NBER WORKING PAPER SERIES

LONDON FOG: A CENTURY OF POLLUTION AND MORTALITY, 1866-1965

W. Walker Hanlon

Working Paper 24488
http://www.nber.org/papers/w24488

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2018

I thank Vellore Arthi, Alastair Ball, Brian Beach, David Cutler, Tatyana Deryugina, Casper Worm Hanson, Stephan Heblich, Richard Hornbeck, David Molitor, Edson Severnini, and Katherin Sudol for helpful comments. I am grateful to Shmuel San and Anna Sudol for excellent research assistance. Funding for this project was provided by National Science Foundation CAREER Grant No. 1552692. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

London Fog: A Century of Pollution and Mortality, 1866-1965
W. Walker Hanlon
NBER Working Paper No. 24488
April 2018
JEL No. I15,N3,Q53

ABSTRACT

This study provides new evidence on the impact of air pollution in London over the century from 1866-1965. To identify weeks with elevated pollution levels I use new data tracking the timing of London's famous fog events, which trapped emissions in the city. These events are compared to detailed new weekly mortality data. My results show that acute pollution exposure due to fog events accounted for at least one out of every 200 deaths in London during this century. I provide evidence that the presence of infectious diseases of the respiratory system, such as measles and tuberculosis, increased the mortality effects of pollution. As a result, success in reducing the infectious diseases burden in London in the 20th century reduced the impact of pollution exposure and shifted the distribution of pollution effects across age groups.

W. Walker Hanlon
Department of Economics
NYU Stern School of Business
Henry Kaufman Management Center
44 West Fourth Street
New York, NY 10012
and NBER
walker.hanlon@gmail.com

1 Introduction

For over a century, from the mid-19th century through the 1960s, London experienced some of the highest sustained levels of air pollution in the world. Today, as modern industrial cities struggle with their own pollution problems, London's experience has the potential to offer useful insights into the cost of high levels of air pollution and how these costs evolve as cities develop. Yet our current understanding of this experience remains limited. This is due largely to the scarcity of direct pollution measures prior to the mid-20th century, which has posed a serious impediment to studying air pollution over longer periods of time.

This study provides novel evidence on the acute impact of air pollution in London over the century from 1866 to 1965. Two new elements allow me to overcome the lack of direct measures of air pollution during most of this period. First, I take advantage of London's famous fog events to infer the timing of weeks with elevated pollution exposure. While individual fog events have been studied in previous work, this study uses newly constructed data tracking every occurrence of fog across over 4,500 weeks. These events allow me to consider the effects of pollution across a much longer period than is possible when relying on direct pollution measures. Second, I draw on a newly digitized data set describing mortality in London at the weekly level, over a consistent geographic area, and broken down by age group and cause of death.

Together, these data sets allow an analysis strategy that uses weekly variation in pollution levels due to fog events to assess the acute impact of pollution exposure on mortality. In the main analysis, identification relies on the fact that the formation of fog depends on the complex interaction of several climatic conditions, including temperature, humidity, wind speed and cloud cover. To strengthen identification, I also offer an instrumental variables strategy that uses the interaction of weather conditions to predict the timing of fog events, while controlling for each individual weather variable. This helps me deal with concerns about the endogeneity of fog formation to pollution levels as well as the possibility that the criteria for reporting fog events may have changed over time.

I present two main sets of results. The first set of results focuses on estimating the overall impact of acute pollution exposure associated with fog events on mortality in London across the century covered by this study. My estimates show that at least one out of every 200 deaths in London during this century was directly attributable to the *acute* effects of pollution exposure, by which I mean the effect occurring within three weeks of a fog event. This corresponds to just over 39,000 deaths. In addition, I estimate that the acute effects of fog events caused over 1,000 additional stillbirths, while in utero exposure led to at least 1,400 infant deaths and an additional 3,500 stillbirths. Overall, these results show that

acute pollution exposure made a substantial contribution to mortality in London across this century. I also provide new evidence on how these effects were distributed across age groups, a topic where existing evidence is sparse. While most deaths were concentrated among the young and elderly, I also find substantial impacts among teenagers and prime-age adults. This suggests that focusing only on the impacts of pollution on infants and the elderly, as is done in many studies, may be missing an important part of the overall effect. In terms of causes-of-death, my analysis shows that the largest effects of pollution occur in the categories we would expect: respiratory diseases such as bronchitis and pneumonia as well as cardiovascular diseases. In contrast, I find no interaction with many other causes of death, such as digestive diseases, where we would not expect an interaction. The fact that the effects of pollution are concentrated in respiratory or cardiovascular channels provides support for the identification strategy.

One may be concerned that many of these deaths were simply due to the harvesting of people who otherwise would have died soon after. To assess this concern, I estimate the impact of fog events up to one year after the acute effect window, which I call *medium-run* effects. If the acute effects were driven by harvesting, then I would expect to see fog events associated with reduced mortality in the medium run. Instead, I estimate that fog events were associated with elevated medium-run mortality. This suggests that either harvesting was not large, or that any harvesting effects were overwhelmed by the medium-run impact of pollution exposure. While not as well identified as the acute effect, the magnitude of the medium-run effect is large, accounting for 1.8% of all deaths in London during the study period, or about 78,000 dead.

The second main set of results presented in this study describe how the impacts of pollution changed as the city developed. Specifically, I provide new evidence on the interaction between pollution and infectious diseases and then consider how this interaction, together with the substantial progress made in reducing infectious diseases in London in the early 20th century, modified the impacts of pollution on health. My results show that acute pollution exposure interacted with a set of infectious diseases of the respiratory system, throat or larynx: measles, pneumonia, tuberculosis (TB), and scarlet fever. The link between elevated pollution exposure and deaths due to these infectious diseases has implications for the impact of development on the costs of pollution. By the middle of the 20th century, substantial progress had been made in reducing mortality from these diseases in Britain. My results suggest that these improvements in the disease environment lowered the mortality cost of acute pollution exposure. For example, acute pollution exposure substantially increased mortality among those infected with the measles. As a result, I estimate that the reduction in measles deaths in London lowered the acute impact of fog events by 18% after

1914 compared to the period from 1865-1914.

Changes in the infectious disease environment also shifted the burden of pollution exposure across age groups. Measles deaths are heavily concentrated among children aged 1-5. Thus, the reduction in measles reduced the effects of acute pollution exposure on this group. A similar result occurs for TB, a disease that was the main killer of prime-aged adults. My results suggest that progress against TB reduced the acute impact of pollution exposure by 5% after WWI compared to before, with much of this benefit concentrated among adults aged 20-40. The fact that much of the effect of air pollution on the mortality of young and prime-aged adults during my study period operated through TB suggests that in modern developed countries, where this disease is relatively uncommon, the impact of pollution on health among these age groups should be much lower. Thus, I provide evidence that changes in the infectious disease environment caused by, for example, investments in public health infrastructure, can reduce the mortality cost of acute pollution exposure while also altering the distribution of these costs across age groups.

This study contributes to a broader literature analyzing the historical impact of pollution. Much of the work in this area has focused on water pollution.[1] A small but growing set of studies address air pollution, including Barreca *et al.* (2014), Clay *et al.* (2016), Beach & Hanlon (Forthcoming), and Bailey *et al.* (2016).[2] This study has a particularly close connection to a paper by Clay *et al.* (2015) which shows that pollution can interact with influenza to increase mortality. While I find similar interactions, I go beyond their results by looking across many diseases, ages and time periods while drawing on a different identification strategy. In addition, I offer evidence on how changes in the disease burden influenced the cost of pollution.

This study also contributes to work using modern data to assess the health effects of pollution.[3] Many studies in this literature use data from modern developed countries and focus on infant mortality (Chay & Greenstone, 2003; Currie & Neidell, 2005). Recently there has been an increase in studies looking at more heavily polluted developing countries (Jayachandran, 2009; Foster *et al.*, 2009; Almond *et al.*, 2009; Chen *et al.*, 2013; Greenstone & Hanna, 2014; Rangel & Vogl, 2016), though research from highly-polluted settings remains somewhat limited. These developing settings are more similar to the context that I consider,

[1] See, e.g., Cutler & Miller (2005), Ferrie & Troesken (2008) and Alsan & Goldin (2014).

[2] An older set of papers in this area focus on individual air pollution events such as the famous episode in Danora, PA (Townsend, 1950), the Great London Fog of 1952 (Logan, 1953; Bell & Davis, 2001; Ball, 2015), the Muesse Valley fog in the 1930s (Nemery *et al.*, 2001), and a pollution incident in New York City in the 1950s (Greenburg *et al.*, 1962). Another related paper looking at fog events is Troesken & Clay (2011) which uses historical sources and aggregate mortality patterns to identify the frequency of polluted fogs in London

[3] Useful reviews of this literature include Currie (2013), Graff Zivin & Neidell (2013) and Rückerl *et al.* (2011).

which was characterized by high levels of pollution and infectious disease mortality. The most closely related studies to this paper are those that focus on the contemporaneous effects of acute air pollution (Pope, 1989; Schlenker & Walker, 2016; Knittel *et al.*, 2016; Jia & Ku, 2017), particularly a pair of papers studying the impact of temperature inversions in Mexico City (Hanna & Oliva, 2015; Arceo *et al.*, 2016). The use of temperature inversions to identify acute pollution effects is similar to using fog events, which were often accompanied by inversions.

This paper differs from existing studies using modern data in that I am able to analyze how pollution impacts change over time, how they vary across age groups, and how they interact with other causes of death, such as infectious diseases. Also, the repeated nature of the events I study allow me to look at whether pollution effects were characterized by harvesting of those who were likely to die soon anyway, or whether repeated exposure had cumulative effects. However, the use of historical data means that my study comes with some caveats. For example, direct pollution measures are not available during most of my study period so I cannot separately identify the impact of different pollutants.

The next section describes the empirical setting, followed by the data, in Section 3. Section 4 outlines the empirical strategy and discusses the main threats to identification. A preliminary analysis showing the relationship between fog events and pollution levels is presented in Section 5 followed by the main analysis in Section 6. Concluding remarks are in Section 7.

2 Setting

London's struggles with air pollution date back at least to the 17th century, when coal replaced wood as the main domestic fuel, but the problem grew in intensity after the Industrial Revolution due to population growth, rising incomes, cheaper access to coal, and industrialization (Brimblecombe, 1987). Though direct pollution measures were unavailable during the 19th and early 20th century, contemporary reports indicate that air pollution levels in London were high. For example, an observer in 1866 wrote,

Here were are in the Metropolis breathing coal-smoke, begrimed by coal-smoke, and sometimes involved in that 'pitchy cloud of infernal darkness' through which we see the sun dimly as a fiery red ball...The fine trees of Kensington are dying apace, and roses bloom not within some miles of Charing Cross...the evil is now caused in great measure by the imperfect combustion of a very large quantity of coal in our domestic fires...[4]

Starting in the middle of the 20th century we begin to have direct pollution measures. Direct pollution measures for 1951-1962, which I describe in more detail in Section 3, show a weekly average of daily maximum total suspended particulate (TSP) levels of 260 $\mu g/m^3$ and a weekly average of daily mean TSP levels of 140 $\mu g/m^3$.[5] Levels over 1000 $\mu g/m^3$ were observed in five different weeks during this decade, including during the Great London Fog of 1952. As a point of comparison, the first set of standards set by the U.S. Environmental Protection Agency in 1971 specified that the annual average of daily maximum pollution levels should not exceed 75 $\mu g/m^3$ and the maximum on the worst day in a year should not exceed 260 $\mu g/m^3$.

The main source of pollution during the study period was residential coal use for home heating and cooking.[6] Residential pollution remained largely unregulated until the Great London Fog of 1952 motivated action by the government, leading to the Clean Air Acts of 1956. This began the slow process of pushing households to convert to smokeless fuels or alternative energy sources such as electricity.[7]

The highest levels of pollution in London occurred during periods of fog. Radiation fogs, the most common type in London, form when humid low-lying air is cooled below the dew point by contact with the ground. If there is sufficient moisture in the air then this cooling causes the water in the air to condense, forming very small suspended water droplets. For radiation fog to occur, the temperature of the air and the ground must be in the right range, the air must contain sufficient moisture, and the air must be relatively

[5]Modern pollution measures generally focus on more specific particulate sizes such as PM10 or PM2.5, rather than TSP. However, for the period I study these more detailed measures are not available.

[6]Automobiles may have been an important contributor towards the end of the study period. One reason for the importance of domestic pollution was that, relative to other British cities, London's industrial structure was concentrated in less polluting sectors such as government, banking, and services (Beach & Hanlon, Forthcoming). Another reason was that regulation had made some progress in reducing industrial pollution sources. Regulation was much less successful in addressing domestic pollution, which was both more onerous to police and more difficult politically (Thorsheim, 2006).

[7]One consequence of the importance of residential pollution related to home heating is that temperature exerted a strong influence on emissions, a pattern that will be visible in the available pollution data from the 1950s and 1960s presented later. This means that temperature will be an important control in my analysis.

stable so that it is in contact with the ground for long enough to cool.[8] Radiation fog also requires fairly clear skies, so that the sun can warm the ground during the day, which then cools through radiation at night. Often, fogs were accompanied by temperature inversions, where colder air above traps warm air below. Under these calm conditions, any emitted pollution remains close to the source and near ground level.[9] Thus, fog events provide a way of identifying periods during which pollution concentrations increased. Later I will show that pollution levels during fog events were much higher than in nearby weeks. During the worst events, TSP levels could reach as high as 1,400-1,600 $\mu g/m^3$, far above modern guidelines and comparable to the levels observed in the most polluted modern cities (Bell & Davis, 2001).

Fog itself, being composed merely of small water droplets suspended in air, is not naturally harmful to health. The coal pollution trapped by the weather conditions that accompanied dense fogs, however, contained a variety of harmful substances. This included suspended particulates of various sizes, sulfur dioxide, nitrogen oxide, carbon monoxide and various metallic compounds such as lead and mercury. These compounds affect human health in many ways. Existing work has highlighted negative effects on infants (Chay & Greenstone, 2003; Currie & Neidell, 2005; Currie *et al.*, 2009) and older adults (Chay *et al.*, 2003). The most pronounced effects operate through the respiratory and cardiovascular systems (Samet *et al.*, 2000; Rückerl *et al.*, 2011). There is also limited evidence that pollution effects may interact with infectious diseases to increase mortality (Clay *et al.*, 2015).

3 Data

All of the data used in this study come from the Weekly Reports generated by the Registrar General's office. The Registrar General's data were collected by trained local registrars. Demographers such as Woods (2000) praise the overall quality of the Registrar's mortality data, even in the 19th century, which he calls "the shining star of Victorian civil registration." The mortality data appear to be particularly accurate in London.[10]

The Registrar's weekly reports include both information on mortality in London during the previous week as well as weather information. The data were collected from original

[8] Meetham *et al.* (1981, p. 172-173).

[9] The effects of these calm conditions were further exacerbated by London's bowl-like topography. London is situated on low flat ground along the Thames river surrounded by higher ground in all directions except in the East, where the Thames estuary opens into the North Sea. This topography had the effect of trapping pollution in the city, though on most days the predominant southwest wind pushes pollution out toward the sea.

[10] See, e.g., the discussion in (Woods, 2000, Ch. 2).

reports stored in the British Library, the LSE Library, and the New York Public Library.[11] The hard copy data were photographed, entered by hand, and the checked for quality. The resulting database covers 4,539 weeks, with between 51 and 53 weeks in each year.[12] A small number of observations are missing because I was unable to find them in the collections of any library or due to factors such as printing errors (though these are quite rare). In total the mortality data include over 350,000 observations. Summary statistics for the main data series are available in Appendix Table 8.

One advantage of this setting is that the geographic area from which the mortality data were drawn, the County of London, remained stable across the study period. After 1965, the County of London was replaced by the much larger Greater London administrative area. To avoid this discontinuity, I end the study period in 1965. While some weekly mortality statistics were reported for cities outside of London, or for neighborhoods within London, to my knowledge the level of detail used in this study is only available for London as a whole.

The mortality data are reported by both age category and cause of death. Typically, the age categories separately identify infants as well as age groups spanning 5 to 15 years, with some changes to the reported age groups across the study period. In order to analyze the impact on different age categories across the study period, I construct the following set of age-group mortality series. For infants aged 0-1 and children 1-5, I have consistent mortality data from 1876-1965. For other ages, changes over time in reporting categories generate some inconsistencies. My analysis will use the following age groupings: a "young" age category which includes deaths in ages 5-20 for 1870-1921 and 5-25 for 1922-1965; an "adult" category that includes ages 20-40 from 1870-1910, ages 20-45 from 1911-1921, and 25-45 for 1922-1965; a "middle age" category that includes adults aged 40-60 from 1870-1910 and ages 45-65 from 1911-1965; an "elderly" age category that includes those over 60 from 1870-1910 and those over 65 from 1911-1965.

The weekly reports also include data on stillbirths starting in 1927. The number of stillbirths is substantial. In the data starting in 1927 I observe that there were two-thirds as many stillbirths as total deaths of infants aged 0-1.[13] I have also collected information on the number of births. This variable provides a useful control because of the high level of infant mortality, particularly early in the study period.

The mortality data also include a substantial amount of cause-of-death (COD) informa-

[11] Alone, none of these libraries has a fully comprehensive set of the weekly reports.

[12] It is worth noting that at the end of years the data in some weeks appear to cover less than a full seven days.

[13] Specifically, my data record 47,787 stillbirths in the years for which those data are available and 71,920 infant deaths in the same time period.

tion. The COD categories can be useful, but also come with important limitations. There were substantial changes in both the COD categories reported and the accuracy of COD diagnosis across the period that I study. To obtain more consistent series, I combine the COD categories into 23 aggregated groups that show fairly consistent patterns over time. A table describing these groups and their subcomponents is available in Appendix A.2.4. Among the available cause-of-death categories, common infectious diseases like measles, scarlet fever, whooping cough, diphtheria, and smallpox are likely to be particularly accurately measured because they were common and present with clear distinguishing symptoms. Because of substantial changes in classifications after WWII, my analysis of the cause-of-death data focuses on the period from 1870-1939.

Most of the deaths reported in the Registrar General's reports would have been registered within 2-3 days of their occurrence.[14] This is an important fact to keep in mind when looking at the lag structure of the impacts of fog events. In particular, even if fog events cause mortality very rapidly, some of the deaths associated with events occurring late in a week may have been registered in the following week.

To identify weather events, we manually reviewed the daily weather notes contained in the Registrar General's Weekly reports for roughly 31,500 days. Two examples of these notes, one from 1880 and another from the week of the famous Great London Fog of 1952, are available in Appendix A.2.2. These show how similar the format and content of the daily notes remained across the century covered by this study. From the daily notes, we identified, for each week, the number of days in which a heavy fog day occurred, i.e., a day in which fog was reported with any indicator of intensity, such as "thick", "heavy", or "dense" fog. This method identifies 932 heavy fog days and 582 weeks in which at least one heavy fog event occurred. Appendix Figure 9 describes the frequency of fog weeks and fog days over the study period.

One important change took place in the weather reports used in this study. Prior to 1950, the weather reports were taken at the Royal Observatory in Greenwich, southeast of Central London along the River Thames. Starting in 1951, the available weather reports come from Kew Gardens which is located just west of London, a bit further from the city center than Greenwich. It is important to keep in mind that this switch may generate a discontinuity in the relationship between fog events and pollution. Because of this change, I will refer to the period starting in 1950 as the "Kew Gardens Period." Appendix Figure 9 shows that the number of reported fog weeks and fog days increased substantially when the reporting moved from Greenwich to Kew Gardens in 1951. It will be important to take this

[14] See (Ministry of Health, 1954, p.11).

discontinuity into account in the analysis.[15]

Seasonality is an important feature of both mortality and fog events. Figure 10 in the Appendix provides a plot of the number of fog events as well as the share of total deaths across weeks of the year. This shows that fog events were more likely to occur in the winter and that overall mortality was higher in the winter as well.[16] Thus, I will include controls for week of the year and temperature when estimating the impact of fog events.

Additional weather data describing weekly mean temperature, humidity, barometric pressure and precipitation were gathered from the Registrar General's reports. These data will provide useful control variables. They will also allow me to generate a quantitative prediction of fog events that will be consistent over time. It is worth noting that these values are from Greenwich Observatory for years before 1951 and from Kew Gardens starting in 1951.

A limited set of data on pollution levels are also available from the Registrar General's reports starting in 1951. These data are available in a consistent way through the 21st week of 1962.[17] The reported values include the weekly average of the daily maximum and mean pollution levels in units that correspond to the total suspended particulate (TSP) values commonly reported during this period.[18] The Online Appendix provides graphs of the pollution data. These show that pollution was highly seasonal and generally declining from 1951-1962.

4 Methodology and identification

4.1 Overview and identification concerns

Figure 1 describes the basic relationships at work in this study. Ideally we would like to identify the impact of pollution exposure on mortality and how this effect is modified by changes in other factors, such as public health infrastructure and medical advances. However,

[15]It is not clear if Kew Gardens really experienced more fog or if there were reporting differences across the two locations.

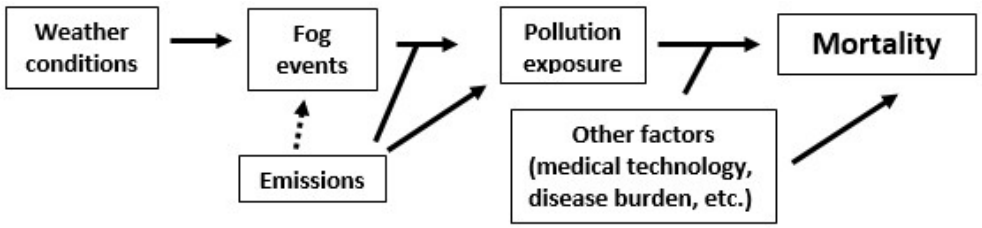
[16]The seasonality of fog events did not change substantially across the study period.

[17]The pollution measures were obtained from a device, called the Owens Smoke Meter, in which a fixed amount of air was passed through a filter paper which captured the particles suspended in the air. The stains left on the filter paper were then compared to a calibrated set of stain ratings. This provided an early direct measure of air pollution. I only use data up to 1962 because it is clear that at that point there was a structural break in the accuracy of the reported values that generates an inconsistency. In particular, from 1951-1962 the readings come in fairly large discrete levels, while after 1962 they become finer. This, discontinuity, which is related to the initiation of the "National Survey" of air quality, has a substantial impact on the consistency of the series which I want to avoid.

[18]This was before pollution measures identified levels of particles of particular sizes, such as PM10 or PM2.5, which are the measures commonly used today.

pollution is not directly observable over such a long period.

Figure 1: Relationships considered in this study



Relying on fog events helps me work around the lack of direct pollution measures in order to study effects over a long period. One primary contribution of this study is estimating the impact of acute pollution effects due to fog events on mortality across the study period. I am also interested in how these effects evolved as the city developed. There are three factors that will influence how the impact of fog events on mortality changes over time. First, this relationship may be modified by factors, such as public health improvements, that affect the relationship between pollution exposure and mortality. This relationship is of primary interest. In addition, the relationship between fog events and mortality will depend on changes in emission levels, which are not observable.[19] Finally, the criteria for reporting fog events may shift over time. This last concern can be addressed by using quantitative weather variables to predict the timing of fog events, but it is not possible to fully separate the influence of changing emissions levels from other factors. However, by using cause of death information, together with what we know about the changing disease environment in London during this period, it is possible to isolate some of the influence of changes in the disease environment on the mortality costs of pollution.

There are several identification concerns that must be addressed in this study. One of these, represented by the dotted line in Figure 1, is that higher levels of emissions may make fog events more likely. More particulates in the air makes it easier for water to condense, forming fog. This possibility raises a concern because emissions are likely to be related to factors, such as income or the price of home heating, that might also influence health. There are several pieces of evidence suggesting that short-run changes in emissions levels are unlikely to be determining the timing of fog events.[20] However, to be sure that this

[19]While there is some data on coal use in London, the type of coal and how it was burned had an important impact on how much pollution was released. This surely changed across the study period in unobservable ways.

[20]For example, Figure 9 shows that changing pollution levels were not a primary driver of fog events, since there is no reduction in the number of fog events in the late 1950s and 1960s despite the consistent decline

is not driving my results I will use data on underlying weather conditions to predict the timing of fog events. These conditions are not affected by emissions levels, so this will deal with endogeneity concerns. To satisfy the exclusion restriction, I will take advantage of the fact that fog formation results from the complex interaction of several weather conditions, including temperature, humidity, cloud cover and wind speed. As discussed below, this allows me to predict the timing of fog formation while controlling flexibly for underlying weather conditions.

Another potential concern is that fog events may have affected mortality through channels other than pollution exposure. One potential channel is accidents or crime, which may have been more likely on foggy days. I will assess this channel directly by separating deaths due to accidents or homicide from deaths due to other cause. Another potential channel is that fog may have made it harder for people to reach the hospital or for doctors to reach their patients, though during much of the study period medical care was rudimentary and often counterproductive (McKeown, 1976; Floud *et al.*, 2011).[21] One way to address this issue is to study whether the effects of fog events are concentrated in causes of death which we know are associated with air pollution, such as diseases of the respiratory or cardiovascular systems. If fog is increasing mortality by making it harder for people to reach medical care, rather than through air pollution exposure, then we should not expect the mortality effects to be concentrated in the diseases typically associated with pollution exposure. Comparing across causes of death is an approach that has been used in several existing studies of the impact of pollution on mortality, including Galiani *et al.* (2005), Alsan & Goldin (2014), Beach & Hanlon (Forthcoming) and Jia & Ku (2017).

Another concern is that fog may have caused people to stay indoors which could have increased mortality by facilitating the spread of disease.[22] However, staying indoors also reduces exposure to people outside of the household, which may have actually reduced infectious disease transmission. In Appendix A.4 I investigate this potential channel by looking at how mortality responded to weeks with heavy rainfall, a factor that was also likely to have caused people to stay indoors.[23] Those results show that, in fact, mortality from airborne

in pollution levels across that period. I will also provide evidence that pollution levels were not higher in the week before fog events occur, as one would expect if the timing of fog events was driven by changes in pollution levels.

[21] Medical care during most of the study period was particularly ineffective for TB and measles, two diseases that play a central role in my results.

[22] Staying indoors may have also exposed people to higher levels of indoor air pollution. Since this is just another channel through which fog increases pollution exposure it does not pose a major concern for my identification strategy.

[23] This exercise also addresses any concerns about the possibility that there may have been something about the moisture content of the air during fog weeks that influenced mortality, since the air would have also had a high moisture content on weeks with heavy rain.

infectious diseases fell following weeks with heavy rain, suggesting that staying indoors may have been healthier than venturing outside. This effect does not appear to be due to the impact of rainfall on outdoor pollution levels, nor is it driven by water-borne diseases that may have been affected by rainfall in other ways. Thus, if anything, the fact that fog may have induced people to stay indoors may bias my results towards finding smaller effects.

4.2 Estimation procedure

This study applies a simple time-series analysis approach that relies on random variation in the week-to-week timing of fog events driven by climatic factors. My baseline regression specification is:

$$\ln(MORT_t) = \alpha + \sum_{s=-\tau}^{\tau'} \beta_s FOG_{t+s} + X_t'\gamma + Y_t + WD_t + e_t \quad (1)$$

where $MORT_t$ is the number of deaths in London in week t , FOG_t is the number of fog days in week t, Y_t is a set of year effects, WD_t is a full set of week-of-the-year by decade effects, and X_t is a vector of control variables. The year effects in this specification absorb changes in mortality patterns over time. The week-of-the-year effects absorb seasonal factors that affect mortality. Allowing these to vary by decade deals with the fact that the seasonality of mortality is likely to change across such a long study period.[24] The dependent variable in this regression is log mortality in London, either total or for a specific age group or cause-of-death. The main set of dependent variables is the number of fog days in a week, as well as leads and lags of that variable.

The most important control variable is temperature. Both high and low temperatures increase mortality, while low temperatures can also affect pollution levels because much of the pollution in London was due to coal burning for home heating. Temperature also plays an important role in fog formation. Thus, in addition to controlling flexibly for temperature in a week I also include controls for leads and lags of temperature and temperature squared. This ensures that, for example, lagged fog effects are not picking up the lagged effect of temperature.[25] The regressions also include controls for other weather variables – pressure,

[24]Barreca *et al.* (2016) suggests that, at least in the U.S., there were dramatic changes in the seasonality of mortality across the 20th century. It is worth noting that 1949 is included as part of the decade of the 1950s when constructing the week-by-decade effects, since that is the only year in the 1940s for which data are available.

[25]I also explore results including additional quadratic temperature terms. These are typically not statistically significant and have no meaningful impact on the results, so I do not include them my preferred specification.

humidity and precipitation – and squared values of these terms. These are typically not strong controls so I don't include a full set of leads and lags of these variables, though leads and lags are included in some robustness exercises. I also include log births as a control in most regressions to account for the high level of infant mortality in much of the study period. This is strongly related to mortality but the inclusion of this variable has no impact on any of the main results.[26]

After establishing the lag structure of the fog effects, I also consider a second specification,

$$\ln(MORT_t) = \alpha + \beta FOG_{t,t-3} + X'\gamma + Y_t + WD_t + e_t \tag{2}$$

where the main explanatory variable is the number of fog days in week t and the three previous weeks. This specification, which is motivated by the lag structure identified using Eq. 1, is useful for simplifying the results so that they can be broken down by age group and cause-of-death in a manageable way.

These specifications generate results in terms of percentage changes in the number of deaths. These can be interacted with baseline mortality to obtain expected changes in the number of deaths. I will, however, avoid looking at changes in death rates because these require population data, which are only observed once every decade (in census years).

One standard concern in an analysis based on time-series data is serial correlation.[27] To

[26]The log births variable should be interpreted with some caution because it will also reflect two other factors. First, this variable may be partially capturing population change within a year that is not soaked up by the year effects. Second, log births may also capture the effect of changes in the number of days included in some weeks at the end of years, when it appears that a week may include fewer than seven days. One may be concerned that the births may be a bad control because they could be affected by fog events. However, in Appendix A.7 I show that there is no evidence of substantial direct or lagged relationship between births and fog events. Consistent with this, the estimated coefficients on pollution or fog events change very little depending on whether births are included as a control. While I do provide evidence that stillbirths increase as a result of fog events, these are rare relative to total births. This, together with the possibility that many stillbirths may occur weeks before the due date explains why I observe effects for stillbirths but not strong effects for live births.

[27]Given the serial correlation structure observed in the data, it is worth considering the appropriateness of including a lagged dependent variable in the regression specification. There are three reasons why including a lagged dependent variable in this specification is not a good idea. First, the autocorrelation observed in the mortality data must be a result of underlying factors that affect mortality levels for multiple weeks at a time, rather than the direct influence of lagged mortality in one week on mortality in the next. Thus, a model that includes lagged mortality is *a priori* misspecified. Second, the identification strategy used in this study means that the key explanatory variables should be independent of any factors that are not directly related to pollution, including those that generate the serial correlation observed in the data. As a result, omitting these factors will not bias my results. Third, including the lagged dependent variable in my regressions is likely to obscure the true effect of pollution on mortality. To see why, suppose that pollution exposure caused by a fog event raises mortality in the week in which the fog occurs but also increased mortality in the next week. If I include a lagged dependent variable in the model, then in the week after the fog event some of the increased mortality will be mis-attributed to the higher mortality in the fog event week, rather than to

account for this, I use Newey-West standard errors that allow for correlation across observations falling within a certain number of weeks of each other. An analysis of the residuals from regressions looking at total mortality using Eq. 2 suggests that serial correlation is a concern in these regressions but that this correlation dies out rapidly (within 1-3 weeks).[28] To be conservative, I allow correlation across windows that are somewhat larger than this – six weeks – except in a few cases where series show no evidence of serial correlation, in which case I calculate robust standard errors.

4.3 Modeling fog formation

To strengthen identification, this study uses weather conditions – temperature, humidity, pressure and precipitation – in order to predict fog weeks. Modeling the formation of fog accurately is notoriously difficult, even with detailed modern weather data, due to the complex set of interactions involved.[29] However, the historical weather series available in this study can be used to generate a rough prediction of fog formation at the week level. Using this predicted model has two main advantages. First, it provides a proxy for fog formation that is independent of pollution levels, addressing potential endogeneity concerns. Second, by using quantitative weather variables to predict fog formation I can generate a measure that does not rely on fog reports being consistent over time. However, the fog event predictions based on weather data are somewhat imprecise and, as a result, in some cases the instrument does not provide sufficient power when cutting the data into particular periods or focusing on particular ages or causes of death. This section briefly summarizes the fog model while further detail is available in Appendix A.3.

I model the conditions that permit fog formation as satisfying a series of necessary conditions characterized by sufficiently low temperature, high humidity, high atmospheric pressure, and low precipitation. Low temperature reduces the amount of moisture that air can hold before condensation occurs. Thus, at a given moisture content the condensation needed for fog formation is more likely when the temperature is lower. Conditional on temperature, higher relative humidity indicates that there is more water in the air that may condense. High atmospheric pressure is associated with fog formation because it typically signals the type of relatively calm conditions needed for radiation fog formation. Precipitation is related to fog formation because high levels of precipitation indicate more and denser cloud cover, while the formation of radiation fog requires clear skies. Thus, the key variable predicting

the lagged effect of the fog, so that the results will fail to capture the true lagged effects of the fog event.

[28] See Appendix A.6.1 for further details.

[29] See, e.g., Gultepe (2007).

fog formation, denoted $PredFOG_t$, is an interaction of four indicator variables based on the four available weather series, while the model includes as controls each of the component indicator variables as well as quadratic controls for each the underlying weather variables.[30]

Conditional on choosing appropriate cutoff values, the interaction of the four indicator variables can provide a sufficiently strong predictor of fog occurrence in a week. However, the choice of appropriate cutoff values involves a trade-off between type I and type II errors. There is no obviously criteria for choosing the weighting between these error types so I explore a variety of alternative cutoff values. In the main text I present results in which fog formation is predicted when all of the following conditions hold in a week: temperature is below 55, humidity is above 85, pressure is above 29.9 and weekly precipitation is below 0.5. These cutoff values do a reasonably good job of predicting fog formation across the full sample period and in several sub-periods as well as pollution levels when direct pollution measures are available. I explore alternative cutoff values in robustness exercises.

It is important to note that the weather variables used to predict fog are based on weekly averages, or a weekly total in the case of precipitation. Even when the weekly averages do not satisfy the conditions for fog formation, the conditions may exist on some subset of days. As a result, despite modeling fog formation using a set of necessary conditions, we should not be surprised that heavy fog events still occur weeks when, on average, these conditions are not satisfied. Similarly, even when all conditions are satisfied, other factors may preclude the occurrence of fog. Thus, as shown in Appendix A.3, the $PredFOG_t$ suffers from both type I and type II errors.

It is possible to use the $PredFOG_t$ variable as an instrument for fog events, but it is important to note that doing so departs somewhat from the standard instrumental variables strategy. In particular, while the occurrence of a heavy fog day is discrete, this masks differences in the intensity of fog events. Because the predicted fog variable identifies the conditions most conducive to fog formation, it is likely to correspond to the most severe fog events. The more restrictive the cutoff values used to generate the $PredFOG_t$ variable, the stronger the set of fog events that this variable identifies. This means that when $PredFOG_t$ is used as an instrument for fog events the resulting coefficients will not reflect the impact of an average fog event, and these coefficients will increase as more restrictive criteria are used to predict fog events.[31]

[30] One may worry that rainfall could be endogenous to pollution levels. If so it would be problematic to use rainfall as an input into the $PredFOG_t$ variable. However, in Appendix A.4 I examine this possibility and find no meaningful relationship between rainfall and measured pollution levels.

[31] This is similar to, but not quite the same as, the more standard impact of instruments on coefficients in the presence of measurement error.

It is interesting to study how the number of predicted fog weeks in a year changes across the study period. This pattern is plotted in Appendix Figure 17. This graph shows a peak in reported fog events in the 1890s followed by a substantial drop in the early 20th century. There has been some speculation that the reduction in the early 20th century may reflect falling pollution levels.[32] However, I find that the same pattern appears in the $PredFOG_t$ variable. This tells us that in fact the reduction in the number of fog events in the early 20th century, relative to the late 19th century, was due to weather conditions that were less favorable for fog formation.

5 Preliminary analysis

As a preliminary step in the analysis, this section uses data from 1951-1962, when consistent direct pollution measures are available, to establish the link between fog events, pollution levels, and mortality.[33] To estimate the relationship between pollution and fog events I use,

$$POL_t = \alpha + \sum_{s=-\tau}^{\tau'} \beta_s FOG_{t+s} + X_t'\gamma + Y_t + W_t + \epsilon_t \tag{3}$$

where POL_t is the level of pollution in week t (either the mean or maximum), FOG_t is a measure of fog events such as the number of heavy fog days in week t, X_t is a set of control variables (temperature, temperature squared, etc.), Y_t is a set of year effects, and W_t is a set of week-of-the-year effects.[34] This specification looks at how leads and lags of fog events are related to pollution levels within the range $[-\tau, \tau']$. An analysis of the residuals provides no evidence of serial correlation, so I use robust standard errors in this specification.[35]

Figure 2 presents results comparing the pollution level to leads and lags of the number of fog days in a week. These results show that pollution levels were substantially elevated in weeks in which heavy fog occurred, while there is no evidence of higher pollution levels

[32] See, e.g., Troesken & Clay (2011).

[33] In the Online Appendix I plot weekly pollution levels from 1951-1952 against heavy fog days as well as predicted fog weeks. These figures show that the highest pollution levels occurred during weeks in which heavy fog was reported. The weeks with the most severe fog events, as well as the highest pollution levels, were also weeks when $PredFOG_t$ predicts fog formation. It is worth noting that the $PredFOG_t$ variable captures mainly the most severe fog events.

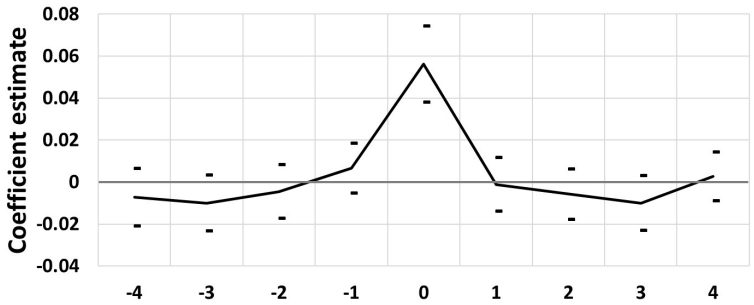
[34] Given the shorter time period over which these data are available I do not allow the week-of-the-year effects to vary by decade in this specification.

[35] I have also generated results using Newey-West standard errors with 6 week lags, following the approach used in the main analysis. These look very similar to the results described in Figure 2 with only slightly larger confidence bands. This is not surprising given that there is little evidence that serial correlation is an important issue.

either in the weeks before or the weeks after a heavy fog week. The fact that there is no evidence of higher pollution levels in the weeks before fog events suggests that the onset of fog in a particular week was not driven by underlying emission levels, providing support for my identification strategy. In Appendix **??** I present results from a variety of alternative specifications, all of which show similar patterns. Appendix Figure 15 shows that similar results are also obtained when I use $PredFOG_t$ as the explanatory variable in place of actual fog events.

Figure 2: Estimated relationship between fog and pollution levels

Results from regressions of pollution levels on heavy fog days



Pollution measures are averages of the maximum daily values reported in the week. Regression includes controls for temperature and temperature squared as well as a full set of year and week-of-the-year effects. Confidence intervals are based on robust standard errors. The Durbin-Watson statistic for these regressions is 1.475, suggesting that serial correlation is not likely to be an important concern.

Next, I consider the impact of pollution, as predicted by fog events, on mortality. The first column of Table 1 presents results from a naive regression comparing the maximum pollution values in each week to total mortality.[36] I include two leads and several lags of the pollution variable to study lagged effects as well as to evaluate whether endogeneity is likely to be a concern. I find that elevated pollution levels were associated with higher mortality across most of the four subsequent weeks. In addition, there is some evidence of a statistically significant negative relationship between mortality and future pollution levels, which suggests that endogeneity may be a concern when pollution is used as the explanatory variable.

In Column 2 of Table 1 I replace pollution levels with the number of fog days in a week. This is the approach that will be used in the main analysis since direct pollution measures are not available for most of the study period. These regressions suggest that fog events

[36]Similar results are obtained if instead I use mean weekly pollution levels.

were associated with increased mortality across the next four weeks. Importantly, unlike the results in Column 1, the estimates show no evidence of a relationship between mortality and future fog events. This suggests that fog events can help deal with the endogeneity concerns that may be present when using pollution as the explanatory variable.

Column 3-4 presents results from an IV regression where contemporaneous and lagged fog days are used to instrument for contemporaneous and lagged pollution levels. In Column 3 I omit leading values in order to increase the strength of the instrument, while Column 4 includes both leading and lagged values. These results show patterns that are similar to the reduced form estimates; higher pollution increases mortality contemporaneously across the next four weeks and this effect fades away by the fifth week. Note that, relative to the results in Column 1, the IV estimates imply a much stronger relationship between pollution levels and mortality. This suggests that either higher pollution levels were endogenously related to other factors that reduced mortality, such as cheap coal prices or higher incomes, or that using fog events as instruments corrects for some measurement error in the pollution variable. Finally, Column 5 presents results that compare the predicted fog event variable to mortality. The results here are fairly similar, with mortality increasing in the three weeks following a fog event. Again, there is no evidence of elevated mortality in the weeks before predicted fog events.[37]

Overall, the results in Figure 2 and Table 1 suggest that fog events are associated with increased pollution levels and that this pollution raised mortality in a window of roughly three to four weeks after each event. Important for the remainder of this study, the reduced form regressions using fog events appear to do a reasonably good job of capturing the pattern of impact of pollution on mortality. In addition, there is no evidence that the week-to-week timing of fog events were endogenously affected by changes in pollution levels.

[37]I do not present results using the predicted fog variable as an instrument for pollution because this instrument is not strong enough across the relatively short period for which pollution data are available.

Table 1: Effect of pollution and fog days on mortality, 1951-1961

	DV: Log total mortality				
	OLS	**Reduced form**	**IV**		**Pred. Fog**
	Using max pollution as the explanatory variable	Using fog days as the explanatory variable	Using fog days as an instrument for pollution		Using predicted fog as the explanatory variable
	(1)	(2)	(3)	(4)	(5)
Future events					
Pollution or fog days in t+2	-0.0535	0.000902		0.0510	0.0257
	(0.0396)	(0.00440)		(0.0708)	(0.0258)
Pollution or fog days in t+1	-0.105**	-0.00190		0.00408	0.0328
	(0.0424)	(0.00475)		(0.0705)	(0.0253)
Contemporaneous					
Pollution or fog days in t	-0.0102	0.00692	0.129	0.146	0.0111
	(0.0394)	(0.00472)	(0.0924)	(0.104)	(0.0316)
Past events					
Pollution or fog days in t-1	0.0930*	0.0148**	0.217**	0.224**	0.0443*
	(0.0561)	(0.00697)	(0.0964)	(0.101)	(0.0251)
Pollution or fog days in t-2	0.0628*	0.0172***	0.235***	0.236***	0.0607**
	(0.0353)	(0.00504)	(0.0649)	(0.0668)	(0.0261)
Pollution or fog days in t-3	0.0222	0.00426	0.0778	0.0887	0.0616*
	(0.0539)	(0.00605)	(0.0861)	(0.0910)	(0.0354)
Pollution or fog days in t-4	0.101**	0.0140**	0.200**	0.199**	0.0224
	(0.0465)	(0.00662)	(0.0850)	(0.0839)	(0.0330)
Pollution or fog days in t-5	0.0258	-0.00318	-0.0134	-0.00712	0.00539
	(0.0426)	(0.00435)	(0.0637)	(0.0647)	(0.0287)
IV f.s. F-stat			11.41	5.71	
Observations	494	494	494	494	494

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. Regression also includes controls for temperature, temperature squared, two leads and five lags of these variables, log births, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared, year effects and week-of-the-year effects. In addition to the controls included in the other regressions, the results in Column 5 include controls for two leads and five lags of the temperature, pollution, precipitation and pressure indicator variables that were interacted to produce the predicted fog event variable. Regressions run on data from the second week of 1951 until the 22nd week of 1962. There are some gaps in the data over weeks in which no pollution data were reported. Pollution values are maximum daily values averaged across each week.

6 Main analysis

The main analysis looks at the impact of acute exposure to elevated levels of pollution due to fog events on mortality across the full 1866-1965 period. I begin by studying total mortality before looking at separate age groups and causes of death.

6.1 Total mortality

The first set of results looking at the impact of fog events on total mortality, in Figure 3, are based on the specification from Eq. 1. The figure describes coefficients and 95% confidence intervals for estimates obtained using the full set of available data.[38] The key explanatory variable in the top panel is the number of fog days in a week, though very similar results or obtained if I use an indicator for whether any fog event was reported in the week. In the bottom panel the explanatory variable is predicted fog events.

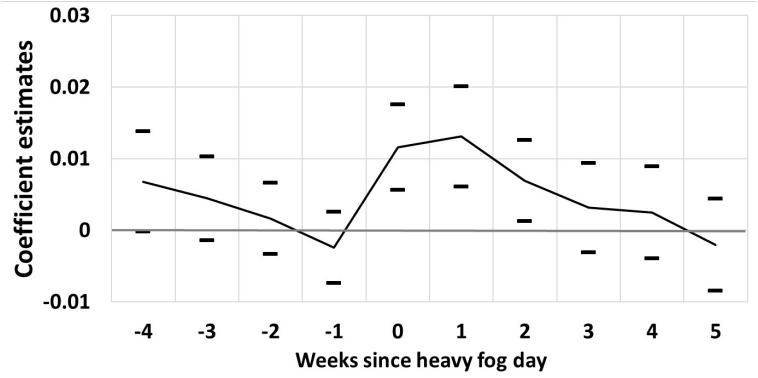
The most obvious feature in Figure 3 is the sharp jump in mortality in the week in which a fog day occurs followed by a peak in mortality in the next week.[39] There is evidence that mortality remains elevated for 2-3 weeks after a fog event. There is no evidence that mortality was higher in the weeks just before a fog event occurs. This provides some confidence in the identification strategy. Instead, in the weeks leading up to fog events the data show a clear downward trend in mortality, though this disappears when focusing on predicted fog events. The cause of this downward trend is not clear, though a likely explanation is that the typically mild weather conditions under which fog events formed were relatively healthy. If so then this suggests that the estimated effects documented in this paper may mildly understate the true impact of pollution exposure.

[38] It is worth noting that this analysis covers fog events in both the Greenwich period and the Kew Gardens period, which introduces some inconsistency. However, the results are similar if I consider only the period during which observations come from Greenwich.

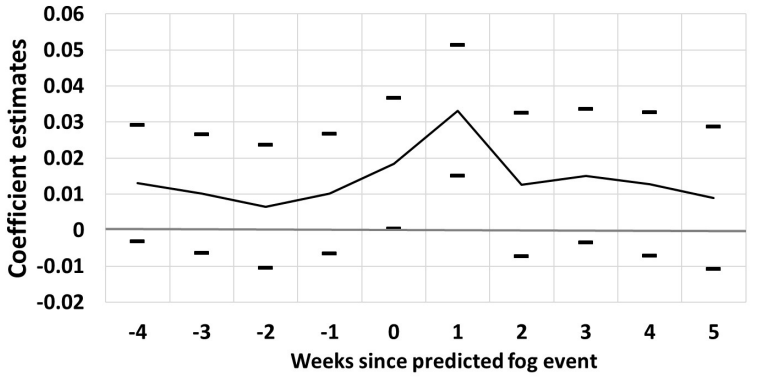
[39] Temperature controls (not reported) show a strong but non-linear relationship to mortality in these results, with both high and low temperatures associated with increased mortality. Leading values of temperature have no relationship to mortality. Lagged values suggest that temperature continued to affect mortality for several weeks, with both low and high temperatures associated with increased mortality. Log births are positively related to mortality but the inclusion of this control has little impact on the results.

Figure 3: Estimated effect of fog events on total mortality, 1866-1965

Using fog days as the key explanatory variable



Using predicted fog event weeks as the key explanatory variable



Coefficient estimates and confidence intervals for a regression of log total mortality on the number of fog days in a week (top panel) or on weeks with predicted fog events (bottom panel). Confidence intervals are based on Newey-West standard errors allowing correlation across observations within 6 weeks of each other. The regression includes controls for temperature, temperature squared, four leads and five lags of temperature and temperature squared, log births, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared, a full set of year effects and a full set of week-of-the-year by decade effect. The bottom panel also includes the temperature, pressure, humidity and precipitation indicator variables used to construct the $PredFOG$ variable as well as leads and lags of these variables. Data cover 1866-1965. N=4,479.

The next set of results, in Table 2, summarize the overall magnitude of the effect of fog events on mortality in the event week and the following three weeks using the specification from Eq. 2.[40] Focusing on this four-week window is motivated by the results in Figure 3. I use this approach in many of the subsequent tables because it provides a simple summary of the acute impact of pollution exposure. In Columns 1-2 the key explanatory variable is

[40]Appendix A.6.2 presents results obtained using windows of different lengths ranging from two to five weeks.

the number of fog days in a week and the three previous weeks. In Column 3 I instead use an indicator for whether a fog event happened in the week or in any of the three previous weeks. In Column 4 I use predicted fog weeks as the main explanatory variable. In Columns 5-6 I use predicted fog weeks as instruments for fog days or fog weeks, respectively, in the four-week window.

Results looking at the impact of fog days and including controls, in Column 2, indicate that a fog day raised mortality by 0.845 percent across a four-week window starting in the fog event week. I can use these estimates to quantify the total share of mortality that can be directly attributed to the acute effects of pollution exposure generated by fog events. In particular, across the study period there were 932 heavy fog days and an average of 1247 deaths in a week. Given the results in Column 2, this implies an additional 39,448 deaths in London across the years studied. There were a total of 5.65 million deaths in London in the weeks covered by these data. Thus, these results suggest that 0.7 percent of all deaths experienced in London during the years covered by this study are directly attributable to the acute effect of heavy fog events. The magnitude is slightly larger (48,887 deaths) if I focus instead on the results using the fog week indicator in Column 3.[41] The effects implied by the results in Column 4 are somewhat smaller, despite the fact that the estimated effect of a predicted fog event is larger, because there are just 282 of these. The estimates in Column 4 imply that fog events led to 30,425 deaths or 0.54 percent of all deaths in London during the study years. In Columns 5-6 I estimate results using predicted fog events to instrument for actual fog events. When the coefficient in Column 5 is applied to the 282 predicted fog events I estimate 42,548 deaths, or 0.75 percent of all deaths. Thus, I conclude that at least one out of every 200 deaths in London during the years covered by my data, and probably more, are associated with acute effect of fog events.

One way to put these magnitudes into context is to compare the deaths associated with acute pollution effects to totals from other important causes of death for the period before 1940, when consistent cause-of-death series are available. Applying the approach in Table 2 to data ending in 1939, I estimate that the acute effects of fog events caused 26,000-35,000 deaths in that period. Thus, the acute effects of pollution were similar in size to total deaths due to suicide (30,400), venereal disease (23,891), or smallpox (23,366), and roughly half as large as the total impact of important infectious diseases like diphtheria (56,848 deaths) or scarlet fever (56,216 deaths), but quite a bit smaller than the most important causes (which air pollution contributed to) such as TB (561,583 deaths), cardiovascular diseases (549,812), pneumonia (331,956), or measles (360,756).

[41] Similar magnitudes are also obtained if I use estimates from each individual lagged fog day variables, such as those shown in Figure 3.

The estimated effects of the control variables (not reported) show reasonable patterns. By far the most important control variables are temperature and temperature squared, which show a clear non-linear relationship, with more deaths at both high and low temperatures. These effects appear both contemporaneously and for several lagged weeks. Mortality is greater in weeks with more births and there is some evidence that high humidity also increases mortality.

Table 2: Effect of fog events on total mortality in four week windows

	DV: Log total mortality					
	OLS	OLS	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Fog days	0.0130***	0.00845***			0.0298***	
(4 week window)	(0.00281)	(0.00234)			(0.00782)	
Fog week ind.			0.0167**			0.137***
(4 week window)			(0.00688)			(0.0381)
Pred. fog weeks				0.0214**		
(4 week window)				(0.00947)		
Additional controls		Yes	Yes	Yes	Yes	Yes
IV f.s. F-stat					114.6	54.7
Observations	4,479	4,479	4,479	4,464	4,464	4,464

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. Regressions run on the full set of available observations from 1866-1965. All regressions include year effects and week-of-the-year by decade effects. Regression in Columns 2, 4 and 6 include controls for log births, humidity, humidity squared, pressure, pressure squared, precipitation, precipitation squared, temperature, temperature squared, and five lags of temperature and temperature squared. Column 6 also includes as controls the temperature, pressure, precipitation and humidity indicator variables used to produce the predicted fog variable, as well as five lags of each of these variables. The inclusion of these lagged terms causes a small reduction in the number of observations.

In Appendix A.6.3 I present some additional robustness results using the approach in Table 2. These show that the main findings are robust to excluding the births control or to including additional temperature quadratic terms. I also estimate results including the squared number of fog days. These provide no strong evidence that fog days had a non-linear impact on total mortality. Other results show that including several lags of the other weather variables (pressure, humidity, precipitation and squared values of these variables) has little impact on the results.

The most interesting results in Appendix A.6.3 show that when fog days are interacted with temperature I see a significant negative coefficient on the interaction term. Because home heating was a primary driver of air pollution emissions, these interactions tell us that fog days had a more severe impact in weeks in which emissions levels were higher. This makes sense given that the primary effect of the weather conditions that accompanied fog

events was to trap in the pollution emitted in the city. In terms of magnitudes, these results suggest that when the temperature was ten degrees (F) lower, the impact of a fog day on mortality increased by 16-18%.[42]

It is also possible to look at how these effects evolved over time, which is done in Appendix A.8. These results suggest that the impact of fog events as a percentage of total mortality was relatively stable over time, with some evidence of a mild (not statistically significant) decline.

Next, I study how these effects were distributed across age groups. This analysis follow the same empirical approach applied to total mortality, but with dependent variables that reflect death within particular age groups. As discussed in Section 3, the age groups I consider aim to provide fairly consistent series despite changes in the age categories reported across time. My main analysis focuses primarily on results looking across four-week windows using the specification in Eq. 2.

Table 3 present estimates of the impact of fog days on mortality by age group.[43] These results show that fog days had a substantial effect on mortality across all age groups except for infants (a group I return to later), with the largest effects, in terms of the percentage increase in mortality, occurring among children aged 1-5. The most important age group in terms of overall number of deaths associated with fog events was older adults, with over 40 percent of fog event deaths occurring among the elderly (those aged over 60 or 65).

The most puzzling finding in Table 3 concerns infants, a group where modern results would lead us to expect strong pollution effects. One factor behind this results is that I am looking across four-week windows. If instead I focus just on the week or two during or just after a fog week I do estimate positive coefficients. Specifically, I estimate that a fog day results in an increase in infant deaths equal to 53 deaths per 100,000 live births in the following week, though this result is not statistically significant at standard confidence levels. Setting aside issues of statistical significance, the magnitude would imply an additional 4,266 deaths across the study period, or about 10% of all acute deaths attributable fog events.

[42]This feature helps explain some of the worst fog events. During the Great London Fog of 1952, for example, temperatures hovered in the mid-30s F. As a point of comparison, a similar number of fog days occurred in late 1953, but with temperatures generally above 50 F the impact on mortality was relatively modest. The increased impact of fog events during periods in which temperatures were low is noted in contemporary sources such as Ministry of Health (1954).

[43]Results using fog week indicator variables are similar. I do not present results obtained using the predicted fog events because using the relatively imprecise fog even predictions together with the noisier data on mortality by age group generates results with standard errors that are too large to draw clear conclusions, though the coefficient estimates are generally similar to those shown in Table 3.

Table 3: Mortality effects by age group

	DV: Log mortality					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	-0.000596	0.0154**	0.0102***	0.00632***	0.00917***	0.00983***
(4 week window)	(0.0104)	(0.00639)	(0.00336)	(0.00231)	(0.00220)	(0.00254)
	Number of additional deaths due to a fog day over a four week period					
		7.87	2.76	3.60	9.63	16.30
	Age group share of deaths associated with a fog event					
		0.20	0.07	0.09	0.24	0.41
	Total implied deaths by age group					
		7,289	2,560	3,338	8,920	15,098
Observations	3,905	3,888	4,192	4,192	4,192	4,192

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. All results include controls for temperature, temperature squared, four lags of temperature and temperature squared, five lags of temperature and temperature squared, log births, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared and a full set of year effects and a week-by-decade effects. The results in Column 1-2 use data from 1876-1965 but a small number of observations are dropped in Column 2 because they have zero deaths. The other columns use data from 1870-1965. The implied number of deaths is calculated by applying the estimated percentage increase in deaths based on the regression coefficient and the average number of deaths in a week in each age group multiplied by four to reflect the four-week window over which effects occur.

As a validation exercise, it is useful to compare my estimates to the results of existing studies focused on infant mortality. A useful point of comparison is Arceo *et al.* (2016) which suggest that a one unit increase in TSP raises infant mortality in a week by 0.42 deaths per 100,000 live births.[44] If I use data from the period for which I have direct pollution measures and regress infant mortality on the weekly average of maximum daily pollution levels while instrumenting for pollution with fog days, I estimate that a one unit increase in TSP raised infant mortality by 0.41 deaths per 100,000 births. While this estimate is not precisely measured due to the limited set of available pollution data, the fact that it is very similar to the results obtained from previous work is comforting. If instead I use the estimated impact of fog days on infant mortality across the full study period and apply the relationship between fog days and pollution obtained from Figure 2, I would conclude that a one unit increase in TSP raised infant mortality by 0.96 deaths per 100,000 live births. This suggests that the impact of fog events on infant mortality may have been larger earlier in the study period, but again, the estimates are not precise enough to draw any strong conclusions.

It is interesting to note that I find some evidence of an increase in infant deaths in

[44]They estimate that a one unit increase in PM10 increases weekly infant mortality by 0.23 deaths per 100,000 births. I convert this to TSP by dividing by 0.55, a conversion factor that is obtained from their study.

the week in which fog events occur and the following week, but estimates across four-week periods are very close to zero. This suggests that infant deaths occurring right after fog events may have been due to the harvesting of infants who would have died rapidly of other causes anyway. This finding contrasts with results from modern studies, such as Arceo *et al.* (2016), which finds very similar results when estimating weekly and monthly effects. One likely explanation for this difference is that more rudimentary medical care in the setting I study meant that more infants were likely to die of causes unrelated to pollution than in modern settings. In this case, pollution exposure in my setting would have accelerated the deaths of infants who would have died anyway, while in modern settings pollution may have led to the deaths of infants who would have survived otherwise thanks to better medical care.

In Appendix A.8 I provide additional results looking at how the distribution of effects across age groups evolved over time. The most striking finding from these results is that there was a substantial decline in the share of deaths associated with fog events accounted for by those from 1-20 years old after 1900. This decline suggests that there was some change that caused a reduction in the impact of pollution exposure on this age group relative to other ages. Later, when I come to the cause-of-death results, I will provide evidence suggesting that a change in the infectious disease environment can account for this shift.

It is also interesting to look at the impact of pollution on stillbirths, which are observable starting in 1927. These results, available in Appendix 17, show that fog days led to a statistically significant increase in stillbirths across a four-week period equal to 0.7 percent, or about 15.7 stillbirths per 100,000 live births. If I apply these estimates across all the years covered by my study, they suggest that acute exposure due to fog days caused to 1,143 additional stillbirths.

I have also looked for evidence of the effects of in utero exposure to fog events. While the data are generally not well suited for looking at in utero effects, because deaths cannot be linked to birth dates, this is possible for two types of deaths. One of these is deaths due to prematurity, which typically occur close to the date of birth. This allows me to roughly infer exposure in earlier trimesters. The second is stillbirths which, conditional on occurring near the end of a regular term, can also allow me to infer first-trimester exposure. In Appendix A.9 I analyze how these series are related to previous fog events. For both series, I observe a strong positive relationship to fog events 25-36 weeks earlier, or roughly in the first trimester. In terms of magnitudes, the estimated impact of first-trimester exposure to fog events on deaths due to prematurity imply 1,436 additional infant deaths across the study period. The first-trimester effect on stillbirths imply an additional 3,509 stillbirths in the years covered

by this study.

To summarize, the results in this subsection show that the effect of acute pollution exposure associated with fog events on health in London was substantial and that these effects were felt across all age categories. However, we may worry that these deaths simply reflected the harvesting of individuals who otherwise would have died soon after. If this were true, then the larger number of deaths associated with fog events may not indicate a large loss in life expectancy. In the next subsection I address this possibility.

6.2 Medium-run effects and harvesting

This section considers the medium-run effects – up to one year – of the high levels of pollution exposure generated by fog events. The analysis follows the same basic approach used in the previous sections, but adds in a variable capturing the number of fog events in the past year. One reason to look for evidence of these medium-run effects is to see whether fog events have some additional consequences not picked up by the analysis of acute effects in the previous sections. A second motivation is that I can look at whether harvesting might be an issue, i.e., whether the acute effect of fog events is simply reflecting mortality of individuals who would have died soon after anyway. If there is a substantial harvesting effect, then I should see that fog events are associated with reductions in mortality in the medium-term.

The identification issues faced when looking at medium-run effects are more substantial than those present when looking only at acute effects. In particular, people may move between cities over the course of a year, which may partially obscure the effect of fog exposure. However, this source of bias is unlikely to be large since a relatively small fraction of the population of a city like London moves in any given year. In addition, there is greater concern that the frequency of fog events in the past year may be correlated with higher baseline emissions levels that are related to other factors that influence mortality. These caveats should be kept in mind when evaluating the results in this section.

To implement the medium-run analysis I include in my preferred regression specification (Eq. 2) an additional variable reflecting the number of fog days reported in a 52-week period starting just after the four week period that I use to identify acute effects. I look across a full year in order to eliminate the possibility that seasonal variation in the recent past can drive the results.[45] Also, in order to exploit variation at this level it is necessary to replace

[45]If instead I looked at, say, events over the past six months, then doing so for a week in December would cover a set of past events that look very different than a week in June in the same year. Since this difference will be systematically related to the season of the observation week, which affects mortality, this could generate bias in the results.

the year effects in Eq. 2 with decade effects. Finally, this analysis focuses only on data up to WWII, because after the war the larger number of fog events observed at Kew Gardens will substantially influence the medium-run effect variable.

Table 4 presents results looking at medium-run effects across all age groups. In Column 1 I include both my standard measure of acute pollution effects (over a four-week window) and a second variable reflecting the sum of all fog days in the 52 weeks prior to that window. Both of these show a positive and statistically significant relationship to mortality. The fact that fog exposure in the recent past does not reduce later mortality runs counter to the idea of harvesting, or at least suggests that any harvesting effects are overwhelmed by the medium-term effect of pollution exposure.[46] In terms of magnitude, the impact of a single fog event on mortality within a week is several times larger if the event occurred in the past four weeks rather than in the 52 weeks before that. However, the overall impact of the estimated medium term effect is about three times as large as the direct acute effect because the acute effects occur across only four weeks while the medium term effect is spread across a full year. In total, the estimates in Table 4 imply that acute effects account for 24% of the total mortality impact of fog events while the medium-term impact accounts for 76%, ignoring any longer-term effects occurring outside of one year. In terms of magnitude, the results in Column 1 suggest that the medium-run effects of fog events accounted for 1.8% of all deaths in London during the period up to WWII.

Table 4: Medium-run results for mortality in all age groups

	DV: Log mortality	
	(1)	(2)
Fog days - acute effects (4 week window)	0.0102***	0.0110*
	(0.00335)	(0.00567)
Fog days - medium run (next 52 weeks)	0.00246**	0.00252**
	(0.00103)	(0.00102)
Fog days (4 weeks) × fog days in recent past		-8.68e-05
		(0.000747)
Observations	3,275	3,275

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. All regressions include controls for temperature and temperature squared, five lags of these variables, log births, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared and a full set of week-by-decade effects. The data cover 1870-1939 but data from 1870 and 1919 are dropped from the analysis due to the need to construct one year lagged fog event counts. For consistency I avoid using data after WWII, when the larger number of fog events affects the variation in the medium-run explanatory variable.

[46] This results is also consistent with Bell & Davis (2001), which link the Great London Smog in December 1952 to high levels of mortality that persisted through February 1953.

Column 2 adds in an interaction between contemporaneous and past fog events in order to look at whether having more fog events in the recent past reduces the contemporaneous population at risk of mortality due to fog exposure. The coefficient term on this interaction is very close to zero, so there is no evidence that fog events in the recent past reduced the population at risk of dying from later fog events. This provides a second and even stronger form of evidence suggesting that harvesting is not driving my results.

Some interesting patterns appear when these medium-run effects are broken down by age group, in Table 5. In particular, while I do not observe strong acute effects for infants (ignoring stillbirths), there is evidence of substantial medium-run impacts. These may be due at least in part to in utero exposure as shown in Appendix A.9. For children aged 1-5, I find no evidence of medium-run effects. This contrasts with the clear evidence of acute effects. For older adults, both acute and medium-run exposure are associated with increased mortality. In the next section I use information on causes of death to shed light on these patterns.

Table 5: Medium-run results by age group

	DV: Log mortality					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	-0.000850	0.0148*	0.000492	0.00637	0.0138***	0.0133***
(4 week window)	(0.00450)	(0.00827)	(0.00535)	(0.00405)	(0.00390)	(0.00435)
Fog days in	0.00324**	-0.00142	-0.000460	0.00130	0.00278**	0.00369***
medium run	(0.00152)	(0.00250)	(0.00177)	(0.00133)	(0.00114)	(0.00137)
Observations	3,039	3,039	3,091	3,091	3,091	3,091

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. All regressions include controls for temperature, temperature squared, five lags of these variables, log births, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared, and a full set of week-of-the-year by decade effects. The data in Columns 1-2 cover 1876-1939 while those in the remaining columns cover 1875-1939.

6.3 Effects by cause of death

To study how acute pollution exposure interacted with the disease environment it is useful to break the effects down by cause of death (COD). This section examines results for 23 aggregated cause of death categories that appear to be fairly consistent over time. I do not examine one category, maternal mortality, which is investigated in more detail in a separate paper (Hanlon & Sudol, 2017).[47] The cause-of-death data cover 1870 to 1939.

[47]Our motivation for offering a separate analysis of the impact of pollution on maternal mortality is that this is an area where the link to pollution has not been previously established in the literature. This calls

Table 6 presents coefficient estimates for regressions comparing weekly mortality in each COD category to the number of fog days in that week and the previous three weeks. The first few rows describe mortality in a set of causes of death related to the respiratory and cardiovascular systems, the areas most clearly linked to the impact of air pollution in modern studies.[48] Consistent with existing results, all of these categories show increases during or just after fog events. Bronchitis deaths show the strongest increase, rising by 3.5% with each additional fog day. In terms of total deaths, this category alone explains more than one-third of all of the deaths associated with fog events during the period covered by the cause-of-death data. Cardiovascular diseases and pneumonia also show large effects. Respectively, these explain 10.0 and 22.2 percent of deaths associated with fog events. Other respiratory causes-of-death, a basket of diseases which includes asthma and influenza, also show a positive association with fog events, though this result is not statistically significant.[49] Together, the four respiratory and cardiovascular categories at the top of Table 6 account for more than two-thirds of the deaths associated with fog events. Outside of fog days these causes account for less than one-third of total mortality.

The next group of major infectious diseases show mixed effects. Diseases affecting the respiratory system, such as measles and tuberculosis, show increases associated with fog events, as does scarlet fever, which is an airborne diseases of the throat and larynx.[50] Diphtheria, another disease of the throat also shows a positive association with fog events which is marginally statistically significant.

It is interesting to note that many other infectious diseases, such as digestive diseases (diarrhea, cholera and typhoid), smallpox, a viral skin disease, or the "other" infectious diseases, do not show increases during fog events. Others, such as diphtheria and typhus, show weak effects that disappear in many specifications. These findings are useful because many of these diseases are highly contagious and at least some of these diseases show very clear symptoms and were therefore relatively easy to diagnose, even in the 19th century. The fact that I see no effect for these infectious diseases provides additional evidence that

for a deeper analysis that considers the physiological channels through which pollution may be affecting mothers' health.

[48] See, e.g., the review by Rückerl *et al.* (2011).

[49] The impact of fog events through respiratory diseases is statistically significant if I focus only on the period before WWI. In the inter-war period there are large outbreaks of influenza which make the data in this category very noisy.

[50] Of course, one may worry that in fact deaths occurring from the direct effects of pollution are being erroneously assigned to these infectious disease categories and that this may be generating the effects shown in Table 6. This is unlikely to be the case for measles or scarlet fever since these diseases show distinct symptoms that would have been well-known during the study period. It is a concern for TB, which was difficult to diagnose and may have been mixed up with deaths due to other respiratory conditions (Woods, 2000).

the increase in mortality associated with fog events is not likely to have been caused by an increased spread of disease as people crowded together indoors on foggy days.

The next grouping contains a random assortment of diseases. Here it is comforting to see that fog events are not positively associated with deaths due to factors such as cancer or neurological diseases. These are causes where, a priori, we would not expect *acute* air pollution exposure to play a major role, though chronic exposure may be an important factor, such as in the development of lung cancer.[51] Deaths due to old age, a category that is somewhat vague, do increase during fog events, though the coefficient is not statistically significant. Premature birth, one of the most important causes of death among infants, does not appear to increase as a result of acute exposure. However, results in Appendix A.9 show that deaths due to prematurity are affected by in utero exposure in the first trimester.

The last group of causes includes accidents, violence, and suicide. Classic stories of fogs during the Victorian era emphasize their contribution to crime and accidental deaths. The results in Table 6 do not show a statistically significant effect of fog events on deaths due to accident or homicide across a four-week period. However, results using separate leads and lags of fog events show strong evidence of an increase deaths due to accidents or violence reported in the week following fog events, consistent with contemporary reports. The magnitude of these effects, however, mean that accidents and violence cannot be a primary driver of the impact of fog events on overall mortality.

We can draw two main lessons from the results in Table 6. First, the fact that the effect of fog events was concentrated in respiratory and cardiovascular diseases, while I find no effect for other causes of death that are unlikely to be associated with pollution exposure – digestive diseases, neurological diseases, venereal diseases, and alcoholism, for example – indicates that the identification strategy is working well. If the effects of fog events were due to factors other than pollution exposure, such as difficulty reaching medical care, then we would not expect to see the effects so concentrated in respiratory and cardiovascular diseases.

Second, these results show that fog events interacted with a specific set of infectious diseases of the respiratory system, throat, or larynx in a way that increased mortality from these diseases. Together, these diseases – measles, scarlet fever, and TB – account for 32 percent of the total deaths associated with fog events.[52] The channels through which pollution increases infectious disease mortality are not currently well-understood. It may be that pollution exposure puts additional stress on those infected, causing a larger fraction to die, or it may be that pollution facilitates the spread of infectious disease. Understanding

[51] In fact, cancer is sometimes used as a placebo category when looking at acute pollution effects, as in Jia & Ku (2017).

[52] These diseases account for 17 percent of deaths overall.

these channels is left for future work.

Table 6: Mortality effects by reported cause of death

DV: Log all-age mortality within disease category				
Cause of death	Coefficient		S.E.	No. deaths due to a fog day in 4 week window
Respiratory & Cardiovascular				
Bronchitis	0.0349	***	(0.00447)	18.9
Cardiovascular	0.0082	***	(0.00237)	5.3
Pneumonia	0.0298	***	(0.00402)	11.7
Misc. respiratory	0.00467		(0.00683)	
Infectious diseases				
Digestive dis.	-0.0042		(0.00373)	
Diphtheria	0.0135	*	(0.00743)	0.91
Measles	0.0791	***	(0.0137)	10.6
Scarlet Fever	0.0349	***	(0.00805)	2.3
Smallpox	-0.00355		(0.0321)	
Tuberculosis	0.00615	***	(0.00214)	4.0
Typhus	0.0340	*	(0.0179)	0.2
Whooping cough	-0.00158		(0.0101)	
Infectious, other	0.00022		(0.0073)	
Other diseases				
Cancer	0.00134		(0.00238)	
Neurological dis.	-0.0021		(0.00215)	
Old age	0.0041		(0.00367)	
Premature birth	-0.00053		(0.00389)	
Venereal diseases	-0.0021		(0.00793)	
Other misc. causes	0.00076		(0.00162)	
Other causes of death				
Accidents/violence	0.0022		(0.00359)	
Alcoholism	0.00832		(0.00941)	
Homicide†	-0.0074		(0.0122)	
Suicide	-0.00151		(0.00837)	
All causes	0.00970	***	(0.00211)	52.8

Estimated coefficients and robust standard errors on the impact of fog events in a week or in the previous three weeks on mortality within each cause-of-death category. Robust standard errors are used because many of the series do not show evidence of serial correlation. For the series where there is evidence of serial correlation (define as a Durbin-Watson statistic under 1.5), results using Newey-West standard errors are available in Appendix Table 24. For those categories that show clear increases during fog events, the last column presents the number of deaths resulting from a fog day across a four-week period. All regressions include controls for temperature, temperature squared, five lags of each of these variables, log births, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared, and a full set of year and week-of-the-year by decade effects. † Homicide data are only separately available from 1870-1921. Homicide deaths are also included in the accidents and violence category.

In Appendix A.10 I present some additional results looking at causes of death. One set of results use an alternative analysis approach in which I estimate the effect of fog days

across a four-week window starting in the week in which the fog occurred and then compare these estimates to the effect in a four-week window just before the fog day. These results are generally similar to those shown in Table 6 except that I no longer find statistically significant effects operating through scarlet fever, while I do find significant effects through deaths of old age. In another set of results, I use the predicted fog variable to instrument for the timing of fog events. This also generates similar results.

The cause-of-death results can help us understand the distribution of fog event mortality across age groups. In Appendix A.2.4 I provide tables showing the share of mortality in each age group broken down by cause of death. This table shows that the most important causes of death for infants are digestive diseases and premature birth. From Table 6 we can see that these causes are not positively associated with the acute effects of fog events. These patterns may help explain why I do not see a strong *acute* effect of fog events on infant mortality. For children aged 1-5, the most important causes of death are infectious diseases, including measles, pneumonia, tuberculosis, whooping cough and bronchitis (in that order). Deaths from all of these, other than whooping cough, increase during fog events. This explains the strong effect of fog events on mortality for the 1-5 age group shown in Table 3. For the young group and prime age adults, the most important cause of death by far was tuberculosis, which accounted for 23% and 39% of mortality in these age groups, respectively. This, together with the connection between tuberculosis mortality and fog events documented in Table 6, helps explain why I find a substantial effect of fog events on adult mortality. Importantly, this also suggests that the impact of pollution on adults in modern developed countries is likely to be weaker, since TB is much less common in developed countries today, though it remains common in the developing world and is still the most deadly single infectious disease (ranking above HIV/AIDS) (World Health Organization, 2017). For older adults, the most important causes of death were cardiovascular diseases and bronchitis while pneumonia was also important. For those in the 40s to 60s, tuberculosis was also an important factor. All of these are linked to fog events in Table 6.

One implication of the results shown in Table 6 is that the presence of infectious diseases increased the costs of pollution exposure. As an example of this point, consider the impact of fog events on children aged 1-5, a group that I find to be strongly affected by pollution events. In the 19th and early 20th century, this group was strongly affected by infectious diseases, particularly measles, which accounted for 14% of deaths in this age group, as well as TB (10% of deaths) and scarlet fever (5%). Table 7 presents estimates of the impact of fog events on this age group using data from 1875-1914 including all cause of death (Column 1), excluding measles mortality (Column 2) and also excluding scarlet fever and TB mortality (Column 3). We can see that once measles deaths are excluded the impact of fog events on

children aged 1-5 drops by half, and this effect declines further when I also exclude scarlet fever and TB. A similar story holds for prime-aged adults and TB, which accounted for 38% of mortality among that age group. The last two columns of the table show the estimated impact of fog events on adults aged 20-40 including all causes of death, in Column 5, and all deaths excluding TB, in Column 6.

Table 7: Effect of certain infectious diseases on mortality by age group

	For ages 1-5			**For ages 20-40**	
Dependent variable:	All deaths	Excluding measles	Also excluding scarlet fever and TB	All deaths	Excluding TB
Fog days	0.0208***	0.00962*	0.00854	0.00564**	0.00378
(4 week window)	(0.00702)	(0.00579)	(0.00690)	(0.00253)	(0.00300)
Observations	2,023	2,023	2,023	1,878	1,878

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. Data in Columns 1-3 cover 1875-1914. Data in Columns 4-5 cover 1875-1910 because age categories change in 1911. Regressions include controls for temperature, temperature squared, five lags of each of these variables, log births, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared, and a full set of year and week-of-the-year by decade effects.

The cause-of-death results can also help us understand the medium-run estimates by age group in Table 5. For example, the results in Table 5 show that children aged 1-5 were affected only by acute and not by medium-run pollution exposure. Regressions looking at medium-run effects by COD (available upon request) show that only acute pollution effects increase deaths due to measles while medium-run pollution exposure does not. Since measles was a primary driver of mortality in this age group, this can explain why this group does not show positive medium-run effects. In contrast, bronchitis and cardiovascular deaths are increasing in both acute and medium-run pollution exposure. This can explain why both acute and medium-run exposure increase mortality among older adults, where these two causes of death are particularly important. Pneumonia deaths, on the other hand, were only influenced by acute pollution exposure.

The connection between pollution exposure and infectious diseases implies that improvements in medical knowledge, public health technologies, or other factors that reduced infectious disease mortality, also have the potential to reduce the health effects of pollution. For example, the number of measles deaths in London decreased dramatically after WWI from about 42 deaths per week from 1870-1914 to 11 deaths per week in the inter-war period.[53] If

[53]See Appendix Figure 12 for a graph of this pattern. The causes of the reduction in measles mortality are not well-identified, but most likely were due to a combination of improved nutrition, better housing conditions and medical care, and efforts to reduce the spread of the disease. This decline appears across the whole of England, so it is not an endogenous response to changes in pollution levels in London.

I apply the reduction in measles deaths to estimates of the impact of fog events on mortality using data up to WWI, the results suggest that the decrease in measles deaths corresponds to a 16.5% reduction in the total number of deaths associated with fog events and this can explain 50% of the total reduction in deaths due to fog events after 1914. Similarly, TB deaths fell from 201 per week before 1914 to 83 per week in the inter-war period. Applying coefficient estimates from data before 1914 to this change implies that the reduction in TB deaths is associated with a 5% reduction in the mortality associated with fog events and accounts for 12.6% of the reduction in mortality due to fog events after WWI.

A second implication of these results is that changes in the infectious disease burden will alter the distribution of the effects of pollution exposure across age categories. For example, the connection between TB and pollution exposure suggest that declines in TB over the 20th century likely reduced substantially the impact of pollution on prime-aged adults. Similarly, reductions in measles and scarlet fever deaths suggest that pollution should have less impact on children aged 1-5 today. This may help explain why, in modern studies, the impact of pollution on mortality seems to be more heavily concentrated in the very young and the very old.

An alternative way to consider the interaction of pollution with infectious disease is to look at whether the prevalence of certain infectious diseases in the weeks before a fog event raises mortality associated with the event. Unfortunately, disease prevalence is not directly observable. While deaths reflect disease prevalence, when deaths are used to infer prevalence it is difficult to separate the impact of disease prevalence, which may raise subsequent mortality due to pollution, from the effect of harvesting, which may reduce pollution effects. Nevertheless, I have examined whether the impact of fog events is larger when there have been more deaths in certain disease categories in the four weeks before the fog, controlling for the direct impact of previous disease prevalence. Of the major COD categories, for only one, whooping cough, do I observe a strong positive association between deaths in the past four weeks and the impact of fog days on overall mortality. Interestingly, more whooping cough deaths in the past four weeks do not increase the impact of fog events on deaths due to whooping cough. Rather, the additional deaths are due mainly to bronchitis, which some also showing up in pneumonia, measles and scarlet fever. This pattern is actually quite sensible, because whooping cough infections are often associated with subsequent respiratory problems. This highlights another avenue, scarring, through which the infectious disease environment may impact the costs of pollution exposure.

7 Concluding remarks

This paper provides evidence on the impact of acute pollution exposure events associated with fog on mortality patterns in London across the century from 1866-1965. The richness of the available data and the repeated nature of the events that I use for identification allow me to generate a wide variety of results, including tracking the impact of acute exposure across different age groups and the interaction of these effects with other causes of death. This paper presents only a small subset of the results that can be generated from this rich dataset. These data should provide fertile ground for further research aimed at improving our understanding of the long-run evolution of health in London.

One of the interesting results reported in this study is that mortality is increased by the interaction of pollution exposure with a specific set of infectious diseases that primarily affect the respiratory system or throat. One implication of this type of interaction is that the health effects of pollution will be modified across contexts depending on the disease environment. For example, the same level of pollution may generate higher levels of mortality in a place were certain infectious diseases are common than in a place where infectious diseases are rare. This implies that even holding constant the level of pollution, one should be careful in extrapolating from the mortality effects observed in one context. A second implication is that reductions in infectious disease mortality will reduce the health impact of pollution, which in turn will alter the benefits we should expect from pollution regulation.

My results point to two infectious diseases, measles and TB, as showing particularly strong interactions with pollution exposure. This is an important finding because these diseases continue to be important causes of mortality today. The WHO reports that there were 6.3 million new cases of TB in 2016 with 1.3 million TB deaths.[54] This made TB the ninth leading cause of death and the most important single infectious diseases (ranking above HIV/AIDS). Measles, while easily preventable, also remains important, killing 89,780 children in 2016, the first recorded year in which global deaths from this disease fell below 100,000.[55] Thus, public health improvements, and particularly progress against these two diseases, offers an avenue for reducing the health costs of pollution. These results raise the possibility that interventions that improve the disease environment may be able to reduce the health costs of pollution more cheaply than direct pollution abatement. Assessing this possibility is an interesting avenue for future research.

This study also provides new evidence on how the impacts of pollution vary across age

[54]Statistics from See World Health Organization (2017). The 1.3 million deaths figure includes only deaths among HIV-negative people.

[55]The measles statistics are from `http://www.who.int/mediacentre/factsheets/fs286/en/`.

group, the lag structure of these effects, and how this distribution evolves over time. In particular, I find that most age groups experience meaningful pollution effects, but the magnitude and timing of these effects varied across ages. Moreover, the distribution of effects across age groups depended on the disease environment. As a result, the distribution of effects across age groups evolved over time as a result of the reduction in infectious disease mortality that took place during the study period.

References

Ahrens, CD. 2007. *Meteorology Today*. Ninth edn. Brooks/Cole, Cengage Learning.

Almond, Douglas, Chen, Yuyu, Greenstone, Michael, & Li, Hongbin. 2009. Winter Heating or Clean Air? Unintended Impacts of China's Huai River Policy. *American Economic Review*, **99**(2), 184–90.

Alsan, Marcella, & Goldin, Claudia. 2014. *Watersheds in Infant Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880-1915*. NBER Working Paper No. 21263.

Arceo, Eva, Hanna, Rema, & Oliva, Paulina. 2016. Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City. *Economic Journal*, **126**(591), 257–280.

Bailey, R., Hatton, T., & Inwood, K. 2016. *Atmospheric Pollution and Child Health in Late Nineteenth Century Britain*. CEPR Discussion Paper No. 11702.

Ball, Alastair. 2015 (November). *The Lifelong Costs of Urban Smogs*. MPRA Working Paper No. 63229.

Barreca, A., Clay, K., Deschenes, O., Greenstone, M., & Shapiro, J.S. 2016. Adapting to Climate Change: The Remarkable Decline in the U.S. Temperature-Mortality Relationship over the 20th Century. *Journal of Political Economy*, **124**(1), 105–159.

Barreca, Alan, Clay, Karen, & Tarr, Joel. 2014 (February). *Coal, Smoke, and Death: Bituminous Coal and American Home Heating*. NBER Working Paper No. 19881.

Beach, B., & Hanlon, W.W. Forthcoming. Coal Smoke and Mortality in an Early industrial Economy. *Economic Journal*.

Bell, M.L., & Davis, D.L. 2001. Reassessment of the lethal London fog of 1952: Novel indicators of acute and chronic consequences of acute exposure to air pollution. *Environmental Health Perspectives*, **109**(3), 389–394.

Brimblecombe, Peter. 1987. *The Big Smoke: A History of Air Pollution in London Since Medieval Times*. Methuen.

Chay, Kenneth, Dobkin, Carlos, & Greenstone, Michael. 2003. The Clean Air Act of 1970 and Adult Mortality. *Journal of Risk and Uncertainty*, **27**(3), 279–300.

Chay, Kenneth Y., & Greenstone, Michael. 2003. The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *The Quarterly Journal of Economics*, **118**(3), pp. 1121–1167.

Chen, Yuyu, Ebenstein, Avraham, Greenstone, Michael, & Li, Hongbin. 2013. Evidence on the impact of sustained exposure to air pollution on life expectancy from Chinas Huai River policy. *Proceedings of the National Academy of Sciences*, **110**(32), 12936–12941.

Clay, K, Lewis, J., & Severnini, E. 2015 (October). *Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic*. NBER Working Paper No. 21635.

Clay, Karen, Lewis, Joshua, & Severnini, Edson. 2016 (April). *Canary in a Coal Mine: Impact of Mid-20th Century Air Pollution Induced by Coal-Fired Power Generation on Infant Mortality and Property Values*. NBER Working Paper No. 22115.

Currie, J. 2013. Pollution and Infant Health. *Child Development Perspectives*, **7**(4), 237–242.

Currie, Janet, & Neidell, Matthew. 2005. Air Pollution and Infant Health: What Can We Learn from California's Recent Experience? *The Quarterly Journal of Economics*, **120**(3), pp. 1003–1030.

Currie, Janet, Neidell, Matthew, & Schmieder, Johannes F. 2009. Air pollution and infant health: Lessons from New Jersey. *Journal of Health Economics*, **28**(3), 688 – 703.

Cutler, D, & Miller, G. 2005. The Role of Public Health Improvements in Health Advances: The Twentieth-

Century United States. *Demography*, **42**(1), 1–22.

Ferrie, Joseph P., & Troesken, Werner. 2008. Water and Chicago's mortality transition, 1850-1925. *Explorations in Economic History*, **45**(1), 1 – 16.

Floud, R., Fogel, R.W., Harris, B., & Hong, S.C. 2011. *The Changing Body: Health, Nutrition and Human Development in the Western World since 1700*. Cambridge: Cambridge University Press.

Foster, Andrew, Gutierrez, Emilio, & Kumar, Naresh. 2009. Voluntary Compliance, Pollution Levels, and Infant Mortality in Mexico. *American Economic Review, Papers and Proceedings*, **99**(2), 191–97.

Galiani, S, Gertler, P, & Schargrodsky, E. 2005. Water for Life: The Impact of the Privatization of Water Services on Child Mortality. *Journal of Political Economy*, **113**(1), pp. 83–120.

Graff Zivin, Joshua, & Neidell, Matthew. 2013. Environment, Health, and Human Capital. *Journal of Economic Literature*, **51**(3), 689–730.

Greenburg, L, Jacobs, MB, Drolette, BM, Field, F, & Braverman, MM. 1962. Report of an air pollution incident in New York City, November 1953. *Public Health Reports*, **77**(1), 7.

Greenstone, Michael, & Hanna, Rema. 2014. Environmental Regulations, Air and Water Pollution, and Infant Mortality in India. *American Economic Review*, **104**(10), 3038–72.

Gultepe, Ismail. 2007. Fog and Boundary Layer Clouds: Introduction. *Pure and Applied Geophysics*, **164**, 1115–1116.

Hanlon, W Walker, & Sudol, Katherin. 2017 (February). *Pollution and Maternal Mortality: Evidence from the London Fog*. Mimeo.

Hanna, Rema, & Oliva, Paulina. 2015. The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City. *Journal of Public Economics*, **122**, 68–79.

Jayachandran, Seema. 2009. Air Quality and Early-Life Mortality: Evidence from Indonesias Wildfires. *Journal of Human Resources*, **44**(4), 916–954.

Jia, R., & Ku, H. 2017 (January). *Is China's Pollution the Culprit for the Choking of South Korea? Evidence from the Asian Dust*. Working Paper.

Knittel, C. R., Miller, D. L., & Sanders, N. J. 2016. Caution, drivers! Children present: Traffic, pollution, and infant health. *Review of Economics and Statistics*, **98**(2), 350–366.

Logan, W.P.D. 1953. Mortality in the London fog incident, 1952. *The Lancet*, **261**(6755), 336–338.

McKeown, Thomas. 1976. *The Modern Rise of Population*. New York: Academic Press.

Meetham, A.R., Bottom, D.W., Crayton, S., Henderson-Sellers, A., & Chambers, D. 1981. *Atmospheric Pollution: Its History, Origins and Prevention*. Fourth edition edn. Oxford: Pergamon Press.

Ministry of Health. 1954. *Mortality and Morbidity during the London fog of December 1952*. Tech. rept. Ministry of Health, London. Reports on Public Health and Medical Subjects No. 95.

Nemery, B, Hoet, PHM, & Nemmar, A. 2001. The Meuse Valley fog of 1930: An air pollution disaster. *Lancet*, **357**(9257), 704–708.

Pope, CA. 1989. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. *American Journal of Public Health*, **79**(5), 623–628.

Rangel, M.A., & Vogl, T. 2016 (December). *Agricultural Fires and Infant Health*. NBER Working Paper No. 22955.

Rückerl, R, Schneider, A, Breitner, S, Cyrys, J, & Peters, A. 2011. Health Effects of Particulate Air Pollution: A Review of Epidemiological Evidence. *Inhalation Toxicology*, **23**(10), 555–592.

Samet, Jonathan M., Dominici, Francesca, Curriero, Frank C., Coursac, Ivan, & Zeger, Scott L. 2000. Fine Particulate Air Pollution and Mortality in 20 U.S. Cities, 1987-1994. *New England Journal of Medicine*,

343(24), 1742–1749. PMID: 11114312.

Schlenker, Wolfram, & Walker, W Reed. 2016. Airports, Air Pollution and Contemporaneous Health. *Review of Economic Studies*, **83**(2).

Thorsheim, Peter. 2006. *Inventing Pollution.* Athens, Ohio: Ohio University Press.

Townsend, James G. 1950. Investigation of the Smog Incident in Donora, Pa., and Vicinity. *American Journal of Public Health*, **40**(2), 183–189.

Troesken, W, & Clay, K. 2011. Did Frederick Brodie Discover the World's First Environmental Kuznets Curve? Coal Smoke and the Rise and Fall of the London Fog. *In:* Libecap, G., & Steckel, R. H. (eds), *The Economics of Climate Change: Adaptations Past and Present.* University of Chicago Press.

Woods, Robert. 2000. *The Demography of Victorian England and Wales.* Cambridge, UK: Cambridge University Press.

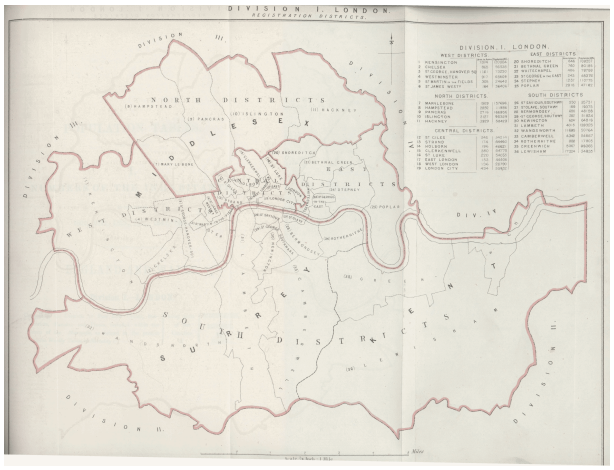
World Health Organization. 2017. *Global Tubercolosis Report.* Tech. rept.

A Appendix

A.1 Population and mortality in London

The area of London covered by this study is depicted in Figure 4. This area stretches from Islington and Hampstead in the North down to Camberwell and Wandsworth in the South. East to West it covers an area running from the western border of Kensington past the edge of Greenwich. This area, now called Inner London, is roughly covers the central districts of Greater London today.

Figure 4: Area of London covered by this study

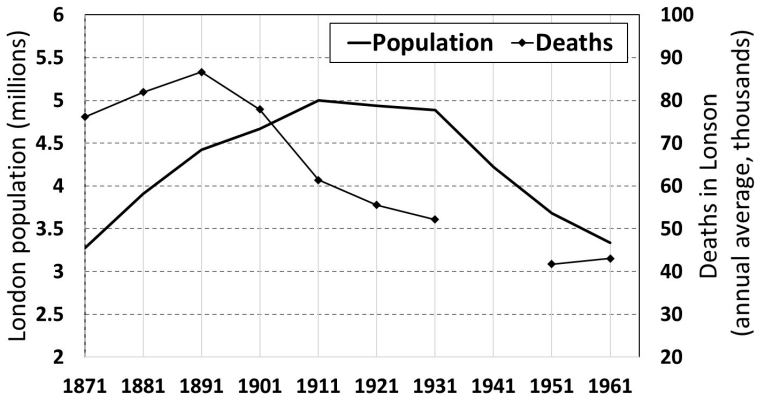


This map is from the 1851 Census of Population.

Figure 5 plots the population of the area of London covered by this study, which is now called Inner London. These population data are from the census, which took place every ten years, and begin with the census of 1871. The population of this area of London peaked in the early 20th century and then began declining as more people moved to the suburbs. The figure also plots annual deaths averaged across the four years starting with each census year. Four-year averages are used to reduce the effect of epidemics. Deaths peaked in the late 19th century and then began falling dramatically at the beginning of the 20th century. This reduction was due mainly to reductions in infectious disease mortality, such as the declines shown in Figure 12. From this figure we can tell that death rates had fallen dramatically. Figure 6 plots the raw death rate in London (without age adjustment) implied by the number of deaths and the population data. The death rate fell by more than half from the beginning of the study period into the inter-war period. If anything this understates the true decline

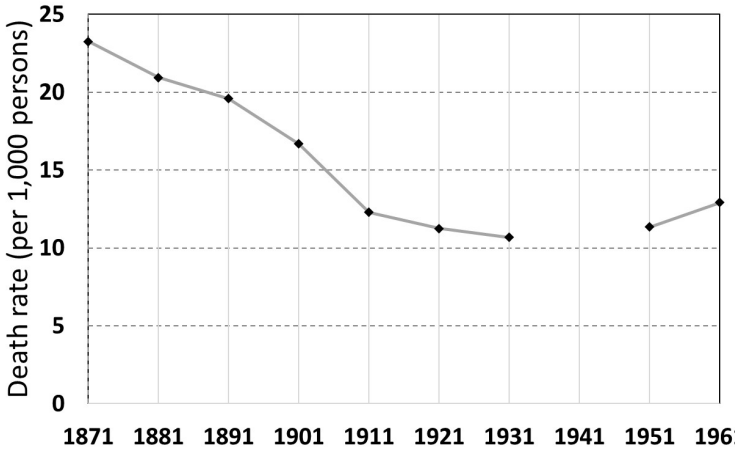
because of the effect of population aging after the demographic transition, which began in Britain in the 1870s.

Figure 5: Population and average annual deaths in London across the study period



Population data are from the census. Deaths data are summed from the Registrar General's weekly reports used in the main analysis. Deaths are average annual deaths across the four years following each census year.

Figure 6: Death rates in London



Population data are from the census. Deaths data are summed from the Registrar General's weekly reports used in the main analysis. Deaths are average annual deaths across the four years following each census year.

A.2 Data appendix

A.2.1 Summary statistics

Table 8 presents summary statistics for the data used in the main analysis, covering 1866-1965 (except 1915-18 and 1940-48).

Table 8: Summary statistics for weekly observations

Variable	Mean	Std. Dev.	Min.	Max.	N
Total deaths	1251.139	407.433	524	3761	4479
Deaths age 0-1	217	174	11	1107	3905
Deaths age 1-5	127	113	0	593	3905
"Young" deaths	67	36	3	216	4192
"Adult" deaths	142	70	20	610	4192
"Middle age" deaths	261	68	130	828	4192
"Elderly" deaths	413	151	181	1666	4192
Infant MR (per 100,000 births)	9693	6299	1044	47380	3905
Fog days	0.207	0.636	0	6	4479
Fog week indicator	0.129	0.335	0	1	4479
Pred. fog week	0.063	0.243	0	1	4479
Avg. max daily pollution	0.269	0.201	0.038	1.437	574
Avg. mean daily pollution	0.122	0.095	0.006	0.732	737
Temperature	50.022	9.438	22.4	73.900	4479
Pressure	29.79	0.241	28.824	30.76	4479
Precipitation	0.456	0.489	0	4.71	4479
Humidity	78.782	8.012	52	99	4479
Births	1963.887	581.87	609	3308	4479
Stillbirths	31.948	12.399	10	71	1499

A.2.2 Weather report examples

Figures 7 and 8 present examples of weather reports from 1880 and 1952 for weeks in which heavy fog occurred.

Figure 7: Example weather report from February 7, 1880

SKY AND ATMOSPHERE.

Hoar frost and dense fog in morning; fine in afternoon and evening; very foggy at night.

Hoar frost and dense fog in morning; fine with light cloud till evening; overcast at night and a little thin rain fell then.

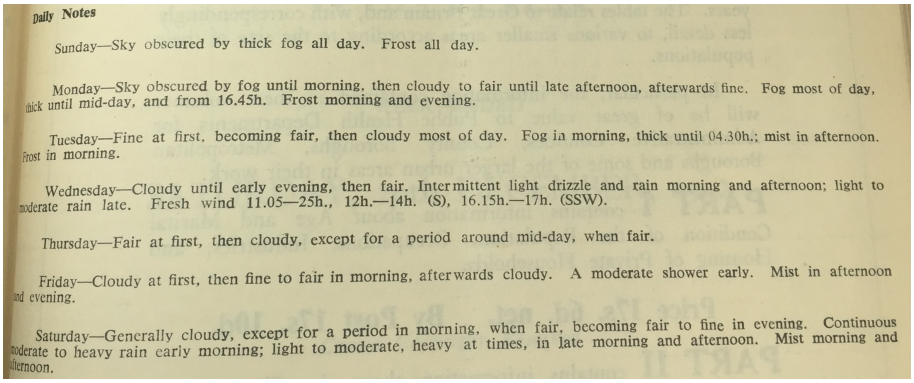
Overcast.

Hoar frost in morning; foggy during the greater part of the day; sky cloudless throughout. (In the lower parts of the Park and in the town of Greenwich the fog was extremely dense at times.)
Misty in morning; very fine and generally cloudless till afternoon; overcast afterwards. A little thin rain fell at night.

Overcast. Frequent thin rain.

Overcast. Frequent rain.

Figure 8: Example weather report from the 50th week of 1952



Daily Notes

Sunday—Sky obscured by thick fog all day. Frost all day.

Monday—Sky obscured by fog until morning, then cloudy to fair until late afternoon, afterwards fine. Fog most of day, thick until mid-day, and from 16.45h. Frost morning and evening.

Tuesday—Fine at first, becoming fair, then cloudy most of day. Fog in morning, thick until 04.30h.; mist in afternoon. Frost in morning.

Wednesday—Cloudy until early evening, then fair. Intermittent light drizzle and rain morning and afternoon; light to moderate rain late. Fresh wind 11.05—25h., 12h.—14h. (S), 16.15h.—17h. (SSW).

Thursday—Fair at first, then cloudy, except for a period around mid-day, when fair.

Friday—Cloudy at first, then fine to fair in morning, afterwards cloudy. A moderate shower early. Mist in afternoon and evening.

Saturday—Generally cloudy, except for a period in morning, when fair, becoming fair to fine in evening. Continuous moderate to heavy rain early morning; light to moderate, heavy at times, in late morning and afternoon. Mist morning and afternoon.

A.2.3 Graphs of the fog event data

Figure 9: Fog weeks and fog days, 1865-1965

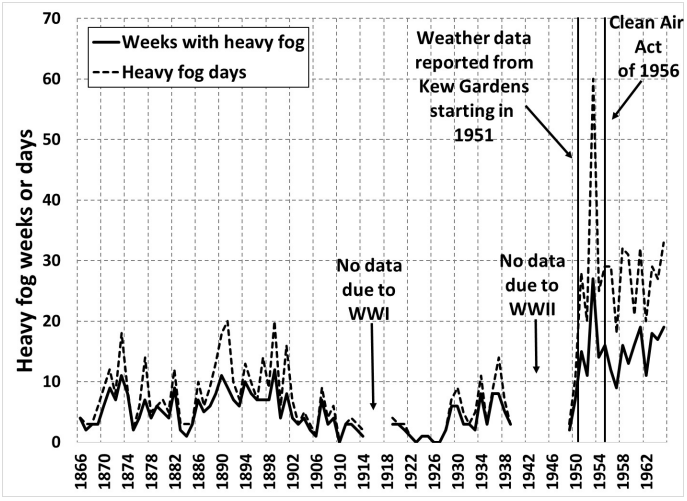
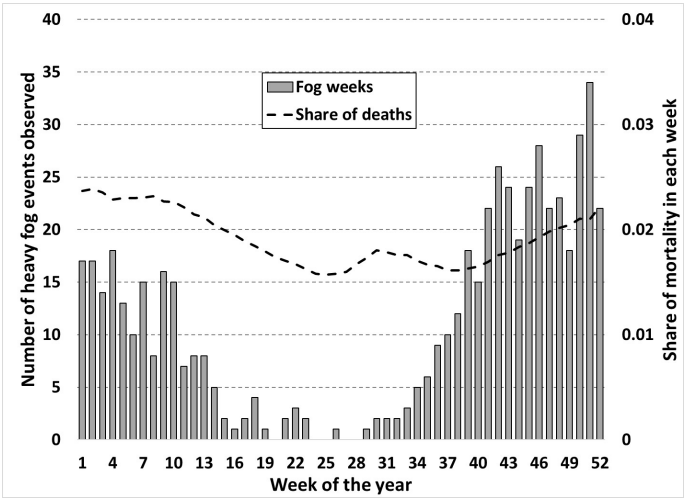


Figure 10: Fog events and total mortality share by week of the year



A.2.4 Cause-of-death data

The cause of death data are reported in a wide variety of categories that change over time. Table 11 provides an overview of how these diseases are grouped to obtain the categories used in the analysis.

Figure 11: Components of the cause of death categories

Accidents and Violence
Accidents (fire, drowning, traffic, etc.)
Suffocation in bed
Homocide
Execution

Alcoholism
Alcoholism and delerium tremens

Bronchitis

Cancer
Cancer, sarcoma
Pancreatic disease
Tumors, undefined

Cardiovascular
Acute endocarditis
Aneurism
Embolism, thrombosis
Heart disease
Pericaritis
Pulmonary apoplexy
Other dis. Circulatory system

Death in childbirth
Peurperal sepsis
Other peurperal causes

Digestive
Cholera
Diarrhea, entiritis
Enteric or typhoid fever
Gastritis
Ileus
Intussusception
Other fevers (remittant, simple cont.)
Stricture of intestines
Ulceration of intestines
Other digestive/stomach diseases

Diphtheria

Measles

Neurologial
Acute poliomyelitis
Apoplexy
Hemiplegia
Cephalitis
Cerebral haemorrhage
Cerebor-spinal fever
Epilepsy
Hydrocephalus
Meningitis
Other dis. of the nervous system

Old age

Other infectious diseases
Ague
Anthrax splenic fever
Carbuncle
Cowpox/effects of vaccination
Erysipelas
Glanders
Hepatitis
Plague
Pyaemia and septicaemia
Pyrexia
Rabies (hydrophobia)
Rheumatic fever
Relapsing fever
Tetanus

Pneumonia

Premature birth

Respiratory diseases
Asthma
Atelectasis (collapsed lung)
Croup
Emphasema
Influenza
Laryngitis
Pleurisy
Other lung/respiratory diseases

Scarlet fever

Smallpox

Suicide

Tuberculosis
Phthisis
Pulmonary tuberculosis
Scrofula
Tabes Mesenterica
Tubercular meningitis
Tuberculous peritonitis

Typhus

Venereal diseases
Syphilis
Other venereal diseases

Whooping cough

Table 9 describes the share of mortality in each age category accounted for by each cause

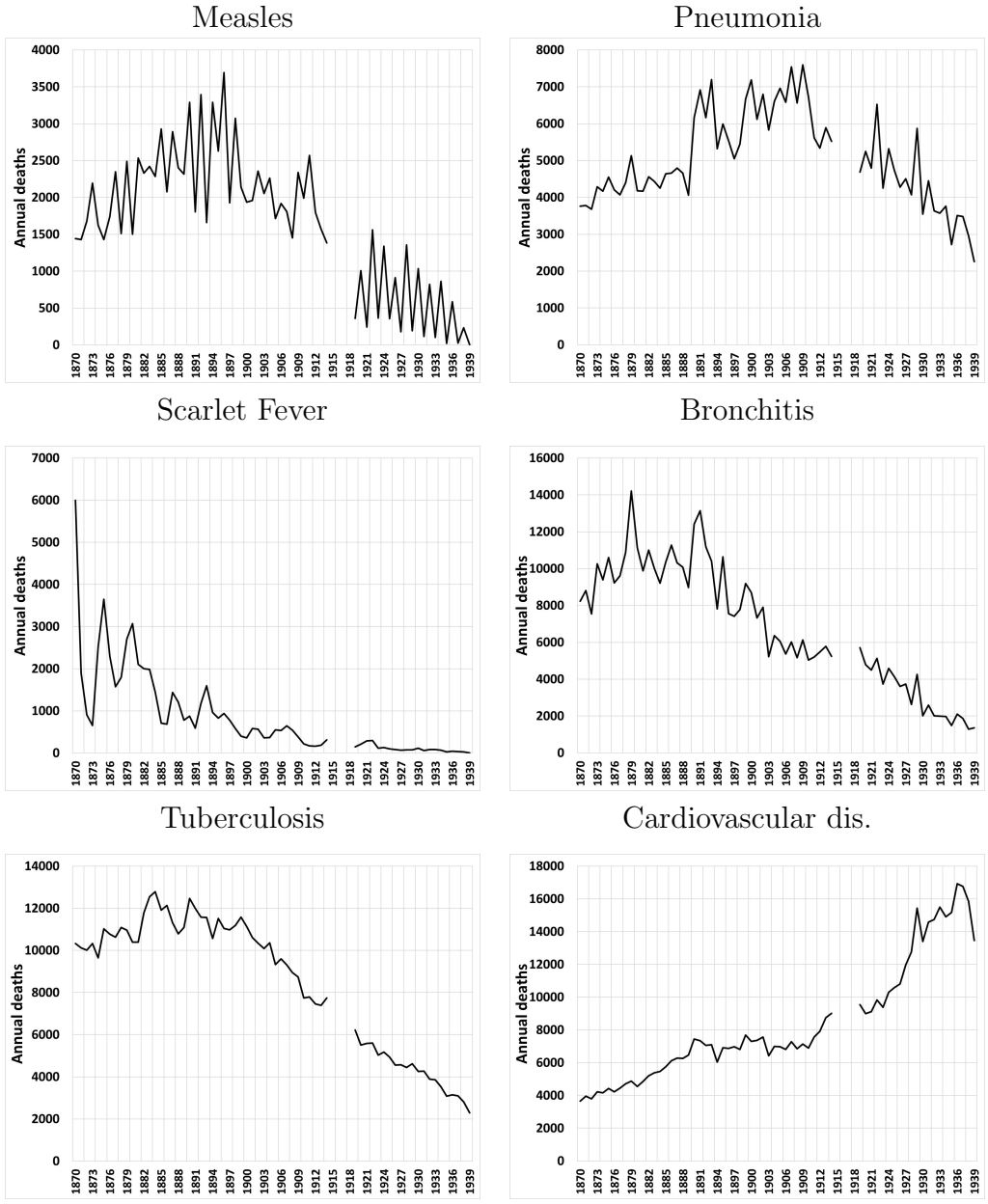
of death.

Table 9: Share of age group mortality accounted for by each COD

	Age group					
Cause of death category	0 - 1	1 - 5	Young (5 to 20/25)	Adult (20/25 to 40/45)	Middle (40/45 to 60/65)	Elderly (over 60 or 65)
BRONCHITIS	9.77%	9.74%	1.85%	2.55%	9.61%	19.94%
CARDIOVASCULAR	0.30%	0.43%	8.53%	8.95%	13.22%	14.18%
PNEUMONIA	8.41%	14.30%	5.36%	6.14%	6.41%	4.46%
RESPIRATORY	1.99%	4.02%	3.08%	2.66%	3.82%	4.02%
DIGESTIVE	17.25%	6.40%	7.81%	5.78%	3.92%	3.85%
DIPHTHERIA	0.42%	6.49%	6.88%	0.23%	0.07%	0.02%
MEASLES	2.53%	14.74%	2.49%	0.06%	0.01%	0.00%
SCARLET_FEVER	0.26%	4.97%	7.52%	0.35%	0.03%	0.00%
SMALLPOX	0.17%	0.38%	2.77%	1.75%	0.36%	0.05%
TUBERCULOSIS	5.79%	10.21%	22.52%	38.58%	19.47%	2.94%
TYPHUS	0.00%	0.02%	0.33%	0.22%	0.12%	0.03%
WHOOPING_COUGH	4.48%	10.62%	1.52%	0.01%	0.00%	0.00%
OTHER_INFECTIOUS	0.67%	0.44%	1.50%	1.06%	1.03%	0.75%
CANCER	0.03%	0.18%	0.65%	2.90%	10.25%	7.55%
NEUROLOGICAL	4.71%	5.91%	7.46%	5.41%	9.35%	12.37%
OLD_AGE	0.00%	0.00%	0.00%	0.00%	0.04%	13.43%
PREMATURE_BIRTH	12.84%	0.01%	0.00%	0.00%	0.00%	0.00%
VENEREAL_DISEASES	1.59%	0.21%	0.08%	0.43%	0.31%	0.10%
OTHER_MISC_COD	25.32%	8.07%	12.12%	11.81%	15.74%	13.67%
ACCIDENT_VIOLENCE	3.33%	2.84%	6.80%	4.17%	3.21%	2.01%
ALCOHOLISM	0.00%	0.00%	0.00%	1.25%	1.43%	0.24%
HOMICIDE	0.15%	0.03%	0.09%	0.16%	0.06%	0.01%
SUICIDE	0.00%	0.00%	0.31%	1.48%	1.21%	0.37%

A.2.5 Cause of death graphs

Figure 12: Deaths in London in select infectious disease categories



A.3 Using weather data to model fog formation

This appendix describes the use of weather data to model fog formation. The formation of fog is a complex event that depends on a number of climatic conditions, including temperature, humidity, wind speed and turbulence, ground temperature, precipitation, radiation, etc.[56] In an attempt to predict the formation of heavy fog I have gathered weather data capturing some of the most important factors. The available data present weekly average levels of mean daily temperature, humidity, and barometric pressure and total weekly precipitation. Table 8 provides summary statistics for the available climate variables using data from all of the years covered by this study.

Figure 13 presents time-series graphs of each of the variables. These show some important patterns; there is clear evidence that temperature rose and humidity fell across the study period. The fall in humidity is in part a natural response to the rise in temperature. This is because humidity in this study is relative humidity, which is defined as the ratio of water in the air to the total holding capacity of water in the air. Warmer air holds more water, so rising temperature mechanically increases the denominator and thus mechanically leads to lower relative humidity. However, note that while the humidity measure partially reflects temperature, it also contains additional information. The correlation between the temperature and humidity variables is -0.51.

There is also some evidence that barometric pressure may have increased, while precipitation appears to have remained fairly steady. Note that we must be somewhat careful in interpreting these patterns because, like the fog reports, the weather data for the years after WWII come from Kew Gardens rather than the Greenwich Observatory.

Figure 14 presents histograms of the climate variables comparing weeks in which fog events did or did not occur. Consistent with the science behind fog formation, these show that fog was more likely to form when temperatures were lower and when humidity was higher. Fog formation was also associated with high pressure, which often signals calm conditions and with lower levels of precipitation, which signals less cloud cover.

[56] See, e.g., Ahrens (2007).

Figure 13: Time-series graphs of climate variables

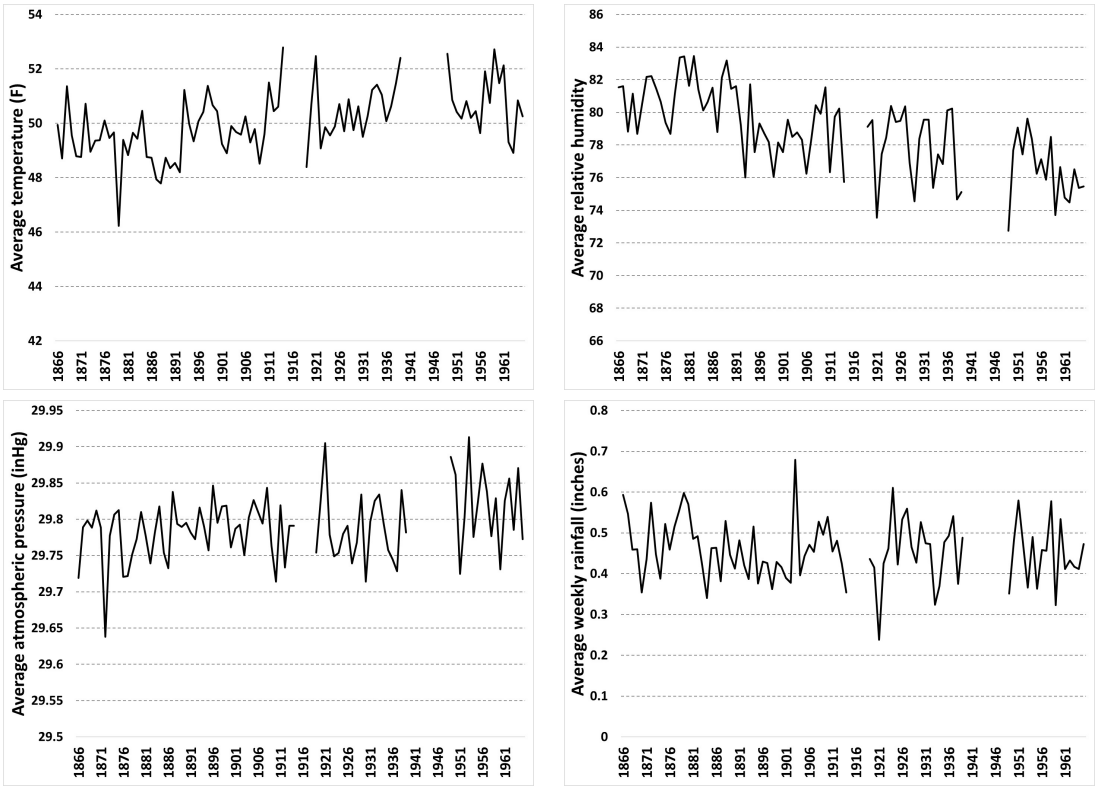
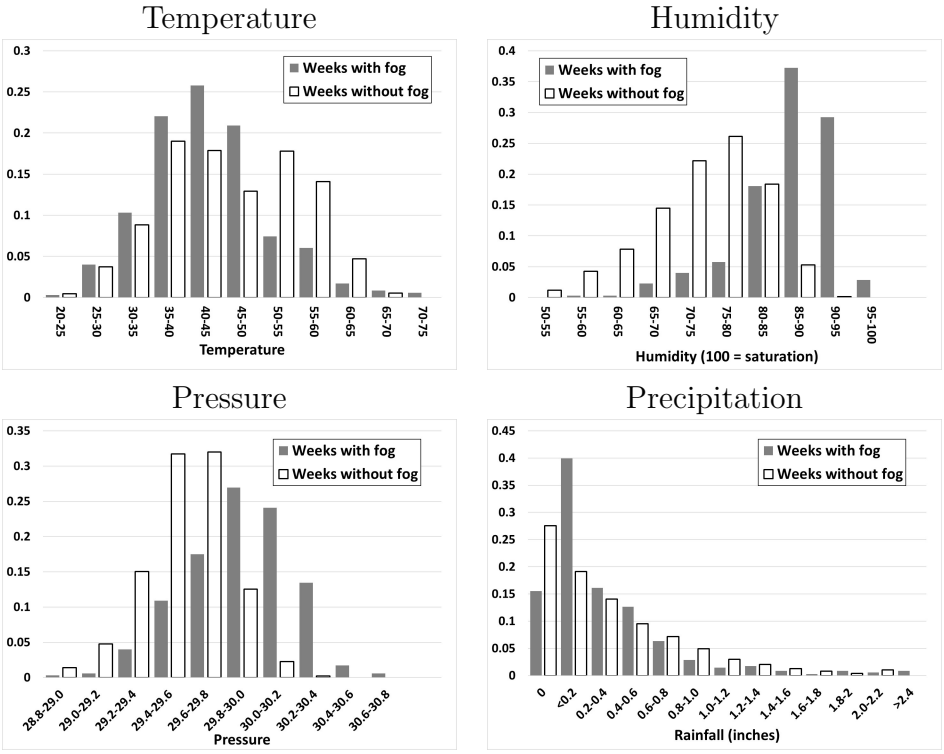


Figure 14: Histograms of climate variables in weeks with and without fog



Histogram plots of weather variables for weeks with our without heavy fog events reported. These graphs only include data up to 1951 to avoid any changes due to the shift of weather reporting from Greenwich to Kew Gardens.

As a simplified way of modeling these interactions I think of the formation of heavy fog as an outcome that occurs when a series of necessary conditions are satisfied. In particular, thick fog is modeled as forming under conditions in which temperature is sufficiently low, humidity in the air is sufficiently high, atmospheric pressure is sufficiently high and precipitation is sufficiently low. Lower temperature means that air can hold less water, making condensation more likely. Conditional on temperature, higher humidity indicates that there is more water in the air which means that condensation is more likely to occur. The type of clear and relatively calm conditions under which fog is mostly likely to form are typically associated with atmospheric high-pressure systems. Precipitation is included here because the presence of rain indicates cloud cover which blocks sunlight from reaching the ground. Thus, the inputs into my simple fog model are a series of indicator variables:

$$TempCUT_t = 1[Temperature < \tilde{T}]$$

$$PresCUT_t = 1[Pressure > \tilde{P}]$$

$$HumCUT_t = 1[Humidity > \tilde{H}]$$

$$RainCUT_t = 1[Precipitation < \tilde{R}]$$

Interaction these indicator variables yields:

$$PredFOG_t(T, P, H, R) = TempCut_t * PresCUT_t * HumCUT_t * RainCUT_t$$

This interaction term provides a way of predicting whether or not conditions within a week favored the formation of thick fog. However, it is still necessary to determine appropriate cutoff points for each variable. The cutoff points used in the main text – $\tilde{T} = 55$, $\tilde{P} = 29.9$, $\tilde{H} = 85$ and $\tilde{R} = 0.5$ – were chosen because the resulting interaction does a reasonable job of predicting fog events. This is shown in Table 10, which compares several leading and lagged values of the predicted fog variable to the number of heavy fog days reported in a week.

Table 11 describe the number of correct predictions as well as counts of type I and type II errors obtained using this prediction. This shows that, while the $PredFOG$ variable predicts heavy fog events, there are still many predicted fog weeks when no heavy fog event was reported (type II errors) and many heavy fog events when no fog event was predicted. It is worth noting that if we compare the predicted fog events to any fog (not just heavy fog events) there are far fewer type II errors; of the 282 predicted fog weeks only 32 had no fog reported at all.

Table 10: Regressions of fog days on the *PredFOG* variable with leads and lags

	DV: Number of heavy fog days in a week				
	All years	1866-1889	1890-1914	1919-1939	1951-1965
	(1)	(2)	(3)	(4)	(5)
PredFOG t+4	-0.0815	-0.0910	0.0435	-0.0902	-0.276
	(0.0540)	(0.0751)	(0.0898)	(0.0840)	(0.244)
PredFOG t+3	0.0145	-0.129*	0.103	0.0445	-0.00824
	(0.0573)	(0.0777)	(0.0999)	(0.0967)	(0.257)
PredFOG t+2	0.0947*	0.0689	0.00381	0.0317	0.211
	(0.0563)	(0.0792)	(0.0981)	(0.102)	(0.236)
PredFOG t+1	-0.0342	-0.126	0.0645	-0.0116	0.0315
	(0.0590)	(0.0776)	(0.123)	(0.0993)	(0.243)
PredFOG t	0.389***	0.221**	0.371***	0.347***	0.814***
	(0.0777)	(0.110)	(0.140)	(0.123)	(0.300)
PredFOG t-1	0.0362	0.0424	0.0931	-0.0399	-0.210
	(0.0602)	(0.0935)	(0.105)	(0.0906)	(0.253)
PredFOG t-2	-0.0264	-0.0165	0.0872	-0.150*	-0.159
	(0.0549)	(0.0759)	(0.0926)	(0.0872)	(0.252)
PredFOG t-3	0.0236	0.0112	0.100	0.106	-0.203
	(0.0548)	(0.0713)	(0.0943)	(0.0984)	(0.240)
PredFOG t-4	-0.0165	0.0751	-0.0603	-0.0373	-0.219
	(0.0527)	(0.0782)	(0.0891)	(0.0942)	(0.220)
PredFOG t-5	0.0848	0.0306	0.165*	-0.0846	0.171
	(0.0538)	(0.0681)	(0.0868)	(0.0846)	(0.255)
Observations	4,479	1,246	1,253	1,066	777

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors, in parenthesis, are used because the data show no evidence of serial correlation (Durbin-Watson statistics are around 1.9). Regressions include controls for pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared, and four leads and five lags of temperature, temperatures squared, and the temperature, humidity, pressure and precipitation cutoffs. Temperature cutoff is 55 degrees f. Pressure cutoff is 29.9. Relative humidity cutoff is 85. Precipitation cutoff is 0.5 inches per week. Regressions also include a full set of year and week-of-the-year by decade effects.

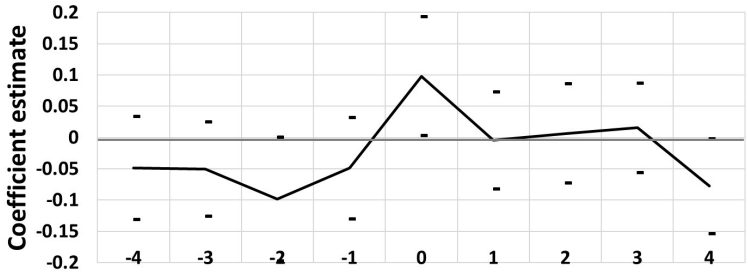
Figure 15 shows that the *PredFOG* variable also does a good job of predicting increases in actual pollution levels, using the direct pollution measures available from 1951-1961.

Table 11: Table showing accuracy of fog week predictions

Period	Heavy fog week	No. pred. fog events	Correct prediction	Type I errors	Type II errors
All years	577	282	144	433	138
1866-1889	125	96	39	86	57
1890-1914	132	85	40	92	45
1919-1939	82	55	29	53	26
1951-1965	229	42	33	196	9

Note that the four periods do not add up to the total for all years because when I focus on the period-by-period results I exclude 1949-1950 from the data since those years differ from the other post-WWII years because the data are based on observations from Greenwich rather than Kew Gardens.

Figure 15: Results from regressions of pollution levels on predicted fog weeks

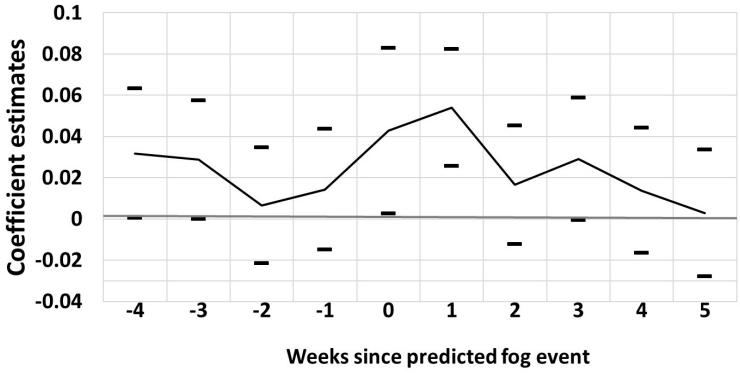


Pollution measures are averages of the maximum daily values reported in the week. Regression includes controls for temperature and temperature squared as well as a full set of year and week-of-the-year effects. Confidence intervals are based on robust standard errors. The Durbin-Watson statistic for these regressions is 1.406, suggesting that serial correlation is not likely to be an important concern.

Figure 16 shows the relationship between total mortality and fog events using predicted fog events as an instrument for the number of fog days. This result looks fairly similar to those obtained when using fog events directly, i.e., mortality tends to increase in the week or two just after fog events occur with some evidence of elevated mortality levels for up to three weeks.

Figure 16: Estimated effect of fog events on total mortality, 1866-1965

Using predicted fog event weeks as an instrument for fog days



Coefficient estimates and confidence intervals for a regression of log total mortality on predicted fog events. Confidence intervals are based on Newey-West standard errors allowing correlation across observations within 6 weeks of each other. Both regressions include controls for temperature, temperature squared, four leads and five lags of temperature and temperature squared, log births, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared, a full set of year effects and a full set of week-of-the-year by decade effect. The regression also includes leads and lags of the temperature, humidity, pressure and precipitation indicator variables that are interacted to generate the predicted fog variable. Data cover 1866-1965. N=4448.

Table 12 presents results using four alternative sets of cutoff values to predict fog event weeks. Columns 1-2 use values that are less restrictive than the set used for the results in the main text while Columns 3-4 use more restrictive values. All of these alternatives can predict the occurrence of heavy fog weeks.

Table 13 presents results from regressions of log mortality on predicted fog events using four week windows as in Eq. 2. All of the alternatives predicted fog event variables are associated with increased mortality. However, note that as I move towards more restrictive cutoff values the effect of a predicted fog event tends to increase. This tells us that predictions based on more restrictive conditions are picking up more severe fog events. As described in the main text, this feature poses a problem if we want to use these predictions to instrument for fog events. In IV regressions, as I use more restrictive fog event prediction variables, the estimated coefficient on the relationship between fog events and mortality tends to increase because the variation picked up by the instrument is increasingly focused on more severe fog events. This makes is impossible to compare the coefficients on the fog variable obtained from IV regressions to those obtained from the OLS regressions.

At the bottom of the table I describe the number of deaths associated with predicted fog events given the estimated coefficients and the number of predicted events (from table above) for each set of parameters. Despite the fact that the estimated effect of a predicted fog event increases as the criteria become more restrictive, the overall number of deaths implied falls because there are fewer events.

It is also interesting to consider the pattern of changes in the number of predicted fog events over time. In Figure 17 I plot of the share of weeks with heavy fog events reported in each decade up through the 1930s. Predicted fog events show a peak in the 1890s followed by a sharp drop in the early 20th century. Thus, after 1900 underlying weather conditions became much less conducive to fog formation. The driving force behind this appears to be increases in temperature and commensurate reductions in relative humidity after 1900 (see Appendix Figure 13). Thus, changing climate conditions meant that London in the early 20th century was naturally less foggy than it had been in the second half of the 19th century.

Table 12: Exploring fog event predictions using alternative cutoff values

	DV: Number of heavy fog days in a week			
	Less restrictive		More restrictive	
	(1)	(2)	(3)	(4)
Temperature cutoff:	60	57.5	50	45
Pressure cutoff:	29.8	29.8	29.9	30
Humidity cutoff:	80	82.5	87.5	90
Precipitation cutoff:	0.7	.6	0.4	0.3
PredFOG t+4	0.00640	0.00650	-0.123*	0.0405
	(0.0364)	(0.0414)	(0.0676)	(0.0947)
PredFOG t+3	0.0293	0.0234	0.00789	(0.110)
	(0.0376)	(0.0430)	(0.0653)	0.105
PredFOG t+2	-0.0553	-0.0577	0.00497	(0.115)
	(0.0357)	(0.0403)	(0.0713)	0.0596
PredFOG t+1	-0.0174	-0.0400	0.0564	(0.119)
	(0.0373)	(0.0407)	(0.0708)	0.0289
PredFOG t	0.197***	0.243***	0.333***	0.489***
	(0.0430)	(0.0504)	(0.0995)	(0.180)
PredFOG t-1	-0.0349	-0.0123	-0.0249	-0.00223
	(0.0381)	(0.0433)	(0.0644)	(0.122)
PredFOG t-2	-0.0445	-0.0364	-0.0979	-0.168*
	(0.0382)	(0.0413)	(0.0640)	(0.0996)
PredFOG t-3	-0.00716	-0.00229	0.0187	0.0671
	(0.0368)	(0.0402)	(0.0650)	(0.123)
PredFOG t-4	-0.0229	-0.0192	-0.0531	0.0751
	(0.0378)	(0.0415)	(0.0617)	(0.0887)
PredFOG t-5	-0.0120	0.0334	0.0331	0.0267
	(0.0373)	(0.0399)	(0.0578)	(0.0933)
Number of weeks with predicted fog events				
	741	572	168	50
Predicted fog weeks with heavy fog events (correct)				
	248	220	92	34
Type I errors				
	329	357	485	543
Type II errors				
	493	352	76	16
Observations	4,479	4,479	4,479	4,479

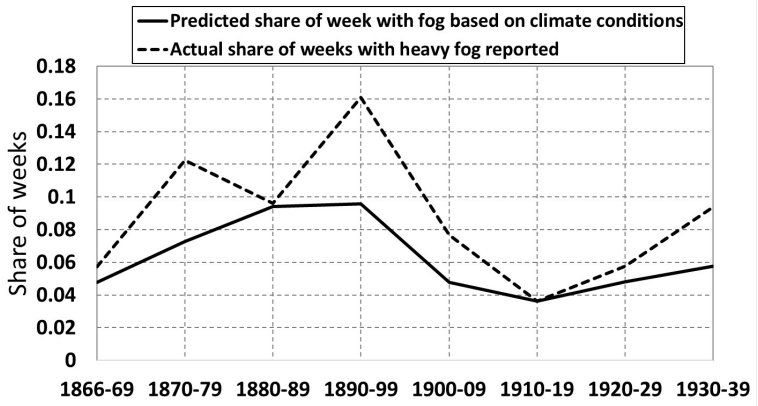
*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors, in parenthesis, are used because the data show no evidence of serial correlation (Durbin-Watson statistics are around 1.9). Regressions include controls for pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared, and four leads and five lags of temperature, temperatures squared, and the temperature, humidity, pressure and precipitation cutoffs used to construct the *PredFOG* variable in each regression. Regressions also include a full set of year and week-of-the-year by decade effects.

Table 13: Alternative fog event predictors and mortality using four week windows

	DV: Number of heavy fog days in a week			
	Less restrictive		More restrictive	
	(1)	(2)	(3)	(4)
Temp. cutoff:	60	57.5	50	45
Pres. cutoff:	29.8	29.8	29.9	30
Humid. cutoff:	80	82.5	87.5	90
Precip. cutoff:	0.7	.6	0.4	0.3
PredFOG	0.0121**	0.0106*	0.0215***	0.0323**
	(0.00525)	(0.00582)	(0.00773)	(0.0130)
	Implied deaths due to predicted fog events			
	45,139	30,501	18,270	8,213
Observations	4,479	4,479	4,479	4,479

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors, in parenthesis, are used because the data show no evidence of serial correlation (Durbin-Watson statistics are around 1.9). Regressions include controls for pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared, and four leads and five lags of temperature, temperatures squared, and the temperature, humidity, pressure and precipitation cutoffs used to construct the *PredFOG* variable in each regression. Regressions also include a full set of year and week-of-the-year by decade effects.

Figure 17: Actual and predicted share of weeks with fog by decade



This graph shows the share of weeks with heavy fog events reported in each decade and the share of weeks with predicted fog events in each decade. Predicted fog events use the following cutoffs: temperature < 55, pressure >29.9, humidity > 85, precipitation < 0.5.

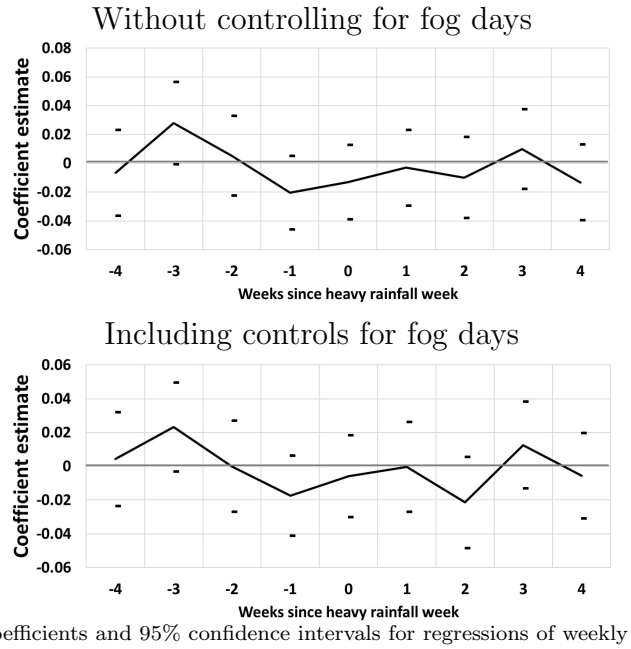
A.4 Impact of heavy rain on infectious disease mortality

One potential identification concern in this study is that fog may have increased infectious disease mortality by causing people to crowd together inside, increasing the spread of disease. However, whether this would increase infectious disease spread is not obvious, since staying indoors may also reduce contact with people outside of the family, which may reduce the transmission of infectious disease. Additionally, we may worry that there is something about the moisture content of the air on foggy days that impacts health. This appendix presents results that can shed light on these issues. Specifically, I look at the relationship between weeks with heavy rain and infectious disease mortality. We would expect people, and particularly non-workers such as children, to stay indoors more during rainy weeks. Thus, looking at rainy weeks offers a way to infer the impact of staying indoors on infectious disease mortality. At the same time, rainy weeks, like foggy weeks, are periods when there is a lot of moisture in the air, so looking at the relationship between rainy weeks and mortality also offers a way to assess the impact of this factor on infectious disease mortality.

I focus on the impact of weeks with heavy rain, which I define as those with more than 0.5 inches of rain, though it is possible to use other values. London being a fairly rainy place, this identifies about one-third of weeks as having substantial amounts of rain. I then look at how mortality in infectious disease causes-of-death are influenced by heavy rain, including leads and lags of the heavy rain variable. I include all of my standard controls except the quadratic controls for rainfall. These are: temperature, temperature squared, four leads and five lags of these variables, humidity, humidity squared, pressure, pressure squared, and log births. I also include as controls leads and lags of the number of fog days in a week, to ensure that the impact of heavy fog weeks is not operating through a relationship between rainfall and fog.

Before moving to the main part of this analysis, it is useful to address the possibility that rainfall might affect pollution levels. I examine this using the period for which I have direct pollution measures. These results, in Figure 18, show that weeks with heavy rain had no strong relationship to pollution levels, so it does not appear that the impact of heavy rainfall on mortality through pollution is a serious concern in this exercise.

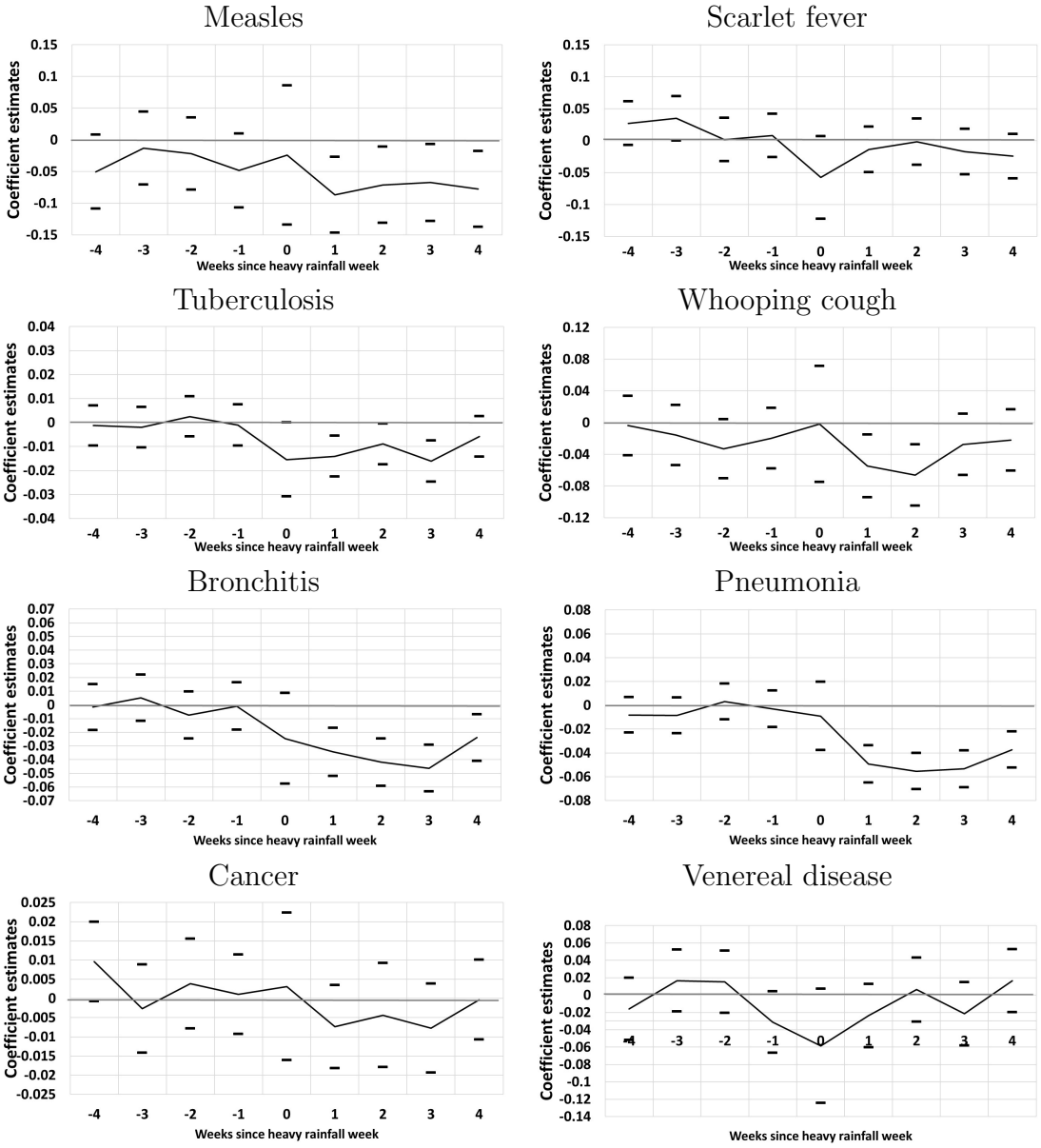
Figure 18: Relationship between pollution levels and heavy rain



The graphs present coefficients and 95% confidence intervals for regressions of weekly average of maximum daily pollution levels on leads and lags of the heavy rain week indicator variable. Robust standard errors. Data cover 1951-1962. I include controls four leads and four lags of temperature, temperature squared, humidity, humidity squared, pressure, and pressure squared as well as a full set of year and week-of-the-year fixed effects. In the bottom panel I also controls for the number of heavy fog days in a week and four leads and lags of that variable.

The results from comparing infectious disease mortality to heavy rainfall are in Figure 19. I focus on the most important infectious diseases, but similar results are obtained for most of the smaller categories as well. The striking thing about these results is that mortality due to infectious diseases appears to fall for several weeks following a week with heavy rain. A likely cause of this is that people stay indoors, reducing infectious disease transmission. These results are inconsistent with the idea that infectious disease mortality increases when people stay indoors. It is also worth noting that there is little evidence of changes in mortality levels in the weeks leading up to weeks with heavy rain, suggesting that these results are well-identified. For comparison purposes, the bottom row of figures present results for cancer and venereal diseases, two non-infectious causes-of-death. Unlike the infectious disease categories, these do not show any clear evidence of a reduction following weeks with heavy rain.

Figure 19: Impact of heavy rainfall on infectious disease mortality



The graphs present coefficients and 95% confidence intervals for regressions of log mortality on leads and lags of the heavy rain week indicator variable. Robust standard errors. Data cover 1870-1939 except 1915-18. I include controls four leads and five lags of heavy fog days, temperature and temperature squared, as well as humidity, humidity squared, pressure, and pressure squared as and full set of year and week-of-the-year-by-decade fixed effects.

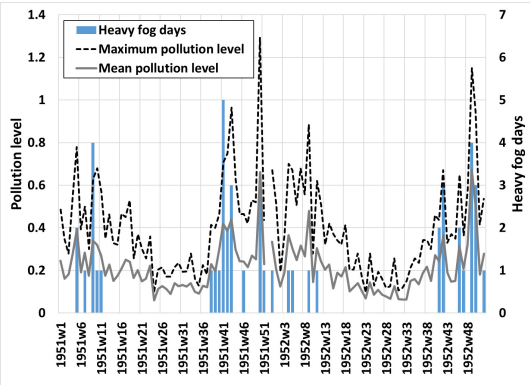
A.5 Preliminary analysis appendix

A.5.1 Graphs comparing pollution and fog events

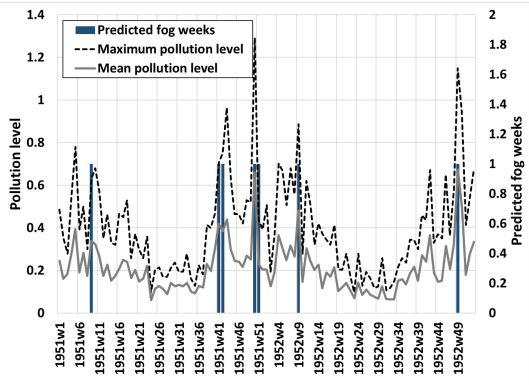
Figure 20 plots pollution levels against fog events for 1951-52, the first two years for which pollution levels are reported. This graph shows a clear correspondence between fog events and pollution levels, with the highest pollution levels obtained during major fog events. It is worth pointing out that the spike at the far right of the chart is the Great London Fog of 1952, but similar pollution levels were measured during earlier fog events. Figure **??** plots pollution levels against predicted fog weeks. We can see that the predicted fog weeks correspond to the highest pollution levels as well as the more severe fog events plotted in Figure 20. The predicted fog variable is clearly more restrictive than the actual data on heavy fog events and identifies the most severe events.

Figure 20: Fog events and reported pollution levels in 1951-52

Compared to fog events



Compared to predicted fog events



Pollution measures are calibrated values from the Owens Smoke Filter. All data are from the Registrar General's Weekly Reports. Predicted fog weeks are based on the interaction of indicator variables for temperature below 55 degrees, humidity above 80, pressure above 29.9 and precipitation below 0.4.

A.6 Appendix to analysis of total mortality

A.6.1 Analysis of serial correlation in total mortality data

Table 21 presents results describing autocorrelation patterns up to ten lags for the residuals from a regression based on Eq. 1. These results suggest that the partial autocorrelation values essentially disappear after two lags, while even the autocorrelation values disappear after five weeks of lags. This suggests that allowing serial correlation up to six lags is a reasonable approach to dealing with the serial correlation found in the data.

Figure 21: Autocorrelation structure of total mortality regression residuals

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.7058	0.7059	2232.7	0.0000		
2	0.5641	0.1314	3659.1	0.0000		
3	0.3929	-0.0933	4351.2	0.0000		
4	0.2440	-0.0862	4618.2	0.0000		
5	0.1400	-0.0162	4706.2	0.0000		
6	0.0640	-0.0068	4724.6	0.0000		
7	0.0208	0.0065	4726.5	0.0000		
8	0.0090	0.0278	4726.9	0.0000		
9	-0.0050	-0.0136	4727	0.0000		
10	-0.0018	0.0061	4727	0.0000		

This table describes the autocorrelation structure of the residuals from a regression based on Eq. 1. AC stands for autocorrelation, i.e., a regression of the residual on separate lags of the residual. PAC indicates the partial autocorrelation, i.e., results from a regression of the residual on all of the lagged values of the residual together. Controls included in the regression are temperature, temperature squared, four leads and five lags of these variables, pressure, pressure squared, humidity, humidity squared, precipitation and precipitation squared.

A.6.2 Results using alternative treatment windows

Table 14 presents regression results showing the effect of fog days over windows starting with the week in which the event occurred and stretching from two to five weeks later. Note that the result in Column 3, which uses a four-week window, corresponds to the specification used in the main text. At the bottom of the table I present the number of deaths implied by each approach. These are calculated accounting for both the estimated coefficient and the number of weeks across which the coefficient is applied.

Table 14: Estimates using effect windows of different lengths

	DV: Log total mortality			
	(1)	(2)	(3)	(4)
Fog days (2 week window)	0.0127***			
	(0.00298)			
Fog days (3 week window)		0.0105***		
		(0.00252)		
Fog days (4-week window)			0.00845***	
			(0.00234)	
Fog days (5 week window)				0.00712***
				(0.00230)
Implied deaths	29,804	36,920	39,575	41,655
Observations	4,479	4,479	4,479	4,479

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. Regressions run on the full set of available observations from 1866-1965. All regressions include year effects and week-of-the-year by decade effects. Regression in Columns 2, 4 and 6 include controls for log births, humidity, humidity squared, pressure, pressure squared, precipitation, precipitation squared, temperature, temperature squared, and five lags of temperature and temperature squared.

A.6.3 Total mortality analysis robustness results

Table 15 presents regressions assessing the robustness of the total mortality results. I focus on variations in my preferred total mortality specification – Column 2 of Table 2 – which looks at the effect of fog events across four-week windows.

Column 1 presents results excluding the log births control. Excluding this control has very little impact on the results. Column 2 includes additional temperature quadratics, $temp^3$ and $temp^4$ and five lags of each of these variables. These additional controls are not statistically significant and including them has little impact on the main results. This motivates my decision to exclude these controls from my main specifications. In Column 3 I look at whether fog days have a non-linear effect on mortality by including the squared number of fog days across the four-week window in the regression. The coefficient on this term is negative but also small and not statistically significant. Thus, I don't find evidence of a clear non-linear relationship between fog days and total mortality. In Column 4 I include controls for five lagged values of pressure, pressure squared, humidity, humidity squared, precipitation and precipitation squared.

The results in Table 16 examine the interaction between fog days and temperature. Since home heating was one of the primary drivers of pollution during the study period, these

interactions reveal whether fog events had more impact during periods in which pollution emission levels were high. In this regression I interact the number of fog days over the four-week window with the mean of the average weekly temperatures observed during the same period. The results show evidence of a negative relationship between fog events and temperature. This pattern indicates that fog events raised mortality more when temperatures were lower and thus emissions levels were higher. This makes sense given that the main impact of weather events associated with fog was to trap emitted pollution in the city. In terms of magnitude, these results suggest that a ten-degree (F) reduction in temperature causes the impact of a fog event to increase by 16-18%.

Table 15: Additional total mortality regression results using four-week windows

	DV: Log total mortality			
	Excluding log births	Temperature quadratics	Fog days squared	Lagged weather vars.
	(1)	(2)	(3)	(4)
Fog days (4-week window)	0.00833***	0.00837***	0.0118***	0.00709***
	(0.00232)	(0.00231)	(0.00411)	(0.00242)
Fog days squared (4-week window)			-0.000469	
			(0.000484)	
Observations	4,479	4,479	4,479	4,464

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. All regressions include a full set of year effects and week-of-the-year by decade effects as well as controls temperature, temperature squared, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared and five lags of temperature and temperature squared. Column 2 includes controls for $temperature^3$ and $temperature^4$ and five lags of both of these variables. Column 4 also includes controls for five lags of humidity, humidity squared, pressure, pressure squared, precipitation and precipitation squared. Data cover the full set of available observations from 1865-1965.

Table 16: Results interacting temperature and fog days

DV: Log total mortality	
Fog days (4-week window)	0.0437**
	(0.0207)
Fog days × mean avg. weekly temp (4-week window)	-0.000771*
	(0.000419)
Observations	4,479

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. Regressions run on the full set of available observations from 1866-1965. Both regressions include controls for year effects, week-of-the-year by decade effects, temperature, temperature squared, five lags of temperature and temperature squared, log births, pressure, pressure squared, humidity, humidity squared, precipitation, and precipitation squared.

A.7 Stillbirths and births

Table 17 presents results showing the impact of fog days on stillbirths. Column 1 shows that fog days did increase the number of stillbirths, by about 0.7 percent. Column 2 looks at the ratio of stillbirths to total births in the past year. This suggests that fog days increased stillbirths by over 15 deaths per 100,000 live births. If I take the magnitude of the estimated increase in stillbirths due to a fog day from Column 2 and apply it to the number of births per week and the number of fog days observed in the full study period then this implies that the acute effects of heavy fog events led to 1,143 additional stillbirths during the full study period, or 1.23 deaths per fog day. Of course, this result is based on the assumption that the relationship between pollution and stillbirths remained constant across the study period, which may not be reasonable.

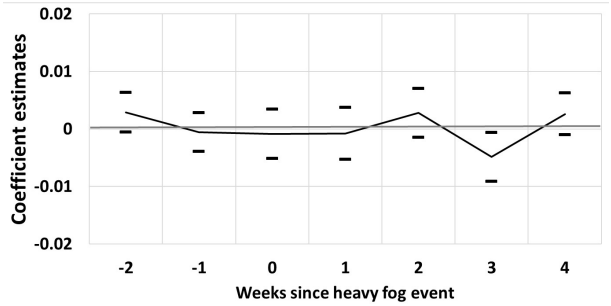
Table 17: Effects on infant mortality accounting for stillbirths

Dependent variable:	Log stillbirths	Stillbirths /total births
Fog days (4-week window)	0.00700*	15.72**
	(0.00373)	(7.991)
Observations	1,499	1,488
DW stat	2.05	2.04

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are presented because the Durbin-Watson statistics do not provide evidence that serial correlation is a major concern. All results include controls for temperature, temperature squared, five lags of temperature and temperature squared, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared and a full set of year effects and a week-by-decade effects. Column 1 also includes a control for log births. Data start in the 41st week of 1927 and continue through 1965 with a break during WWII. Results in Column 3 are per 100,000 births.

The next set of results looks at whether births were affected by fog days. The analysis follows the approach used for total mortality in Section 6.1. The first set of results, in Figure 22, compare log births to the number of fog days as well as leads and lags of that variable. These results provide no evidence that birth were affected by fog event. Table 18 provides some additional results looking at the effect across four-week windows, as in Table 2 in the main analysis. Again, I find no evidence that births were affected by fog events. This may seem surprising given that stillbirths were affected. However, stillbirths are only a small fraction of overall births, about 2.5% in the years in which stillbirths are reported.

Figure 22: Estimated effect of fog days on log births, 1866-1965



Coefficient estimates and confidence intervals for a regression of log births on the number of fog days in a week. Confidence intervals are based on Newey-West standard errors allowing correlation across observations within 6 weeks of each other. The births data do not show any evidence that serial correlation is an issue (Durbin-Watson statistics are around 1.9), but for consistency I follow the approach used in the main analysis. Regressions include controls for temperature, temperature squared, four leads and lags of these variables, as well as a full set of year effects and week-of-the-year by decade effect. Data cover 1866-1965.

Table 18: Effect of fog events on births in four week windows

	DV: Log births	
	(1)	(2)
Fog days (4-week window)	-0.000824	
	(0.000996)	
Fog indicator (4-week window)		0.00380
		(0.00351)
Observations	4,479	4,479

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. Regressions run on the full set of available observations from 1865-1965. Regressions include controls for temperature, temperature squared, five lags of these variables, pressure, pressure squared, humidity, humidity square, precipitation, precipitation squared and a full set of year effects and week-of-the-year by decade effects.

A.8 Changes over time

This section looks at how the effect of fog events on mortality changed across the study period. However, it is important to recognize that it is not possible to identify the extent to which these changes are due to changes in emission levels vs. changes in other conditions that influence the relationship between pollution and mortality.

Before estimating the impact of fog events on mortality in different time periods, it is useful to look at whether the reporting of these events appears to have changed over time. One way to do this is to look at the relationship between fog events and the quantitative weather variables in different periods. Table 19 presents averages of the weather variables on weeks in which heavy fog was reported broken down by different time periods. The notable feature here is that the characteristics of fog weeks was reasonably stable in the first three periods but changed in the last period (when fog reports came from Kew Gardens). The biggest change appears for temperature, which was typically in the low 40s during fog weeks in the late 19th and early 20th century but averaged 47.5 in the 1950s-60s. Similarly, average humidity during weeks with fog was much lower in the last period than in the three preceding periods. The main take-away from this table is that the relationship between fog events and underlying weather conditions appears to have been reasonably stable up to WWII but changed after that.

Table 19: Average weather conditions in weeks with heavy fog reported

	1865-1889	1890-1914	1918-1939	1951-1965
Avg. temp.	42.9627	44.09621	41.32805	47.53863
Avg. humidity	86.55556	85.43939	86.40244	81.82764
Avg. pressure	29.89972	29.91818	29.89161	29.8425
Avg. precipitation	0.28328	0.3246212	0.3332927	0.4115006

Next, I estimate the effect of fog events in different time periods. I break the sample up into three sub-periods: 1866-1900, 1900-1939 and 1951-1965. The first two periods are chosen to be roughly equal with a natural break at 1900, while last period covers the years for which weather observations are obtained from Kew Gardens rather than Greenwich.[57]

Table 20 presents results examining the evolution of the effect of fog events over time. The top panel looks at how the impact of reported fog days changed over time. The results

[57]Note that this analysis does not include 1949-1950. Data are available for these years, but I do not want to include these with observations after 1951 because the weather observations for these years come from Greenwich, while these observations are separated from the second period by the long break in the data from 1939-1949.

for the period before WWII show evidence that the impact of fog events, as a percentage of total mortality, was stable or slightly decreasing over time. In the bottom panel I use predicted fog events as an instrument for the number of fog days. These results show a similar pattern over time. The IV results are particularly useful when looking at the impact over time, since the instrument is derived from quantitative weather variables that were measured in a fairly consistent way across the study period. Note that the coefficients obtained in the IV regressions are larger than those in the OLS regressions. The most likely explanation for this is that the predicted fog weeks tend to identify the most severe fog events.

Table 20: Estimated mortality effects for different time periods

	DV: Log total mortality		
A. Using fog days as the key explanatory variable			
	1866-1889	**1900-1939**	**1951-1965**
Fog days (4-week window)	0.0121***	0.0101*	0.00869***
	(0.00381)	(0.00517)	(0.00298)
No weeks:	1768	1835	777
No. of fog days	304	179	429
Avg. fog days per week	0.172	0.098	0.552
Total implied deaths	23,033	8,512	12,090
Share of all deaths due to fog	0.0084	0.00396	0.0193
B. Using *predicted* fog weeks as an instrument for fog days			
	1866-1889	**1900-1939**	**1951-1965**
Fog days (4-week window)	0.0420***	0.0329*	0.0302***
	(0.0120)	(0.0173)	(0.00789)
Observations	1,763	1,835	767

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. All results include a full set of year effects and a week-of-the-year by decade effects as well as controls for log births, temperature, temperature squared, five lags of temperature and temperature squared, pressure, pressure squared, precipitation, precipitation squared, humidity and humidity squared. The regressions in the bottom panel also include controls for the temperature, pressure, humidity and precipitation indicator variables used to construct the predicted fog variable as well as five lags of these variables.

For the top panel I have also included calculations showing the number of deaths implied by the estimates in each period as well as the share of all deaths that these represent.[58] The number of deaths associated with fog events was greatest in the 19th century and then fell over time, due in part to a reduction in the number of fog events and in part to a reduction

[58]It is not realistic to do this with the coefficients estimated in the bottom panel. This is because if the predicted fog event instrument is identifying the most severe fog events then it is unrealistic to attribute this coefficient to all fog days and then use that to estimate the overall impact of fog events.

in the overall number of deaths occurring in London. As a share of total deaths in London, however, the impact of fog events was greatest in the last period. This reflects the fact that by this time London had become much healthier and other major causes of death, particularly infectious diseases, had become much less important.[59]

The next set of results look at how the distribution of effects across age groups evolved over time. These results split the available data into two periods, using 1900 as the cutoff, and focus on age groups over one. Before 1900, I find statistically significant effects across all age groups over age one, with particularly strong effects among children aged 1-5 and older adults. After 1900, I observe weaker effects for young and prime-age adults.

Table 21: Changes in the effects by age group over time

	DV: Log mortality					
	Estimates from 1866-1899					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	-0.000898	0.0212**	0.00945**	0.00640*	0.0113***	0.0133***
(4-week window)	(0.00486)	(0.00866)	(0.00403)	(0.00358)	(0.00364)	(0.00418)
Observations	1,252	1,252	1,539	1,539	1,539	1,539
	Estimates from 1900-1939					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	0.00496	0.0254*	0.000115	0.00474	0.0131**	0.0116*
(4-week window)	(0.00674)	(0.0136)	(0.00444)	(0.00575)	(0.00535)	(0.00695)
Observations	1,840	1,840	1,840	1,840	1,840	1,840

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. The results for children ages 5 and below use data from 1876-1965. Results for other age groups use data from 1870-1965. All regressions include controls for temperature, temperature squared, five lags of each of these variables, log births, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared, and a full set of year and week-of-the-year by decade effects. The number of deaths in the bottom row is calculated by applying the estimated percentage increase in deaths based on the regression coefficient and the average number of deaths in a week in each age group multiplied by four to reflect the four-week window over which effects occur.

One potential issue with the results in Table 21 is the impact of fog days as a percentage of total mortality may increase simply because mortality due to other causes fell. If mortality fell differentially across age groups, then this can make it difficult to make comparisons across age groups using this approach. As an alternative, Table 22 presents the impact of fog days on the share of overall deaths accounted for by each age group. These results suggest that in the period up to 1900, children aged 1-5 were relatively more susceptible to the effect of fog events compared to other age groups while infants and adults were relatively less susceptibility. After 1900 this changes: both children aged 1-5 and young adults aged 5-20

[59]See Appendix A.1 for a discussion of the changing mortality rate in London during the study period.

become much less susceptible to the effect of fog events, relative to other populations, than they were before 1900. This pattern suggests that something changed in the period after 1900 that reduced the impact of fog events on ages 1-20 more than on other populations. This is consistent with the effect of reductions in infectious diseases that particularly affected children and teenagers – measles and TB – which reduced the impact of fog events for these populations.

Table 22: Changes in the share of fog deaths in each age group over time

	DV: Log mortality					
	Estimates from 1866-1899					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	-0.0022***	0.00196**	-5.92e-05	-0.00065*	-1.15e-05	0.00046
(4-week window)	(0.0007)	(0.00097)	(0.00026)	(0.00035)	(0.0004)	(0.00061)
Observations	1,252	1,539	1,539	1,539	1,539	
	Estimates from 1900-1939					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	-0.00067	0.00035	-0.00046**	-0.00055	0.00069	0.00064
(4-week window)	(0.00096)	(0.00062)	(0.0002)	(0.0004)	(0.00045)	(0.0011)
Observations						

*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. The results for children ages 5 and below use data from 1876-1965. Results for other age groups use data from 1870-1965. All regressions include controls for temperature, temperature squared, five lags of each of these variables, log births, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared, and a full set of year and week-of-the-year by decade effects. The number of deaths in the bottom row is calculated by applying the estimated percentage increase in deaths based on the regression coefficient and the average number of deaths in a week in each age group multiplied by four to reflect the four-week window over which effects occur.

A.9 In utero exposure

A broad set of existing research shows that pre-natal exposure can increase infant mortality as well as impacting health later in life.[60] In general, my data are not well suited for assessing the impact of pre-natal exposure because I do not observe the date of birth of those who die. This makes it impossible to identify when they were in utero. However, it is possible to partially overcome this challenge by focusing on two series that allow me to roughly infer the period of birth. The first series that I consider is deaths due to premature birth, which generally occurred soon after birth. By using the week of death in which premature birth occurred as a proxy for the week of birth I am able to roughly identify the period in which each child who died prematurely was in utero. I also consider the impact of in utero exposure on stillbirths, many of which would have been near the time of birth.

Results looking at the impact of in utero exposure are presented in Table 23. Column 1 looks at the impact of the number of fog days on premature births in each previous twelve-week period. The main finding is that fog events occurring 25-36 weeks in the past, roughly the first trimester, are associated with increased infant mortality due to prematurity. In Column 2 I study the impact on stillbirths. Here I see effects from both very recent fog events and fog events in roughly the first trimester. Thus, both series provide evidence of the impact of in utero exposure. Note also that fog events more than 36 weeks in the past, as well as those in the period roughly corresponding to the second trimester, do not have any strong effects.

In terms of magnitude, these results suggest that an additional fog day in the first trimester increased the number of deaths due to prematurity by 0.74 percent. On average, there were 1.67 fog days in each twelve-week period for the years covered by the cause-of-death data, so on average in utero fog exposure in the first trimester increased mortality due to prematurity by about 1.23%. Prematurity accounted for 13.3% of deaths among those aged 0-1 during the period for which cause-of-death data are available, so this implies that deaths due to prematurity because of in utero exposure to fog events account for 1.6 out of every 1000 infant deaths during the years covered by this data series (1,352 deaths in total up to 1939). Put another way, this comes to 16.2 deaths per 100,000 live births. If I apply this figure to the number of births observed in all years covered by my data, assuming the same effect holds across the full study period, then I estimate that in utero exposure to fog events led to 1,436 additional infant deaths due to prematurity.

[60] See, e.g., Chay & Greenstone (2003), Currie & Neidell (2005), and Arceo *et al.* (2016). Jayachandran (2009) provides evidence suggesting that pre-natal exposure may be important. Currie (2013) provides a review of this literature.

During the period covered by the stillbirths data, which begins in 1927, there were on average 2.47 fog events in each 12 week period. This suggests that fog events raised the total number of stillbirths by 1.55%, equivalent to 750 deaths during the period covered by this series. Put another way, fog events led to 39.6 stillbirths for every 100,000 live births during the years from 1927 covered by my data. Applying the same figure to births across the full study period I estimate that first-trimester in utero exposure led to around 3,509 additional stillbirths in the years covered by this study.

Table 23: Evidence of in utero exposure

Dep. Var.:	Log mortality from prematurity	Log stillbirths
Years:	1866-1939 (except 1915-18)	1927-65 (except 1939-49)
Fog events this week	-0.00191	0.00354
	(0.00946)	(0.00857)
Fog events 1-12 weeks ago	0.00391	0.00423*
	(0.00243)	(0.00237)
Fog events 13-24 weeks ago	0.00105	0.00277
	(0.00268)	(0.00233)
Fog events 25-36 weeks ago	0.00738***	0.00629***
	(0.00284)	(0.00233)
Fog events 37-48 weeks ago	0.00281	0.000787
	(0.00291)	(0.00238)
Durbin-Watson stat.	1.93	1.98
Observations	3,283	1,452

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis. Robust standard errors are used because these data do not show strong evidence of serial correlation, as suggested by the Durbin-Watson statistics at the bottom of the table. All regressions include controls for temperature and temperature squared, five lags of these variables, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared and a full set of year effects and a week-of-the-year by decade effects. The data on premature births cover 1870-1939 (except 1915-1919). The data on stillbirths run from 1927-1965 (except 1940-49).

A.10 Additional cause of death results

Table 24 presents cause-of-death results using Newey-West standard errors allowing serial correlation across observations within six weeks of each other for those series that show evidence of serial correlation, which I define conservatively as a Durbin-Watson statistic below 1.5.

Table 24: Cause of death results with Newey-West standard errors

Cause of death	Coefficient		S.E.
Bronchitis	0.0349	***	(0.00690)
Measles	0.0791	***	(0.0219)
Pneumonia	0.0298	***	(0.00590)
Respiratory, other	0.00467		(0.0106)
Scarlet Fever	0.0349	***	(0.0106)
Smallpox	-0.00355		(0.0475)
Whooping cough	-0.0158		(0.0150)

Estimated coefficients and Newey-West standard errors with 6 week lag lengths for the impact of fog events in a week or in the previous three weeks on mortality within each cause-of-death category. All regressions include controls for temperature, temperature squared, five lags of both of these variables, log births, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared and a full set of year and week-of-the-year by decade effects.

Table 25 presents additional cause of death results. In these results, I run regressions that include one variable reflecting the number of fog days in a week or the previous three weeks, as in the main results, and a second variable reflecting the number of fog days in the next four weeks (i.e., in the future). If fog events increase mortality due to a particular cause of death, then we should see this increase in weeks during or after the occurrence of a fog event but not weeks just before a fog event. Thus, these results account for the possibility that mortality in some causes of death may be elevated even in weeks just before fog events (e.g., because of associated weather conditions). Table 25 presents the difference between the coefficient reflecting the impact of a fog event that has happened and the impact of a fog event up to four weeks in the future. The next two columns present test statistics from a Wald test for equality of these two coefficients.

Overall, results obtained using this approach are similar to the results shown in the main text. In particular, I find evidence that fog events were associated with substantial increases in mortality due to bronchitis, pneumonia, cardiovascular diseases, measles, and TB. The main difference, relative to the results in the main text, is that I no longer find

statistically significant evidence that fog events raised mortality due to scarlet fever while I do find evidence of elevated mortality from old age.

I have also generated cause-of-death results using the $PredFOG_t$ variable to construct an instrument for the number of fog days (available upon request). These results look fairly similar to those reported in the main text. The main differences are that the impact on deaths due to cardiovascular diseases is no longer statistically significant, though the coefficient is still positive, while there is evidence of a positive effect of fog events on mortality due to whooping cough.

Table 25: COD results comparing estimates from weeks just before and just after fog events

	Coefficient	Testing significance of difference	
Cause of death	difference	p-value	F-statistic
All causes	0.013	0.000	22.95
Respiratory & Cardiovascular			
Bronchitis	0.039	0.000	42.61
Pneumonia	0.027	0.000	22.88
Cardiovascular	0.009	0.011	6.43
Other respiratory	0.011	0.227	1.46
Infectious diseases			
Digestive	0.006	0.249	1.33
Diphtheria	-0.005	0.563	0.33
Measles	0.035	0.074	3.19
Scarlet fever	0.007	0.559	0.34
Smallpox	0.051	0.236	1.41
Tuberculosis	0.009	0.003	8.66
Typhus	0.019	0.444	0.59
Whooping cough	0.006	0.688	0.16
Other diseases			
Cancer	-0.002	0.656	0.20
Neurological	0.002	0.464	0.54
Old age	0.011	0.049	3.88
Premature birth	0.007	0.280	1.17
Venereal diseases	0.010	0.390	0.74
Other infectious dis.	-0.009	0.393	0.73
Other causes of death			
Accidents & violence	0.000	0.978	0.00
Alcoholism	0.018	0.201	1.64
Homicide†	0.010	0.576	0.31
Suicide	0.004	0.762	0.09
Other misc. CODs	0.001	0.669	0.18

See text for a description. † Homicide data are only separately available from 1870-1921. Homicide deaths are also included in the accidents and violence category.