I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12177

What Explains Cross-City Variation in Mortality During the 1918 Influenza Pandemic? Evidence from 438 U.S. Cities

Karen Clay
Joshua Lewis
Edson Severnini

FEBRUARY 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12177

What Explains Cross-City Variation in Mortality During the 1918 Influenza Pandemic? Evidence from 438 U.S. Cities

Karen Clay
Carnegie Mellon University

Joshua Lewis
University of Montreal

Edson Severnini
Carnegie Mellon University and IZA

FEBRUARY 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

What Explains Cross-City Variation in Mortality During the 1918 Influenza Pandemic? Evidence from 438 U.S. Cities

Disparities in cross-city pandemic severity during the 1918 Influenza Pandemic remain poorly understood. This paper uses newly assembled historical data on annual mortality across 438 U.S. cities to explore the determinants of pandemic mortality. We assess the role of three broad factors: i) pre-pandemic population health and poverty, ii) air pollution, and iii) the timing of onset and proximity to military bases. Using regression analysis, we find that cities in the top tercile of the distribution of pre-pandemic infant mortality had 21 excess deaths per 10,000 residents in 1918 relative to cities in the bottom tercile. Similarly, cities in the top tercile of the distribution of proportion of illiterate residents had 21.3 excess deaths per 10,000 residents during the pandemic relative to cities in the bottom tercile. Cities in the top tercile of the distribution of coal-fired electricity generating capacity, an important source of urban air pollution, had 9.1 excess deaths per 10,000 residents in 1918 relative to cities in the bottom tercile. There was no statistically significant relationship between excess mortality and city proximity to World War I bases or the timing of onset. Together the three statistically significant factors accounted for 50 percent of cross-city variation in excess mortality in 1918.

JEL Classification: N32, N52, N72, Q40, Q53, O13

Keywords: influenza, pandemic, mortality, air pollution

Corresponding author:
Edson Severnini
Heinz College
Carnegie Mellon University
4800 Forbes Avenue
Pittsburgh, PA, 15213
USA

E-mail: edsons@andrew.cmu.edu

Introduction

The 1918-1919 Influenza Pandemic was a global health catastrophe that is estimated to have killed 50 million people worldwide. The pandemic spread rapidly across the United States during the fall of 1918, killing more Americans than all wars in the twentieth century. In the U.S. and elsewhere there were significant cross-city differences in pandemic severity. Although there has been considerable speculation regarding the factors that contributed to pandemic severity, only a small number of studies have examined the determinants of cross-city severity. Acuna-Soto et al. (2011) combine data for 66 cities to investigate the ability of pre-pandemic pneumonia and influenza mortality, city size, longitude, and latitude to explain pandemic pneumonia and influenza mortality in 1918. Bootsma and Ferguson (2007) and Markel et al (2007) examine the effects of public health measures on pandemic severity drawing on data for the 43 cities for which weekly influenza and pneumonia mortality data is available. Grantz et al. (2016) and Tuckel et al. (2006) detail temporal and spatial data from Chicago and Hartford to examine within city variation in mortality. Their analysis points to the importance of poverty-related factors as contributing to pandemic severity.

We investigate the determinants of cross-city differences in pandemic severity. The paper relies on a new dataset of annual mortality in 438 U.S. cities that represent two-thirds of the urban population for the period 1915 to 1925. The panel structure of the dataset allows us to construct a measure of pandemic severity for a large sample of cities. The empirical analysis involves two steps. First, we estimate excess pandemic mortality in every city as the difference between observed and predicted mortality in 1918. Second, we estimate cross-sectional regressions to

assess the importance of three broad determinants of excess pandemic mortality across cities: i) measures of pre-pandemic health and poverty, ii) the timing of onset and proximity to military bases, and iii) air pollution. The first two factors have been discussed in the historical and medical literatures. The third has received far less attention, although there is growing biological (e.g., Jakab 1993; Jaspers et al., 2005), animal (e.g., Hahon et al., 1985; Harrod et al., 2003; Lee et al., 2014), and epidemiological evidence that air pollution can increase susceptibility to influenza (e.g., Ciencewicki and Jaspers 2007). Recent empirical evidence suggests that air pollution interacts with infectious disease. Hanlon (2018) finds that the higher underlying rates of measles and tuberculosis (TB) increased the mortality effects of pollution episodes in London from 1866 to 1965. Similarly, Clay, Lewis, and Severnini (2018) show that coal-fired generating capacity led to significantly higher mortality rates during the pandemic.[1]

This study builds on and complements previous statistical analyses of the factors influencing cross-city variation in mortality in 1918. The dataset includes a much larger sample of cities than has been previous studied, allowing us to examine multiple factors simultaneously to determine their relative importance. Motivated by the existing historical and medical literatures, our analysis focuses on three broad categories that may have influenced pandemic severity across cities: underlying population health and poverty, the timing of onset, and air pollution, although we also explore the role of additional factors related to trade, religious homogeneity, and public health infrastructure. By quantifying the influence of several distinct sociodemographic and

[1] Whereas Clay, Lewis, and Severnini (2018) focus more narrowly on the impact of coal capacity, this paper explores the role of multiple underlying determinants of cross-city pandemic mortality, and seeks to quantify their respective influence on excess mortality in 1918. Additionally, our expanded sample of 438 cities (versus 180 cities) allows us to assess these relationships over a much broader population.

environmental factors on pandemic mortality, our analysis sheds new light on the disparities in death rates across U.S. cities during pandemic, and may offer insights into how policymakers should allocate resources in response to future pandemics.

Methods

Data

Data on all-cause deaths for the years 1915 to 1925 were assembled from the *Mortality Statistics* for the 438 cities with at least 10,000 residents in 1920 mapped in Figure S1. We combined these data with city population in 1920 to construct the mortality rate per 10,000 city residents. Cities are linked to county-level pre-pandemic demographic and economic characteristics from the 1910 census of population and manufacturing (Haines and ICPSR, 2010). We assembled the following variables: percent urban residents, percent illiterate, percent foreign born, percent homeowner, and percent of employment in manufacturing. Information on the week of pandemic onset was obtained from Sydenstricker (1918). City proximity to military bases was obtained from the U.S. War Department (1919). Data on coal-fired generating capacity were obtained from the U.S. Department of Agriculture (1916). We collected information on all coal-fired power stations with at least 5 megawatts of installed capacity, and calculated total coal-fired

capacity within a 30-mile radius of each city-centroid.[2,3] Figure S1 also displays coal power plants and hydroelectric dams in 1915 by terciles of electricity generating capacity.

Regression Analysis

We use annual data on all-cause mortality for the years 1915 to 1925 to derive an estimate of excess mortality in 1918 for the sample of cities. Excess mortality is calculated based on the difference between observed mortality in 1918 and an expected baseline mortality level (absent the pandemic). This approach has been widely used to estimate the impact of influenza epidemics (e.g., Serfling, 1963; Housworth and Langmuir, 1974; Vibroud et al., 2004; Olson et al., 2005). First, we construct a measure of predicted 1918 mortality (absent the pandemic), based on a city-specific linear trend for the period 1915 to 1925. We exclude the year 1918 from the analysis, and estimate the following regression model:

$$M_{ct} = \alpha_c + \beta_c t + e_{ct},$$

where M_{ct} denotes all-cause mortality in city c in year t, the variable t represents year, and e_{ct} denotes an error term.[4] The coefficient α_c is a city-specific intercept that allows for different

[2] This distance was chosen to capture the fact that the majority of power plant emissions were dispersed locally (Levy et al., 2012; Seinfeld and Pandis, 2012). Qualitatively similar results were found when the analysis was run with a 50-mile radius. Regressions estimates available upon request.

[3] We do not consider interventions during the pandemic, because data on these interventions are only available for a small number of cities (e.g., Bootsma and Ferguson, 2007; Markel et al., 2007).

[4] The year 1918 is excluded from the analysis to ensure that mortality during the pandemic does not influence the estimation of the city trend in mortality. We also explore the sensitivity of the

baseline mortality rates across cities, and the coefficient β_c allows for different trends in mortality across cities. The slope and intercept estimates are used to predict 1918 mortality by city. Excess mortality in 1918 is constructed as the difference between actual and predicted mortality in city c in 1918.

We estimate multivariate regressions that relate excess mortality to pre-pandemic infant mortality, measures of city poverty, timing of pandemic onset, and air pollution. Controls for percent urban are included in all specifications. All explanatory variables included in the main regression model were selected by F-test and partial R-squared, as reported in Tables S3, S5, and S7. For each broad category of predictors, the two variables with the strongest explanatory power were kept in the model. They had large F-statistic, p-value for the F-test smaller than 0.05, and non-negligible partial R-squared. All explanatory variables are included in a tercile specification (high vs. low, middle vs. low). This specification is more flexible than using the variables continuously as it allows for nonlinear relationships, and the regression coefficients for each explanatory variable are easier to interpret, since the coefficients for high (middle) reflect the difference in pandemic mortality for cities in the high (middle) tercile relative to cities in the low tercile.[5] All estimated standard errors are robust to heteroscedasticity.[6]

---

results to alternative measures of excess pandemic mortality, which do not depend on post-1918 mortality rates.

[5] Qualitatively similar results were found when the analysis was run with continuous explanatory variables. Regressions estimates available upon request.

[6] Conley standard errors are also reported in Table 1. The Conley method allows for outcomes to be correlated among nearby cities, with the degree of correlation declining linearly until some cutoff distance (Conley 1999). We allow for spatial correlation up to 200 miles. Because the Conley standard errors are generally smaller than the robust standard errors, we report the more conservative robust standard errors throughout the analysis.

Results

Figure 1a shows mortality in our 438 sample cities. Mortality rates rose sharply in 1918, exceeding their pre-pandemic level by 35 percent. Figure 1b shows cross-city variation in excess mortality in 1918. Again, excess mortality is calculated as the difference between observed all-cause mortality in 1918 and predicted all-cause mortality in 1918, based on a city-specific linear trend. This variable reflects the extent to which city mortality rates in 1918 differed from their predicted values.[7] The median city experienced excess mortality of 57.7 per 10,000 residents. Applying the estimates of excess mortality across the entire U.S. population, we calculate that the pandemic was responsible for 615,000 American deaths, similar to previous estimates of pandemic severity (Crosby, 1989, p. 206). There was wide variation in pandemic mortality across cities. The inter-quartile range for excess mortality is 38.5 - 78.0.

Pre-pandemic Health and Poverty

Our analysis examines the relationship between pre-pandemic health and poverty and excess mortality in 1918 across the sample of U.S. cities. Previous authors have used different measures of population health to predict mortality in 1918. Acuna-Soto et al. (2011) examine the relationship between pre-pandemic influenza and pneumonia mortality and pandemic influenza and pneumonia mortality, finding that pre-pandemic and pandemic pneumonia mortality are highly correlated. Bootsma and Ferguson (2007) show that 1918 mortality is correlated with 1917 mortality. Other researchers have explored the relationship between poverty markers and

[7] This measure of excess mortality will not be affected by misclassification in the cause of death records (see Vital Statistics 1957, pp. 18-26).

pandemic severity. Grantz et al. (2016) examine the ability of percent illiterate, percent homeowners, percent unemployed and population density to predict pandemic influenza mortality across census tracts in Chicago. Tuckel et al. (2006) explore the relationship between the percent foreign born and ward-level influenza mortality in Hartford.

Our analysis builds on the previous research, examining the relationship between a number of socioeconomic variables – percent foreign born, percent illiterate, percent homeowners, and percent urban – and excess 1918 mortality across a much larger sample of cities. In addition, we include the infant mortality rate in the years 1915 and 1916 as an explanatory variable. The infant mortality rate is widely used as a measure of population health, since the link between infant deaths and contemporaneous health conditions – including disease, pollution, and nutrition – is immediate, whereas adult mortality reflects an accumulation of lifetime exposure (Chay and Greenstone, 2003; Currie and Neidell, 2005).

Table S1 shows the correlation among selected pre-pandemic measures of health and poverty. There is a strong correlation across many of the explanatory variables, suggesting that they are all capturing factors related to health and poverty. For example, the infant mortality rate was elevated in cities with a higher percent of foreign residents, lower rates of home ownership, and higher rates of illiteracy. The results also show that percent urban is correlated with percent foreign born and percent homeowner, although baseline infant mortality is largely unrelated to city size.

Table S2 reports the multivariate regression estimates. All models include controls for percent urban. Consistent with Acuna-Soto et al (2011), cities in more urban counties had statistically significantly lower mortality rates. This may have occurred for a number of reasons. More urban areas may have greater exposure to the milder spring wave of influenza and so have greater immunity. More urban areas also may have been more able to implement non-pharmaceutical interventions such as isolation and quarantine of victims, school closure, and cancelation of public gatherings (e.g., Bootsma and Ferguson, 2007; Markel et al., 2007). Pre-pandemic infant mortality rates are positively and statistically significantly related to excess 1918 mortality. Markers for city poverty are generally associated with higher pandemic mortality (cols 2, 4, and 6), although once we control for baseline infant mortality, the estimates on the various markers of city poverty decrease in magnitude (cols 3, 5, and 7). Together, the results suggest that baseline population health had an impact on pandemic severity independent of other poverty markers. In contrast, it appears that much of the relationship between poverty and pandemic mortality can be explained by the poor health in low-income populations. These findings suggest that other characteristics associated with urban poverty, such as the higher rates of disease transmission in crowded neighborhoods, may have been less important determinants of mortality during the pandemic. Guided by these results, we include percent urban, infant mortality in 1915-1916, and percent illiterate in the main specification in Table 1, discussed below.

Timing of Pandemic

The timing of pandemic onset is thought to be an important predictor of mortality, because the virulence may have declined over time (Crosby, 1989; Barry, 2004). The movement of military personnel is also believed to have influenced severity through its role in spreading the virus

across the country (see Crosby, 1989; Kolata, 2001; Barry, 2004; and Byerly, 2010 for accounts of the pandemic in the military).

We assess the impact of the timing of onset and city proximity World War I military bases on pandemic severity. Consistent with the historical narrative, the pandemic arrived earlier to cities near a military base (Table S1). Table S4 shows the relationship between the two explanatory variables and excess mortality in 1918. We find some evidence that proximity to World War I bases affected pandemic severity. The coefficient estimates for cities in the high tercile are positive and generally statistically significant. In contrast, there is no consistent relationship between the week of pandemic onset and excess mortality. Given these results, we focus on proximity to the World War I base as the main explanatory variable in the regressions in Table 1 below.

Coal-Fired Electricity Capacity

We assess the impact of city-level air pollution on pandemic severity. Our analysis is motivated by an emerging body of evidence suggesting that air pollution may exacerbate pandemic mortality. 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 (Jakab 1993; Jaspers et al., 2005). Ciencewicki and Jaspers (2007) review a number of epidemiological studies showing associations between exposure to air pollutants and increased risk for respiratory virus infections. Similarly, both Hanlon (2018) and Clay, Lewis, and Severnini (2018) find

evidence that pollution interacts with infectious disease to affect mortality, both in London and during the 1918 pandemic.

Historical evidence suggests that air pollution was severe and varied widely across cities (Flagg, 1912; Ives et al., 1936; Stern, 1982). Average levels of total suspended particulates (TSP) across a sample of 15 large American cities was seven times higher than the annual thresholds initially set under the Clean Air Act Amendments of 1970. Electricity generation was a significant contributor to urban air pollution. A 1912 study of Chicago found that electricity-generating plants accounted for 44 percent of visible smoke (Goss, 1915). In addition, there was wide variation in coal-fired generating capacity across cities depending on local availability of coal and the proximity to hydroelectric power.

We use coal-fired capacity and percent of employment in manufacturing as two different proxies for city air pollution.[8] These two variables are positively related (Table S1), since locations with more abundant coal resources tended to have larger manufacturing sectors. There is a positive relationship between coal-fired capacity and excess pandemic mortality, and the coefficient estimates on the high tercile are statistically significant and stable across the different specifications (Table S6). The coefficients are only modestly reduced once pre-pandemic infant mortality is included as a control (col. 1 and 2), indicating that effects arose primarily through a direct contemporaneous link rather than through the indirect impact of coal capacity on baseline health. There is some evidence that the manufacturing employment share was related to

[8] Coal was often burned as part of the manufacturing process. Manufacturing may also have impacted pandemic severity through the close working conditions.

pandemic severity, although this relationship weakens once infant mortality is included as a control. Given these findings, we include coal-fired generating capacity as our preferred measure of air pollution in the main regressions in Table 1 below.

Multiple Determinants of Pandemic Mortality

We take advantage of the large sample of cities to explore the joint influence of pre-pandemic health and poverty, pandemic timing, and pollution on excess mortality in 1918 in a multivariate regression framework. The estimates are presented in Table 1. For reference, columns 1-4 report the individual estimates for each variable, controlling for percent urban. Column 5 reports the results from the multi-factor model. When all of the covariates are included, the coefficients on the top and middle terciles of proximity to World War I bases are statistically insignificant. Although the point estimates are reduced, high coal cities experienced statistically significantly higher mortality rates in 1918. Meanwhile, the estimated effects for pre-pandemic infant mortality and percent illiterate are large and statistically significant for both the high and medium terciles.

Pre-pandemic health and poverty and coal capacity appear to have been significant determinants of mortality during the pandemic. Figure 2 reports the corresponding magnitudes for the main explanatory variables from Table 1, column 5. The R-squared from the regression model implies that the three factors accounted for 25 percent of the total cross-city variation in excess pandemic mortality. This is despite the fact that the tercile specification limits the explanatory power of the model, since the framework does not exploit the substantial *within*-tercile variation in each explanatory variable.

To quantify the role of pre-pandemic health, air pollution, and proximity to World War I bases on pandemic severity, we re-estimate the distribution of excess mortality across cities under several alternative counterfactual scenarios. In particular, we use regression coefficients (Table 1, col. 5) to ask the following question: what is the counterfactual distribution of excess pandemic mortality if the explanatory variables for cities in the high and middle terciles are each reduced to the low tercile. This counterfactual exercise would probably have been infeasible. For example, cities with high pre-pandemic infant mortality may have lacked access to the public health and medical resources necessary to improve public health. Nevertheless, the analysis allows us to quantitative assess the importance of the various determinants of pandemic mortality, and may provide insight into the benefits of investments – such as poverty reduction campaigns or air pollution abatement – that could be made in modern developing countries to mitigate the harm from future pandemics.

We simulate the counterfactual excess mortality distribution under four scenarios: a) reducing pre-pandemic infant mortality in all cities to the low tercile, b) reducing coal capacity in all cities to the low tercile, c) decreasing World War I base proximity to the low tercile, and d) policies a) through c) and decreasing the percent illiterate to the low tercile. Figure 3a reports the effects for the reduction in pre-pandemic infant mortality. This change would have led to a 17 percent decrease in average excess mortality in 1918 across the sample. The magnitude of this decrease is striking given that mortality in one third of cities – those in the lowest tercile of pre-pandemic infant mortality – remains unchanged under the policy change. The counterfactual mortality distribution is compressed, indicating that this scenario would have resulted in a decrease in

pandemic mortality disparities across cities. Figure 3b shows the counterfactual distribution for coal-fired capacity. This policy change is associated with an 8 percent reduction in excess mortality, although the shape of the mortality distribution remains similar. Meanwhile, a decrease in city proximity to World War I bases is associated with a modest 5 percent decrease in average pandemic mortality and has little impact on the shape of the mortality distribution (Figure 3c). Figure 3d reports the distribution under the fourth counterfactual in which all factors are reduced to the low tercile. This policy would have resulted in a 50 percent decrease in pandemic mortality and substantial narrowing of the cross-city distribution. The magnitude of the mortality decrease demonstrates the importance of pre-pandemic mortality, sociodemographic factors captured by the percent illiterate, and to a lesser extent the influence of coal capacity, in influencing the severity of the pandemic. The results also highlight the independent influence of each of these factors. The broad leftward shift of the mortality distribution shows that the three factors affected pandemic severity across a large segment of the urban population, and that the relationship was not confined to a handful of heavily polluted cities with high pre-pandemic mortality and low literacy rates. In fact, less than 9 percent of cities fell into the lowest tercile for all three explanatory variables, indicating that there was scope to mitigate pandemic mortality across the vast majority of American cities.

Robustness Checks and Alternative Determinants of Pandemic Mortality

Table 2 reports the results from a series of robustness exercises. Columns 1 and 2 report the results from principal component analysis for each of the three broad categories: health and

poverty, coal, and pandemic timing.[9] Consistent with the baseline findings, factors related to both health and poverty and coal are associated with significantly higher rates of pandemic mortality, whereas there are no significant differences based on the timing of onset. Figure S2 shows that the relative coefficient magnitudes mirror those from baseline findings: the estimates for health and poverty exceed those for coal, which exceed those for timing.

Columns 4-7 explore the sensitivity of the results to alternate measures of excess mortality in 1918 (for reference, column 3 reports the baseline estimates). We find similar effects for mortality rates derived based on 1910 city population (col. 4).

One concern is that the pandemic influenced post-1918 mortality trends through selective mortality, by killing less-healthy individuals who would have died in subsequent years. This hypothesis is supported by Noymer and Garenne (2000) and Noymer (2011), who find that the pandemic led to decreases in TB mortality in post-1918 years. We take several steps to address this concern. First, we re-estimate excess mortality in 1918, excluding the years 1918-1920 from the trend calculation; second, we calculate excess mortality as the deviation from the average morality rates pre-pandemic (1915-1917); third, we recalculate excess mortality as the deviation from the predicted mortality trend based solely on the period 1915-1917.[10] The coefficient estimates based on these alternative measures are similar to the original estimates (cols. 5-7).

---

[9] For each category we calculate the principal component – the latent variable that accounts for the largest variance – based on a linear combination of variables included in Tables S.2, S.4, and S.6. We then estimate the role of each of these three factors in explaining excess mortality in 1918. The estimates in column (1) are based on factor components constructed based on a linear specification of the underlying explanatory variables; the estimates in column (2) are based on factor components derived from the middle and high tercile specification of the independent variables.

[10] All three measures are highly correlated with the original measure of excess mortality (Figure S2), although given the limited number of years over which the pre-trend is calculated, estimates based on the third approach are significantly noisier.

Our analysis has focused on the role of factors related to health and poverty, air pollution, and the timing of onset in explaining the wide variation in cross-city mortality during the 1918 pandemic. The focus is motivated by the historical and medical literature, and these factors account for a significant fraction of the cross-city variation in mortality. Notwithstanding these results, half of the cross-city differences in excess mortality remain unexplained.

To conclude the empirical analysis, we explore the other potential sources of cross-city differences in pandemic severity (Table 3). In columns 1 and 2, we explore the role of religiosity and religious homogeneity in influencing pandemic severity.[11] We find no significant differences according to church membership, although there is some evidence that greater religious fractionalization is associated with higher pandemic mortality. These results could suggest that higher levels of population homogeneity may have aided the local response to the pandemic or that concentrated leadership of religious officials mitigated its impacts. In contrast, we find no evidence that pre-pandemic local public health infrastructure had any impact on pandemic mortality (cols. 5-7), consistent with local public health response having been overwhelmed by the magnitude of the pandemic (Crosby, 1989). Finally, we find some evidence that greater access to trade, as measured by total miles of railway in 1911, is associated with increased pandemic severity, consistent with recent evidence on the role of transportation in accelerating the spread of influenza (Adda, 2016). While not exhaustive, this evidence points to potentially fruitful new areas of scholarship on the 1918 pandemic.

[11] Data on religious membership come from the 1916 Census of Religion. To measure religiosity, we calculate fraction of individuals who belong to a religious organization. We measure religious fractionalization as one minus the square of the shares of each of the ten major religious denominations (Alesina et al., 2003).

Discussion

This study examined the determinants of excess mortality during the 1918 Influenza pandemic across a large sample of U.S. cities. We found that cities with higher pre-pandemic infant mortality rates and more illiterate residents had statistically significantly higher mortality rates in 1918, suggesting that low levels of health and poverty contributed to pandemic severity. These results correspond with Grantz et al.'s (2016) findings for Chicago, and Noymer (2011), who found that the pandemic disproportionately affected individuals with tuberculosis, who were disproportionately poor. In contrast, there is little evidence that the timing of pandemic onset influenced excess mortality in 1918. Lastly, we found that cities with high levels of air pollution, as measured by coal-fired capacity, experienced significant higher mortality rates during the pandemic.

These findings not only improve our understanding of the disparities in cross-city pandemic severity during the 1918 Spanish Influenza Pandemic, but also shed light on factors that might have mediated the long-run effects of that pandemic found by Almond (2006), Beach, Ferrie, and Saavedra (2018), Fletcher (2018), and Ogasawara (2018), and the long-term impacts of the 1889 Russian Influenza Pandemic found by Riggs and Cuff (2013).

This study has several limitations. First, although we have done considerable exploration of alternative factors, this study does not definitely demonstrate causality of any of the factors. It may be possible to establish a causal relationship in the context of later pandemics or annual influenza in later periods, but this analysis is limited by data availability and the uniquely severe

nature of the 1918 pandemic. Second, because excess mortality may capture several causes of death, one should use caution in interpreting our findings. That measure could include mortality due to influenza, or a combination of influenza and other diseases such as tuberculosis (e.g., Noymer 2009, 2011). Unfortunately, data (un)availability precludes us to estimate effects on different causes of death. Third, our measure of air pollution, coal-fired electricity generation is limited by data availability. Systematic air pollution monitor data did not become available in the U.S. until the 1950s and coverage is sparse until the 1970s. Finally, while we have tried to address a range of factors identified as important in a variety of context, our analysis may not capture all of the factors that contributed to excess mortality for cities in our sample. In particular, our analysis does not account for local public interventions in response to the pandemic that could have influenced the variation in excess mortality across cities.

Understanding the social and environmental determinants of mortality during the 1918 Influenza Pandemic can provide useful insights for the policy response to future pandemics. In particular, our findings on the relationship between pre-pandemic health, poverty and pandemic mortality may have implications for the distribution of scarce medical resources across locations during a future outbreak. The relationship between coal capacity and pandemic mortality may be particularly relevant in modern developing countries, where urban pollution is severe and comparable to the levels in the early 20th century America. Despite dramatic improvements in the quality of medical care and public health infrastructure in the hundred years since the pandemic, the risks of a future outbreak are significant and are unlikely to be met by existing medical infrastructure. As Taubenberger and Morens (2006, p. 77) note: "Even with modern antiviral and antibacterial drugs, vaccines, and prevention knowledge, the return of a pandemic virus

equivalent in pathogenicity to the virus of 1918 would likely kill >100 million people worldwide. A pandemic virus with the (alleged) pathogenic potential of some recent H5N1 outbreaks could cause substantially more deaths."

Acknowledgements

Karen Clay and Edson Severnini acknowledge financial support from the Heinz College at Carnegie Mellon University, and from the National Science Foundation Grant SES-1627432. Joshua Lewis acknowledges financial support from the University of Montreal.

References

Acuna-Soto, Rodolfo, Cecile Viboud, and Gerardo Chowell. "Influenza and pneumonia mortality in 66 large cities in the United States in years surrounding the 1918 pandemic." *PLoS One* 6.8 (2011): e23467.

Adda, Jerome. "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data." *Quarterly Journal of Economic* 131 No. 2 (2016): 891-941.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. "Fractionalization." *Journal of Economic Growth* 8 (2003): 155-194.

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.

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

Beach, Brian, Joseph P. Ferrie, and Martin H. Saavedra. “Fetal Shock or Selection? The 1918 Influenza Pandemic and Human Capital Development.” *NBER Working Paper* No. 24725 (2018).

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.

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 Economic* 118 (2003): 1121-1167.

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

Clay, Karen, Joshua Lewis, and Edson Severnini. “Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic.” *Journal of Economic History* 78 No. 4 (2018): 1179-1209.

Conley, Timothy G. 1999. “GMM Estimation with Cross-Sectional Dependence.” *Journal of Econometrics* 92(1): 1-45.

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

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.

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

Fletcher, Jason M. “The effects of in utero exposure to the 1918 influenza pandemic on family formation.” *Economics & Human Biology* 30 (2018): 59-68.

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.

Hanlon, W. Walker. "London Fog: A Century of Pollution and Mortality, 1866-1965." NBER Working Paper No. 24488, 2018.

Harrod, Kevin S., Richard J. Jaramillo, Cynthia L. Rosenberger, Shan-Ze Wang, Jennifer A. Berger, Jacob D. McDonald, et al. "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.

Housworth, Jere and Alexander D. Langmuire. "Excess Mortality from Epidemic Influenza, 1957-1966." *American Journal of Epidemiology* 100 No. 1 (1974): 40-48.

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: U.S. 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, et al. "Diesel Exhaust Enhances Influenza Virus Infections in Respiratory Epithelial Cells." *Toxicology Sciences* 85 No. 2 (2005): 990-1002.

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, et al. "Exposure to combustion generated environmentally persistent free radicals enhances severity of influenza virus infection." *Particle and Fibre Toxicity* 11 No. 1 (2014): 57.

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, et al. "Nonpharmaceutical Interventions Implemented by US Cities During the 1918-1919 Influenza Pandemic." *JAMA* 298 No. 6 (2007): 644-654.

Noymer, Andrew and Michel Garenne. "The 1918 Influenza Epidemic's Effects on Sex Differentials in Mortality in the United States." *Population and Development Review* 26 No. 3 (2000): 565-581.

Noymer, Andrew. "The 1918 influenza pandemic hastened the decline of tuberculosis in the United States: an age, period, cohort analysis." *Vaccine* 29 (2011): B38-B41.

Noymer, Andrew. "Testing the influenza-tuberculosis selective mortality hypothesis with Union Army data." *Social Science & Medicine* 68 No. 9 (2009): 1599-608.

Ogasawara, Kota. “The long-run effects of pandemic influenza on the development of children from elite backgrounds: Evidence from industrializing Japan.” *Economics & Human Biology* 31 (2018): 125-137.

Olson, Donald R., Lone Simonsen, Paul J. Edelson, and Stephen S. Morse. “Epidemiological Evidence of an Early Wave of the 1918 Influenza Pandemic in New York City.” *PNAS* 102 No. 31 (2005): 11059-63.

Riggs, Paul, and Timothy Cuff. “Ladies from Hell, Aberdeen Free Gardeners, and the Russian influenza: An anthropometric analysis of WWI-era Scottish soldiers and civilians.” *Economics & Human Biology* 11 No. 1 (2013): 69-77.

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

Serfling, Robert E. “Methods for Current Statistical Analysis of Pneumonia-Influenza Deaths.” *Public Health Reports* 78 (1963): 494-506.

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

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.

Vibroud, Cécile, Pierre-Yves Boelle, Khashayar Pakdaman, Fabrice Carrat, Alain-Jacques Valleron, and Antoine Flahault. "Influenza Epidemics in the United States, France, and Australia, 1972-1997." *Emerging Infectious Disease* 10 No. 1 (2004): 32-39.

Figures

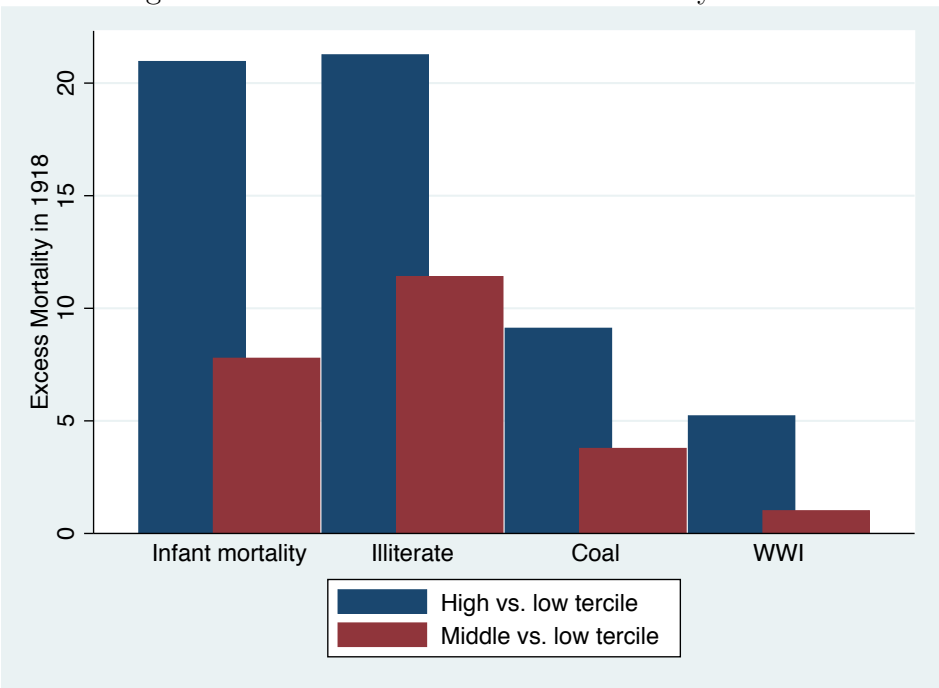
Figure 1: All-age Mortality

(a) Mortality, 1915-1925

(b) Excess Mortality in 1918

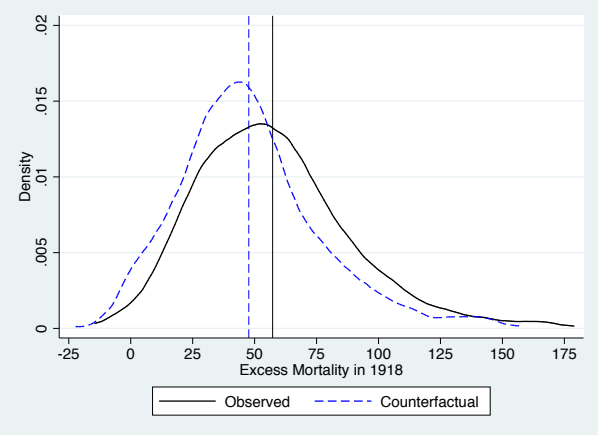
Notes: Panel (a) reports the all-age mortality rate per 10,000 city residents. Panel (b) reports the distribution of excess all-age mortality in 1918 across cities. Excess mortality is calculated as the difference between observed mortality and predicted mortality in 1918, where predicted mortality is calculated based on a linear city-specific trend. The solid and dotted lines denote the mean and interquartile range of excess mortality in the sample.

Figure 2: Determinants of Excess Mortality in 1918

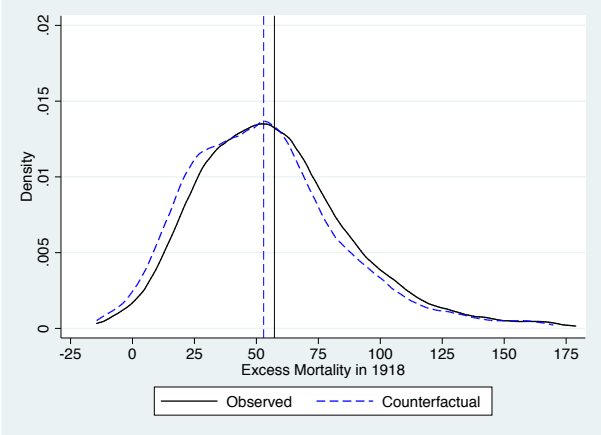


Notes: This figure reports the coefficient estimates from Table 1 col. 5. Excess mortality is calculated as the difference between observed mortality and predicted mortality in 1918, where predicted mortality is calculated based on a linear city-specific trend.

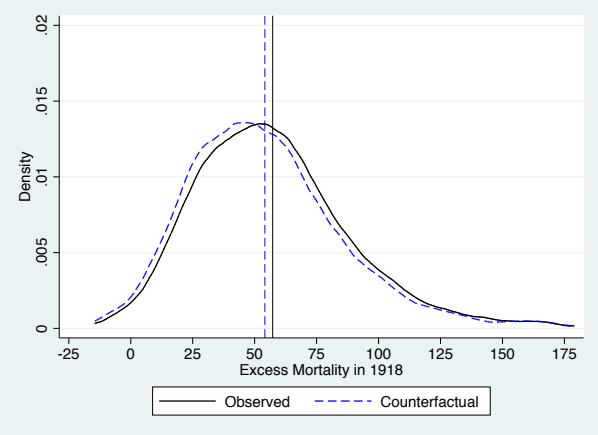
Figure 3: Excess Mortality and Counterfactual Excess Mortality in 1918



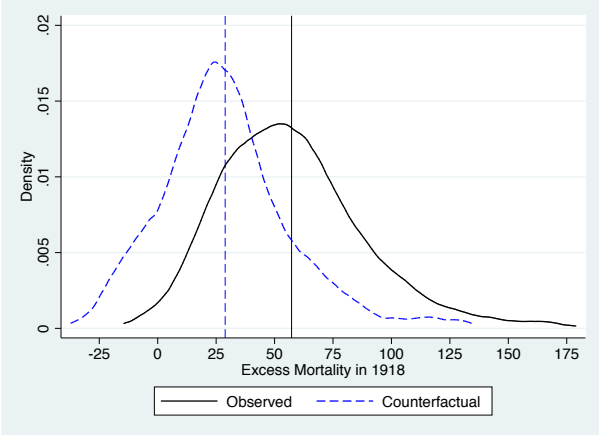
(a) Baseline Infant Mortality



(b) Coal Capacity



(c) Proximity to World War I Bases



(d) All Factors

Notes: This figures reports the distribution of excess mortality and counterfactual excess mortality in 1918 across the sample of 438 cities. Panel (a) assumes that baseline infant mortality was reduced to the lowest tercile. Panel (b) assumes that coal-fired capacity was reduced to the lowest tercile. Panel (c) assumes that proximity to World War I bases was reduced to the lowest tercile. Panel (d) combines (a) through (c) and assumes that fraction illiterate was reduced to the lowest tercile. Vertical lines denote mean excess mortality rates.

Tables

Table 1: Determinants of Excess Mortality in 1918

	Dependent variable: Excess Mortality in 1918				
	(1)	(2)	(3)	(4)	(5)
% urban residents in 1910					
High vs. low	-7.556**	-12.37***	-6.937*	-14.03***	-18.46***
	[3.420]	[3.434]	[4.104]	[4.825]	[4.176]
	(3.400)	*(3.414)*	*(4.080)*	*(4.798)*	*(4.123)*
Middle vs. low	-7.725**	-5.649	-6.522*	-8.484**	-8.181**
	[3.501]	[3.572]	[3.845]	[3.890]	[3.464]
	(3.481)	*(3.552)*	*(3.823)*	*(3.868)*	*(3.420)*
Infant mortality, 1915-1916					
High vs. low	31.35***				20.96***
	[3.408]				[4.033]
	(3.389)				*(3.982)*
Middle vs. low	13.88***				7.782**
	[3.063]				[3.156]
	(3.046)				*(3.116)*
% illiterate in 1910					
High vs. low		33.09***			21.26***
		[3.461]			[3.993]
		(3.442)			*(3.942)*
Middle vs. low		19.35***			11.41***
		[3.345]			[3.693]
		(3.326)			*(3.646)*
Proximity to WWI base					
High vs. low			7.078*		5.225
			[4.088]		[3.825]
			(4.065)		*(3.776)*
Middle vs. low			2.223		1.012
			[3.608]		[3.456]
			(3.588)		*(3.413)*
Coal capacity					
High vs. low				15.89***	9.116**
				[4.991]	[4.472]
				(4.962)	*(4.416)*
Middle vs. low				7.230*	3.778
				[3.706]	[3.428]
				(3.685)	*(3.384)*
Observations	438	438	438	438	438
R-squared	0.177	0.178	0.016	0.035	0.253

Notes: The dependent variable is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. The coefficient estimates report difference for the middle and high tercile of each explanatory variable, relative to the lowest tercile. Columns 1-4 report the individual estimates for each variable, controlling for percent urban. Column 5 reports the results from the multi-factor model. Robust standard errors are reported in brackets. *** represents statistical significance at 1 percent level, ** 5 percent level, and * 10 percent level. (For comparison, Conley standard errors are reported in parentheses (in italic). The Conley method allows for outcomes to be correlated among nearby cities, with the degree of correlation declining linearly until some cutoff distance, 200 miles in this case.)

Table 2: Alternative Specifications

	Dependent variable: Excess Mortality in 1918						
	Principal component analysis		Alternate measures of excess mortality				
			Baseline estimates	1910 city pop denom	Construct mort counterfactual without years 1918-1920	Construct mort counterfactual from 1915-1917 mean	Construct mort counterfactual from 1915-1917 trend
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Health & poverty	9.404***						
	[1.079]						
High vs. low		11.09***					
		[1.185]					
Middle vs. low		3.128**					
		[1.357]					
Coal	2.384*						
	[1.428]						
High vs. low		2.913*					
		[1.667]					
Middle vs. low		1.624					
		[1.502]					
Pandemic timing	0.999						
	[1.469]						
High vs. low		-0.682					
		[1.642]					
Middle vs. low		1.857					
		[1.647]					
Infant mortality, 1915-1916							
High vs. low			20.96***	17.26***	19.27***	10.77***	13.77***
			[4.033]	[5.275]	[4.033]	[3.975]	[4.418]
Middle vs. low			7.782**	5.457	6.666**	2.219	0.534
			[3.156]	[4.469]	[3.173]	[3.114]	[3.363]
% illiterate in 1910							
High vs. low			21.26***	29.93***	21.91***	21.43***	19.58***
			[3.993]	[5.359]	[4.003]	[3.966]	[4.599]
Middle vs. low			11.41***	16.29***	11.61***	10.82***	7.239*
			[3.693]	[5.116]	[3.705]	[3.618]	[3.948]
Proximity to WWI base							
High vs. low			5.225	5.129	4.073	3.418	9.174**
			[3.825]	[5.047]	[3.893]	[3.815]	[4.044]
Middle vs. low			1.012	-2.501	0.729	0.770	7.200*
			[3.456]	[5.121]	[3.509]	[3.448]	[3.767]
Coal capacity							
High vs. low			9.116**	16.21**	9.177**	9.936**	8.950*
			[4.472]	[6.322]	[4.504]	[4.499]	[4.715]
Middle vs. low			3.778	7.149	3.873	4.665	4.499
			[3.428]	[4.650]	[3.474]	[3.388]	[3.665]
Observations	438	438	438	435	438	438	438
R-squared	0.148	0.212	0.253	0.200	0.238	0.181	0.180

Notes: The dependent variable is calculated as the difference between observed and predicted all-age mortality in 1918. Columns 1 and 2 report the estimates for explanatory variables constructed from principal components analyses of the underlying explanatory variables in Tables S2, S4, and S6. The main component from each group of variables is included in the regression. Columns 4-7 report the results based on alternative measures of excess mortality. In column 4, mortality rates are constructed based on 1910 city population; in column 5, predicted 1918 mortality is constructed based on a linear city-specific trend from 1915-1925 (excluding years 1918-1920); in column 6, predicted 1918 mortality is calculated as average city-level mortality for the years 1915-1917; in column 7, predicted 1918 mortality is extrapolated from a linear city-specific trend for the years 1915-1917. Robust standard errors are reported in brackets. *** represents statistical significance at 1 percent level, ** 5 percent level, and * 10 percent level.

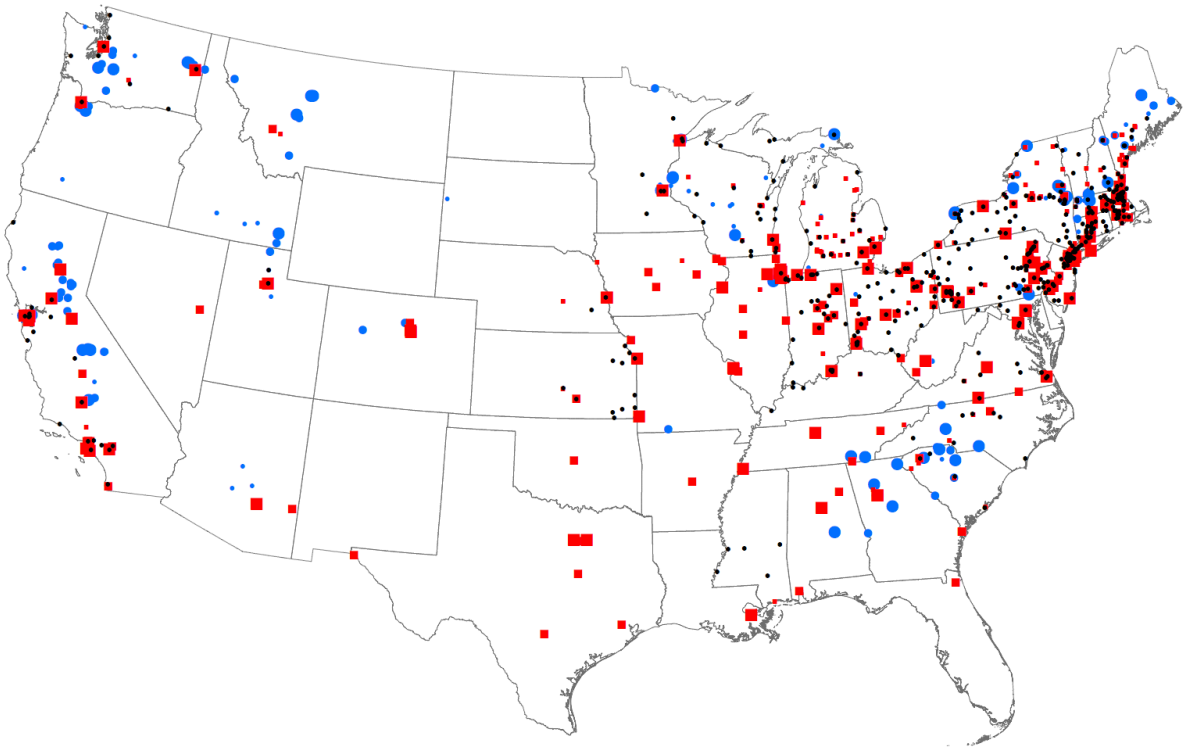
Table 3: Additional Determinants of Excess Mortality in 1918

	Dependent variable: Excess Mortality in 1918						
	Religion		Trade	City expenditure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Infant mortality, 1915-1916							
High vs. low	21.10***	21.37***	20.16***	15.91**	16.41**	15.92**	14.91**
	[4.157]	[4.205]	[4.014]	[6.488]	[6.514]	[6.528]	[7.014]
Middle vs. low	7.720**	7.295**	6.670**	5.886	5.917	5.887	5.440
	[3.179]	[3.197]	[3.151]	[4.844]	[4.854]	[4.865]	[5.293]
% illiterate in 1910							
High vs. low	21.53***	24.51***	20.90***	14.72**	15.14***	14.72**	15.55**
	[4.226]	[4.448]	[3.966]	[5.824]	[5.744]	[5.798]	[6.080]
Middle vs. low	11.46***	14.03***	11.04***	9.373*	10.05*	9.397*	10.61*
	[3.892]	[4.030]	[3.691]	[5.574]	[5.666]	[5.593]	[5.742]
Proximity to WWI base							
High vs. low	5.126	6.024	5.057	6.613	7.389	6.652	8.234*
	[3.861]	[3.829]	[3.771]	[4.352]	[4.483]	[4.632]	[4.799]
Middle vs. low	0.915	2.423	0.721	4.431	4.536	4.428	4.752
	[3.476]	[3.461]	[3.406]	[4.995]	[4.980]	[4.973]	[4.987]
Coal capacity							
High vs. low	9.327**	10.36**	9.418**	10.75*	10.69*	10.76*	9.689*
	[4.533]	[4.449]	[4.399]	[5.604]	[5.557]	[5.718]	[5.622]
Middle vs. low	3.752	3.406	4.386	6.896	6.759	6.909	6.430
	[3.450]	[3.448]	[3.395]	[4.702]	[4.689]	[4.756]	[4.859]
% congregation members	-2.645	13.29					
	[15.11]	[17.42]					
Religious fractionalization		36.60***					
		[12.96]					
Log(miles of rail)			2.148**				
			[0.871]				
Expend on health p.c.					-8.962		
					[6.477]		
Expend on sanitation p.c.						-0.147	
						[4.560]	
Expend on hospitals & charities							-1.907
							[3.182]
Observations	437	437	435	157	157	157	152
R-squared	0.253	0.270	0.264	0.295	0.301	0.295	0.296

Notes: The dependent variable is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. Columns 1 and 2 include controls for the percent of population that belonged to a religious organization and religious fractionalization in 1916. Column 3 controls for total miles of railway within the county in 1911. Columns 5-7 control for average per capita city expenditure on health, sanitation, and hospitals and charities for the years 1915-1917. Robust standard errors are reported in brackets. *** represents statistical significance at 1 percent level, ** 5 percent level, and * 10 percent level.

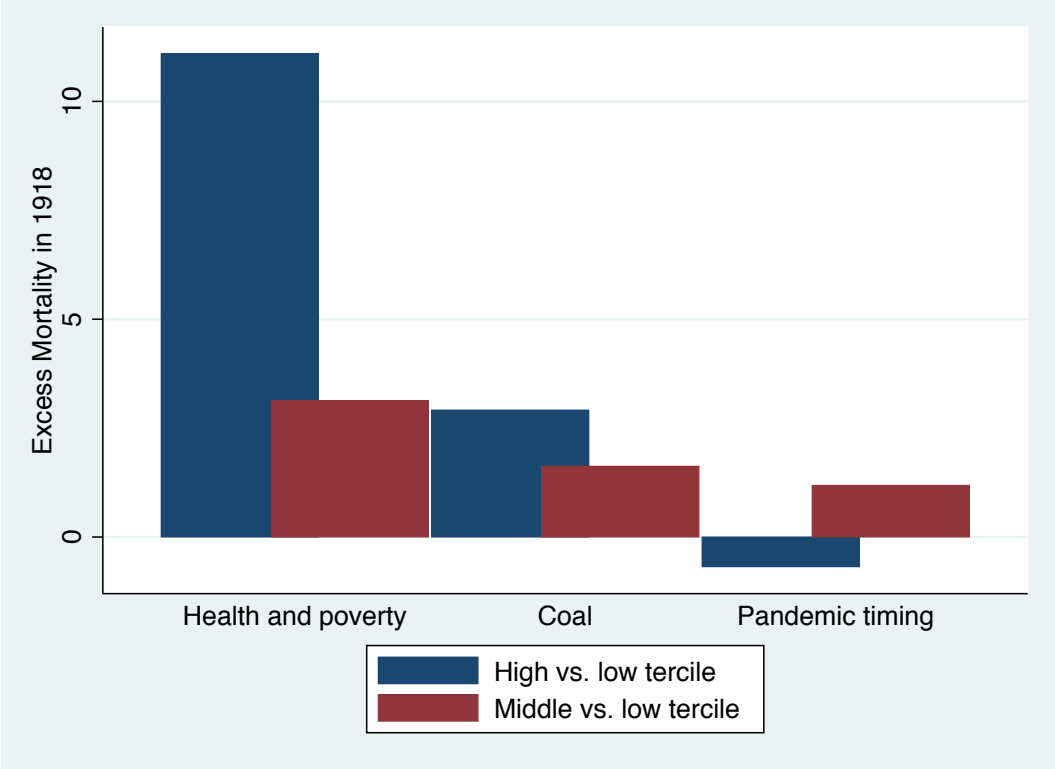
Supporting Information (Appendix)

Figure S1: Sample Cities and Power Plants in 1915



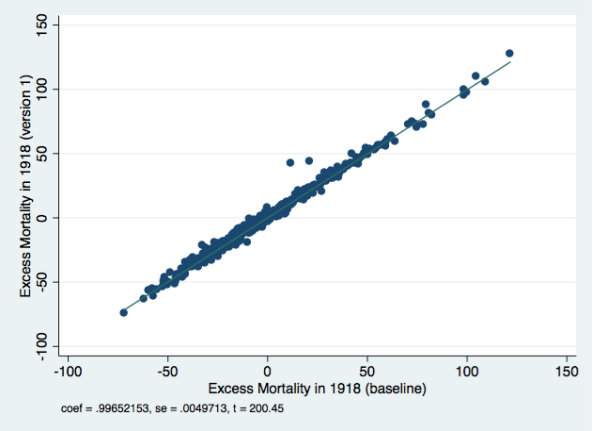
Notes: This figure maps the 438 cities in our sample (small black asterisks) as well as the coal-fired power plants (red squares) and hydroelectric dams (blue circles) in 1915 by terciles of electricity generating capacity (larger symbols represent higher terciles of capacity).

Figure S2: Determinants of Excess Mortality, Principal Components Analysis

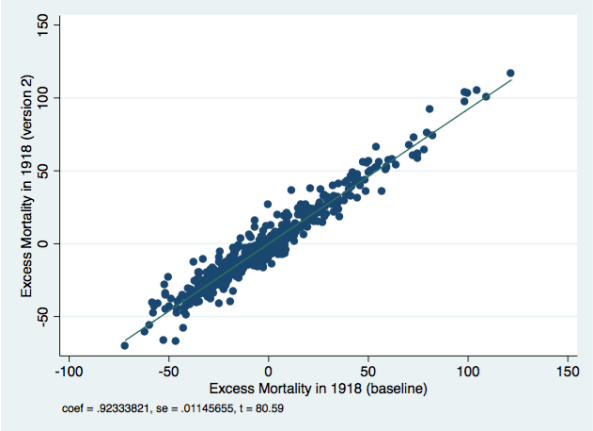


Notes: This figure reports the coefficient estimates from Table 2, col. 2.

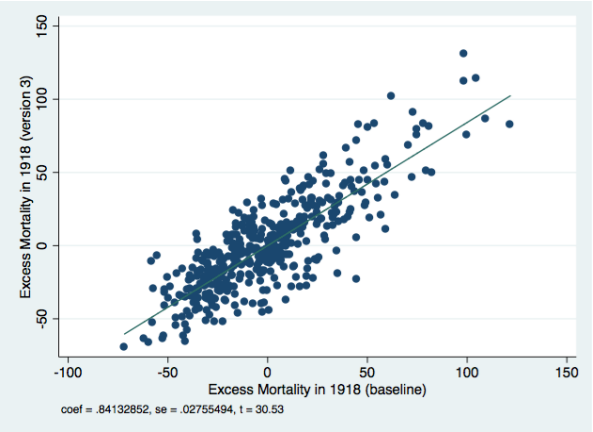
Figure S3: Alternate Measures of Excess Mortality in 1918



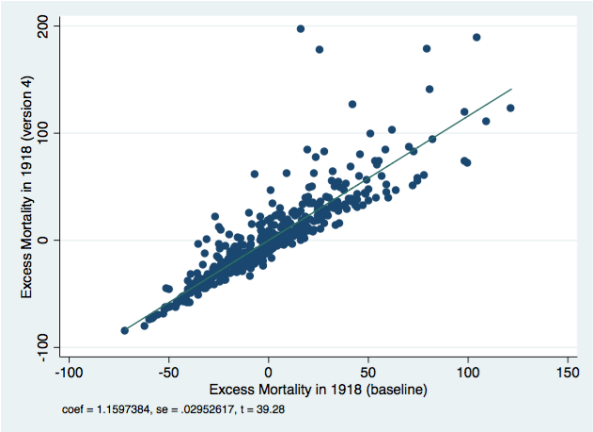
(a) Baseline excludes years 1918-1920



(b) Baseline from 1915-1917 avg. mortality



(c) Baseline from 1915-1917 mortality trend



(d) Use 1910 population denominator

Notes: This figure reports the relationship between alternate measures of excess mortality in 1918 and the original estimates. In panel (a), predicted 1918 mortality is constructed based on a linear city-specific trend from 1915-1925 (excluding years 1918-1920); in panel (b), predicted 1918 mortality is calculated as average city-level mortality for the years 1915-1917; in panel (c), predicted 1918 mortality is extrapolated from a linear city-specific trend for the years 1915-1917; in panel (d) mortality rates are constructed based on 1910 city population.

Table S1: Correlation Between Explanatory Variables

		Health and poverty				Pandemic timing		Coal	
	% urban	Infant mortality	% foreign	% own home	% illiterate	Timing of onset	Prox. WWI base	Coal capacity	% mfg emp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
% urban	1								
Infant mortality	0.09	1							
% foreign	0.57	0.17	1						
% own home	-0.63	-0.26	-0.48	1					
% illiterate	0.23	0.49	0.41	-0.49	1				
Timing of onset	-0.41	-0.18	-0.43	0.40	-0.26	1			
Proximity WWI base	0.34	-0.02	0.18	-0.40	0.15	-0.51	1		
Coal capacity	0.58	0.13	0.50	-0.56	0.24	-0.46	0.43	1	
% mfg employment	0.33	0.24	0.38	-0.30	0.15	-0.37	0.24	0.44	1

Notes: The table reports the correlation coefficient between each independent variable. All variables are measured in terciles.

Table S2: Health, Poverty, and Excess Mortality in 1918

	Dependent variable: Excess Mortality in 1918							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% urban residents in 1910								
High vs. low	-7.556**	-12.37***	-11.88***	-19.49***	-15.67***	-10.54**	-9.502**	-13.16***
	[3.420]	[3.434]	[3.406]	[4.433]	[4.252]	[4.477]	[4.062]	[4.700]
Middle vs. low	-7.725**	-5.649	-6.824**	-10.51***	-10.26***	-8.031**	-8.267**	-7.531**
	[3.501]	[3.572]	[3.441]	[3.658]	[3.460]	[4.014]	[3.670]	[3.714]
Infant mortality, 1915-1916								
High vs. low	31.35***		20.90***		28.09***		30.95***	19.85***
	[3.408]		[3.923]		[3.611]		[3.493]	[4.063]
Middle vs. low	13.88***		8.073**		11.76***		13.84***	6.811**
	[3.063]		[3.161]		[3.201]		[3.049]	[3.166]
% illiterate in 1910								
High vs. low		33.09***	22.33***					22.20***
		[3.461]	[3.921]					[4.463]
Middle vs. low		19.35***	13.33***					13.52***
		[3.345]	[3.498]					[3.737]
% homeowner in 1910								
High vs. low				-24.07***	-13.78***			-4.826
				[4.332]	[4.364]			[4.909]
Middle vs. low				-4.138	0.422			5.702
				[3.713]	[3.405]			[3.949]
% foreign born in 1910								
High vs. low						9.108**	3.053	-3.040
						[4.070]	[3.695]	[3.994]
Middle vs. low						2.063	0.486	-2.654
						[3.943]	[3.528]	[3.467]
Observations	438	438	438	438	438	438	438	438
R-squared	0.177	0.178	0.236	0.082	0.208	0.017	0.179	0.256

Notes: The dependent variable is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. The coefficient estimates report difference for the middle and high tercile of each explanatory variable, relative to the lowest tercile. Columns 1-7 report the estimates for each variable, or subset of variables, controlling for percent urban. Column 8 reports the results from the multi-factor model. Robust standard errors are reported in brackets. *** represents statistical significance at 1 percent level, ** 5 percent level, and * 10 percent level.

Table S3: Health, Poverty, and Excess Mortality in 1918 – Diagnostics

	Dependent variable: Excess Mortality in 1918							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Infant mortality, 1915-1916								
F-statistic	42.550		14.240		30.290		39.470	12.250
P-value (F-statistic)	0.000		0.000		0.000		0.000	0.000
Partial R-squared	0.170		0.059		0.127		0.161	0.054
% illiterate in 1910								
F-statistic		46.450	16.370					12.480
P-value (F-statistic)		0.000	0.000					0.000
Partial R-squared		0.170	0.059					0.046
% homeowner in 1910								
F-statistic				17.940	7.260			4.250
P-value (F-statistic)				0.000	0.001			0.010
Partial R-squared				0.075	0.031			0.017
% foreign born in 1910								
F-statistic						2.650	0.370	0.420
P-value (F-statistic)						0.072	0.689	0.660
Partial R-squared						0.010	0.001	0.001

Notes: The diagnostic tests reported in this table refer to the corresponding regression specifications from the previous table. The dependent variable in each regression is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. Controls for the high and medium terciles of percent urban are included in all models. Each F-statistic refers to the hypothesis test for the joint significance of the coefficients of the medium and high terciles (relative to the low tercile) of each explanatory variable. Each partial R-squared refers to the proportion of the variation in excess mortality in 1918 explained by the high and medium terciles (relative to the low tercile) of each explanatory variable.

Table S4: Pandemic Timing and Excess Mortality in 1918

	Dependent variable: Excess Mortality in 1918					
	(1)	(2)	(3)	(4)	(5)	(6)
% urban residents in 1910						
High vs. low	-6.937*	-10.79***	-6.196	-7.683**	-6.838	-9.316**
	[4.104]	[3.707]	[4.234]	[3.820]	[4.470]	[3.967]
Middle vs. low	-6.522*	-8.024**	-6.557*	-7.715**	-6.490*	-7.646**
	[3.845]	[3.439]	[3.871]	[3.537]	[3.854]	[3.486]
Infant mortality, 1915-1916						
High vs. low		31.85***		30.69***		31.85***
		[3.423]		[3.485]		[3.468]
Middle vs. low		13.49***		13.56***		13.44***
		[2.992]		[3.102]		[3.032]
Proximity to WWI base						
High vs. low	7.078*	9.356***			6.045	10.98***
	[4.088]	[3.614]			[4.228]	[3.728]
Middle vs. low	2.223	4.600			1.899	5.247
	[3.608]	[3.317]			[3.599]	[3.355]
Timing of pandemic onset						
High vs. low			-6.782*	-1.841	-4.163	2.908
			[3.923]	[3.623]	[4.005]	[3.701]
Middle vs. low			3.292	1.871	5.162	5.086
			[4.114]	[3.668]	[4.126]	[3.815]
Observations	438	438	438	438	438	438
R-squared	0.016	0.190	0.023	0.179	0.028	0.194

Notes: The dependent variable is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. The coefficient estimates report difference for the middle and high tercile of each explanatory variable, relative to the lowest tercile. Columns 1-5 report the estimates for each variable, or subset of variables, controlling for percent urban. Column 6 reports the results from the multi-factor model. Robust standard errors are reported in brackets. *** represents statistical significance at 1 percent level, ** 5 percent level, and * 10 percent level.

Table S5: Pandemic Timing and Excess Mortality in 1918 – Diagnostics

	Dependent variable: Excess Mortality in 1918					
	(1)	(2)	(3)	(4)	(5)	(6)
Infant mortality, 1915-1916						
F-statistic		43.490		38.890		42.300
P-value (F-statistic)		0.000		0.000		0.000
Partial R-squared		0.175		0.156		0.166
Proximity to WWI base						
F-statistic	1.530	3.360			1.040	4.340
P-value (F-statistic)	0.218	0.036			0.353	0.014
Partial R-squared	0.008	0.013			0.005	0.014
Timing of pandemic onset						
F-statistic			2.770	0.430	2.190	0.910
P-value (F-statistic)			0.064	0.650	0.114	0.404
Partial R-squared			0.016	0.002	0.013	0.003

Notes: The diagnostic tests reported in this table refer to the corresponding regression specifications from the previous table. The dependent variable in each regression is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. Controls for the high and medium terciles of percent urban are included in all models. Each F-statistic refers to the hypothesis test for the joint significance of the coefficients of the medium and high terciles (relative to the low tercile) of each explanatory variable. Each partial R-squared refers to the proportion of the variation in excess mortality in 1918 explained by the high and medium terciles (relative to the low tercile) of each explanatory variable.

Table S6: Coal and Excess Mortality in 1918

	Dependent variable: Excess Mortality in 1918					
	(1)	(2)	(3)	(4)	(5)	(6)
% urban residents in 1910						
High vs. low	-14.03***	-14.88***	-8.715**	-9.045***	-15.00***	-15.13***
	[4.825]	[4.265]	[3.708]	[3.465]	[4.802]	[4.211]
Middle vs. low	-8.484**	-9.190***	-8.744**	-8.674**	-9.705**	-9.464***
	[3.890]	[3.547]	[4.048]	[3.655]	[3.987]	[3.632]
Infant mortality, 1915-1916						
High vs. low		30.53***		30.50***		30.40***
		[3.367]		[3.628]		[3.584]
Middle vs. low		13.40***		13.30***		13.25***
		[3.105]		[3.128]		[3.157]
Coal capacity						
High vs. low	15.89***	12.37***			12.78**	12.04**
	[4.991]	[4.252]			[5.673]	[4.851]
Middle vs. low	7.230*	4.090			5.375	3.965
	[3.706]	[3.500]			[4.015]	[3.682]
% manufacturing in 1910						
High vs. low			10.85***	3.614	7.260*	0.375
			[3.825]	[3.665]	[4.363]	[4.147]
Middle vs. low			1.373	-1.640	0.0451	-2.807
			[3.876]	[3.506]	[3.936]	[3.551]
Observations	438	438	438	438	438	438
R-squared	0.035	0.194	0.029	0.182	0.045	0.196

Notes: The dependent variable is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. The coefficient estimates report difference for the middle and high tercile of each explanatory variable, relative to the lowest tercile. Columns 1-5 report the estimates for each variable, or subset of variables, controlling for percent urban. Column 6 reports the results from the multi-factor model. Robust standard errors are reported in brackets. *** represents statistical significance at 1 percent level, ** 5 percent level, and * 10 percent level.

Table S7: Coal and Excess Mortality in 1918 – Diagnostics

	Dependent variable: Excess Mortality in 1918					
	(1)	(2)	(3)	(4)	(5)	(6)
Infant mortality, 1915-1916						
F-statistic		41.150		35.440		36.000
P-value (F-statistic)		0.000		0.000		0.000
Partial R-squared		0.159		0.170		0.152
Coal capacity						
F-statistic	5.100	4.390			2.540	3.250
P-value (F-statistic)	0.007	0.013			0.080	0.040
Partial R-squared	0.028	0.017			0.016	0.015
% manufacturing in 1910						
F-statistic			5.970	1.390	2.410	0.590
P-value (F-statistic)			0.003	0.249	0.091	0.554
Partial R-squared			0.022	0.005	0.010	0.002

Notes: The diagnostic tests reported in this table refer to the corresponding regression specifications from the previous table. The dependent variable in each regression is calculated as the difference between observed and predicted all-age mortality in 1918, where predicted mortality is calculated based in a linear city-specific trend for the period 1915 to 1925. Controls for the high and medium terciles of percent urban are included in all models. Each F-statistic refers to the hypothesis test for the joint significance of the coefficients of the medium and high terciles (relative to the low tercile) of each explanatory variable. Each partial R-squared refers to the proportion of the variation in excess mortality in 1918 explained by the high and medium terciles (relative to the low tercile) of each explanatory variable.