London Fog: A Century of Pollution and Mortality, 1866-1965 *

W. Walker Hanlon
NYU Stern School of Business
and NBER

August 14, 2019

Abstract

This study draws on London's long experience with air pollution in order to improve our understanding of the overall effects of pollution exposure and how and why these effects evolve as locations develop. I compare uniquely detailed new mortality data covering 1866-1965 to the timing of London's famous fog events, which trapped emissions in the city. I show that air pollution was a major contributor to mortality in London over this period and that it interacted strongly with specific infectious diseases. As a consequence of this interaction, reductions in the infectious disease burden substantially altered the health costs of pollution.

*I thank Eva Arceo, Vellore Arthi, Alastair Ball, Brian Beach, David Cutler, Tatyana Deryugina, Casper Worm Hanson, Stephan Heblich, Richard Hornbeck, Maggie Jones, David Molitor, Matthew Neidell, Felix Pretis, Edson Severnini, Katherin Sudol and seminar participants at Bristol, CIDE, Simon Fraser, UCLA, the University of Illinois, the University of Chicago and the Economic History Society for helpful comments. I am grateful to Shmuel San, Giorgio Ravalli, and Anna Sudol for excellent research assistance. Funding for this project was provided by National Science Foundation CAREER Grant No. 1552692. Author contact information: NYU Stern School of Business, Henry Kaufman Management Center 7-160, 44 West 4th St., New York, NY 10012; whanlon@stern.nyu.edu.

1 Introduction

For over a century 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. This study focuses on understanding the acute impact of exposure to high-levels of pollution in London over the century from 1866 to 1965. Studying London, particularly during the 19th century, offers an opportunity to extend our evidence on air pollution into a setting with a more substantial disease burden and higher levels of pollution than most existing work, features comparable to the poorest modern developing countries. In addition, the availability of detailed mortality statistics allows me to study how these effects vary across age groups, how they interact with infectious diseases, and how these interactions influenced the costs of pollution as the city developed. This allows me to generate a number of new results which, largely for reasons of data availability, have not been considered in previous work.

One contribution of this study is to provide a novel assessment of the overall effect of acute exposure to air pollution in a highly polluted city over a long period of time. My results show that at least one out of every 200 deaths across the century I study is directly attributable to acute air pollution exposure, which I define as deaths occurring within four weeks of exposure to highly elevated pollution levels. In my preferred estimates this corresponds to around 39,000 deaths. In addition, I estimate that the acute effects of fog events caused over 1,000 additional stillbirths, while in utero exposure in the first trimester led to at least 1,400 infant deaths and an additional 3,500 stillbirths. To put this into perspective, this is roughly similar in magnitude to all deaths due to suicide, or to important infectious diseases such as smallpox. To my knowledge these are the first well-identified estimates of the acute impact of air pollution exposure on mortality across all age groups in such a heavily polluted environment.

A second contribution of this study is to provide new evidence on how the underlying disease burden in a location modifies the effect of pollution exposure. This allows us to begin understanding the mechanisms through which pollution effects in developing countries are likely to differ from those measured in the developed world. Specifically, I provide evidence that the presence of two infectious diseases of the respiratory system, measles and tuberculosis (TB), substantially increase the mortality effects of pollution. To my knowledge this is the first study to draw a clear causal link between pollution exposure and mortality from these diseases.[1] The interactions have not been cleanly identified in existing studies

[1] Appendix A.12 reviews a set of existing public health studies showing correlations between TB risk and

primarily because few of the poor developing countries where these diseases remain important also offer the detailed mortality data needed to study this link.[2]

The interaction between pollution exposure and the infectious disease environment implies that the impact of pollution exposure in London changed as the city developed. In particular, I show that progress toward eliminating these diseases in London after WWI substantially reduced the mortality effects of acute air pollution exposure. These results have implications for modern developing countries where both diseases remain prevalent. TB, in particular, remains the largest single-organism killer, accounting for 1.3 million deaths in 2017 (more than HIV/AIDS). Most TB cases are concentrated in locations, such as India and China, that also experience high levels of air pollution.[3] A simple back-of-the-envelope calculation based on my results suggests that the interaction between air pollution and TB accounts for around 16,000 deaths globally each year, with over 8,000 occurring in India, more than 1,000 deaths in each of Bangladesh, Nigeria and Pakistan, and around 600 deaths in China.

To generate these results, the key challenge is getting around the lack of direct pollution measures in the historical setting that I consider. Two new elements allow me to overcome this issue. 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 while also providing the statistical power needed to obtain reliable results. Second, I draw on a large 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. Leveraging this relatively high-frequency variation helps me avoid a number of potential identification concerns. 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

pollution and argues that it is difficult to draw causal conclusions from the existing evidence.

[2]For example, mortality data by cause of death is spotty in many countries, so in places like India the WHO is only able to make rough estimates of the overall burden of diseases like TB and measles.

[3](WHO, 2018). India and China account for 27% and 9% of the estimated 10 million global TB cases respectively.

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 conduct a number of tests to verify that my identification strategy is working well. For example, I show that the largest effects of pollution occur in exactly the cause-of-death categories suggested by the modern literature: respiratory diseases such as bronchitis and pneumonia as well as cardiovascular causes. In contrast, I find no interaction with many other causes of death, such as digestive diseases, where we would not expect acute pollution exposure to matter. I also conduct a check using heavy rainfall, which shares many features of fog–it is damp, causes people to stay indoors and impedes transport–but which differs from fog in that it acts to clean pollution from the air, rather than trap it. If these potential sources of bias are not too important, then we should expect estimates based on rainfall to deliver roughly mirror-image effects to those based on fog events. This is largely what I find. These checks help ensure that my estimates are not being driven by mechanisms other than pollution, such as changes in behavior, access to medical care, increased accidents or crime, etc.

One may wonder whether many of these deaths were simply due to the death of people who otherwise would have died soon after (harvesting). To assess this possibility, 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 find that fog events were associated with elevated medium-run mortality, implying 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 1.6% of all deaths in London during the study period.

In addition to my main results, I also provide novel evidence on the distribution of pollution effects across age groups and how this distribution is modified by the disease environment. These results differ from findings of the existing literature in two mains ways. First, while most deaths were concentrated among the elderly, I also find substantial impacts among children over age one, teenagers, and prime-age adults, groups that are not typically thought to suffer substantial mortality effects from short-run pollution exposure. I show that the impact of pollution on these populations depends heavily on the presence of measles and TB, which were major killers within these age categories. Second, I find that infant deaths are relatively less important than commonly thought and examine several possible explanations for this result. Together, these results indicate that focusing only on the impacts of pollution

on infants and the elderly, as is done in many existing studies, may be missing an important part of the overall effect of pollution in settings with high infectious disease burdens.

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.[4] A small but growing set of studies address air pollution, including Barreca *et al.* (2014), Clay *et al.* (2016), Beach & Hanlon (2018), Heblich *et al.* (2016), and Bailey *et al.* (2018).[5] This study has a particularly close connection to a paper by Clay *et al.* (2018) which shows that pollution can interact with influenza to increase mortality. While I find similar interactions (though with different diseases), I go beyond their results by looking across many ages and time periods while drawing on a different identification strategy. More importantly, I offer new evidence on how changes in the disease burden caused the cost of pollution to evolve over time.

This study also contributes to work using modern data to assess the health effects of pollution.[6] Evidence on the health effects of pollution in developing countries with high disease burdens remains limited (Greenstone & Jack, 2015). At the same time, developing countries typically face the highest pollution levels. Thus, there is a need for evidence documenting the effects of pollution exposure in developing settings, and, more importantly, understanding the mechanisms through which these effects might differ from those observed in the developed world.

One study that is particularly closely related to this paper is Jia & Ku (Forthcoming), which uses wind patterns to study the impact of Chinese pollution on mortality in South Korea. Another related study is Arceo *et al.* (2016), which uses temperature inversions to study the impact of pollution on infant mortality in Mexico City.[7] The use of temperature inversions to identify acute pollution effects is similar to using fog events, which were often accompanied by inversions.[8] These studies provide some of the best available evidence on the acute impact of pollution on mortality outside of the U.S. and Europe. However, both of these studies consider environments that are substantially healthier, and less polluted, than

[4]See, e.g., Cutler & Miller (2005), Ferrie & Troesken (2008) and Alsan & Goldin (2019).

[5]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, 2018), 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.

[6]Useful reviews of this literature include Currie (2013), Currie *et al.* (2014), Graff Zivin & Neidell (2013) and Rückerl *et al.* (2011). This study is also similar to work that focuses on the effects of acute air pollution exposure (Pope, 1989; Schlenker & Walker, 2016; Knittel *et al.*, 2016; Jia & Ku, Forthcoming; Halliday *et al.*, Forthcoming).

[7]Another related paper is Hanna & Oliva (2015).

[8]In fact, the temperature inversions in Mexico City that they study are often accompanied by fog (Gonzalez-Viveros *et al.*, 2018).

either the setting I study or industrialized modern developing countries. For example, in the setting I consider there were 96.7 infant deaths per 1,000 births, a value that is similar to the levels observed in the very poorest countries today. In contrast, infant mortality in Mexico City in the period studied by Arceo *et al.* (2016) was just 19.9 per thousand births, well below the current global average. In addition, because I study a much more polluted setting than these studies, my results extend our understanding of pollution effects beyond the levels examined in existing work.[9] A second important difference between this study and existing work such as Jia & Ku (Forthcoming) or Arceo *et al.* (2016) is my ability to break down effects into detailed age and cause-of-death categories. This allows me to identify previously unknown interactions between pollution and certain infectious diseases.

A small number of other papers also investigate the health effects of air pollution in developing settings.[10] Greenstone & Hanna (2014), for instance, studies the effect of air pollution regulation on infant mortality in India. They find modest and not statistically significant declines in infant mortality as a result of reductions in air pollution. Another related paper, also focused on infants, is Jayachandran (2009), which finds evidence that wildfires in Indonesia caused a substantial reduction in birth cohort size, with the largest effects associated with third-trimester exposure. Rangel & Vogl (2016) also find that third-trimester exposure to smoke from agricultural fires reduces in utero survival.[11] One thing to note about all three of these studies is that they focus only on infant mortality. My results suggest that studies focused only on infants this may be missing important aspect of the health effects of pollution exposure in developing countries.

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 of the relationship between fog events and pollution is conducted in Section 5, followed by the main analysis in Section 6, a discussion of implications for modern developing countries, in Section 7, and concluding remarks in Section 8.

[9]London during the period for which I have pollution data also appears to have been more polluted than Mexico City during the period studied by Arceo *et al.* (2016), more closely resembling the most highly polluted developing cities today. Note that the data used by Arceo *et al.* (2016) have a weekly average of maximum 24-hour pollution levels of 67 $\mu g/m^3$ of PM10, equivalent to roughly 122 $\mu g/m^3$ of TSP (this calculation uses the conversion ratio of PM10=0.55*TSP from their study). In London from 1951-62, the weekly average of daily maximum pollution levels is 260 $\mu g/m^3$ of TSP.

[10]Other studies of air pollution in developing countries are Foster *et al.* (2009) and Chen *et al.* (2013).

[11]Like these studies, I find that third-trimester pollution exposure can increase stillbirths, though my results show larger first-trimester effects. A possible reason for this difference is that I study the effects of coal smoke, which releases a different set of pollutants than forest or agricultural fires.

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,[12]

> *Here we 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...*

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 $\mu g/m^3$.[13] 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 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.[14] 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.[15]

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

[12] *Little's The Living Age*, 26 May, 1866.

[13] 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.

[14] Automobiles may have been an important contributor toward 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, 2018). 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).

[15] 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.

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 stable so that it is in contact with the ground for long enough to cool.[16] 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.[17] 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, 2003b; 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; Brook *et al.*, 2010; Rückerl *et al.*, 2011; Rajagopalan *et al.*, 2018). There is also limited evidence that pollution effects may interact with infectious diseases to increase mortality (Clay *et al.*, 2018).

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

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

[17] 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.

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.[18]

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 reports stored in the British Library, the LSE Library, and the New York Public Library.[19] 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. In total the mortality data include over 350,000 observations. Summary statistics for the main data series are available in Appendix Table 7.

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. Another reason to end the study at this point is that after 1965, rising temperatures mean that there were fewer and fewer fogs in London, weakening my ability to leverage fog events in order to study pollution effects. 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. This motivates my use of London as a whole as the unit of analysis.

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

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

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

as many stillbirths as total deaths of infants aged 0-1.[20] 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) information. 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 21 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, 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.

Naturally, one may worry about the accuracy of cause of death data drawn from a historical setting. However, results obtained from the cause-of-death data consistently match the patterns that we expect, suggesting that these categories are reasonable. For example, comparing the COD data to temperature yields reasonable results, with low temperatures associated with deaths due to respiratory diseases, mainly among the elderly, while high temperatures are associated with deaths due to digestive diseases, particularly among infants. In addition, the distribution of causes of death across age groups is consistent with what we would expect. These patterns, together with those I will report later in this study, indicate that the cause-of-death data are able to deliver reasonable results.

Most of the deaths reported in the Registrar General's reports would have been registered within 2-3 days of their occurrence.[21] 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,

[20]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.

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

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 11 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 11 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 discontinuity into account in the analysis.[22]

Seasonality is an important feature of both mortality and fog events. Figure 12 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.[23] 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, relative humidity (hereafter just 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. Finally, note that wind speed and direction, while reported in some of the Registrar General's reports, is not available in a systematic way, so it is not possible to use this to generate additional variation in pollution levels (as done recently by Deryugina *et al.* (Forthcoming)).

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.[24] The reported values include the weekly average of the daily maximum and

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

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

[24]The pollution measures were obtained from a device, called the Owens Smoke Meter, in which a fixed

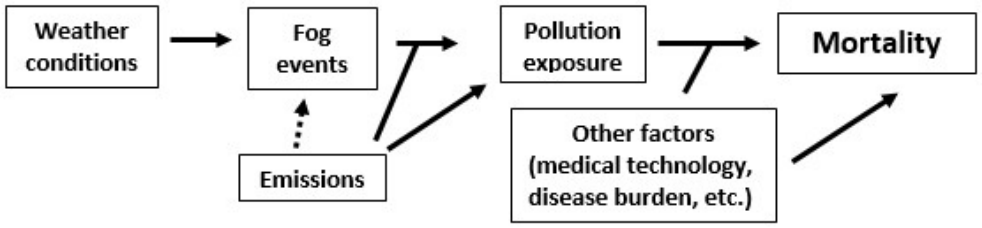
mean pollution levels in units that correspond to the total suspended particulate (TSP) values commonly reported during this period.[25] These data show that pollution was highly seasonal and generally declining from 1951-1962.

4 Methodology and identification

4.1 Overview and identification concerns

Figure 1 describes the basic relationships at work in this study. Ideally I 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, 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

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.

[25]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.

on changes in emission levels, which are not observable.[26] 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.

It is possible to isolate one factor that changed the effect of pollution exposure on health: the infectious disease environment. Specifically, by estimating the interaction of pollution exposure with mortality due to specific infectious diseases, and then applying these estimates to the changing disease environment in London across the study period, it is possible to isolate the influence of changes in the infectious disease environment on the mortality costs of pollution. While the infectious disease environment is only one factor that affected how the relationship between fog events, pollution exposure, and health evolved across the study period, it turns out to be an important part of the story.

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.[27] However, to be sure that this is not driving my results I will use data on underlying weather conditions to predict the timing of fog events. These climate conditions are generally not thought to be affected by emissions levels, which is why they are commonly relied on to identify pollution effects.[28] Using climate conditions to predict fog events 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

[26]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.

[27]For example, Figure 11 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 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.

[28]For example, Arceo *et al.* (2016) rely on temperature inversions to identify pollution effects, while Deryugina *et al.* (Forthcoming) exploit wind direction. Both of these are closely related to climate conditions such as pressure and temperature.

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).[29] 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 (2019), Beach & Hanlon (2018) and Jia & Ku (Forthcoming).

Another concern is that fog may have caused people to stay indoors which could have increased mortality by facilitating the spread of disease.[30] However, staying indoors also reduces exposure to people outside of the household, which may have actually reduced infectious disease transmission.

As a check on all of the results obtained from the analysis of fog events I conduct a second analysis focused on the effect of heavy rain. Heavy rainfall is similar to fog in a number of ways: it is damp, causes people to stay indoors, and can disrupt transportation. However, unlike fog, rain is known to reduce pollution levels by "washing" particulates and other pollutants from the air.[31] Thus, rainfall provides a useful check on the fog results. If the mortality effects of fog are being driven by factors like people staying indoors, increased moisture in the air, or transportation disruptions, then the effects for fog should look similar to those obtained for heavy rain. If instead, the effect of fog events that I estimate are reflecting the impact of pollution, then the estimates for heavy rain should look like a mirror image of the fog effects. In fact, that is exactly what I find. This identification check is

[29] 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.

[30] 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.

[31] Note that a previous version of this paper reported no strong impact of rainfall on pollution levels. That result was obtained from regressions that controlled for humidity levels, which are strongly correlated with rain. Once humidity is omitted as a control, the data show a clear reduction in pollution associated with rain, as described in Appendix A.11. A second important difference between fog events and heavy rain, which I discuss in the Appendix, is that rain can affect water quality and increase digestive diseases.

discussed in more detail in Appendix A.11.

4.2 Estimation procedure

This study begins with a simple time-series analysis approach that relies on random variation in the week-to-week timing of fog events driven by climatic factors. In robustness exercises I also move beyond pure time-series identification by comparing across cause-of-death categories in a difference-in-difference style framework. As a starting point I consider the following specification:

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

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.[32] 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.[33] The regressions also include controls for other weather variables – pressure, relative humidity and precipitation – and squared values of these terms.[34] These do not have

[32]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.

[33]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.

[34]In robustness exercises, I study the impact of using absolute rather than relative humidity. This alternative makes little difference, which is not surprising given that the relative humidity control does not have

as strong an impact as the temperature 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.

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).

Serial correlation is a potential concern in this study. To deal with 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).[35] 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. I have also tested the series for stationarity. Standard tests strongly reject the null hypothesis of a unit root for both total mortality and mortality within each age category.

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.[36] 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

a substantial impact on my results.

[35] See Appendix A.5.1 for further details.

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

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 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.

At first glance it may seem concerning that this instrument includes variables, such as temperature, which is also an important control variable. However, the fact that my instrument is based on a four-way interaction is critical. The way to interpret one component, say temperature, in this instrument is that there are many days with low temperature, but those days are only likely to see fog formation if the other variables are also conducive to fog. Thus, along one dimension the instrument is comparing days with similar temperature, but where other conditions are or are not conducive to fog formation.

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. To choose appropriate cutoff values I conduct a grid search across different combinations of cutoff values and run regressions looking at how well each combination identifies the timing of fog events. As decision criteria, I focus on the t-statistic of the coefficient on the $PredFOG_t$ term and the increase in the R-squared that is achieved by adding this term to the regression. Both criteria lead to the same preferred combination of cutoff values: average temperature below 40, humidity above 82.5, pressure above 29.9 and rainfall below 0.5. This combination

identifies 145 predicted fog weeks across the study period. 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.

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.[37] Put another way, the IV estimates reflect a local average treatment effect, and there is a clear reason to expect that this will be larger than the average treatment effect associated with fog days, with the size of the difference increasing in the restrictiveness of the criteria used to construct the instrument.

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 19. 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.[38] 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.

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

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

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.[39] 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.[40] 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.

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 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. Appendix Figure 18 shows that similar results are also obtained when I use $PredFOG_t$ as the explanatory variable in place of actual fog events.

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.[41] 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. It is

[39] In Appendix A.4.1 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.

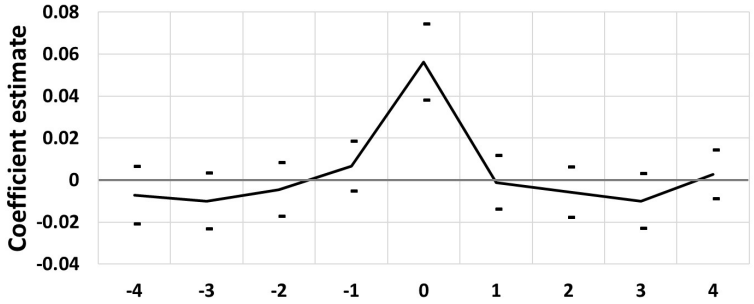
[40] Given the shorter time period over which these data are available I do not allow the week-of-the-year effects to vary by decade in this specification.

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

important not to over-interpret these results, since they are based on a small sample relative to the main results, may be affected by endogeneity bias, and rely on rudimentary pollution measures (see footnote 24). One way to improve on these results is to use fog days as an instrument in order to reduce the effects of measurement error in the pollution measure as well as concerns about endogeneity.

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.

In Column 2 of Table 1 I present reduced form results using the number of fog days in a week as the key explanatory variable. 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 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.

Columns 3-4 present 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.

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)
Impact on weeks before fog event or pollution measurement					
Pollution or fog days on t-2	-0.0534	0.00109		0.0471	0.0348
	(0.0462)	(0.00453)		(0.0731)	(0.0356)
Pollution or fog days on t-1	-0.0983**	-0.000498		0.0310	-0.0186
	(0.0490)	(0.00504)		(0.0782)	(0.0384)
Contemporaneous					
Pollution or fog days on t	-0.0126	0.00662	0.124	0.141	0.0141
	(0.0394)	(0.00488)	(0.0929)	(0.105)	(0.0582)
Impact on weeks after fog event or pollution measurement					
Pollution or fog days on t+1	0.0848	0.0134*	0.189*	0.197*	0.131**
	(0.0615)	(0.00701)	(0.0974)	(0.103)	(0.0548)
Pollution or fog days on t+2	0.0585	0.0184***	0.264***	0.266***	0.188***
	(0.0379)	(0.00501)	(0.0666)	(0.0696)	(0.0310)
Pollution or fog days on t+3	0.0178	0.00149	0.0231	0.0356	0.117**
	(0.0563)	(0.00660)	(0.0959)	(0.101)	(0.0483)
Pollution or fog days on t+4	0.0958**	0.0149**	0.217**	0.218**	0.120***
	(0.0459)	(0.00678)	(0.0888)	(0.0891)	(0.0303)
Pollution or fog days on t+5	0.0331	-0.00437	-0.0363	-0.0286	0.0361
	(0.0482)	(0.00451)	(0.0695)	(0.0697)	(0.0274)
IV f.s. F-stat			11.483	5.74	
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, 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.

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.[42]

Importantly 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 was endogenously affected by changes in pollution levels.

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.[43] In the top panel the key explanatory variable is the number of fog days in a week (left panel) or an indicator for any heavy fog in a week (right panel). In the bottom left panel the explanatory variable is predicted fog events while the bottom right panel uses predicted fog as an IV for the number of fog days in a week.

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.[44] 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

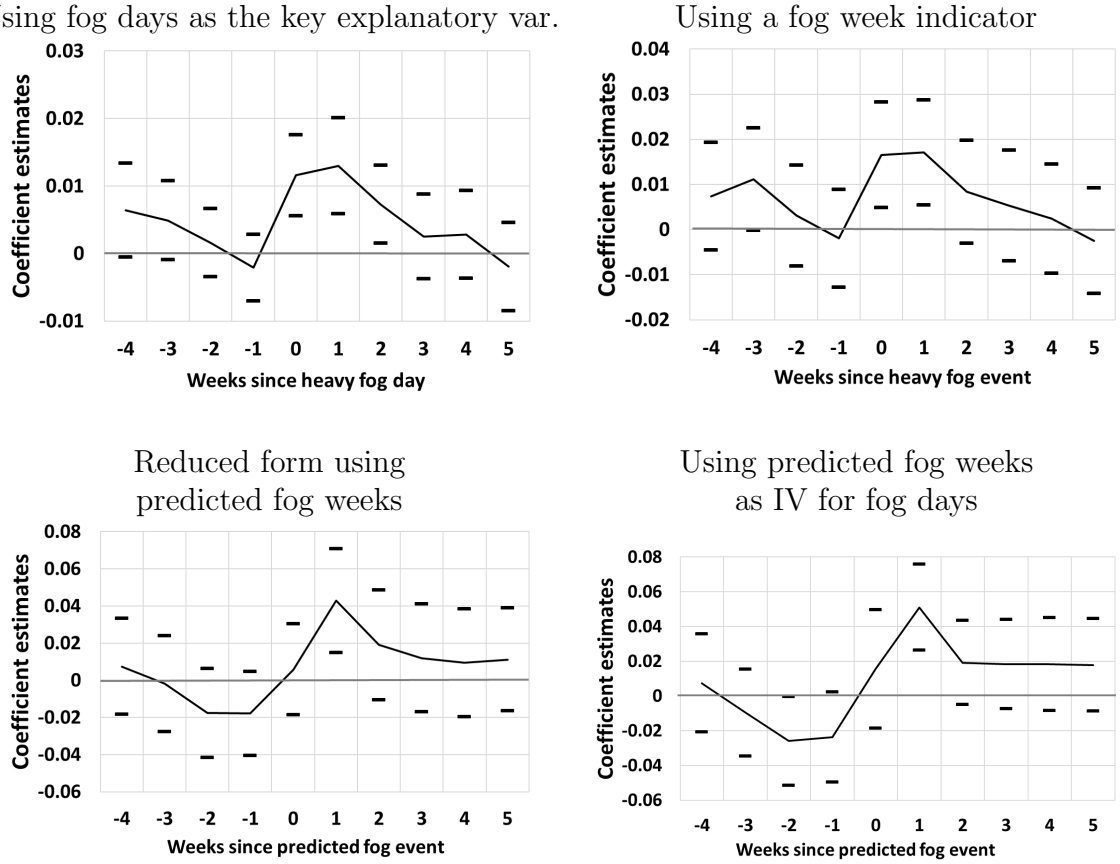
[42] I do not present results using the predicted fog variable as an instrument for pollution because this instrument is not strong enough across the relatively short period for which pollution data are available.

[43] 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.

[44] 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.

downward trend in mortality. 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.[45] This pattern suggests that the estimated effects documented in this paper may mildly understate the true impact of pollution exposure.

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



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, 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 set of results also include 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 first stage F-statistic for the results in the bottom-right panel is 10.028.

[45] Appendix A.2.6 presents some evidence showing that weather conditions such as temperature and pressure were trending mildly in the run-up to fog events. Of course, I control for temperature and pressure, but the trends in those variables suggests that there may be trends in other related weather conditions that I cannot observe.

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.[46] 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 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 0.84 percent across a four-week window starting in the fog event week. This is my preferred specification. 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 38,980 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 (50,363 deaths) if I focus instead on the results using the fog week indicator in Column 3.[47] 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 145 of predicted fog events. The estimates in Column 4 imply that fog events led to 24,714 deaths or 0.44 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 145 predicted fog events I estimate 27,637 deaths, or 0.49 percent of all deaths. Thus, I conclude that, even using this conservative specification, roughly one out of every 200 deaths in London during the years covered by my data are associated with the acute effect of fog events.

The fact that the IV estimate is much larger than the OLS is not surprising and does not indicate that the IV is invalid. The difference between the IV and OLS estimates is a

[46] Appendix A.5.2 presents results obtained using windows of different lengths ranging from two to five weeks.

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

natural and expected result of the fact that the instrument is identifying weeks with the most severe fog events, which we would expect to have a greater mortality effect. This fact does, however, mean that we should not interpret the IV estimate as reflecting the average effect of a fog event. This is fine, since the goal of this study is to estimate the overall magnitude of pollution effects associated with fog events, rather than obtaining a precise estimate of the impact of the average fog event, which naturally depends on the specific definition of an event. When handled correctly, both the IV and OLS yield similar predictions of the overall number of deaths associated with fog events.

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***	0.00835***			0.0375***	
(4 week window)	(0.00281)	(0.00232)			(0.00771)	
Fog week ind.			0.0172**			0.166***
(4 week window)			(0.00683)			(0.0367)
Pred. fog weeks				0.0336***		
(4 week window)				(0.0127)		
Additional controls		Yes	Yes	Yes	Yes	Yes
IV f.s. F-stat					123.3	64.3
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 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.

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 (Column 2) to data ending in 1939, I estimate that the acute effects of fog events caused around 26,000 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).

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. In Appendix Figure 23 and Table 16 I examine the impact of including a richer set of controls for temperature and absolute humidity. I find that controlling for these weather variables in a more non-parametric way has only a small impact on my results, even with a very flexible specification. Given this, in the main analysis I opt for a more parsimonious set of temperature and humidity controls.

In Appendix A.5.4 I present some additional robustness results using the approach in Table 2. These show that the main findings are robust to including additional temperature quadratic terms, controlling for the number of births, etc. I also look for non-linearity in the effect of fog days, by estimating separate coefficients depending on the number of days in a week. These results, in Appendix A.5.3, show that the contemporaneous or one-week lagged effect of fog are fairly linear in the number of fog days, but that longer lagged effects are driven primarily by the worst fog weeks, those with four or more fog days. 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.5.4 show that when fog days are interacted with temperature there is 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 18%.[48]

It is also possible to look at how these effects evolved over time, which is done in Appendix A.7. 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 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

[48]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).

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 presents estimates of the impact of fog days on mortality by age group.[49] 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. Before considering the causes of this surprising pattern, it is useful to consider the lag structure of the results, shown in Figure 4.

Table 3: Mortality effects by age group

	DV: Log mortality					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	-0.0006	0.0154**	0.0102***	0.0063***	0.0092***	0.0098***
(4 week window)	(0.0028)	(0.0064)	(0.0034)	(0.0023)	(0.0022)	(0.0025)
	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

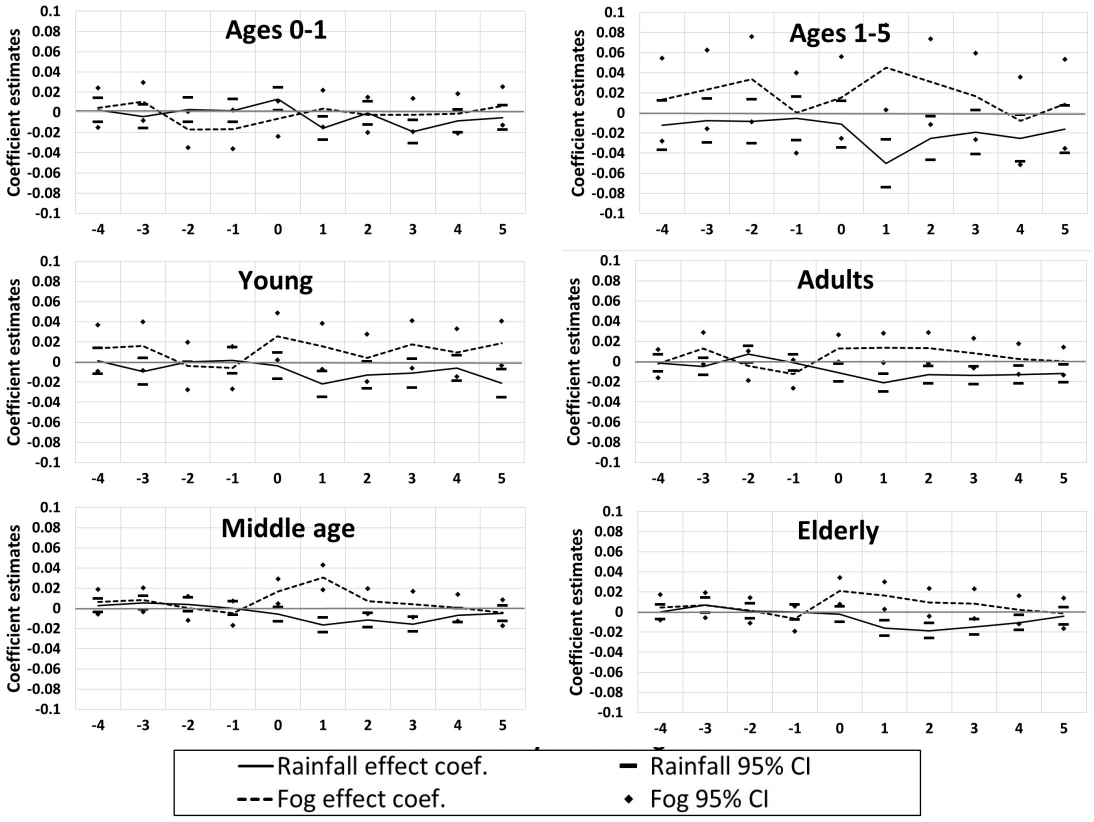
*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. All results include controls for temperature, temperature squared, four lags of temperature and temperature squared, five lags of temperature and temperature squared, 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.

Figure 4 provides further detail on the by-age results. For each age category, this figure presents the estimated effect of fog events, including leads and lags, on mortality. Results

[49]Results using fog week indicator variables are similar. Results obtained using predicted fog as an instrument for fog days generate similar patterns as well, though the estimates are less precise so the increase in mortality associated with fog days is only statistically significant for the middle-age and elderly groups.

are also reported for regressions using heavy rainfall in place of fog events. As discussed above, if the fog analysis is performing well, then we should expect the rainfall check to provide mirror image results. For most age categories we see that the rainfall check confirms the results obtained using fog events. The main difference between the rainfall and fog event results is in the week in which the event occurred. As discussed in Appendix A.11, this difference is due to an increase in digestive diseases in weeks in which heavy rain occurred. We can see this most clearly in the statistically significant increases in deaths in the 0-1 age group, the category most affected by digestive diseases, in weeks when heavy rainfall occurs. This effect provides a strong reason to prefer the fog event results over the rainfall check results, since fog events have no similar impact on digestive deaths.

Figure 4: Results by age group including the rainfall check



Each graph presents results for two regressions. The dotted lines describe coefficient estimates for how the occurrence of any fog day in a week affects mortality in each age group. For comparability to the rainfall results I focus on an indicator variable for any heavy fog event, rather than the number of fog days. The solid lines describe coefficient estimates for how the occurrence of heavy rainfall affects mortality. All results include controls four leads and four lags of temperature and temperature squared as well as a full set of year and week-of-the-year fixed effects. The fog results also include controls for pressure, pressure squared, humidity, humidity squared, rainfall and rainfall squared. Confidence intervals are generated using Newey-West standard errors allowing correlation across observations up to six weeks apart.

The fact that in both Table 3 and Figure 4 I observe no evidence of an increase in infant mortality associated with fog events is very surprising given existing results from developed and middle-income countries. One potential explanation for this difference is that the heavier disease burden in the setting I study may have reduced the number of infants at risk of dying from pollution exposure. If this is the key mechanism then we should see more infant deaths associated with fog events later in the study period, since infant mortality from other causes fell substantially over time. While I do see some increase over time in the impact of fog events on infant mortality (see Appendix Table 21), even after 1900 infants remained on average less vulnerable to the effects of fog events than people in other age categories (see Appendix Table 22). Thus, this does not appear to be the key channel at work. A second factor that may also be contributing to the lack of infant deaths associated with fog events is the high baseline pollution level experienced in London during non-fog weeks. If this high baseline level killed off most of the infants that were vulnerable to pollution effects, that may explain why we do not see elevated levels of infant mortality in weeks when pollution levels were particularly high.

A third potential explanation for the lack of infant mortality effects is that many infant deaths may have been classified as stillbirths. Indeed, when I do observe statistics on stillbirths, starting in 1927, there is evidence of increases associated with fog events. Results 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. Once I include stillbirths the magnitude of the effects on infant mortality appear roughly similar to previous work.[50] This suggests that part of the difference between my results and those observed in existing work may be attributable to stillbirths. Extrapolating the estimated stillbirth effects across the full study period, I estimate that acute exposure due to fog days caused 1,143 additional stillbirths. Thus, in terms of magnitude, even when I include stillbirths as infant mortality, the impact of fog events on infant deaths is still substantially smaller than the impact observed for other age groups.

In Appendix A.7 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

[50]Using the estimated relationship between fog events and pollution levels from the period in which I have pollution data, it is possible to produce a rough comparison between my results and the estimates from Mexico City reported by Arceo *et al.* (2016). Their results suggest that a one-unit increase in TSP raised infant mortality by 0.42 deaths per 100,000 births. Using the relationship between fog events and stillbirths I estimate using data after 1927, my results imply that a one-unit increase in TSP raises stillbirths by 0.27 per 100,000 births. If I combine stillbirths and infant deaths during this period, then my estimates suggest that a one-unit increase in TSP raises the combination of these by 0.57 per 100,000 births.

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.

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 fetal underdevelopment, which typically occur close to the date of birth.[51] 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.8 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 fetal underdevelopment 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 a broad set of 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 years of life. 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

[51] These are labeled "prematurity" in the Registrar General's data, but because it would have been hard for contemporaries to establish the pregnancy term during much of the study period, this most likely includes a broader set of deaths due to fetal underdevelopment than what we would label prematurity today.

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. However, as I will discuss in Section 6.3, the breakdown of medium-run effects by cause of death looks reasonable, which suggests that these identification concerns are not too substantial.

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 prior to 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. Also, in order to exploit variation at this level it is necessary to replace 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.[52] 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 25% of the total mortality impact of fog events while the medium-term impact accounts for 75%, 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 1.6% of

[52]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.

all deaths in London during the period up to WWII.

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. Additional results breaking these effects down by age group are available in Appendix A.9.

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

	DV: Log mortality	
	(1)	(2)
Fog days - acute effects (4 week window)	0.00965***	0.00953*
	(0.00334)	(0.00566)
Fog days - medium run (past 52 weeks)	0.00225**	0.00224**
	(0.00107)	(0.00107)
Fog days (4 weeks) × fog days in medium run (past 52 weeks)		1.45e-05
		(0.000755)

*** p<0.01, ** p<0.05, * p<0.1. N=3,275. 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, 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.

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 21 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).[53] The cause-of-death data cover 1870 to 1939.

Table 5 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

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

first few rows describe mortality in a set of causes of death related to the respiratory and cardiovascular systems, the categories most clearly linked to the impact of air pollution in modern studies.[54] 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 10.1 and 21.8 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.[55] Together, the four respiratory and cardiovascular categories at the top of Table 5 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. I also observe effects associated with "other infectious diseases," a category that contains a range of infectious diseases such as diphtheria and typhus. However, as discussed below, this estimate disappears when using alternative estimating procedures. It is interesting to note that many other important infectious diseases, such as digestive diseases (diarrhea, cholera and typhoid), smallpox, a viral skin disease, and whooping cough do not show increases during fog events. 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 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 miscellaneous 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.[56] Deaths due to old age, a category that

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

[55]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.

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

is somewhat vague, do increase during fog events, though the coefficient is not statistically significant. Fetal underdevelopment (e.g., prematurity), 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.8 show that deaths due to fetal underdevelopment 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 5 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, means that accidents and violence cannot be a primary driver of the impact of fog events on overall mortality.

As a check on these results, Appendix A.11 presents results examining the relationship between heavy rain events and mortality in different cause of death categories. In almost all cases the results based on heavy rain support the patterns identified in Table 5. The two exceptions are scarlet fever and other infectious disease deaths, which do not appear to be impacted by heavy rain events. This suggests that the significant results reported for these categories in Table 5 may be due to random chance. These results also disappear in some additional robustness exercises, as discussed below. Thus, I interpret the estimated effects operating through scarlet fever and other infectious diseases as spurious results.

We can draw two main lessons from the results in Table 5. 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.

Table 5: 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	***	(0.0045)	18.9
Cardiovascular	0.0084	***	(0.0024)	5.4
Pneumonia	0.0298	***	(0.0040)	11.7
Misc. respiratory	0.0045		(0.0069)	
Infectious diseases				
Digestive dis.	-0.0037		(0.0037)	
Measles	0.0796	***	(0.0137)	10.6
Scarlet Fever	0.0351	***	(0.0081)	2.3
Smallpox	-0.00398		(0.0320)	
Tuberculosis	0.00615	***	(0.0021)	4.0
Whooping cough	-0.000186		(0.0101)	
Infectious, other	0.0138	***	(0.0046)	
Other diseases				
Cancer	0.00151		(0.0024)	
Neurological dis.	-0.0018		(0.0022)	
Old age	0.0044		(0.0037)	
Fetal underdevelopment	0.00027		(0.0039)	
Venereal diseases	-0.0016		(0.0079)	
Other misc. causes	0.0012		(0.0017)	
Other causes of death				
Accidents/violence	0.0026		(0.0036)	
Alcoholism	0.00832		(0.0094)	
Homicide†	-0.0072		(0.0122)	
Suicide	-0.0012		(0.0083)	
All causes	0.0099	***	(0.0021)	53.9

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 25. 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, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared, and a full set of year and week-of-the-year by decade effects. Data cover 1870 to 1939. † Homicide data are only separately available from 1870-1921. Homicide deaths are also included in the accidents and violence category.

Second, these results show that fog events interacted with a specific set of infectious diseases of the respiratory system in a way that increased mortality from these diseases. Together, these diseases – measles and TB – account for 27 percent of the total deaths associated with fog events.[57] While a number of studies, discussed in Appendix A.12, show

[57]These diseases account for 14 percent of deaths overall.

a correlation between pollution exposure based on location of residence and TB mortality, to my knowledge this is the first study to draw a clear causal link between pollution and TB or measles mortality. The channels through which pollution increases infectious disease mortality are not currently well-understood, but the timing of the results I provide can help shed some light on this connection. In particular, the typical period between infection and death for TB and measles tells us that the acute effects of pollution documented in my main results are unlikely to be due to an increase in disease transmission due to pollution exposure. Instead, it must be the case that pollution increases the probability of death of individuals that already have the disease.

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 5 except that I no longer find statistically significant effects operating through scarlet fever or other infectious diseases. This provides further evidence that those causes of death were not being impacted by pollution exposure. In another set of results, I use the predicted fog variable to instrument for the timing of fog events. This also generates similar results.

Appendix A.10 also presents results where I focus on respiratory causes of death and include as controls causes of death not thought to be associated with acute pollution effects, such as cancer, neurological disorders, or venereal diseases. This framework mimics a difference-in-difference approach comparing across cause of death categories. The results are very similar to those shown in Table 5.

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 fetal underdevelopment. From Table 5 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 5, 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. However, it remains common in the developing world. 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 5.

One implication of the results shown in Table 5 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.[58] In the 19th and early 20th century, this group faced substantial mortality risk from infectious diseases, particularly measles, which accounted for 14% of deaths in this age group. Table 6 presents estimates of the impact of fog events on this age group using data from 1875-1914 including all cause of death (Column 1) and excluding measles mortality (Column 2). We can see that once measles deaths are excluded the impact of fog events on children aged 1-5 drops by half. A similar story holds for prime-aged adults and TB, which accounted for 39% of mortality among that age group. The next two columns of the table show the estimated impact of fog events on adults aged 20-40 including all causes of death, in Column 3, and all deaths excluding TB, in Column 4. Columns 5-6 show the effect of excluding these causes of death when looking at all-age mortality. This shows that the estimated effect of pollution on overall mortality drops by roughly one-quarter when deaths due to these two diseases are excluded from the analysis.

Further analysis (available upon request) shows that the medium-run increase in mortality is being driven by deaths due to bronchitis, cardiovascular deaths, whooping cough, and other respiratory deaths. These patterns seem reasonable and the medium-run increase in other respiratory deaths, which includes influenza, is consistent with the results found by Clay *et al.* (2018). There is also evidence of an increase in neurological deaths, a common cause of death for infants. Only three major COD categories show medium-run reductions associated with fog events: cancer, scarlet fever, and the undefined "old age" category. This suggests that there may have been some harvesting effects among the very old or among young children, but they appear to be more than offset by the medium-run increase in mortality in other categories. There is no evidence of harvesting effects for deaths due to TB or measles. All of the primary cause-of-death patterns discussed above survive when the

[58] I study the impact of measles on the 1-5 age group because that was the group where that disease was most important. I study the impact of TB on the 20-40 age group for the same reason.

regressions also account for medium-run effects.

Table 6: Effect of certain infectious diseases on mortality by age group

	Ages 1-5		**Ages 20-40**		**All Ages**	
Dependent variable:	All deaths	Excluding measles	All deaths	Excluding TB	All deaths	Excluding measles and TB
Fog days (4 week window)	0.0209*** (0.00703)	0.00972* (0.00581)	0.00579** (0.00254)	0.00395 (0.00301)	0.00921** (0.00366)	0.00693* (0.00376)
Observations	2,023	2,023	1,878	1,878	1,878	1,878

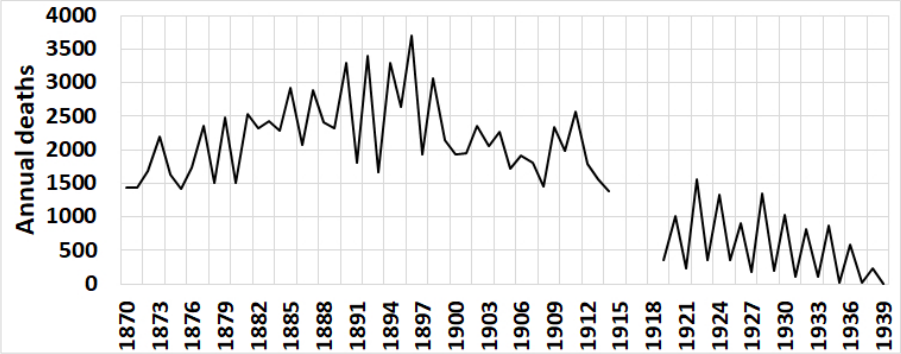
*** p<0.01, ** p<0.05, * p<0.1. Newey-West standard errors allowing correlation across observations up to six weeks apart in parenthesis. Data in Columns 1-3 cover 1875-1914. Data in Columns 4-5 cover 1875-1910 because age categories change in 1911. Data in Columns 5-6 cover 1875-1911 to keep the results comparable to the previous columns. Regressions include controls for temperature, temperature squared, five lags of each of these variables, 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 24. For example, the results in Table 24 show that children aged 1-5 were affected only by acute and not by medium-run pollution exposure. Since measles was a primary driver of mortality in the 1-5 age group and I find no evidence of medium-run effects on measles deaths, this explains 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. This effect is described for measles in Figure 5. The top part of the figure describes the sharp change in measles deaths from around 42 per week from 1870-1914 to under 11 per week from 1919-1939. This decline appears to have been due to a variety of factors, including reductions in overcrowding in dwellings, improved nursing care, better nutrition, and changing disease virulence (Woods, 2000). A measles vaccine was not introduced until after this period. At the bottom of the figure I calculate the impact that this change had on the mortality effects associated with fog events. Using data up to 1914, I estimate that a fog day increased measles deaths by 9% across a four-week period. This relationship, together with the fall in underlying measles prevalence, implies that a fog event in 1919-39 led to 12 fewer deaths than one occurring in the 1870-1914 period. This explains 50.6 percent of the

reduction in the number of deaths associated with a fog day in 1919-39 compared to the pre-WWI period. A similar calculation shows that reductions in TB rates account for an additional 12% of the reduction in deaths due to fog events after WWI.[59]

Figure 5: Effect of falling measles mortality on effect of fog events



	1870-1914	1919-1939	Change
Measles deaths per week:	41.6	10.7	30.9
Measles deaths per fog (over 4 weeks)*:	16.2	4.2	12
Total deaths per fog (all causes, 4 weeks):	74.6	50.3	24.3
Share of overall decline in fog event mortality due to decline in measles prevalence:			0.494

* Note: This row applies the estimated 9.3 percent increase in measles deaths per week due to fog obtained using data up to 1914

7 Implications for today

The results presented above indicate that pollution interacted with two infectious diseases, measles and TB. This finding has implications for modern developing countries where these diseases, particularly TB, remain common. In this section I conduct a back-of-the-envelope calculation in order to get an idea of how many deaths might be attributable to the interaction between TB and pollution in modern developing countries.

The starting point for this analysis is the number of TB deaths by country in 2017 reported by the WHO.[60] Importantly, these TB deaths already reflect the level of available health care in each country, as well as the local disease environment, so my analysis is not

[59]TB deaths fell from 201 per week before 1914 to 83 per week in the inter-war period, most likely due to improved public health measures (the TB vaccine BCG only came into widespread use after WWII.) Applying coefficient estimates from data before 1914 to this change implies that the reduction in TB deaths is associated with a 4.1% reduction in the mortality associated with fog events and accounts for 12.5% of the reduction in mortality due to fog events after WWI.

[60]WHO (2018). I focus on TB because it is more common than measles and country-level TB mortality estimates are readily available.

assuming that the level of health care in, say, India today, is the same as the level during my study period. Rather, the key assumption I needs is that for those with TB in those poor countries where most TB deaths occur, the available health care is not substantially better at protecting them from the additional effect of pollution exposure.

The second element I need is an idea of pollution levels in each country, which also come from the WHO.[61] Using these data, together with the estimated percent increase in TB mortality associated with fog events from Table 5 and the relationship between fog events and pollution levels from Figure 2, I can calculate the number of TB deaths that would be averted in each country if the country's pollution level was lowered to the U.S. level. This simple calculation indicates that the interaction between pollution and TB accounts for over 16,000 deaths globally, including over 8,000 deaths in India each year, 1,600 deaths in Bangladesh, 1,350 in Nigeria, 1,000 in Pakistan, and 600 in China. This is a substantial number of lives lost; as a point of comparison, Chay & Greenstone (2003a) estimate that the landmark Clean Air Act in the U.S. reduced infant mortality by 1,300 in 1972.

My estimates suggest that the interaction of TB and pollution accounts for around 1-3 percent of the total number of TB deaths in high-pollution countries such as India, Pakistan, Bangladesh, and Nigeria, and around 0.7-2.0 percent of the overall number of deaths associated with air pollution in these locations (based on WHO estimates).[62] Clearly eliminating pollution is not a solution to the problem of TB, nor is eliminating TB a solution to the costs of air pollution exposure. After all, these are two of the most important health challenges facing the developing world today. However, the interaction between pollution and TB does appear to be meaningfully raising the health burden of each of these individual factors in a way that magnifies the health costs of pollution in developing settings.

8 Conclusions

This paper provides evidence on the impact of acute pollution exposure associated with fog events 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

[61] WHO (2014).

[62] It is notable these additional deaths will be missed in most calculations of the global burden of pollution, since most of those calculations, such as the one in WHO (2014), do not account for the way that the underlying disease environment increases the health effects of pollution exposure.

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. These diseases, TB and measles, remain among the most important health challenges in modern developing countries. By identifying the interaction of these diseases with air pollution exposure, this study provides a first step toward understanding the mechanisms behind the changing impact of air pollution on health as countries develop.

This study extends our understanding of air pollution effects to incorporate highly polluted environments with substantial infectious disease burdens, exactly the types of developing environments where air pollution is likely to pose the most severe challenge. My results provide some useful lessons informing future work on this topic. One of these is that much of the impact of air pollution exposure may be found in age groups other than infants and the elderly, the populations where, motivated by results from much less polluted settings, much of the attention in existing work has been focused.

References

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

Alsan, Marcella, & Goldin, Claudia. 2019. Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880-1920. *Journal of Political Economy*, **127**(2), 586–638.

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. 2018. Atmospheric Pollution, Health and Height in Late Nineteenth Century Britain. *Journal of Economics History*, **78**(4), 1210–1247.

Ball, Alastair. 2018. Hidden Costs of the Great London Smog: Evidence from Missing Births. *Economics Bulletin*, **38**(4), 1823–1830.

Barmpadimos, I, Keller, J, Oderbolz, D, Hueglin, C, & Prévôt, ASH. 2012. One decade of parallel fine (PM 2.5) and coarse (PM 10–PM 2.5) particulate matter measurements in Europe: trends and variability. *Atmospheric Chemistry and Physics*, **12**(7), 3189–3203.

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.

Barreca, Alan I. 2012. Climate change, humidity, and mortality in the United States. *Journal of Environmental Economics and Management*, **63**(1), 19–34.

Beach, B., & Hanlon, W.W. 2018. Coal Smoke and Mortality in an Early industrial Economy. *Economic Journal*, **128**(615), 2652–2675.

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.

Brook, RD, Rajagopalan, S, Pope, CA, Brook, JR, Bhatnagar, A, Diez-Roux, AV, Holguin, F, Hong, Y, Luepker, RV, Mittleman, MA, & Peters, A. 2010. Particulate Matter Air Pollution and Cardiovascular Disease. *Circulation*, **121**(21), 2331–2378.

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. 2003a. *Air Quality, Infant Mortality, and the Clean Air Act of 1970*. NBER Working Paper No. 10053.

Chay, Kenneth Y., & Greenstone, Michael. 2003b. 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 China's Huai River policy. *Proceedings of the National Academy of Sciences*, **110**(32), 12936–12941.

Clay, K, Lewis, J., & Severnini, E. 2018. Pollution, Infectious Disease, and Infant Mortality: Evidence from the 1918-1919 Spanish Influenza Pandemic. *Journal of Economics History*, **78**(4), 1179–1209.

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, J., Graff Ziven, J., Mullins, Jamie, & Neidell, M. 2014. What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution? *Annual Review of Resource Economics*, **6**(1), 217–247.

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, & Walker, Reed. 2011. Traffic Congestion and Infant Health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, **3**(1), 65–90.

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.

Deryugina, T., Huetel, G., Miller, N.H., Molitor, D., & Reif, J. Forthcoming. The Mortality and Medical Costs of Air Pollution: Evidence from changes in Wind Direction. *American Economic Review*.

Deschênes, Olivier, & Greenstone, Michael. 2011. Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *American Economic Journal: Applied Economics*, **3**(4), 152–185.

Feng, Xinyuan, & Wang, Shigong. 2012. Influence of different weather events on concentrations of particulate matter with different sizes in Lanzhou, China. *Journal of Environmental Sciences*, **24**(4), 665–674.

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.

Gonzalez-Viveros, P, Caetano, E, & Garcia-Garcia, F. 2018. Characterization and Modeling of Fog in the Mexico Basin. *Aerosol and Air Quality Research*, **18**, 79–90.

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.

Greenstone, Michael, & Jack, B. Kelsey. 2015. Envirodevonomics: A Research Agenda for a Young Field. *Journal of Economic Literature*, **53**(1), 5–42.

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

Halliday, TJ, Lynham, J, & de Paula, A. Forthcoming. Vog: Using Volcanic Eruptions to Estimate the

Health Costs of Particulates. *Economic Journal*.

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.

Heblich, Stephan, Trew, Andrew, & Zylberberg, Yanos. 2016 (June). *East Side Story: Historical Pollution and Neighborhood Segregation*. Mimeo.

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

Jia, R., & Ku, H. Forthcoming. Is China's Pollution the Culprit for the Choking of South Korea? Evidence from the Asian Dust. *Economic Journal*.

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 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.

Rajagopalan, Sanjay, Al-Kindi, Sadeer G, & Brook, Robert D. 2018. Air pollution and cardiovascular disease: JACC state-of-the-art review. *Journal of the American College of Cardiology*, **72**(17), 2054–2070.

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).

Seinfeld, John H., & Pandis, Spyros N. 2016. *Atmospheric Chemistry and Physics : From Air Pollution to Climate Change*. John Wiley & Sons.

Siddika, Nazeeba, Balogun, Hamudat A, Amegah, Adeladza K, & Jaakkola, Jouni JK. 2016. Prenatal ambient air pollution exposure and the risk of stillbirth: systematic review and meta-analysis of the empirical evidence. *Occup Environ Med*, **73**(9), 573–581.

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.

WHO. 2014. *Burden of Disease from Ambient Air Pollution for 2012*. Tech. rept. World Health Organization, Geneva.

WHO. 2018. *Global Tuberculosis Report*. Tech. rept. World Health Organization.

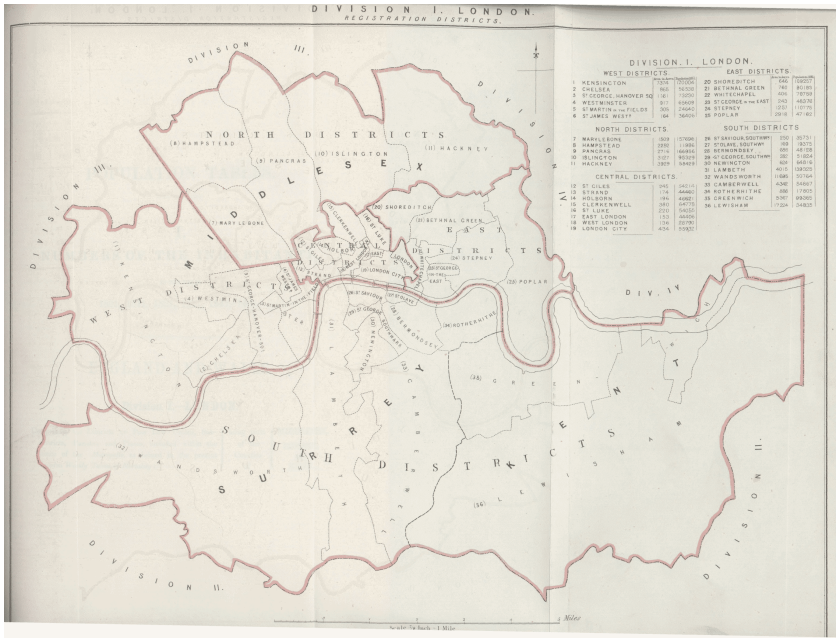
Woods, Robert. 2000. *The Demography of Victorian England and Wales*. Cambridge, UK: Cambridge University Press.

A Appendix

A.1 Population and mortality in London

The area of London covered by this study is depicted in Figure 6. 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, roughly covers the central districts of Greater London today.

Figure 6: Area of London covered by this study

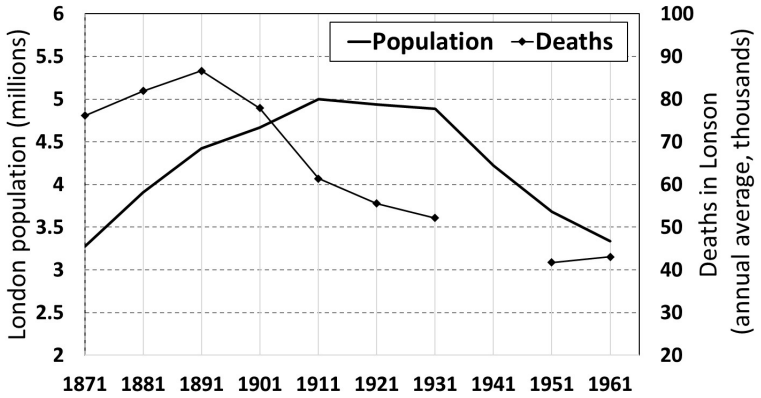


This map is from the 1851 Census of Population.

Figure 7 plots the population of the area of London covered by this study. The 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 14. Figure 8 plots the raw death rate in London (without age adjustment) implied by the number of

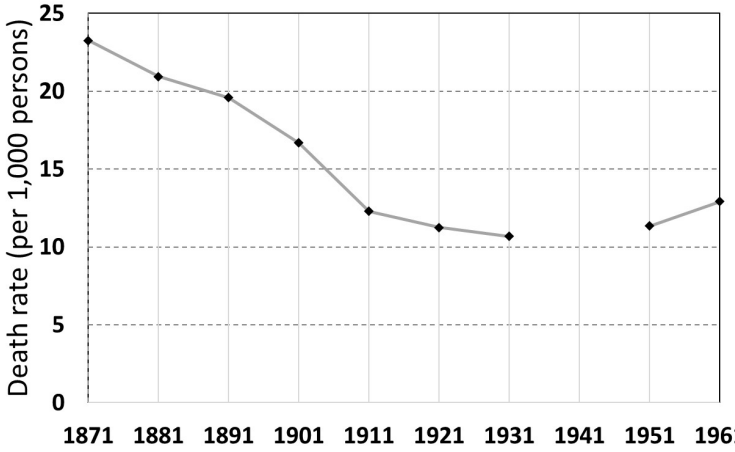
deaths and the population data. The death rate fell by more than half from the beginning of the study period into the inter-war period. If anything this understates the true decline because of the effect of population aging after the demographic transition, which began in Britain in the 1870s.

Figure 7: 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 8: 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 7 presents summary statistics for the data used in the main analysis, covering 1866-1965 (except 1915-18 and 1940-48).

Table 7: 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 9 and 10 present examples of weather reports from 1880 and 1952 for weeks in which heavy fog occurred.

Figure 9: 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.

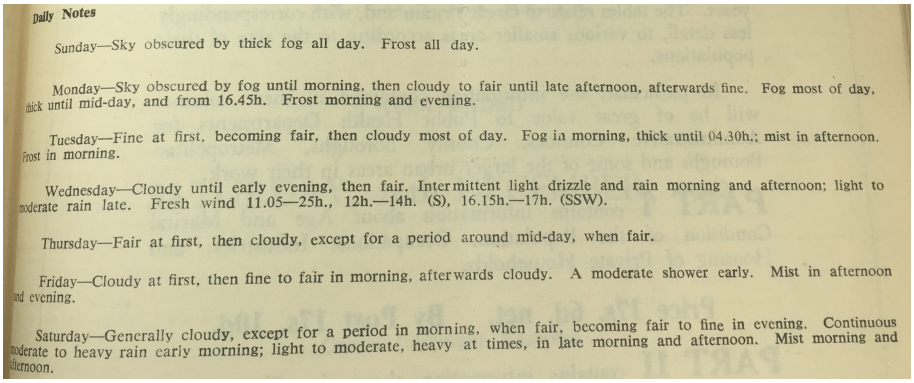
Overcast.

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 10: 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 11: Fog weeks and fog days, 1865-1965

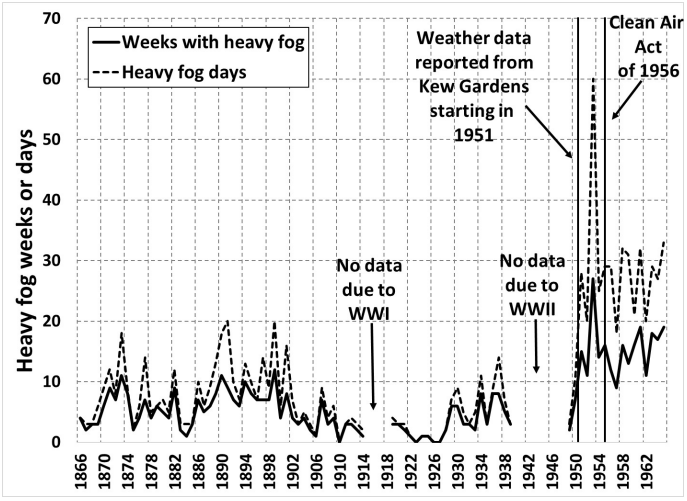
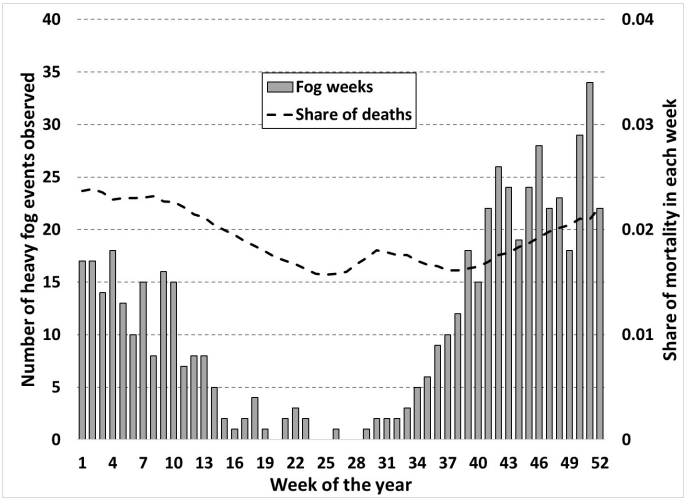


Figure 12: Fog events and total mortality share by week of the year



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 13 provides an overview of how these diseases are grouped to obtain the categories used in the analysis.

Figure 13: 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

Measles

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

Old age

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

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

Venereal diseases
Syphilis
Other venereal diseases

Whooping cough

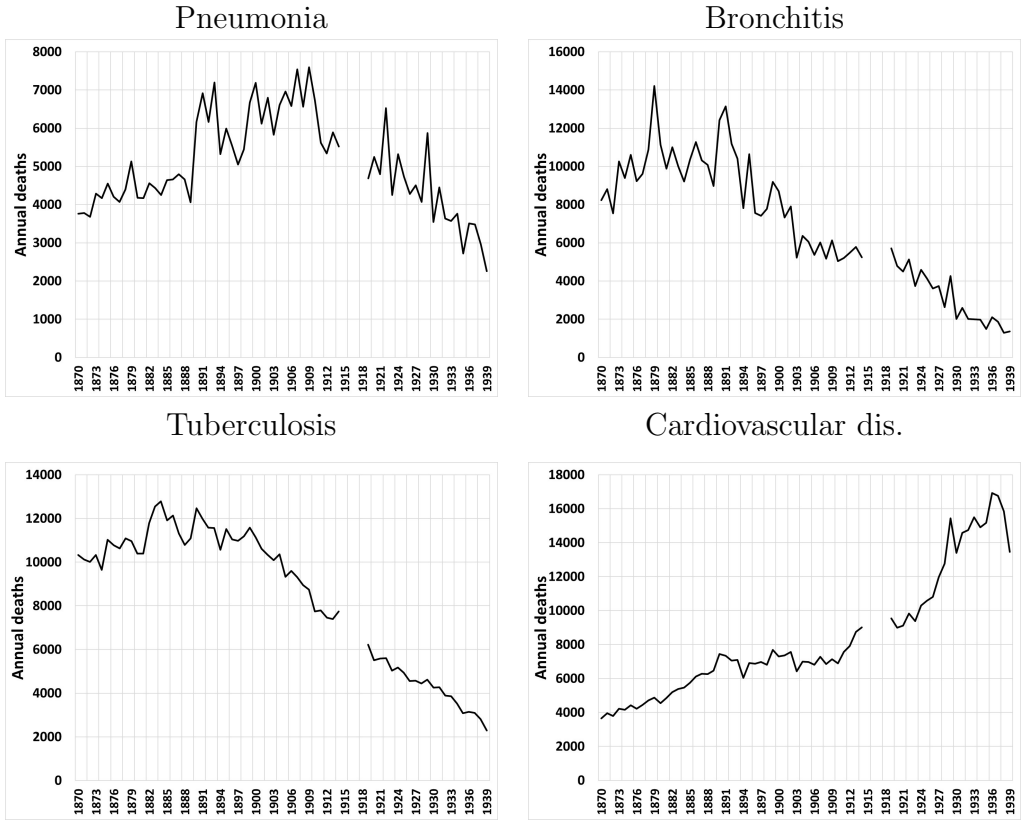
Table 8 describes the share of mortality in each age category accounted for by each cause of death.

Table 8: 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%	13.23%	14.18%
PNEUMONIA	8.41%	14.31%	5.37%	6.14%	6.41%	4.46%
RESPIRATORY	1.99%	4.02%	3.08%	2.66%	3.82%	4.02%
DIGESTIVE	17.30%	6.44%	7.85%	5.82%	3.95%	3.87%
MEASLES	2.53%	14.75%	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%
WHOOPING COUGH	4.48%	10.61%	1.52%	0.01%	0.00%	0.00%
OTHER INFECTIOUS	1.03%	6.91%	8.67%	1.48%	1.18%	0.78%
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%
FETAL UNDERDEVELOPMENT	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

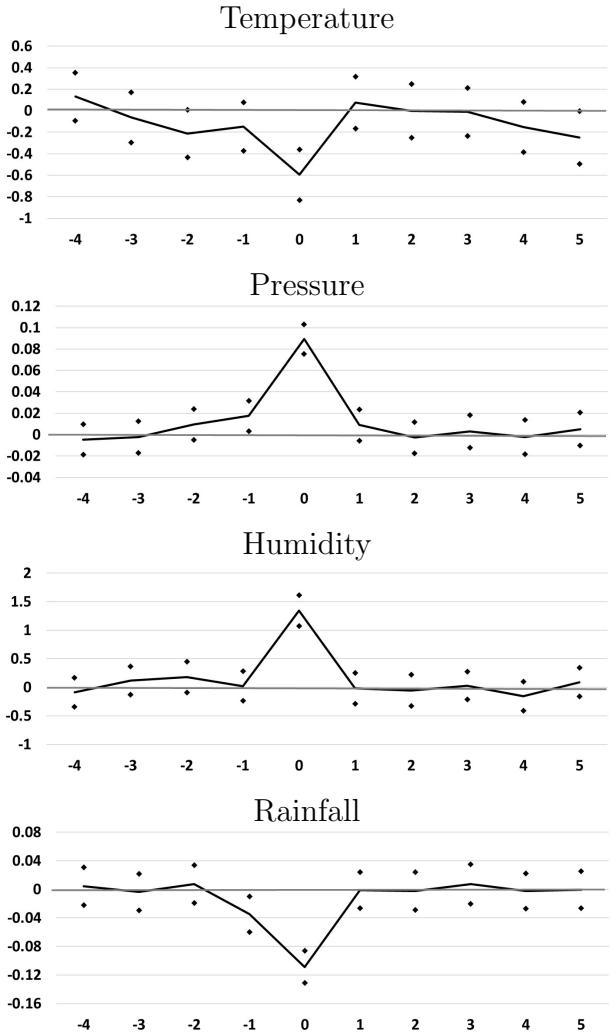
Figure 14: Deaths in London in select infectious disease categories



A.2.6 Relationship between fog events and other weather conditions

Figure 15 presents regressions of the available weather variables–temperature, pressure, humidity and rainfall–on fog events. This is useful because it allows us to look at whether there appear to be trends in weather in the weeks leading up to fog events, which may help explain the downward trend in mortality observed in Figure 3.

Figure 15: Estimated relationship between fog events and leads and lags of weather variables



Coefficient estimates and confidence intervals for a regression of weather variables 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 regression includes controls for a full set of year and week-of-the-year by decade effect. Data cover 1866-1965. N=4,479.

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.[63] 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 7 provides summary statistics for the available climate variables using data from all of the years covered by this study.

Figure 16 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 17 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.

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

Figure 16: Time-series graphs of climate variables

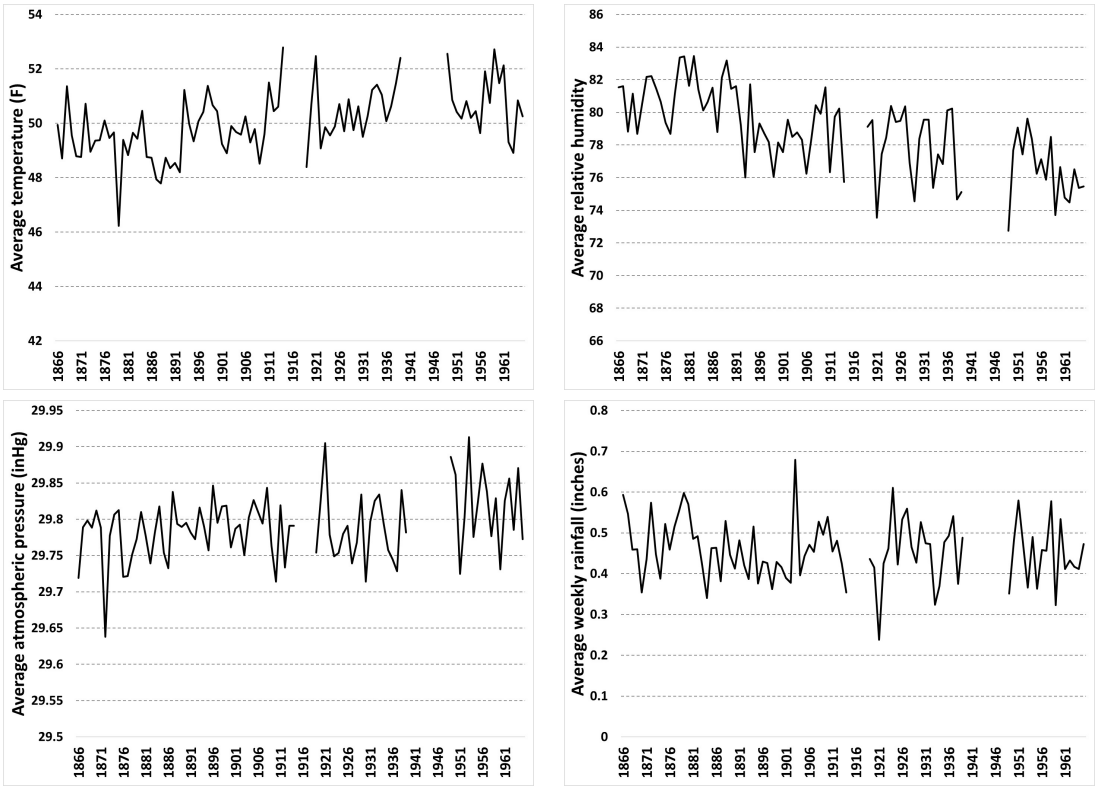
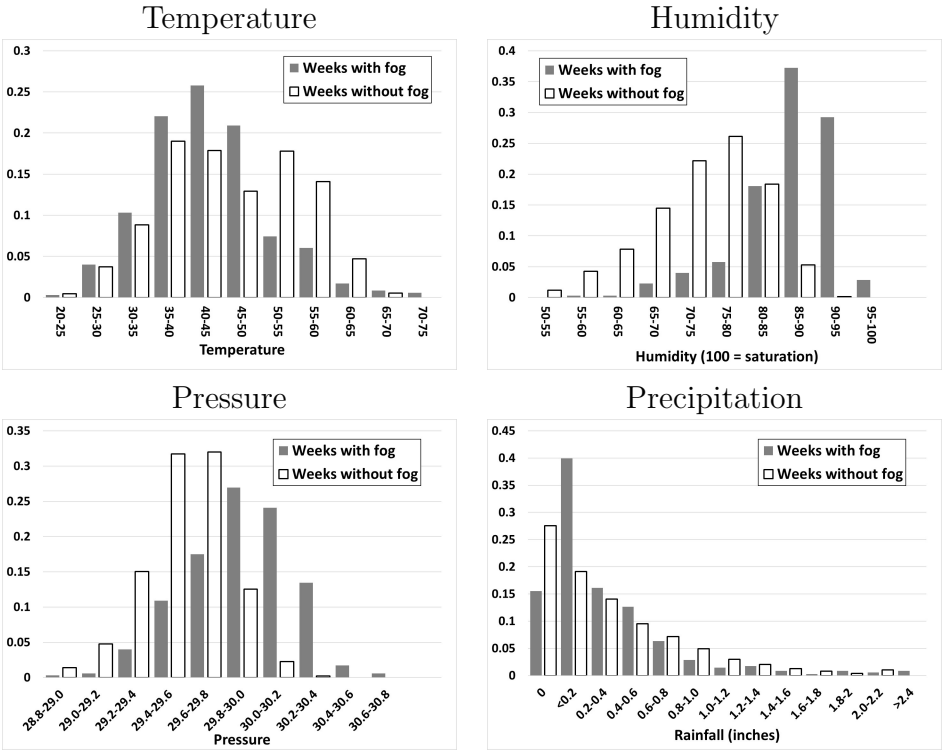


Figure 17: Histograms of climate variables in weeks with and without fog



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, heavy 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:

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

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

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

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

The interaction of 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 heavy fog. However, it is still necessary to determine appropriate cutoff points for each variable. To choose the appropriate cutoff values, I conducted a grid search over possible combinations of values and then ran regression using each combination to predict the timing of fog formation. As decision criteria I consider the increase in R-squared generated by adding each $PredFOG$ variable to the regression as well as the t-statistic on the estimated coefficient for that variable. Along both criteria, the cutoff values that performed the best were: $\tilde{T} = 40$, $\tilde{P} = 29.9$, $\tilde{H} = 82.5$ and $\tilde{R} = 0.5$. Thus, these are the values used in the main analysis.

Table 9 describes the relationship between the $PredFOG$ variable obtained using these cutoff values and the timing of fog events across the full study period as well as several sub-periods. Table 10 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 a few predicted fog weeks when no heavy fog event was reported (type II errors) and many heavy fog events when no fog event was predicted.

Table 9: 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	0.0488	0.0218	0.0714	0.0416	0.313
	(0.0689)	(0.0920)	(0.117)	(0.110)	(0.297)
PredFOG t+3	-0.00422	-0.0528	0.290**	-0.102	-0.193
	(0.0686)	(0.109)	(0.138)	(0.101)	(0.239)
PredFOG t+2	0.0274	0.0457	-0.143	-0.0508	0.361
	(0.0746)	(0.0899)	(0.154)	(0.124)	(0.257)
PredFOG t+1	-0.134**	-0.147*	-0.184	-0.0178	-0.479*
	(0.0681)	(0.0852)	(0.136)	(0.112)	(0.270)
PredFOG t	0.603***	0.286**	0.643***	0.778***	0.923***
	(0.0991)	(0.123)	(0.189)	(0.159)	(0.352)
PredFOG t-1	0.0823	0.108	0.337**	-0.126	-0.0685
	(0.0680)	(0.0992)	(0.151)	(0.103)	(0.232)
PredFOG t-2	-0.0679	-0.00145	0.0821	-0.0222	-0.400*
	(0.0624)	(0.0947)	(0.137)	(0.0917)	(0.220)
PredFOG t-3	0.0315	0.101	3.02e-05	-0.0275	0.0678
	(0.0639)	(0.104)	(0.106)	(0.101)	(0.255)
PredFOG t-4	-0.0540	-0.00934	-0.108	-0.0711	-0.165
	(0.0579)	(0.0870)	(0.0925)	(0.0948)	(0.242)
PredFOG t-5	0.0486	0.178*	-0.0542	-0.0442	0.0892
	(0.0589)	(0.0920)	(0.118)	(0.0795)	(0.226)
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 40 degrees f. Pressure cutoff is 29.9. Relative humidity cutoff is 82.5. Precipitation cutoff is 0.5 inches per week. Regressions also include a full set of year and week-of-the-year by decade effects.

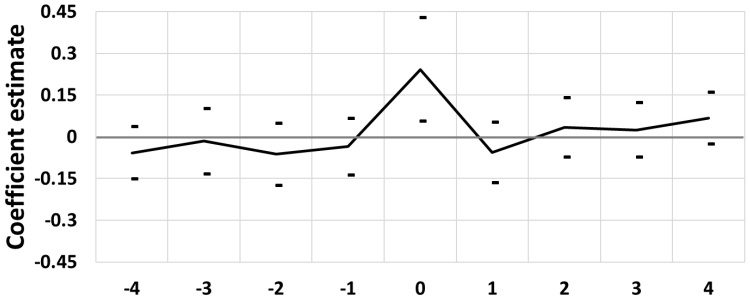
Figure 18 shows that the *PredFOG* variable also does a good job of predicting increases in actual pollution levels, using the direct pollution measures available from 1951-1961.

Table 10: 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	145	86	491	59
1866-1889	125	54	26	99	28
1890-1914	132	37	21	111	16
1919-1939	82	32	22	60	10
1951-1965	229	22	17	212	5

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 18: 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 1.56, suggesting that serial correlation is not likely to be an important concern.

As a robustness check, it is useful to consider some alternative combinations of cutoff values. Table 11 presents results using four alternative sets of cutoff values to predict fog event weeks. All of these alternatives can predict the occurrence of heavy fog weeks.

Table 12 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 toward 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 19 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

Table 11: 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.

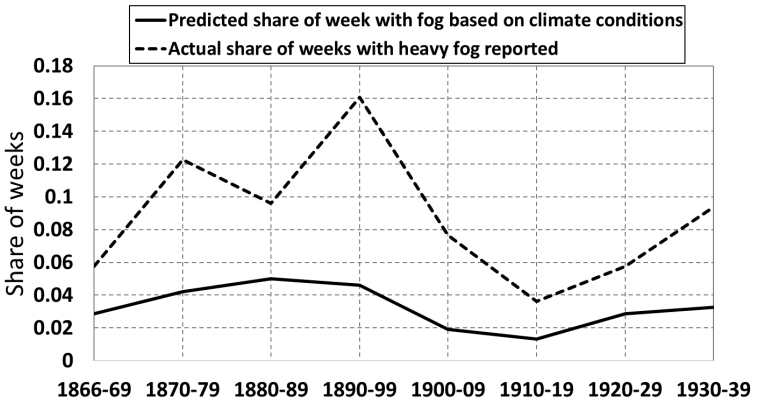
Table 12: Alternative fog event predictors and mortality using four week windows

	DV: Log total mortality			
	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.

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 16). 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.

Figure 19: 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 < 40, pressure >29.9, humidity > 82.5, precipitation < 0.5.

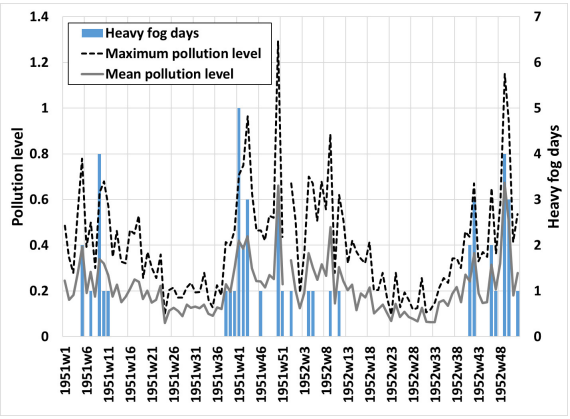
A.4 Preliminary analysis appendix

A.4.1 Graphs comparing pollution and fog events

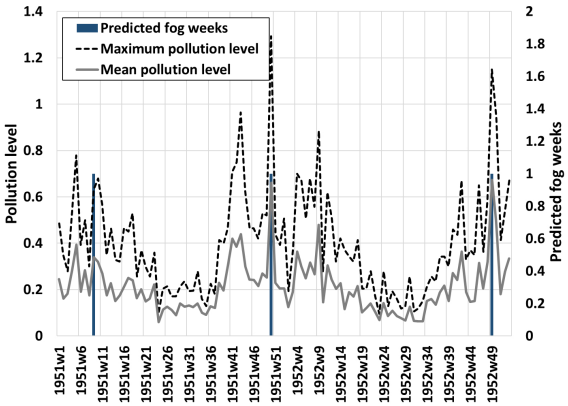
The left-hand panel of Figure 20 plots pollution levels against fog events for 1951-52, the first two years for which pollution levels are reported. This graph shows a clear correspondence between fog events and pollution levels, with the highest pollution levels obtained during major fog events. It 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. The right-hand panel 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. 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 40 degrees, humidity above 82.5, pressure above 29.9 and precipitation below 0.5.

A.5 Appendix to analysis of total mortality

A.5.1 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.5.2 Results using alternative treatment windows

Table 13 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 13: Estimates using effect windows of different lengths

	DV: Log total mortality			
	(1)	(2)	(3)	(4)
Fog days (2 week window)	0.0126***			
	(0.00298)			
Fog days (3 week window)		0.0105***		
		(0.00252)		
Fog days (4-week window)			0.00835***	
			(0.00232)	
Fog days (5 week window)				0.00709***
				(0.00228)
Implied deaths	29,472	36,802	38,980	41,347
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 humidity, humidity squared, pressure, pressure squared, precipitation, precipitation squared, temperature, temperature squared, and five lags of temperature and temperature squared.

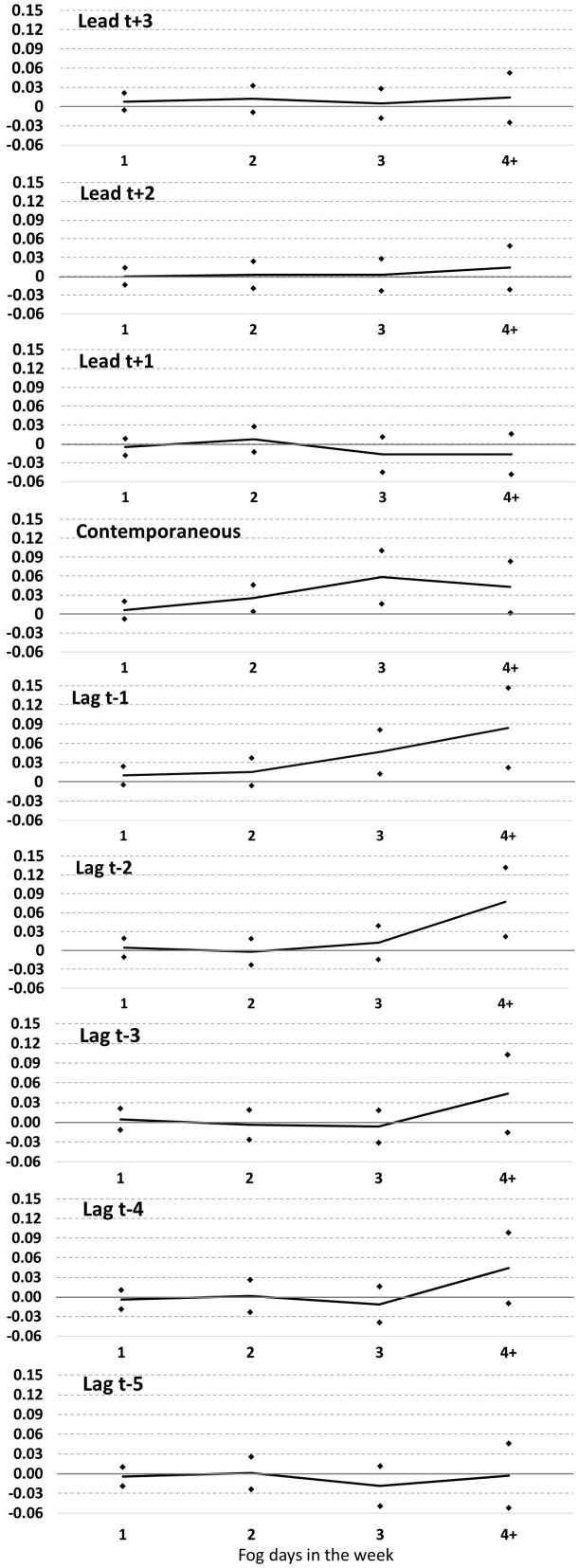
A.5.3 Non-parametric estimates of fog effects

This appendix presents estimates of fog effects where I bin weeks based on the number of fog days and then estimate separate effects, including leads and lags, for each level of fog days in a week. This is useful for looking at whether the effects are non-linear in the number of fog days in a week. I estimate results for weeks with one, two, three, or four or more fog days. The latter category is used because there are not enough weeks with five, six, or seven fog days to obtain separate estimates for that groups.

The estimated effect of fog days for up to three leads and five lags are presented in Figure 22. Figure 22 presents coefficient estimates and confidence intervals for a regression of log total mortality on fog days, with separate effects estimated for weeks with one, two, three, or four or more fog days. The number of fog days in a week is on the x-axis while the y-axis documents the size of the estimated coefficients. 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, 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 data cover 1866-1965 and N=4,479.

The first three panels present leading effects, i.e., the estimated relationship between fog days in a week and mortality in previous weeks. These estimates are close to zero and there is no clear gradient with the number of fog days in a week. This suggests that the identification strategy is working well. The fourth panel presents the effects of fog days on the weeks in which they occurred, while the lower five panels estimate effects in the following weeks. The contemporaneous and one-week lagged effects show a fairly linear increase in mortality associated with the number of fog days in a week. Interestingly, for later lags, this linearity disappears and we see that it is only the worst fog weeks, those with four or more days of fog reported, that exhibit these longer-lagged effects on mortality. By the fifth week after the fog event, even this longer-lagged effect weakens.

Figure 22: Estimated effect by number of fog days in a week



See previous page for table notes.

A.5.4 Total mortality analysis robustness results

Table 14 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 including a control for log births. Including 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.

Table 14: Additional total mortality regression results using four-week windows

	DV: Log total mortality			
	With log births control	Temperature quadratics	Fog days squared	Lagged weather vars.
	(1)	(2)	(3)	(4)
Fog days (4-week window)	0.00845***	0.00829***	0.0118***	0.00653***
	(0.00234)	(0.00229)	(0.00407)	(0.00239)
Fog days squared (4-week window)			-0.000483	
			(0.000480)	
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.

The results in Table 15 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 about 18%.

Table 15: Results interacting temperature and fog days

DV: Log total mortality	
Fog days (4-week window)	0.0427**
	(0.0206)
Fog days × mean avg. weekly temp	-0.000752*
(4-week window)	(0.000417)
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, pressure, pressure squared, humidity, humidity squared, precipitation, and precipitation squared.

The next set of results allow more flexible controls for temperature and humidity. Existing work by Deschênes & Greenstone (2011) has highlighted the fact that extreme temperatures can have large effects while Barreca (2012) highlights the non-linear impact of humidity on mortality. We may therefore worry that the simple quadratic controls included in the main specification are not sufficiently flexible to deal with these potential sources of bias, though it is important to note that London has a much milder climate than most of the U.S., the subject of those existing papers.

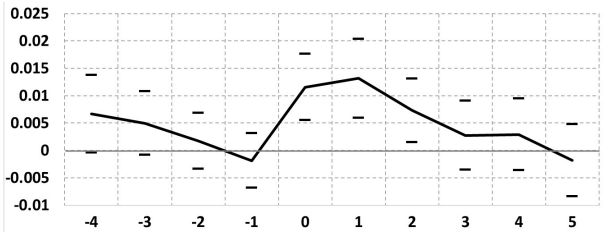
The graphs in Figure 23 examine the impact of including more flexible controls for temperature and (absolute) humidity on the estimated impact of fog events on mortality. An important thing to note here is that, following Barreca (2012), I focus on absolute humidity in these exercises.

As a point of comparison, the top figure reproduces results obtained with my baseline specification, which includes as controls contemporaneous temperature, relative humidity, rainfall and pressure, the square of each of those variables, as well as a full set of leads and lags of the temperature and temperature squared variables. In Panel B, I instead bin weeks into eight temperature categories and include a full set of leads and lags for each of these temperature bin indicators. In Panel C, I also bin weeks into six absolute humidity categories and include a full set of leads and lags of those humidity bin indicators (together

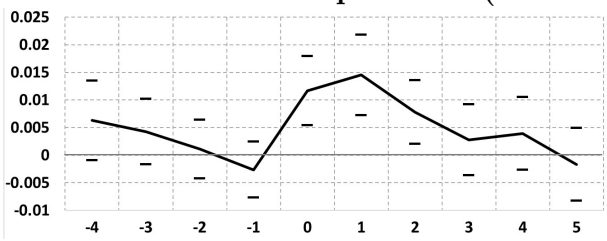
with the temperature bin indicators). The similarity of the results across these alternative specifications is clear.

Figure 23: Effect of fog events with more flexible weather controls

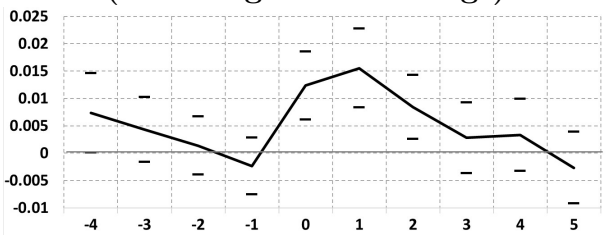
Baseline: With quadratic controls for temperature (including leads and lags)



Panel B: With binned controls for temperature (including leads and lags)



Panel C: With binned controls for temperature and absolute humidity (including leads and lags)



Coefficient estimates and confidence intervals for a regression of log deaths 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. All regressions include controls for pressure, pressure squared, rainfall, and rainfall squared as well as year and week-of-the-month by decade fixed effects. Data cover 1866-1965. N=4,479.

Table 16 makes it easier to assess how the inclusion of this large set of additional temperature and absolute humidity controls impacts the magnitude of the effect of fog days. For comparison purposes, the first column reproduces the results obtained using my main analysis approach. In Column 2, I replace the quadratic temperature controls with controls for eight temperature bins, including leads and lags of each temperature bin variable. This slightly increases the magnitude of the estimated coefficient, though the estimate is not statistically distinguishable from the coefficient in Column 1. In Column 3, I also include more

flexible controls for absolute humidity in place of the quadratic relative humidity controls included in the main specification. Again, the estimated coefficient is similar to, if slight larger than, the estimate in Column 1, but the difference is not statistically distinguishable.

The take-away from all of these results is that including this very rich set of controls (130 additional variables in Column 3 of Table 16) has only a modest impact on the results. Given this, in my main analysis I opt for the more parsimonious specification. One reason for this choice is that the inclusion of the much richer set of binned temperature and humidity control variables leads to concern about power in regressions where I am breaking the data down by period, age group, or cause of death.

Table 16: Estimates with more flexible temperature and humidity controls

	DV: Log total mortality		
	Baseline specification	Binned temperature controls	Binned temp. and absolute humidity controls
	(1)	(2)	(3)
Fog days	0.00835***	0.00885***	0.00950***
(4 week window)	(0.00232)	(0.00235)	(0.00238)
Observations	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 as well as humidity, humidity squared, pressure, pressure squared, precipitation, and precipitation squared. Column 1 includes controls for temperature, temperature squared, and four leads and five lags of these variables. Column 2 replaces those with binned temperature controls reflecting eight levels of temperature, including four leads and five lags of each. Column 3 replaces the relative humidity and humidity squared variables with controls for six bins of absolute humidity, as well as four leads and five lags of those controls.

Finally, it is worth noting that analysis of the relationship between the temperature bins and total mortality also shows reasonable results. In particular, we see no relationship between temperature in a week and mortality in previous weeks, which suggests that the identification strategy is working well. We see that both unusually high and unusually low temperatures are associated with increased mortality in the week in which they occur. There is also evidence that these effects persist for several weeks. This persistence makes sense given that many of the deaths associated with temperature are due to infectious diseases, which may continue to spread and cause mortality in following weeks.

A.6 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. These results are consistent with modern public health studies that find that pollution increases stillbirths (see Siddika *et al.* (2016) for a review of this literature).

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.

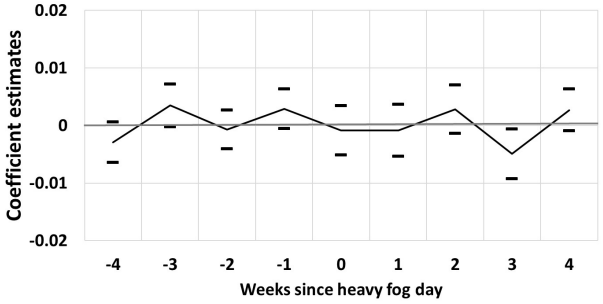
Table 17: Effects on infant mortality accounting for stillbirths

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. 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 24, 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 events. 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 24: 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 1.94), 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. N=4,479.

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.7 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: 1866-1899, 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. 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.

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 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. In order to maintain sufficient power, these results focus on longer time periods before or after WWI. The IV results also show evidence that the impact of fog events was decreasing over time. These 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 because 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			
	1866-1899	**1900-1939**	**1951-1965**
Fog days (4-week window)	0.0115***	0.0106**	0.00700**
	(0.00379)	(0.00525)	(0.00283)
No weeks:	1,763	1,835	767
No. of fog days	304	179	429
Avg. fog days per week	0.172	0.098	0.552
Total implied deaths	21,885	8,936	9,731
Share of all deaths due to fog	0.00795	0.00416	0.0155
B. Using *predicted* fog weeks as an instrument for fog days			
	1866-1914	**1918-1965**	
Fog days (4-week window)	0.0439***	0.0159*	
	(0.0148)	(0.00899)	
No. pred. fog weeks	91	54	
First-stage F-stat	128.65	85.28	
Observations	2,532	1,932	

*** 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 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.[64] The

[64]It is not realistic to do this with the coefficients estimated in the bottom panel. This is because if 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 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.[65]

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	-0.000975	0.0212**	0.00937**	0.00629*	0.0112***	0.0132***
(4-week window)	(0.00492)	(0.00868)	(0.00403)	(0.00358)	(0.00365)	(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	0.00643	0.0260*	0.000117	0.00475	0.0133**	0.0120*
(4-week window)	(0.00670)	(0.0136)	(0.00448)	(0.00602)	(0.00546)	(0.00705)
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, 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

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.

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

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 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**	-5.54e-05	-0.00065*	-4.88e-06	0.00046
(4-week window)	(0.0007)	(0.00096)	(0.00027)	(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	-0.00054	0.00035	-0.00043**	-0.00063	0.00063	0.00062
(4-week window)	(0.00097)	(0.00062)	(0.0002)	(0.0004)	(0.00044)	(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, 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.8 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 (see Currie (2013) for a review of this literature). 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 fetal underdevelopment, which generally occurred soon after birth. In fact, this series is called prematurity in the data, though this description is probably somewhat misleading given the difficulty of identifying term length in the past, so I use the terminology fetal underdevelopment instead. By using the week of death as a proxy for the week of birth I am able to roughly identify the period in which each child who died of fetal underdevelopment 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 deaths due to fetal underdevelopment 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 fetal underdevelopment. One likely channel is through prematurity, which has been linked to pollution exposure by a number of studies, including Currie & Walker (2011), though that study does not examine how the effect of exposure varies across trimester. 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. These results are consistent with modern public health studies. A recent meta-analysis by Siddika *et al.* (2016) suggests that pollution has substantial effects on stillbirths, and particularly exposure in the first and third trimesters. Note also that for both series, fog events more than 36 weeks in the past, as well as those in the period roughly corresponding to the second trimester, did 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 fetal underdevelopment 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 fetal underdevelopment by about 1.23%. Fetal underdevelopment 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 fetal underdevelopment because of in utero exposure to fog events account for 1.6 out of every 1000 infant deaths during the years cov-

ered 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 fetal underdevelopment.

During the period covered by the stillbirth 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.9 Medium-run effects by age group

Table 24 reports medium-run effects broken down by age group. The clearest pattern here is that older adults suffer from both the acute and medium-run effects of fog days. For ages 0-1, while I do not observe strong acute effects for infants (ignoring stillbirths), there is some evidence of medium-run impacts. The medium-run impacts are not statistically significant, though they are significant when log births is included as a control variable. This medium-run effect may reflect at least in part in utero exposure, as suggested by the analysis in Appendix A.8.

For children aged 1-5, I find no evidence of medium-run effects. This is consistent with results showing that most of the acute effect on this age group occurred through interactions with the measles. It is somewhat surprising to see that the estimated acute effect on young adults disappears when controlling for the medium-run effects, though we have to be somewhat careful with that group because relatively few deaths occur at these ages, making estimates somewhat fragile.

For prime-age and older adults, the estimated acute effect of pollution exposure is relatively unchanged when the medium-run effect is also included in the regression, though for prime-age adults the coefficient is no longer statistically significant.

Table 24: Medium-run results by age group

	DV: Log mortality					
Age group:	0-1	1-5	Young	Adult	Middle age	Elderly
Fog days	-0.00288	0.0126	-0.000136	0.00536	0.0137***	0.0136***
(4 week window)	(0.00524)	(0.00864)	(0.00554)	(0.00432)	(0.00391)	(0.00440)
Fog days in	0.00298	-0.00169	-0.000531	0.00119	0.00277**	0.00372***
medium run	(0.00189)	(0.00279)	(0.00183)	(0.00150)	(0.00115)	(0.00138)
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, 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.

A.10 Additional cause of death results

Table 25 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 25: Cause of death results with Newey-West standard errors

Cause of death	Coefficient		S.E.
Bronchitis	0.0349	***	(0.00692)
Measles	0.0796	***	(0.0219)
Pneumonia	0.0298	***	(0.00592)
Respiratory, other	0.00447		(0.0107)
Scarlet Fever	0.0351	***	(0.0106)
Smallpox	-0.00398		(0.0474)
Whooping cough	-0.00019		(0.0151)

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, pressure, pressure squared, humidity, humidity squared, precipitation, precipitation squared and a full set of year and week-of-the-year by decade effects.

Table 26 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 26 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 or other infectious diseases, 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 26: 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	0.0132	0.0000	22.87
Respiratory & Cardiovascular			
Bronchitis	0.0392	0.0000	42.58
Pneumonia	0.0271	0.0000	22.88
Cardiovascular	0.0086	0.0110	6.47
Other respiratory	0.0111	0.2301	1.44
Infectious diseases			
Digestive	0.0062	0.2353	1.41
Measles	0.0350	0.0709	3.26
Scarlet fever	0.0066	0.5552	0.35
Smallpox	0.0511	0.2345	1.42
Tuberculosis	0.0087	0.0033	8.66
Whooping cough	0.0063	0.6527	0.20
Other diseases			
Cancer	-0.0016	0.6711	0.18
Neurological	0.0023	0.4470	0.58
Old age	0.0112	0.0454	4.01
Fetal underdevelopment	0.0071	0.2649	1.24
Venereal diseases	0.0101	0.3814	0.77
Other infectious dis.	-0.0023	0.7161	0.13
Other causes of death			
Accidents & violence	0.0003	0.9494	0.00
Alcoholism	0.0178	0.2003	1.64
Homicide†	0.0098	0.5754	0.31
Suicide	0.0039	0.7530	0.10
Other misc. CODs	0.0011	0.6265	0.24

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.

In the next set of results, I focus on the causes of death thought to be associated with pollution effects, such as bronchitis, as well as infectious diseases, while including as controls

those causes not thought to be associated with the acute effects of pollution. Including these as controls means my regressions are essentially mimicking a difference-in-difference framework, with causes of death not thought to be associated with pollution effects treated as the control group. This provides an alternative to treating these causes as a placebo, as was done in the main text.

I consider two alternatives for causes not likely to be directly affected by pollution. First, I follow Jia & Ku (Forthcoming) and focus on cancer as a control category ("Cancer control"), since cancer deaths are unlikely to be related to *acute* pollution effects. Second, I consider a broader basket of control diseases ("Broader control") including cancer as well as digestive, neurological, and venereal diseases and deaths due to accidents, violence and suicide.

The results are presented in Table 27. These results look very similar to the results presented in the main text. That is not surprising given that I found that, as expected, the control diseases did not show elevated mortality in weeks in which fog events occurred.

Table 27: Mortality effects by reported cause of death with COD controls

DV: Log all-age mortality within disease category						
	Cancer control			Broader control		
Cause of death	Coefficient		S.E.	Coefficient		S.E.
Respiratory & Cardiovascular						
Bronchitis	0.0349	***	(0.00448)	0.0360	***	(0.00438)
Cardiovascular	0.00835	***	(0.00241)	0.00914	***	(0.00238)
Pneumonia	0.0297	***	(0.00403)	0.0311	***	(0.00393)
Misc. respiratory	0.00423		(0.00686)	0.00626		(0.00677)
Infectious diseases						
Measles	0.0796	***	(0.0137)	0.0811	***	(0.0136)
Scarlet Fever	0.0351	***	(0.00806)	0.0350	***	(0.00807)
Smallpox	-0.00301		(0.0320)	-0.00389		(0.0321)
Tuberculosis	0.00610	***	(0.00214)	0.00681	***	(0.00209)
Whooping cough	-0.000250		(0.0101)	0.00116		(0.0101)
Infectious, other	0.0137	***	(0.00464)	0.0143	***	(0.00462)
All causes	0.00973	***	(0.00212)	0.0119	***	(0.00186)

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. Results in the left columns include as a control log deaths due to cancer in each week. Results in the right columns include as a control log deaths due to cancer, neurological disorders, digestive diseases, venereal diseases, accidence, violence and suicide in each week. All regressions include controls for temperature, temperature squared, five lags of each of these variables, pressure, pressure squared, precipitation, precipitation squared, humidity, humidity squared, and a full set of year and week-of-the-year by decade effects. Data cover 1870 to 1939.

A.11 Impact of heavy rain

This section takes an alternative approach to examining the health effects of pollution, using heavy rain as a source of exogenous variation in pollution levels. Rain is useful because it is similar to fog in a number of respects. It can disrupt or slow down transportation, cause people to stay indoors, and is associated with damp weather. However, rain differs from fog in one crucial way. Unlike fog, which increases pollution levels, rain decreases pollution. This occurs because, as raindrops fall, along the way they pick up particulate matter, removing it from the air.[66] This washing effect has been documented in a number of modern studies.[67] This means that studying the mortality effects of rainfall can provide a check on whether the mortality effects associated with fog events are driven by pollution or other factors. If the fog effects documented in the main text were driven by pollution, then we should observe the opposite effects for rainfall. However, if the fog effects were driven by other factors, such as difficulty in getting to medical care, or the transmission of infectious diseases because people stayed indoors, then the effect of rainfall should be similar to the effects estimated for fog.

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 controls for rainfall (of course), as well as pressure and humidity. It is important to omit the pressure and humidity variables because these are closely associated with rainfall, so if they are included as controls they absorb much of the relationship between rainfall and pollution.[68]

As a starting point it is useful to establish that the rainfall washing effect is present in the setting I consider. Figure 25 presents estimates showing the response of pollution to rain using the period from 1951-1962 when consistent pollution data are available. In the top panel, the key explanatory variable is an indicator for heavy rain in a week (>0.5 inches) while in the bottom panel I use a continuous rainfall measure. This figure shows that rainfall has a strong negative effect on pollution in the week in which it occurs. No effect is observed in the following weeks. There is some limited and not statistically significant evidence that

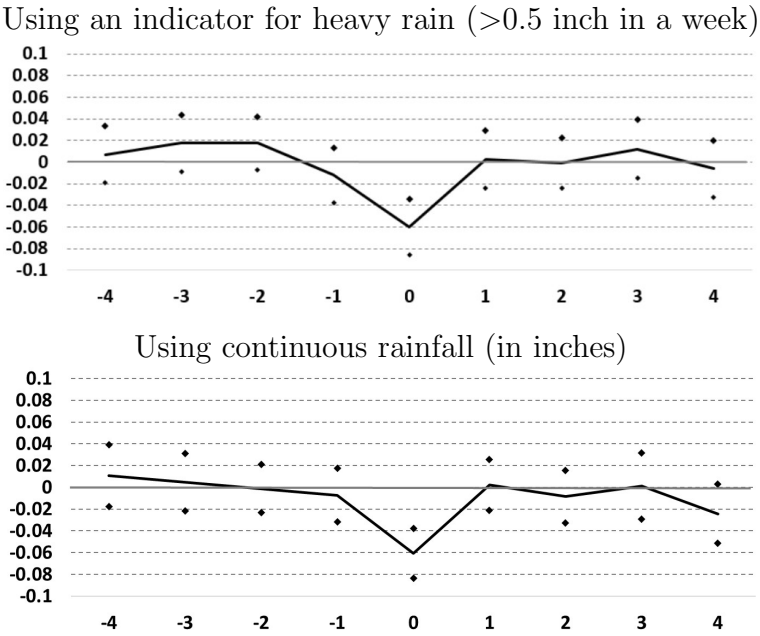
[66] Seinfeld & Pandis (2016).

[67] See, e.g., Barmpadimos *et al.* (2012) and Feng & Wang (2012).

[68] In a previous version of the paper, I reported results suggesting that rainfall had no statistically significant relationship with pollution. This was due to the inclusion of the pressure and humidity variables, which are closely related to rainfall (particularly humidity). These soaked up most of the relationship between pollution and rainfall when using the simple 0.5-inch cut-off for heavy rain days.

pollution is lower in the week just before a heavy rain when using the discrete explanatory variable. This is because weeks of heavy rain are typically preceded by weeks with lighter rain, which is why no similar effect is observed when using the continuous measure.

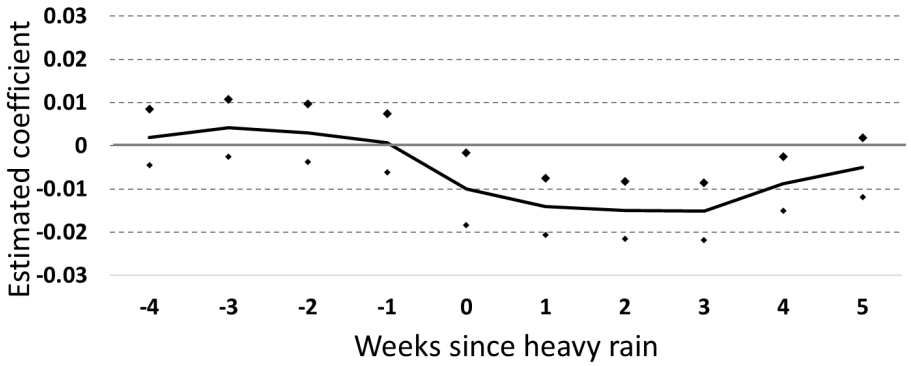
Figure 25: Relationship between pollution levels and heavy rain



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 and temperature squared as well as a full set of year and week-of-the-year fixed effects. Robust standard errors are used since there is no evidence of serial correlation in the data.

Next, I consider the impact of heavy rain on overall mortality. These results, in Figure 26, show that there was no relationship between heavy rainfall and mortality in the preceding weeks but that heavy rainfall was associated with reduced mortality for several following weeks.

Figure 26: Impact of heavy rain on total mortality



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 robust standard errors. The regression includes controls for temperature, temperature squared, four leads and five lags of temperature and temperature squared, a full set of year effects and a full set of week-of-the-year by decade effect. Data cover 1866-1965. N=4,479.

Table 28 allows us to compare the estimated effect of heavy rainfall on mortality in different causes of death to the effect of fog events. Note that the magnitudes are not directly comparable, since the rainfall cutoff level of 0.5 inches per week is arbitrary, but the patterns are instructive. We can see that the results obtained using the rainfall check are broadly similar to what we observe when using fog days (though of course in the opposite direction). For example, there are substantial effects on mortality due to bronchitis, cardiovascular diseases, pneumonia, measles and TB. There are, however, some differences that are worth considering.

An important factor in the rainfall results, which is not a concern with fog days, is that rainfall causes an increase in deaths due to digestive diseases in the week in which heavy rain occurs, as shown in Figure 27. This is most likely due to the fact that heavy rain may cause water sources to become contaminated. While the impact of heavy rainfall on digestive deaths shown in Figure 27 appears small in percentage terms, this represents a substantial number of deaths because digestive diseases were a major killer. As a result, this effect causes the impact of rainfall on mortality in the week in which rainfall occurred to be less comparable to the cleaner estimates generated using fog events. Another consequence of these rapid digestive deaths is that they may have reduced deaths from other causes in the following weeks by reducing the at-risk population. This effect is likely to be particularly important for infants, where digestive deaths were particularly important. This likely explains why, in Table 28, we see that heavy rainfall was associated decreases in mortality in causes of death that were concentrated in the 0-1 age group, such as neurological diseases, whooping cough,

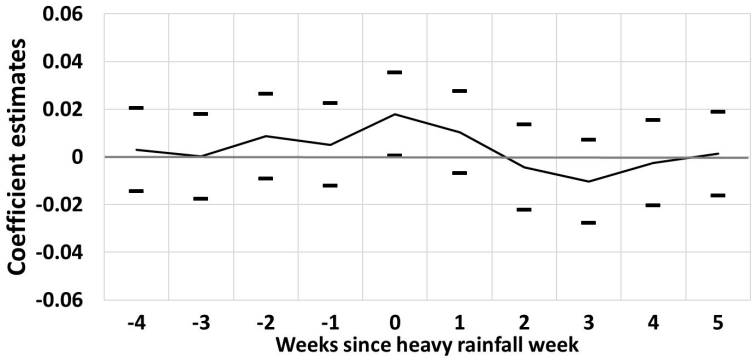
and the "other miscellaneous causes of death" category. With many young children killed by digestive diseases, fewer would have been at risk of dying from other cause. This feature also highlights why the fog analysis is likely to produce cleaner results than the rainfall analysis.

Table 28: Comparing fog event and rainfall effects by cause of death

DV: Log all-age mortality within disease category						
	Fog days			Heavy rainfall flag		
Cause of death	Coefficient		S.E.	Coefficient		S.E.
Respiratory & Cardiovascular						
Bronchitis	0.0349	***	(0.0045)	-0.0718	***	(0.0110)
Cardiovascular	0.0084	***	(0.0024)	-0.0265	***	(0.0052)
Pneumonia	0.0298	***	(0.0040)	-0.0809	***	(0.0092)
Misc. respiratory	0.0045		(0.0069)	-0.0985	***	(0.0178)
Infectious diseases						
Digestive dis.	-0.0037		(0.0037)	0.00361		(0.0106)
Measles	0.0796	***	(0.0137)	-0.123	***	(0.0345)
Scarlet Fever	0.0351	***	(0.0081)	-0.0248		(0.0206)
Smallpox	-0.00398		(0.0320)	0.163	**	(0.0702)
Tuberculosis	0.00615	***	(0.0021)	-0.0255	***	(0.0052)
Whooping cough	-0.000186		(0.0101)	-0.0720	***	(0.0230)
Infectious, other	0.0138	***	(0.0046)	0.00279		(0.0129)
Other diseases						
Cancer	0.00151		(0.0024)	-0.0097		(0.0060)
Neurological dis.	-0.0018		(0.0022)	-0.0153	***	(0.0056)
Old age	0.0044		(0.0037)	-0.0273	***	(0.0093)
Fetal underdevelopment	0.00027		(0.0039)	-0.0079		(0.0102)
Venereal diseases	-0.0016		(0.0079)	-0.0198		(0.0210)
Other misc. causes	0.0012		(0.0017)	-0.0130	***	(0.0043)
Other causes of death						
Accidents/violence	0.0026		(0.0036)	-0.00078		(0.0086)
Alcoholism	0.00832		(0.0094)	-0.0265		(0.0246)
Homicide†	-0.0072		(0.0122)	-0.0178		(0.0355)
Suicide	-0.0012		(0.0083)	-0.0178		(0.0185)
All causes	0.0099	***	(0.0021)	-0.0326	***	(0.00483)

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 25. All regressions include controls for temperature, temperature squared, five lags of each of these variables, and a full set of year and week-of-the-year by decade effects. The fog effects also include controls for pressure, pressure squared, precipitation, precipitation squared, humidity, and humidity squared. Data cover 1870 to 1939. † Homicide data are only separately available from 1870-1921. Homicide deaths are also included in the accidents and violence category.

Figure 27: Impact of heavy rain on deaths due to digestive diseases



This graph present coefficients and 95% confidence intervals for regressions of deaths due to digestive diseases on leads and lags of the heavy rain week indicator variable. Robust standard errors. The regressions include controls four leads and four lags of temperature and temperature squared as well as a full set of year and week-of-the-year fixed effects. Data cover 1870-1939.

A.12 Discussion of public health studies linking TB and pollution

This section offers a discussion of existing studies in the medical and public health literature examining the link between pollution and TB. This discussion is included in the Appendix rather than the main text because the analysis approach used in these studies is likely to struggle to identify causal effects.

There exists a relatively small set of public health studies that study association between the mortality of those with TB and their level of pollution exposure. A typical example of studies of this type is Blount, et al. (2017, "Traffic-Related Air Pollution and All-Cause Mortality during Tuberculosis Treatment in California," *Environmental Health Perspectives*, 125:9) which looks at mortality among 32,875 patients being treated for active TB in California from 2000-2012. Mortality patterns among this population are compared to traffic volumes and densities in buffers around the residence of each patient. They find that mortality risk increases with nearby traffic. This suggests a link between pollution exposure and TB mortality. While the study does include controls for demographic and socioeconomic factors, a natural concern with this approach is that patients who live in high-traffic areas differ from those living in low-traffic areas on dimension which cannot be observed or controlled for. As a result, while the evidence from this study is suggestive, it is difficult to draw strong causal conclusions.

A very similar approach is applied by Peng, et al. (2016, "Long-term exposure to ambient air pollution and mortality in a Chinese tuberculosis cohort," *Science of the Total Environ-*

ment, 580, p. 1483-1488) to a cohort of TB patients in China, which are observed from 2003-2013. As in the Blount, et al. study, Peng, et al. compare mortality rates to annual average pollution measures using patient's residence location. Similar to the previous study, they find evidence that those TB patients who lived in locations with more pollution had higher overall mortality and TB mortality.

A different set of public health studies argues that air pollution exposure may be associated with increased TB prevalence. One example in this area is Jassal, Bakman & Jones (2013, "Correlation of ambient pollution levels and heavily trafficked roadway proximity on the prevalence of smear-positive tuberculosis," *Public Health*, 127, p. 268-274), which uses medical records from Los Angeles County and compares the TB diagnosis to pollution exposure. As in the studies discussed above, pollution exposure is inferred from patient's residential location, which raises concerns that those living in more-polluted locations may differ in meaningful ways from those living in less-polluted locations. Their results indicate that air pollution exposure is correlated with TB risk.

A similar methodology is applied by Lai, et al. (2016, "Ambient air pollution and risk of tuberculosis: a cohort study," *Occup Environ Med*, 73, p. 56-61) using a cohort of patients in Taiwan tracked from 2005-2012. They also find evidence of a correlation between pollution exposure, based on residential address, and the chance that a patient was diagnosed with TB.

An earlier study, Smith, et al. (2014, "Particulate air pollution and susceptibility to the development of pulmonary tuberculosis disease in North Carolina: an ecological study," *International Journal of Environmental Health Research,* 24:2, p. 103-112) uses data covering all residents of North Carolina. Instead of residential location, this study uses county of residence to infer pollution exposure and then looks at whether those living in counties with higher pollution levels were more likely to be diagnosed with TB.

You, et al. (2016, "On the association between outdoor PM2.5 concentration and the seasonality of tuberculosis for Beijing and Hong Kong," *Environmental Pollution,* 218, p. 1170-1179) differs from other studies in this area in that they compare the seasonality of TB and air pollution. Their study shows that the seasonality of TB notifications is positively correlated with the seasonality of pollution levels. It is not clear if this reflects an increase in the number of people who acquire TB or whether it is instead due to more people seeking medical care for respiratory distress which leads to an increase in TB diagnoses. A fairly similar approach is taken by Zhu, et al. (2018, "Ambient air pollutants are associated with newly diagnosed tuberculosis: A time-series study in Chengdu, China," *Science of the Total Environment,* 631-632, p. 47-55) using data from Chengdu, China.

To summarize, there are a number of studies suggesting that TB prevalence and TB mortality risk are positively correlated with air pollution exposure. However, it is difficult to draw strong causal conclusions from this literature because pollution exposure is almost universally based on location of residence, and those who choose to live in more polluted locations are likely to differ from those living in less polluted locations along a range of unobservable dimensions which may affect exposure to TB or TB mortality.