CORONAVIRUS

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

X. Wu[1]*, R. C. Nethery[1]*, M. B. Sabath[1], D. Braun[1,2], F. Dominici[1]†

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 expo sure 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 challenges, and outline promising directions and opportunities.

INTRODUCTION

The suddenness and global scope of the coronavirus disease 2019 (COVID-19) pandemic have raised urgent questions that require coordinated investigation to slow the disease's devastation. A critically important public health objective is to identify key modifiable environ mental factors that may contribute to the severity of health outcomes [e.g., intensive care unit (ICU) hospitalization and death] among individuals with COVID-19. Numerous scientific studies reviewed by the U.S. Environmental Protection Agency (EPA) have linked fine particles ($PM_{2.5}$; particles with diameter, ≤ 2.5 μm) to a variety of adverse health events (*1*) including death (*2*). It has been hypothe sized that because long-term exposure to $PM_{2.5}$ adversely affects the respiratory and cardiovascular systems and increases mortality risk (*3*–*5*), it may also exacerbate the severity of COVID-19 symp toms and worsen the prognosis of this disease (*6*).

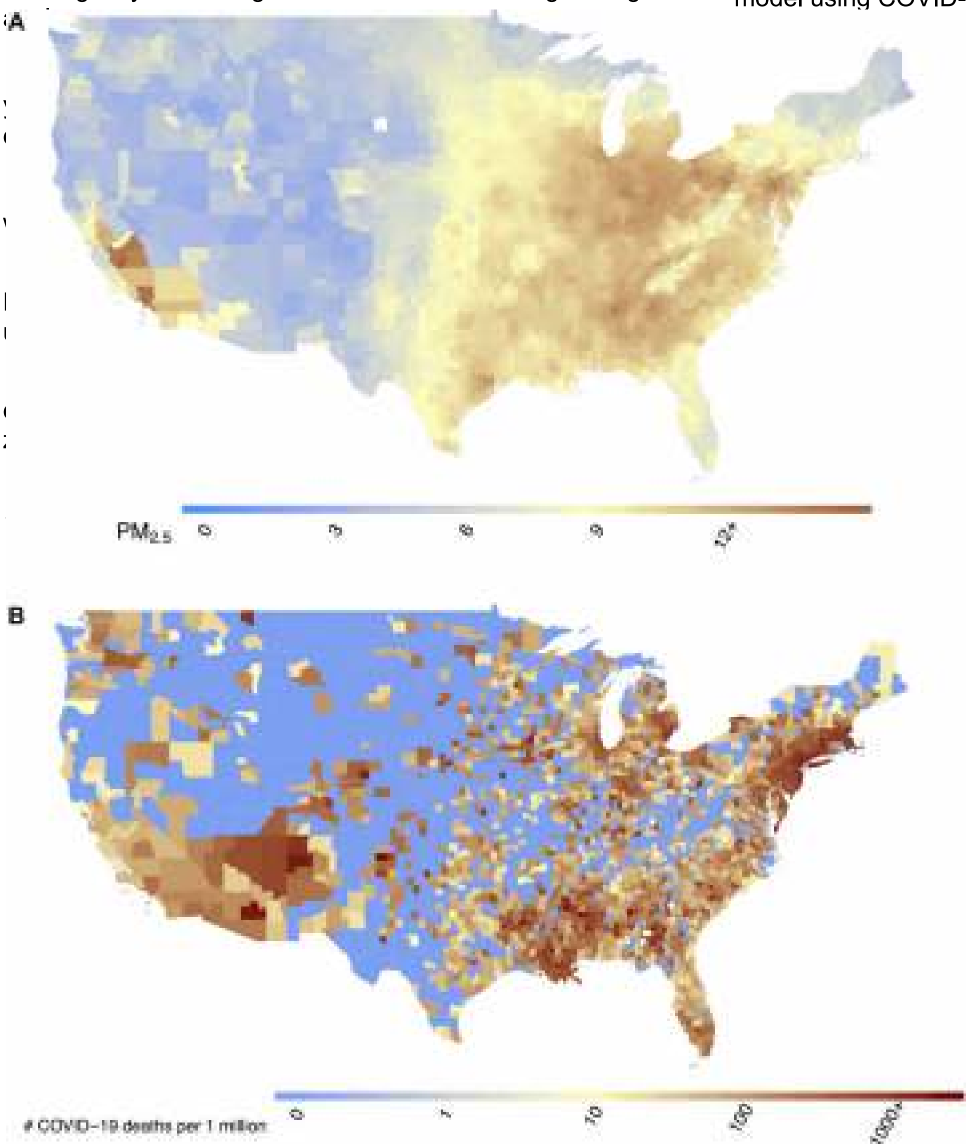
Epidemiological studies to estimate the association between long-term exposure to air pollution and COVID-19 hospitalization and death is a rapidly expanding area of research that is attracting attention around the world. Two studies have been published using data from European countries (*7*, *8*), and many more are available as preprints. However, because of the unprecedented nature of the pandemic, researchers face serious challenges when conducting these studies. One key challenge is that, to our knowledge, individual-level data on COVID-19 health outcomes for large, representative popula tions are not publicly available or accessible to the scientific com munity. Therefore, the only way to generate preliminary evidence on the link between $PM_{2.5}$ and COVID-19 severity and outcomes using these aggregate data is to use an ecological regression analysis. With this study design, publicly available area-level COVID-19 mortality rates are regressed against area-level air pollution concentra tions while accounting for area-level potential confounding factors. Here, we discuss the strengths and limitations of conducting eco

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [2]Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA. *These authors contributed equally to this work. †Corresponding author. Email: fdominic@hsph.harvard.edu

logical regression analyses of air pollution and COVID-19 health outcomes and describe additional challenges related to evolving data quality, statistical modeling, and control of measured and un measured confounding, paving the way for future research on this topic. We discuss these challenges and illustrate them in the context of a specific study, in which we investigated the impact of long-term $PM_{2.5}$ exposure on COVID-19 mortality rates in 3089 counties in the United States, covering 98% of the population.

Illustration of an ecological regression analysis of historical exposure to $PM_{2.5}$ and COVID-19 mortality rate

We begin by describing how to conduct an ecological regression



COVID-19 deaths at the time of our analysis. Daily $PM_{2.5}$ concentrations were estimated across the United States on a 0.01° × 0.01° grid for the period 2000–2016 using well-validated atmospheric chemistry and machine learning models (*9*). We used zonal statistics to aggregate $PM_{2.5}$ concentration estimates to the county level and then averaged across the period 2000–2016 to perform health outcome analyses. Figure 1 illustrates the spatial variation in 2000–2016 average (hereafter referred to as "long-term average") $PM_{2.5}$ concentrations and COVID-19 mortality rates (per 1 million population) by county. We fit a negative binomial mixed model using COVID-19 mortality rates as the outcome and

e $PM_{2.5}$ as the exposure of interest,
unty-level covariates. We conducted
vity analyses to assess the robustness
ious modeling assumptions. We found
l µg/m^3 in the long-term average $PM_{2.5}$
statistically significant 11% (95% CI, 6
the county's COVID-19 mortality rate
association continues to be stable as
ulate (fig. S3). We also found that
ays since the first COVID-19 case was
an household income, percent of
using, percent of the adult population
less than high school

Fig. 1. National maps of historical $PM_{2.5}$ concentrations and COVID-19 deaths. Maps show (**A**) county-level 17-year long-term average of $PM_{2.5}$ concentrations (2000–2016) in the United States in $\mu g/m^3$ and (**B**) county-level number of COVID-19 deaths per 1 million population in the United States up to and including 18 June 2020.

education, age distribution, and percent of Black residents are im portant predictors of the COVID-19 mortality rate in the model. We found a 49% (95% CI, 38 and 61%) increase in COVID-19 mortality rate associated with a 1-SD (per 14.1%) increase in percent Black resi dents of the county. Details on the data sources, statistical methods, and analyses are summarized in the Supplementary Materials. All data sources used in the analyses, along with fully reproducible code, are publicly available at https://github.com/wxwx1993/PM_COVID.

Strengths and limitations of an ecological regression analysis

Ecological regression analysis provides a simple and cost-effective approach for studying potential associations between historical ex posure to air pollution and increased vulnerability to COVID-19 in large representative populations, as illustrated in our study in the

previous section. This approach is regularly applied in many areas of research (*10*). Using our study as an example, we summarize in Table 2 the strengths, limitations, and opportunities considering (i) study design, (ii) COVID-19 health outcome data, (iii) historical exposure to air pollution, and (iv) measured and unmeasured confounders, with the goal of paving the way for future research.

Among the key limitations, by design, ecological regression analyses are unable to adjust for individual-level risk factors (e.g., age, race, and smoking status); when individual-level data are unavailable, this approach leaves us unable to make conclusions re garding individual-level associations. In the context of COVID-19 health outcomes, this is a severe limitation, as individual-level risk factors are known to affect COVID-19 health outcomes. It is im portant to note that confusion between ecological associations and individual associations may present an ecological fallacy. In extreme

cases, this fallacy can lead to associations detected in ecological regression that do not exist or are in the opposite direction of true associations at the individual level. However, ecological regression analyses still allow us to make conclusions at the area level, which can be useful for policy-making (*11*). For the association between COVID-19 health outcomes and $PM_{2.5}$ exposure, we argue that area-level conclusions are valuable, as they can inform important immediate policy actions that will benefit public health, such as

Table 1. Mortality rate ratios (MRR), 95% confidence intervals (CI), and *P* values for all variables in the main analysis. Details of the statistical models are available in section S2. Q, quintile.

	MRR	95% CI	*P* value
$PM_{2.5}$	1.11	(1.06–1.17)	0.00
Population density (Q2)	0.91	(0.71–1.15)	0.42
Population density (Q3)	0.91	(0.71–1.16)	0.45
Population density (Q4)	0.74	(0.57–0.95)	0.02
Population density (Q5)	0.92	(0.69–1.23)	0.56
% In poverty	1.04	(0.96–1.12)	0.31
Log(median house value)	1.13	(0.99–1.29)	0.07
Log(median household income)	1.19	(1.04–1.35)	0.01
% Owner-occupied housing	1.12	(1.04–1.20)	0.00
% Less than high school education	1.20	(1.10–1.32)	0.00
% Black	1.49	(1.38–1.61)	0.00
% Hispanic	1.06	(0.97–1.16)	0.23
% ≥ 65 years of age	1.04	(0.93–1.17)	0.46
% 45–64 years of age	0.77	(0.67–0.90)	0.00
% 15–44 years of age	0.76	(0.68–0.85)	0.00
Days since stay-at-home order	1.18	(0.92–1.52)	0.20
Days since first case	2.40	(2.05–2.80)	0.00
Rate of hospital beds	1.00	(0.93–1.08)	0.95
% Obese	0.96	(0.90–1.03)	0.32
% Smokers	1.13	(1.00–1.28)	0.05
Average summer temperature (°F)	1.11	(0.95–1.30)	0.20
Average winter temperature (°F)	0.86	(0.69–1.07)	0.19
Average			

(i) prioritization of precautionary measures [e.g., personal protective equipment (PPE) allocations and hospital beds] to areas with historical higher air pollution and (ii) further strengthening the scientific argument for lowering the U.S. National Ambient Air Quality Standards for $PM_{2.5}$ and other pollutants. To completely avoid potential ecological bias, a representative sample of individual-level data is necessary. While this may not be feasible in the near future, as some COVID-19 outcome data become available at the individual level, existing approaches that augment county-level data with individual-level data (*12*) could be used to correct for ecological bias.

Furthermore, air pollution exposure misclassification, due to between-area mobility and within-area variation, is another potential source of bias that could affect the ecological regression results described in our example study. Methods to account for the propagation of exposure error into the ecological regression model (*13*) could be applied to help mitigate the impact of measurement error. Outcome misclassification is another limitation that can be partially overcome by accessing nationwide registry data with the validated cause of death (*14*). As in all observational studies, adjustment for measured and unmeasured confounding presents another key challenge in ecological regression analyses, which may be exacerbated when dealing with dynamic pandemic data, as in our study. Conducting studies using both traditional regressions and methods for causal inference as in Wu *et al.* (*2*) is necessary to assess the robustness of the findings.

Increasing the scientific rigor of research in this area requires access to representative, individual-level data on COVID-19 health outcomes, including information about patients' residential address, demographics, and individual-level confounders. This is an enormous challenge that will require consideration of many privacy, legal, and ethical trade-offs (*14*). Future areas of research also include the application of statistical methods to quantify and correct for ecological bias and measurement error, reproducible methods for causal inference, and sensitivity analysis of measured and unmeasured confounding bias as suggested above. These strengths and limitations are illustrated further in the context of our own study (see the Supplementary Materials).

summer relative humidity (%)				
Average winter relative humidity (%)	0.93 (0.80–1.09)	0.38	0.97 (0.87–1.07)	0.52

DISCUSSION

Ecological regression analyses are crucial to stimulate innovations in a rapidly evolving area of research. Ongoing research has already focused on overcoming some aspects of these limitations (*8*, *15*). For example, ecological regression analysis of air pollution and COVID-19, using data with finer geographic resolution, is being conducted for different countries and regions around the world. Cole *et al.* (*8*) published an ecological regression analysis using data in Dutch municipalities and found results consistent with our own investigation; the California Air Resources Board (CARB) is planning to conduct a similar study at the census tract level (*15*). Although an ecological regression analysis cannot provide insight into the mechanisms underlying the relationship between $PM_{2.5}$ exposure and COVID-19 mortality, studies are starting to shed light on the potential biological mechanisms that may explain the relationship between air pollution and viral infection outcomes (*16*). For example, it has been hypothesized that chronic exposure to $PM_{2.5}$ causes alveolar angiotensin-converting enzyme 2 (ACE-2) receptor overexpression and impairs host defenses (*17*). This could cause a more severe form

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 methods as in Wu *et al.* (*2*)	Feasible, timely, and cost-effective Data are representative of the entire U.S. population Allows inference at the area level, which can be useful for policy-making (*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 Publicly available data updated almost daily	Cannot be used to make inference about individual-level associations, doing so leads to ecological fallacy Cannot adjust for individual-level risk factors such as age, gender, and race (*19*–*21*) Results are sensitive to the assumptions of the statistical model (*11*)	Augment county-level data with individual-level data to adjust for ecological bias (*12*) Conduct studies of individual-level health records using traditional regression and causal inference
Outcome: COVID-19 deaths aggregated at the county level		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*) Individual-level data on COVID-19 deaths with geocoded addresses to link to air pollution data at the place of residence
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*) $PM_{2.5}$ exposure estimated at fine grids, which can be aggregated to the county level to assess exposure even in unmonitored areas (*24*) 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	Aggregation assumes that everyone in a county experiences the same exposures, leading to exposure misclassification, especially for the largest counties 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*)

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)

These overlap with the confounder sets used in much of the previous literature on air pollution and health (*25*, *26*)

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)

County average features may not represent the features of COVID-19 patients, leading to inadequate adjustment

Difficult to formalize the notion of “epidemic stage,” which may be an important confounder

The threat of unmeasured confounding bias still present

Sensitive to the form of the statistical model specified (i.e., assumptions of linearity and no effect modification)

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

Causal inference approaches to adjust for measured confounding bias, producing results that are less sensitive to statistical modeling assumptions

Causal inference approaches to assess covariate balance (*2*)

Individual-level data on key measured confounders such as smoking and body mass index

Natural experiment designs and instrumental variables can be used to reduce the threat of unmeasured confounding but are less common

of COVID-19 in ACE-2–depleted lungs, increasing the likelihood of poor outcomes, including death (*18*).

The associations detected in ecological regression analyses provide strong justification for follow-up investigations as more and higher quality COVID-19 data become available. Such studies would include validation of our findings with other data sources and study types, as well as investigations into mediating factors and effect modifiers, biological mechanisms, impacts of $PM_{2.5}$ exposure timing, and re lationships between $PM_{2.5}$ and other COVID-19 outcomes such as hospitalization. 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 pandemic. Such research could also provide a strong scientific argument for revision of the U.S. Ambient Air Quality Standards for $PM_{2.5}$ and other environmental policies in the midst of a pandemic.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/ content/full/6/45/eabd4049/DC1

REFERENCES AND NOTES

1. U.S. Environmental Protection Agency (EPA), Integrated Science Assessment (ISA) for Particulate Matter (Final Report, 2019). U.S. Environmental Protection Agency, EPA/600/R-19/188 (2019); https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=347534 [accessed 15 September 2020].
2. X. Wu, D. Braun, J. Schwartz, M. A. Kioumourtzoglou, F. Dominici, Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci. Adv.* **6**, eaba5692 (2020).
3. R. D. Brook, S. Rajagopalan, C. Arden Pope III, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman, A. Peters, D. Siscovick, S. C. Smith Jr., L. Whitsel, J. D. Kaufman, Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation* **121**, 2331–2378 (2010).
4. C. A. Pope III, R. T. Burnett, G. D. Thurston, M. J. Thun, E. E. Calle, D. Krewski, J. J. Godleski, Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. *Circulation* **109**, 71–77 (2004).
5. C. A. Pope III, N. Coleman, Z. A. Pond, R. T. Burnett, Fine particulate air pollution and human mortality: 25+ years of cohort studies. *Environ. Res.* **183**, 108924 (2020). 6. T. Benmarhnia, Linkages between air pollution and the health burden from COVID-19: Methodological challenges and opportunities. *Am. J. Epidemiol.* , kwaa148 (2020). 7. E. Conticini, B. Frediani, D. Caro, Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ. Pollut.* **261**, 114465 (2020).
8. M. A. Cole, C. Ozgen, E. Strobl, Air pollution exposure and COVID-19 in Dutch municipalities. *Environ. Resour. Econ. (Dordr)* , 1–30 (2020).
9. A. van Donkelaar, 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.* **53**, 2595–2611 (2019).
10. G. King, M. A. Tanner, O. Rosen, *Ecological Inference: New Methdological Strategies* (Cambridge Univ. Press, 2004).
11. A. Gelman, D. K. Park, S. Ansolabehere, P. N. Price, L. C. Minnite, Models, assumptions and model checking in ecological regressions. *J. Royal Stat. Soc.*

Ser. A **164**, 101–118 (2001).
12. C. Jackson, N. Best, S. Richardson, Improving ecological inference using individual-level data. *Stat. Med.* **25**, 2136–2159 (2006).
13. J. Richmond-Bryant, T. C. Long, Influence of exposure measurement errors on results from epidemiologic studies of different designs. *J. Expo. Sci. Environ. Epidemiol.* **30**, 420–429 (2020).
14. D. F. Sittig, H. Singh, COVID-19 and the need for a national health information technology infrastructure. *JAMA* **323**, 2373–2374 (2020).
15. InisdeEPA.com, CARB study on pollution link to virus deaths may spur push for PM cuts (May 6, 2020); https://insideepa.com/daily-news/carb-study-pollution-link-virus-deaths may-spur-push-pm-cuts.
16. J. Ciencewicki, I. Jaspers, Air pollution and respiratory viral infection. *Inhal. Toxicol.* **19**, 1135–1146 (2007).

17. L. Miyashita, G. Foley, S. Semple, J. Grigg, Traffic-derived particulate matterand angiotensin-converting enzyme 2 expression in human airway epithelial cells. *bioRxiv* 2020.2005.2015.097501 , (2020).
18. A. Frontera, L. Cianfanelli, K. Vlachos, G. Landoni, G. Cremona, Severe air pollution links to higher mortality in COVID-19 patients: The "double-hit" hypothesis. *J. Infect.* **81**, 255–259 (2020).
19. S. Greenland, H. Morgenstern, Ecological bias, confounding, and effect modification. *Int. J. Epidemiol.* **18**, 269–274 (1989).
20. S. Greenland, J. Robins, Invited commentary: Ecologic studies–Biases, misconceptions, and counterexamples. *Am. J. Epidemiol.* **139**, 747–760 (1994).
21. S. Richardson, I. Stücker, D. Hémon, Comparison of relative risks obtained in ecological and individual studies: Some methodological considerations. *Int. J. Epidemiol.* **16**, 111–120 (1987).
22. New York City Department of Health; Mental Hygiene Covid-Response Team, Preliminary estimate of excess mortality during the COVID-19 outbreak–New York City, March 11–May 2, 2020. *MMWR Morb. Mortal Wkly. Rep.* **69**, 603–605 (2020).
23. R. J. Acosta, R. A. Irizarry, Monitoring health systems by estimating excess mortality. *medRxiv* 2020.06.06.20120857 , (2020).
24. Q. Di, I. Kloog, P. Koutrakis, A. Lyapustin, Y. Wang, J. Schwartz, Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* **50**, 4712–4721 (2016).
25. Q. Di, Y. Wang, A. Zanobetti, Y. Wang, P. Koutrakis, C. Choirat, F. Dominici, J. D. Schwartz, Air pollution and mortality in the Medicare population. *N. Engl. J. Med.* **376**, 2513–2522 (2017).
26. D. W. Dockery, C. A. Pope III, X. Xu, J. D. Spengler, J. H. Ware, M. E. Fay, B. G. Ferris Jr., F. E. Speizer, An association between air pollution andmortality insix US cities. *N. Engl. J. Med.* **329**, 1753–1759 (1993).
27. S. Haneuse, T. J. VanderWeele, D. Arterburn, Using the E-value to assess the potential effect of unmeasured confounding in observational studies. *JAMA* **321**, 602–603 (2019).
28. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
29. Q. Di, H. Amini, L. Shi, I. Kloog, R. Silvern, J. Kelly, M. B. Sabath, C. Choirat, P. Koutrakis, A. Lyapustin, Y. Wang, L. J. Mickley, J. Schwartz, An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. *Environ. Int.* **130**, 104909 (2019).
30. J. G. Booth, G. Casella, H. Friedl, J. P. Hobert, Negative binomial loglinear mixed models. *Stat. Model.* **3**, 179–191 (2003).
31. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2018); www.R-project.org/.
32. X. Zhang, H. Mallick, Z. Tang, L. Zhang, X. Cui, A. K. Benson, N. Yi, Negative binomial mixed models for analyzing microbiome count data. *BMC Bioinformatics* **18**, 4 (2017).
33. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
34. Q. H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333 (1989).
35. T. J. VanderWeele, P. Ding, Sensitivity analysis in observational research: Introducing the E-value. *Ann. Intern. Med.* **167**, 268–274 (2017).
36. M. B. Mathur, P. Ding, C. A. Riddell, T. J. VanderWeele, Web site and R package for computing E-values. *Epidemiology* **29**, e45–e47 (2018).
37. D. D. Ingram, S. J. Franco, 2013 NCHS urban-rural classification scheme for counties. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics (2014).
38. Health Effects Institute, State of global air 2019: A special report on global exposure to air pollution and its disease burden (2019); www.stateofglobalair.org/sites/default/files/ soga_2019_report.pdf [accessed 15 September 2020].
39. R. D. Brook, B. Franklin, W. Cascio, Y. Hong, G. Howard, M. Lipsett, R. Luepker, M. Mittleman, J. Samet, S. C. Smith Jr., I. Tager, Air pollution and cardiovascular disease: A statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation* **109**, 2655–2671 (2004).
40. F. Dominici, R. D. Peng, M. L. Bell, L. Pham, A. McDermott, S. L. Zeger, J. M. Samet, Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA* **295**, 1127–1134 (2006).
41. R. C. Puett, J. E. Hart, J. D. Yanosky, C. Paciorek, J. Schwartz, H. Suh, F. E. Speizer, F. Laden, Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Environ. Health Perspect.* **117**, 1697–1701 (2009).
42. R. J. Šrám, B. Binková, J. Dejmek, M. Bobak, Ambient air pollution and pregnancy outcomes: A review of the literature. *Environ. Health Perspect.* **113**, 375–382 (2005).
43. G. A. Wellenius, M. R. Burger, B. A. Coull, J. Schwartz, H. H. Suh, P. Koutrakis, G. Schlaug, D. R. Gold, M. A. Mittleman, Ambient air pollution and the risk of acute ischemic stroke. *Arch. Intern. Med.* **172**, 229–234 (2012).

44. J. Rhee, F. Dominici, A. Zanobetti, J. Schwartz, Y. Wang, Q. Di, J. Balmes, D. C. Christiani, Impact of long-term exposures to ambient $PM_{2.5}$ and ozone on ARDS risk for older adults in the United States. *Chest* **156**, 71–79 (2019).
45. Y. Cui, Z.-F. Zhang, J. Froines, J. Zhao, H. Wang, S.-Z. Yu, R. Detels, Air pollution and case fatality of SARS in the People's Republic of China: An ecologic study. *Environ. Health* **2**, 15 (2003).
46. C. A. Pope III, Respiratory disease associated with community air pollution and a steel mill, Utah Valley. *Am. J. Public Health* **79**, 623–628 (1989).
47. D. P. Croft, W. Zhang, S. Lin, S. W. Thurston, P. K. Hopke, E. van Wijngaarden, S. Squizzato, M. Masiol, M. J. Utell, D. Q. Rich, Associations between source-specific particulate matter and respiratory infections in New York State adults. *Environ. Sci. Technol.* **54**, 975–984 (2020).
48. B. D. Horne, E. A. Joy, M. G. Hofmann, P. H. Gesteland, J. B. Cannon, J. S. Lefler,

D. P. Blagev, E. K. Korgenski, N. Torosyan, G. I. Hansen, D. Kartchner, C. A. Pope III, Short-term elevation of fine particulate matter air pollution and acute lower respiratory infection. *Am. J. Respir. Crit. Care Med.* **198**, 759–766 (2018).

49. Q. Di, L. Dai, Y. Wang, A. Zanobetti, C. Choirat, J. D. Schwartz, F. Dominici, Association ofshort-term exposure to air pollution with mortality in older adults. *JAMA* **318**, 2446–2456 (2017).

50. D.-H. Tsai, M. Riediker, A. Berchet, F. Paccaud, G. Waeber, P. Vollenweider, M. Bochud, Effects ofshort- and long-term exposures to particulate matter on inflammatory marker levels in the general population. *Environ. Sci. Pollut. Res. Int.* **26**, 19697–19704 (2019).

51. K. F. Morales, J. Paget, P. Spreeuwenberg, Possible explanations for why some countries were harder hit by the pandemic influenza virus in 2009 – A global mortality impact modeling study. *BMC Infect. Dis.* **17**, 642 (2017).

52. Z. Xu, W. Hu, G. Williams, A. C. A. Clements, H. Kan, S. Tong, Air pollution, temperature and pediatric influenza in Brisbane, Australia. *Environ. Int.* **59**, 384–388 (2013). 53. K. Clay, J. Lewis, E. Severnini, Pollution, infectious disease, and mortality: Evidence from the 1918 Spanish influenza pandemic. *J. Econ. Hist.* **78**, 1179–1209 (2018). 54. C. A. Pope III, A. Bhatnagar, J. P. McCracken, W. Abplanalp, D. J. Conklin, T. O'Toole, Exposure to fine particulate air pollution is associated with endothelial injury and systemic inflammation. *Circ. Res.* **119**, 1204–1214 (2016).

55. S. Becker, J. M. Soukup, Exposure to urban air particulates alters the macrophage mediated inflammatory response to respiratory viral infection. *J. Toxicol. Environ. Health A* **57**, 445–457 (1999).

56. P. M. Kaan, R. G. Hegele, Interaction between respiratory syncytial virus and particulate matter in guinea pig alveolar macrophages. *Am. J. Respir. Cell Mol. Biol.* **28**, 697–704 (2003). 57. A. L. Lambert, J. B. Mangum, M. P. DeLorme, J. I. Everitt, Ultrafine carbon black particles enhance respiratory syncytial virus-induced airway reactivity, pulmonary inflammation, and chemokine expression. *Toxicol. Sci.* **72**, 339–346 (2003).

58. L. Peng, X. Zhao, Y. Tao, S. Mi, J. Huang, Q. Zhang, The effects of air pollution and meteorological factors on measles cases in Lanzhou, China. *Environ. Sci. Pollut. Res. Int.* **27**, 13524–13533 (2020).

59. Q. Ye, J.-f. Fu, J.-h. Mao, S.-q. Shang, Haze is a risk factor contributing to the rapid spread of respiratory syncytial virus in children. *Environ. Sci. Pollut. Res. Int.* **23**, 20178–20185 (2016).

Acknowledgments: The computations in this paper were run on (i) the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University and (ii) the Research Computing Environment supported by Institute for Quantitative Social Science in the Faculty of Arts and Sciences at Harvard University. We gratefully acknowledge support from the 2020 Star-Friedman Challenge for Promising Scientific Research, the Climate Change Solutions Fund at Harvard University, and the Fernholz Foundation. We would like to thank L. Goodwin and S. Tobin for editorial assistance in the preparation of this manuscript. **Funding:** This work was made possible by support from NIH grants R01 ES024332-01A1, P50MD010428, RO1 ES026217, RO1 ES028033, R01 ES030616, R01 AG066793-01, and R01 MD012769; Health Effects Institute grant (HEI) 4953-RFA14-3/16-4: and US EPA grant 83587201-0. The funding sources did not participate in the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript. The research described in this article was conducted under contract to the HEI, an organization jointly funded by the EPA (Assistance Award No. R-83467701), and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers. **Author contributions:** X.W. and R.C.N. contributed equally to the paper.

X.W. and R.C.N. contributed to formulation of the idea, data preparation, data analysis, data interpretation, and writing of the manuscript. M.B.S. and D.B. contributed to data preparation, data interpretation, and review of the manuscript. F.D. contributed to formulation of the idea, study design, data interpretation, funding, and writing of the manuscript. All authors contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. F.D. is the guarantor. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data and code are publicly available at https://github.com/wxwx1993/PM_COVID. Additional data related to this paper may be requested from the authors.

Submitted 25 June 2020

Accepted 18 September 2020

Published 4 November 2020

10.1126/sciadv.abd4049

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

Science Advances

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

X. Wu, R. C. Nethery, M. B. Sabath, D. Braun and F. Dominici

Sci Adv **6** (45), eabd4049.
DOI: 10.1126/sciadv.abd4049

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title Science Advances is a registered trademark of AAAS.