Journal of Environmental Science and Economics
https://doi.org/10.56556/jescae.v3i1.751 ISSN: 2832-6032

RESEARCH ARTICLE

Correlation or Causation: Unraveling the Relationship between PM2.5 Air Pollution and COVID-19 Spread Across the United States

Mohammad Maniat[1]*, Hossein Habibi[1], Elham Manshoorinia[1], Parisa Raufi[2], Payam Marous[1], Masoud Omraninaini[1]

[1]Department Civil Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran
[2]Department of Civil Engineering, Faculty of Technical and Engineering, Islamic Azad University, Tehran North Branch, Tehran, Iran

Corresponding Author: Mohammad Maniat: mohammad.maniat@srbiau.ac.ir
Received: 13 January, 2024, Accepted: 13 February,2024, Published: 17 February, 2024

Abstract
Numerous studies have examined the potential connection between air pollution, particularly PM2.5, and the incidence of COVID-19 cases during the pandemic. While several studies have demonstrated a strong correlation, caution is advised as correlation does not imply causation. To address this concern, our two-year observational study employs a comprehensive approach that utilizes a large sample size and draws on temporal and spatial data across the United States, surpassing the limitations of previous studies restricted to specific locations. Through rigorous correlation and regression analyses, we control for potential confounding factors. Air pollution data, a crucial component of our study, has been sourced from the United States Environmental Protection Agency (EPA). Additionally, COVID-19 case data is extracted from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, providing a robust and widely recognized dataset for our analyses. Notably, a significant spatial correlation exists between COVID-19 cases and population size ($r=0.98$, $p\text{-value} <0.01$), as confirmed by multivariate regression analysis, suggesting a confounding influence of population. It is crucial to emphasize that correlation does not automatically imply a direct cause-and-effect relationship. Moreover, to minimize the impact of population, we employ rates (COVID-19 cases/population of States), demonstrating that the rate of COVID-19 cases is independent of PM2.5 and population. Additionally, the rate of COVID-19 infection is not correlated with population density, implying the population's influence on infection is more likely due to probability rather than being a direct cause. In summary, while many studies report a correlation between air pollution and COVID-19 cases, the influence of confounding factors like population density necessitates further investigation to establish a definitive causal relationship. In conclusion, while many studies report a correlation between air pollution and COVID-19 cases, the influence of confounding factors like population density necessitates further investigation to establish a definitive causal relationship.

Keywords: COVID-19; population; air pollution; PM2.5; confounding

Introduction

A study conducted by Doremalen et al.(Van Doremalen et al., 2020)has demonstrated that SARS-CoV-2 can remain viable and infectious in aerosols for several hours and on certain surfaces. Building upon this research, the hypothesis arises for other researchers that COVID-19, caused by the coronavirus, might potentially interact with air pollution. Groulx et al(Groulx, Urch, Duchaine, Mubareka, & Scott, 2018)confirm that microbial agents of communicable diseases, such as viruses, have interactions with air pollution, affecting public health. A study conducted in Poland found a significant association between particulate matter and the number of new COVID-19 infections (Czwojdzińska, Terpińska, Kuźniarski, Płaczkowska, & Piwowar, 2021a). Similar studies across Europe suggest that short-term exposure to particulate matter (PM) is related to the spread of SARS-CoV-2, with PM levels in England and Italy specifically implicated(Renard et al., 2022; Zoran, Savastru, Savastru, & Tautan, 2020). In the Middle East, a study of Baghdad and Kuwait found that PM2.5 levels were positively related to deaths caused by COVID-19, with a decrease in particulate matter leading to a significant decrease in the death rate. In Kuwait, a 38.4% decrease in deaths was observed during the travel ban period, with an average decrease of 22.3% in PM2.5 levels. This study also found a positive relationship between air temperature and a negative relationship between humidity and the number of deaths (Halos, Al-Dousari, Anwer, & Anwer, 2021). Therefore, some studies have found a relationship between PM and COVID-19(Czwojdzińska, Terpińska, Kuźniarski, Płaczkowska, & Piwowar, 2021b; Renard, Surcin, Annesi-Maesano, & Poincelet, 2023b; Setti et al., 2020), while others have not found any significant association between the two (Bontempi, 2020). Some studies have merely identified a correlation between PM and the daily number of confirmed cases without providing a p-value(Zoran et al., 2020). In a study conducted in Delhi, researchers found that the number of COVID-19 cases exhibited a significant negative correlation with PM2.5 levels (correlation=−0.63, p-value<0.01) during the pre-lockdown phase. However, the number of COVID-19 cases during the lockdown phase also showed a positive correlation with PM2.5, with a correlation value of 0.56. Despite these contrasting correlations, the researchers concluded that there is a dependence of COVID-19 transmission on the concentration of PM2.5 in Delhi's environment(Chaudhary et al., 2022).

The study aims to investigate the reason behind the varied correlations in existing research, exploring the potential role of confounding factors, notably population, in influencing whether some studies observe a positive correlation while others find a negative association. The study unfolds systematically, commencing with a thorough introduction to the global impact of COVID-19 and its potential connection to air pollution. A comprehensive literature review examines existing research, paving the way for a detailed methodology encompassing study design, data sources, and statistical analyses. The data sources section clarifies the origins and reliability of air pollution and COVID-19 data. The ensuing analysis meticulously presents statistical findings while addressing potential biases. A nuanced discussion interprets results, exploring implications and limitations, and the conclusion succinctly summarizes key findings while proposing avenues for future research.

Literature review

The COVID-19 pandemic will have long-term effects on the worldwide economy(Al-kasasbeh, 2022). Meanwhile, Various studies have explored factors influencing COVID-19 transmission, including air pollution (Maniat et al., 2023)and preventative measures like handwashing (Otto, Opatoki, & Luyi, 2022). An observational study in USA California, using data from the Environmental Pollution Agency (EPA), reported negative correlations between PM2.5 levels and both COVID-19 cases (-0.45) and mortality (-0.42) (Bashir, Jiang, et al., 2020). Researchers Adhikari and Yin studied air pollution in Queens, New York, comparing levels of PM2.5 with

COVID-19 infection and mortality rates. While they found no significant relationship between daily PM2.5 and either COVID-19 infection or mortality, they did uncover a significant positive association with new confirmed cases(Adhikari & Yin, 2020). A study of 14,783 COVID-19 patients found long-term exposure to fine particulate matter (PM2.5) is associated with increased hospitalization risk. Among the participants, 13.6% were hospitalized. Researchers analyzed both average PM2.5 exposure over the past 10 years and estimated exposure for the year 2018. The study found that for every 1 $\mu g/m^3$ increase in PM2.5, the odds of hospitalization rose by 18% (10-year average) and 14% (2018 estimate). While this suggests a link, further research is needed to confirm causation and explore the underlying mechanisms(Mendy et al., 2021). A study has revealed a potential link between increased air pollution and higher COVID-19 death rates. Researchers found that every 1 microgram per cubic meter ($\mu g/m^3$) increase in fine particulate matter (PM2.5) was associated with an 8% rise in COVID-19 deaths. This association was statistically significant and remained consistent even after accounting for other potential influencing factors.(Wu, Nethery, Sabath, Braun, & Dominici, 2020). While some studies indicate a link between air pollution and COVID-19 severity, findings remain mixed. One study found no significant association between long-term exposure to PM2.5 or ozone (O3) and COVID-19 case-fatality rate. However, they did observe a weak but potentially important connection between higher PM2.5 levels (an increase of 2.6 micrograms per cubic meter) and a 14.9% increase in COVID-19 mortality rate, even after adjusting for other air pollutants. This suggests further investigation is needed to clarify the complex relationship between air pollution and COVID-19 outcomes(Liang et al., 2020). A study found a 10.5% ± 2.5% increase in mortality per 1 $\mu g/m^3$ increase in air pollution. However, this impact lessened over time, suggesting potential factors like improved pandemic management and broader vaccination after mid-2021. Interestingly, despite potential differences in initial conditions, the relative trend of mortality increase with higher air pollution was consistent across the studied countries(Renard et al., 2022). A review paper by Arun Srivastava explores the relationship between various pollution parameters and the number of COVID-19 cases. The findings reveal diverse correlations, including some with no correlation, others exhibiting a negative relationship, and some indicating a positive association(Srivastava, 2021). The reason why some studies find a positive relationship between PM and COVID-19 cases, while others do not, can be attributed to the fact that correlation does not imply causation. To establish causation, researchers need to conduct carefully designed studies, such as randomized controlled trials or longitudinal studies, to demonstrate a direct cause-and-effect relationship between PM levels and COVID-19 outcomes.

Indeed, emissions from the combustion of diesel fuel in cars and other vehicles are recognized as a significant source of particulate matter (PM) in urban areas (McDuffie et al., 2021; Nava et al., 2020). As a result, regions with higher population density tend to have more transportation activities, contributing to increased levels of PM(Aljoufie, Zuidgeest, Brussel, & Van Maarseveen, 2011; Maniat, Abdoli, Raufi, & Marous) During the COVID-19 lockdowns implemented in response to the pandemic, there were significant reductions in urban activity, including a decrease in transportation and industrial activities. As a result, there was a noticeable reduction in emissions, including those of particulate matter. This reduction in human activity led to improvements in air quality in many urban areas during the lockdown periods(Manjeet, Airon, Kumar, & Saifi, 2022). Population is a crucial factor in urban areas, as it reflects the concentration of individuals in a given space. Areas with higher populations are more likely to experience quick spreading of infectious diseases, including COVID-19(Ahmed, Jaman, Saha, & Ghosh, 2021). While areas with larger populations tend to have more reported COVID-19 cases (correlation), it does not necessarily mean that, the population itself directly causes the spread of the virus (causation). Just like flipping a coin multiple times increases the likelihood of observing both heads and tails, having a larger population in an area might lead to more reported COVID-19 cases due to an increased chance of encountering infected individuals. However, this correlation does not imply that population size directly causes the occurrence of COVID-19 cases. Two studies Malaysia found a strong positive and statistically significant correlation between the total population and COVID-19 cases, indicating that larger populations were associated with higher case numbers. However, the relationship between population density and the spread of COVID-19 was weaker(Aw et al., 2021; H. S. Wong, Hasan, Sharif, & Rahman, 2023). Using cumulative frequency reports of COVID-19 cases or deaths in research studies can lead to several common mistakes and misinterpretations. Cumulative data grows over time, and using it directly in analysis may introduce a time-dependent bias. Cumulative data may not adequately control for confounding factors such as public health interventions, population mobility, healthcare capacity, and socioeconomic variables. Failing to account for these factors can lead to spurious correlations. For instance, two studies found a correlation between population density and COVID-19 in the USA (Sy, White, & Nichols, 2021; D. W. Wong & Li, 2020), using cumulative frequency reports of COVID-19. Of course, the population density in specific places, such as hospitals, public transportation, and cruise ships(Rocklöv & Sjödin, 2020), can significantly contribute to the transmission of COVID-19 in localized settings, it is crucial to clarify that our study's primary objective is to investigate this phenomenon on a broader macro scale, covering provinces, cities, and countries. We seek to discern the distinction between physical distancing and population density. It is imperative to recognize that while the density, calculated as city population divided by area, may be high in a city, it does not necessarily correlate with low levels of physical distancing. In the study (D. W. Wong & Li, 2020)there is an assumption that the level of physical distancing is contingent on population density, implying that areas with higher population density experience a greater incidence of the coronavirus. Consequently, the study concludes that population density is a significant variable influencing COVID-19 cases. However, it's essential to approach this assumption with a nuanced perspective. While there may be a correlation between population density and COVID-19 cases, establishing a direct causation is complex. The relationship is influenced by various factors, including local public health interventions, cultural practices, healthcare infrastructure, and individual behaviors. Our research seeks to explore this intricate relationship on a broader macro scale, encompassing provinces, cities, and countries. By considering multiple variables and potential confounders, we aim to contribute to a more comprehensive understanding of the factors influencing COVID-19 transmission dynamics. In another study conducted in America, focusing on 913 counties, they found that metropolitan population density played a significant role as a predictor of infection rates. However, they observed that county density, by itself, was not significantly related to the infection rate. Instead, the study highlighted that connectivity, which involves factors beyond just density, appears to have a more significant impact on infection rates(Hamidi, Sabouri, & Ewing, 2020).

Considering the complexities of the association between air pollution and the spread of COVID-19, it would be reasonable to expect that regions with higher wind speeds, resulting in lower pollution levels, would also have fewer COVID-19 cases if all other factors were equal. However, despite this logical expectation, studies have not consistently shown a correlation between wind speed, pollution, and COVID-19 cases. The Gaussian air pollutant dispersion equation is indeed one of the earliest and simplest forms of pollutant dispersion modeling. It describes how air pollutants disperse and spread in the atmosphere under the influence of wind and other meteorological factors(Abdel-Rahman, 2008) Higher wind speeds can enhance the dispersion of air pollutants, leading to lower local pollution levels in densely populated cities. In areas with high wind speeds, it is expected that air pollutants would disperse more effectively, potentially reducing the concentration of pollutants in the air. In the study conducted in New York, the Spearman Correlation Coefficient of +0.172 suggests a positive correlation between wind speed and COVID-19 cases. This means that higher wind speeds were associated with higher COVID-19 case counts in that particular area(Bashir, Ma, et al., 2020). On the other hand, the study in Jakarta, Indonesia, revealed a significantly negative correlation ($r = -0.314$; $p < 0.05$) between low wind speed and higher COVID-19 cases(Rendana, 2020). Moreover, the study by Shao et al. found a positive and negative correlation between wind and the number of infected, indicating a connection between pollution and COVID-19 (Shao et al., 2022). The limitations observed in existing investigations stem from the fact that both air pollution and COVID-19 infections are correlated both spatially and temporally. Both spatial and temporal correlations between air pollution and infections can introduce biases in the estimation of results. Typically, researchers choose to consider either spatial or temporal correlations, depending on the research question and the nature of the data being analyzed. Our study possesses several advantages. Firstly, it benefits from a large number of statistical samples, which enhances the robustness and reliability of the findings. Additionally, the research employs two different types of correlations, namely spatial and temporal to thoroughly investigate the relationship between air pollution and COVID-19. This comprehensive approach allows for a more comprehensive understanding of the potential link between air pollution and the incidence of the disease. By utilizing various correlation methods and a substantial dataset, this study aims to provide valuable insights into the impact of air pollution on COVID-19.

Methodology

Sample

This study centers on fifty-one (N= 51) states in the USA, one of the countries significantly impacted by the COVID-19 pandemic, with over 54 million cases reported over the course of two years (2020 and 2021). Due to the larger dataset of people infected with COVID-19 compared to the number of deaths, this study utilized data on the number of infected individuals for analysis.

Sources

Wind speed and air pollution data were obtained from the United States Environmental Protection Agency (EPA) website (Agency, 2020,2021) the study also obtained temperature data in Fahrenheit from the National Centers for Environmental Information(Information). COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University(jhu, 2022).

Measurements

The data of PM2.5 is often reported using the Air Quality Index (AQI), which provides an overall measure of air quality based on various pollutants, including PM2.5. However, the AQI is a dimensionless index and not directly usable for quantitative analyses due to its scale and unitless nature. To facilitate statistical analysis and comparisons, researchers often convert AQI values to a more quantitative and usable unit such as micrograms per cubic meter (µg/m³) using appropriate conversion equations. This conversion allows for the data to be expressed in a standard unit that can be utilized in statistical models and helps to establish a more meaningful relationship between PM2.5 concentrations and other variables. While the correlation between AQI and µg/m³ values not be 1, converting AQI to µg/m³ provides a more accurate representation of PM2.5 concentrations, enabling researchers to better understand its relationship with other variables in quantitative analyses. The AQI is given by Equation (1)(Kanchan, Gorai, & Goyal, 2015).

$$AQI = \frac{AQI_{Hi} - AQI_{Lo}}{conc_{Hi} - conc_{Lo}} \times (conc_i - conc_{Lo}) + AQI_{Lo} \quad (1)$$

Where;
$Conc_i$(PM2.5)= input concentration for a given pollutant(pm2.5)
$Conc_{Lo}$= the concentration breakpoint that is less than or equal to $Conc_i$

$Conc_{Hi}$(PM2.5)= the concentration breakpoint that is greater than or equal to $Conc_i$
AQI_{Lo}= the AQI breakpoint corresponding to $Conc_{Lo}$
AQI_{Hi}= the AQI breakpoint corresponding to $Conc_{Hi}$
The average wind speed is measured in meters per second (m/s) using the Instrumental - RM Young Model 05103, which is designed to measure wind speed at low altitudes. It is important to note that wind speed can vary with height, and therefore, different devices and methods may yield different results due to the variations in wind patterns at different altitudes.
Time series data for COVID-19 confirmed cases in the United States for the years 2020 and 2021 can be obtained from the CSSE (Center for Systems Science and Engineering) at Johns Hopkins University public archive data (University). In the archive, the data is initially provided as cumulative frequency, which represents the total number of COVID-19 cases up to a specific date. To use this data for analysis, it needs to be transformed into daily frequency by taking the difference between consecutive data points. To clarify, for each day, the number of new COVID-19 cases (frequency) can be calculated by subtracting the cumulative count on the previous day (t_0) from the cumulative count on the current day (t_1), denoted as $x(t_1) - x(t_0)$. In addition, the ratio of the number of cases to the total time the population is at risk of disease can also be calculated. This ratio provides insights into the incidence rate of COVID-19 cases per unit of time for each state. Furthermore, to determine population density, one can obtain the population of each state and divide it by the area of each state. In the majority of studies, researchers commonly employ Pearson correlation for assessing the relationship between variables. While some studies use Kendall and Spearman correlation, the differences in results are not significant. To facilitate comparison with other research, we also utilize Pearson correlation. Pearson's correlation coefficient (r) is a widely used measure that evaluates the strength, type, and direction of the relationship between two variables. The Pearson correlation (r) is defined as shown in Equation (2)(Akoglu, 2018).

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2(y_i-\bar{y})^2}} \tag{2}$$

where:
r=correlation coefficient,
x_i , y_i are the values of the variable in a sample i,
$\bar{x}$, $\bar{y}$= mean of the values of the y-variable.
In research that investigates a potential cause-and-effect relationship, a confounding variable is an unmeasured third variable that influences both the supposed cause and the supposed effect. Confounding is one of three types of bias that can distort the results of epidemiologic studies and potentially lead to erroneous conclusions(Howards, 2018)
It’s important to consider potential confounding variables and account for them in your research design to ensure your results are valid. Left unchecked, confounding variables can introduce many research biases to your work, causing you to misinterpret your results. Confounding variables (a.k.a. confounders or confounding factors) are a type of extraneous variable that are related to a study’s independent and dependent variables. A variable must meet two conditions to be a confounder(McNamee, 2003):
It must be correlated with the independent variable. This may be a causal relationship, but it does not have to be.
It must be causally related to the dependent variable.
The conceptual model incorporates the idea of these two conditions, with the confounding variable being present in Figure 1.

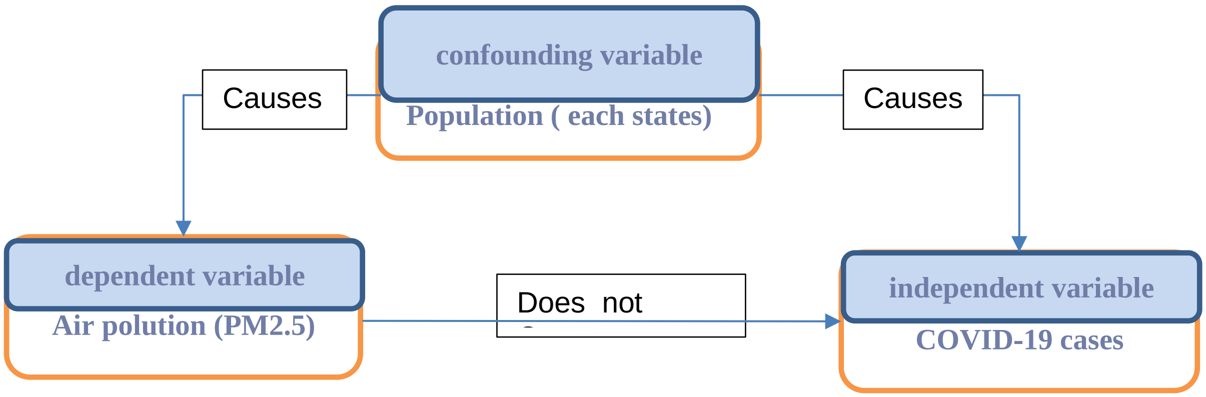

Figure 1. Conceptual model confounding variable

The technique of multivariable regression analysis has been extensively employed to manage confounding factors, and its utilization saw significant augmentation, especially when modeling tools became easily accessible(Kahlert, Gribsholt, Gammelager, Dekkers, & Luta, 2017). Multiple regression analysis serves the purpose of evaluating the presence of confounding. Through multiple linear regression analysis, we can estimate the relationship between a specific independent variable and the outcome while keeping all other variables constant. This approach allows for the adjustment or accounting of potential confounding variables incorporated into the model. Consider a scenario with a risk factor or exposure variable denoted as X1 (e.g., X_1=Air pollution or X_1=Temperature) and an outcome or dependent variable denoted as Y. The estimation of a simple linear regression equation relating the risk factor to the dependent variable is expressed as follows in equation (3).

$$Y=b_0+b_1X \tag{3}$$

Suppose the aim is to assess whether a third variable (e.g., population) acts as a confounder. This potential confounder is denoted as X_2, and the estimation involves a multiple linear regression (4).

$$Y=b_0+b_1X+b_2X_2 \tag{4}$$

Some researchers evaluate confounding by examining the extent of change in the regression coefficient associated with the risk factor after adjusting for the potential confounder. In this context, a comparison is made between b1 from the simple linear regression model and b1 from the multiple linear regression model. As a general guideline, when there is a shift of more than 10% in the regression coefficient derived from the simple linear regression model, it is commonly considered that X_2 functions as a confounding variable(Harrell Jr, Lee, & Mark, 1996; Sudin, Aziz, Saad, Khalid, & Ibrahim, 2021; Vittinghoff, Shiboski, Glidden, & McCulloch, 2005).

Results

Figure 2 depicts the number of confirmed COVID-19 cases in the United States throughout the years 2020 and 2021. The data shows that the peak of COVID-19 infections in 2020 occurred in December, while in 2021, the highest number of cases was reported in January. Over the entire year of 2020, a total of 20,126,950 confirmed COVID-19 cases were recorded in the United States, and this number surged to 34,505,103 in 2021. The Fig2 effectively presents the overall trend of COVID-19 cases over the two-year period, highlighting fluctuations and changes in infection rates across different months in both years.

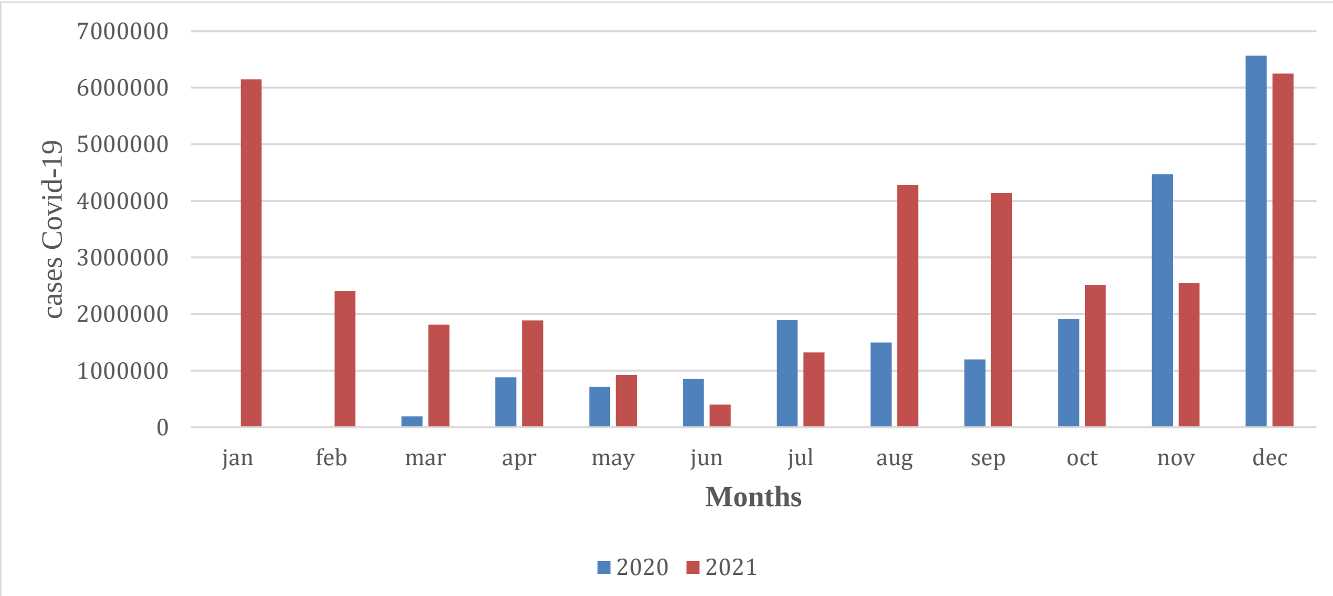

Figure 2. The number of confirmed COVID-19 cases in the years 2020 and 2021 Source authors `s analysis

The data analysis presented in Figure 3 consistently demonstrates a high prevalence of COVID-19 cases in California, Florida, New York, and Texas throughout the two-year period. The three graphs indicate that the pattern of COVID-19 cases in these states closely correlates with their respective population sizes. States with larger populations tend to have a higher number of COVID-19 cases.

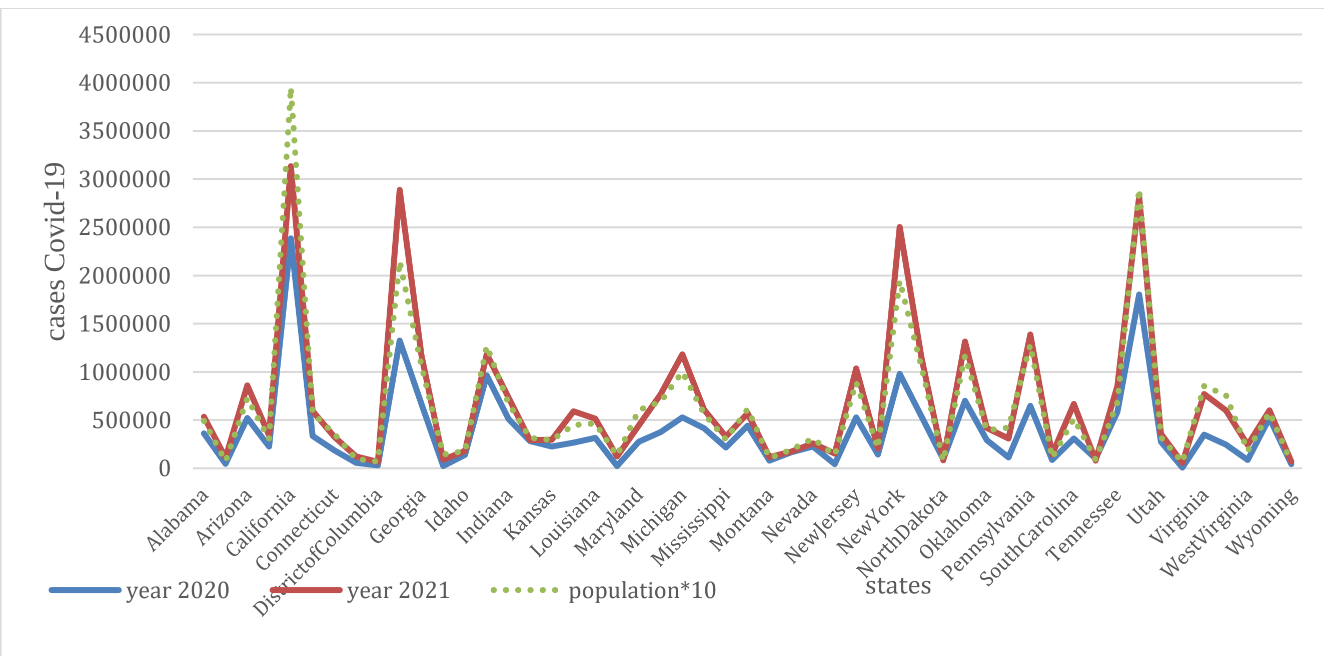

Figure 3. Number of confirmed COVID-19 in the years 2020 and 2021 Source authors `s analysis

The strong spatial correlation between COVID-19 cases in 2020 and 2021 suggests that the pattern of infections for each state repeated in the following year (Table1). There is a significant positive correlation between the population and COVID-19 cases(r=0.98), supporting the idea discussed in the introduction that population size can influence the likelihood of infection. The weak correlations, close to zero, between the rate of COVID-19 cases and population, as well as population density and COVID-19 cases. Wind speed shows no correlation with COVID-19 cases, indicating it has little impact on transmission dynamics. Temperature, on the other hand, exhibits a positive correlation with COVID-19 cases. Regarding PM2.5, COVID-19 cases in 2020 show a significant positive correlation (r=0.468) with PM2.5, while in 2021, the correlation remains positive (r=0.168) but not significant. Additionally, the correlation between the rate of COVID-19 cases and PM2.5 is close to zero, suggesting their independence.

Table 1. Spatial correlation and COVID-19 cases in different states Source authors `s analysis

	COVID-20	COVID-21	r2020	r2021	pop	density	pm2020	pm2021	temp2020	temp2021	wind2020	wind2021
COVID-20	1	.948**	0.045	-0.023	.982**	-0.095	.468**	.289*	.338*	.333*	-0.012	-0.011
COVID-21	.948**	1	-0.071	0.1	.967**	-0.084	.340*	0.168	.349*	.338*	-0.092	-0.101
rate2020	0.045	-0.071	1	0.253	-0.083	-0.169	0.09	.368**	-0.13	-0.113	.374**	.392**
rate2021	-0.023	0.1	0.253	1	-0.054	-0.082	-0.139	-0.067	-0.105	-0.136	-.286*	-.304*
population	.982**	.967**	-0.083	-0.054	1	-0.082	.450**	0.244	.324*	.316*	-0.065	-0.065
density	-0.095	-0.084	-0.169	-0.082	-0.082	1	0.092	0.09	0.12	0.104	-0.1	-0.104
pm2020	.468**	.340*	0.09	-0.139	.450**	0.092	1	.803**	0.076	0.071	0.032	0.057
pm2021	.289*	0.168	.368**	-0.067	0.244	0.09	.803**	1	-0.052	-0.046	0.17	0.188
temp2020	.338*	.349*	-0.13	-0.105	.324*	0.12	0.076	-0.052	1	.998**	-0.115	-0.083
temp2021	.333*	.338*	-0.113	-0.136	.316*	0.104	0.071	-0.046	.998**	1	-0.078	-0.046
wind2020	-0.012	-0.092	.374**	-.286*	-0.065	-0.1	0.032	0.17	-0.115	-0.078	1	.973**
wind2021	-0.011	-0.101	.392**	-.304*	-0.065	-0.104	0.057	0.188	-0.083	-0.046	.973**	1

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

Table 2 displays the Temporal correlation between different variables. The correlation between COVID-19 cases in 2020 and 2021 is found to be r=0.384, which is much weaker than the spatial correlation observed earlier. This suggests that the relationship between COVID-19 cases is dependent on spatial variables, not temporal variables. . The correlation between temperatures in 2020 and 2021 is high, indicating that the temperature pattern remains consistent in most states of America and is repeated year after year. The 7th and 8th months of the year are typically the hottest months. Additionally, there is a high and significant correlation between wind speed in 2020 and 2021 (r=0.899). Wind speed and temperature tend to have an inverse relationship, where higher wind speeds are associated with cooler temperatures. Furthermore, the correlation between wind speed and PM2.5 is -0.685 and -0.613 (p-value <0.01) for the years 2020 and 2021, respectively. This indicates that when wind speed is higher, PM2.5 levels tend to be lower. Regarding COVID-19 cases, there is a positive correlation with PM2.5 in both 2020 (r=0.111) and 2021 (r=0.235).

Table 2. Temporal correlation between pollution and COVID-19 cases Source authors `s analysis

	COVID-2020	COVID-2021	temp2020	temp2021	wind2020	wind2021	pm2020	pm2021
COVID-20	1	0.384	-0.175	-0.104	-0.273	-0.182	0.111	-0.005
COVID-21	0.384	1	-0.455	-0.398	-0.375	-0.176	0.355	0.235
temp2020	-0.175	-0.455	1	.986**	-0.529	-.620*	0.295	0.528
temp2021	-0.104	-0.398	.986**	1	-0.551	-.603*	0.327	0.477
wind2020	-0.273	-0.375	-0.529	-0.551	1	.899**	-.685*	-.689*
wind2021	-0.182	-0.176	-.620*	-.603*	.899**	1	-0.556	-.613*
pm2020	0.111	0.355	0.295	0.327	-.685*	-0.556	1	0.331
pm2021	-0.005	0.235	0.528	0.477	-.689*	-.613*	0.331	1

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

In the context of multiple regression, the Table 3 provides an overview of the R-Square, Std. Error of the Estimate, R-Square Change, F Change, and Significance of F Change for each model, incorporating various sets of predictors such as the constant, pm, temperature, wind, and population. Model 4 is the best model as it has the highest R-squared value of 0.847. the R-squared change value of 0.656 suggests that population explains 65.6% of the remaining variance in the dependent variable after accounting for the other independent variables in the model. This is a significant increase, and it suggests that population is indeed a confounding variable.

Table 3. Model Summary Source authors `s analysis

Model		R Square	Std. Error of the Estimate	R Square Change	F Change	Sig. F Change
1	a. Predictors: (Constant), pm	0.087	596876.16057	0.087	9.517	0.003
2	b. Predictors: (Constant), pm, temperature	0.185	566575.96405	0.099	11.982	0.001
3	c. Predictors: (Constant), pm, temperature, wind	0.190	567708.50660	0.005	0.605	0.438
4	d. Predictors: (Constant), pm, temperature, wind, population	0.847	248202.09317	0.656	415.703	0.000

Table 4. regression results Coefficients Source authors `s analysis

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	-79946.814	207796.301		-0.385	0.701	-492208.757	332315.130
	pm	59529.328	19297.089	0.295	3.085	0.003	21244.453	97814.202
2	(Constant)	-1255637.504	392769.598		-3.197	0.002	-2034977.599	-476297.410
	pm	58455.988	18320.104	0.289	3.191	0.002	22104.928	94807.048
	temperature	21971.717	6347.476	0.314	3.461	0.001	9376.948	34566.487
3	(Constant)	-1139117.126	421084.237		-2.705	0.008	-1974745.090	-303489.162
	pm	60440.713	18533.106	0.299	3.261	0.002	23662.368	97219.059
	temperature	21588.773	6379.179	0.309	3.384	0.001	8929.500	34248.045
	wind	-27986.691	35969.267	-0.072	-0.778	0.438	-99366.529	43393.148
4	(Constant)	-16004.504	192162.430		-0.083	0.934	-397393.734	365384.725
	pm	-4459.725	8705.497	-0.022	-0.512	0.610	-21737.727	12818.277
	temperature	1756.935	2953.726	0.025	0.595	0.553	-4105.394	7619.265
	wind	-65.921	15785.272	0.000	-0.004	0.997	-31395.316	31263.473
	population	0.078	0.004	0.919	20.389	0.000	0.070	0.086

The standardized coefficient for population in model 4 is 0.919, which is very significant. Indeed, based on Table 4, it is evident that the population (variable) exhibits a significant influence on the dependent variable. This indicates that population is a confounding variable, meaning that it is an extraneous factor that is correlated with both the independent variable (PM) and the dependent variable (COVID-19 cases). This can make it difficult to isolate the true relationship between PM and COVID-19 cases. The fact that the coefficient for PM decreases by more than 10% after controlling for population suggests that population is indeed a confounding variable. This means that PM is not the sole cause of COVID-19 cases, and that population must also be considered a factor.

Discussion

Our study employed spatial and temporal correlation analyses to explore the relationships between wind, temperature, pollution, population density, and COVID-19 cases. The findings suggest correlations between pollution and COVID-19 cases but caution against making direct causative conclusions. While many studies have shown a correlation between air pollution and the number of COVID-19 infections, it does not imply causality. During lockdown periods, we observed a decrease in pollution, and studies have shown that the disease itself caused a decrease in air pollution(Su et al., 2023). However, this correlation does not indicate causation but rather reflects the simultaneous occurrence of two phenomena. Observing similar patterns between the graphs of mortality and infection rates in Europe(Renard, Surcin, Annesi-Maesano, & Poincelet, 2023a), researchers may be inclined to automatically assume that pollution has a strong effect on COVID-19. There are several reasons why caution is necessary in making such conclusions:

1-Correlation does not imply causation: Just because two variables (in this case, air pollution and COVID-19 outcomes) show similar patterns does not necessarily mean that one directly causes the other. There could be other factors at play that are responsible for the observed associations. To demonstrate the potential for such errors, you used the rate of infected people (the number of infected individuals divided by the population of the state) and found that its correlation with air pollution was close to zero. This finding suggests that there is no strong linear relationship between air pollution and the rate of COVID-19 infections.

2-Confounding factors: The observed patterns in COVID-19 cases could be influenced by numerous confounding factors, such as population. These factors may influence both air pollution levels and the spread of COVID-19 independently(Kelly et al., 2023). Although the spatial correlation in Table 1 shows the effect of population on COVID-19 and pollution at a significant level (p-value<0.01). Population is one such confounding factor that can impact both air pollution levels and the spread of COVID-19 independently. A larger population in an area may lead to more reported COVID-19 cases due to the increased likelihood of encountering infected individuals. However, this correlation does not imply that population size directly causes the occurrence of COVID-19 cases. If population size were the primary determinant of COVID-19 cases, then population density would also have a similar effect on both COVID-19 cases and air pollution (But the correlation is close to zero).

3-Regional variations: Consistent with previous research(Coşkun, Yıldırım, & Gündüz, 2021; Rendana, 2020) areas experiencing higher wind speeds tend to have lower levels of PM2.5 pollution. Interestingly, we also observed a temporal correlation between lower wind speeds and increased COVID-19 cases. This temporal correlation suggests that reduced wind speeds might contribute to higher COVID-19 case numbers. However, when examining the spatial correlation, we found a positive association. This suggests that factors beyond just wind speed and pollution may influence the spatial distribution of COVID-19 cases.

Conclusions

The global impact of the COVID-19 pandemic, stemming from a highly contagious virus within the SARS family, has been widespread, affecting over 200 countries and leading to more than 6.9 million deaths as of the current date (Rahimi, Chen, & Gandomi, 2023).The study identifies a correlation between air pollution and COVID-19 cases, emphasizing the need for cautious interpretation. Although a correlation exists, it does not necessarily imply a causal relationship, prompting consideration of other variables such as population and wind speed. The intricate relationship among air pollution, COVID-19, and various factors requires further research. It is stressed that the correlation between two variables does not automatically suggest a direct cause-and-effect connection; additional factors may account for the observed correlation. The study recognizes confounding factors, with population identified as one such factor, correlated with both air pollution and COVID-19 cases, while wind speed shows the correlation solely with air pollution. While exposure to air pollution is linked to heightened vulnerability in COVID-19 patients, it cannot be definitively stated that pollution directly causes exacerbation of COVID-19. Various contributors, including temperature, lifestyle, population density, and nutrition, play roles in the incidence of COVID-19. Notably, the rate of COVID-19 infection is not correlated with population and population density, categorizing the impact of population on infection as a probability effect rather than an effective and causal variable. To achieve a more comprehensive understanding of the intricate interactions between air pollution and COVID-19, it is essential to collect data from different states or cities. Establishing a robust causal relationship demands rigorous scientific investigations, including longitudinal studies with meticulous control of confounding factors, as well as experimental studies and causal modeling. While mounting evidence suggests that air pollution may exacerbate respiratory conditions and increase vulnerability to infections, including COVID-19, it is crucial to refrain from drawing definitive conclusions solely based on visual observations of graphs. A careful and nuanced approach is essential in unraveling the complexities of the relationship between air pollution and COVID-19 outcomes.

Declaration: We (all authors) declare that the paper is our original work and is not published anywhere.

Acknowledgment: None

Funding: There is no funding for this study

Conflict of Interest: The authors declare that they have no conflict of interest.

Authors contribution: Conceptualization: Mohammad Maniat, Methodology: Hosein Habibi, Mohammad Maniat; Software: Payam Marous, Validation: Elham Manshoorinia, Resources; Data Curation: Masoud Omrani; Writing—Original Draft Preparation: Parisa Raufi, Writing—Review

Data availability: All the code files necessary to reproduce the results of this study are available at https://doi.org/10.5281/zenodo.8197105

References

Abdel-Rahman, A. A. (2008). *On the atmospheric dispersion and Gaussian plume model.* Paper presented at the Proceedings of the 2nd International Conference on Waste Management, Water Pollution, Air Pollution, Indoor Climate, Corfu, Greece.

Adhikari, A., & Yin, J. (2020). Short-term effects of ambient ozone, PM2. 5, and meteorological factors on COVID-19 confirmed cases and deaths in Queens, New York. *International journal of environmental research and public health, 17*(11), 4047.

Agency, U. S. E. P. (2020,2021). Pre-Generated Data Files. https://aqs.epa.gov/aqsweb/airdata/download_files.html#Annual.

Ahmed, J., Jaman, M. H., Saha, G., & Ghosh, P. (2021). Effect of environmental and socio-economic factors on the spreading of COVID-19 at 70 cities/provinces. *Heliyon, 7*(5).

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(13): p. 91-93.

Al-kasasbeh, O. (2022). COVID-19 pandemic: macroeconomic impacts and understanding its implications for Jordan. *Journal of Environmental Science and Economics, 1*(2), 51-57.

Aljoufie, M., Zuidgeest, M., Brussel, M., & Van Maarseveen, M. (2011). Urban growth and transport: understanding the spatial temporal relationship. *Urban transport XVII: urban transport and the environment in the 21st Century. WIT press, Southampton*, 315-328.

Aw, S. B., Teh, B. T., Ling, G. H. T., Leng, P. C., Chan, W. H., & Ahmad, M. H. (2021). The covid-19 pandemic situation in malaysia: Lessons learned from the perspective of population density. *International journal of environmental research and public health, 18*(12), 6566.

Bashir, M. F., Jiang, B., Komal, B., Bashir, M. A., Farooq, T. H., Iqbal, N., & Bashir, M. (2020). Correlation between environmental pollution indicators and COVID-19 pandemic: a brief study in Californian context. *Environmental Research, 187*, 109652.

Bashir, M. F., Ma, B., Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of the total environment, 728*, 138835.

Bontempi, E. (2020). First data analysis about possible COVID-19 virus airborne diffusion due to air particulate matter (PM): the case of Lombardy (Italy). *Environmental Research, 186*, 109639.

Chaudhary, V., Bhadola, P., Kaushik, A., Khalid, M., Furukawa, H., & Khosla, A. (2022). Assessing temporal correlation in environmental risk factors to design efficient area-specific COVID-19 regulations: Delhi based case study. *Scientific Reports, 12*(1), 12949. doi:10.1038/s41598-022-16781-4

Coşkun, H., Yıldırım, N., & Gündüz, S. (2021). The spread of COVID-19 virus through population density and wind in Turkey cities. *Science of the total environment, 751*, 141663.

Czwojdzińska, M., Terpińska, M., Kuźniarski, A., Płaczkowska, S., & Piwowar, A. (2021a). Exposure to PM2. 5 and PM10 and COVID-19 infection rates and mortality: a one-year observational study in Poland. *Biomedical journal*.

Czwojdzińska, M., Terpińska, M., Kuźniarski, A., Płaczkowska, S., & Piwowar, A. (2021b). Exposure to PM2. 5 and PM10 and COVID-19 infection rates and mortality: A one-year observational study in Poland. *biomedical journal, 44*(6), S25-S36.

Groulx, N., Urch, B., Duchaine, C., Mubareka, S., & Scott, J. A. (2018). The Pollution Particulate Concentrator (PoPCon): A platform to investigate the effects of particulate air pollutants on viral infectivity. *Science of the total environment, 628*, 1101-1107.

Halos, S. H., Al-Dousari, A., Anwer, G. R., & Anwer, A. R. (2021). Impact of PM2. 5 concentration, weather and population on COVID-19 morbidity and mortality in Baghdad and Kuwait cities. *Modeling Earth Systems and Environment*, 1-10.

Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *Journal of the American Planning Association, 86*(4), 495-509.

Harrell Jr, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine, 15*(4), 361-387.

Howards, P. P. (2018). An overview of confounding. Part 2: how to identify it and special situations. *Acta obstetricia et gynecologica Scandinavica, 97*(4), 400-406.

Information, N. C. f. E. https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series.

jhu. (2022). Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. *https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_US.csv*.

Kahlert, J., Gribsholt, S. B., Gammelager, H., Dekkers, O. M., & Luta, G. (2017). Control of confounding in the analysis phase–an overview for clinicians. *Clinical epidemiology*, 195-204.

Kanchan, K., Gorai, A. K., & Goyal, P. (2015). A review on air quality indexing system. *Asian Journal of Atmospheric Environment, 9*(2), 101-113.

Kelly, S. L., Shattock, A. J., Ragettli, M. S., Vienneau, D., Vicedo-Cabrera, A. M., & de Hoogh, K. (2023). The Air and Viruses We Breathe: Assessing the Effect the PM2.5 Air Pollutant Has on the Burden of COVID-19. *Atmosphere, 14*(5), 887.

Liang, D., Shi, L., Zhao, J., Liu, P., Sarnat, J. A., Gao, S., . . . Scovronick, N. (2020). Urban air pollution may enhance COVID-19 case-fatality and mortality rates in the United States. *The Innovation, 1*(3).

Maniat, M., Abdoli, R., Raufi, P., & Marous, P. Trip Distribution Modeling Using Neural Network and Direct Demand Model.

Maniat, M., Habibi, H., Manshoorinia, E., Marous, P., Pirayvatlou, P. S., & Majidi, A. (2023). Temporal and Spatial Correlation of Air Pollution with COVID-19 in the USA: Challenges and Implications.

Manjeet, Airon, A., Kumar, R., & Saifi, R. (2022). Temporal and spatial impact of lockdown during COVID-19 on air quality index in Haryana, India. *Scientific Reports, 12*(1), 20046. doi:10.1038/s41598-022-20885-2

McDuffie, E. E., Martin, R. V., Spadaro, J. V., Burnett, R., Smith, S. J., O'Rourke, P., . . . Shah, V. (2021). Source sector and fuel contributions to ambient PM2. 5 and attributable mortality across multiple spatial scales. *Nature communications, 12*(1), 3594.

McNamee, R. (2003). Confounding and confounders. *Occupational and environmental medicine, 60*(3), 227-234.

Mendy, A., Wu, X., Keller, J. L., Fassler, C. S., Apewokin, S., Mersha, T. B., . . . Pinney, S. M. (2021). Air pollution and the pandemic: Long-term PM2. 5 exposure and disease severity in COVID-19 patients. *Respirology, 26*(12), 1181-1187.

Nava, S., Calzolai, G., Chiari, M., Giannoni, M., Giardi, F., Becagli, S., . . . Lucarelli, F. (2020). Source apportionment of PM2. 5 in Florence (Italy) by PMF analysis of aerosol composition records. *Atmosphere, 11*(5), 484.

Otto, E., Opatoki, A., & Luyi, D. (2022). Water, sanitation and hygiene practice among students in secondary school, Ijebu Ode, Nigeria. *Journal of Environmental Science and Economics, 1*(3), 15-19.

Rahimi, I., Chen, F., & Gandomi, A. H. (2023). A review on COVID-19 forecasting models. *Neural Computing and Applications, 35*(33), 23671-23681.

Renard, J.-B., Surcin, J., Annesi-Maesano, I., Delaunay, G., Poincelet, E., & Dixsaut, G. (2022). Relation between PM2. 5 pollution and Covid-19 mortality in Western Europe for the 2020–2022 period. *Science of the total environment, 848*, 157579.

Renard, J.-B., Surcin, J., Annesi-Maesano, I., & Poincelet, E. (2023a). Temporal Evolution of PM2.5 Levels and COVID-19 Mortality in Europe for the 2020–2022 Period. *Atmosphere, 14*(8), 1222.

Renard, J.-B., Surcin, J., Annesi-Maesano, I., & Poincelet, E. (2023b). Temporal Evolution of PM2. 5 Levels and COVID-19 Mortality in Europe for the 2020–2022 Period. *Atmosphere, 14*(8), 1222.

Rendana, M. (2020). Impact of the wind conditions on COVID-19 pandemic: a new insight for direction of the spread of the virus. *Urban climate, 34*, 100680.

Rocklöv, J., & Sjödin, H. (2020). High population densities catalyse the spread of COVID-19. *Journal of travel medicine, 27*(3), taaa038.

Setti, L., Passarini, F., De Gennaro, G., Barbieri, P., Licen, S., Perrone, M. G., . . . Di Gilio, A. (2020). Potential role of particulate matter in the spreading of COVID-19 in Northern Italy: first observational study based on initial epidemic diffusion. *BMJ open, 10*(9), e039338.

Shao, L., Cao, Y., Jones, T., Santosh, M., Silva, L. F., Ge, S., . . . BéruBé, K. (2022). COVID-19 mortality and exposure to airborne PM2. 5: A lag time correlation. *Science of the total environment, 806*, 151286.

Srivastava, A. (2021). COVID-19 and air pollution and meteorology-an intricate relationship: A review. *Chemosphere, 263*, 128297.

Su, Z., Lin, L., Xu, Z., Chen, Y., Yang, L., Hu, H., . . . Luo, S. (2023). Modeling the Effects of Drivers on PM2.5 in the Yangtze River Delta with Geographically Weighted Random Forest. *Remote Sensing, 15*(15), 3826.

Sudin, S., Aziz, A. N. A., Saad, F. S. A., Khalid, N. S., & Ibrahim, I. I. (2021). *Cycling performance prediction based on cadence analysis by using multiple regression.* Paper presented at the Journal of Physics: Conference Series.

Sy, K. T. L., White, L. F., & Nichols, B. E. (2021). Population density and basic reproductive number of COVID-19 across United States counties. *Plos one, 16*(4), e0249271.

University, J. H. time series covid19 confirmed US *https://github.com/CSSEGISandData*.

Van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., . . . Gerber, S. I. (2020). Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England journal of medicine, 382*(16), 1564-1567.

Vittinghoff, E., Shiboski, S. C., Glidden, D. V., & McCulloch, C. E. (2005). Linear regression. *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models*, 69-131.

Wong, D. W., & Li, Y. (2020). Spreading of COVID-19: Density matters. *Plos one, 15*(12), e0242398.

Wong, H. S., Hasan, M. Z., Sharif, O., & Rahman, A. (2023). Effect of total population, population density and weighted population density on the spread of Covid-19 in Malaysia. *Plos one, 18*(4), e0284157.

Wu, X., Nethery, R. C., Sabath, M. B., Braun, D., & Dominici, F. (2020). Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. *MedRxiv*, 2020.2004. 2005.20054502.

Zoran, M. A., Savastru, R. S., Savastru, D. M., & Tautan, M. N. (2020). Assessing the relationship between surface levels of PM2. 5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. *Science of the total environment, 738*, 139825.