Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic

Karen Clay
Joshua Lewis
Edson Severnini

October 2015

Pollution, Infectious Disease, and Mortality:
Evidence from the
1918 Spanish Influenza Pandemic

Karen Clay
Carnegie Mellon University and NBER

Joshua Lewis
University of Montreal

Edson Severnini
Carnegie Mellon University and IZA

ABSTRACT

Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic[*]

This paper uses the 1918 influenza pandemic as a natural experiment to examine whether air pollution affects susceptibility to infectious disease. The empirical analysis combines the sharp timing of the pandemic with large cross-city differences in baseline pollution measures based on coal-fired electricity generating capacity for a sample 183 American cities. The findings suggest that air pollution exacerbated the impact of the pandemic. Proximity to World War I military bases and baseline city health conditions also contributed to pandemic severity. The effects of air pollution are quantitatively important. Had coal-fired capacity in above-median cities been reduced to the median level, 3,400-5,860 pandemic-related infant deaths and 15,575-23,686 pandemic-related all-age deaths would have been averted. These results highlight the complementarity between air pollution and infectious disease on health, and suggest that there may be large co-benefits associated with pollution abatement policies.

JEL Classification: N32, N52, I15, I18, Q53, Q56, Q58

Keywords: pollution, infectious disease, mortality, 1918 influenza pandemic

Corresponding author:

Edson Severnini
Heinz College
Carnegie Mellon University
4800 Forbes Ave
Pittsburgh, PA 15213
USA
E-mail: ersevernini@gmail.com

1 Introduction

Together environmental air pollution and infectious diseases accounted for almost 25 percent of all global deaths in 2012 (WHO, 2014). While the health impacts of air pollution and infectious disease are often assumed to be distinct and are typically studied separately, a small emerging literature has begun to examine the extent to which pollution exacerbates infectious disease. The literature includes correlational human studies, animal studies, and cellular-level studies. The interaction of air pollution and infectious disease has significant policy implications for both pandemic prevention and environmental regulation. A key challenge is that unobserved confounding factors such as prior exposure to infectious disease or vaccination make it difficult to identify the causal effect of pollution on infectious disease in human populations.

The 1918-1919 Spanish influenza pandemic presents an exceptional natural experiment that can be used to evaluate health interaction effects. The unanticipated, short-lived, and severe outbreak killed an estimated 675,000 Americans. The rapid spread of the disease limited the scope for viral evolution, and previous exposure to influenza provided no protection against the new H1N1 strain. Thus cohorts exposed to the 1918 pandemic experienced a common health shock. The pandemic also occurred during a period in which there was little scope for medical interventions, and there were wide differences in pandemic severity.

The research design combines the sharp timing of the pandemic with large cross-city differences in baseline pollution levels. The empirical analysis draws on information on infant and all-age mortality for a panel of 183 American cities, representing more than one-half of the urban population and one-quarter of the total population, for the years 1915 to 1925. These data are linked to newly digitized detailed 1915 data on the location, capacity and type of generation for all electricity plants with at least 5 megawatts of capacity. Our preferred measure of exposure is total coal-fired electricity capacity within 30 miles of a city. It is highly correlated with measures of bituminous coal consumption and can be thought of as a proxy for city-level pollution. Contemporaneous historical evidence and air quality

data from the early 1930s suggest that were wide differences in air pollution across cities, in part due to differences in the available inputs for electricity generation.

Based on the historical and medical literatures, two other potential contributors to the severity of the 1918-1919 influenza pandemic were identified: distance to World War I military bases and baseline city-level health conditions. Because of the close living conditions and mobility of troops across bases and into neighboring towns, World War I bases were believed to have contributed to the spread of influenza to nearby areas (Crosby 1989, Kolata 2001, Byerly 2010). City-level health conditions are related to population health capital, which may have also influenced pandemic-severity (Pearl, 1921; Acuna-Soto et al., 2011; Chowell et al., 2008, Bootsma and Ferguson, 2007).

The analysis suggests three main findings. First, coal-fired capacity is positively and statistically significantly related to both infant and all-age mortality in 1918. The effect in 1919 is also positive, albeit smaller in magnitude, and there are no effects in other years. These results are consistent with the timing of the pandemic, which was most severe in the fall of 1918, but continued to affect mortality through the first quarter of 1919. Second, there is no significant relationship between hydroelectric capacity and mortality in 1918 or any other year. Hydroelectric capacity serves as a placebo test, because it also generates electricity but is emission free. Third, both proximity to World War I military bases and baseline health conditions contributed to the severity of the pandemic. Together with capacity, these factors can account for roughly 76 and 44 percent of the cross-city variation in infant and all-age pandemic mortality. To illustrate the magnitude of our main estimates, we compute the effect of moving all cities above the median in coal-fired generating capacity to the median. In this scenario, 3,400-5,860 infant deaths and 15,575-23,686 all-age deaths would have been averted.

This study contributes to the literature on air pollution and mortality by providing some of the first evidence of the health interaction between air pollution and infectious disease in human populations. A number of studies have shown a causal link between air

2

pollution and infant mortality.[1] However, the joint impact of air pollution and infectious disease on mortality is not well understood. Our findings show that exposure to air pollution exacerbated the 1918-1919 influenza pandemic. From a policy perspective, the presence of these complementarities implies that there may be considerable co-benefits of pollution abatement policies that are not accounted for by conventional cost-benefit evaluations.

This paper also adds to the large literature on the 1918-1919 pandemic by investigating the causes of pandemic severity. This pandemic continues to be widely studied because of its relevance for the prevention of future outbreaks. Medical experts wrote in 2006 "The "Spanish" influenza pandemic of 1918-1919, which caused approximately 50 million deaths worldwide, remains an ominous warning to public health. Many questions about its origins, its unusual epidemiologic features, and the basis of its pathogenicity remain unanswered."[2] Medical research has primarily focused on the virus, the immune response, transmission, and understanding the extent to which victims died from the influenza virus or bacterial pneumonia that often followed infection. Beginning with Almond (2006), economists have used the natural experiment provided by the pandemic to examine long-term outcomes of survivors.[3] This paper provides new evidence on the contribution of air pollution, proximity to World War I bases, and city-level health conditions to pandemic severity.

The paper proceeds as follows. Section 2 discusses the history of the 1918 influenza pandemic and air pollution in early 20th century United States. Section 3 describes the data. Section 4 introduces the empirical strategy. Section 5 presents the main findings. Section 6 reports a variety of robustness exercises; and section 7 concludes.

---

[1]See Chay and Greenstone (2003a, 2003b); Currie and Neidell (2005); Currie and Walker (2011); and the studies summarized in Currie et al (2014).

[2]Taubenberger and Morens (2006, p. 69).

[3]See Nelson (2010); Neelsen and Stratmann (2012); and Lin and Liu (2014).

2 Historical Context

This section reviews the historical context for the 1918-1919 influenza pandemic. The first part describes the pandemic. The second part reviews evidence on air pollution in American cities. The third part discusses the direct and indirect effects of air pollution on mortality, and the fourth discusses two other factors that potentially contributed to the severity of the pandemic.

2.1 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 influenza and pneumonia deaths by month for the 1918-1919 period and the corresponding months for the previous 5 years. Pandemic-related mortality was particularly elevated between October 1918 and January 1919. This four-month period accounted for over 90 percent of pandemic-related deaths.

The 1918 pandemic was caused by the H1N1 virus. Unlike the seasonal flu, which is typically caused by slight variations in pre-existing strains, the H1N1 virus had not previously been introduced to the human population, a process known as antigenic shift. As a result, all individuals lacked immunity to the virus. Antigenic shift is typically characterized by very high incidence rates. Approximately 30 percent of the U.S. population contracted the H1N1 virus in 1918-1919 (Collins, 1931).

Case fatality rates in 1918 were greater than 2.5 percent, which is far higher than usual (0.1 percent). Figure 2 reports case fatality rates by age. The figure reveals the well-known 'W' pattern of mortality. Although commentators have focused on the increased death rates for young adults, the rates were highest for infants.

The pandemic spread rapidly throughout the country. The most serious second wave

4

originated in Camp Devens near Boston in the first week of September 1918. Figure 3 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. The rapid spread meant that there was limited evolution of the strain.

Medical and public health interventions were largely ineffective. Antibiotics had not yet been developed and so could not be used to treat the bacterial pneumonia that often developed. Medicine had little to offer beyond palliative care, and hospitals were quickly overwhelmed. Preventative measures – such as bans of public gatherings, regulations against spitting in public, and encouragement to wear masks – were inadequate and adopted too late to have any meaningful effects on transmission rates (Brainerd and Siegler, 2003; Crosby, 1989). One exception is Bootsma and Ferguson (2007), who find that the extent of public health interventions had an impact on total pandemic mortality across 43 U.S. cities. The role of public health interventions is explored further in the empirical analysis.

There were wide cross-state and cross-city differences in pandemic severity. For example, mortality was more than twice as high in Pennsylvania relative to neighboring Ohio. Mortality in Dayton, Ohio was 80 percent higher than in Columbus, Ohio (Huntington 1923, table 7). While researchers have commented on the differences, there is little understanding of the underlying causes (Huntington 1923, Crosby 1989, Kolata 1999, Brainerd and Siegler 2003).

2.2   Air Pollution in Early 20th Century American Cities

Historical evidence suggests that air pollution was a problem and that there were substantial differences in air quality across cities. As smoke became significant, cities often passed legislation aimed at reducing it. In 1912, the Bureau of Mines reported that 23 of 28 cities with populations over 200,000 were trying to combat smoke (Goklany, 1999, p. 15). The top half of Table A.1 lists the 23 cities with smoke problems and the 5 cities that did

not have smoke problems, because they used relatively little coal. The bottom half of Table A.1 reports cities that adopted legislation to combat smoke problems. Dozens of smaller cities also passed legislation.

Newspapers frequently discussed the 'smoke nuisance'. Although systematic cross-city information on pollution levels was not available until the mid-1950s, intermittent monitor readings during the early 20th century suggest the problem was severe and varied widely across cities. The top half of Table A.2 reports TSP concentrations across various cities in the early 20th century. TSP concentrations in these cities were similar to levels in Chinese cities from 1980-1993. These levels are six times higher than the average annual TSP threshold and twice the maximum daily threshold initially set under the Clean Air Act Amendements of 1970.[4] There was considerable variation in air quality across cities. For example, TSP concentrations in Chicago were twice the levels found in Detroit.

Coal consumption for electricity was an important contributor to urban air pollution. Cities near large bituminous coal deposits were heavily dependent on coal-fired capacity, while cities in areas where topographical conditions favored its use were heavily dependent on hydroelectric capacity. The majority of power plants were in urban areas, and emissions were dispersed locally through stacks that were below 75 meters in height on average (Hales, 1976, Figure 4, p.10).[5]

2.3   Air Pollution and Influenza

Air pollution has direct effects on infant mortality and mortality at all ages (Chay and Greenstone, 2003a, 2003b; Currie and Neidell, 2005; Pope et al, 2004; DelFino et al, 2005). For infants, prenatal exposure affects health through the timing of birth and birthweight (Currie and Walker, 2011), and postnatal exposure has been linked to Sudden Infant Death

---

[4]Under the 1970 CAAA, the EPA designated a county as nonattainment if TSP concentrations exceeded either 1) an annual geometric mean concentration of 75 μg/m^3, or 2) the second highest daily concentration exceeded 260 μg/m^3.

[5]Electricity generation accounted for 20 percent of total domestic coal consumption in 1920 (Historical Statistics, 1976, p.591, 824).

Sydrome and to respiratory illness (Woodruff et al, 2008; Arceo-Gomez et al, 2012). Air pollution affects adult mortality primarily through cardiovascular disease and respiratory disease (Hoek et al, 2013).[6]

There is emerging evidence that air pollution can *indirectly* affect health by making individuals more susceptible to respiratory infections. The literature consists of three types of studies i) epidemiological time-series analyses of the correlation between child mortality rates from influenza and pollution levels within a given site (Xu et al, 2013), ii) randomized studies with mice (Harrod et al, 2003; Lee et al, 2014), and iii) microbiology studies of respiratory cells (Jaspers et al, 2005). Respiratory epithelial cells are the primary site for influenza virus infection and replication. Studies suggest that exposure to particulate matter (PM) enhances host susceptibility to influenza infection (Hahon et al, 1985; Harrod et al, 2003) and increases the viral-load post-infection (Jaspers et al, 2005). In experimental models, PM exposure has also been shown to reduce the host response to bacterial infections of the lungs through altered bacterial clearance (Jakab, 1993), an effect that may have been particularly severe during the 1918 epidemic, when mortality was often caused by a secondary infection, such as bacterial pneumonia.[7]

2.4 Other Potential Determinants of Pandemic Severity

Based on the historical and medical literatures, we identified two other potential contributors to the severity of the 1918-1919 influenza pandemic. These factors are distance to the nearest World War I military base and baseline city health conditions. World War

---

[6]In their review and American Heart Association statement, Brook et al (2010) find: "Most, but not all, epidemiological studies corroborate the elevated risk for cardiovascular events associated with exposure to fine PM <2.5 μm in aerodynamic diameter ($PM_{2.5}$). $PM_{2.5}$ generally has been associated with increased risks of myocardial infarction (MI), stroke, arrhythmia, and heart failure exacerbation within hours to days of exposure in susceptible individuals." The mechanism are still being explored, but they summarized the evidence up to that point: "Air pollutants have been linked with endothelial dysfunction and vasoconstriction, increased blood pressure (BP), prothrombotic and coagulant changes, systemic inflammatory and oxidative stress responses, autonomic imbalance and arrhythmias, and the progression of atherosclerosis."

[7]Contemporary researchers noted the devastation to the lungs of influenza victims. At a discussion reported in the *Journal of the American Medical Association*, pathologists noted that "the 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).

I bases were believed to have contributed to the spread of influenza. Crosby (1989) has a very detailed discussion of the spread of influenza by the Navy and Army. The rapidity with which it spread within the military appears to have been because of close living conditions both on the bases and in transit. Because of movements of troops and some civilians along major transportation routes, infected individuals spread it to friends, family, and anyone else they encountered. Kolata (2001) and Byerly (2010) also provide detailed accounts of pandemic in the military.

Previous research has argued that cities with higher baseline health as measured by health and sanitation expenditures and water quality were less affected by the pandemic (Pearl 1921, Acuna-Soto et al, 2011; Chowell et al, 2008; Bootsma and Ferguson, 2007). In principal, the relationship between baseline local health conditions and pandemic severity is ambiguous. For example, high pre-pandemic mortality rates could indicate that a population was particularly vulnerable to the effects of a negative health shock. On the other hand, low levels of baseline health could mollify the impact of the pandemic through culling of the weakest.[8]

3 Data

To examine the interaction effect between pollution and influenza on health, information on infant and all-age mortality was combined with newly digitized data on electricity generation and additional census data on city characteristics. Infant and all-age mortality rates were digitized for a panel of 183 American cities for the period 1915-1925 (excluding the year 1920, a year for which no data were available).[9] This data comes from the *Mortality Statistics*, and covers over one quarter of the U.S. population and over half of the urban

---

[8]See Mamelund (2006) for a discussion.

[9]Price Fishback generously provided this data. Infant mortality is defined as the number of infant deaths per 1,000 live births, all-age mortality is the number of total deaths (including infant deaths) per 10,000 population. Because we lack annual city-level information on births and population, the rates are calculated by dividing annual deaths by total births and population in 1921, the first year that this data was reported. Similar unreported results were found when mortality rates were constructed using 1910 county-level population.

population.

These health outcomes are linked to a measure of city-level pollution. Data 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). Using GIS software, these data were combined with information on city locations to construct a measure of local exposure to pollution from coal-fired electricity generation. Our preferred measures are total coal-fired capacity and hydroelectric capacity within 30 miles of each city-centroid. This distance was chosen to capture the fact that the majority of power plant emissions are dispersed locally.[10] The sensitivity of the results to this particular cutoff is explored in the empirical analysis.

The effects of coal-fired capacity can be contrasted with hydroelectric capacity to disentangle the impact of pollution from the broader effects of local electricity generation. Figure 4 reports the relationship between total state level coal consumption and coal-fired and hydroelectric capacity in 1917. There is a strong positive relationship between coal-fired capacity and total coal consumption at the state-level, but no link between hydroelectric capacity and overall coal use. Figure 5 displays the sample of cities overlaid with a map of hydroelectric and coal-fired power plants, and coal deposits. The geographic pattern in the use of coal and hydropower reflects the availability of inputs. Coal-fired power was concentrated in the Midwestern states with abundant coal resources. Pennsylvania, Ohio, and Indiana were almost entirely reliant on coal for electricity. Hydroelectricity was prevalent where topographical conditions favored its use.

These data are linked to county-level demographic and economic characteristics drawn from the census of population, and census of manufacturing (Haines and ICPSR, 2010). Demographic controls include total population, urban population, and the share of white residents in 1910. Economic covariates include employment in manufacturing in 1910, and

---

[10]For a review of the mechanics of airborne pollutant transport, see Seinfeld and Pandis (2012). Recent evidence from Illinois found that over 40 percent of $PM_{2.5}$ exposure occurs 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, Figure 4, p.10).

manufacturing payroll per worker in 1900.[11] Additional controls include county-centroid longitude and latitude.

We also include information on the two other contributors to pandemic severity: proximity to a World War I military base and city-level health conditions. We digitized information on the locations of all major army training camps in 1918 (U.S. War Department, 1919, p.1519), and calculated the city-centroid distance to the nearest base. We also include three measures of city-level health conditions: spending on health and sanitation; typhoid mortality, which is a marker for water quality; and pre-pandemic infant mortality. The annual city spending on health and sanitation are from Miller (2008). Typhoid mortality is measured as the average annual deaths per 100,000 population for the period 1900 to 1905, compiled from Whipple (1908).

Table 1 reports mean characteristics (column 1) and estimated differences between cities above and below median coal-fired capacity (column 2). Panel A reports these summary statistics for mortality. Infant mortality is not statistically different across the two groups of cities, and all-age mortality is slightly lower in above-median cities. Rows 3 and 4 report the logarithm of excess infant and all-age mortality in 1918. This variable is calculated as the residual mortality rate after controlling for a city-specific linear trend. In 1918, infant mortality exceeded its trend by 20 percent and all-age mortality exceeded its trend by 36 percent. Figures 6 presents the city-level distributions of mortality by year. The deviation from trend is clearly evident in the rightward shift of the two mortality distributions in 1918. The pandemic was particularly severe in above-median coal-capacity cities. Pandemic-related mortality rates were 6 to 7 percent higher in above-median coal capacity cities. Figures 7 plots residual infant and all-age mortality in 1918 for cities above and below median coal-fired capacity. In both cases, the density in above-median capacity cities is shifted even further to the right, suggesting that the mortality differences were not driven by outliers.

Above-median capacity cities were closer to World War I bases, had higher health expen-

---

[11]Data on manufacturing payroll is not available in 1910.

diture before and during the pandemic, had lower typhoid mortality rates, and were more populous. These cities also had a greater fraction of white residents, and higher levels of industrial activity, measured by manufacturing employment and payroll per worker. The goal of the empirical analysis is to disentangle these potentially competing determinants of pandemic-severity.

4 Empirical Strategy

The empirical analysis will evaluate whether heterogeneity in mortality during the influenza pandemic was related to local levels of pollution. We estimate the following model:

$$
\begin{aligned}
Log(MortRate_{ct}) = \alpha &+ \beta_{Coal}\big(I(Year_t = 1918) \times Log(CoalCap30mile_c)\big) \\
&+ \beta_{Hydro}\big(I(Year_t = 1918) \times Log(HydroCap30mile_c)\big) \\
&+ \lambda_t + \lambda_t Z_c + \lambda_t X_c + \lambda_t W_c + \eta_c + \phi_s t + \epsilon_{ct}
\end{aligned}
\tag{1}
$$

where the dependent variable, $Log(MortRate_{ct})$, denotes the logarithm of the infant or all-age mortality rate in city c in year t. The term η_c denotes a vector of city fixed effects, λ_t denotes a set of year fixed effects, and $\phi_s t$ is a linear state trend which allows for differential trends in mortality across states.[12] Standard errors are clustered at the city-level to adjust for heteroskedasticity and within-city autocorrelation.

The term $I(Year_t = 1918)$ is a dummy variable for the year 1918, which is meant to proxy pandemic exposure. The term $Log(CoalCap30mile_c)$ denotes the logarithm of coal-fired capacity within 30 miles of the city-centroid, our measure of local exposure to air pollution.[13] The interaction term of interest, $I(Year_t = 1918) \times Log(CoalCap30mile_c)$, identifies the differential impact of the pandemic on mortality in cities differing in levels of

---

[12]In some specifications, this state-year trend is replaced by a city-year trend or by a state-year fixed effect.

[13]The log-log specification is motivated by recent research that documents a concave ("supralinear") concentration-response relationship between pollution and mortality (Goodkind, Cogglin, and Marshall, 2014; Pope, Cropper, Coggins, and Cohen, 2015). Alternative functional forms are explored in the robustness checks.

coal-fired capacity. A positive estimate of β_{Coal} would suggest that exposure to power plant emissions exacerbated the impact of the influenza outbreak.[14] Meanwhile, the estimate β_{Hydro} captures the impact of emissions-free capacity on pandemic severity.

Equation (1) takes the form of a Bartik-style estimator, in which a time-varying national health shock affects locations differentially according to initial local differences in air pollution. The identifying assumption is that, conditional on covariates, unobservable determinants of mortality are not correlated with $I(Year_t = 1918) \times Log(CoalCap30mile_c)$. In practice, this condition requires that pandemic-related mortality did not differ systematically across high and low coal cities for reasons unrelated to air quality. There are three primary concerns regarding this hypothesis, which we address in turn.

First, geography may have influenced both the spread and severity of the pandemic. For example, some accounts suggest that the virulence of the H1N1 strain weakened by late September (Sydenstricker, 1918). Since the outbreak occurred later in western states, where access to coal was limited, viral evolution could lead to a spurious relationship between pandemic-severity and city-level pollution. To address this concern, equation (1) includes the term $\lambda_t Z_c$, which denotes the city-centroid longitude and latitude interacted with year. These terms allow for heterogeneity in pandemic-severity based on city location. Additionally, we estimate models that control for state-year fixed effects, which rely solely on within-state variation in pandemic-severity. These models address the concern of viral evolution, since the *timing* of pandemic onset varied little within states. Moreover, other determinants of disease transmission, such as weather, should be similar within states.[15]

A second possibility is that local population and demographic factors influenced the spread of the pandemic and hence the size of the infected population. We interact year fixed effects with a vector of baseline county-level population characteristics, X_c, that include the logarithm of total population in 1910, percent urban in 1910, and the share of white residents

---

[14]In some specifications, we also include the full vector of coal-fired and hydroelectric capacity-year fixed effect interaction terms. The coal capacity interaction effect in 1919 is of particular interest, since 17 percent of pandemic-related mortality occurred between January and April, 1919.

[15]There is a limited number of monitoring stations in this time period.

in 1910. These terms control for the fact that local pollution levels may be correlated with determinants of viral transmission. In addition, we control for annual city-level expenditure on health and sanitation, to allow for cross-city differences in the public health response to the pandemic.[16]

A final concern is that mortality was particularly elevated in 1918 in highly polluted cities as a result of the war. Although nationwide bituminous coal consumption rose just slightly in 1918, it is still possible that coal-intensive cities experienced a disproportionate increase in wartime production (and hence local pollution levels), which might have contributed to 1918 mortality, independently of the pandemic. To address this issue, W_c controls for baseline economic conditions (log manufacturing wages in 1900 and log employment in manufacturing in 1910), which we interact with year. We also examine the impact of coal-fired capacity on mortality in 1917, when the country was mobilizing for the war effort, but did not experience an influenza outbreak. In addition, we explore the sensitivity of the results to a range of alternative specifications and controls.

5   Results

5.1   Coal-Fired Capacity and Pandemic-Related Mortality

In Table 2, we examine the extent to which mortality in 1918 was related to local coal-fired and hydroelectric capacity. The top panel presents results for infant mortality and the bottom panel presents results for all-age mortality. Equation (1) is estimated across several different specifications. Column (1) includes city and year fixed effects and a linear state trend; column (2) adds controls for geography and population covariates; in column (3) we include a control for annual city-level expenditure on health and sanitation and controls for local economic conditions; and in column (4), we include a full set of interaction effects

---

[16]We restrict the sample to cities reporting expenditure for at least 5 years between 1915 and 1925, interpolating between missing observations. The results are not sensitive to this sample restriction.

between coal-fired and hydroelectric capacity and year fixed effects.

In the top and bottom panels, across all four specifications, the coefficients on $I(Year_t = 1918) \times Log(CoalCap30mile_c)$ are positive and highly significant. For infants, they range from 0.0285 to 0.0341. For all-age mortality, the coefficients range from 0.0209 to 0.0217. In both cases, the point estimates change relatively little with the inclusion of demographic and economic covariates, which provides confidence that the local coal-fired capacity is not simply a proxy for some other determinant of pandemic severity. The coefficients on $I(Year_t = 1918) \times Log(HydroCap30mile_c)$ are small and not statistically significant. In all but one specification, the estimates of β_{Coal} and β_{Hydro} are statistically different from each other.[17] Together, these results provide confidence that coal-fired capacity captures exposure to air pollution and not some other characteristic related to electricity production.

Industrial mobilization for World War I may have contributed to mortality in 1918 in heavily polluted cities independently of the pandemic. To disentangle these two effects, Table 3 reports the coefficients for 1917, 1918, and 1919. In 1917, mobilization for World War I was under way, but there was no pandemic. The pandemic was most severe in 1918, although pandemic-related mortality persisted into the spring of 1919, well after the armistice on November 11. Columns 1 and 3 include just those three years are interacted with coal-fired capacity, and columns 2 and 4 include a full set of year-capacity interactions. There is no evidence of an interaction effect in 1917, suggesting that wartime mobilization did not differentially affect mortality based on coal-fired capacity. The interaction effects are significant in both 1918 and 1919. The coefficient on 1919 is substantially smaller in magnitude, consistent with the timing of mortality during the pandemic. The inclusion of the full set of interaction effects has very little effect on the point estimates in these regressions.[18]

To quantify the role of air pollution in exacerbating the pandemic, we assess the severity of the influenza pandemic in two alternative scenarios. We first consider a setting in which

---

[17]Regressions weighted by city population yield qualitatively similar results.

[18]The coefficients for the full set of interaction effects are reported in appendix Table A.3.

all above-median cities reduced coal-fired capacity to the median level. In this scenario, individuals in above-median cities would experience a mean reduction in coal-fired capacity of 1.76 log points.[19] Applying the point estimates from Table 3, we calculate the number of infant deaths and all-age deaths that would have been averted in this scenario.

Table 4 reports these mortality reductions, with square brackets reporting the 95 percent confidence interval. Columns (1) and (3) report the number of deaths averted under the assumption that the effect of coal capacity on pandemic mortality occurred solely in 1918. We calculate that 3,400 infant deaths and 15,575 all-age deaths would have been averted had coal capacity been reduced to the median-city level, a decrease of roughly 37 infant deaths and 169 all-age deaths per city.[20] In columns (2) and (4) of Table 4, we allow coal capacity to affect excess mortality in both 1918 and 1919, based on the linear combination of the interaction effects founds in Table 3. In this scenario, we calculate that a reduction in coal capacity to the median level would have averted 5,860 infant deaths and 23,686 all-age deaths, a decrease of 64 infant deaths and 257 all-age deaths per city.

The results suggest that air pollution played an important role in exacerbating the pandemic. By reducing coal capacity to the median level, above median cities would have experienced a 26 percent decrease in pandemic-related infant mortality and a 15 percent decrease in all-age pandemic-related mortality. The large effects found for infants are consistent with early-life health outcomes being particularly sensitive to environmental air quality (Currie et al, 2014). The fact that pre-1915 air quality is unobserved will also tend to downward bias the estimates for all-age mortality.[21]

---

[19]To derive the change in individual-level exposure to coal-fired capacity, cities are weighted by city population (births) in 1921. Intuitively, this scenario is equivalent to moving all individuals in above-median coal capacity to a median capacity city. As a result, we want to rely on the change in individual-level exposure rather than the change in city-level exposure when calculating the number of lives saved.

[20]To derive these estimates, we first calculate the change in death probability associated with the pollution reduction and then multiply by the total population in high coal cities: ΔProbability of a death $= \hat{\beta}_{Coal} \cdot \Delta$Log(CoalCap30mile) $\cdot \overline{\text{MortRate}}_{1918}/10,000 = 0.0218 \times 1.764 \times 191.05/10,000 = 0.0007346$. Given a total population 21.3 million in above-median cities, the number of deaths averted is calculated to be 15,575.

[21]Currie et al (2014) show that exposure to pollution in early childhood is particularly important for later health outcomes. The large rural outmigration that occurred during the late 19th and early 20th century implies that many city residents were not exposed to urban air pollution during childhood.

Although the previous scenario provided insight into the impact of air pollution on pandemic mortality, in practice, it would have been difficult to achieve these reductions in coal-fired capacity. Hydroelectricity was an alternative to coal-fired power, although the feasibility of this substitute depended crucially on local topography. We next consider a setting in which hydroelectric capacity added between 1920 and 1930 was instead been installed prior to the pandemic. This scenario has the advantage that it simply alters the *timing* of capacity installations, and does not impose implausible assumptions on the local availability of hydroelectric potential. We digitized information on the location of hydroelectric power plants built between 1920 and 1930 (Federal Power Commission, 1962). For each city in the sample, we calculate the change in hydroelectric capacity within 50 miles between 1920 and 1930.[22] We then calculate the required coal-fired capacity necessary to maintain energy production at its 1918 level.

This scenario is associated with an average city-level decrease in coal-fired capacity of 0.551 log points. Combining the health interaction effects in 1918 and 1919, we calculate that 1,838 infant deaths and 7,430 all-age deaths could have been averted had coal capacity been replaced with readily available hydropower. These estimates correspond to a 16 percent decrease in pandemic-related infant mortality and a 5 percent decrease in all-age pandemic-related mortality.

5.2   Other Determinants of Pandemic Severity

A number of other factors may have contributed to mortality during the pandemic, including proximity to World War I bases and city baseline health conditions. To assess the effect of these factors, we add interaction terms for these variables to the main specification. The factors include the logarithm of distance to the nearest World War I base, the logarithm of city-level expenditure on health and sanitation during the pandemic year, the logarithm of average health and sanitation expenditure between 1915 and 1917, and the logarithm

---

[22]The expanded radius reflects the fact that available hydroelectric capacity tends to be located further outside cities.

of baseline typhoid mortality. Table 5 reports the results for infant mortality and Table 6 reports the results for all-age mortality.

In Table 5 across columns (1)-(5), the estimates of the effect of coal-fired capacity on infant mortality are robust to the inclusion of these other determinants. In column (5), the coefficients on proximity to World War I bases and baseline typhoid mortality are both statistically significant and have the expected signs. In the bottom half of Table 5, we assess the relative magnitude of these various factors for the cross-city variation in pandemic severity. Each row reports the fraction of the standard deviation in excess pandemic mortality that can be explained by a one standard deviation change in each independent variable. Together, coal-fired capacity, proximity to World War I bases, and baseline typhoid mortality can account for 76 percent of the cross-city variation in excess infant mortality.

In Table 6, the estimates of the effect of coal-fired capacity on all mortality are also robust to the inclusion of these other determinants. The magnitude of the coefficient in column (5) is, however, smaller than in column (1). In column (5), the coefficient on baseline typhoid mortality is statistically significant. The coefficient on distance to a World War I base is negative, but not statistically significant. In the bottom half of Table 6, coal-fired capacity, and baseline typhoid mortality can account for 44 percent of the cross-city variation in excess infant mortality. Together, these findings support previous research that argues that baseline health capital influenced population susceptibility to the pandemic.

6 Robustness Checks

6.1 Heterogeneity by Baseline Capacity Levels and Pre-Pandemic Infant Mortality

In Table 7, we assess the robustness of the main findings to the functional form specification. The top panel reports the interaction effect by tercile of coal-fired capacity for both infant mortality and all-age mortality. In general, the elasticities increase across terciles.

For infant mortality the effects are significant in the top two terciles, for all-age mortality significant effects are found only in the top tercile. These findings indicate that the main results are not driven by outliers. These estimates are also consistent with previous research that shows that the concentration-response relationship between pollution and mortality is concave (Goodkind, Cogglin, and Marshall, 2014; Pope, Cropper, and Cohen, 2015). This research finds that marginal changes in air quality have small health impacts at higher pollution levels, motivating the choice of the log-log functional form specification.

We next examine the interaction effect between coal-fired capacity and pandemic-severity across each tercile of pre-pandemic infant mortality.[23] This analysis offers three advantages. First, the results address the selection of less healthy individuals – who may have been particularly vulnerable to the pandemic – into heavily polluted cities. By estimating the impact of coal-fired capacity on pandemic-related mortality across cities with comparable levels of baseline health, we are able to mitigate these selection concerns. Second, the analysis allows us to distinguish the impact of air pollution from the health effects of other factors present in heavily polluted cities – such as polluted water. Third, the results shed light on how air pollution and health capital interact more broadly to influence individual susceptibility to an infectious disease outbreak.

The bottom panel of Table 7 reports these results. For both infant and all-age mortality, the estimated effects of coal capacity on pandemic-severity are significant in all but the lowest tercile. The significant effects of pollution on pandemic mortality across cities with comparable baseline levels of health provides strong evidence that the original findings were not driven by the selection of unhealthy individuals into heavily polluted cities. These findings also suggest that air pollution affected the severity of the pandemic in ways that are not fully captured by pre-pandemic mortality. Finally, the interaction effects increase with tercile of pre-pandemic mortality, which suggests that exposure to air pollution may interact with baseline health capital more broadly to make individuals more susceptible to

---

[23]Pre-pandemic infant mortality is defined as average city-level infant mortality between 1915 and 1917.

the consequences of a negative health shock.

6.2 Additional Specification Checks

Table 8 examines the robustness of the main findings to several other specifications and controls. For reference, column (1) of Table 8 reports the baseline estimates from column (3) of Table 2.

A concern with the baseline findings is that evolution in the severity of the influenza strain may be spuriously correlated with local pollution levels. Despite the fact that the pandemic was a very short-lived phenomenon, previous research suggests that the virus may have weakened in the weeks following the initial outbreak in early September (Syndenstricker, 1918). If the virus did evolve as it spread across the country, and the timing of pandemic onset is correlated with city-level pollution levels, the baseline results may be biased. For example, if the virus weakened as it spread to western cities, which were simultaneously less reliant on coal, the baseline results might overestimate the relationship between air pollution and pandemic severity.

To address this issue, we construct a measure of the timing of pandemic onset. The analysis relies on a map by Sydenstricker (1918), which identifies the week of approximate pandemic onset across 376 localities (see Figure 3). We digitized this information, identifying the week of pandemic onset for each city in the sample, and interact this variable with the vector of year fixed effects. These models identify the impact of pollution across population exposed to similar strains of the virus. The results are reported in column (2). Controlling for the timing of pandemic onset has little effect on main findings. The point estimates are somewhat smaller in magnitude but remain significant.

To further address the issue of viral evolution, column (3) controls for state-year fixed effects. These models rely solely on within-state differences in pandemic severity, which greatly limits the importance of timing, since the onset occurred within a very short time frame within states. The results are highly significant, and slightly stronger than the baseline

findings. Together, these results provide confidence that the interaction effect between coal-fired capacity and the 1918 pandemic was not driven by evolution in strain severity.

Table A.1 shows the timing of local smoke-abatement legislation adopted prior to 1930. These policies may have led to secular changes in air pollution across cities. In column (4), we allow for differential trends in pollution across cities, replacing the state-specific trend with a city-specific trend. The results are robust to this alternative specification. In column (5) we restrict the sample to the balanced panel of 129 cities reporting mortality in each year between 1915 and 1925. The point estimates are similar to the baseline results, indicating that non-random sample selection is not driving the original findings. Columns (6)-(8) examine the sensitivity of the estimates to alternative measures of local pollution exposure. Column (6) reports the results based on coal-fired capacity in levels (100s of megawatts). The results are again sizable, although the coefficient for all-age mortality is imprecisely estimated. Columns (7) and (8) measure coal-fired capacity within 50 and 100 miles of each city-centroid. The point estimates are again large and statistically significant, confirming that the results are not sensitive to the particular distance cutoff. The effect of a one standard deviation change in coal-fired capacity is similar across the three distances. In unreported regression models that include both coal-fired capacity within 30 miles and 100 miles, only the former has a significant impact on pandemic-related mortality. These findings support our choice of distance, and are consistent with evidence suggesting that the historical dispersion of power plant emissions was highly localized.

7   Conclusion

This paper provided new evidence on the extent to which air pollution exacerbated infant and all-age mortality during 1918-1919 influenza pandemic. The effects of air pollution are sizeable. Cities with above median levels of coal-fired capacity collectively experienced thousands of excess infant deaths and tens of thousands of excess all-age deaths during the

pandemic. Baseline city health conditions, as measured by typhoid deaths, also played an important role. Together with distance to World War I bases, these factors explain 76 and 44 percent of the cross-city variation in infant and all-age pandemic mortality.

The 1918 influenza pandemic was an exceptional episode, with death rates 5 to 20 times higher than subsequent pandemics. Our findings highlight the need for research on the impact of air pollution on later pandemics, including the 1957-1958, 1968-1969, and 2008-2009 pandemics, when healthcare was more readily available. Because conventional studies often exploit quasi-experimental annual variation in air pollution, they may understate the true benefits of abatement policies, which arise only during these exceptional years.

Air pollution is particularly important today, because a far greater share of the global population live in heavily polluted cities than did in the early 20th century. Moreover, many cities in the developing world do not yet have reliable clean drinking water. Even with modern antiviral and antibacterial drugs, a pandemic virus with similar pathogenicity to the 1918 virus would quickly overwhelm the existing medical infrastructure, and would likely kill more than 100 million people worldwide (Taubenberger and Morens, 2006). Thus, preventative approaches including pollution abatement, improvements in access to clean drinking water, and strategic allocation of vaccination efforts are likely to be critical for mitigating mortality.

References

Acuna-Soto, Rodolfo, Cecilia Viboud, and Gerardo Chowell. 2011. "Influenza and pneumonia mortality in 66 large cities in the United States in years surrounding the 1918 pandemic." *PloS one*, 6(8): e23467.

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

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

Arceo-Gomez, Eva, Rema Hanna, and Paulina Oliva. 2012. "Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City." NBER Working Paper #18349.

Bootsma, Martin, and Neil Ferguson. 2007. "The effect of public health measures on the 1918 influenza pandemic in U.S. cities." *PNAS*, 104(18): 7588-7593.

Brainerd, Elizabeth, and Mark V. Siegler. 2003. "The Economic Effects of the 1918 Influenza Epidemic." Discussion Paper no. 3791, Centre Econ. Policy Res., Paris.

Brook, Robert, et al. 2010. "Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association," *Circulation*, 121: 2331-2378.

Byerly, Carol. 2010. "The U.S. Military and the Influenza Pandemic of 1918-1919," *Public Health Reports*, 125(3): 82-91.

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

Chay, Kenneth Y. and Michael Greenstone. 2003b. "Air Quality, Infant Mortality, and the Clean Air Act of 1970." MIT Department of Economics Working Paper No. 04-08.

Chowell, Gerardo, et al. 2008. "The 1918-1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact." *Proceedings of the Royal Society B: Biological Sciences*, 275.1634: 501-509.

Cohen, Aaron et al. 2004. "Chapter 17: Urban Air Pollution" in *Comparative Quantification of Health Risks, Volume 2*. Geneva: World Health Organization.

Collins, Selwyn D. 1930. "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(33): 1909-1937.

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

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

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

Currie, Janet, and Reed Walker. 2011. "Traffic Congestion and Infant Health: Evidence from E-Z Pass," *American Economic Journal: Applied Economics*, 3(1): 65-90.

Delfino, Ralph J., Constantinos Sioutas, and Shaista Malik. 2005. "Potential Role of Ultrafine Particles in Associations between Airborne Particle Mass and Cardiovascular Health," *Environmental Health Perspectives*, 113: 934-946.

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

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

Goklany, Indur M. *Clearing the air: the real story of the war on air pollution*. Cato Institute, 1999.

Goodkind, Andrew L., Jay S. Coggins, and Julian D. Marshall. 2014. "A Spatial Model of Air Pollution: The Impact of the Concentration-Response Function," *Journal of the Association of Environmental and Resource Economists*, 1(4): 451-479.

Hahon, Nicholas, et al. 1985. "Influenza virus infection in mice after exposure to coal dust and diesel engine emissions," *Environmental Research*, 37(1): 44-60.

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

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

Harrod, Kevin, et al. 2003. "Increased Susceptibility to RSV Infection by Exposure to Inhaled Diesel Engine Emissions," *American Journal of Respiratory Cell and Molecular Biology*, 28(4): 451-463.

Hoek, Gerard, et al. 2013. "Long-term air pollution exposure and cardio-respiratory mortality: a review," *Environmental Health*, 12(43): 1-15.

Huntington, Ellsworth. 1923. "Causes of Geographical Variation in the Influenza Epidemic in the Cities of the United States," *Bull. Nat. Res. Council*, 6: 1-36.

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

Ives, James E., et al. 1936. *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.

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

Jaspers, Ilona, et al. 2005. "Diesel Exhaust Enhances Influenza Virus Infections in Respiratory Epithelial Cells," *Toxicology Sciences*, 85(2): 990-1002.

Johnson, Niall. and Juergen Mueller. 2002. *Bulletin of Historical Medicine*, 76: 105-115.

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

Lee, Greg, et al. 2014. "Exposure to combustion generated environmentally persistent free radicals enhances severity of influenza virus infection," *Particle and fibre toxicity*, 11(1): 57.

Levy, Jonathan I., et al. 2002. "Using CALPUFF to evaluate the impacts of power plant emissions in Illinois: model sensitivity and implications," *Atmospheric Environment*, 36: 1063-1075.

Lin, Ming-Jen, and Elaine M. Liu. 2014. "Does in *utero* exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment," *Journal of Health Economics*, 37: 152-163.

Mamelund, Svenn-Erik. 2006. "A Socially Neutral Disease? Individual Social Class, Household Wealth and Mortality from Spanish Influenza in Two Socially Contrasting Parishes in Kristiania 1918-19," *Social Science and Medicine* 62: 923-40.

Miller, Grant. "Women's Suffrage, Political Responsiveness, and Child Survival in American History," *Quarterly Journal of Economics*, 123(3): 1287-1327.

Neelsen, Sven, and Thomas Stratmann. 2012. "Long-Run Effects of Fetal Influenza Exposure: Evidence from Switzerland," *Social Science and Medicine*, 74(1): 58-66.

Nelson, Richard E. 2010. "Testing the fetal origins hypothesis in a developing country: Evidence from the 1918 influenza pandemic," *Health Economics*, 19(10): 1181-1192.

Pearl, Raymond. 1921. "Influenza studies: further data on the correlation of explosiveness of outbreak of the 1918 epidemic." *Public Health Report*, 36: 273-298.

Pope, C. Arden III, et al. 2004. "Cardiovascular mortality and longterm exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease," *Ciruculation*, 109: 71-77.

Pope, C. Arden III, Maureen Cropper, Jay Coggins, and Aaron Cohen. 2015. "Health Benefits of Air Pollution Abatement Policy: Role of the Shape of the Concentration-Response Function," *Journal of the Air & Waste Management Association*, 65(5): 516-522.

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

Stern, Arthur C. "History of Air Pollution Legislation in the United States." *Journal of the Air Pollution Control Association* 32.1 (1982): 44-61.

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

Taubenberger, Jeffery, and David M. Morens. 2006. "1918 Influenza: the Mother of All Pandemics," *Emerging Infectious Diseases*, 12(1): 15-22.

U.S. Census Bureau. 1919. *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.

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

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

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

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

U.S. Federal Power Commission. 1962. *Principal Electric Power Facilities in the United States (map).* Washington D.C.: U.S. Federal Power Commission.

U.S. War Department. 1919. *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.

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

World Health Organzation. 2014. World Health Statistics: 2014. http://www.who.int/gho/publications/world_health_statistics/en/ (accessed September 2015).

Woodruff, Tracey, Lyndsey Darrow, and Jennifer Parker. 2008. "Air Pollution and Post-neonatal Infant Mortality in the United States, 1999-2002." *Environmental Health Perspectives* 116: 110-115.

Xu, Zhiwei, et al. 2013. "Air pollution, temperature, and pediatric influenza in Brisbane, Australia," *Environmenta International*, 59: 384-388.

Figure 1: Influenza and Pneumonia Deaths by Month
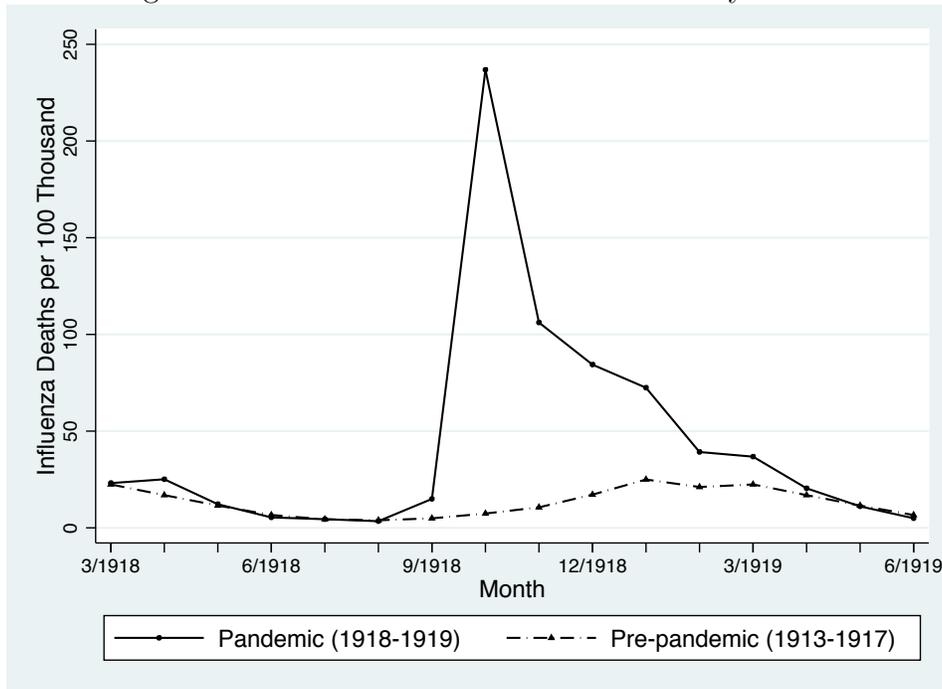
Source: Vital Statistics, 1913-1919.

Figure 2: Case-fatality Rate from Influenza and Pneumonia by Age
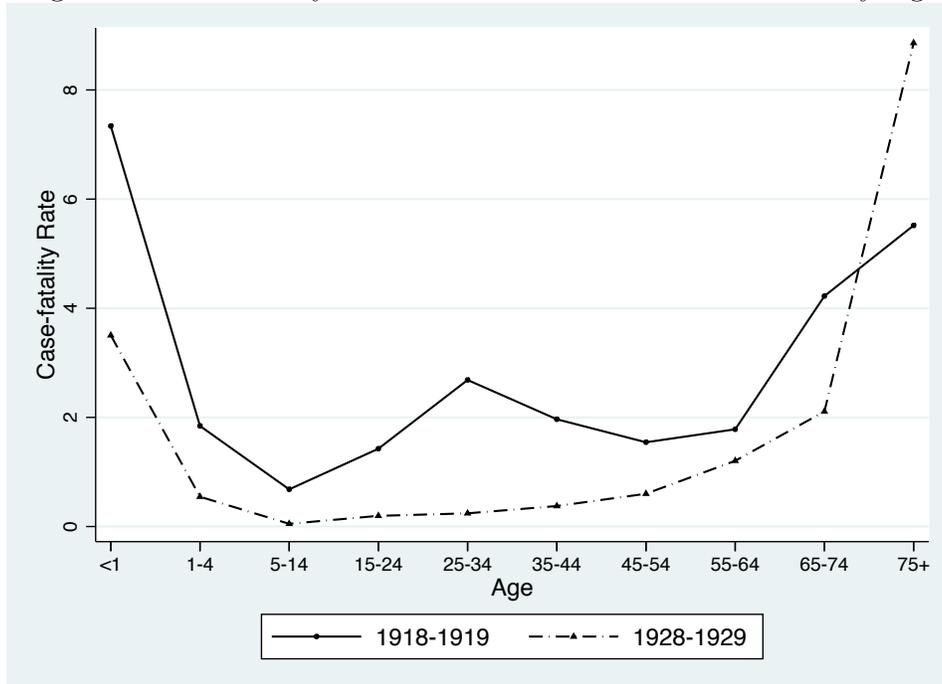
Notes: Based on surveys conducted by the United States Public Health Service in 12 localities in 1918-1919 and 14 localities in 1928-1929. Source: Collins (1931).

Figure 3: Timing of Pandemic Onset for 376 Localities
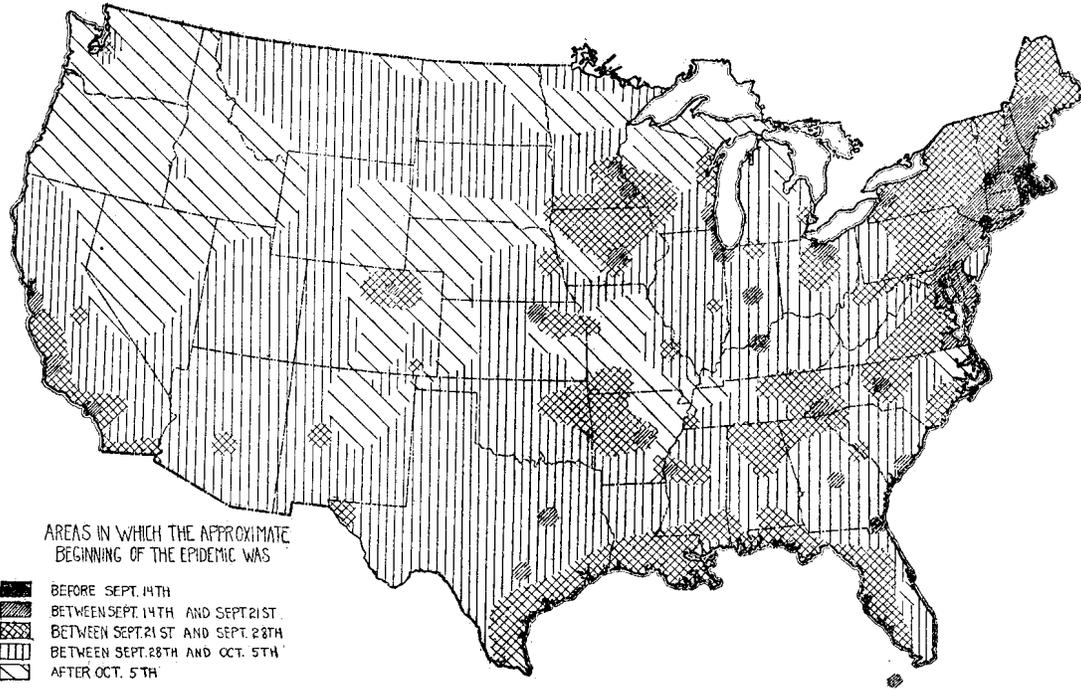
Source: Sydenstricker (1918).

Figure 4: State-level Coal Consumption, Coal-Fired Capacity, and Hydroelectric Capacity
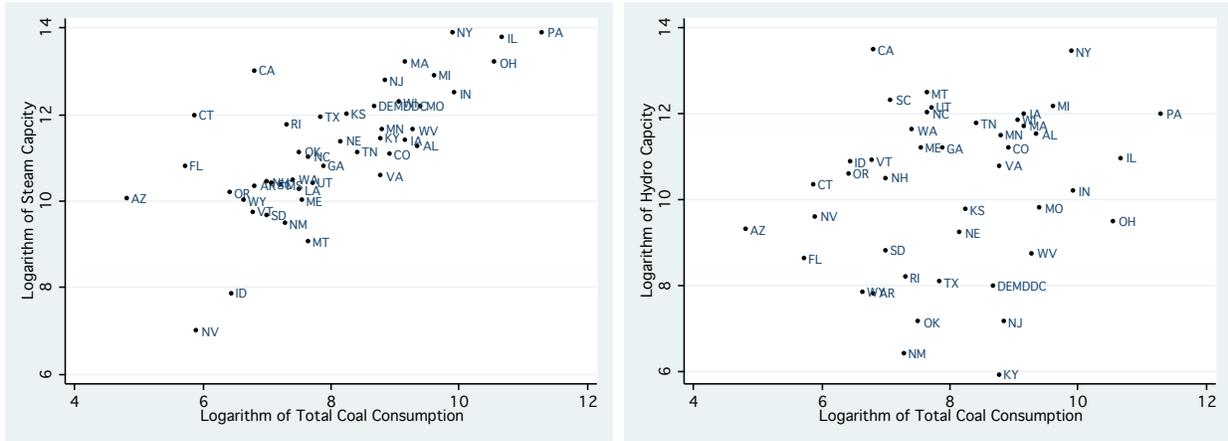
Notes: These figures report the relationship between the logarithm of total state-level coal consumption, coal-fired capacity and hydroelectric capacity in 1917. Source: Mineral Resources of the United States (1917, p.1254).

Figure 5: Sample selection and the location coal-fired and hydroelectric power plants
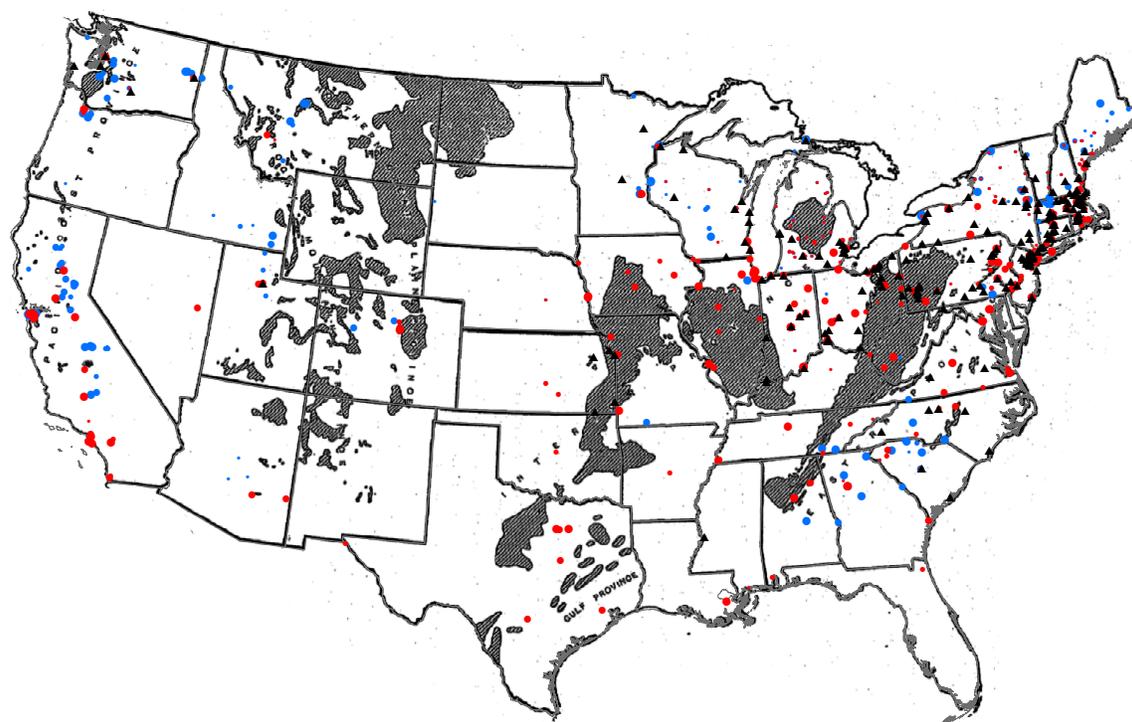
Notes: This figures presents the location of 183 cities in the sample, the location of coal-fired and hydroelectric power plants in 1915 overlayed on a map of major coal fields in 1919. Red and blue circles denote steam and hydroelectric power plants by quartile of capacity ($<$7mw, 7-11mw, 11-18mw, $>$18mw), and shaded areas denote major coal fields. Source: Fourteenth Census of the United States, Vol. XI, Mines and Quarries, 1919, p.254.

Figure 6: Excess Mortality by Year
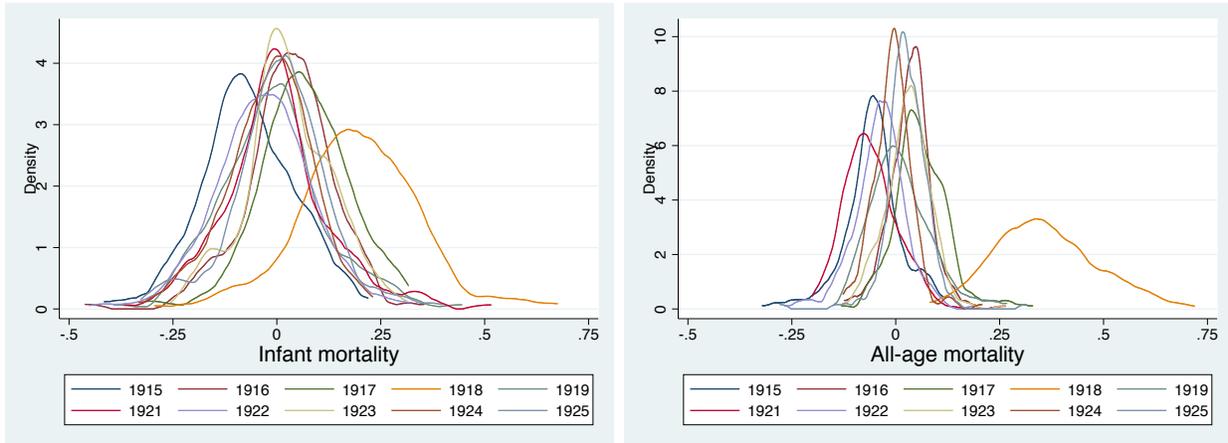
Notes: This figure reports the density of residual infant mortality and residual all-age mortality from regressions that controls for a linear city-specific trend. Mortality is plotted separately for each year in the sample between 1915 and 1925.

Figure 7: Excess Mortality in 1918 for Cities Above and Below Median Steam Capacity
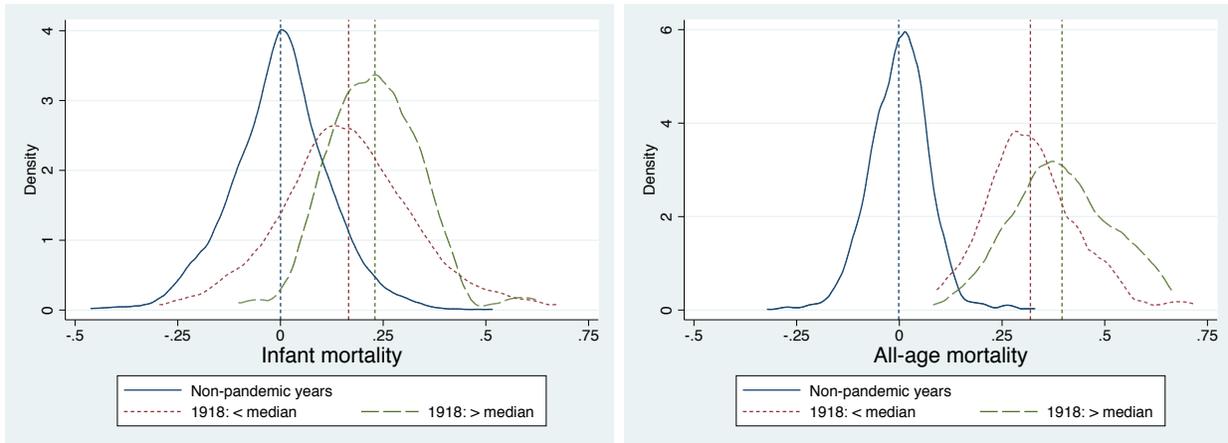
Notes: This figure reports the density of residual mortality from a regression that controls for a linear city-specific trend. The distributions are plotted in non-pandemic years and in pandemic years for cities above and below median coal-fired capacity.

Table 1: Summary statistics

|	Full sample	Difference, by capacity: (Above median - Below median)
Panel A: Mortality		
Log(infant mortality)	4.413	0.042
	[0.283]	[0.031]
Log(all-age mortality)	7.196	-0.072
	[0.256]	[0.031]**
Log(excess infant mortality in 1918)	0.196	0.060
	[0.146]	[0.022]***
Log(excess all-age mortality in 1918)	0.359	0.070
	[0.123]	[0.018]***
Panel B: Hydroelectric and coal-fired capacity		
Log(coal-fired capacity within 30 miles)	3.828	3.030
	[1.866]	[0.166]***
Coal-fired capacity within 30 miles	180.3	328.8
	[324.2]	[41.9]***
Hydro capacity within 30 miles	11.6	-14.0
	[29.8]	[4.6]***
Panel C: Other determinants of pandemic severity		
Distance to nearest WWI base	87.7	-42.5
	[84.0]	[12.8]***
Health expenditure in 1918 (per capita)	1.58	0.41
	[0.89]	[0.14]***
Pre-pandemic health expenditure (per capita)	1.38	0.39
	[0.91]	[0.14]***
Pre-pandemic typhoid mortality rate (per 100,000)	35.85	-16.89
	[24.46]	[4.03]***
Panel D: City characteristics		
City population in 1921	155,204	151,733
	[466,137]	[69,324]**
City births in 1921	3,766	3,720
	[10,953]	[1,628]**
Panel E: Pre-pandemic county characteristics		
Log(county population in 1910)	11.921	1.231
	[0.949]	[0.111]***
Share urban in 1910	0.756	0.156
	[0.167]	[0.023]***
Share white in 1910	0.950	0.053
	[0.105]	[0.015]***
Log(total manufacturing payroll per population in 1900)	0.959	0.046
	[0.049]	[0.007]***
Employment share in manufacturing in 1910	0.143	0.041
	[0.059]	[0.009]***
Number of cities	183	
Observations	1,611	

Notes: Column 1 reports the sample means. Column 2 reports the regression estimates for the difference between cities above and below median coal-fired capacity (31.7mw). Standard errors are clustered at the city-level. ***,**,* denote significance at the 1%, 5%, and 10%, respectively.

Table 2: The effect of the pandemic on mortality, by coal-fired and hydroelectric capacity

	Estimated effect			
	(1)	(2)	(3)	(4)
Dependent variable: Log(infant mortality)				
I(Year=1918) × Log(CoalCap30mile)	0.0285***	0.0309***	0.0324***	0.0341***
	[0.00592]	[0.00754]	[0.00831]	[0.0127]
I(Year=1918) × Log(HydroCap30mile)	-0.000296	-0.00392	-0.000949	0.0126
	[0.00651]	[0.00830]	[0.00908]	[0.0123]
P-value from test of $\beta_{Coal} = \beta_{Hydro}$	<0.001	<0.001	0.001	0.099
Observations	1,611	1,611	1,611	1,611
R-squared	0.767	0.787	0.794	0.797
Number of clusters	183	183	183	183
Sample mean				
Log(excess infant mortality in 1918)		Mean = 0.196, S.D. = 0.146		
Dependent variable: Log(all-age mortality)				
I(Year=1918) × Log(CoalCap30mile)	0.0209***	0.0240***	0.0217***	0.0218**
	[0.00550]	[0.00757]	[0.00777]	[0.00935]
I(Year=1918) × Log(HydroCap30mile)	0.00117	0.00166	0.00253	0.00583
	[0.00570]	[0.00627]	[0.00684]	[0.00786]
P-value from test of $\beta_{Coal} = \beta_{Hydro}$	0.009	0.025	0.053	0.128
Observations	1,610	1,610	1,610	1,610
R-squared	0.907	0.915	0.920	0.922
Number of clusters	183	183	183	183
Sample mean				
Log(excess all-age mortality in 1918)		Mean = 0.359, S.D. = 0.123		
Controls				
City & year FE + Linear state trend	Y	Y	Y	Y
Demographic & geographic covariates × year		Y	Y	Y
Annual city-level health expenditure			Y	Y
Manufacturing covariates × year			Y	Y
Coal & hydro capacity × year				Y

Notes: Each column reports the point estimates from a different regression. The variables Log(CoalCap30mile) and Log(HydroCap30mile) denote logarithm of coal-fired and hydroelectric capacity within 30 miles of the city-centroid. Demographic and geographic covariates include county-level controls for the logarithm of population, percent urban, and share white in 1910, and longitude and latitude. Manufacturing covariates include the logarithm of manufacturing wages in 1900 and the logarithm of manufacturing employment in 1910. The final column includes the full interaction effects of Log(CoalCap30mile) and Log(HydroCap30mile) with year fixed effects. Standard errors are clustered at the city-level. ***,**,* denote significance at the 1%, 5%, and 10%, respectively.

Table 3: The effect of coal-fired capacity on mortality in 1917, 1918, and 1919

	Estimated effect			
	(1)	(2)	(3)	(4)
	Dependent variable: **Log(infant mortality)**		**Dependent variable:** **Log(all-age mortality)**	
Log(CoalCap30mile) ×				
I(Year=1917)	0.0078	0.00935	-0.0030	-0.00267
	[0.00961]	[0.0162]	[0.00601]	[0.00858]
I(Year=1918)	0.0352***	0.0341***	0.0222***	0.0218**
	[0.00861]	[0.0127]	[0.00827]	[0.00935]
I(Year=1919)	0.0205**	0.0209*	0.0107*	0.0112*
	[0.00914]	[0.0116]	[0.00536]	[0.00671]
Full controls	Y	Y	Y	Y
Coal and hydro capacity × year FE		Y		Y

Notes: Each column reports the point estimates from a different regression. The variable Log(CoalCap30mile) denotes the logarithm of coal-fired capacity within 30 miles of the city-centroid. All models include include the full set of controls reported in column (3) of Table 2. Columns (2) and (4) include the full interaction effects of Log(CoalCap30mile) and Log(HydroCap30mile) with year fixed effects. Standard errors are clustered at the city-level. ***,**,* denote significance at the 1%, 5%, and 10%, respectively.

Table 4: Pandemic-related deaths averted by reducing coal-fired capacity in cities

	Infant deaths averted		**All-age deaths averted**	
	Health impact only in 1918	Health impact in 1918 & 1919	Health impact only in 1918	Health impact in 1918 & 1919
	(1)	(2)	(3)	(4)
Scenario 1: *Reduce coal-fired capacity in above-median cities to median level (51.7mw)*				
# deaths averted	3,400	5,860	15,575	23,686
	[1,676 5,125]	[2,957 8,763]	[4,553 26,599]	[7,653 39,719]
Scenario 2: *Replace coal-fired capacity with available hydroelectric capacity*				
# deaths averted	1,067	1,838	4,886	7,430
	[526 1,608]	[926 2,749]	[1,428 8,344]	[2,400 12,459]

Notes: Each column reports the number of deaths averted under each scenario, based on the regression estimates reported in columns (2) and (4) of Table 3. The number of deaths averted in columns (1) and (3) are derived solely from the 1918 interaction effect. The number of deaths averted in columns (2) and (4) are derived based on the linear combination of the effects in 1918 and 1919. Square brackets report the 95% confidence interval for the number of deaths averted.

Table 5: Other determinants of pandemic severity

Dep variable: Log(infant mortality)

	Estimated effect				
	(1)	(2)	(3)	(4)	(5)
I(Year=1918) ×					
Log(CoalCap30mile)	0.0279*** [0.00798]	0.0324*** [0.00818]	0.0291*** [0.00762]	0.0334*** [0.00928]	0.0254*** [0.00809]
Log(Distance to nearest WWI base)	-0.0366** [0.0168]				-0.0456** [0.0205]
Log(Health expenditure in 1918)		0.000339 [0.0139]			0.0165 [0.0643]
Log(Pre-pandemic health expenditure)			0.0145 [0.0142]		-0.00486 [0.0628]
Log(Baseline typhoid mortality)				0.0290 [0.0213]	0.0371* [0.0207]

Fraction of variation in excess 1918 infant mortality explained by 1 s.d. Δ indep vbl (× 100)

I(Year=1918) ×					
Log(CoalCap30mile)	**35.6%**	**41.3%**	**37.1%**	**42.6%**	**32.4%**
Log(Distance to nearest WWI base)	**22.2%**				**27.6%**
Log(Health expenditure in 1918)		0.3%			13.7%
Log(Pre-pandemic health expenditure)			12.3%		4.1%
Log(Baseline typhoid mortality)				12.3%	**15.7%**
Observations	1,611	1,611	1,528	1,311	1,266
R-squared	0.795	0.794	0.814	0.808	0.809
Number of clusters	183	183	156	141	129

Notes: Each column of the top panel reports the point estimates from a different regression. All models include the full set of controls reported in column (3) of Table 2. Standard errors are clustered at the city-level. ***,**,* denote significance at the 1%, 5%, and 10%, respectively. The bottom panel reports the fraction of the cross-city standard deviation in Log(excess infant mortality in 1918) that is explained by a one standard deviation change in each indepenent variable.

Dep variable: Log(all-age mortality)

	Estimated effect				
	(1)	(2)	(3)	(4)	(5)
I(Year=1918) ×					
Log(CoalCap30mile)	0.0213*** [0.00788]	0.0208*** [0.00755]	0.0171** [0.00727]	0.0197** [0.00813]	0.0133* [0.00762]
Log(Distance to nearest WWI base)	-0.00301 [0.0141]				-0.0124 [0.0144]
Log(Health expenditure in 1918)		-0.0102 [0.0140]			-0.0276 [0.0610]
Log(Pre-pandemic health expenditure)			-0.00738 [0.0139]		0.0141 [0.0613]
Log(Baseline typhoid mortality)				0.0502*** [0.0188]	0.0469** [0.0198]

Fraction of variation in excess 1918 all-age mortality
explained by 1 s.d. Δ indep vbl (× 100)

	(1)	(2)	(3)	(4)	(5)
I(Year=1918) ×					
Log(CoalCap30mile)	**32.3%**	**31.5%**	**25.9%**	**29.9%**	**20.2%**
Log(Distance to nearest WWI base)	2.2%				8.9%
Log(Health expenditure in 1918)		10.1%			27.3%
Log(Pre-pandemic health expenditure)			7.4%		14.2%
Log(Baseline typhoid mortality)				**25.3%**	**23.7%**
Observations	1,610	1,610	1,527	1,310	1,265
R-squared	0.920	0.920	0.913	0.915	0.915
Number of clusters	183	183	156	141	129

Notes: Each column of the top panel reports the point estimates from a different regression. All models include the full set of controls reported in column (3) of Table 2. Standard errors are clustered at the city-level. ***,**,* denote significance at the 1%, 5%, and 10%, respectively. The bottom panel reports the fraction of the cross-city standard deviation in Log(excess all-age mortality in 1918) that is explained by a one standard deviation change in each indepenent variable.

Table 7: Heterogeneous effects by coal-fired capacity and pre-pandemic mortality

	Estimated effect			
	(1)	(2)	(3)	(4)
	Dependent variable: Log(infant mortality)		Dependent variable: Log(all-age mortality)	
Heterogeneity by tercile of coal-fired capacity				
I(Year=1918) × Log(CoalCap30mile) ×				
Lower tercile	0.00936	0.0142	-0.0105	-0.0137
	[0.0175]	[0.0192]	[0.0125]	[0.0142]
Middle tercile	0.0221**	0.0247**	0.00433	0.00584
	[0.00915]	[0.0102]	[0.00729]	[0.00899]
Upper tercile	0.0252***	0.0312***	0.0159***	0.0196***
	[0.00642]	[0.00821]	[0.00542]	[0.00742]
Observations	1,611	1,611	1,610	1,610
R-squared	0.767	0.794	0.907	0.921
Number of clusters	183	183	183	183
Controls				
City & Year FE + Linear state trend	Y	Y	Y	Y
Full controls		Y		Y
Heterogeneity by tercile of pre-pandemic infant mortality				
I(Year=1918) × Log(CoalCap30mile) ×				
Lower tercile	0.0160**	0.0234**	0.00146	0.00766
	[0.00770]	[0.00929]	[0.00739]	[0.00826]
Middle tercile	0.0263***	0.0310***	0.0181***	0.0211***
	[0.00638]	[0.00810]	[0.00540]	[0.00692]
Upper tercile	0.0346***	0.0416***	0.0300***	0.0341***
	[0.00642]	[0.00898]	[0.00655]	[0.00839]
Observations	1,611	1,611	1,610	1,610
R-squared	0.768	0.795	0.910	0.923
Number of clusters	183	183	183	183

Notes: Each column reports the point estimates from a different regression. Columns (2) and (4) include the full set of controls reported in column (3) of Table 2. The top panel reports the interaction effects across the three terciles of coal-fired capacity in 1915. The bottom panel reports the interaction effects across the three terciles of pre-pandemic infant mortality (between 1915 and 1917). Standard errors are clustered at the city-level. ***,**,* denote significance at the 1%, 5%, and 10%, respectively.

Table 8: Robustness checks

| | | Estimated effect | | | | Alternative measures of capacity | | |
	Baseline	Control for weekly timing pandemic onset	Control for state × year FE	Control for city-level linear trend	Balanced panel	Capacity (100s of mw)	Within 50 miles	Within 100 miles
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: Log(infant mortality)								
I(Year=1918) ×								
Log(CoalCap30mile)	0.0324***	0.0266***	0.0328***	0.0304***	0.0275***	0.00836**	0.0351***	0.0370***
	[0.00831]	[0.00864]	[0.00959]	[0.0097]	[0.00899]	[0.00326]	[0.00961]	[0.0138]
Observations	1,611	1,611	1,611	1,611	1,290	1,611	1,611	1,611
R-squared	0.794	0.794	0.818	0.861	0.777	0.793	0.794	0.793
Number of clusters	183	183	183	183	129	183	183	183
Dependent variable: Log(all-age mortality)								
I(Year=1918) ×								
Log(CoalCap30mile)	0.0217***	0.0151*	0.0228***	0.0215***	0.0196**	0.00560	0.0306***	0.0289**
	[0.00777]	[0.00795]	[0.00857]	[0.0076]	[0.00819]	[0.00428]	[0.00858]	[0.0119]
Observations	1,610	1,610	1,610	1,610	1,290	1,610	1,610	1,610
R-squared	0.920	0.921	0.933	0.946	0.904	0.920	0.921	0.920
Number of clusters	183	183	183	183	129	183	183	183

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

A   Appendix

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

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

Table A.2: TSP Concentration 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

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

Table A.3: The effect of coal-fired capacity on excess mortality, by year

	Estimated effect			
	(1)	(2)	(3)	(4)
	Dependent variable:		**Dependent variable:**	
	Log(infant mortality)		**Log(all-age mortality)**	
Log(CoalCap30mile) ×				
I(Year=1915)	0.0021	0.0060	-0.0074	-0.0097
	[0.0188]	[0.0202]	[0.0108]	[0.0125]
I(Year=1916)	0.0019	0.0005	-0.0034	-0.0064
	[0.0177]	[0.0208]	[0.00950]	[0.0106]
I(Year=1917)	0.0094	0.0091	-0.0027	0.0001
	[0.0162]	[0.0187]	[0.0086]	[0.0099]
I(Year=1918)	0.0341***	0.0362***	0.0218**	0.0237**
	[0.0127]	[0.0137]	[0.00935]	[0.0109]
I(Year=1919)	0.0209*	0.0243*	0.0112*	0.0111
	[0.0116]	[0.0137]	[0.00671]	[0.00705]
I(Year=1921) – Omitted Category				
I(Year=1922)	-0.0004	-0.0029	-0.0030	-0.0058
	[0.0126]	[0.0145]	[0.0054]	[0.0053]
I(Year=1923)	-0.0120	-0.0066	0.0007	0.0011
	[0.0123]	[0.0146]	[0.0058]	[0.0060]
I(Year=1924)	0.0051	0.0072	0.0054	0.0065
	[0.011]	[0.012]	[0.0055]	[0.0064]
I(Year=1925)	-0.0047	-0.0039	0.0128**	0.0113*
	[0.0118]	[0.0133]	[0.0059]	[0.0063]
Observations	1,611	1,611	1,610	1,610
R-squared	0.797	0.820	0.922	0.934
Number of clusters	183	183	183	183
Full controls	Y	Y	Y	Y
Control for state × year FE		Y		Y

Notes: Each column reports the point estimates from a different regression. The models are estimated for the period 1915 to 1925 (exclusive of 1920, for which we lack data on mortality). The variable Log(CoalCap30mile) denotes the logarithm of coal-fired capacity within 30 miles of the city-centroid. All models include include the full set of controls reported in column (3) of Table 2. Columns (2) and (4) replace the state-year trend with a state-year fixed effect. Standard errors are clustered at the city-level. ***,**,* denote significance at the 1%, 5%, and 10%, respectively.