Outdoor Air Pollutant Concentration and COVID-19 Infection in Wuhan, China

Yang Han[1], Jacqueline CK Lam[1,2,3*], Victor OK Li[1*], Peiyang Guo[1], Qi Zhang[1], Andong Wang[1], Jon Crowcroft[2], Illana Gozes[4], Jinqi Fu[5], Zafar Gilani[1], Shanshan Wang[1]

[1]Department of Electrical and Electronic Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong

[2]Department of Computer Science and Technology, The University of Cambridge, Cambridge, UK

[3]CEEPR, MIT Energy Initiative, MIT, Cambridge, Massachusetts, USA

[4]Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Adams Super Center for Brain Studies and Sagol School of Neuroscience, Tel Aviv University, Israel

[5]MRC Cancer Unit, Department of Oncology, The University of Cambridge, Cambridge, UK

*Corresponding authors. Email: jcklam@eee.hku.hk; vli@eee.hku.hk

Abstract

COVID-19 infection, first reported in Wuhan, China in December 2019, has become a global pandemic, causing significantly high infections and mortalities in Italy, the UK, the US, and other parts of the world. Based on the statistics reported by John Hopkins University, 4.7M people worldwide and 84,054 people in China have been confirmed positive and infected with COVID-19, as of 18 May 2020. Motivated by the previous studies which show that the exposures to air pollutants may increase the risk of influenza infection, our study examines if such exposures will also affect Covid-19 infection. To the best of our understanding, we are the first group in the world to rigorously explore the effects of outdoor air pollutant concentrations, meteorological conditions and their interactions, and lockdown interventions, on Covid-19 infection in China. Since the number of confirmed cases is likely to be under-reported due to the lack of testing capacity, the change in confirmed case definition, and the undiscovered and unreported asymptotic cases, we use the rate of change in the daily number of confirmed infection cases instead as our dependent variable. Even if the number of reported infections is under-reported, the rate of change will still accurately reflect the relative change in infection, provided that the trend of under-reporting remains the same. In addition, the rate of change in daily infection cases can be distorted by the government imposed public health interventions, including the lockdown policy, inter-city and

intra-city mobility, and the change in testing capacity and case definition. Hence, the effects of the lockdown policy and the inter-city and intra-city mobility, and the change in testing capacity and case definition are all taken into account in our statistical modelling. Furthermore, we adopt the generalized linear regression models covering both the Negative Binomial Regression and the Poisson Regression. These two regression models, when combined with different time-lags (to reflect the COVID-19 incubation period and delay due to official confirmation) in air pollutant exposure ($PM_{2.5}$), are used to fit the COVID-19 infection model. Our statistical study has shown that higher $PM_{2.5}$ concentration is significantly correlated with a higher rate of change in the daily number of confirmed infection cases in Wuhan, China ($p < 0.05$). We also determine that a higher dew point interacting with a higher $PM_{2.5}$ concentration is correlated with a higher rate of change in the daily number of confirmed infection cases, while a higher UV index and a higher $PM_{2.5}$ concentration are correlated with a lower rate of change. Furthermore, we find that $PM_{2.5}$ concentration eight days ago has the strongest predictive power for COVID-19 Infection. Our study bears significance to the understanding of the effect of air pollutant ($PM_{2.5}$) on COVID-19 infection, the interaction effects of both the air pollutant concentration ($PM_{2.5}$) and the meteorological conditions on the rate of change in infection, as well as the insights into whether lockdown should have an effect on COVID-19 infection.

Keywords: COVID-19, rate of change in infection rate, air pollution, dew point, ultraviolet index, inter-city and intra-city mobility, lockdown policy, testing capacity, infection case definition, public health, Wuhan, China

1. Introduction

COVID-19 infection was first reported in Wuhan, China in December 2019.[1] It has been declared a global pandemic by WHO and been transmitted to all parts of the world, causing significantly high infections and mortalities in Italy, the UK, the US, and other parts of the world. Based on the statistics reported by John Hopkins University (JHU), 4.7M people worldwide have been confirmed positive with COVID-19, and 84,054 people have been tested positive in China, as of 18 May 2020.

Previous studies on COVID-19 infection have examined a number of key factors, including demographics, meteorology, and lockdown measures, to determine whether they are strongly

correlated with COVID infection.[2–5] To the best of our understanding, this is the first rigorous study that investigates the effects of outdoor air pollution concentrations and lockdown on Covid-19 infections in China. Previous studies have suggested that the exposures to air pollutants with and/or without interacting with meteorological conditions may increase the risk of influenza infection.[6–8] Such relationships have also been observed in SARS and MERS.[9,10] More recently, research studies have suggested that meteorological conditions are associated with the spread of COVID-19.[4,11] In the US and Europe, the long-term air pollution exposure has been identified as a predictor of COVID-19 mortality.[12,13] Some studies have started hypothesizing that air pollution is a significant attribute to COVID-19 infection in China and Italy.[14–16] A recent study has suggested that air pollution is associated with COVID-19 infection after the lockdown has been exercised in Wuhan, China,[17] but has neither accounted for the effect of lockdown in Wuhan, the change in testing capacity and the inconsistency in COVID-19 confirmed case definition, nor the interactive effect of air pollutants concentration and meteorological conditions on COVID-19 infection. Their studies are still in the stage of infancy, lacking rigorous statistical modelling and control methodology for substantiation.

In this study, we will examine the statistical relationship between the outdoor air pollutant concentration ($PM_{2.5}$) and the rate of change in the daily number of COVID-19 confirmed infection cases in Wuhan, while accounting for any inconsistency in COVID-19 confirmed case definition, testing capacity, and the potential effect of lockdown policy and associated lock down indicators, such as inter-city and intra-city movements in Wuhan, China. Our study bears significance to the understanding of the effect of air pollutant ($PM_{2.5}$) on COVID-19 infection, the interaction effects of both the air pollutant concentration ($PM_{2.5}$) and the meteorological conditions on the rate of change in infection, as well as the insights concerning the lockdown policy on COVID-19 infection, based on the rate of change, instead of the actual daily number of COVID-19 confirmed cases reported in Wuhan, China, as the dependent variable.

2. Method

2.1 Unit of Analysis and Data Collection

Our study examines COVID-infection and its relation to air pollutants concentration in Wuhan from the period of 1 January to 20 March 2020. This was the period when COVID-19 was first announced officially in China, the lockdown measures were strictly exercised in Wuhan and other

parts of China, and the number of confirmed cases peaked and dropped subsequently. We examine the relationship between air pollutant concentration ($PM_{2.5}$) and lockdown vs. COVID-19 Infection in Wuhan. Data is collected on a daily basis at the city level. Our data are sourced from the following websites: Daily confirmed COVID-19 cases are collected from a popular online platform which aggregates the cases reported by the Chinese national/provincial health authorities.[18] The air pollutants concentration data are collected from the Chinese National Environmental Monitoring Center.[19] The meteorological data, including the temperature, the dew point (the temperature to achieve a relative humidity of 100%),[20] the ultraviolet (UV) index, the precipitation (including rain and snow), and the wind speed, are collected from the US National Climatic Data Center and a weather data API owned by Apple, Inc.[21,22] The mobility data, including the inter- and the intra-city movements index, are collected from Baidu, Inc.[23]

2.2 Statistical Analysis

As suggested in previous studies, the number of confirmed cases is likely to be under-reported due to the lack of testing capacity, the change in confirmed case definition, and the undiscovered and unreported asymptotic cases.[5,24,25] Hence, we adjust the confirmed infection case data as follows: First, since the number of reported infections might not be reliable, we use the rate of change in our statistical analysis, i.e., the rate of change in the daily number of confirmed cases as compared to that of the previous day (Eq (1)), to reflect the relative variation in COVID-19 infection during the study period. Even if the number of reported infections is incorrect (under-reported), the rate of change will still accurately reflect the relative change in infection, provided the trend of under-reporting remains the same. However, the rate of change can still be distorted by China's public health interventions such as the lockdown policy and the change in testing capacity.[26] Therefore, the effects of both the lockdown policy and the change in testing capacity are taken into account in our statistical analysis. In addition, the change in confirmed case definition can distort the actual epidemic curve.[25] Hence, in our infection case modelling, we take into account the most significant discontinuity in the infection curve contributed by the change in definition on confirmed case during the study period.

Rate of Change in the Daily Number of Confirmed Infection Cases = (The Daily Number of Confirmed Infection Cases in Day *t* – The Daily Number of Confirmed Infection Cases in Day *t*-1) / The Daily Number of Daily Confirmed Infection Cases in Day *t*-1 Eq (1)

Regression models are constructed to examine the relationship between the outdoor air pollution concentration and the COVID-19 Infection. The dependent and independent variables are listed as follows.

COVID-19 Infection: We use an adjusted dependent variable, namely, the rate of change in the daily number of confirmed infection cases.

Air Pollution Concentration: Due to the incubation period and the delay in reporting before the symptom onset and the confirmation, we must account for the corresponding time-lag in the air pollutant concentration ($PM_{2.5}$) examined in our statistical models, based on two exposure assessment methods, which are deployed for air pollution-related influenza studies.[6–8] First, we collect $PM_{2.5}$ Pollutant Concentration *t* days prior to the COVID-19 Infection. Second, the daily $PM_{2.5}$ Pollutant Concentration are averaged over *t* days prior to the COVID-19 Infection. In our analysis, the lag-time *t* ranges from Day 1 to Day 12, assuming that the mean incubation period and mean reporting delay are 12 days in total.[25]

Meteorological Condition: Daily meteorological conditions are examined, including the temperature, the dew point, the UV index, the precipitation, and the wind speed. In order to account for the interaction between the air pollutant concentration ($PM_{2.5}$) and the meteorology, for each weather condition variable, an interaction term is included in the statistical modelling.

Lockdown Policy: It is represented as a binary variable to indicate whether the city lockdown measure is imposed on a particular day or not.

Inter-city and Intra-city Mobility: The effect of lockdown in Wuhan can be measured by three mobility indexes, representing the number of people moving into the city, moving out of the city, and moving within the city. The details of how the three indexes are defined are not publicly

available from Baidu, Inc. Therefore, for data consistency, these indexes are normalized in the scale of 0 to 1. Moreover, to capture the relative change in mobility during the Chinese New Year (25 January 2020), for each index, we deduct the index recorded on one date in 2019 (the mobility from the index recorded on the same date in 2020, based on the number of days before and after the Chinese New Year, which, traditionally, is the period of intensive movement in China).

Dummy Variable: Two dummy variables are introduced to account for the change in COVID-19 testing capacity and the definition on COVID-19 confirmed infection case definition in Wuhan.

The concept model is shown in Eq (2). Linear Regression models are selected for modelling the relationship between the Rate of Change in COVID-19 Infection Case and the Air Pollutant Concentration ($PM_{2.5}$). In addition, following the previous research capturing the relationship between air pollutant exposure and influenza, we adopt the generalized linear regression models covering both the negative Binomial Regression and the Poisson Regression.[6–8] These regression models, combined with different time-lags in air pollutant exposure ($PM_{2.5}$), are used to fit the COVID-19 data. In total, 100 regression models have been constructed (see Table 1 for the list of regression models). The best-fit model is selected based on Akaike Information Criterion (AIC).[6,7]

Rate of Change in the Daily Number of Confirmed Infection Cases = Intercept + Air Pollutant Concentration + Air Pollutant Concentration* Meteorology Condition + Lockdown Policy + Inter-city and Intra-city Mobility + Change in Testing Capacity + Change in Infection Case Definition + Error Term Eq (2)

Table 1. The list of regression models that are used to fit the confirmed infection cases data

Regression	Dependent Variable	Air Pollutant Exposure	Number of Models
Linear	Daily Number of Confirmed Infection Cases Rate of Change in the Daily Number of Confirmed Infection Cases	Current day Previous 1-12 days Averaged over previous 1-12 days	2 x (1+12+12) = 50

Negative Binomial	Daily Number of Confirmed Infection Cases	Current day Previous 1-12 days Averaged over previous 1-12 days	1 x (1+12+12) = 25
Poisson	Daily Number of Confirmed Infection Cases	Current day Previous 1-12 days Averaged over previous 1-12 days	1 x (1+12+12) = 25
Notes 1. The Rate of Change is not used in the Negative Binomial Regression or Poisson Regression because these models assume that the dependent variable is positive.			

3. Result

The cumulative number of confirmed cases and the Air Pollutant Concentration ($PM_{2.5}$) at the daily level in Wuhan, China, are shown in Figure 1. The level of $PM_{2.5}$ was on a decreasing trend after the city had entered into a complete lockdown from 23 January 2020 onwards. The mean $PM_{2.5}$ Pollutant Concentration was 63 $\mu g/m^3$ and 39 $\mu g/m^3$, before and after the city's lockdown, respectively. In addition, the case definition change exercised on 12 February 2020 (as indicated by the vertical line in Figure 1) led to a significant discontinuity in Wuhan's epidemic curve. Linear Regression, Negative Binomial Regression, and Poisson Regression models were used to fit the confirmed cases. Different lags of $PM_{2.5}$ pollutant exposure are tested. The results of the best-fit regression model are shown in Table 2.

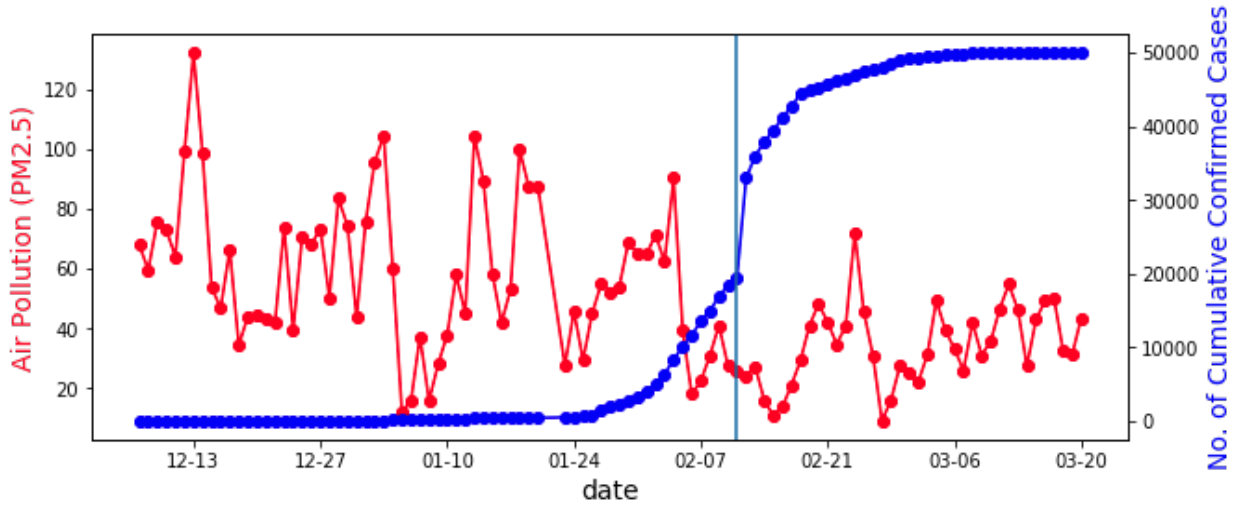

Figure 1. The cumulative number of confirmed cases vs. the air pollutant concentration ($PM_{2.5}$) in Wuhan, China

Table 2. Significant factors on COVID-19 infection in Wuhan (from 1 January to 20 March, 2020)

Dependent Variable: The Rate of Change in the Daily Number of Confirmed Infection Cases in Wuhan		
Regression Model: Linear Regression		
Variable	**Coefficient**	***p*-value**
Intercept	-4.971	0.365
$PM_{2.5}$ (t = 8 days)	0.146	0.031**
$PM_{2.5}$ * Temperature	-0.011	0.280
$PM_{2.5}$ * Dew Point	0.016	0.079*
$PM_{2.5}$ * UV Index	-0.023	0.061*
$PM_{2.5}$ * Precipitation	-0.000	0.846
$PM_{2.5}$ * Wind Speed	-0.011	0.376
Lockdown Policy	-0.657	0.854
Move-in Index	-2.631	0.320
Move-out Index	-4.007	0.194

Intra-city Movement Index	3.866	0.216
Change in Testing Capacity on 19 January 2020	0.153	0.891
Change in Testing Capacity on 12 February 2020	1.002	0.291
Notes 1. Air pollutant exposure is lagged by *t* days. 2. **p*-value < 0.1, ***p*-value < 0.05, ****p*-value < 0.01		

The statistical relationship between the air pollutant concentration ($PM_{2.5}$) and the rate of change in the daily number of confirmed infection cases is statistically significant and positive ($p < 0.05$). This suggests that a higher $PM_{2.5}$ concentration is associated with a higher rate of change in the daily number of confirmed infection cases in Wuhan. The interaction between the $PM_{2.5}$ concentration and the dew point and the interaction between $PM_{2.5}$ concentration and the UV index, are both significant ($p < 0.1$). A higher dew point and a higher $PM_{2.5}$ concentration are correlated with a higher rate of change in the daily number of confirmed infection cases, while a higher UV index and a higher $PM_{2.5}$ concentration are correlated with a lower rate of change. Furthermore, as compared to the average $PM_{2.5}$ concentration over multiple days, $PM_{2.5}$ concentration of a particular day has a stronger predictive power for COVID-19 Infection. The best-fit lagged time in $PM_{2.5}$ concentration is eight days. Based on this model, lockdown policy, and inter-city and intra-city mobility have no significant statistical correlation with the rate of change in COVID-19 infection cases in Wuhan.

Reference

1 Li Q, Guan X, Wu P, *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine* 2020; **382**: 1199–207.

2 Chen N, Zhou M, Dong X, *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 2020; **395**: 507–13.

3 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine* 2020; **382**: 1708–20.

4 Liu J, Zhou J, Yao J, *et al.* Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. *Science of The Total Environment* 2020; **726**: 138513.

5 Kraemer MUG, Yang C-H, Gutierrez B, *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020; **368**: 493.

6 Feng C, Li J, Sun W, Zhang Y, Wang Q. Impact of ambient fine particulate matter (PM2.5) exposure on the risk of influenza-like-illness: a time-series analysis in Beijing, China. *Environmental Health* 2016; **15**: 17.

7 Su W, Wu X, Geng X, Zhao X, Liu Q, Liu T. The short-term effects of air pollutants on influenza-like illness in Jinan, China. *BMC Public Health* 2019; **19**: 1319.

8 Chen G, Zhang W, Li S, *et al.* The impact of ambient fine particles on influenza transmission and the modification effects of temperature in China: A multi-city study. *Environment International* 2017; **98**: 82–8.

9 Cui Y, Zhang Z-F, Froines J, *et al.* Air pollution and case fatality of SARS in the People's Republic of China: an ecologic study. *Environmental Health* 2003; **2**: 15.

10 Gardner EG, Kelton D, Poljak Z, Van Kerkhove M, von Dobschuetz S, Greer AL. A case-crossover analysis of the impact of weather on primary cases of Middle East respiratory syndrome. *BMC Infectious Diseases* 2019; **19**: 113.

11 Bu J, Peng D-D, Xiao H, *et al.* Analysis of meteorological conditions and prediction of epidemic trend of 2019-nCoV infection in 2020. *medRxiv* 2020; : 2020.02.13.20022715.

12 Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. *medRxiv* 2020; : 2020.04.05.20054502.

13 Ogen Y. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. *Science of The Total Environment* 2020; **726**: 138605.

14 Han Y, Lam JCK, Li VOK, *et al.* The Effects of Outdoor Air Pollution Concentrations and Lockdowns on Covid-19 Infections in Wuhan and Other Provincial Capitals in China. 2020; published online March 24. DOI:10.20944/preprints202003.0364.v1.

15 Conticini E, Frediani B, Caro D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environmental Pollution* 2020; **261**: 114465.

16 Contini D, Costabile F. Does Air Pollution Influence COVID-19 Outbreaks? *Atmosphere* 2020; **11**: 377.

17 Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. *Science of The Total Environment* 2020; **727**: 138704.

18 Xinxing Guanzhuang Bingdu Feiyan Yiqing Shishi Dongtai [Novel Coronavirus Pneumonia Real-Time Epidemic Information]. https://ncov.dxy.cn/ncovh5/view/pneumonia (accessed May 13, 2020).

19 Quanguo Kongqi Zhiliang Yubao Xinxi Fabu Xitong [China National Air Quality Reporting System]. http://106.37.208.228:8082/ (accessed May 13, 2020).

20 National Weather Service, NOAA. Dew Point vs Humidity. https://www.weather.gov/arx/why_dewpoint_vs_humidity (accessed May 13, 2020).

21 Data Access | National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC). https://www.ncdc.noaa.gov/isd/data-access (accessed May 13, 2020).

22 Dark Sky API. https://darksky.net/dev (accessed May 13, 2020).

23 Baidu Qianxi [Baidu Mobility Map]. https://qianxi.baidu.com/ (accessed May 13, 2020).

24 Li R, Pei S, Chen B, *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 2020; **368**: 489–93.

25 Tsang TK, Wu P, Lin Y, Lau EHY, Leung GM, Cowling BJ. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *The Lancet Public Health* 2020; **5**: e289–96.

26 Omori R, Mizumoto K, Chowell G. Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. *International Journal of Infectious Diseases* 2020; **94**: 116–8.