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 August 2017

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 would also like to thank the Energy Policy Institute at the University of Chicago (EPIC) for research support and its generous hospitality during his semester-long visit, when part of this project was executed. Karen Clay and Edson Severnini acknowledge financial support from Heinz College and the Berkman Fund at Carnegie Mellon. Joshua 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 August 2017
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 experimental 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

INTRODUCTION

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, and more than the Black Death of 1347-1351 (Crosby 1989). In the U.S., more than 30 percent of the population was infected, and 675,000 died. 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) 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 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 (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.[1] Given historical limitations in electricity transmission, coal-fired plants were typically located near urban areas, producing large volumes of unregulated emissions. Coal-fired capacity varied widely across cities, in part, based on differences in the availability of fuel. The empirical analysis is based on a difference-in-difference 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 estimate that air pollution exacerbated the impact of the 1918 Influenza Pandemic. More polluted cities experienced large relative increases in 1918 infant and all-age mortality:

[1] 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.

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.[2] This paper contributes to this literature by providing evidence on another factor, air pollution. Drawing on a new panel dataset on mortality that covers a large sample of American cities, we are also able to empirically evaluate the importance of a number

[2] 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. There has been some debate about the size of the effects (see Brown and Thomas 2013; Beach et al 2017).

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 (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. Our results demonstrate how exposure to air pollution can exacerbate the mortality effects of severe if less frequent health shocks.[3]

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. Figure 1 reports national influenza and pneumonia death rates by month for the 1918-1919 period and the average over the previous 5 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.

[3] Although a number of epidemiological studies indicate associations between exposure to air pollutants and increased risk for respiratory virus infections – see a review by Ciencewicki and Jaspers (2007) – it is unclear whether those correlations have a causal interpretation.

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.[4] 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 Figure A.1).[5]

The pandemic spread rapidly throughout the country. The most serious wave originated in Camp Devens near Boston in the first week of September 1918.[6] Figure 2 documents the timing of pandemic onset, based on information compiled across 376 localities by Sydenstricker (1918). The pandemic had surfaced in most East Coast cities by mid-September and then moved westward, diffusing nationwide by early October.

Determinants of Pandemic Mortality

[4] 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).

[5] 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).

[6] The pandemic first appeared in the U.S. as part of a mild 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.

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.[7] Differences in pandemic mortality were large even among cities within the same state. For example, mortality rates in Gary 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).[8]

Historians have argued that the timing of local onset influenced pandemic mortality (Sydenstricker 1918; Crosby 1989). Some accounts suggest that cities experiencing later outbreak were exposed to a less virulent strain as the virus weakened over the fall of 1918. Proximity to World War I bases may also have influenced the spread of disease through the movement of soldiers and the high transmission on military bases (Barry 2004; Byerly 2010).

Poverty is another factor that has been linked to pandemic mortality, both because of higher transmission in crowded areas and lower levels of baseline health capital among the poor.

---

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

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

Grantz et al (2016) use detailed spatial and temporal data across census tracts in Chicago to examine the socioeconomic factors that influenced pandemic mortality. They find that markers for poverty were positively related to pandemic severity. Tracts with higher rates of illiteracy and unemployment and lower rates of homeownership experienced significantly higher levels of pandemic mortality. Notably, no relationship has been found between crowding – measured by tract population density – and pandemic mortality.[9]

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

[9] Using tract-level data from Hartford, Tuckel et al (2006) find that pandemic mortality was primarily influenced by the fraction of foreign born residents.

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 (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 (Table A.1). Dozens of smaller cities also passed legislation.

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). 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).

Coal-fired generating capacity varied widely across cities depending on local availability of inputs. For example, both Grand Rapids and Lansing had similar installed electricity capacity in 1915, however, Grand Rapids – which had more abundant sources hydro potential – generated more than twice as much power from hydroelectricity. In general, coal-fired power was concentrated in the Midwestern states with abundant coal resources (see Figures A.2 and A.3).

DATA CONSTRUCTION AND CITY CHARACTERISTICS BY COAL-FIRED CAPACITY

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 mortality rates were

digitized for a panel of 180 American cities for the period 1915-1925.[10] These data come from the *Mortality Statistics*, and cover roughly 30 percent of the U.S. population and 60 percent of the urban population. Infant mortality is calculated per 1,000 live births, and all-age mortality is calculated per 10,000 residents.[11][12]

To construct a measure of city-level pollution, we digitized information from a 1915 federal report on the location and capacity of coal-fired and hydroelectric power stations with installed capacity of at least 5 megawatts were digitized (U.S. Department of Agriculture 1916). 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.[13] Cities are linked to pre-pandemic

[10] Price Fishback generously provided these data. 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 for generation, leaving a final sample of 180 cities.

[11] 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 2015).

[12] Because there is no annual data on births and city population, mortality rates are calculated using the number of births and population in 1921. Similar results were found when mortality rates are constructed using 1910 county-level population.

[13] The 30-mile radius was chosen to capture the fact that the majority of power plant emissions are dispersed locally (Seinfeld and Pandis 2012). Recent evidence from Illinois found that over 40 percent of exposure occurred within 30 miles of a power plant (Levy et al 2002). Historically,

county-level demographic and economic characteristics drawn from the census of population, and census of manufacturing (Haines and ICPSR 2010).

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 3a). The all-age mortality rate was 138 per 10,000 residents, and remained roughly stable in non-pandemic years (Figure 3b). 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 (col. 2) and high coal capacity (col. 3) relative to low coal capacity cities. There is no significant relationship between infant mortality and coal capacity and a slightly negative relationship between all-age mortality and coal capacity, indicating that healthier workers were somewhat more likely to reside in highly polluted cities, potentially drawn to better labor market opportunities. Excess infant and all-age mortality in 1918 were particularly elevated in high pollution cities, consistent with a possible causal impact of air pollution on pandemic severity. There were other differences city characteristics by levels 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). Because these baseline characteristics might have influenced mortality in both pandemic and non-pandemic years, the empirical specification controls for city fixed effects and allows for differential changes over time in mortality according to observable pre-pandemic characteristics and baseline dependent variables.

---

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).

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.[14] Formally, outcome Y_{ct} in city *c* and year *t* is regressed on city and year fixed effects (μ_c and λ_t), indicators for high coal capacity (H_c) and medium coal capacity (M_c) that are each interacted with year fixed effects, pre-pandemic mortality in 1915 and 1916 ($Y_{c,pre}$) 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 (ε_{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 (β_{1t} and β_{2t}) are allowed to vary in each year. The coefficient estimates capture the change in mortality in year t relative to the reference year of 1917 for high and medium coal cities relative to low coal cities.[15]

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

[15] The specification addresses the possibility that a differential rise in pollution in high coal cities led to a rise in mortality that was unrelated to the pandemic. By setting 1917 as the reference year, the analysis controls for differential expansions in wartime production in high coal cities that were already under way in 1917.

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. For example, if less healthy individuals were more likely to reside in heavily polluted cities, and were more vulnerable to the consequences of a negative health shock.

The baseline demographic and economic controls are the variables reported in Table 1, panels C, each of which is interacted with a linear time trend and a dummy variable for 1918.[16] These controls allow for both differential trends in mortality and differential changes in mortality during the pandemic year. We also include controls for city longitude and latitude to allow for differences in pandemic severity based on its geographic spread throughout the country.

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 air pollution differences. In practice, this assumption must hold after controlling 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.

[16] 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.

Two other estimation details are worth noting. First, the regressions are unweighted.[17] Standard errors are clustered at the city level to adjust for heteroskedasticity and within-city correlation over time.

RESULTS

Infant and All-Age Mortality

To illustrate the empirical approach, Figure 4 graphs estimated β s with different sets of controls (see equation (1)). The left-hand figures include city fixed effects, year fixed effects and latitude and longitude, which captures the spread of the virus. The right-hand figures include 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. Table A.3 reports the corresponding coefficient estimates. The observed relationship between coal capacity and changes in mortality is specific to 1918, and there is no statistically significant effect in non-pandemic years. These patterns support the identifying assumption underlying the difference-in-differences strategy that mortality would have changed similarly in 1918 in the absence of the pandemic.

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

[17] In unreported regressions, weighting by population gives qualitatively similar results.

longitude and latitude. In column 2, we add controls for baseline city demographic characteristics, and, in column 3, we include the full set of economic controls. 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 (panel A, col. 3).

There were similarly large relative increases in pandemic all-age mortality in heavily polluted cities. Table 2, columns. 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.

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. Moreover, the fact that the coefficient estimates are not sensitive to baseline demographic and economic controls suggests that the relationship between coal capacity and pandemic mortality 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 excess deaths in 1918 (see Appendix B). We calculate that there were 158,000 excess deaths in 1918.[18] Given that the 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 3 of Table 2. In this scenario, we calculate that 30,000 to 42,000 total deaths (5,600 to 6,500 infant deaths) would have been averted, a 19 to 26 percent reduction in pandemic mortality (see Appendix B for calculations). In comparison, total all-age mortality in New Jersey in 1920 was 41,326. Thus, excess mortality associated with coal-fired capacity in cities in our sample was comparable to all mortality in New Jersey two years later. The majority of the deaths averted would have occurred in high capacity cities. These cities were more populous and the health impacts of air pollution 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, 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

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

Having established a link between pollution 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. Previous research on mortality rates in Chicago and Hartford show a relationship between markers of poverty and pandemic severity (Tuckel et al 2006; Grantz et al 2016). 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.[19] In order 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 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 is also positively related to all-age pandemic mortality.

A number of 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. It has also been suggested that local public health interventions,

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

such as quarantines, mitigated pandemic mortality. 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 pandemic onset influenced 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 (see Figure 2). 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. The results (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.[20]

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 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.[21]

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 coefficient for high capacity remain statistically significant and similar in magnitude to the coefficients 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.

---

[20] The coefficient estimates for coal capacity are also unaffected by controls for the fraction of males aged 15-25 in the population, suggesting that differential selection of young men into military service across cities with higher and lower levels of pollution cannot account for the main results.

[21] 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).

Robustness Checks

Table 6 provides a number of robustness checks. For reference, column 1 reports the baseline specification. 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 explore the sensitivity of the results to the log specification, re-estimating the model based on the mortality rate in levels. 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.

To further assess the validity of the research design, column 4 reports the results from a set of placebo regressions. In these regressions we interact indicators for medium and high hydroelectric capacity with 1918. Hydroelectric capacity serves as a placebo because it generates electricity but is emissions free. The results show no significant relationship between hydro capacity and excess infant or all-age mortality in 1918.

We assess the sensitivity of the results to various sample restrictions. In column 5, we report the results, dropping cities for which more than one year of mortality data is missing. In column 6, we drop cities in the South. The coefficients on coal-fired capacity remain similar to the baseline values in sign, significance, and magnitude.

To conclude the analysis, we assess the robustness of the main findings to controls for city population density. Although we lack information on city size for the main sample, the 1910 Census of Population reports information on city population density for cities with at least 100,000 residents. In the final two columns, we focus on the sample of 71 cities with populations

over 100,000. Column 7 presents the baseline specification for these 71 cities. Despite the restricted sample size, the qualitative results are similar to the main findings. In column 8, we replace the city population control with a control for city population density from the census. The coefficient estimates are very similar across the two specifications.[22]

CONCLUDING REMARKS

The 1918 Influenza Pandemic remains an area of active research, particularly because of its relevance for understanding and mitigating contemporary pandemics. The 'Spanish Flu' was an exceptional episode, with death rates 5 to 20 times higher than typical pandemics. Nevertheless, even 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).

This paper provides new evidence on role of air pollution in exacerbating pandemic mortality. The effects of air pollution were sizeable. Cities with high levels of coal capacity collectively experienced tens of thousands of total deaths during the pandemic. 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.

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 20th century.

[22] In addition to these robustness tests, we have also explored the sensitivity of the main findings to a number of additional controls (results available upon request). 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 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.

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.

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

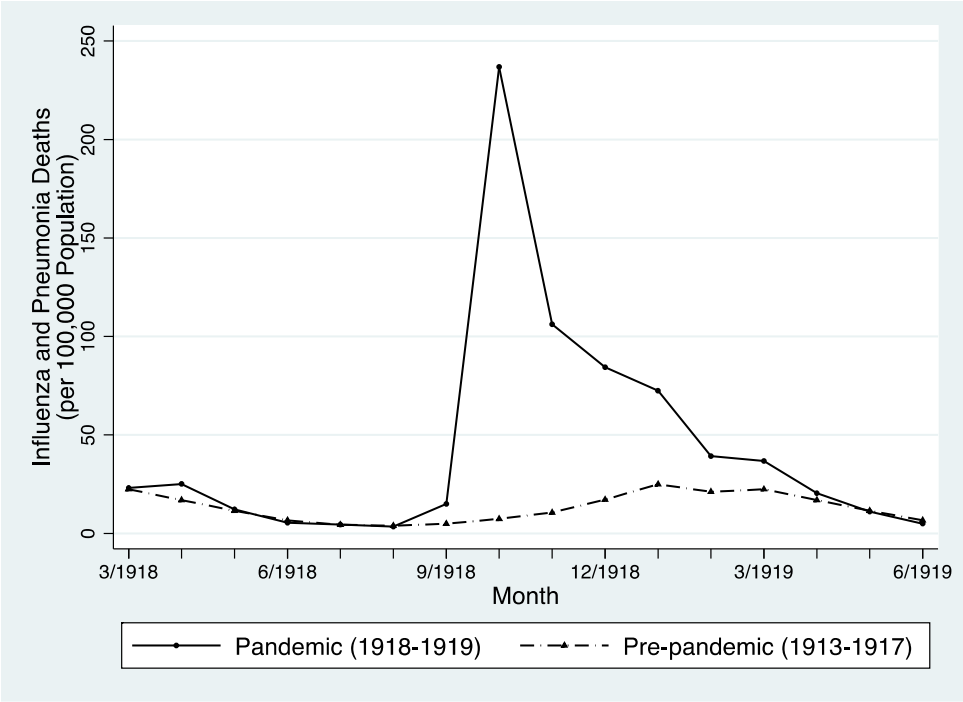
U.S. National Office of Vital Statistics. *Vital Statistics of the United States*. Washington D.C.: U.S. Government Printing Office, various years.

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.

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

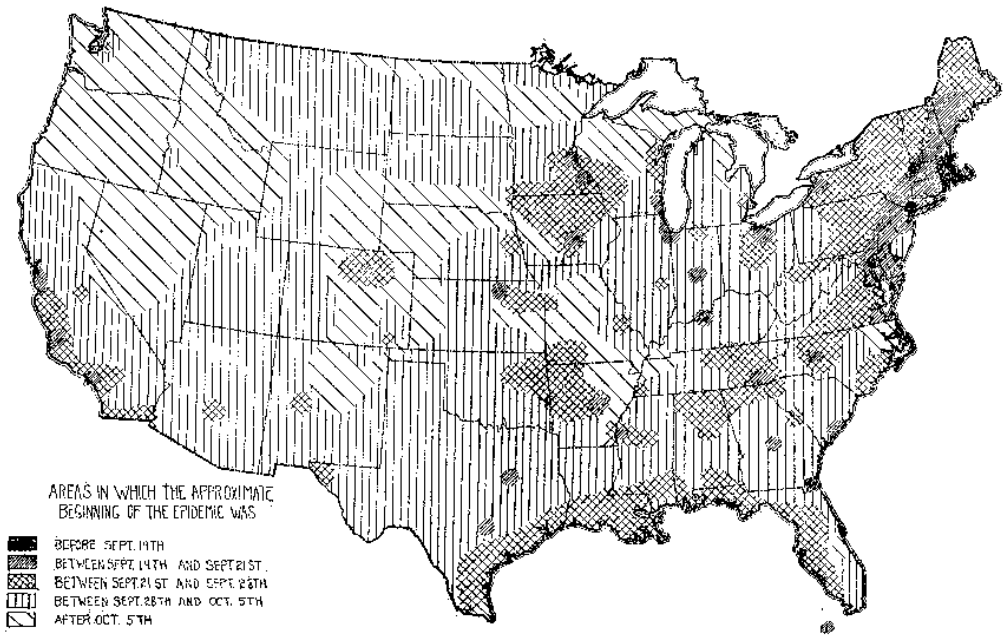
FIGURES AND TABLES

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



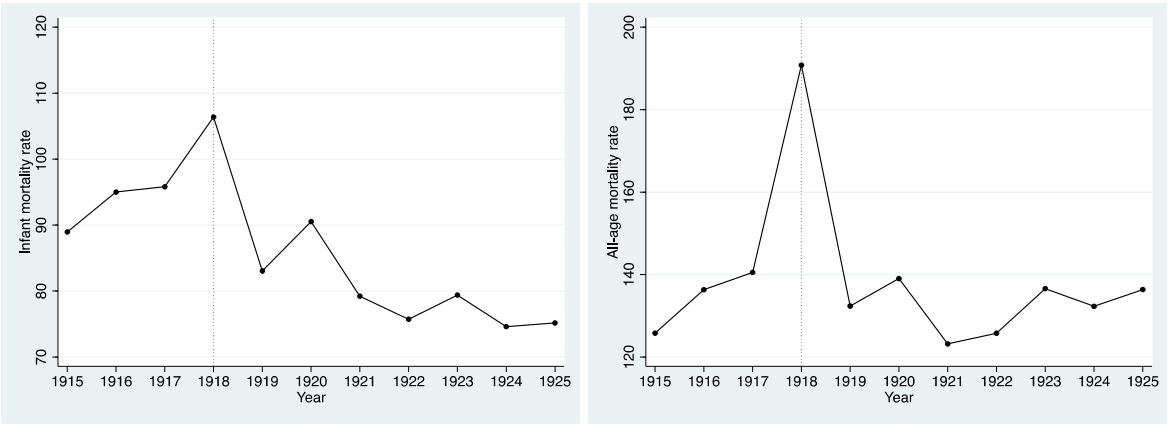
Source: Vital Statistics (1913-1919).

Figure 2: Timing of Pandemic Onset in 376 Localities, September - October 1918



Source: Sydenstricker (1918).

Figure 3: Infant and All-age Mortality, 1915-1925

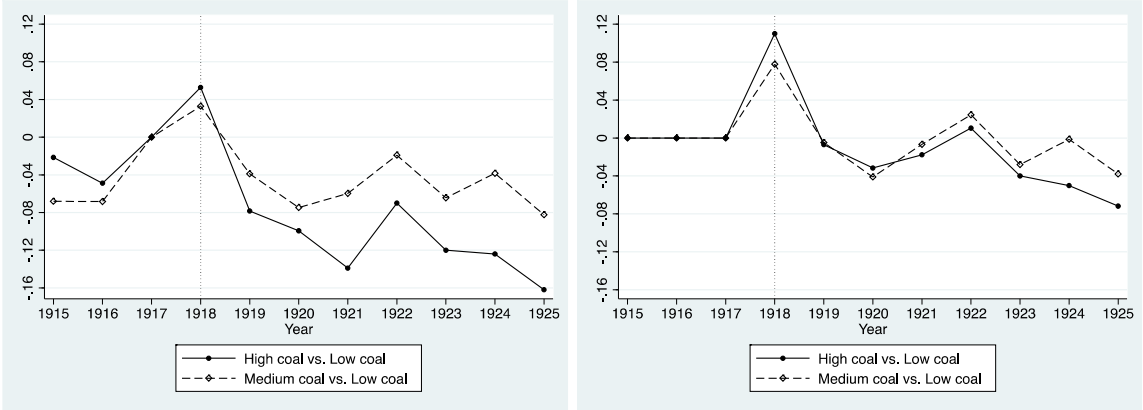


(a) Infant mortality

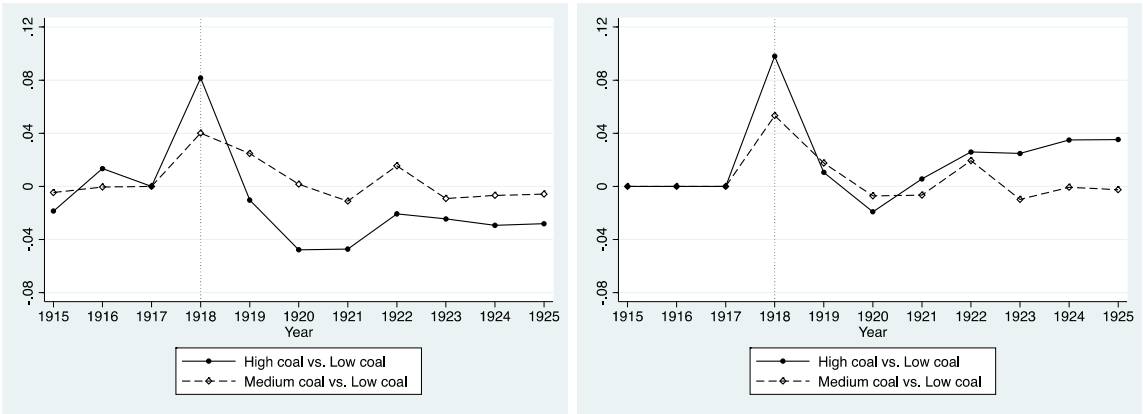
(b) All-age mortality

Notes: The infant mortality rate is per 1,000 live births. The all-age mortality rate is per 10,000 population.
Source: See text.

Figure 4: Estimated Differences in Mortality by Coal-fired Capacity, 1915-1925



Panel A: Infant mortality



Panel B: All-age mortality

Notes: Each panel graphs the estimated coefficients (β_1 and β_2) from versions of equation (1) in the text. The left figures include city fixed effects, year fixed effects, and latitude and longitude controls. The right figures include additional controls for 1915 and 1916 mortality interacted with year along with the full set of demographic and economic covariates described in the text.
Sources: See text.

Table 1: Summary Statistics

	Mean (all cities) (1)	Differences relative to low coal: Medium coal (2)	Differences relative to low coal: High coal (3)	Difference (3)-(2) (4)
Panel A: Infant and All-age Mortality				
Infant mortality rate, per 1,000 births	85.6	-1.53	-1.8	-0.2
		(2.4)	(3.5)	(3.8)
All-age mortality rate, per 10,000 population	138.2	-112.0**	-143.9***	-31.9
		(49.9)	(49.7)	(50.8)
Excess infant mortality in 1918 (%)	18.5	7.5***	10.5***	3.0
		(2.8)	(2.7)	(2.4)
Excess all-age mortality in 1918 (%)	35.1	3.5	9.3***	5.8**
		(2.2)	(2.4)	(2.5)
Panel B: Coal and hydro capacity				
Coal capacity within 30 miles (MWs)	182.8	50.7***	473.3***	422.6***
		(6.0)	(56.3)	(55.6)
Hydro capacity within 30 miles (MWs)	11.7	-8.7	-25.0***	-16.4***
		(6.8)	(7.6)	(3.9)
Panel C: Baseline population and socioeconomic characteristics				
City population (000s)	169.2	59.3***	287.1**	227.7*
		(20.9)	(119.9)	(115.9)
Fraction urban	0.76	0.10***	0.20***	0.10***
		(0.03)	(0.03)	(0.03)
Fraction white	0.95	0.02	0.02*	0.00
		(0.02)	(0.01)	(0.01)
Fraction foreign born	0.21	0.03**	0.12***	0.09***
		(0.01)	(0.01)	(0.02)
Fraction workers in manufacturing	0.34	0.05**	0.09***	0.03
		(0.02)	(0.02)	(0.02)
Manufacturing payroll per worker	5.25	0.23	-0.09	-0.32
		(0.46)	(0.46)	(0.35)
Residential coal per capita, by tercile 0-2	1	0.58***	1.21***	0.63***
		(0.13)	(0.13)	(0.13)
Panel D: Other determinants of pandemic severity				
Baseline typhoid mortality rate, per 100,000 population	35.8	-9.8*	-10.6*	-0.8
		(5.0)	(6.2)	(4.7)
Week of pandemic onset (0-4)	1.63	-0.54***	-1.33***	-0.79***
		(0.15)	(0.15)	(0.15)
Distance to nearest WW1 base (miles)	87.5	-21.3	-37.5**	-16.2
		(16.6)	(15.1)	(10.1)

Notes: Column 1 reports unweighted average values for the 180 sample cities. Columns 2 and 3 report the coefficient estimates from regressions of the relevant characteristic on indicators for medium and high coal capacity (relative to low coal capacity), conditional on city longitude and latitude. Column 4 reports the difference between the coefficient estimates in columns 2 and 3. Robust standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.
Sources: See text.

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
Longitude & latitude 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. 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.
Sources: See text.

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. 3. Estimates for approach 2 are calculated by subtracting the coefficient estimates from Table 2, col. 3 from observed excess mortality in 1918 and multiplying by total population. See Appendix B for calculations.
Sources: See Table 2 and text.

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: All models include the full set of controls reported in Table 2, col. 3. 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.
Sources: See text.

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: All models are estimated for a restricted set of controls that include city and year fixed effects, and longitude/latitude and city population (each interacted with a linear time trend and an indicator for 1918). The quarantine sample is a group of 32 cities for which we have information on local city interventions from Market et al (2007). Early intervention and long intervention are dummy variables based on the classification in Market et al (2007). Standard errors are clustered at the city-level. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.
Sources: See text.

Table 6: Robustness tests

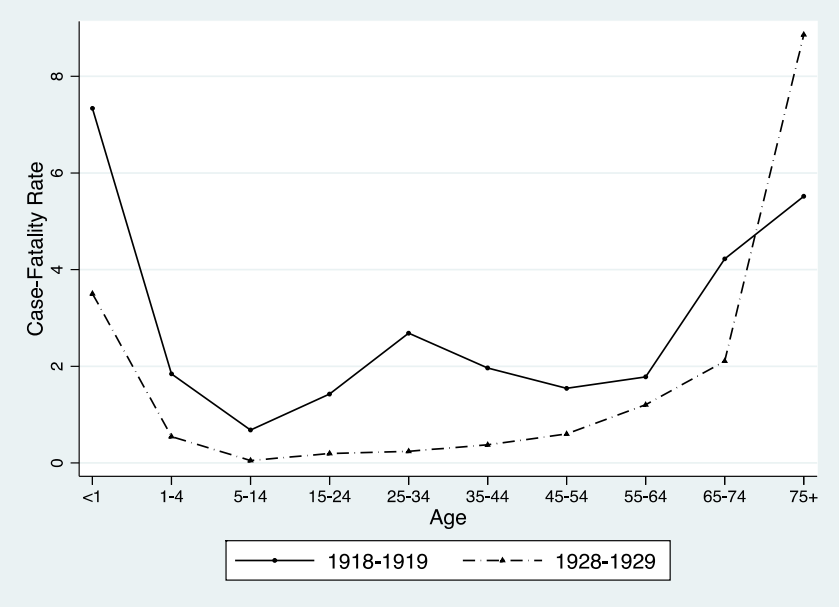
					Alternate samples			
			Alternate dep. var:	Placebo test:			Cities with pop >100,000	
	Baseline	Control for state trend	Mortality rate	Coal vs. hydro capacity	Balanced panel	Drop south		Control for city pop density
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var: Log(infant mortality rate)								
I(Year=1918) ×								
Med vs. low coal	0.0784**	0.0839**	7.615**	0.0691**	0.0755**	0.0712*	0.0173	0.0190
	(0.0340)	(0.0359)	(3.404)	(0.0328)	(0.0339)	(0.0370)	(0.0381)	(0.0371)
High vs. low coal	0.109**	0.121***	13.01***	0.108**	0.104**	0.101**	0.0954**	0.0906**
	(0.0424)	(0.0459)	(4.625)	(0.0421)	(0.0423)	(0.0474)	(0.0450)	(0.0451)
Med vs. low hydro				-0.0362				
				(0.0288)				
High vs. low hydro				0.00835				
				(0.0296)				
Observations	1,771	1,771	1,771	1,771	1,675	1,588	766	766
Cities	180	180	180	180	155	159	71	71
Dep var: Log(all-age mortality rate)								
I(Year=1918) ×								
Med vs. low coal	0.0544**	0.0605**	10.96**	0.0455*	0.0498*	0.0384	0.0261	0.0228
	(0.0258)	(0.0269)	(4.58)	(0.0253)	(0.0257)	(0.0273)	(0.0327)	(0.0333)
High vs. low coal	0.0964***	0.104***	21.14***	0.101***	0.0913**	0.0771**	0.0756	0.0986**
	(0.0350)	(0.0368)	(6.86)	(0.0343)	(0.0361)	(0.0378)	(0.0459)	(0.0438)
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	1,674	1,587	765	765
Cities	180	180	180	180	155	159	71	71

Notes: 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.
Sources: See text.

APPENDIX

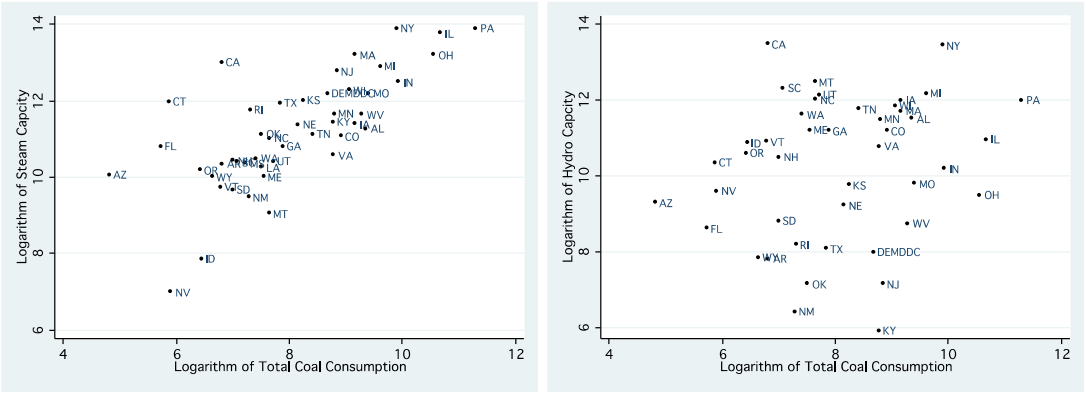
A. Figures and Tables

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



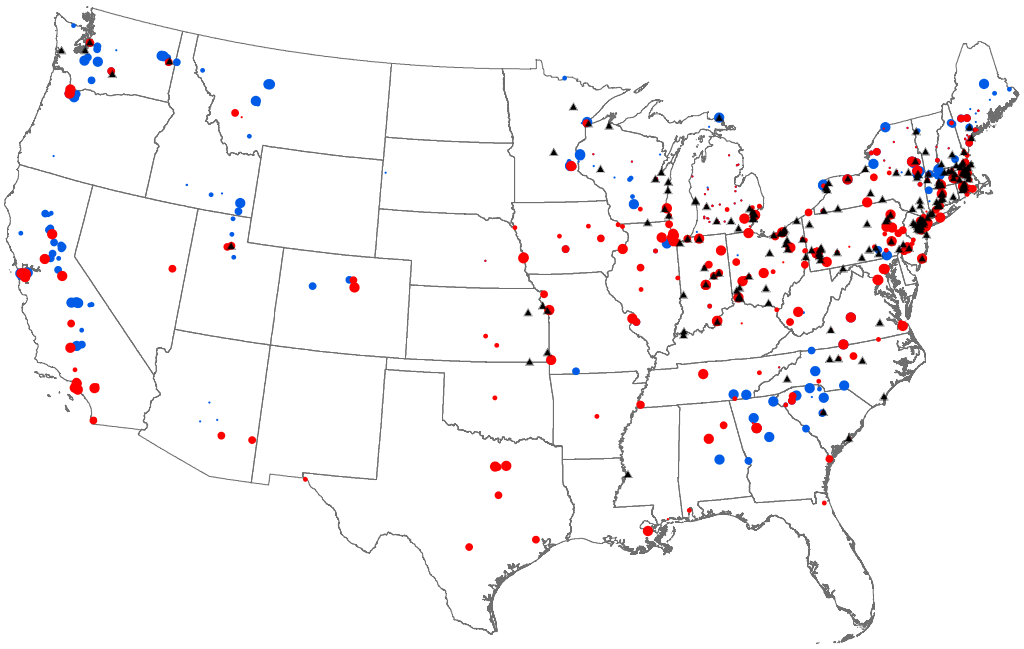
Notes: Based on surveys conducted by the United States Public Health Service in 12 localities in 1918-1919 and 14 localities in 1928-1929.
Source: Collins (1931).

Figure A.2: State-level Coal Consumption, Coal-Fired Capacity, and Hydroelectric Capacity



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

Figure A.3: Cities in Sample and the Location of Coal-Fired and Hydroelectric Power Plants



Notes: This figures presents the location of 180 cities in the sample, the location of coal-fired and hydroelectric power plants in 1915. Red and blue circles denote steam and hydroelectric power plants by quartile of capacity (<7mw, 7-11mw, 11-18mw, >18mw).

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: Estimated Changes in Mortality by Coal-Fired Capacity, 1915-1925

	Dependent Variable:	
	Log(infant mortality)	Log(all-age mortality)
Medium vs. low coal ×		
I(Year=1918)	0.0784**	0.0544**
	(0.0340)	(0.0258)
I(Year=1919)	-0.0049	0.0179
	(0.0315)	(0.0194)
I(Year=1920)	-0.0411	-0.0068
	(0.0313)	(0.0229)
I(Year=1921)	-0.0069	-0.0062
	(0.0312)	(0.0221)
I(Year=1922)	0.0242	0.0198
	(0.0350)	(0.0202)
I(Year=1923)	-0.0283	-0.0093
	(0.0360)	(0.0232)
I(Year=1924)	-0.0016	-0.0002
	(0.0374)	(0.0216)
I(Year=1925)	-0.0382	-0.0019
	(0.0373)	(0.0209)
High vs. low coal ×		
I(Year=1918)	0.109**	0.0964***
	(0.0424)	(0.0350)
I(Year=1919)	-0.0069	0.0100
	(0.0377)	(0.0230)
I(Year=1920)	-0.0318	-0.0197
	(0.0378)	(0.0256)
I(Year=1921)	-0.0175	0.0050
	(0.0402)	(0.0251)
I(Year=1922)	0.0109	0.0250
	(0.0400)	(0.0277)
I(Year=1923)	-0.0395	0.0241
	(0.0453)	(0.0297)
I(Year=1924)	-0.0496	0.0342
	(0.0524)	(0.0309)
I(Year=1925)	-0.0712	0.0346
	(0.0518)	(0.0316)
Observations	1,771	1,770
Cities	180	180
Full controls	Y	Y

Notes: Each column reports the estimated coefficients (β_1 and β_2) from equation (1). The regressions include the full set of controls reported in Table 2, col. 3. Robust standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.
Sources: See Table 2 and text.

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 in calculating 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. 3. 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. 3 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 $\Delta Deaths_{H,1918} = 106,896 - 81,701 = 25,195$.