

# The Causal Effect of Air Pollution on COVID-19 Transmission: Evidence from China

Guojun He, Yuhang Pan, and Takanao Tanaka<sup>1</sup>

There is increasing concern that ambient air pollution could exacerbate COVID-19 transmission. However, estimating the relationship is challenging because it requires one to account for epidemiological characteristics, to isolate the impact of air pollution from potential confounders, and to capture the dynamic impact. We propose a new econometric framework to address these challenges: we rely on the epidemiological Susceptible-Infectious-Recovered-Deceased (SIRD) model to construct the outcome of interest, the Instrument Variable (IV) model to estimate the causal relationship, and the Flexible-Distributed-Lag (FDL) model to understand the dynamics. Using data covering all prefectural Chinese cities, we find that a 10-point (14.3%) increase in the Air Quality Index would lead to a 2.80 percentage point increase in the daily COVID-19 growth rate with 2 to 13 days of delay (0.14 ~ 0.22 increase in the reproduction number:  $R_0$ ). These results imply that improving air quality can be a powerful tool to contain the spread of COVID-19.

First version: 19, Oct, 2020

This version: 19, Oct, 2020

---

<sup>1</sup> All authors have contributed equally to the manuscript. Corresponding author: Dr. Guojun He. He: Division of Social Science, Division of Environment and Sustainability, and Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: gjhe@me.com); Pan: Division of Environment and Sustainability, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: yhyhpan@gmail.com); Tanaka: Division of Social Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: takanao\_tanaka@hotmail.co.jp).

# Introduction

The rapid spread of COVID-19 has become a massive global public health and economic crisis, bringing millions of infections<sup>1</sup> and massive layoffs<sup>2</sup>. To design an effective response to this unprecedented pandemic, it is important to identify factors affecting virus transmission. Existing studies have documented that age<sup>3,4</sup>, gender<sup>4,5</sup>, comorbidities<sup>3,5,6</sup>, and climatic conditions<sup>7</sup> affect the virus symptoms. However, less is known about how ambient air pollution, which causes severe damage on various health outcomes<sup>8</sup>, can affect virus transmission.

Ambient air pollution could affect the spread of infections through increasing both exposure and susceptibility to the virus (see Supplementary Note 1 for details). Air pollution can cause a persistent inflammatory response and impair the respiratory and immune systems<sup>9,10</sup>, making it more challenging for an individual to resist infection. Besides, recent studies suggest that aerosols in the air may maintain the viability and transmissibility of the virus<sup>11-13</sup>. Therefore, degraded air quality, dominated by particulate matter (i.e., aerosols), may extend the survival of the virus in the air, which amplifies the chances of infection. Several previous papers suggest that air pollution may increase the spread of influenza<sup>14,15</sup> and SARS<sup>16</sup>; more recently, it has been found that air pollution is correlated with COVID-19 incidence in the U.S.<sup>17,18</sup>, Germany<sup>19</sup>, the Netherlands<sup>20</sup>, Italy<sup>21</sup>, the U.K.<sup>22</sup>, and China<sup>23</sup>.

However, at least two limitations have plagued the existing studies linking ambient air pollution to the COVID-19 outbreak. First, unlike other health outcomes, the virus growth is exponential, and failure to account for such non-linearity could easily provide biased estimates. For example, most existing studies use linear regression models to quantify the relationship between air pollution and COVID-19 prevalence and use new COVID-19 confirmed cases or deaths as the outcomes<sup>17-23</sup>. However, such models do not consider the epidemiological features

of disease transmission and may mistakenly generate spurious correlations between air pollution and factors that could affect the virus spread. (In the Methods section, we describe in detail why this is the case).

Second, quantifying the causal effect of air pollution on COVID-19 transmission is challenging. It is well documented in the literature that omitted variables and measurement errors could generate substantial biases in estimating the air pollution effects<sup>24-31</sup>. Such concerns are exacerbated in the case of COVID-19, because economic activities (e.g., opening industries and schools), health interventions (e.g., social distancing and business closure), and avoidance behaviors (e.g., wearing masks) not only change the transmission of COVID-19, but also could be correlated with air pollution exposure<sup>32,33</sup>. Existing evidence on the relationship between air pollution and COVID-19 relies mostly on associational approaches to quantify the impact and does not explicitly address the endogeneity of air pollution levels<sup>17,19,21-23</sup>. Hence, it remains unclear to policymakers, healthcare professionals, and researchers whether air pollution can causally affect the transmission of COVID-19.

In this study, we estimate the plausibly causal effect of air pollution on the transmission of COVID-19 across China, where the levels of air pollution are orders of magnitude higher than those in developed countries. Our empirical strategy is designed to overcome the two challenges. First, our econometric specification is derived from the Susceptible-Infectious-Recovered-Deceased (SIRD) model, which is widely used by epidemiologists to characterize the transmission of infectious disease. Based on this model, we show that the use of daily confirmed cases (or deaths) as the outcome variable in the regressions, as commonly employed by previous studies, could be problematic<sup>17-23</sup> (see Methods for details). Instead, the daily growth rate of the confirmed active infections is used as the outcome variable; it is obtained by taking the first difference in the natural logarithm of daily confirmed active cases. The model implies that using growth rate as

the outcome variable allows us to account for the exponential epidemic growth.

To address the second challenge, we adopt an Instrumental Variable (IV) approach to isolate the impact of air pollution from potential confounders. Specifically, we employ thermal inversions to instrument air pollution. Thermal inversion is a natural phenomenon that involves changes in the normal tendency of the air to cool down with altitude. When a thermal inversion occurs, a layer of warmer air overlays a layer of cooler air in the atmosphere<sup>25,34</sup>. Because warmer air has a lower density, the air pollutants emitted from the ground surface are “trapped,” which eventually leads to higher levels of local air pollution. The intuition of the IV model is that, while observed air pollution could be correlated with local economic activities, health interventions, individuals’ averting behaviors, etc., part of the variation in air pollution can be exogenously changed by thermal inversions. Thermal inversion is a complicated meteorological phenomenon and is unlikely to be correlated with the confounders mentioned above (see Methods). Therefore, by exploiting changes in air pollution that are induced by the plausibly random occurrences of thermal inversions, we can credibly isolate the causal impact of air pollution from the confounding factors.

In addition, to examine the dynamic impact of air pollution on the transmission of COVID-19, we incorporate the Flexible Distributed-Lag (FDL) models into our IV estimation<sup>38</sup>. Accounting for its dynamic component is important because there are time delays between infections and case confirmations due to the period of incubation, testing, and reporting<sup>35-37</sup>. Moreover, to ensure that our regression controls for potential confounders, we include date fixed effects to account for shocks identical to all cities but unique to each date (such as macroeconomic condition, the national level change in the definition of the disease, and the nation’s virus containment policies) and city fixed effects to account for time-invariant characteristics unique to each city (such as local healthcare resources, local testing capacity, and short-term population structure). We also include

time-varying weather variables (temperature, precipitation, and snow depth) in the regressions.

The core of our analysis uses comprehensive COVID-19 incidence and air quality data at a day-by-city level from 1 January to 1 April 2020 (N=30,360) in China (Figure 1). We focus on China because COVID-19 was largely controlled in the country by 1 April 2020, so we can observe the entire cycle of the virus transmission. The air quality data are collected from 1,605 monitoring stations covering all the 330 prefectural cities. We focus on the Air Quality Index (AQI), a composite measure of air pollution adopted by the Chinese government (see Supplementary Figure 1 and Supplementary Table 1 for details). The daily number of confirmed COVID-19 cases in each city is obtained from the National Health Commission of China. There were 49,982 confirmed cases in Wuhan city and 30,441 cases in other cities by 1 April 2020 (Figure 2 and Supplementary Figure 2). In our baseline analysis, we exclude Wuhan city from the regressions because of concerns about the city's COVID-19 data quality<sup>39</sup>, but our results remain robust when we include data from Wuhan.

## Results

The IV estimates can be obtained through a Two-Stage-Least-Squares (2SLS) procedure. In the first stage, we predict the air pollution concentration using thermal inversion (temperature difference between the ground surface and the upper layer). In the second stage, we regress the daily virus growth rate on the predicted pollution concentration rather than on observed pollution concentration. While observed air pollution could be correlated with factors potentially affecting the virus prevalence (economic activities, health intervention, and averting behaviors), air pollution concentration derived from the thermal inversion is plausibly uncorrelated with such confounders.

### **First-stage: Thermal inversion and air quality**

For the 2SLS procedure to be valid, it has to be the case that the thermal inversion is a strong predictor of the air pollution concentration. We find that thermal inversion is strongly correlated with variation in local air pollution, even when conditioning on all the time-varying control variables and a set of fixed effects (Figure 3A). Specifically, a 1°C increase in the temperature inversion is associated with a 3.03 point increase in the Air Quality Index (AQI) (Figure 3B) – where a higher AQI means worse air quality. Such patterns are also found when we further control for the city lockdown indicator and the days since the disease outbreak, and when we include Wuhan in the regression (Supplementary Figure 3). To confirm that this strong relationship is not driven by the local time trends or the spatial distribution of frequency of thermal inversions, we randomly shuffle the observed thermal inversions within the same location or within the same day 1000 times. The average estimates using the placebo sample in each case are close to zero, implying that the relationship is not spurious (Figure 3B).

### **Second-stage: Dynamic relationship between air quality and COVID-19 growth rate**

In the second stage, we use predicted air pollution from the first-stage regression to estimate the pollution-transmission relationship. We include up to 22 days of lags in the regressions (current + previous 21 days) to capture its dynamic impacts.

Our second-stage regression demonstrates positive impacts of air pollution on the COVID-19 growth rates with 2 to 13 days of delay (Figure 4A, see Supplementary Table 2 for full results). This delay is consistent with epidemiological observations that the disease is usually confirmed after incubation, testing, and reporting<sup>35-37</sup>. Specifically, a 10 point (14.3%) increase in AQI during these windows raises the growth rate by 2.80 percentage points. In contrast, before and after these

periods of higher pollution, air pollution does not have a consistent and meaningful effect: if we add the impact before t-2 and after t-14, the size of joint coefficients varies only slightly, but its standard error becomes larger.

These results are robust to a number of different model specifications. We add the city's lockdown status (Supplementary Figure 4A) and days since the outbreak as control variables (4B), include Wuhan in our regression (4C), and use different FDL model segments settings (4D, 4E). All results remain similar. In addition, we add three days of future air pollution as another placebo test and find that, as expected, its effects are close to zero and statistically insignificant (4F). When we change the lengths of the lags of air pollution from 16 days to 24 days, we also observe similar patterns (Supplementary Figure 5). (see Supplementary Note 2 and Supplementary Table 3 for the city's lockdown data<sup>40</sup>).

We observe that the average daily growth rate in the first week of the epidemic outbreak is 24.9% across cities, with the doubling time of infections at 2.78 days. If AQI increases 10 points, the doubling time would be shortened to 2.50 days. Alternatively, assuming the removal rate (total of recovered rate ( $\gamma$ ) and death rate ( $\rho$ ) in our SIRD model) is 13%~20%, which means the patients recover or die in 5.1~7.1 days (see Supplementary Note 3), a 10-point increase in AQI would lead to 0.14~0.22 higher reproduction number ( $R_0 = \beta/(\gamma + \rho)$ ).

Existing studies linking air pollution to various health outcomes often find that ordinary least squares (OLS) estimates understate the impact of air pollution<sup>24-31</sup> (Supplementary Figure 6), in that observed air pollution is directly used. We also find that the OLS estimates are substantially smaller than the IV estimates (even though the dynamic patterns are similar). OLS regression shows that a 10-point increase in AQI between 2 days and 13 days before the case is reported is associated with a 0.77 percentage point increase in the growth rate. This is about one-fourth of the IV estimate, suggesting that the OLS estimate can be biased downward. Supplementary

Figures 5 and 7 provide more robustness checks using different model specifications.

## Results by different air pollutants

We also provide the results using the specific air pollutants ( $PM_{2.5}$ ,  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $O_3$ ). The results show similar patterns for most of the pollutants: higher pollution levels increase the growth rate of COVID-19 with 2-13 days of delays (Supplementary Figure 8A). For example, a 10% increase in  $PM_{2.5}$  (A1),  $SO_2$  (A3), and  $CO$  (A5) statistically significantly leads to a 1.4 ppt, 1.1 ppt, and 3.0 ppt rise in the virus growth rate. The only exception is ozone, for which we find a higher concentration of  $O_3$  (A6) slightly decreases the disease transmission, even though the relationship is not statistically significant. This might be because (1) the concentration of ozone is often negatively correlated with other pollutants<sup>41</sup>, and (2) ambient ozone can inactivate the virus by disrupting the virus structure<sup>42</sup>. Moreover, across most pollutants, OLS estimates (Supplementary Figure 8B) are substantially smaller than the IV estimates, consistent with our baseline findings.

## Back-of-the-envelope calculations for the “blue sky” policy scenarios

We estimate the excess COVID-19 cases attributable to poor air quality. In China, when daily AQI is below 100, it is regarded as “good” or “moderate” air quality (also called a “blue sky” day). During our study period, 18.6% of the city-by-day observations do not meet this standard, and most of the “above-standard” (worse air quality) readings are obtained from northern Chinese cities. Here, we ask what would happen if we were able to bring the air quality index in all Chinese cities to below 100, holding other things constant.

Our estimates imply that the daily virus growth rate would have been slowed by 2.11% on average if all the cities had met the AQI=100 standard (Figure 5A). As a consequence, the number



of active infections would have been reduced substantially. For instance, the number of active infections would have dropped from 21,855 to 16,714 (23.5%) on 14 February, when we recorded the highest active infection number in China (Figure 5B). Applying our estimates to the observed removal rate, we expect that the cumulative confirmed cases would have been reduced by 25.7% (30,376 to 22,578) during our study period (Figure 5C).

The simulation shows that air pollution reshaped the exponential growth of the virus. We observe that the difference in the confirmed cases is small in the initial outbreak, but gradually becomes larger as the outbreak progresses. If we do not model the virus transmission process using the SIRD model, we will not be able to capture the dynamic, cumulative, and non-linear impacts of air pollution on COVID-19 cases.

## Discussion

Accurately estimating the effect of air pollution on COVID-19 transmission requires researchers to account for the exponentiality in virus growth and to introduce exogenous shocks to local air quality. Based on the SIRD model, we show that a non-model-based framework could produce biased estimates. This finding suggests that many studies linking air pollution to COVID-19 could be problematic<sup>17-23</sup>. The same concern also applies to studies that try to assess the impacts of virus containment policies<sup>43-45</sup> and weather conditions<sup>46,47</sup> on COVID-19 spread. We also documented that associational analysis based on simple OLS regression models will understate the true impact of air pollution on COVID-19 transmission<sup>17,19,21-23</sup>, as is consistent with the previous environmental health literature<sup>24-31</sup>.

Our analyses find that a 10-point (14.3%) increase in AQI leads to a 2.80 percentage point increase in the daily growth rate over 2 to 13 days after the pollution exposure. The effect is

statistically significant and economically meaningful. For example, holding all else constant, if we were able to bring all the air quality levels to the country's standard (AQI = 100, mean AQI would have been reduced by 13.2%), a back-of-envelope calculation reveals that the cumulative cases of COVID-19 would have been reduced by 25.7%

Knowing that air pollution can increase the transmission of COVID-19 is particularly relevant to developing countries that rely heavily on manufacturing and coal (such as India, Indonesia, and Pakistan). These countries face significant challenges in controlling COVID-19, in part because of the faster transmission rate caused by high levels of air pollution. Policymakers should thus consider adopting more stringent pollution control policies in their war to combat COVID-19.

Our study has two caveats. First, we use confirmed active infections to create our outcome variables. As is common in any study of infectious disease<sup>39</sup>, the number of confirmed cases could be much lower than the actual cases, and confirmed cases might not reflect the real epidemic outbreak. This concern is partially alleviated by our use of the growth rate as the outcome variable because our results will hold as long as under-reporting is constant (for example, if 50% of infected cases are always confirmed) within a city. In addition, our regressions include date fixed effects that can control for nationwide events specific to each date, such as national testing policies or revision of the disease classification. While controlling the variation in testing capacity can partially mitigate the concern about under-reporting, such data are not available at the city-by-day level.

Second, we do not have sufficient statistical power to investigate whether air pollution affects COVID-19 deaths. This is because there have been only a few COVID-19 deaths in most Chinese cities. Outside Wuhan, more than 90% of Chinese cities have recorded only 0 or 1 death. While more than 3,000 people died from COVID-19 in Wuhan city, the data were not accurate at the

early stage of the outbreak in the city<sup>39</sup>. Therefore, we refrain from discussing the relationship between air pollution and the COVID-19 death rate. Nevertheless, we investigate this issue in Supplementary Figure 9 and Supplementary Note 4 and do not find any statistically significant relationship. Future research is warranted on this issue if richer data from other countries become available.

# Methods

## Materials

*COVID-19 data.* The COVID-19 data are retrieved from the China National Health Commission (CNHC)<sup>48</sup>. The data comprise newly infected, recovered, and death cases from 20 January to 1 April in 2020 in 330 prefectural cities in China. These periods are overlapped with the disease outbreak in China. This includes 49,982 cases and 2,553 deaths in Wuhan and 30,441 cases and 727 deaths in other cities. More than 95% of confirmed cases recovered (tested as negative) during our study period. In our baseline analyses, we exclude Wuhan city due to concerns about the city's data quality. Wuhan was the epicenter of COVID-19 in China. During the first few weeks after the COVID-19 outbreak, the city faced severe medical resource shortages, and many patients could not get immediate diagnosis and treatment. In fact, a retrospective survey conducted by the Chinese government shows that many COVID-19 deaths in Wuhan may have been misclassified. On 17 April, the Chinese government added another 1,290 COVID-19 deaths in Wuhan (without reporting the timing of the fatalities). COVID-19 data outside Wuhan do not suffer from similar problems because there were far fewer COVID-19 cases in these cities, and COVID-19 tests became widely available soon after scientists learned about the situation in Wuhan.

*Air quality data.* The air quality data are obtained from 1,605 air quality monitoring stations covering all of the prefectural cities in China. These data are available from the Ministry of Ecology and Environment<sup>49</sup>. The AQI is a comprehensive measure of air pollution: the index is constructed using PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, CO, O<sub>3</sub>, and NO<sub>2</sub> concentrations, with a lower AQI meaning better air quality. In China, the AQI is determined by the maximum concentration of different air pollutants. We describe the relationship between the AQI and each pollutant in

Supplementary Figure 1 and Supplementary Table 1. To create the city-level air quality data, we first calculated the distance from a city's population center to all monitoring stations within the corresponding city. We then aggregated station-level air pollution data to city-level data using inverse distance weights. For this process, stations closer to the population center are given higher weights so that city-level air pollution data can better represent the population of each city. The weights are inversely proportional to square distance.

*Thermal inversion data.* The thermal inversion data are obtained from the MERRA-2<sup>50</sup>. The data include the temperature in 42 atmospheric layers (110m to 36,000m). We use the difference in temperature between the first layer (110m) and the second layer (320m) because this is expected to be associated with the air pollution at ground level. The raw data include the information for each 50\*60 km grid. We aggregate the grid-level data to the city level using the same methodology as the air pollution data.

*Weather data.* Weather data include temperature, precipitation, and snow depth. These data were obtained from the Global Historical Climatology Network (GHCN) from the U.S. National Oceanic and Atmospheric Administration (NOAA)<sup>51</sup>. We collapse these data to the city by day level using the same method as the air quality data.

## Methods

### SIRD Model

Our empirical analyses are based on the Susceptible-Infectious-Recovered-Deceased Model (SIRD Model), in that the individuals are classified into either Susceptible ( $S_t$ ), Infected ( $I_t$ ), Recovered ( $R_t$ ), or Deceased ( $D_t$ ). The number of each compartment evolves as follows:

$$S'_t = -\beta_t S_t I_t \quad (1)$$

$$I'_t = (\beta_t S_t - \gamma - \rho) I_t \quad (2)$$

$$R'_t = \gamma I_t \quad (3)$$

$$D'_t = \rho I_t \quad (4)$$

where  $\beta_t$  is transmission rate,  $\gamma$  is the recovered rate, and  $\rho$  is the death rate. Our empirics aim to recover how air pollution affects the transmission rate ( $\beta_t$ ), which is a widely used parameter to measure the spread of the epidemic because it deterministically affects disease development. Here, the reproduction number ( $R_{0t}$ ) is proportional to the transmission rate ( $R_{0t} = \beta_t/(\gamma + \rho)$ ). Using equation (2), the active epidemic growth can be modeled as:

$$I'_t = g_t I_t = (\beta_t S_t - \gamma - \rho) I_t \underset{S_t \rightarrow 1}{=} (\beta_t - \gamma - \rho) I_t \quad (5)$$

In our analyses, we assume the proportion of the susceptibles is close to 1 ( $S_t \rightarrow 1$ ), i.e., almost the entire population can be thought of as susceptible.

The solution of equation (5) can be described by the following exponential function:

$$\frac{I_t}{I_{t-1}} = e^{\beta_t - \gamma - \rho} = e^{g_t} \quad (6)$$

where  $I_t$  is active infections at time  $t$ , and the infection growth rate ( $g_t$ ) is proportional to the transmission rate ( $\beta_t$ ). Taking the natural logarithms of this equation, we get

$$\log(I_{t+1}) - \log(I_t) = g_t = g_0 + \alpha * f(AP_t) + \varepsilon_t \quad (7)$$

$$\text{where, } g_t = \beta_t - \gamma - \rho$$

We expect ambient air pollution to alter the COVID-19 growth rate ( $g_t$ ), through changes to the virus transmission rate ( $\beta_t$ ). Thus, we model the growth rate as a function of air pollution exposure  $f(AP_t)$  and its average treatment effect ( $\alpha$ ), in addition to the baseline growth rate ( $g_0$ ), which measures the growth rate without any exposure to air pollution, and a mean-zero error term ( $\varepsilon_{it}$ ).

We use the growth rate as the outcome variable because it helps us understand how pollution changes the transmission rate. In contrast, the use of new confirmed cases (or deaths) as an

outcome variable, as commonly employed by previous studies<sup>17-23</sup>, could produce estimates that are difficult to interpret. The number of new confirmed cases ( $C_t - C_{t-1}$ ) could be written as

$$\text{new confirmed cases}_t = C_t - C_{t-1} = (\beta_t S_{t-1}) I_{t-1} \quad (8)$$

Taking the natural logarithms of this equation, we get

$$\log(\text{new confirmed cases}_t) = \log(\beta_t S_{t-1}) + \log(I_{t-1}) = \vartheta'_0 + \alpha' * f(AP_t) + \varepsilon_t \quad (9)$$

As represented, the number of new confirmed cases depends on both virus transmission rate ( $\beta_t S_{t-1}$ ) and the current number of active infections ( $I_{t-1}$ ). Therefore, if we model the new cases as a function of air pollution exposure, the average impact of air pollution ( $\alpha'$ ) reflects not only the effect on virus transmission but on the current level of disease spread. Because air pollution could be associated with factors affecting the ongoing disease outbreak, such as the timing of the virus arrival, government interventions, and local economic activities, there could exist spurious correlations between air pollution and these factors.

### Estimation with Flexible Distributed Lag Model

To estimate equation (7), we start by fitting the simple multi-variable regression using Ordinary Least Squares (OLS). The effect of air pollution can appear with lags because the cases will be confirmed after a period of incubation, testing, and reporting. Therefore, we use a Flexible Distributed Lag (FDL) model to allow a dynamic time course of the relationship between air pollution exposure and virus transmission. This can be described as:

$$g_{it} = \log(I_{it}) - \log(I_{i,t-1}) = \sum_k AQI_{i,t-k} * f(\alpha_{t-k}^Z) + \sum_k X_{i,t-k} * f(\sigma_{t-k}^Z) + \theta_i + \pi_t + \varepsilon_{i,t} \quad (10)$$

where  $g_{it}$  is the daily growth rate of active infections in city  $i$  at day  $t$ .  $AQI_{i,t-k}$  denotes Air Quality Index (AQI) in city  $i$  at day  $t - k$ , and  $X_{i,t-k}$  is a set of control variables in city  $i$  at day  $t - k$ . We include temperature, precipitation, and snow depth as control variables because these

climatic conditions could affect virus transmission<sup>7</sup>. We include the AQI and these control variables during the period between  $k = 0$ , and  $k = 21$  (previous 3 weeks). Existing studies suggest that the virus incubation period is often around 5-6 days, with 2 days being the lower bound<sup>34-36</sup>. Therefore, we are particularly interested in the effect starting from  $k = 2$  to  $k = 13$ . Indeed, in our estimation, the impact during this window is the most salient, and including more periods in the regressions does not change the magnitude of the effect but enlarges its standard errors.

$\alpha_{t-k}$  represents the effect of air pollution on the daily virus growth rate in period  $t-k$ . To investigate the dynamic impact of air pollution, we adopt the FDL model, which approximates the set of coefficients  $\alpha_{t-k}$  as a cubic B-spline function with  $z$  segments, as denoted by  $f(\alpha_{t-k}^z)$ . Because daily variation in air pollution generally is highly correlated, this smoothing process helps us reduce artificial oscillations from our parameter estimates and interpret the coefficients easily; without this process, we might observe some random spikes and descents in the estimated parameters, which could not be rationalized<sup>38</sup>. In our primary model, we adopt  $z = 3$  segments. For the robustness check (Supplementary Figures 4 and 7), we also choose  $z = 2$  and  $z = 4$ .

$\theta_i$  and  $\pi_t$  denote city and time fixed effects, which are a set of city-specific and time-specific dummy variables. The inclusion of these sets of fixed effects helps isolate variation in air pollution exposure from time-invariant, time-trending, or seasonal confounders, which could be correlated with virus transmission. Specifically, city fixed effects ( $\theta_i$ ) account for time-invariant confounders specific to each city (e.g., the city's income level, natural endowments, and short-term industrial and economic structure), while time fixed effects ( $\pi_t$ ) account for shocks that are common to all cities on a given day (e.g., national virus containment policies, macroeconomic conditions, and the national air pollution time trend). Conditional on these fixed effects and the time-varying control variables, we assess how air pollution affects virus growth rate. If air pollution accelerates



virus transmission,  $\alpha_{t-k}$  will be positive.

Note that, as is common in any infectious disease literature, the confirmed active cases could be substantially lower than the true cases, and the use of the confirmed cases might not be precise<sup>39</sup>. In our study, this concern can be partially alleviated for at least three reasons. First, we use the growth rate of confirmed active infections as our outcome variable, so our findings will not be affected as long as the under-reporting is constant (for example, 50% of infections are always confirmed) within a city. Second, our regression includes date fixed effects, and this can absorb the national-level testing policies or revision of the disease classification. For example, the national government changed the disease classification on 18 February, but this variation can be absorbed by including date fixed effects. Finally, we exclude Wuhan, in which misreporting was more frequent<sup>39</sup>.

## Two-Stage Least Squares Strategy

Isolating the effect of air pollution from other confounding factors that could also affect the virus spread is a central empirical challenge in estimating the causal relationship between air pollution and COVID-19 transmission. In the literature, it is well documented that omitted variables and measurement errors could under-estimate the true pollution effects on health (also see Supplementary Figure 6)<sup>24-31</sup>. In addition, there are at least two additional potential sources of endogeneity when investigating the effects on infectious diseases.

First, the variation in air pollution could be correlated with individuals' averting behaviors, which could also affect virus transmission. When people are aware of severe air pollution, they might wear masks, purchase air purifiers, or reduce outdoor activities<sup>52,53</sup>, which could reduce virus transmission. If this is the case, the OLS estimators might understate the true impact of air pollution. Second, economic activities (e.g., opening industries and schools), human mobility (e.g., use of transportation), and health interventions (e.g., social distancing and business closures) can

affect the disease spread, and these could also be correlated with air quality<sup>33</sup>. This implies that even if we observe a positive association between air pollution and the epidemic's growth (with potential delays), the association might not necessarily reflect the pollution-disease relationship.

Our solution to address these empirical threats is to adopt the Instrumental Variable (IV) approach. We use thermal inversion to instrument air pollution<sup>25,34</sup>. Thermal inversion is a natural phenomenon in which a layer of cooler air is overlaid by a layer of warmer air in the atmosphere. When a thermal inversion occurs, the air pollutants emitted from the ground surface will be trapped, which raises the air pollution concentration. We use a 2SLS procedure to estimate the IV model. The first stage in 2SLS can be described as

$$AQI_{i,t-k} = \sum_k TInversion_{i,t-k} * f(\delta_{t-k}^Z) + \sum_k X_{i,t-k} * f(\sigma_{t-k}^Z) + \theta_i + \pi_t + \varepsilon_{i,t} \quad (11)$$

where  $TInversion_{i,t-k}$  denotes the difference in temperature between surface ground and upper layer in city  $i$  in time  $t-k$ , with the larger number representing the warmer temperature in the upper layer. (Therefore, we expect  $\delta_k^Z$  to be positive.) Other variables are analogous to equation (10).

In the second stage, we regress the daily virus growth rate on this predicted pollution concentration rather than on observed pollution concentration using the following equation:

$$g_t = \sum_k \widehat{AQI}_{i,t-k} * f(\widehat{\alpha}_{t-k}^Z) + \sum_k X_{i,t-k} * f(\sigma_{t-k}^Z) + \theta_i + \pi_t + \varepsilon_{i,t} \quad (12)$$

where  $\widehat{\alpha}_k$  measures the plausibly causal impact of air quality on virus transmission. In other words, the thermal inversion-driven air pollution has to be uncorrelated with the error term  $\varepsilon_{i,t}$ , conditional on the set of fixed effects and time-varying control variables, which is not directly testable. Here, the inclusion of these control variables can help isolate the variation in air pollution from other confounders at the cost of reducing much of the variation in pollution exposure. We cluster the standard errors at the city level.

## Placebo Test

To examine the relationship between thermal inversion and air pollution, we regress daily air pollution levels on the occurrence of thermal inversions, conditional on the set of control variables and fixed effects used in equation (10) - (12). We further conduct a “placebo” test to rule out the possibility that the strong inversion-pollution relationship is driven by the local trends/seasonality, the spatial distribution of frequency of the thermal inversion, or other factors. Specifically, we randomly shuffle the observed thermal inversions within the same location or within the same day by 1000 times and re-estimate the relationship between pollution and placebo inversions. We plot the distribution of the estimated coefficients and find that their average effects are close to zero.

## Back-of-the-envelope calculation

We estimate the excess COVID-19 cases attributable to poor air quality. Developing countries, including China, have suffered greatly from poor environmental quality as the cost of rapid economic growth<sup>54</sup>. In China, AQI lower than 100 is regarded as “moderate” or “good” air quality (See Supplementary Table 1) and is recognized as the “blue sky”. However, during our study period, 18.6% of the samples (city-by-day) did not meet this standard, and 30% of the cities explains 72.9% of the share with poor air quality. To estimate the excess COVID-19 cases, we first estimate their daily growth rate in two scenarios: observed air quality and the blue sky scenario (AQI is always lower than 100) by calculating the following equations.

$$\hat{g}_{it}^b = g_{0it} - \sum_k (AQI_{i,t-k} - AQI_{i,t-k}^b) * f(\hat{\alpha}_{t-k}^z) \quad (13)$$

where  $g_{0it}$  denotes the observed daily growth rate in city  $i$  at time  $t$ , and  $\hat{g}_{it}^b$  denotes the growth

rate if the AQI were always lower than 100.  $AQI_{i,t-k}$  is observed AQI and  $AQI_{i,t-k}^b$  is the hypothetical AQI, in that AQI higher than 100 is replaced with 100.  $f(\hat{\alpha}_{t-k}^z)$  represents the estimated coefficients from our main estimates (equation (12)). As discussed in the main analyses, we only consider impacts between  $k = 2$  and  $k = 13$ .

Using the predicted growth rate, the active infections in the blue sky scenario would have been:

$$\widehat{active}_{it}^b = \widehat{active}_{i,t-1}^b * \exp(\hat{g}_{it}^b) \quad (14)$$

The predicted number of active infections in city  $i$  at time  $t$  builds on that number in the previous period. To predict these numbers, we use the initial active cases in each city when it exceeds 10. We then use the observed removal rate in equations (3) and (4) to estimate the cumulative confirmed cases.

For the removal rate, we use the following equation.

$$removal\ rate_t (\gamma_t + \rho_t) = \frac{recovered\ individual + dead\ individual\ at\ time\ t}{infections\ at\ t - 1} \quad (15)$$

For this projection, we allow the removal rate to vary over time because it should vary depending on the stage of the outbreak. Note that the individuals are classified as recovered when they are tested as negative in our dataset. Therefore, its definition is slightly different from that in the classical SIRD model (See Supplementary Note 3).

## References

1. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
2. Forsythe, E. *et al.* Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. *J. Public Econ.* **189**, 104238 (2020).
3. Guan, W. *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
4. Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).
5. Williamson, E. J. *et al.* OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* (2020) doi:10.1038/s41586-020-2521-4.
6. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054–1062 (2020).
7. Carleton, T. *et al.* Ultraviolet radiation decreases COVID-19 growth rates: global causal estimates and seasonal implications. *SSRN Electron. J.* (2020) doi:10.2139/ssrn.3588601.
8. Zivin, J. G. & Neidell, M. Environment, health and human capital. *J. Econ. Lit.* **51**, 689–730 (2013).
9. Diamond, G., Legarda, D. & Ryan, L. K. The innate immune response of the respiratory epithelium. *Immunol. Rev.* **173**, 27–38 (2000).
10. Ciencewicz, J. & Jaspers, I. Air pollution and respiratory viral infection. *Inhal. Toxicol.* **19**, 1135–1146 (2007).
11. van Doremalen, N. *et al.* Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **382**, 1564–1567 (2020).
12. Zhang, R. *et al.* Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 14857–14863 (2020).
13. Liu, Y. *et al.* Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature* **582**, 557–560 (2020).

14. Clay, K., Lewis, J. & Severnini, E. Pollution, infectious disease, and mortality: Evidence from the 1918 Spanish influenza pandemic. *J. Econ. Hist.* **78**, 1179–1209 (2018).
15. Singer, G. *et al.* Air pollution increases influenza hospitalizations. (2020) doi:10.1101/2020.04.07.20057216.
16. Cui, Y. *et al.* Air pollution and case fatality of SARS in the People's Republic of China: an ecologic study. *Environ. Health.* **2**, 1-15 (2003).
17. X, Wu. *et al.* Exposure to air pollution and COVID-19 mortality in the United States. *medRxiv* (2020) doi.org/10.1101/2020.04.05.20054502
18. Persico, C. & Johnson, K. R. Deregulation in a time of pandemic: does pollution increase coronavirus cases or deaths? *IZA Discuss. Pap.* (2020).
19. Isphording, I. E. & Pestel, N. Pandemic meets pollution: poor air quality increases deaths by COVID-19. *IZA Discuss. Pap.* (2020).
20. Cole, M. A. & Strobl, E. Air Pollution exposure and COVID-19 air pollution exposure and COVID-19. *IZA Discuss. Pap.* (2020).
21. Ogen, Y. Assessing nitrogen dioxide (NO<sub>2</sub>) levels as a contributing factor to coronavirus (COVID-19) fatality. *Sci. Total Environ.* **726**, 138605 (2020).
22. Travaglio, M. *et al.* Links between air pollution and COVID-19 in England. *medRxiv* (2020) doi:10.1101/2020.04.16.20067405.
23. Zhu, Y. *et al.* Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. *Sci. Total Environ.* **727**, 138704 (2020).
24. Chay, K. Y. & Greenstone, M. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *Q. J. Econ.* **118**, 1121–1167 (2003).
25. Arceo, E., Hanna, R. & Oliva, P. Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. *Econ. J.* **126**, 257–280 (2016).
26. Knittel, C. R., Miller, D. L. & Sanders, N. J. Caution, drivers! children present: traffic, pollution, and infant health. *Rev. Econ. Stat.* **98**, 350–366 (2016).
27. Deschênes, O., Greenstone, M. & Shapiro, J. S. Defensive investments and the demand for air quality: Evidence from the NO<sub>x</sub> budget program. *Am. Econ. Rev.* **107**, 2958–2989 (2017).

28. Deryugina, T. *et al.* The mortality and medical costs of air pollution: Evidence from changes in wind direction. *Am. Econ. Rev.* **109**, 4178–4219 (2019).
29. Cheung, C. W., He, G. & Pan, Y. Mitigating the air pollution effect? The remarkable decline in the pollution-mortality relationship in Hong Kong. *J. Environ. Econ. Manage.* **101**, 102316 (2020).
30. He, G., Liu, T. & Zhou, M. Straw burning, PM<sub>2.5</sub>, and death: Evidence from China. *J. Dev. Econ.* **145**, 102468 (2020).
31. Heft-Neal, S. *et al.* M. Dust pollution from the Sahara and African infant mortality. *Nat. Sustain.* (2020) doi:10.1038/s41893-020-0562-1.
32. He, G., Pan, Y. & Tanaka, T. The short-term impacts of COVID-19 lockdown on urban air pollution in China. *Nat. Sustain.* (2020) doi:10.1038/s41893-020-0581-y.
33. Venter, Z. S. *et al.* COVID-19 lockdowns cause global air pollution declines. *Proc. Natl. Acad. Sci.* **5**, 18984–18990 (2020).
34. He, J., Liu, H. & Salvo, A. Severe air pollution and labor productivity: Evidence from industrial towns in China. *Am. Econ. J. Appl. Econ.* **11**, 173–201 (2019).
35. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020).
36. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
37. Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
38. Barwick, P. J. *et al.* The morbidity cost of air pollution: evidence from consumer spending in China. *NBER Working Paper* w24688 (2018).
39. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
40. Qi, J. *et al.* Do Lockdowns Bring about Additional Mortality Benefits or Costs? Evidence based on Death Records from 300 Million Chinese People. *medRxiv* (2020)  
doi:10.1101/2020.08.28.20183699.

41. Sillman, Sanford. 1999. "The Relation between Ozone, NO<sub>x</sub> and Hydrocarbons in Urban and Polluted Rural Environments." *Atmospheric Environment* 33(12): 1821-1845.
42. Tseng, C.-C. & Li, C.-S. Ozone for Inactivation of Aerosolized Bacteriophages. *Aerosol Sci. Technol.* 40, 683–689 (2006).
43. Auger, K. A. *et al.* Association between statewide school closure and COVID-19 incidence and mortality in the US. *JAMA* **45229**, (2020).
44. Dave, D. *et al.* When do shelter-in-place orders fight COVID-19 best? Policy heterogeneity across states and adoption time. *Econ. Inq.* 1–60 (2020) doi:10.1111/ecin.12944.
45. Islam, N. *et al.* Physical distancing interventions and incidence of coronavirus disease 2019: Natural experiment in 149 countries. *BMJ* **370**, 1–10 (2020).
46. Xie, J. & Zhu, Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Sci. Total Environ.* **724**, 138201 (2020).
47. Ma, Y. *et al.* Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Sci. Total Environ.* **724**, 138226 (2020).

## Methods References

48. China National Health Commission (accessed 1 July 2020); <http://en.nhc.gov.cn/index.html>,
49. China National Urban Air Quality Real-Time Publishing Platform (Ministry of Ecology and Environment of China, accessed 1 July 2020); <http://106.37.208.233:20035>
50. MERRA-2 (National Aeronautics and Space Administration, Goddard Space Flight Center, accessed 1 July 2020); <https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/>
51. Global Historical Climatology Network (National Oceanic and Atmospheric Administration, accessed 1 July 2020); <https://go.nature.com/2Vy9fez>
52. Liu, T., He, G. & Lau, A. Avoidance behavior against air pollution: evidence from online search indices for anti-PM<sub>2.5</sub> masks and air filters in Chinese cities. *Environ. Econ. Policy Stud.* **20**, 325–363 (2018).
53. Ito, K. & Zhang, S. Willingness to pay for clean air: Evidence from air purifier markets in China. *J. Polit. Econ.* **128**, 1627–1672 (2020).



54. Greenstone, M., He, G., Li, S. & Zou, E. China 's War on Pollution : Evidence from the First Five Years. *Working Paper* (2020).

## EndNote

### Competing Interests

We declare that none of the authors have competing financial or non-financial interests.

### Correspondence and Requests for Materials:

GH (email: [gjhe@me.com](mailto:gjhe@me.com)) addresses the correspondence and requests for materials.

### Acknowledgment

We would like to thank Wenbo Gong, Yatang Lin, and Deyu Rao for their valuable comments and suggestions. TT also thanks the Bai Xian Asia Institute for the scholarship support as a Bai Xian Scholar.

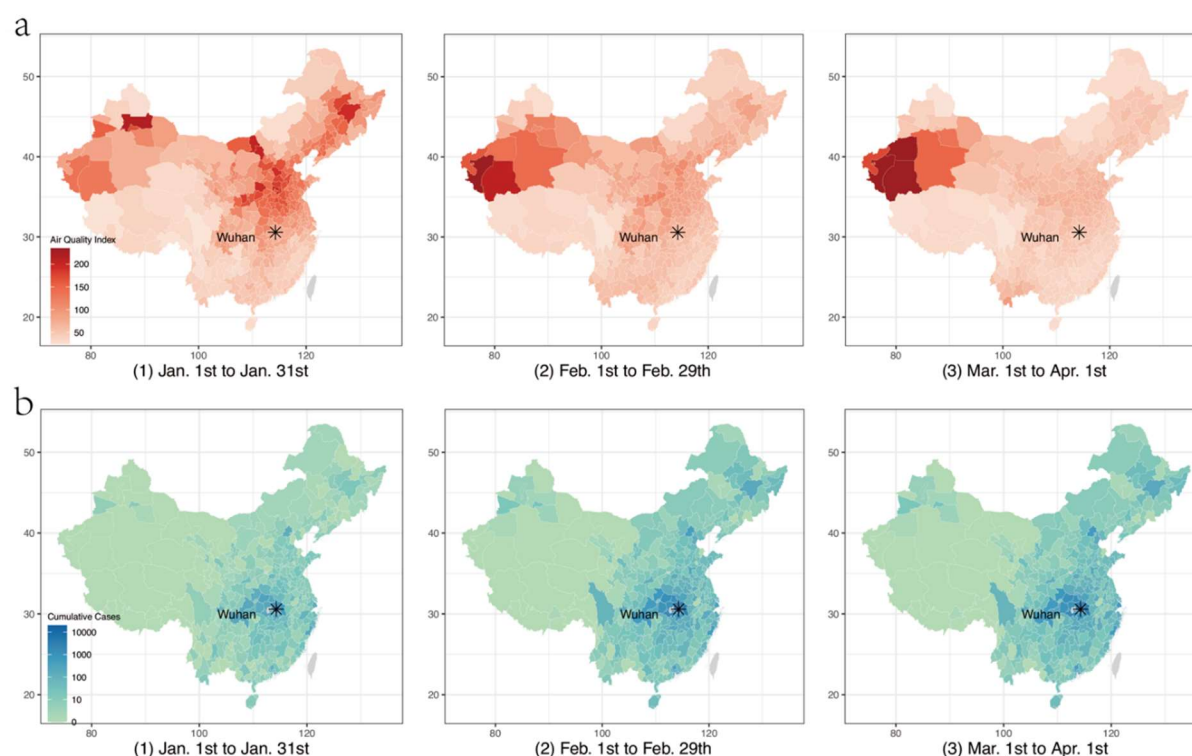
### Author Contributions

All authors (GH, PY, and TT) contributed equally to the paper. GH, PY, and TT conceptualized the study and carried out initial planning. PY retrieved and constructed the data set. PY and TT carried out the statistical analysis, which was refined by GH for the final version. TT prepared the first draft of the report, which was revised by PY and GH. All authors reviewed and contributed to a final draft and approved the final version for publication.

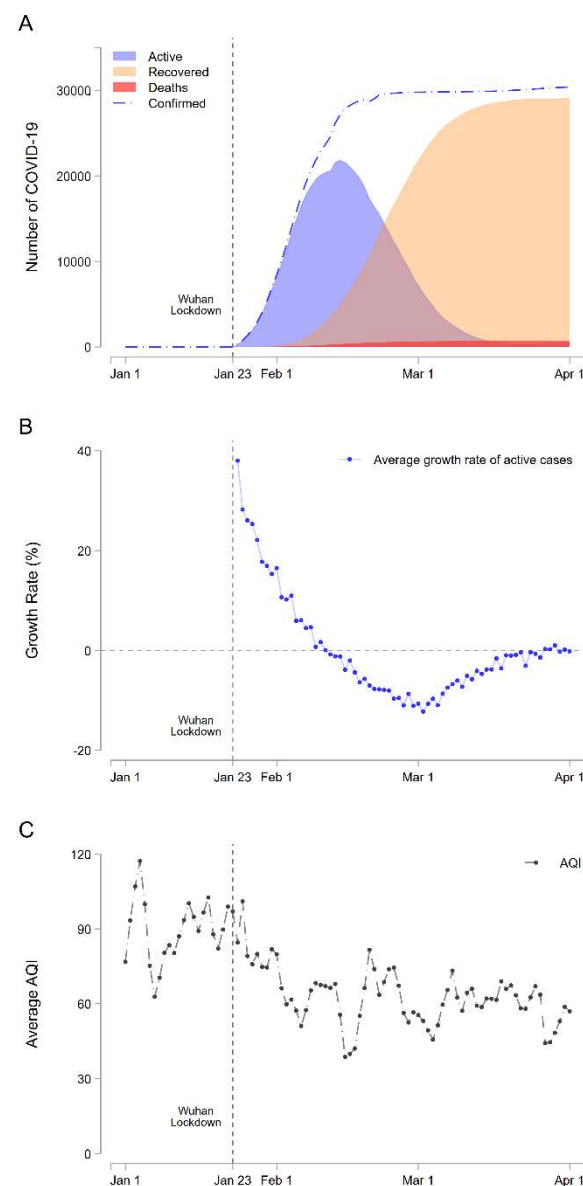
### Data and Code Availability Statement

All data and code necessary for replication will be available at a public repository upon publication.

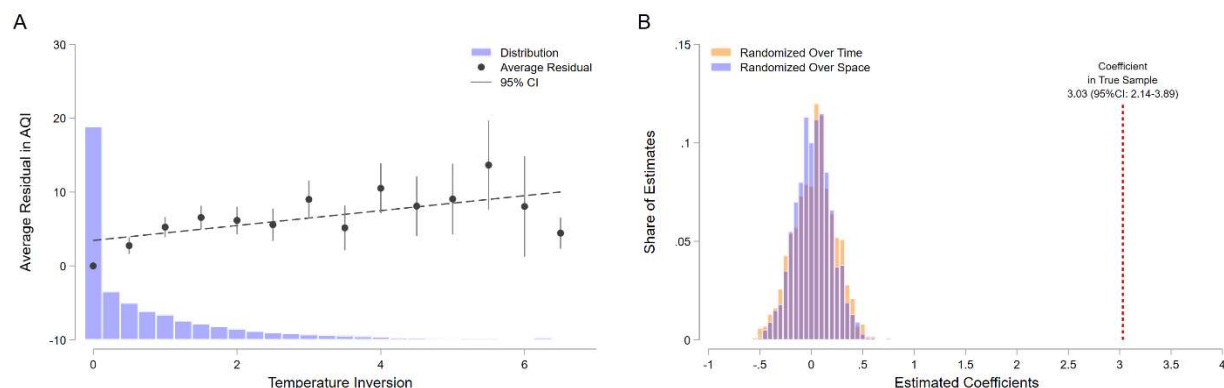
# Figures



**Figure 1 | Data on COVID-19 infections and air pollution in China. A.** The trend of the Air Quality Index (AQI). Higher AQI means worse air pollution. AQI is a comprehensive measure of air pollution: the index is constructed using PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, CO, O<sub>3</sub>, and NO<sub>2</sub> concentrations (See Methods Materials). **B.** The Confirmed COVID-19 cases. The gray color denotes the no-data area.

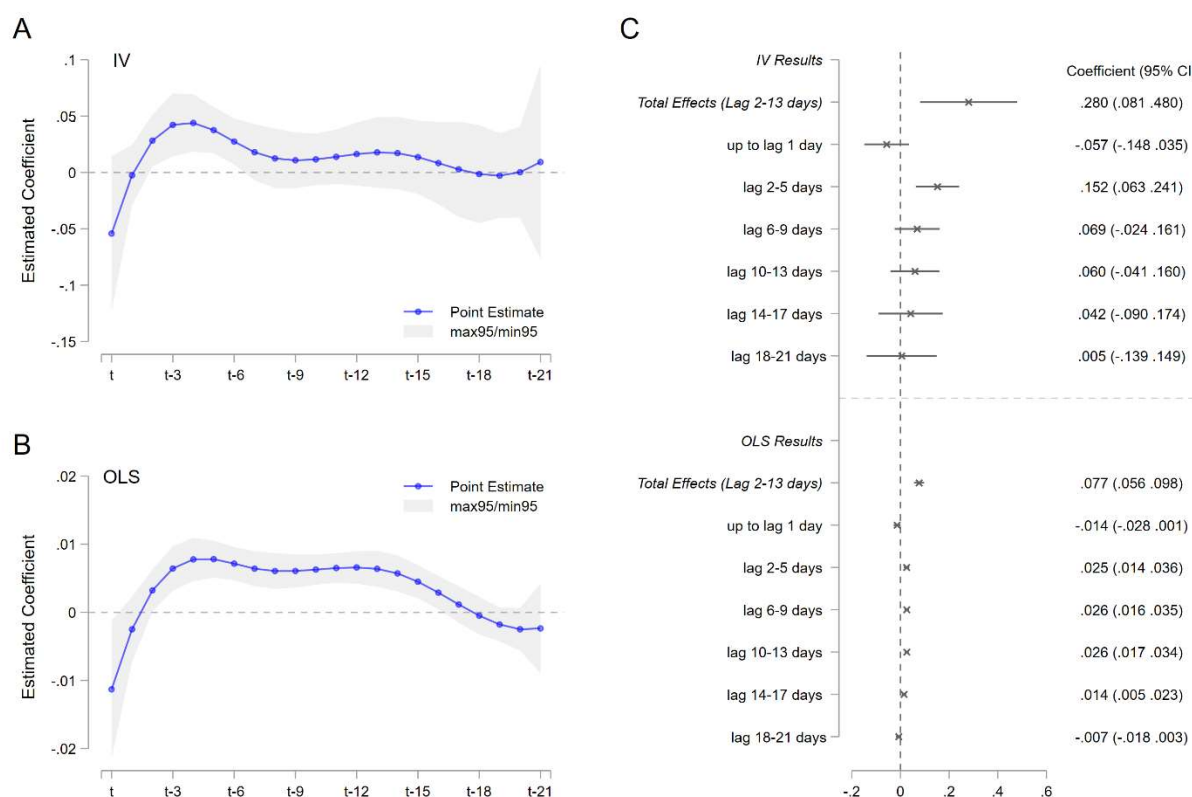


**Figure 2 | Data on COVID-19 infections and air pollution. A.** Daily confirmed cases (blue dash line), active infections (blue area), recovered (orange area), and deaths (red area) from 1 January to 1 April. The vertical line is the date of the Wuhan lockdown (23 January). **B.** The average growth rate of active infections. If it is larger than 0, the active infections increase. **C.** The trend of the Air Quality Index (AQI). Higher AQI means worse air pollution. AQI is a comprehensive measure of air pollution; the index is constructed using PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, CO, O<sub>3</sub>, and NO<sub>2</sub> concentrations (See Methods Materials).

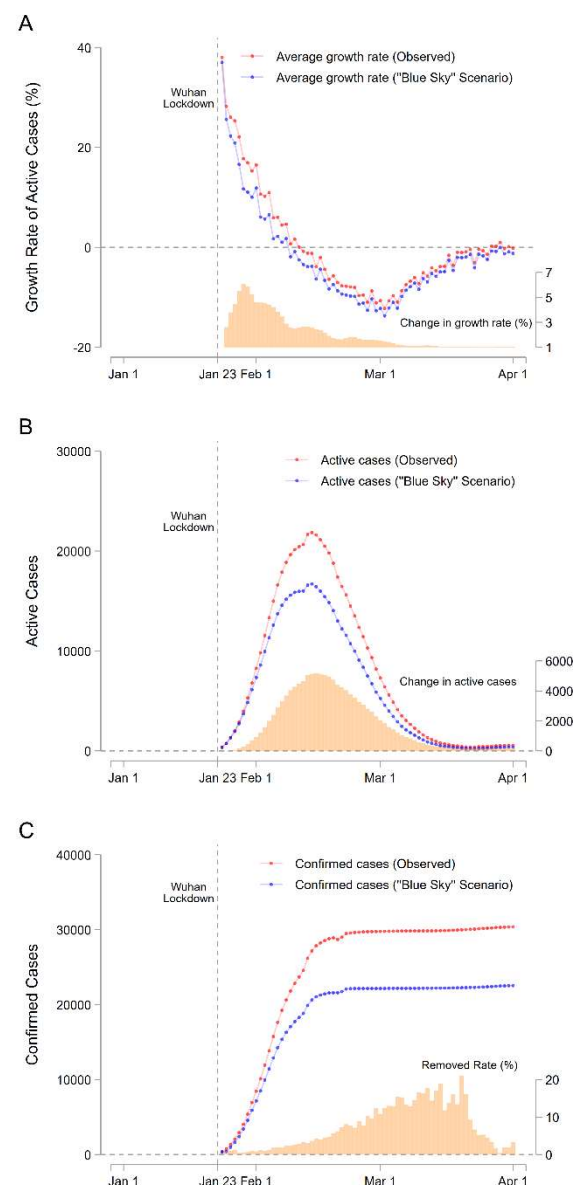


**Figure 3 | Variation in thermal inversion is strongly correlated with the Air Quality Index. A.**

The graph represents the distribution of thermal inversions, defined as the difference between the temperature in the second layer (320m) and that in the first layer (110m). The Y-axis represents the residual in variation in the AQI after controlling for temperature, precipitation, snow depth, city fixed effects, and date fixed effects. A higher temperature in the second layer is strongly associated with a larger residual in the AQI (worse air quality). **B.** A 1°C increase in temperature inversion is associated with a 3.03 increase in the AQI. In contrast, we find no statistically significant association between placebo thermal inversion and the AQI. In the placebo sample, we randomized the thermal inversions across different days within the city, or across different cities within a specific day.



**Figure 4 | Severe air pollution amplifies COVID-19 transmissivity with 2-13 days of delay.** **A.** shows the results using 2SLS. Weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects are included in both the first and second stage regression. The blue line represents the point estimates, while the gray area denotes the 95% confidence interval. **B.** directly regresses the daily disease growth rate on the observed air pollution rather than predicted air pollution. The regression includes the same controls as IV estimates. **C.** compares the magnitude of the effects between IV results and OLS results. The gray crosses represent the point estimates, and the gray line is the 95% confidence interval. In all regression, standard errors are clustered at the city level.



**Figure 5 | Simulated impacts of air pollution on the daily growth rate, active infections, and confirmed cases of COVID-19.** **A.** represents the observed daily growth rate of COVID-19 and the hypothetical rate in the “blue sky” scenario. In the blue sky scenario, an AQI exceeding 100 is replaced with 100. The orange bar represents the change in the growth rate (right axis). **B** shows the active cases in the two scenarios and the orange bar shows the difference. **C.** predicts the estimated cumulative confirmed cases over time. To compute these numbers, we use the observed removal rate, which is shown as the orange bar graphs. See Methods for details.

# Supplementary Materials for The Effect of Air Pollution on COVID-19 Transmission in China

Guojun He, Yuhang Pan, and Takanao Tanaka<sup>2</sup>

**Correspondence:** [ghje@me.com](mailto:ghje@me.com)

## **Contents:**

Supplementary Notes 1-4

Supplementary Figures 1-9

Supplementary Tables 1-3

Supplementary References 54~65

---

<sup>2</sup> Corresponding Author: Dr Guojun He. He: Division of Social Science, Division of Environment and Sustainability, and Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: [ghje@me.com](mailto:ghje@me.com)); Pan: Division of Environment and Sustainability, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: [yhyhpan@gmail.com](mailto:yhyhpan@gmail.com)); Tanaka: Division of Social Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (email: [takanao\\_tanaka@hotmail.co.jp](mailto:takanao_tanaka@hotmail.co.jp)).

# Supplementary Notes

## Supplementary Note 1: Mechanisms of air pollution and COVID-19 transmission

In this note, we elaborate how air pollution could affect COVID-19 virus transmission. First, short-term exposure to severe air pollution can weaken the immune system's response to the virus and thereby increase the chance of infection<sup>55,56</sup>. In physiological studies, inhaled particulates and other air pollutants can interact with immune cells within the airways. In particular, it is found that pollutants can trigger cellular signaling pathways, leading to multicellular immune responses and perturbation, and eventually cause disease or fail to prevent disease<sup>57,58</sup>. In epidemiological studies, air pollutants such as PMs, NO<sub>2</sub>, SO<sub>2</sub>, and O<sub>3</sub> are also found to be associated with inflammatory and immune responses<sup>59,60</sup>.

Second, aerosols in air pollutants may maintain virus activity and facilitate virus transmission. Aerosols are suspensions of solid or liquid particles in the air. These particles are small and have a low settling velocity; therefore, they can remain airborne for prolonged periods. For example, coughing and sneezing can generate a substantial quantity of particles. Because of evaporation, a large number of these particles shrink and then behave as aerosols. Aerosols due to coughing, sneezing and breathing can result in virus transmission, in which the virus from an infected person can be carried over a fairly long distance, compared to transmission through large droplets and direct contact. Both laboratory and epidemiological studies have shown that aerosol transmission can be an important mode of transmissions of influenza, chickenpox, measles, tuberculosis, smallpox, H5N1, MERS, Ebola<sup>61-66</sup>, and SARS-CoV-2 (i.e., COVID-19)<sup>11-13</sup>.

## Supplementary Note 2: COVID-19 Lockdown Data

We collected local governments' lockdown information city by city from various news media



and government announcements (Supplementary Table 3). Most of the cities' lockdown policies were directly issued by the city-level governments, although a few were promulgated by the provincial governments. To ensure compliance, civil servants and volunteers were assigned to communities, firms, business centers, and traffic checkpoints. Local governments also penalized offenders if the rules were violated. There were some variations in rules and in the degree of the lockdown. For example, in some cities, individuals were not allowed to go out (food and daily necessities were delivered to them), while in other cities, they could go out if they did not have a fever. In this paper, we designated a city as locked down when the following three measures were all enforced: (1) prohibition of unnecessary commercial activities in people's daily lives, (2) prohibition of any type of gathering by residents, (3) restrictions on private vehicles and public transportation. Following our definition, 95 out of 324 cities were locked down in our study period.

### **Supplementary Note 3: Removal rate**

For the back-of-the-envelope calculation, we use the observed removal rate to consistently project the epidemic growth in different policy scenarios. However, in our dataset, the individuals with infections are classified as recovered when they test negative, rather than when they recover from symptoms. Therefore, to estimate the reproduction number from our main results, we adopt the removal rate which is reported by existing studies. They suggest that an average interval is 5.1 days<sup>34</sup>, 5.8 days<sup>35</sup>, or 7.5 days<sup>36</sup>, making the removal rate 13.3%-19.6%. Therefore, we use a removal rate between 13%-20% to compute the reproduction number ( $R_0 = \beta/(\gamma + \rho)$ ).

### **Supplementary Note 4: The Effect of Air Pollution on the COVID-19 Death Rate.**

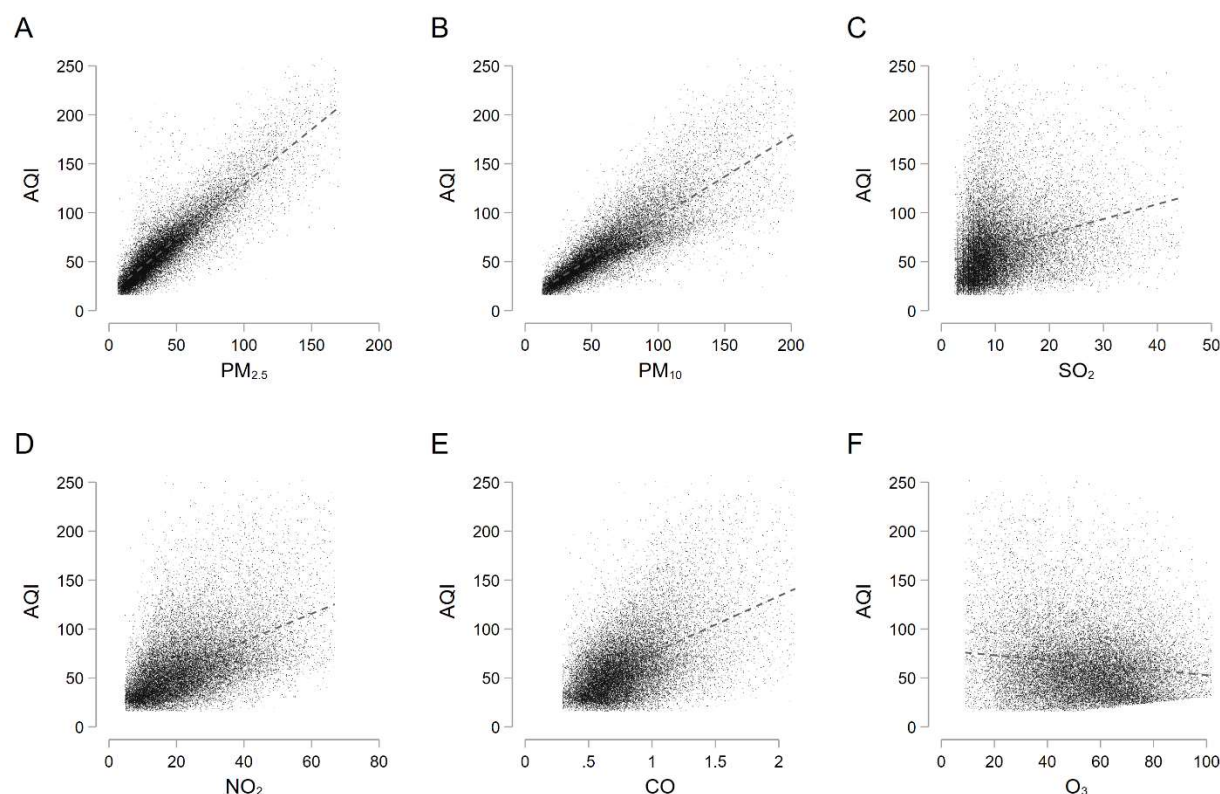
Existing studies suggest that air pollution increases deaths from COVID-19, likely because air pollution could impair the physical capability to recover from the infection. This is policy-relevant, but our dataset does not have sufficient statistical power to investigate the relationship. Outside Wuhan, there have been only around 700 deaths from the pandemic, with about 90% of cities having only 0 or 1 death. While more than 3,000 deaths are recorded in Wuhan, the data may not be credible. For example, on 17 April, its official death record was revised, with 1,290 deaths being added but without the timing of the deaths.

Nevertheless, here we examine the relationship between air pollution and death rate by aggregating the daily level data to the weekly level so that we can have a wider variation in the outcome. Our outcome variable is not the growth rate of deaths because it should be proportional to the growth rate of the infections. Instead, we use the death rate, which is defined as the probability of death given infection. Using this outcome variable, we fit the following model, analogous to equation (7),

$$\rho_t = \frac{D_t - D_{t-1}}{I_{t-1}} = \frac{\text{new deaths}_t}{\text{active infection}_{t-1}} = \rho_{t0} + f(AP_{it}) + \varepsilon_{it} \quad (S2)$$

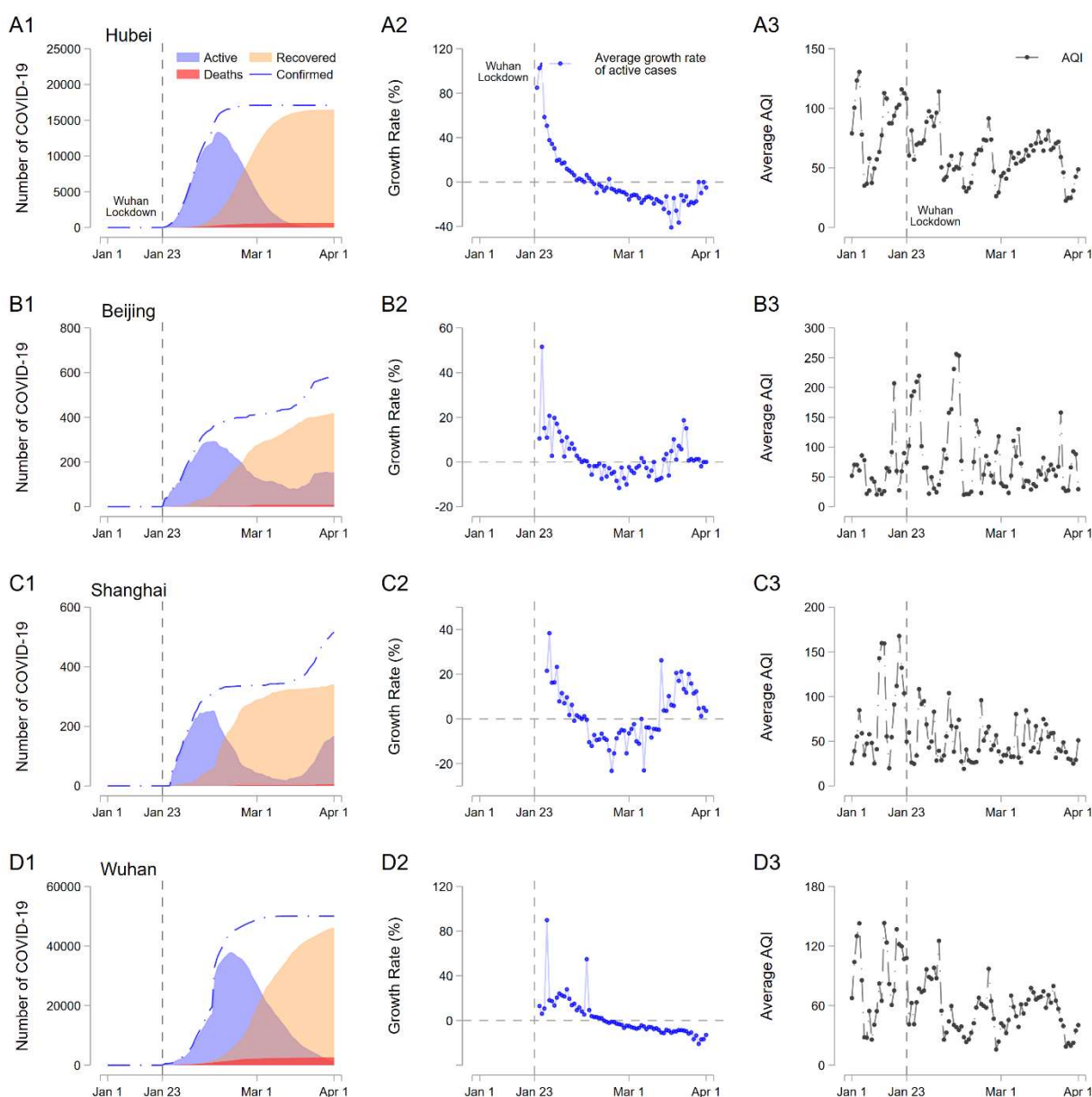
where the death rate is a function of baseline death rate, air pollution, and mean zero error term. We use the same control variables and a set of fixed effects as in equations (9) - (11).

We do not find any suggestive evidence that air pollution increases the death rate of the patients. All coefficients seem to be very small, and the sign is not consistent. As expected, our data may not have sufficient statistical power to investigate the relationship accurately.

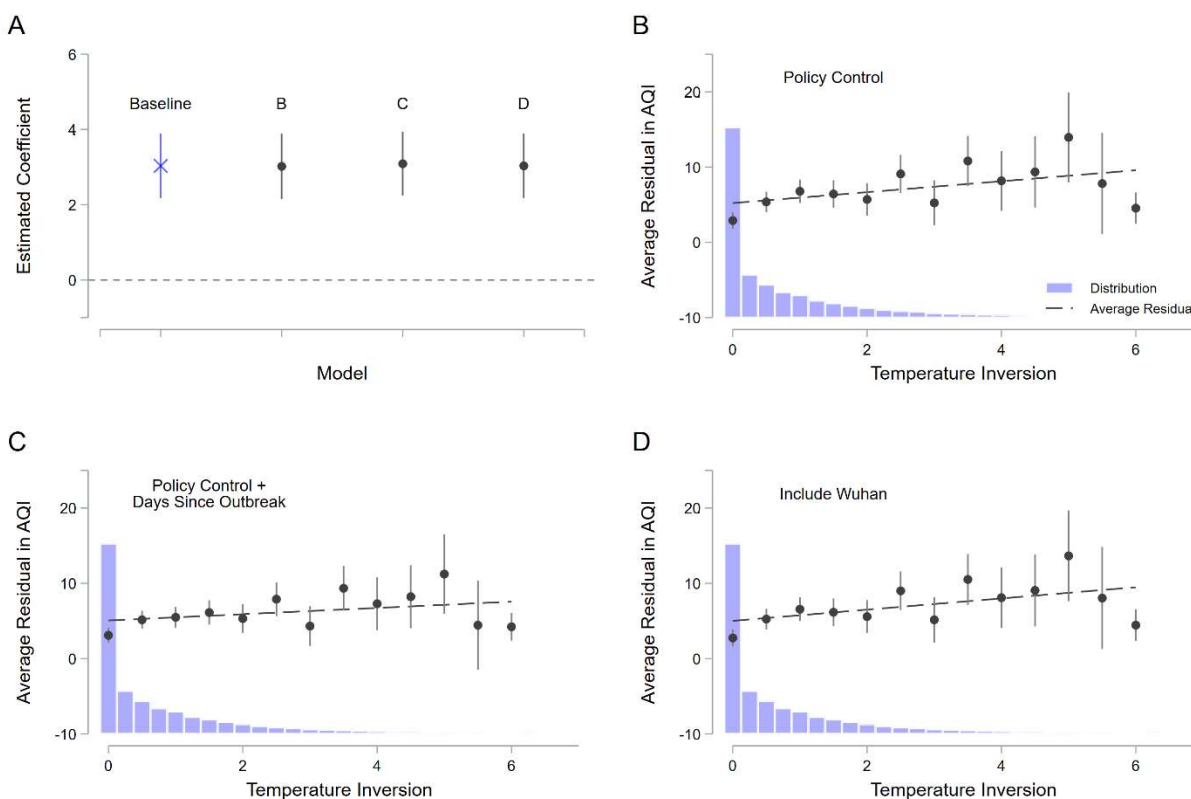


# **Supplementary Figure 1 | Air Quality Index is correlated with the most important pollutants.**

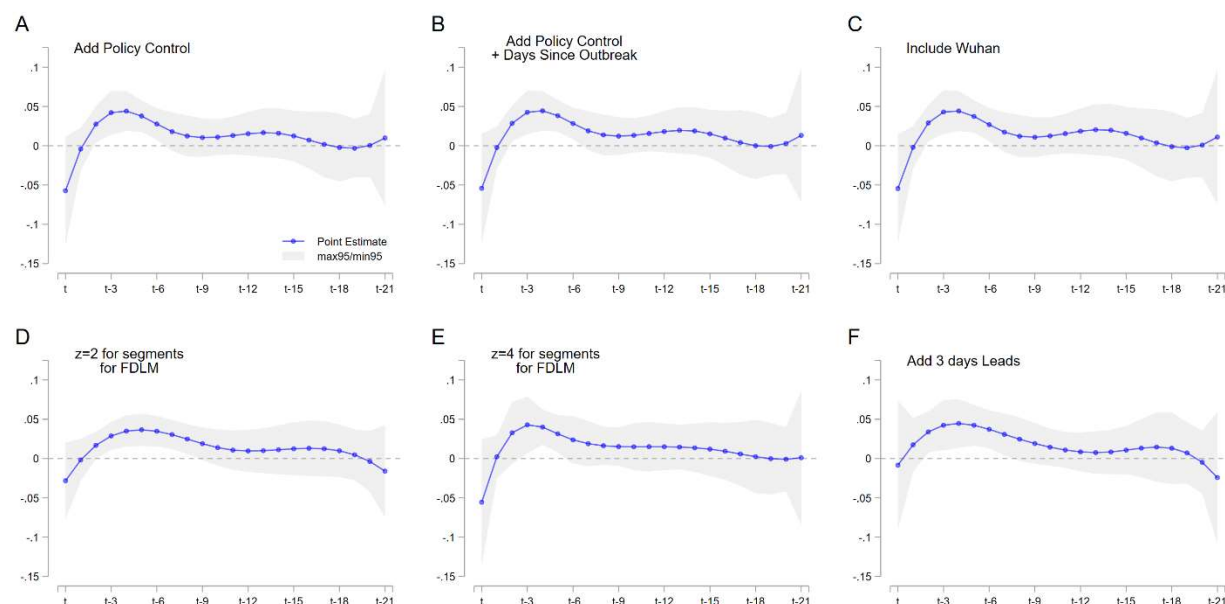
A-F represents the relationship between the AQI and each pollutant, including PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub>, which are components of the index. During the wintertime in China, PM<sub>2.5</sub>, PM<sub>10</sub> domain the primary pollutants of AQI in most cities. There is a strong correlation between the AQI and each pollutant, except for O<sub>3</sub> (ozone). This might be because ozone is mechanically negatively correlated with some of the primary pollutants. We trim the observations below 1 percentile and above 99 percentile in each pollutant. See Extended Data Table 1 for the definition of the AQI.



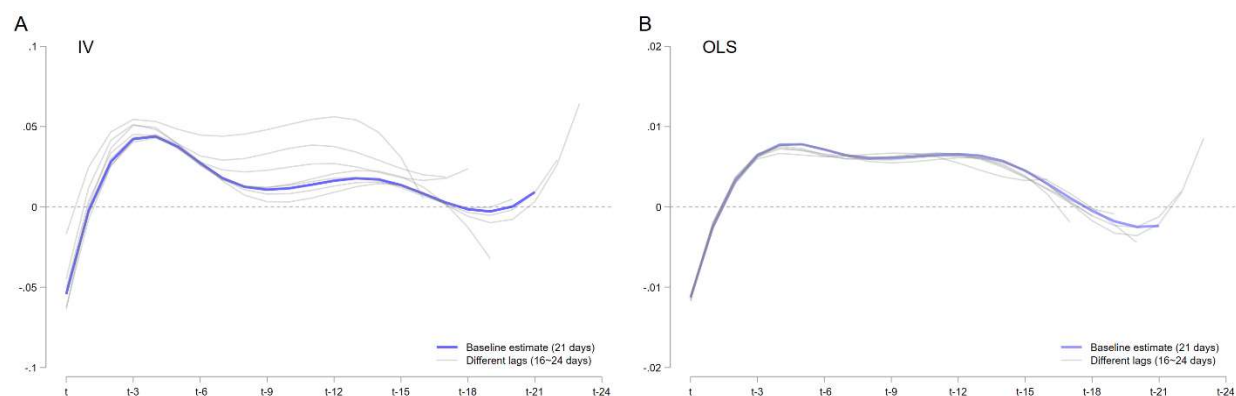
**Supplementary Figure 2 | COVID-19, its growth rate, and the Air Quality Index in Hubei Province, Beijing, Shanghai, and Wuhan.** These graphs show the COVID-19 outbreak (number of active, recovered, and deceased from COVID-19), the infection growth rate, and AQI in each region.



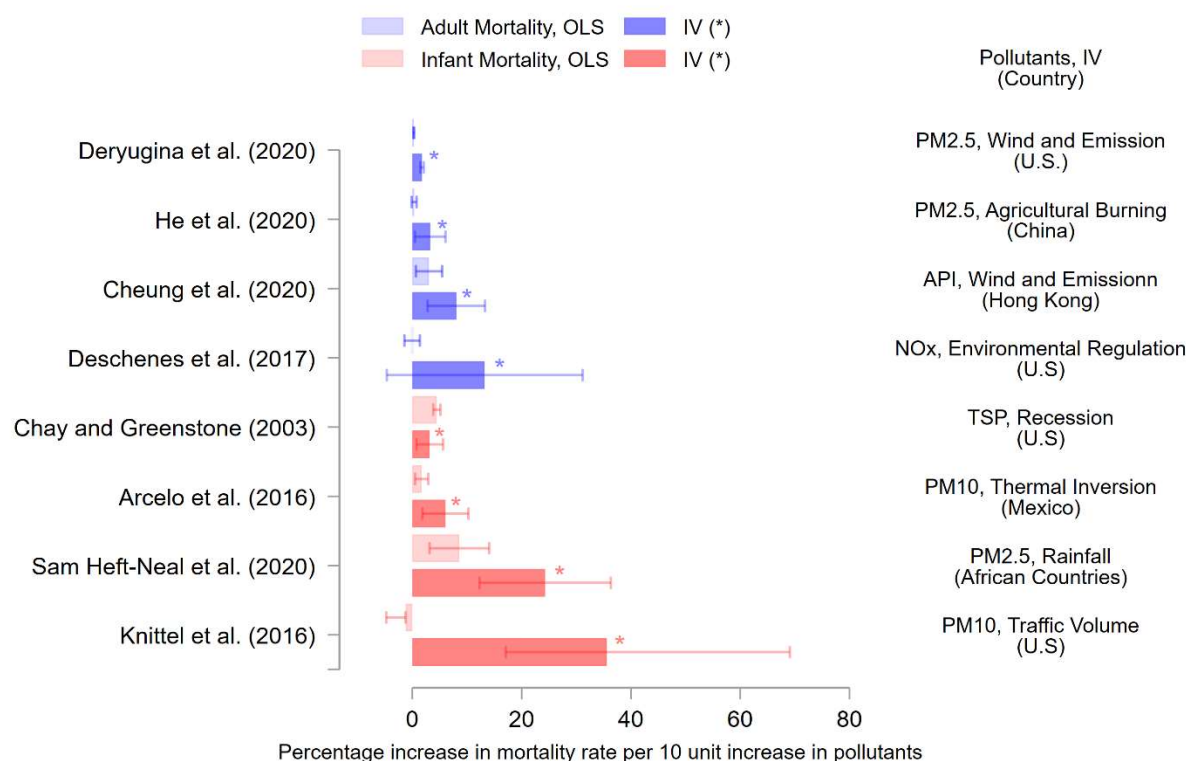
**Supplementary Figure 3 | Strong correlation between thermal inversion and air pollution is robust to different specifications.** These graphs show the correlation between temperature inversion and residual in the Air Quality Index, after controlling for the weather variables, city fixed effects, and date fixed effects. **A.** shows the coefficients of the contemporaneous relationship for each model, which corresponds to Figure 3A. **B-D.** show the distribution of thermal inversion and the average residual. **B.** includes lockdown status for the control variable, **C.** includes lockdown status and days since the outbreak (first confirmed cases). **D.** includes Wuhan city in the regression.



**Supplementary Figure 4 | The effects of air pollution on the COVID-19 growth rate using the IV estimates are robust to a number of model specifications. A.** lockdown status is added as a control variable. **B.** the lockdown status and days since the outbreak (the first case confirmed) are added in the regression. **C.** We include Wuhan. **D and E.** We adopt different segments for the Flexible Distributed Lag Model. **F.** We add three days of future air pollution. The joint coefficient for the three days lead is 0.002, with the standard error at 0.051, suggesting that future air pollution does not affect the disease growth rate. Thermal inversions and the same controls are used for all first-stage regressions. All regressions include weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects. Standard errors are clustered at the city level.

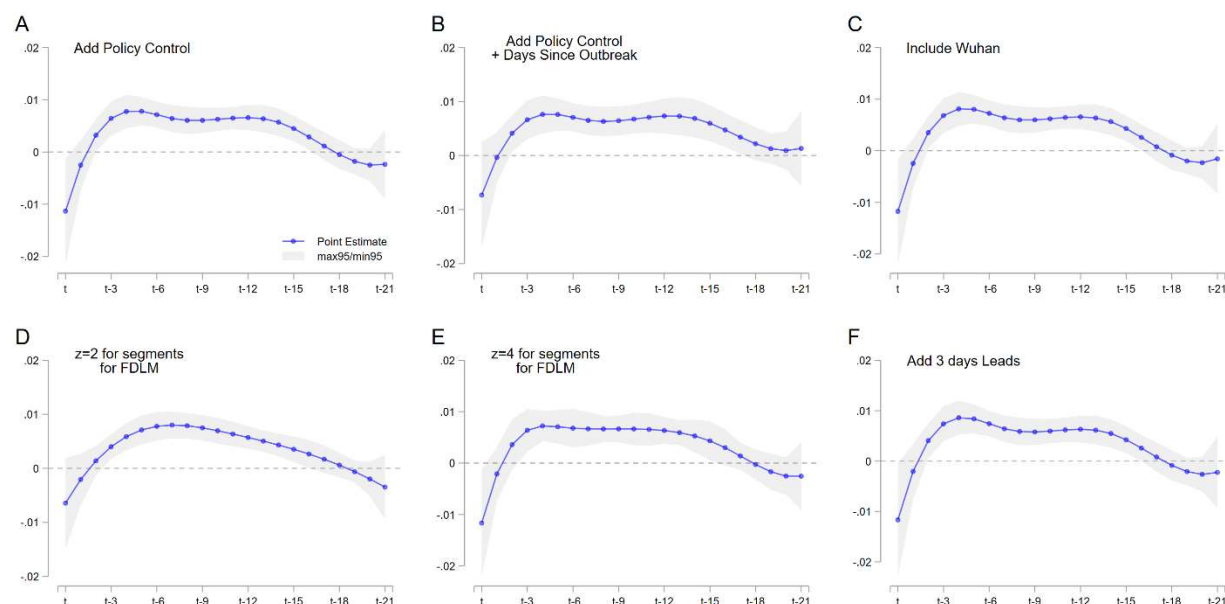


**Supplementary Figure 5 | The effects of air pollution on the COVID-19 growth rate using different lags. A.** represents the results using IV. We use different lengths of lags from 16 days to 24 days, while the main estimation uses 21 days. The blue line shows the baseline estimates, and the gray line shows the results with different lags. **B.** shows the results using OLS estimates.

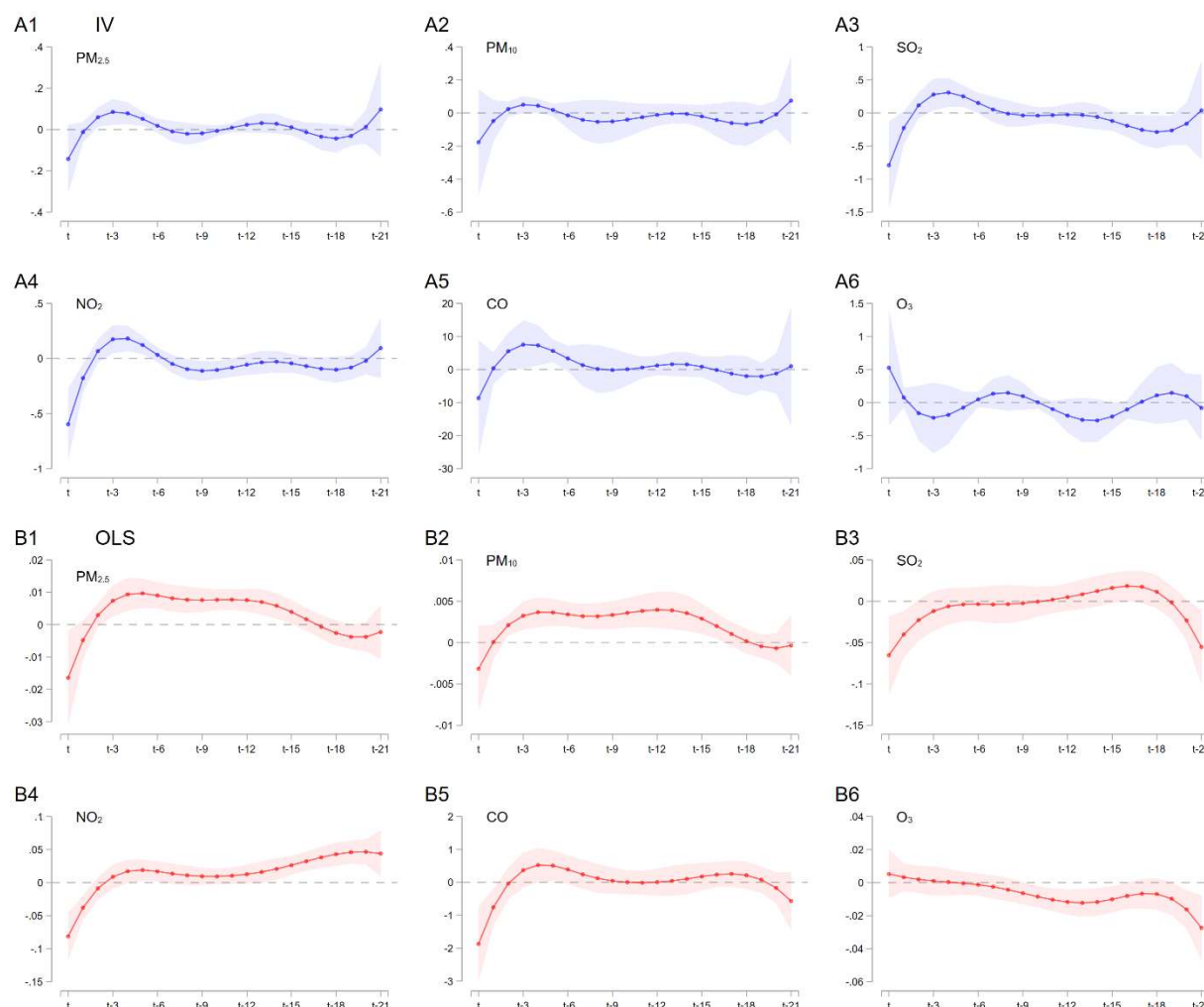


**Supplementary Figure 6 | Instrumental variables (IV) estimates are consistently larger than OLS estimates in existing studies linking air pollution and health outcomes.** The graph represents the estimates of the effect per 10 unit increase in air pollution on mortality rate (%). Deryugina et al. (2020) use mortality among aged above 65. Except for Chay and Greenstone (2003), existing studies report the OLS estimates are smaller than the IV estimates. Note that different studies focus on different pollutants and use different instrumental variables. Therefore, we do not compare estimates across different studies but across different methods within each study.

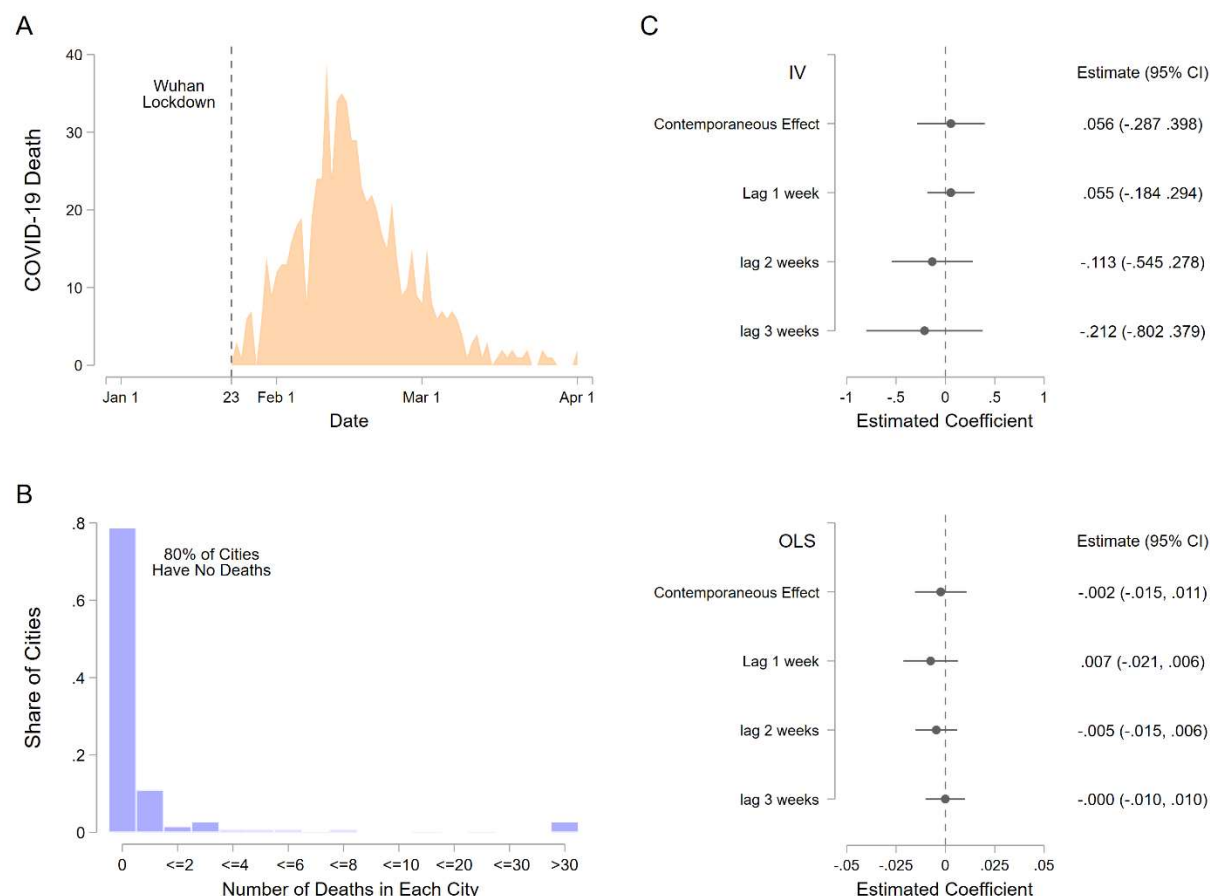




**Supplementary Figure 7 | The correlations between air pollution and COVID-19 growth rate using OLS estimates are robust to a number of model specifications. A.** lockdown status is added as a control variable. **B.** the lockdown status and days since the outbreak (the first case confirmed) are added in the regression. **C.** We include Wuhan. **D and E.** We adopt different segments for the Flexible Distributed Lag Model. **F.** We add three days of future air pollution. The joint coefficient for the three days lead is 0.010, with the standard error at 0.010, suggesting that future air pollution does not affect the disease growth rate. All regressions include weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects. Standard errors are clustered at the city level.



**Supplementary Figure 8 | The effects of air pollution on the COVID-19 growth rate using different air pollutants. A1-A6.** Represents the results using IV. Each graph shows the coefficient of a 1 point increase in each pollutant in each day. Weather controls (temperature, precipitation, and snow depth), date fixed effects, and city fixed effects are included in both the first and second stage regression. The blue line represents the point estimates, while the blue area denotes the 95% confidence interval. **B.** represents the results using OLS. The regression includes the same controls as the IV estimates. The red line represents the point estimates, while the red area denotes the 95% confidence interval. In all regressions, standard errors are clustered at the city level.



**Supplementary Figure 9 | Impacts of air pollution on the COVID-19 death rate.** **A.** The trend of the COVID-19 deaths over time outside Wuhan. **B.** Distribution of deaths in each city. **C.** The graph above shows the effect of air pollution on death rate using IV, while the graph below shows the effect using OLS. To keep the variation in death rate, we aggregate data to the week-by-city level. The results do not show a statistically significant relationship between AQI and the death rate.

## Supplementary Tables

AQI	Each Air Pollutant						Air Quality Levels
	PM <sub>10</sub> (24hr)	PM <sub>2.5</sub> (24hr)	NO <sub>2</sub> (24hr)	O <sub>3</sub> (8hr)	CO (24hr)	SO <sub>2</sub> (24hr)	
0-50	0-50	0-35	0-40	0-100	0-2	0-50	Excellent
50-100	50-150	35-75	40-80	100-160	2-4	50-150	Good
100-200	150-350	75-150	80-280	160-265	4-24	150-800	Slightly Polluted
200-300	350-420	150-250	280-565	265-800	24-36	800-1600	Moderately Polluted
300-400	420-500	250-350	565-750	/	36-48	1600-2100	Severely Polluted
400-500	500-600	350-500	750-940	/	48-60	2100-2620	Severely Polluted

**Supplementary Table 1 | The relationship between the AQI and different air pollutants.** This table reports the AQI sub-index levels for each air pollutant. The sub-index with the highest value will then be used as the AQI. For CO, the unit is mg/m<sup>3</sup>, and for other pollutants, the units are µg/m<sup>3</sup>.

Dependent Variable: Daily disease growth rate	OLS		IV	
	coefficients	s.e.	coefficients	s.e.
	(1)	(2)	(3)	(4)
Current Day	-0.011**	[0.005]	-0.054	[0.035]
lag: 1 day	-0.003	[0.002]	-0.002	[0.014]
lag: 2 days	0.003**	[0.002]	0.028**	[0.012]
lag: 3 days	0.006***	[0.002]	0.042***	[0.014]
lag: 4 days	0.008***	[0.002]	0.044***	[0.013]
lag: 5 days	0.008***	[0.001]	0.038***	[0.010]
lag: 6 days	0.007***	[0.001]	0.028***	[0.010]
lag: 7 days	0.006***	[0.001]	0.018	[0.013]
lag: 8 days	0.006***	[0.001]	0.012	[0.013]
lag: 9 days	0.006***	[0.001]	0.011	[0.013]
lag: 10 days	0.006***	[0.001]	0.012	[0.012]
lag: 11 days	0.006***	[0.001]	0.014	[0.012]
lag: 12 days	0.007***	[0.001]	0.016	[0.014]
lag: 13 days	0.006***	[0.001]	0.018	[0.016]
lag: 14 days	0.006***	[0.001]	0.017	[0.016]
lag: 15 days	0.004***	[0.001]	0.014	[0.016]
lag: 16 days	0.003**	[0.001]	0.008	[0.019]
lag: 17 days	0.001	[0.001]	0.003	[0.021]
lag: 18 days	-0.000	[0.001]	-0.001	[0.022]
lag: 19 days	-0.002	[0.001]	-0.003	[0.019]
lag: 20 days	-0.002	[0.002]	0.000	[0.020]
lag: 21 days	-0.002	[0.003]	0.009	[0.044]
z-Segment, k-order	3, 3		3, 3	
Observations (cities)	22,701 (329)		22,701 (329)	
Weather Control	Y		Y	
Date fixed effects	Y		Y	
City fixed effects	Y		Y	

**Supplementary Table 2 | Full results of the effect of air quality on the COVID-19 growth rate.**

The results correspond to Figure 4. The dependent variable is the day-by-city level growth rate of the activated COVID-19 cases. Each estimate indicates the effect of the current and past air pollution (Air Quality Index) on the growth rate of COVID-19. Weather controls include temperature, precipitation, and snow depth. Standard errors are clustered at the city level and shown in the right-side brackets. Significance levels are indicated by \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , and \*  $p < 0.01$ .

Starting Date	Cities
23-Jan	Wuhan
24-Jan	Huangshi, Shiyan, Yichang, Ezhou, Jingmen, Xiaogan, Huanggang, Xianning, Enshi
25-Jan	Qinhuangdao
26-Jan	Xiangyang, Jingzhou, Xiantao
28-Jan	Tangshan
30-Jan	Dongying
31-Jan	Chongqing, Yinchuan, Wuzhong
2-Feb	Wenzhou
3-Feb	Wuxi, Jining
4-Feb	Harbin, Nanjing, Xuzhou, Changzhou, Nantong, Hangzhou, Ningbo, Fuzhou, Jingdezhen, Zaozhuang, Linyi, Zhengzhou, Zhumadian
5-Feb	Shenyang, Dalian, Anshun, Fushun, Benxi, Dandong, Jinzhou, Fuxin, Liaoyang, Panjin, Tieling, Chaoyang, Huludao, Yangzhou, Hefei, Quanzhou, Nanchang, Jinan, Qingdao, Taian, Rizhao, Laiwu, Nanning
6-Feb	Tianjin, Shijiazhuang, Suzhou, Pingxiang, Jiujiang, Xinyu, Yingtan, Ganzhou, Ji'an, Yichun, Fuzhou, Shangrao, Neijiang, Yibin, Xinyang
7-Feb	Suzhou, Guangzhou
8-Feb	Shenzhen, Foshan, Fangchenggang,
9-Feb	Cangzhou, Huaibei
10-Feb	Beijing, Shanghai
13-Feb	Hohhot, Baotou, Wuhai, Chifeng, Tongliao, Ordos, Hulun Buir, Bayan Nur, Ulanqab, Xing'an League, Xilingol League, Alxa League

**Supplementary Table 3 | List of locked-down cities.** The lockdown information is from local government and various media news in 2020.

# Supplementary References

55. Ciencewicki, J. & Jaspers, I. Air Pollution and Respiratory Viral Infection. *Inhal. Toxicol.* 19, 1135–1146 (2007).
56. Diamond, G. et al. The innate immune response of the respiratory epithelium. *Immunol. Rev.* 173, 27–38 (2000).
57. Ghio, A. J. et al. Composition of Air Pollution Particles and Oxidative Stress in Cells, Tissues, and Living Systems. *J. Toxicol. Environ. Heal. Part B* 15, 1–21 (2012).
58. Li, Y.-F. et al. Associations of Tumor Necrosis Factor G-308A with Childhood Asthma and Wheezing. *Am. J. Respir. Crit. Care Med.* 173, 970–976 (2006).
59. Vawda, S. et al. Associations Between Inflammatory and Immune Response Genes and Adverse Respiratory Outcomes Following Exposure to Outdoor Air Pollution: A HuGE Systematic Review. *Am. J. Epidemiol.* 179, 432–442 (2014).
60. Tellier, R. et al. Recognition of aerosol transmission of infectious agents: a commentary. *BMC Infect. Dis.* 19, 101 (2019).
61. Tellier, R. Aerosol transmission of influenza A virus: a review of new studies. *J. R. Soc. Interface* 6, (2009).
62. Tellier, R. Review of Aerosol Transmission of Influenza A Virus. *Emerg. Infect. Dis.* 12, 1657–1662 (2006).
63. Escombe, A. R. et al. The Detection of Airborne Transmission of Tuberculosis from HIV-Infected Patients, Using an In Vivo Air Sampling Model. *Clin. Infect. Dis.* 44, 1349–1357 (2007).
64. Yu, I. T.-S. et al. Severe Acute Respiratory Syndrome Beyond Amoy Gardens: Completing the Incomplete Legacy. *Clin. Infect. Dis.* 58, 683–686 (2014).
65. Leclair, J. M. et al. Airborne Transmission of Chickenpox in a Hospital. *N. Engl. J. Med.* 302, 450–453 (1980).
66. Chen, P.-S. et al. Ambient Influenza and Avian Influenza Virus during Dust Storm Days and Background Days. *Environ. Health Perspect.* 118, 1211–1216 (2010).