Updated April 5, 2020

Exposure to air pollution and COVID-19 mortality in the United States

Xiao Wu MS, Rachel C. Nethery PhD, M. Benjamin Sabath MA, Danielle Braun PhD, Francesca Dominici PhD

All authors are part of the Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, 02115, USA

Lead authors: Xiao Wu and Rachel C. Nethery

Corresponding and senior author: Francesca Dominici, PhD
Clarence James Gamble Professor of Biostatistics, Population and Data Science
Harvard T.H. Chan School of Public Health
Co-Director Harvard Data Science Initiative
677 Huntington Avenue Boston, MA 02115
410.258.5886
Email: fdominic@hsph.harvard.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

**Background:** United States government scientists estimate that COVID-19 may kill between 100,000 and 240,000 Americans. The majority of the pre-existing conditions that increase the risk of death for COVID-19 are the same diseases that are affected by long-term exposure to air pollution. We investigate whether long-term average exposure to fine particulate matter (PM$_{2.5}$) increases the risk of COVID-19 deaths in the United States.

**Methods:** Data was collected for approximately 3,000 counties in the United States (98% of the population) up to April 04, 2020. We fit zero-inflated negative binomial mixed models using county level COVID-19 deaths as the outcome and county level long-term average of PM$_{2.5}$ as the exposure. We adjust by population size, hospital beds, number of individuals tested, weather, and socioeconomic and behavioral variables including, but not limited to obesity and smoking. We include a random intercept by state to account for potential correlation in counties within the same state.

**Results:** We found that an increase of only 1 $\mu$g/m$^3$ in PM$_{2.5}$ is associated with a 15% increase in the COVID-19 death rate, 95% confidence interval (CI) (5%, 25%). Results are statistically significant and robust to secondary and sensitivity analyses.

**Conclusions:** A small increase in long-term exposure to PM$_{2.5}$ leads to a large increase in COVID-19 death rate, with the magnitude of increase 20 times that observed for PM$_{2.5}$ and all-cause mortality. The study results underscore the importance of continuing to enforce existing air pollution regulations to protect human health both during and after the COVID-19 crisis. The data and code are publicly available.
Introduction

The scale of the COVID-19 public health emergency is an unmatched one in our lifetime. It will have grave social and economic consequences. The suddenness and global scope of this pandemic has raised urgent questions that require coordinated and credentialed information to slow its devastation. A critically important public health objective is to identify key modifiable environmental factors, such as ambient air pollution, that could increase the severity of the health outcomes (e.g., ICU hospitalization and death) among individuals with COVID-19.

Our understanding of what causes death in COVID-19 patients is evolving. Early data from China suggests that a majority of COVID-19 deaths occurred in adults aged ≥60 years and among persons with serious underlying health conditions.\(^1\,^2\) Preliminary analyses of outcomes among COVID-19 patients in the United States largely agree, reporting the highest fatality rates in persons aged ≥65. A report by last month’s joint World Health Organization\(^3\) reported that one in seven patients develops difficulty breathing and other severe complications. These patients typically suffer respiratory failure and failure of other vital systems. COVID-19 can cause viral pneumonia with additional extrapulmonary manifestations and complications including acute respiratory distress syndrome (ARDS) which has a mortality rate ranging from 27% to 45%.\(^4\) Factors associated with mortality include sex (male), advanced age (≥65), and the presence of comorbidities including hypertension, diabetes mellitus, cardiovascular diseases, and cerebrovascular diseases. COVID-19 is also associated with a high inflammatory burden that can induce vascular inflammation, myocarditis, and cardiac arrhythmias.\(^5\)

Although the epidemiology of COVID-19 is evolving, we have determined that there is a large overlap between causes of deaths of COVID-19 patients and the diseases that are affected by
long-term exposure to fine particulate matter (PM$_{2.5}$). The Global Burden of Disease Study identified air pollution as a risk factor for total and cardiovascular disease mortality and is believed to be responsible for 5.5 million premature deaths worldwide a year. PM$_{2.5}$ contains microscopic solids or liquid droplets that are so small they can be inhaled and cause serious health problems. On Thursday, March 26, 2020 the US EPA announced a sweeping relaxation of environmental rules in response to the coronavirus pandemic, allowing power plants, factories and other facilities to determine for themselves if they are able to meet legal requirements on reporting air and water pollution.

We hypothesize that because long-term exposure to PM$_{2.5}$ adversely affects the respiratory and cardiovascular system, it can also exacerbate the severity of the COVID-19 infection symptoms and may increase the risk of death in COVID-19 patients. The association between PM$_{2.5}$ and health including pregnancy outcomes, respiratory diseases, cardiovascular diseases, neurocognitive disease in the United States and worldwide is well established.

A recent study by our group also documented a statistically significant association between long-term exposures to PM$_{2.5}$ and ozone and risk of ARDS among older adults in the United States. Numerous scientific studies reviewed by the US Environmental Protection Agency (US EPA) have linked PM$_{2.5}$ to a variety of health concerns including: premature death in people with heart or lung disease, non-fatal heart attacks, irregular heartbeats, aggravated asthma, decreased lung function, and increased respiratory symptoms such as inflammation, airway irritations, coughing, or difficulty breathing.

Our study includes 3,080 counties in the United States and covering 98% of the United States population. We leverage our previous efforts that focused on estimating the long-term effects of PM$_{2.5}$ on mortality among 60 million United States’ Medicare enrollees. We used a well-
tested research data platform that gathers, harmonizes, and links nationwide air pollution data, census data, and other potential confounding variables with health outcome data. We augment this platform with newly collected COVID-19 data from authoritative data sources. All data sources used in these analyses, along with fully reproducible code, are publicly available to facilitate continued investigation of these relationships as the COVID-19 outbreak evolves and more data become available.

Methods

Table 1 summarizes our data sources and their provenance, including links where the raw data can be extracted directly.

COVID-19 deaths: We obtain COVID-19 death counts for each county in the United States from Johns Hopkins University, Center for Systems Science and Engineering Coronavirus Resource Center. This source provides the most comprehensive county level COVID-19 data to date reported by the US Centers for Disease Control and Prevention and state health departments. This includes real-time number of deaths and number of confirmed cases in each county across the United States. We collect the cumulative number of deaths for each county up to and including April 4, 2020. County level COVID-19 mortality rates are defined for our analyses as the ratio of COVID-19 deaths to county level population size.

Exposure to air pollution: We calculate county level long-term exposure to PM$_{2.5}$ (averaged for 2000 to 2016) from an established exposure prediction models. The PM$_{2.5}$ exposure levels are estimated monthly at 0.01° × 0.01° grid resolution across the entire continental United States by combining satellite, modeled, and monitored PM$_{2.5}$ data in a geographically-weighted regression. These estimates have been extensively cross-validated. We obtained temporally averaged PM$_{2.5}$ values (2000-2016) at the county level by averaging estimated PM$_{2.5}$ values.
within a given county. We compute average 2016 PM$_{2.5}$ exposure analogously for each county to use in sensitivity analyses.

**Potential Confounders:** We consider the following sixteen county level variables and one state level variable as potential confounders: population density, percent of the population $\geq 65$, percent living in poverty, median household income, percent black, percent Hispanic, percent of the adult population with less than a high school education, median house value, percent of owner-occupied housing, population mean BMI (an indicator of obesity), percent ever-smokers, number of hospital beds, and average daily temperature and relative humidity for summer (June-September) and winter (December-February) for each county, and state level number of COVID-19 tests performed. Additional detail on the creation of all variables used in the analysis is available in the Supplementary Materials.

**Statistical methods**

We fit zero-inflated negative binomial mixed models (ZNB).$^{22,23,24}$ using COVID-19 deaths as the outcome and PM$_{2.5}$ as the exposure of interest. The ZNB is composed of two sub-models. The first is a count sub-model that estimates the association between COVID-19 deaths and PM$_{2.5}$ (adjusted by covariates) among counties eligible (e.g., confirmed COVID-19 cases) to experience a COVID-19 death. The second is a zero sub-model that accounts for the excess of zeros that may be generated by counties not yet eligible for COVID-19 deaths (e.g., due to the absence of confirmed COVID-19 cases) and unlikely to have COVID-19 deaths as of April 4, 2020. Additional modeling details are provided in the Supplementary Materials. We include a population size offset and we adjust for all variables listed above. We also include a random intercept by state to account for potential correlation in counties within the same state, due to similar socio-cultural, behavioral, and healthcare system features and similar COVID-19 response and testing policies. We only report the result from the count sub-model. More
specifically we report exponentiated parameter estimates, the mortality rate ratios (MRR) and 95% CI from the count sub-model. The MRR can be interpreted as the relative increase in the COVID-19 death rate associated with a 1 μg/m³ increase in long-term average PM$_{2.5}$ exposure among the counties eligible to experience a COVID-19 death. We do not report results from the zero sub-model. We carried out all analyses in R statistical software and performed model fitting using the NBZIMM package.²⁵

**Secondary Analysis**

We conduct six secondary analyses to assess the robustness of our results to the confounder set used, potential unmeasured confounders, and outliers.

- First, because New York state has experienced the most severe COVID-19 outbreak in the United States to date and has five times higher COVID-19 deaths than the next highest state, we anticipate that it will strongly influence our analysis. As a result, we repeat the analysis excluding all counties in New York state.

- Second, existing COVID-19 testing and case count data are unable to accurately capture the size of an outbreak in a given county, and the inability to fully adjust for this factor could induce confounding in our analyses (e.g., if counties with high PM$_{2.5}$ exposure also tend to have large outbreaks relative to the population size, then their death rates per unit population could appear differentially elevated, inducing a spurious correlation with PM$_{2.5}$). To explore how this may impact our results, we conduct analyses excluding counties with less than 10 confirmed COVID-19 cases.

- Finally, we fit models omitting the following potential confounders from the model (separately):
  - Number of hospital beds
  - Number of COVID-19 tests performed
  - Population mean BMI and percent smokers (BRFSS)
Summer and winter temperature and relative humidity (Weather)

Sensitivity analysis

We conduct several sensitivity analyses to assess the robustness of our results to data and modeling choices.

- First, we repeat all the analyses using alternative methods to estimate exposure to PM$_{2.5}$.26
- Second, because our study relies on observational data, our results could be sensitive to modeling choices (e.g., distributional assumptions or assumptions of linearity). We evaluate sensitivity to such choices by conducting analyses:
  - Treating PM$_{2.5}$ as a categorical variable (categorized at empirical quintiles)
  - Adjusting for population density as a categorical variable (categorized at empirical quintiles)
  - Using a negative binomial model without accounting for zero-inflation.

Results for the sensitivity analyses are shown in Supplementary Materials.

Results

Our study utilized data from 3,080 counties, of which 2,395 (77.8%) have reported zero COVID-19 deaths at the time of this analysis. Table 2 describes the data used in our analyses. All COVID-19 death counts are cumulative counts up to April 4, 2020.

Figure 1 visualizes the spatial variation of long-term average exposure to PM$_{2.5}$ and COVID-19 death rates (per 1 million population) by county. Visual inspection suggests that COVID-19 death rates are higher in the Mid-Atlantic, upper Midwest, gulf coast, and west coast regions. These spatial patterns in COVID-19 death rates generally mimic patterns in both high population density and high PM$_{2.5}$ exposure areas. In the Supplementary Materials, we provide additional visualizations and data diagnostics that justify the use of the ZNB model for our analyses. After
removal of counties with missing confounder values, our main analysis was conducted using
1,783 counties. These counties account for 90% of COVID-19 deaths up to April 4, 2020. In
**Figure 2** we report the mortality rate ratios (MRR) and 95% CI for our main and secondary
analyses. The results of all sensitivity analyses are shown in the Supplementary Materials. In
the main analysis, the estimated MRR is 1.15 (1.05, 1.25). That is, we found that an increase of
only 1 μg/m³ in long-term average PM$_{2.5}$ is associated with a statistically significant increase of
15% in the COVID-19 death rate. This suggests that in New York County (Manhattan) if we
were to lower the long-term average PM$_{2.5}$ exposure by only 1 μg/m³, we would have expected
to see 248 (= 1905 × (1 − 1/1.15)) fewer COVID-19 deaths among a total of 1905 deaths up to
April 04, 2020.

In secondary analyses shown in **Figure 2**, we (a) omit number of hospital beds from the
adjustment set (b) omit number of tested cases in each state from the adjustment set (c) omit
smoking rate and BMI from Behavioral Risk Factor Surveillance System (BRFSS) from the
adjustment set (d) omit summer and winter temperature and relative humidity from the
adjustment set (e) exclude data from New York state (f) exclude counties with less than 10
confirmed COVID-19 cases. The results of these analyses are consistent with the main
analysis, except when the number of hospital beds is omitted from the adjustment set. In that
case, the point estimate decreases, and statistical significance is attenuated. This suggests that
number of hospital beds is a strong confounder.

**Discussion**

This is the first nationwide study in the United States that estimates the relationship between
long-term exposure to PM$_{2.5}$ and COVID-19 death rates. We found statistically significant
evidence that an increase of 1 μg/m³ in long-term PM$_{2.5}$ exposure is associated with a 15%
increase in the COVID-19 mortality rate. Our results are adjusted for a large set of
socioeconomic, demographic, weather, behavioral, and healthcare-related confounders and demonstrated robustness across a wide range of sensitivity analyses.

In our previous study\textsuperscript{27} of 60 million Americans older than 65 years, we found that a 1 mg/m\textsuperscript{3} in long-term PM\textsubscript{2.5} exposure is associated with a 0.73\% increase in the rate of all-cause mortality. Therefore, a small increase in long-term exposure to PM\textsubscript{2.5} leads to a large increase in COVID-19 death rate of a magnitude that is 20 times the one estimated for all-cause mortality.

This analysis provides a timely characterization of the relationship between historical exposure to air pollution COVID-19 deaths in the United States. Research on how modifiable factors may exacerbate COVID-19 symptoms and increase mortality risk is essential to guide policies and behaviors to minimize fatality related to the outbreak. Our analysis relies on up-to-date population-level COVID-19 data and well-validated air pollution exposure measures.\textsuperscript{27}

Strengths of this analysis include adjusting for a wide range of potential confounders and a demonstrated robustness of results to different model choices. Moreover, the analyses rely exclusively on data and code that are publicly available. This provides a platform for the scientific community to continue updating and expanding these analyses as the pandemic evolves, and data accumulate.

The inability to accurately quantify the number of COVID-19 cases due to limited testing capacity presents a potential limitation in that we are unable to fully adjust for the size of the outbreak in a given county, which could be a source of unmeasured confounding. We instead adjust by total population size. However, we anticipate that by using a zero-inflated model we provide partial adjustment for this, as we estimate the MRR accounting for the fact that some counties were not eligible to experience a COVID-19 death because they still had zero
confirmed cases. Moreover, the constantly evolving COVID-19 response measures creates a fluid scenario with regards to the availability of medical resources in a given area. The availability of these resources is likely to impact fatality rates. Because real-time, nationwide data on such measures and resources are not available, we are unable to adjust for this feature. As a result, this could represent another source of confounding.

The results of this paper suggest that long-term exposure to air pollution increases vulnerability to experiencing the most severe COVID-19 outcomes. These findings align with the known relationship between PM$_{2.5}$ exposure and many of the cardiovascular and respiratory comorbidities that dramatically increase the risk of death in COVID-19 patients. They are also consistent with findings that air pollution exposure dramatically increased the risk of death during the Severe Acute Respiratory Syndrome (SARS) outbreak in 2003, which is caused by another type of coronavirus.\textsuperscript{28} This study provides a motive for expanded follow-up investigations as more and higher quality COVID-19 data become available. These would include validating these results in other data sources and study types and studies of biological mechanisms, impacts of PM$_{2.5}$ exposure timing, and relationships between PM$_{2.5}$ and other COVID-19 outcomes such as hospitalization.

The results of this study also underscore the importance of continuing to enforce existing air pollution regulations during the COVID-19 crisis. Based on our result, we anticipate a failure to do so can potentially increase the COVID-19 death toll and hospitalizations, further burdening our healthcare system and drawing resources away from COVID-19 patients.

Acknowledgments:
This work was made possible by the support from the NIH grant R01 ES024332-01A1, P50MD010428, ES024012, ES026217, ES028033; MD012769, HEI grant 4953-RFA14-3/16-4, and USEPA grants 83587201-0, RD-83479801. The contents are solely the responsibility of the grantee and do not necessarily represent the official views of the funding agencies. Further, funding agencies do not endorse the purchase of any commercial products or services related to this publication. The computations in this paper were run on (1) the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University; (2) the Research Computing Environment supported by the Institute for Quantitative Social Science in the Faculty of Arts and Sciences at Harvard University. The authors would like to thank Lena Goodwin, for editorial assistance in the preparation of this manuscript.
<table>
<thead>
<tr>
<th>Source Data</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome: COVID-19 Deaths</strong></td>
<td>Johns Hopkins University the Center for Systems Science and Engineering (CSSE) Coronavirus Resource Center (<a href="https://coronavirus.jhu.edu/">https://coronavirus.jhu.edu/</a>)</td>
</tr>
<tr>
<td><strong>Exposure: PM$_{2.5}$ concentrations</strong></td>
<td>Atmospheric Composition Analysis Group (<a href="http://fizz.phys.dal.ca/~atmos/martin">http://fizz.phys.dal.ca/~atmos/martin</a>)</td>
</tr>
<tr>
<td><strong>Confounders</strong></td>
<td>US Census/American Community Survey (<a href="https://www.census.gov/programs-surveys/acs/data.html">https://www.census.gov/programs-surveys/acs/data.html</a>)</td>
</tr>
<tr>
<td></td>
<td>US CDC BRFSS (<a href="https://www.cdc.gov/brfss/">https://www.cdc.gov/brfss/</a>)</td>
</tr>
<tr>
<td></td>
<td>Homeland Infrastructure Foundation-Level Data (HIFLD) (<a href="https://hifld-geoplatform.opendata.arcgis.com/datasets/hospitals">https://hifld-geoplatform.opendata.arcgis.com/datasets/hospitals</a>)</td>
</tr>
<tr>
<td></td>
<td>The COVID tracking project (<a href="https://covidtracking.com/">https://covidtracking.com/</a>)</td>
</tr>
<tr>
<td></td>
<td>Gridmet via google earth engine (<a href="https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSC_OR_GRIDMET">https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSC_OR_GRIDMET</a>)</td>
</tr>
</tbody>
</table>
### County level risk factors characteristics (SD)

| % Ever smoked | 47.0 (9.2) | 47.0 (9.2) | 45.9 (8.0) | 46.8 (8.9) | 47.0 (9.2) |
| % Below poverty level | 10.5 (5.7) | 10.6 (5.7) | 9.7 (4.7) | 10.3 (5.3) | 10.3 (5.1) |
| % Below high school | 21.2 (10.5) | 21.3 (10.5) | 19.5 (8.6) | 20.6 (9.8) | 21.4 (9.7) |

| PM$_{2.5}$ ($\mu$g/m$^3$) (SD, standard deviation) | 8.4 (2.5) | 8.4 (2.5) | 9.5 (2.3) | 8.5 (2.5) | 9.0 (2.3) |

| Number of deaths | 7,054 | 4,662 | 6,968 | 6,342 | 7,024 |
| Number of hospital beds | 827,799 | 770,716 | 706,051 | 827,799 | 811,070 |
| Number of testing (State level) | 1,620,460 | 1,336,839 |  |  |  |

All rights reserved. No reuse allowed without permission.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Owner-occupied housing</td>
<td>74.2 (8.9)</td>
<td>74.2 (8.8)</td>
<td>70.1 (10.3)</td>
<td>74.1 (9.0)</td>
<td>73.4 (9.1)</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>7.7 (12.3)</td>
<td>7.7 (12.4)</td>
<td>8.6 (11.3)</td>
<td>7.3 (11.4)</td>
<td>7.5 (11.5)</td>
</tr>
<tr>
<td>% Black</td>
<td>8.2 (14.2)</td>
<td>8.3 (14.3)</td>
<td>12.1 (15.4)</td>
<td>8.0 (13.6)</td>
<td>9.2 (13.9)</td>
</tr>
<tr>
<td>% older than 65 year-old</td>
<td>15.9 (4.1)</td>
<td>16.0 (4.1)</td>
<td>13.7 (3.6)</td>
<td>15.8 (4.0)</td>
<td>14.9 (3.5)</td>
</tr>
<tr>
<td>Population density (persons/mile²)</td>
<td>446.0 (2121.62)</td>
<td>377.4 (1162.27)</td>
<td>1044.7 (2854.5)</td>
<td>514.1 (2384.8)</td>
<td>608.3 (2479.9)</td>
</tr>
<tr>
<td>Median household income (1000 $)</td>
<td>49.0 (13.1)</td>
<td>48.8 (13.0)</td>
<td>54.7 (15.6)</td>
<td>49.6 (12.7)</td>
<td>50.1 (13.6)</td>
</tr>
<tr>
<td>Median home value (1000 $)</td>
<td>136.6 (90.5)</td>
<td>135.2 (87.9)</td>
<td>179.8 (114.7)</td>
<td>134.7 (80.4)</td>
<td>148.8 (94.5)</td>
</tr>
</tbody>
</table>
Figure 1: Maps show (a) county level 17-year long-term average of PM$_{2.5}$ concentrations (2000-2016) in the US in $\text{g/m}^3$ and (b) county level number of COVID-19 deaths per one million population in the US up to and including April 4, 2020.
Figure 2: Mortality Risk Ratios (MRR) and 95% confidence intervals. The MRR can be interpreted as percentage increase in the COVID-19 death rate associated with a 1 $\mu g/m^3$ increase in long term average PM$_{2.5}$. The MRR from the main analysis is adjusted by seventeen potential confounders (population density, percent of the population >65, percent living in poverty, median household income, percent black, percent Hispanic, percent of the adult population with less than a high school education, median house value, percent of owner-occupied housing, population mean BMI, percent ever-smokers, summer/winter temperature/relative humidity, and number of hospital beds for each county, and number of COVID-19 tests performed for each state). In secondary analyses, we (a) omit number of hospital beds; (b) omit number of tested cases in each state; (c) omit smoking rate and BMI from BRFSS; (d) omit summer/winter temperature/relative humidity; (e) exclude data from New York state; (f) exclude data from counties with less than 10 confirmed cases. All COVID-19 death counts are cumulative counts up to and including April 04, 2020.

References


(https://www.nature.com/articles/s41591-020-0822-7)


14. Rhee, Jongeun et al. Impact of Long-Term Exposures to Ambient PM$_{2.5}$ and Ozone on ARDS Risk for Older Adults in the United States CHEST, Volume 156, Issue 1, 71 - 79 ([https://journal.chestnet.org/article/S0012-3692(19)30711-1/abstract](https://journal.chestnet.org/article/S0012-3692(19)30711-1/abstract))


20. [https://coronavirus.jhu.edu](https://coronavirus.jhu.edu)

21. Van Donkelaar A, Martin RV, Li C, Burnett RT. Regional estimates of chemical com-position of fine particulate matter using a combined geoscience-statistical method with information


27. http://fizz.phys.dal.ca/~atmos/martin/?page_id=140