

Fine particulate matter and COVID-19 mortality in the United States

A national study on long-term exposure to air pollution and COVID-19 mortality in the United States

Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis

By using the contents on this website and the Github repo, you agree to cite:

Wu, X., Nethery, R. C., Sabath, M. B., Braun, D. and Dominici, F., 2020. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Science advances*, 6(45), p.eabd4049.

Abstract

Assessing whether long-term exposure to air pollution increases the severity of COVID-19 health outcomes, including death, is an important public health objective. Limitations in COVID-19 data availability and quality remain obstacles to conducting conclusive studies on this topic. At present, publicly available COVID-19 outcome data for representative populations are available only as area-level counts. Therefore, studies of long-term exposure to air pollution and COVID-19 outcomes using these data must use an ecological regression analysis, which precludes controlling for individual-level COVID-19 risk factors. We describe these challenges in the context of one of the first preliminary investigations of this question in the United States, where we found that higher historical PM_{2.5} exposures are positively associated with higher county-level COVID-19 mortality rates after accounting for many area-level confounders. Motivated by this study, we lay the groundwork for future research on this important topic, describe the complex challenges, and outline promising directions and opportunities.

Strengths and Limitations

Table 2. Strengths and limitations of ecological regression analyses applied to research on air pollution and COVID-19 and opportunities for future research.

	Strengths	Limitations	Future research
Study design: ecological regression	Feasible, timely, and cost-effective	Cannot be used to make inference about individual-level associations, doing so leads to ecological fallacy	Augment county-level data with individual-level data to adjust for ecological bias (12)
	Data are representative of the entire U.S. population	Cannot adjust for individual-level risk factors such as age, gender, and race (19–21)	Conduct studies of individual-level health records using traditional regression and causal inference methods as in Wu <i>et al.</i> (2)
	Allows inference at the area level, which can be useful for policy-making (11)	Results are sensitive to the assumptions of the statistical model (11)	
	Computationally efficient and can be conducted daily to allow for the dynamic nature of the data and observe temporal trends; see fig. S3		
Facilitates comparison of results across countries			
Outcome: COVID-19 deaths aggregated at the county level	Publicly available data updated almost daily	Potential for outcome misclassification (22), particularly differential misclassification over time and space, which could bias results	Access to nationwide registry data with the validated cause of death (14) Analyses using county excess deaths as the outcome (23)
Exposure: 2000–2016 average exposure to PM _{2.5} at the county level	Use of well-validated atmospheric chemistry models and machine learning models (9, 24)	Aggregation assumes that everyone in a county experiences the same exposures, leading to exposure misclassification, especially for the largest counties	Individual-level data on COVID-19 deaths with geocoded addresses to link to air pollution data at the place of residence
	PM _{2.5} exposure estimated at fine grids, which can be aggregated to the county level to assess exposure even in unmonitored areas (24)	Can be used to assess historical exposures to air pollution but not real-time exposures	Additional statistical methods to account for the propagation of exposure error into the ecological regression model (13)
	As opposed to using monitor data, aggregation of modeled estimates ensures that county PM _{2.5} exposure estimates represent the distribution across the entire area		
Measured confounders	More than 20 area-level variables capture age distribution, race distribution, socioeconomic status, population density, behavioral risk factors, epidemic stage, and stay-at-home orders (see tables S1 and S2)	County average features may not represent the features of COVID-19 patients, leading to inadequate adjustment	Causal inference approaches to adjust for measured confounding bias, producing results that are less sensitive to statistical modeling assumptions
	These overlap with the confounder sets used in much of the previous literature on air pollution and health (25, 26)	Difficult to formalize the notion of “epidemic stage,” which may be an important confounder	
		The threat of unmeasured confounding bias still present	Causal inference approaches to assess covariate balance (2)
		Sensitive to the form of the statistical model specified (i.e., assumptions of linearity and no effect modification)	Individual-level data on key measured confounders such as smoking and body mass index
Unmeasured confounders	Leverage existing approaches, such as the calculation of the E-value (27), to assess how strong the effect of an unmeasured confounder would need to be to explain away the associations detected (see section S3)	The most important threat to the validity of any observational study Even measures like the E-value cannot inform us about the likelihood that a strong unmeasured confounder exists; this must be evaluated on the basis of subject matter knowledge	Natural experiment designs and instrumental variables can be used to reduce the threat of unmeasured confounding but are less common

Summary of the Analysis

Background: United States government scientists estimate that COVID-19 may kill tens of thousands of Americans. Many of the pre-existing conditions that increase the risk of death in those with COVID-19 are the same diseases that are affected by long-term exposure to air pollution. We investigated whether long-term average exposure to fine particulate matter (PM_{2.5}) is associated with an increased risk of COVID-19 death in the United States.

Design: A nationwide, cross-sectional ecological study using county-level data.

Data sources: COVID-19 death counts were collected for more than 3,000 counties in the United States (representing 98% of the population) from Johns Hopkins University, Center for Systems Science and Engineering Coronavirus Resource Center.

Methods: We fit 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. In the main analysis, we adjusted by 20 potential confounding factors including population size, age distribution, population density, time since the beginning of the outbreak, time since state's issuance of stay-at-home order, hospital beds, number of individuals tested, weather, and socioeconomic and behavioral variables such as obesity and smoking. We included a random intercept by state to account for potential correlation in counties within the same state. We conducted more than 80 additional sensitivity analyses.

Results: We found that higher historical PM_{2.5} exposures are positively associated with higher county-level COVID-19 mortality rates after accounting for many area-level confounders. The results were 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 the COVID-19 death rate. Despite inherent limitations of the ecological study design, our 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 so our analyses can be updated routinely.

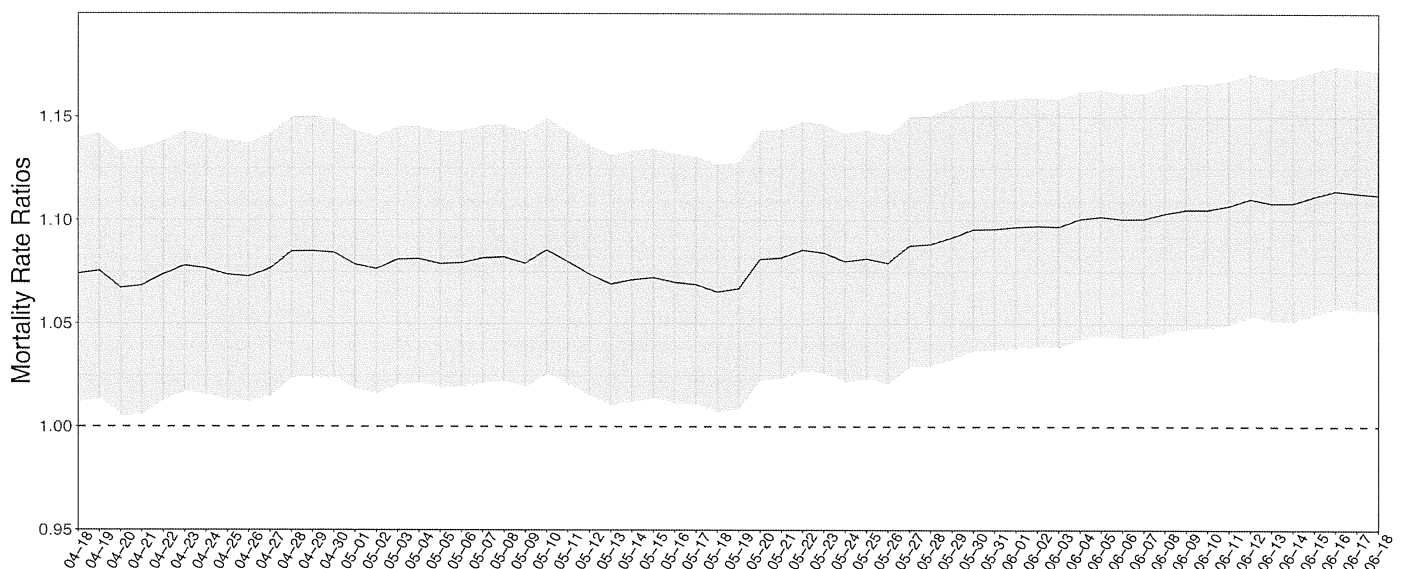


Figure: COVID-19 mortality rate ratios (MRR) per 1 µg/m³ increase in PM_{2.5} and 95% CI using daily cumulative COVID-19 death counts from April 18, 2020 to June 18, 2020. Our published paper included results until June 18, 2020, and we provide the **unpublished updated results** until September 7, 2020.

Data and Code:

Our data and code is available on github [here](#). (Updated Oct 23, 2020)

Manuscript and Supplemental Material

- *Science Advances* Manuscript: <https://advances.sciencemag.org/content/6/45/eabd4049>
- *Science Advances* Editorial: <https://advances.sciencemag.org/content/6/45/eabf1897>
- By using the contents on this website and the Github repo, you agree to cite:

1. Wu, X., Nethery, R. C., Sabath, M. B., Braun, D. and Dominici, F., 2020. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Science advances*, 6(45), p.eabd4049.
2. A earlier pre-print version can be found at: Exposure to air pollution and COVID-19 mortality in the United States. Xiao Wu, Rachel C. Nethery, Benjamin M. Sabath, Danielle Braun, Francesca Dominici. medRxiv 2020.04.05.20054502; doi: <https://doi.org/10.1101/2020.04.05.20054502>

Acknowledgments

We appreciate the work of Aaron Van Donkelaar, Randall Martin, and his team for providing us with access to their estimates of PM_{2.5} exposure. Their data (V4.NA.02.MAPLE) can be found on Randall Martin's website here: <https://sites.wustl.edu/acag/datasets/surface-pm2-5/>

The data was produced as part of the following paper:

van Donkelaar, A., R. V. Martin, C. Li, R. T. Burnett, Regional Estimates of Chemical Composition of Fine Particulate Matter using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors, *Environ. Sci. Technol.*, doi: 10.1021/acs.est.8b06392, 2019.

We would like to thank Lena Goodwin and Stacey Tobin for editorial assistance in the preparation of this manuscript.

Website updated on November 4, 2020.