medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Outdoor $PM_{2.5}$ Concentration and Rate of Change in COVID-19 Infection in 31 Provincial Capital Cities in China

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

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

[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]Department of Computer Science, Queen Mary University, London, UK

*Corresponding authors. Address: Department of Electrical and Electronic Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong. Email: jcklam@eee.hku.hk; vli@eee.hku.hk Tel: +85239174843; +85239178425

[#]The first three authors have equal contributions.

Abstract

Background: Motivated by the findings that exposure to daily outdoor $PM_{2.5}$ (P) may increase the risk of influenza infection, our study examines if immediate exposure to outdoor P will modify the rate of change in the daily number of COVID-19 infections (R), for (1) 31 Chinese provincial capital cities and (2) Wuhan, China, using regression modelling. Effective public health measures for reducing R are recommended.

Method: A multiple linear regression model was constructed to model the statistical relationship between P and R in China and in Wuhan, from 1 January to 20 March 2020. We carefully accounted for potential key confounders and addressed collinearity. The causal relationship between P and R, and the interaction effect between key variables were investigated.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Finding: A causal relationship between P and R across 31 provincial capital cities in China was established via matching. A higher P resulted in a higher R in China. A 10 μg/m^3 increase in P gave a 0.9% increase in R ($p < 0.05$). An interaction analysis between P and absolute humidity (AH) showed a statistically significant positive relationship between P × AH and R ($p < 0.01$). When AH was ≤ 8.6 g/m^3, higher P and AH gave a higher R ($p < 0.01$).

Interpretation: Given that P can exacerbate R, we recommend the installation of air purifiers and improving air ventilation to reduce the effect of P on R. Given that an increasing AH that falls below 8.6 g/m^3 exacerbated R, dehumidifiers can be used to reduce AH and R. Further, given the increasing discussions/observations that COVID-19 can be airborne, surgical masks should be used to protect one from contracting the virus via the viral-particulate transmission pathway.

Keywords: COVID-19, rate of change in the daily number of confirmed infection cases, $PM_{2.5}$, absolute humidity, net move-in mobility, lockdown policy, testing capacity, infection case definition, public health, China

1. Introduction

COVID-19 was first reported in Wuhan, China in December 2019. It is a global viral pandemic that has resulted in more than 23-million infections and 800,000 deaths across different parts of the world, including 89,549 reported cases in China.

Recent COVID-19 studies investigated whether demography (D), co-morbidity (CM), meteorology, and lockdown generated any significant statistical effects on the viral infection.[1–3] Evidence has, therefore, been presented to suggest that weather effects producing low temperature and humidity is associated with COVID-19 transmission.[2] This is consistent with earlier epidemiological studies that implicated air pollution (PM_{10} as the primary pollutant) and meteorological effects with changes in SARS/MERS infection.[4,5] Furthermore, influenza studies have suggested that exposure to $PM_{2.5}$ (P) with and without interacting with meteorology may increase the risks of flu infection.[6] In the US and Europe, long-term exposures to P and NO_2 were reported as the predictors of COVID-19 mortality.[7,8] An earlier study reported that air pollutants, such as particulates presented independent risks to Vitamin-D deficiency and could be an important factor driving the severity of infection.[9] Recently, increasing evidence in China and Italy showed that air pollution contributes significantly to cases of COVID-19 infection.[10,11] The

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

study conducted in China concluded that P, NO_2 and O_3 were associated with cases of COVID-19,[12] though it failed to fully account for the change in testing capacity and the inconsistency in COVID-19 confirmed case definition, and the confounding effect of D and CM. To investigate whether environmental P exposure will affect COVID-19 infection, more rigorous statistical modelling and control methodologies that reduce the confounding biases and remove the collinearities across different independent variables are needed.

In the following, we examine the immediate effect of outdoor P on the rate of change in the daily number of COVID-19 confirmed infections (R), across 31 provincial capital cities in China, while seeking to address inadequacies in official case reporting, and taking into account as many confounders as possible, including D, CM, meteorology, net move-in mobility (NM), time-lag in case reporting due to testing capacity (TC) and inconsistency in case definition (CD), trends over time (T), and day-of-the-week (DOW) to reflect the recurrent weekly effect.

Outdoor P is chosen as the focus of our study given the assumption that R may be increased due to the potential deposition of viral droplets on P.[13] A recent rigorous study on COVID-19 aerodynamics ascertained that viral aerosol droplets 0.25-1 μm in size can remain suspended in air.[14] When such viral droplets are combined with suspending particles, P, they can travel greater distances, remain viable in the air for hours, and be inhaled deeply into the lungs, thus increasing the potential of airborne viral infection.[15]

Our study sheds new light on the effect of P in an outdoor environment, the interaction effect between P and meteorological effects such as absolute humidity (AH), and the effect of NM (lockdown), on R (the dependent variable). Our work adds weights to the recent discussions/observations that COVID-19 droplets are airborne[14,16], can suspend in the air and combine with the particulates, promoting infection via the airborne transmission pathway.[17]

2. Method

Data Collection and Procedure

We collected data covering the daily P and the daily number of confirmed infections of 31 provincial capital cities in China, covering the period from 1 January to 20 March 2020. This was the period when COVID-19 infection was first officially announced in China, the lockdown measures were strictly exercised in Wuhan and other parts of China, and the number of confirmed cases peaked and dropped. Other data at the provincial city-level were also collected

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

on a daily basis (including meteorology and NM) or on a yearly basis (including D and CM) from internet sources and official statistical documents. Table 1 highlights our research objectives and procedures.

Table 1. Research objectives and procedures

Primary Objective	• Explore the statistical relationship, and determine the causal effects, if any, between daily outdoor P and R across 31 provincial capital cities in China, including Wuhan, the origin of COVID-19 outbreak.
	• To achieve this objective, we built two statistical models that can best address the following challenges in statistical modelling:
	  ○ Redefinitions and potential inadequacies in infection case reporting
	  ○ Non-linearity
	  ○ Collinearity
	  ○ Confounders and confounding biases
	  ○ Interaction between statistically significant factors

Secondary Objective	• Highlight the conditions under which R can be reduced, and effective public health measures that can be employed to facilitate this. • Add weight to the current observations that COVID-19 can be airborne and that particulates can be carriers of the viral droplets.

Statistical Analysis

A final multiple linear regression model including only the statistically significant variables was constructed to model the relationship between daily outdoor P and daily R across 31 provincial capital cities in China (see Eq (1)), and in Wuhan, China (see Eq (2)), while taking into account potential confounders, in order to determine if any causal relationship exists between the two variables (see Table 2). In addition, the interaction between the key statistically significant variables and their effect on R was also investigated via a new regression model (see Table 3).

$$\mathrm{Log}(R_{c,t}) = \alpha + \beta_1 * \mathrm{Log}(R_{c,t-1}) + \beta_2 * P_{c,t-L} + \beta_3 * AH_{c,t-L} + \beta_4 * AH^2_{c,t-L} + \beta_5 * \mathrm{Age}(\text{0-14 years old})_c + \beta_6 * T_t + \varepsilon \quad Eq\ (1)$$

$$\mathrm{Log}(R_{c,t}) = \alpha + \beta_1 * \mathrm{Log}(R_{c,t-1}) + \beta_2 * P_{c,t-L} + \beta_3 * AH_{c,t-L} + \beta_4 * CD_t + \varepsilon \quad Eq\ (2)$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

where α is the intercept, the subscript *c* denotes a city, subscript *t* denotes a day, subscript *L* denotes the time lag for P and AH, and ε serves as the error term. *L* ranges from one to fourteen days. R denotes the rate of change in the daily number of confirmed COVID-19 infections. P denotes the $PM_{2.5}$ concentration. AH denotes the absolute humidity. Age (0-14 years old) represents the percentage of population with age below 15 years old at the provincial city-level. T denotes the time trend fixed effect (such as variation of lockdown effects over time). CD is a dummy binary variables to reflect the change in the confirmed case definition[18] during the period of study.

Earlier COVID-19 studies had expressed reservations concerning the number of infection cases reported, given inadequate TC, the change in confirmed CD, and undiscovered and undocumented asymptomatic cases.[3,18,19] Apart from adding two dummy variables, TC and CD, to address the delay in testing capacity and the change in case definition and their effects on reported cases, we used R, rate of change, as the dependent variable, in order to capture the relative change in COVID-19 infection during the study period. By using R, even if the number of reported infections might deviate, the relative change in infection could still be accounted for, provided that the reporting trends remain consistent. Further, to account for the skewed distribution of R, a log transformation was applied to the linear regression. A one day time-lag variable representing the R of the previous day was also included in the model as an autoregressive term to account for the temporal auto-correlation among R time-series.

To estimate the causal effect of P on R, our model for China and Wuhan had to cover the potential confounders. Independent variables, including meteorology (AH, temperature (TEMP), ultraviolet (UV) index, air pressure (AP), and wind speed (WS)) and NM, were included in the model for China and Wuhan as the confounders. Moreover, D (population density, age, sex, income, and education) and CM (high blood pressure, diabetes, chronic obstructive pulmonary disease (COPD), stroke, and obesity) were included in the model for China to control for city-level fixed effects. T and day of week were included in the model for China and Wuhan to control for the time-varying fixed effects and the recurrent fixed effects. Furthermore, matching was adopted to further reduce the confounding biases, by matching a high P day with a low P day, based on the similarities of corresponding confounders, thereby helping one more accurately estimate the causal relationship between P and R in China.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

To account for the collinearity in our model for China and Wuhan, a Spearman correlation was conducted to remove meteorological and CM variables that present collinearity. We tested the collinearity between TEMP, UV index, AP, WS, and AH, and removed all of them, except WS, due to their high collinearity with AH, which was able to account for the transmission of a flu virus, and hence may also be used to account for R[20] (Spearman coefficient > 0.5). WS remained in our model due to its low collinearity with AH. Similarly, we tested the collinearity between CM variables, diabetes, stroke, obesity, COPD, and high blood pressure. We found that all CM variables, except for COPD, correlated highly with high blood pressure, a more common CM identified from recent COVID infection cases, and might account for R.[21] These (except for COPD) were removed, due to their high collinearity with high blood pressure (Spearman coefficient > 0.5). COPD was retained in our model as it had low collinearity with high blood pressure.

To account for the non-linear relationship between the meteorological variables and R in our model for China and Wuhan, a non-linear transformation was applied to the meteorological variables, including AH and WS. Two transformation functions, a second order polynomial function and a natural spline function with two degrees of freedom, were attempted to address non-linearity, based on the goodness of fit. From this it was determined that the second order polynomial function provided the best fit and was adopted to address the non-linear relationship between AH/WS with R.

Due to the lengthy asymptomatic incubation period and the delay in case reporting before the onset and confirmation of COVID-19 symptoms, the corresponding time-lag in P, meteorology, and NM was accounted for by our statistical model for both China and Wuhan, namely, the single-day lag and the multi-day average lag models, based on previous air-pollution related epidemiological studies.[6] We determined the best fit lag-time from day 1 to day 14, with the assumption that the mean incubation period and the mean delay in official case reporting could cover a maximum of 14 days.[18]

Finally, to investigate whether there was any interaction between P and AH on R across 31 provincial capital cities in China during the infection cycle, a new regression model was developed for China (Eq (3)), based on the statistically significant factors that associate with R (see Table 2).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

$$\mathrm{Log}(R_{c,t}) = \alpha + \beta_1 * \mathrm{Log}(R_{c,t-1}) + \beta_2 * P_{c,t-L} + \beta_3 * AH_{c,t-L} + \beta_4 * \mathrm{Age}(0\text{-}14\ \mathrm{years\ old})_c + \beta_5 * T_t + \beta_6 * P_{c,t-L} * AH_{c,t-L} + \varepsilon \quad Eq\ (3)$$

Role of the funding source

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all data in the study and held the final responsibility for the decision to submit for publication.

3. Results

Based on two best-fit regression models (below), the results of the statistically significant factors that associate with R across the 31 provincial capital cities in China and in Wuhan are shown in Table 2(a) and 2(b), respectively. No collinearity was identified in the best-fit regression models via variance inflation factor (VIF) analysis. In order to better illustrate the relationship between P/AH and R, the conditional plots of P and AH are shown in Figure 1. For conditional plots, when examining the relationship between the P/AH and the predicted R, the median values of the remaining independent variables were used.

Table 2. Statistically significant factors that associate with R (a) across all 31 provincial capital cities in China and (b) in Wuhan (from 1 January to 20 March 2020).

(a) 31 Provincial Capital Cities in China			**(b) Wuhan Only**		
Dependent Variable: Log(R_t)			**Dependent Variable:** Log(R_t)		
Independent Variables	**Coefficient**	***p*-value**	**Independent Variables**	**Coefficient**	***p*-value**
Intercept	1.2568	0.0000 ****	Intercept	-0.6724	0.185
Log(R_{t-1})	-0.3025	0.0000 ****	Log(R_{t-1})	-0.2776	0.020 **
P_{t-L}	0.0009	0.036 **	P_{t-L}	0.0261	0.0041 ***

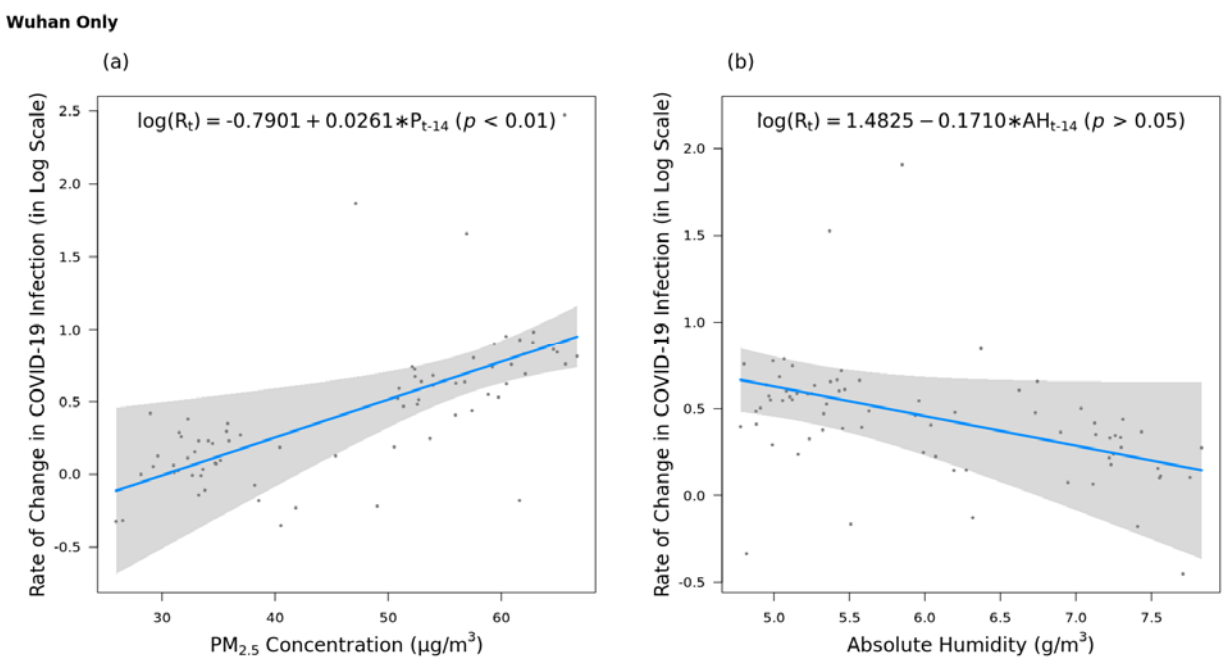
medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

AH_{t-L}	1.1046	0.024 **	AH_{t-L}	-1.4330	0.063
AH^2_{t-L}	-1.1195	0.0081 ***	CD_t	0.7174	0.0089 ***
Age (0-14 years old)	-1.2416	0.019 **			
T_t	-0.0069	0.0000 ****			

Notes

1. For (a) 31 provincial capital cities in China, P and AH were lagged by L=9 days. For (b) Wuhan, P and AH were lagged and averaged by L=14 days.
2. For (b) Wuhan, D and CM were not included in the regression model since they were constant within the city.
3. ***p*-value < 0.05, ****p*-value < 0.01, **** *p*-value < 0.001



medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

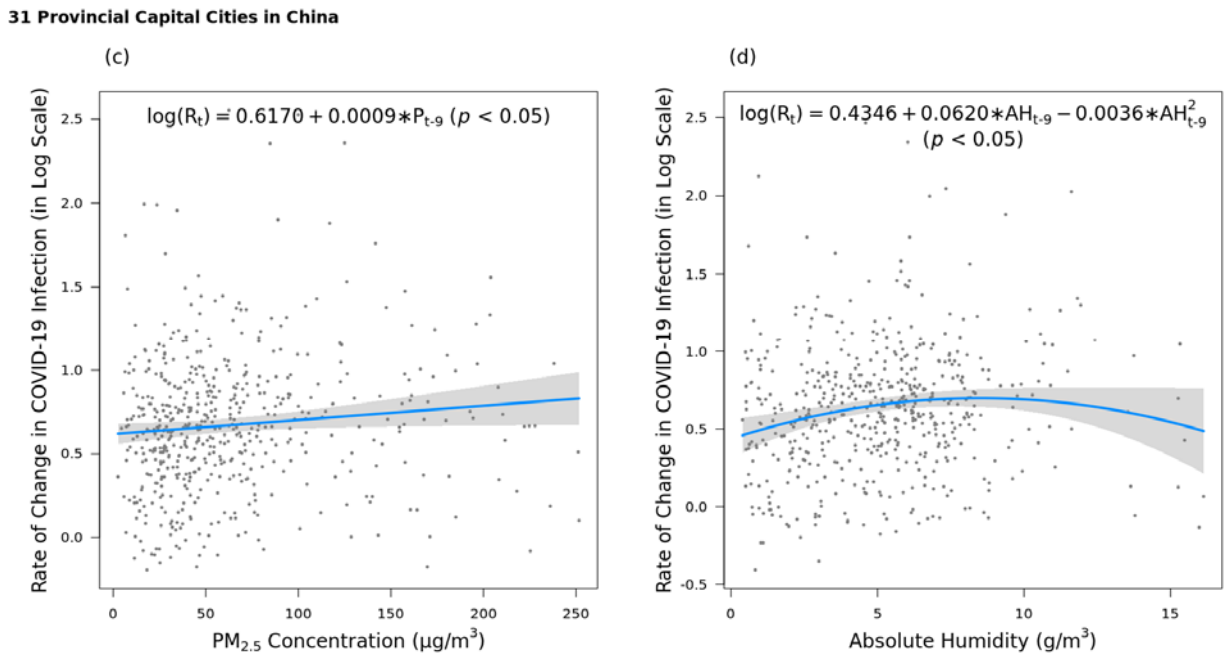


Figure 1. The statistical relationship between P/AH vs. R in Wuhan, China (Figure 1(a) and 1(b)) and across the 31 provincial capital cities in China (Figure 1(c) and 1(d))

As shown in Table 2 (a), there was a positive, statistically significant, relationship between P and AH with R across 31 provincial capital cities in China ($p < 0.05$). A higher P was associated with a higher R in China (see Figure 1 (c)). A 10 μg/m^3 increase in P was associated with a 0.9% increase in R. Moreover, AH with its second order term was a significant factor for accounting R in China ($p < 0.05$). As shown in Figure 1 (d), an inverted U-shape is observed; a higher AH increased R when the value of AH ≤ 8.6 g/m^3, while a higher AH decreased R when the value of AH > 8.6 g/m^3. As compared to the average P over multiple days, P of a single day had a stronger explanatory power for R. The best-fit day-lag for P was nine. Based on the regression model for the 31 provincial capital cities in China, T had a significant statistical correlation with R ($p < 0.001$). A decrease in R was observed along with the increase in T. Cities having a higher proportion of younger population (0-14 years old) tended to give a lower R ($p < 0.01$).

When Wuhan, the city with the greatest number of confirmed cases during the period of study, was included in our analysis, similar results were observed. As shown in Table 2, and Figure 1, the effects of P on R in Wuhan and in all cities were similar. However, there was a stronger correlation between P and R (Coefficient = 0.0261, $p < 0.01$) in Wuhan as compared to all cities in China (Coefficient = 0.0009, $p < 0.05$); while the lag time was fourteen days instead of nine; unlike China, both AH and AH in the second form, was not statistically associated with R in a

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

log-linear relationship ($p > 0.05$), while CD was a significant factor that associated with R in Wuhan ($p < 0.01$). Compared to all cities in China, the effect of P on R in Wuhan was higher. A 10 μg/m^3 increase in P was associated with a 2.6% increase in R in Wuhan ($p < 0.01$). Moreover, the change in confirmed case definition increased R in Wuhan ($p < 0.01$), suggesting that our model has properly addressed any inadequacy due to changes in case definition in Wuhan.

Apart from this, the causal effect of P on R was established via matching, by addressing the confounding biases. The result was consistent with our main findings. On average, across 31 provincial capital cities, days with higher P (≥ 40 μg/m^3) resulted in a 15.1% increase in R compared to days with lower P (< 40 μg/m^3). Furthermore, the interaction between P and AH was significant across 31 provincial capital cities in China. When a higher P interacts with a higher AH, a higher R was generally observed (Coefficient = 0.0005, $p < 0.01$, see Table 3). Taking a closer look, when AH ≤ 8.6g/m^3, higher P and AH gave a higher R. When AH > 8.6 g/m^3, higher P and AH no longer resulted in a higher R. The effect of a higher P on R (in increasing trend) was counteracted by the effect of a higher AH on R (in decreasing trend), though the interaction effect of higher P and AH was not statistically significant ($p > 0.05$, see Figures 2 (e) – 2 (f)). When only Wuhan was included in our interaction analysis, the interaction effect of P and AH was not statistically significant ($p > 0.05$).

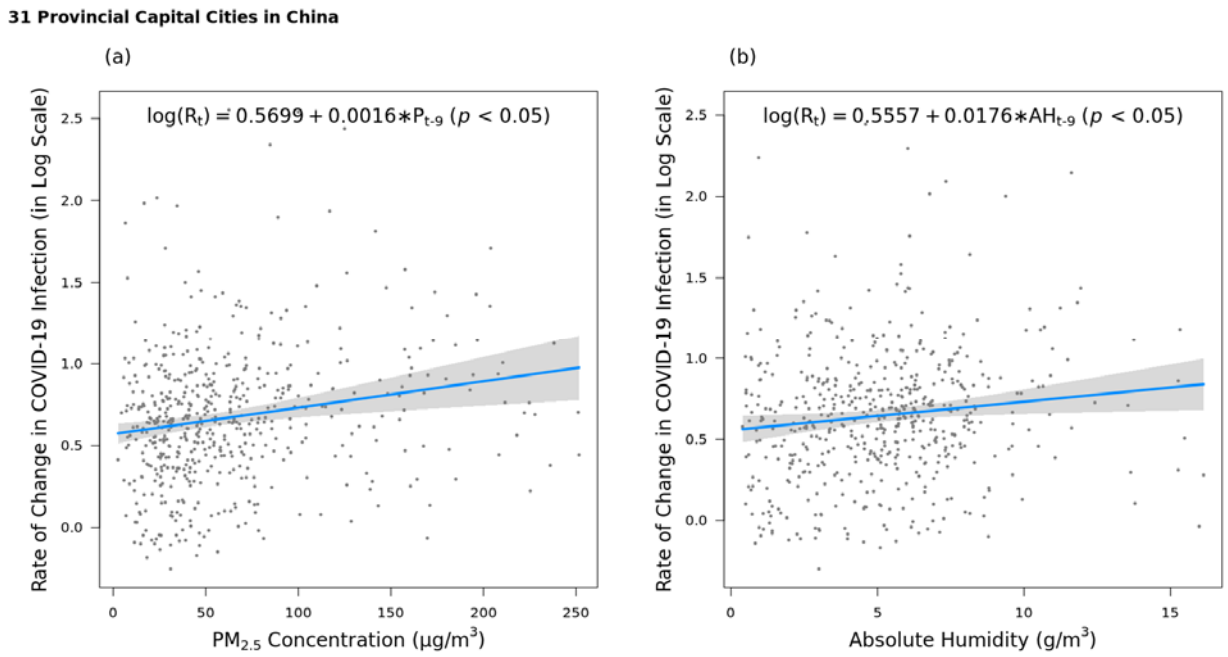
Table 3. Statistically significant factors covering P × AH that associate with R across all 31 provincial capital cities in China (from 1 January to 20 March 2020)

31 Provincial Capital Cities in China		
Dependent Variable: Log(R_t)		
Independent Variables	**Coefficient**	***p*-value**
Intercept	1.3068	0.0000 ****
Log(R_{t-1})	-0.3015	0.0000 ****

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

$P_{t\text{-}L}$	0.0016	0.0012 ***
AH_{t-L}	0.0255	0.0017 ***
Age (0-14 years old)	-1.1028	0.035 **
T_t	-0.0066	0.0000 ****
$P_{t\text{-}L} \times AH_{t-L}$	0.0005	0.0024 ***
Notes 1. P and AH were lagged by L=9 days. 2. ***p*-value < 0.05, ****p*-value < 0.01, **** *p*-value < 0.001		



medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

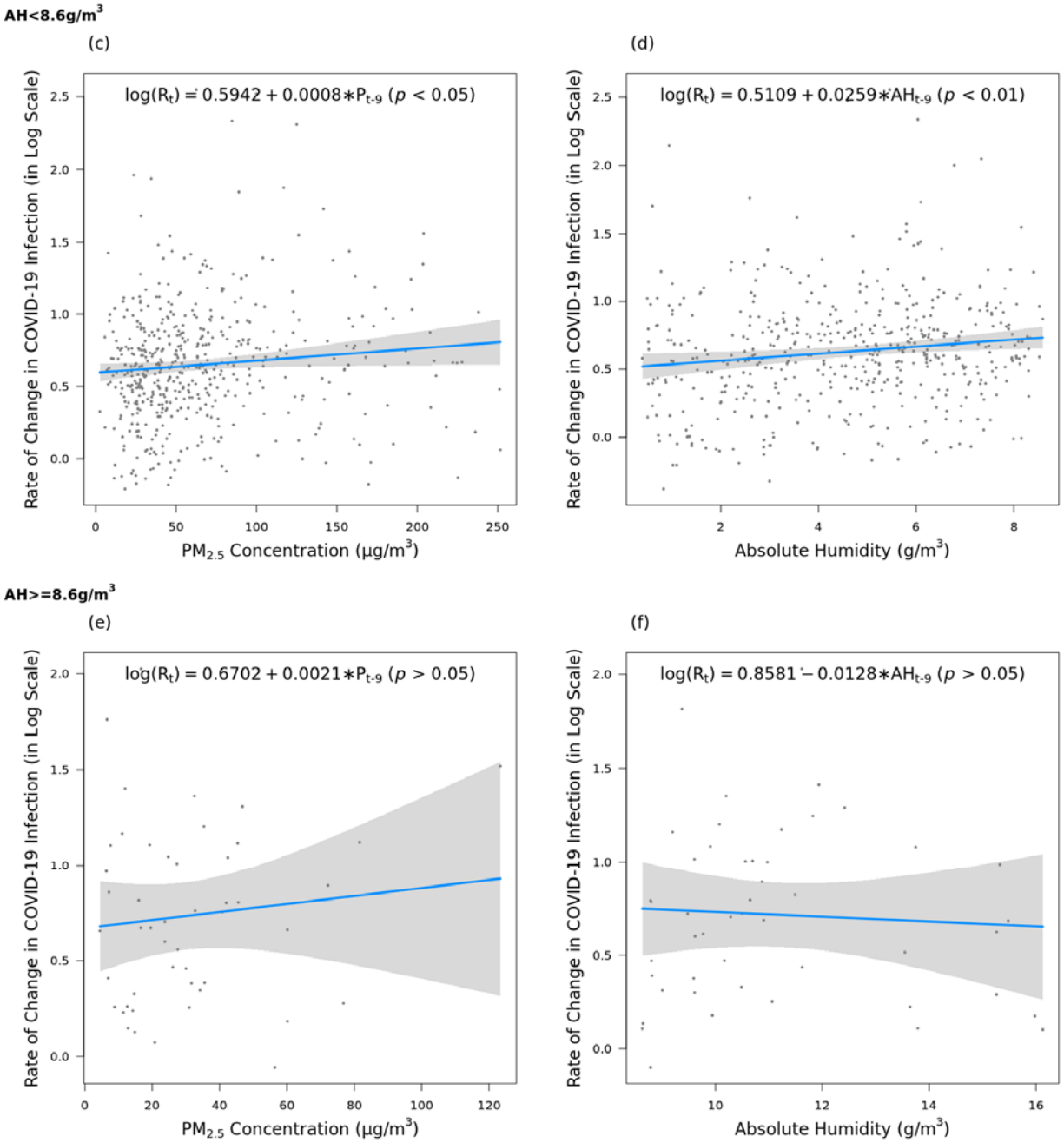


Figure 2. Statistical relationship between P/AH/P x AH vs. R across 31 provincial capital cities in China

4. Discussion

Recent COVID-19 studies investigated whether D, CM, meteorology, and lockdown were associated with significant statistical effects on viral infection. Some studies ascertained that meteorological effects demonstrate an association with COVID-19 transmission.[2] Earlier influenza studies suggested that exposure to P with and without interactions with meteorological

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

effects might increase the risks of flu infection. Earlier epidemiological studies also found that the air pollution index (measuring PM_{10} as the primary pollutant) and meteorological effects associated with SARS/MERS. In the US and Europe, long-term exposures to P and to NO_2 were reported to be predictors of COVID-19 mortality. Recently, increasing evidence in China and Italy argued that air pollution was a significant attributing factor to the rate of COVID-19 infection. A previous study conducted in China concluded that P was associated with COVID-19 infection, though it had not comprehensively accounted for the change in TC and the inadequacy in COVID-19 confirmed case definition, and the confounding effect of D and CM. Recent scientific studies also pointed towards the significant potential for COVID-19 to be transmitted via airborne routes.[22]

To identify whether P affects R across 31 provincial cities in China, including Wuhan, our regression model took into account all possible confounders, including meteorological variables, NM, D at the provincial city-level, and CM at the provincial level, including five major diseases that can potentially decrease immunities and increase the risk of COVID-19 infection.[21,23] Further, to reflect potential changes in TC and CD, two dummy variables were added to the regression model. In addition, the single day time-lag effect on P, meteorology, and NM, were also addressed.

Our model outperforms other existing air-pollution related COVID-19 epidemiological studies in four ways. First, instead of observing the absolute number of infections, which can be inadequate due to possible human or systemic deficiencies, in relation to testing methods and changes in case definition, our study examines R. Hence, even if the number of reported infections may be inadequate, R can sufficiently reflect the relative change in infection numbers, provided that the reporting trends are consistent. Therefore, our approach is able to provide greater insight than previous studies on air pollution[12] and other COVID-19 epidemiological or modelling studies,[18,24] which focussed on the absolute number of infections instead of R.

Secondly, our study has addressed a wide spectrum of possible confounders that may affect observations concerning the effect of P on R, including meteorological, NM, D, and key CM variables. This stands in contrast to existing work that has explored the effect of air pollution on COVID-19 infection/mortality by controlling for only the meteorological variables,[12,25] or controlling for the meteorological variables and simple D variables without considering the lockdown and the CM confounders.[26] Further, our work has in parallel sought to address issues

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

of non-linearity, collinearity, time-lag, and other possible inadequacies resulting from changes in case definition and testing capacity via two dummy variables (see Section 2). This is particularly critical for precision modelling when some statistical relationships that cover meteorology, such as AH, and R, have been shown to be non-linear (see Figure 1 (d)). By accounting for the non-linearity between some of our confounders and R, our model provides a more reliable and rigorous scientific prediction concerning how and when P will affect R across the 31 Chinese provincial capital cities in China, as well as in Wuhan, in contrast to prior air pollution-related COVID-19 infection/mortality models which have yet taken non-linearity or other statistical characteristics into account.[7,26]

Third, to the best of our understanding, this is the first international study that claimed a causal relationship between P and R across 31 Chinese provincial capital cities, via matching. Each high P exposure day was matched with a low P exposure day sharing similar background covariates such as meteorology and NM to estimate the causal effect. This causal relationship between immediate P exposure and R (i.e. a higher P can increase R, see Table 2), when combined with the recent reports that particulates less than 10 μm in size could facilitate the deposition of COVID-19 viral droplets and be suspended in the air,[13] further substantiates the recent observations regarding the risks of airborne infection.[14,16]

Finally, this is the first study that pursues the individual effects of P and AH on R, as well as the interaction effect of P and AH on R, covering (1) 31 provincial capital cities in China, and (2) data focussing on Wuhan only. Our study based on (1) ascertained that a 10 $\mu g/m^3$ increase in P was associated with a 0.9% increase in R. In addition, an inverted U-shape was identified between AH and R. A higher AH increased R when its value exceeded 8.6 g/m^3, while a higher AH decreased R when its value < 8.6 g/m^3. Further, when P interacted with AH, their interaction effect on R was significantly positive in (1) overall (Coefficient = $3.900 * 10^{-4}$, $p < 0.05$). When breaking down into two groups, if AH $\leq$ 8.6 g/m^3, higher P (Coefficient = 0.0008, $p < 0.05$) and AH (Coefficient = 0.0259, $p < 0.01$) led to a higher R (see Figures 2 (c) – (d)). However, if AH > 8.6 g/m^3, the interaction effect was statistically insignificant (see Figures 2 (e) – (f)). For (2), only P is significantly associated with R, and had a significantly positive and much stronger effect on R than in (1). However, unlike (1), AH was not significant ($p > 0.05$). Further, unlike (1) that a statistically significant interaction effect between P and AH was observed, the interaction

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

effect between the two variables was not statistically significant in (2) (Coefficient = -0.006, $p > 0.05$).

Further, for (1), NM was not a significant predictor of R. However, unlike the data for (2), T was significantly associated with a negative R in (1), suggesting that unobserved time-varying effects such as change in mobility within a city due to lockdown could reduce R. Further, the D variable age (0-14 years old), had a significant effect on R, implying that cities with a higher proportion of young people tended to have a lower R (see Table 2(a)).

A limitation of this study is that the statistical relationship between P and R in (1) (Coefficient = 0.0009, $p < 0.05$) was less strong as compared to (2) (Coefficient = 0.0261, $p < 0.01$). This might be due to the lack of datapoints at higher P and AH in (1) (see Figures 1 (c) – (d)), though the statistical relationship between P/AH and R was significant ($p < 0.05$, see Table 1(a)).

Our findings call for immediate medical and public health attention concerning the significance of P in exacerbating R. Controlling and reducing outdoor P, and reducing the possibility for outdoor P to be used as a carrier for COVID-19 viruses, have never been as urgent as they are now. Public health measures such as installing air purifiers, both indoors and outdoors, can help reduce P and alleviate the situation.[27] Alternatively, improving air ventilation, both indoors and outdoors, presents another possibility.[28] Despite previous claims that TEMP, UV, and WS will temper the COVID-19 infection,[29,30] our result showed no statistical significance for these factors. Nevertheless, our study supports recent findings on AH.[2] Given the result in China that AH < 8.6 g/m^3 exacerbated the rate of change, dehumidifiers may help to reduce infection rates. Moreover, the possibility for airborne infection of COVID-19 is too high a cost to be ignored and proper public health measures, such as requiring citizens to wear face masks, should be established to reduce the possibility of COVID-19 infection through air, especially for countries with high population densities and mobilities, and high particulate pollution. Given that the best fit model was obtained with a nine-day delay between infection and case reporting in China, this implies on average, it would take seven days for a COVID-19 patient to become symptomatic, assuming case reporting takes two days. This will have important public health implications on the quarantine period of those who are infected.

Contributor

JCL and VOL were responsible for conceptualization and initial framework development. YH, PG, QZ and AW collected the statistical data. SW conducted literature review. JCL, VOL and

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

YH developed the methodology. YH processed the data and conducted the statistical analysis. YH, JCL and VOL interpreted the results and wrote the full manuscript. JC, PG, JF, SW, IG and ZG provided valuable suggestions and comments on the data input and analysis, and the methodology. JCL and VOL applied for funding. YH, JCL and VOL contributed equally.

Declaration of Interest

We declare no competing interests.

Acknowledgement

This research is supported in part by the Theme-based Research Scheme of the Research Grants Council of Hong Kong, under Grant No. T41-709/17-N. We also acknowledge the help of J. Downey in editing this article. We also acknowledge the valuable comments from Joseph Hui and K.W. Wu on the non-linear relationship between the independent variables and the dependent variable.

References

1 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.

2 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.

3 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.

4 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.

5 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.

6 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.

7 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.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

9 Hoseinzadeh E, Taha P, Wei C, *et al.* The impact of air pollutants, UV exposure and geographic location on vitamin D deficiency. *Food and Chemical Toxicology* 2018; **113**: 241–54.

10 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.

11 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.

12 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.

13 Setti L, Passarini F, De Gennaro G, *et al.* SARS-Cov-2RNA found on particulate matter of Bergamo in Northern Italy: First evidence. *Environmental Research* 2020; **188**: 109754.

14 Liu Y, Ning Z, Chen Y, *et al.* Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature* 2020; **582**: 557–60.

15 Prather KA, Wang CC, Schooley RT. Reducing transmission of SARS-CoV-2. *Science* 2020; **368**: 1422–4.

16 Guo Z-D, Wang Z-Y, Zhang S-F, *et al.* Aerosol and Surface Distribution of Severe Acute Respiratory Syndrome Coronavirus 2 in Hospital Wards, Wuhan, China, 2020 - Volume 26, Number 7—July 2020 - Emerging Infectious Diseases journal - CDC. DOI:10.3201/eid2607.200885.

17 Kim Y-I, Kim S-G, Kim S-M, *et al.* Infection and Rapid Transmission of SARS-CoV-2 in Ferrets. *Cell Host & Microbe* 2020; **27**: 704-709.e2.

18 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.

19 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.

20 Shaman J, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. *PNAS* 2009; **106**: 3243.

21 Yang J, Zheng Y, Gou X, *et al.* Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases* 2020; **94**: 91–5.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.19.20106484; this version posted September 11, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

22 Morawska L, Milton DK. It is Time to Address Airborne Transmission of COVID-19. *Clin Infect Dis* DOI:10.1093/cid/ciaa939.

23 Sattar Naveed, McInnes Iain B., McMurray John J.V. Obesity Is a Risk Factor for Severe COVID-19 Infection. *Circulation* 2020; **142**: 4–6.

24 Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 2020; **395**: 689–97.

25 Ma Y, Zhao Y, Liu J, *et al.* Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of The Total Environment* 2020; **724**: 138226.

26 Bianconi V, Bronzo P, Banach M, Sahebkar A, Mannarino MR, Pirro M. Particulate matter pollution and the COVID-19 outbreak: results from Italian regions and provinces. *Archives of Medical Science* 2020; **16**.

27 Liang C-S, Duan F-K, He K-B, Ma Y-L. Review on recent progress in observations, source identifications and countermeasures of PM2.5. *Environment International* 2016; **86**: 150–70.

28 Martins NR, Carrilho da Graça G. Impact of PM2.5 in indoor urban environments: A review. *Sustainable Cities and Society* 2018; **42**: 259–75.

29 Gunthe SS, Swain B, Patra SS, Amte A. On the global trends and spread of the COVID-19 outbreak: preliminary assessment of the potential relation between location-specific temperature and UV index. *J Public Health* 2020; published online April 24. DOI:10.1007/s10389-020-01279-y.

30 Ahmadi M, Sharifi A, Dorosti S, Jafarzadeh Ghoushchi S, Ghanbari N. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Science of The Total Environment* 2020; **729**: 138705.