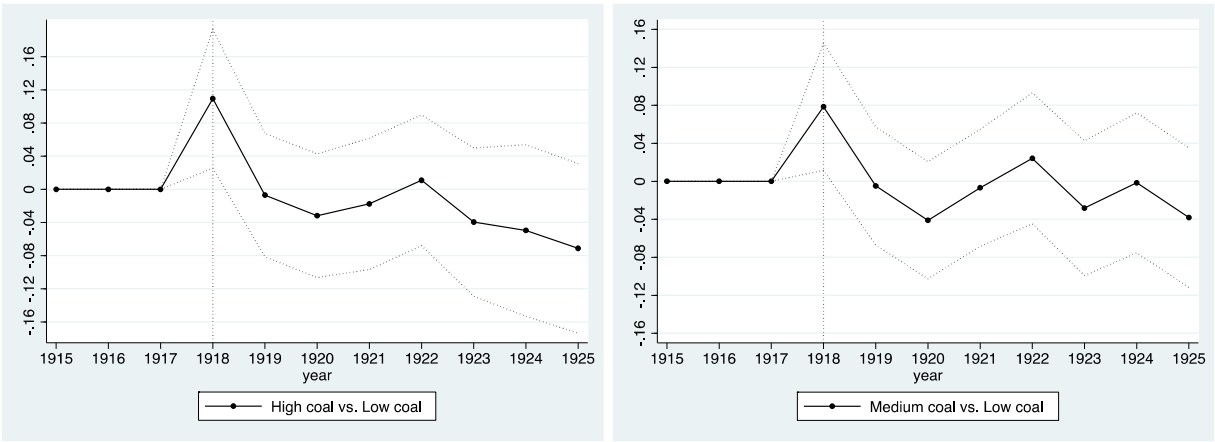
ric power plants by quartile of capacity

(<7mw, 7-11mw, 11-18mw, >18mw).
*Source:* U.S. Department of Agriculture (1916).

Figure A.5: Estimated Differences in Infant Mortality by Coal-fired Capacity, 1915-1925



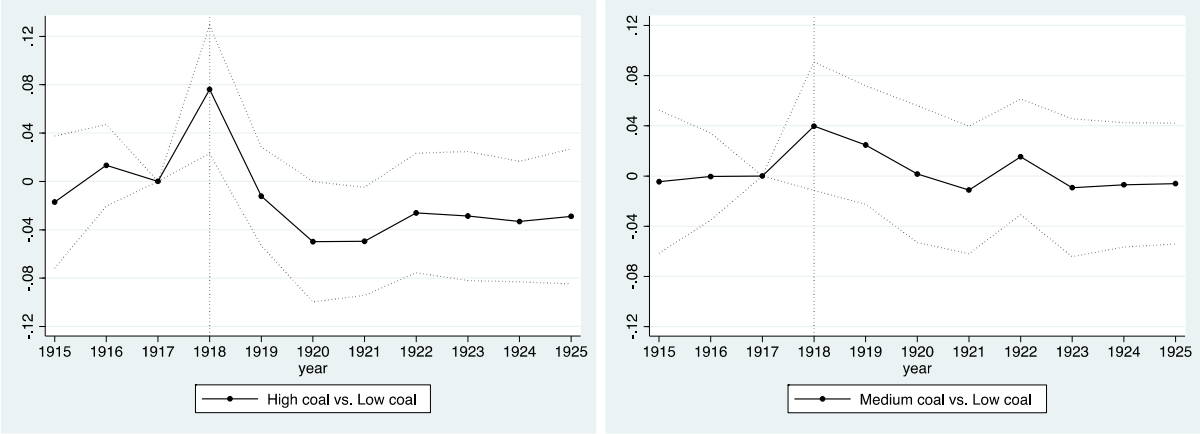

Panel A: Baseline Estimates



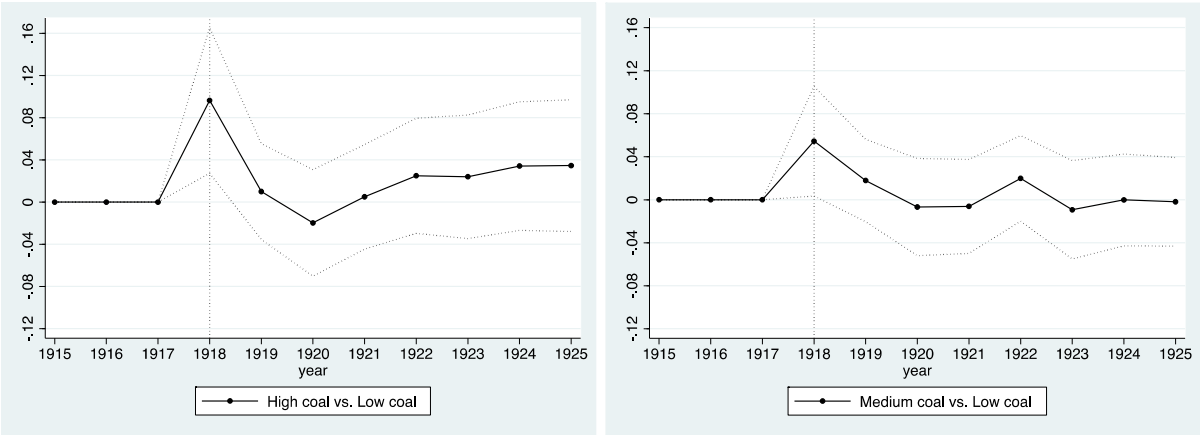

Panel B: Full Controls

*Notes:* This figure reports the coefficient estimates (solid line) and 95% confidence interval (dotted line) from event study regressions based on different versions of equation (1) in the text. The coefficients estimates are relative to the 1917 baseline year. The lefthand figures report the differential change in log mortality in high coal relative to low coal cities ($\beta_{1t}$), and the righthand figures report the differential change in medium coal relative to low coal cities ($\beta_{2t}$). Panel A reports the estimates based on regression models that include only city fixed effects, year fixed effects, and geographic controls (longitude and latitude interacted with a linear time trend and an indicator for 1918). Panel B reports the estimates based on the full set of controls from equation (1).

Figure A.6: Estimated Differences in All-Age Mortality by Coal-fired Capacity, 1915-1925




Panel A: Baseline Estimates




Panel B: Full Controls

*Notes:* This figure reports the coefficient estimates (solid line) and 95% confidence interval (dotted line) from event study regressions based on different versions of equation (1) in the text. The coefficients estimates are relative to the 1917 baseline year. The lefthand figures report the differential change in log mortality in high coal relative to low coal cities ($\beta_{1t}$), and the righthand figures report the differential change in medium coal relative to low coal cities ($\beta_{2t}$). Panel A reports the estimates based on regression models that include only city fixed effects, year fixed effects, and geographic controls (longitude and latitude interacted with a linear time trend and an indicator for 1918). Panel B reports the estimates based on the full set of controls from equation (1).

Table A.1: Cities with Smoke Problems and Municipal Smoke Abatement Legislation

| Year | Cities with Smoke Problems |
|---|---|
| 1912 | *Large Cities with Smoke Problems* |
| | Baltimore, Boston, Buffalo, Chicago, Cincinnati, Cleveland, Denver, Detroit, Indianapolis, Jersey City, Kansas City, Louisville, Milwaukee, Minneapolis, Newark, New York, Philadelphia, Pittsburgh, Providence, Rochester, St. Louis, St. Paul, Washington |
| | *Large Cities without Smoke Problems* |
| | Los Angeles, New Orleans, Portland, San Francisco, Seattle |

| Decade | Cities Passing Smoke Legislation |
|---|---|
| 1880-1890 | Chicago, Cincinnati |
| 1890-1900 | Cleveland, Pittsburgh, St. Paul |
| 1900-1910 | Akron, Baltimore, Boston, Buffalo, Dayton, Detroit, Indianapolis, Los Angeles, Milwaukee, Minneapolis, New York, Newark, Philadelphia, Rochester, St. Louis, Springfield (MA), Syracuse, Washington |
| 1910-1920 | Albany County (NY), Atlanta, Birmingham, Columbus, Denver, Des Moines, Duluth, Flint, Hartford, Jersey City, Kansas City, Louisville, Lowell, Nashville, Portland (OR), Providence, Richmond, Toledo |

*Sources:* Top: Flag (1912); Bottom: Stern (1982, Table III, p.45).

Table A.2: Total Suspended Particulates (TSP) Concentrations in Various Years

| Year | Location | TSP |
|---|---|---|
| 1912-1913 | Chicago | 760 |
| 1931-1933 | Baltimore, Boston, Chicago, Pittsburgh, St. Louis | 630 |
| 1931-1933 | Buffalo, Cleveland, New Orleans, New York, Philadelphia | 520 |
| 1931-1933 | Detroit, Los Angeles, San Francisco, Washington | 350 |
| 1990 | US National Average | 60 |
| 1980-1993 | 58 Chinese Cities | 538 |
| 1999 | Worldwide | 18% urban population $>$ 240 |

*Sources:* Eisenbud (1978); Ives et al (1936); Chay and Greenstone (2003a); Almond et al (2009); Cohen et al (2004).

Table A.3: Correlates of Coal-fired and Hydro Capacity in 1915

| | Dependent Variable: | |
|---|---|---|
| | Log(Coal-fired Capacity) | Log(Hydro Capacity) |
| Log population, in 1921 | -0.174*** | 0.103 |
| | (0.044) | (0.076) |
| Fraction urban | -0.060 | 0.600 |
| | (0.497) | (0.642) |
| Fraction white | -0.663 | 0.437 |
| | (0.419) | (0.824) |
| Fraction foreign born | 2.083*** | 0.642 |
| | (0.667) | (0.898) |
| Log manufacturing | 0.322*** | -0.015 |
| employment | (0.057) | (0.086) |
| Manufacturing payroll | 0.013 | -0.022 |
| per worker, in 1900 | (0.020) | (0.023) |
| Log infant mortality rate, | -0.124 | 0.130 |
| in 1915 | (0.126) | (0.206) |
| Tercile of residential coal | 0.305*** | -0.399*** |
| per capita, in 1918 | (0.094) | (0.124) |
| Observations | 180 | 180 |
| R-squared | 0.57 | 0.13 |

*Notes:* This table reports the coefficient estimates from city-level regressions for coal-fired and hydro capacity in 1915. Unless otherwise indicated, explanatory variables are from 1910. Robust standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

### *B. Calculating Pandemic-Related Deaths Averted by Reducing Coal-Fired Capacity*

This section describes the calculations for the number of pandemic-related deaths averted by reducing coal-fired capacity that are reported in Table 3. Total excess deaths in 1918 are calculated as the difference between observed mortality in 1918 and predicted mortality in 1918 based on a city-specific linear trend regression for the period 1915 to 1925. For example, in high coal cities, all-age mortality exceeded its predicted value in 1918 (138.4 per 10,000 residents) by 40.9 percent. Given a total population of 18.9 million in high coal cities, we estimate that there were (138.4/10,000)*0.409*18,884,435 = 106,896 excess deaths in 1918 that are attributable to the pandemic.

We rely on two different approaches to calculate the number of deaths averted. For the first approach, the estimates are calculated by multiplying the total exposed population by the change in mortality probability implied by the coefficient estimates in Table 2, col. 6. For example, in high coal cities, we calculate the number of deaths averted as follows: $\Delta Deaths_{H,1918} = \beta_1 \cdot \overline{Mortrate}_{H,1918} \cdot Population_H = 0.0964 \cdot \left(\frac{191.4}{10{,}000}\right) \cdot 18{,}884{,}435 = 34{,}844.$

To derive the counterfactual deaths for approach 2 we subtract the coefficient estimates from Table 2, col. 6 from the observed excess mortality in 1918 and then multiplying by the total population. The counterfactual number of pandemic deaths in high coal cities are given by the following expression: $Counterfactualdeaths_{H,1918} = \overline{PredictedMort}_{H,1918} \cdot \left(ExcessMort_{H,1918} - \beta_1\right) \cdot Population_H = \left(\frac{138.4}{10{,}000}\right) \cdot (0.409 - 0.0964) \cdot 18{,}884{,}435 = 81{,}701$, implying that the change in pandemic mortality in high coal cities is $\Delta Deaths_{H,1918} = 106{,}896 - 81{,}701 = 25{,}195.$