

HEALTH AND MEDICINE

Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly

X. Wu^{1*}, D. Braun^{1,2*}, J. Schwartz³, M. A. Kioumourtzoglou⁴, F. Dominici^{1†}

Many studies link long-term fine particle (PM_{2.5}) exposure to mortality, even at levels below current U.S. air quality standards (12 micrograms per cubic meter). These findings have been disputed with claims that the use of traditional statistical approaches does not guarantee causality. Leveraging 16 years of data—68.5 million Medicare enrollees—we provide strong evidence of the causal link between long-term PM_{2.5} exposure and mortality under a set of causal inference assumptions. Using five distinct approaches, we found that a decrease in PM_{2.5} (by 10 micrograms per cubic meter) leads to a statistically significant 6 to 7% decrease in mortality risk. Based on these models, lowering the air quality standard to 10 micrograms per cubic meter would save 143,257 lives (95% confidence interval, 115,581 to 170,645) in one decade. Our study provides the most comprehensive evidence to date of the link between long-term PM_{2.5} exposure and mortality, even at levels below current standards.

INTRODUCTION

The Clean Air Act requires the U.S. Environmental Protection Agency (EPA) to set National Ambient Air Quality Standards (NAAQS). In 1971, EPA set the first NAAQS to best protect public health and welfare. NAAQS are periodically reviewed/revised on the basis of scientific evidence, resulting in a steady decrease in fine particle (PM_{2.5}; particles with diameter, $\leq 2.5 \mu\text{m}$) concentrations. The association between long-term exposure to PM_{2.5} and mortality is well documented (1–4). Recent studies have examined effect estimates among people who are exposed to PM_{2.5} concentrations below the current NAAQS (3, 5). These studies found that exposure to PM_{2.5} below the U.S. standard is associated with an increased mortality risk.

Some scientists, including the current chair of EPA's Clean Air Scientific Advisory Committee, have argued against including studies that use traditional statistical approaches to informing revisions of the NAAQS and propose focusing on studies that apply causal inference approaches (6). Their main criticism is that traditional approaches that include potential confounders as covariates in the regression model do not inform causality. Recently, Goldman and Dominici (7) argued that although “causal inference approaches tend to be more robust to violation of assumptions, [...] air pollution regulations must be based on existing evidence and demonstrated inference methods that arise from review of existing literature.”

Our study bridges this divide. Using the largest air pollution study cohort to date (nationwide Medicare enrollees 2000–2016), we provide strong evidence of the causal link between long-term PM_{2.5} exposure and mortality under a set of assumptions necessary for causal inference (described in detail in Materials and Methods). We applied traditional and causal inference approaches to the same data, assessed sensitivity to modeling assumptions and study period, and made statistical software available.

RESULTS

We obtained open cohort data for more than 68.5 million Medicare enrollees (65 years of age or older) from 2000 to 2016 (8), including demographic information on age, sex, race/ethnicity, date of death, and residential zip code. Each person was tracked through a unique patient ID. Annual PM_{2.5} exposure estimates were obtained from a previously developed and validated ensemble prediction model (Fig. 1) (9). The model ensembles three machine learning algorithms to predict daily PM_{2.5} at 1-km² grid resolution across the United States using remote sensing, output from chemical transport models, meteorological and land-use variables, and other predictors. As residential addresses for Medicare enrollees are available at the zip code level, we used zonal statistics to aggregate exposure estimates to the zip code level to perform the health analyses. To adjust for confounding, we considered 10 zip code- and county-level confounders including zip code-level socioeconomic status (SES) indicators from the 2000 and 2010 Census and the 2005–2012 American Community Surveys (ACS) and county-level information from the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS). We considered the following zip code-level meteorological variables: summer (June to September) and winter (December to February) average of (i) maximum daily temperatures and (ii) relative humidity in each zip code obtained from gridMET via Google Earth Engine. Materials and Methods and table S1 include details on the variables included in the analyses and the R code for the implementation of the methods (section S4). All study data sources are publicly available.

We implemented five statistical approaches to estimating the effect of PM_{2.5} exposure on mortality, accounting for potential confounders. The two traditional approaches rely on regression modeling for confounding adjustment: (i) Cox proportional hazards model and (ii) Poisson regression. We created multiple observations for each individual, each observation representing a person year of follow-up. We fit Cox hazards models, using follow-up year as the time metric and annual PM_{2.5} as the time-varying exposure, stratifying by age (5-year categories), sex, race/ethnicity, and Medicaid eligibility (a surrogate for individual-level SES). We fit Poisson models for aggregated outcomes (i.e., counts of death) by zip code and year, stratified by the same individual-level characteristics and follow-up year. We adjusted both for confounding by including 10 zip code- or county-level risk

¹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ²Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA. ³Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁴Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: fdominici@hsph.harvard.edu

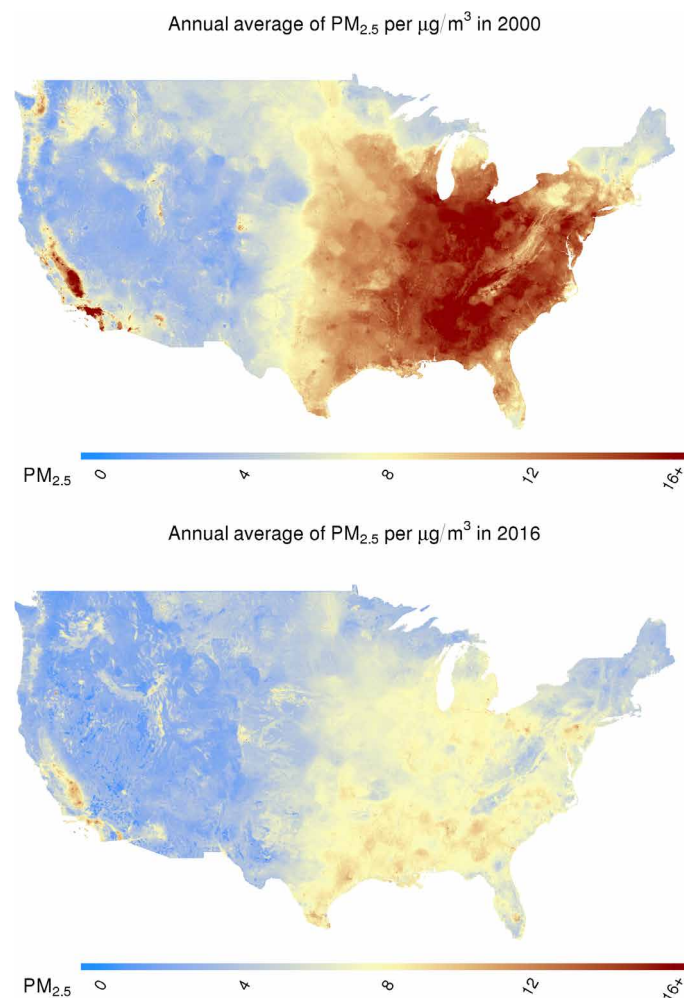


Fig. 1. Annual average $PM_{2.5}$ concentrations in the continental United States for 2000 and 2016.

factors, four zip code-level meteorological variables, and indicators for geographic region (Northeast, South, Midwest, and West). To account for long-term time trends, we included calendar year as a categorical variable.

We also considered three approaches for causal inference that rely on the potential outcomes framework and generalized propensity scores (GPSs). These approaches adjust for confounding using (i) matching by GPS, (ii) weighting by GPS, and (iii) adjustment by GPS, by including GPS as a covariate in the health outcome model. The GPSs were estimated by regressing the $PM_{2.5}$ exposure on confounders using a gradient boosting machine, including the same confounders as in the traditional approaches. Following GPS implementation, we fit a Poisson regression model stratified by individual-level characteristics and follow-up year, as if it was a stratified randomized experiment. All approaches are described in detail in Materials and Methods and in section S1.

All five approaches were fit on the 2000–2016 data. The 2000–2016 cohort consisted of 68,503,979 individuals (573,370,257 person years); we observed 27,106,639 deaths (39.6%; Table 1). Figure 1 shows the average $PM_{2.5}$ concentrations in 2000 and 2016. To estimate low-level $PM_{2.5}$ effects on mortality, we applied the five statistical approaches,

restricting analyses to the subpopulation of Medicare enrollees who were always exposed to $PM_{2.5}$ levels lower than $12 \mu\text{g}/\text{m}^3$ over the entire study period. Additional analysis was conducted on the previously used 2000–2012 cohort. To evaluate the model sensitivity to potential unmeasured confounders that vary over time, all five approaches were fit twice, once with year as a covariate (the main analysis) and once without (as a sensitivity analysis). Results for all additional analyses are shown in section S3. Effect estimates are presented as hazard ratios (HRs) per increase ($10 \mu\text{g}/\text{m}^3$) in annual $PM_{2.5}$. Ninety-five percent confidence intervals (CIs) for all models were evaluated by m-out-n blocked bootstrap to account for spatial correlation. More specifically, we recalculated the GPS and refit the outcome model in each bootstrapped sample to ensure that the bootstrapping procedure jointly accounted for the variability associated both with the GPS estimation and with the outcome model (see section S1).

The causal inference framework lends itself to the evaluation of covariate balance for measured confounders. The covariate balance indicates the quality of the causal inference approach at recovering randomized experiments and informs the degree to which we can make a valid causal assessment. A detailed discussion is provided in section S1. Covariate balance was evaluated using mean absolute correlation (AC), with values of <0.1 indicating high quality in recovering randomized experiments. Figure 2 shows that AC is smaller than 0.1 using causal inference GPS methods (matching and weighting), thus strengthening the interpretability and validity of our analyses as providing evidence of causality. Additional evaluation of covariate balance estimating the standardized mean difference after dichotomizing the continuous exposure is provided in fig. S3.

For the period 2000–2016, we found that all statistical approaches provide consistent results: A decrease ($10 \mu\text{g}/\text{m}^3$) in $PM_{2.5}$ led to a statistically significant decrease in mortality rate ranging between 6 and 7% ($= 1 - 1/\text{HR}$) [HR estimates, 1.06 (95% CI, 1.05 to 1.08) to 1.08 (95% CI, 1.07 to 1.09)]. The estimated HRs were larger when studying the cohort of Medicare enrollees that were always exposed to $PM_{2.5}$ levels lower than $12 \mu\text{g}/\text{m}^3$ [1.23 (95% CI, 1.18 to 1.28) to 1.37 (95% CI, 1.34 to 1.40)] (Fig. 3; corresponding numbers are shown in table S3). Similar results were found for the period 2000–2012 (table S3), also showing consistency with Di *et al.* (3). We reanalyzed all data excluding year as a covariate. The estimated HRs were larger in magnitude, potentially indicating residual confounding bias by some unmeasured confounders with time trends that covary with time trends in the outcome and exposure (table S3). We conducted further sensitivity analyses to unmeasured confounding by calculating the E-value (10, 11). The results, shown in table S5, suggest that our conclusions are overall robust to unmeasured confounding bias.

We estimated the total number of deaths avoided among elderly in a decade if, hypothetically, the U.S. standards followed the World Health Organization (WHO) annual guideline of $10 \mu\text{g}/\text{m}^3$ and all zip codes complied. For this calculation, we used the most conservative HR estimate across all statistical approaches [HR, 1.06 (95% CI, 1.05 to 1.08) and 1.23 (95% CI, 1.18 to 1.28)]. We found that lowering the standards to $10 \mu\text{g}/\text{m}^3$ would have saved 143,257 lives (95% CI, 115,581 to 170,645) in one decade.

DISCUSSION

This study provides the most robust and reproducible evidence to date on the causal link between exposure to $PM_{2.5}$, even at levels below

Table 1. Characteristics for the study cohorts. Note that mean (SD) is presented for continuous variables.

Variables	Entire Medicare enrollees	Medicare enrollees exposed to $PM_{2.5} \leq 12 \mu g/m^3$
Number of individuals	68,503,979	38,366,800
Number of deaths	27,106,639	10,124,409
Total person years	573,370,257	259,469,768
Median years of follow-up	8.0	8.0
Individual-level characteristics		
Age at entry (years)		
65–74 (%)	80.6	88.1
75–84 (%)	14.9	9.0
85–94 (%)	4.1	2.6
95 or above (%)	0.4	0.2
Mean (SD)	69.2 (6.7)	67.6 (5.6)
Sex		
Female (%)	55.5	53.8
Male (%)	44.5	46.2
Race		
White (%)	83.9	84.7
Black (%)	9.1	7.3
Asian (%)	1.8	1.8
Hispanic (%)	2.0	2.2
North American Native (%)	0.3	0.4
Medicaid eligibility		
Eligible (%)	11.7	10.9
Area-level risk factors characteristics		
Ever smoked (%)	47.3	47.3
Below poverty level (%)	10.5	10.1
Below high school education (%)	28.5	25.6
Owner-occupied housing (%)	72.0	72.9
Hispanic (%)	8.9	7.5
Black (%)	8.9	9.2
Population density (persons/km ²)	600.0 (1953.9)	489.1 (1634.0)
Mean BMI (kg/m ²)	27.6 (1.1)	27.6 (1.1)
Median household income (\$1000)	48.9 (21.7)	50.3 (22.0)
Median home value (\$1000)	162.5 (140.9)	170.9 (146.2)
Meteorological variables		
Summer temperature (°C)	29.5 (3.7)	29.5 (3.9)
Winter temperature (°C)	7.6 (7.2)	7.4 (7.6)
Summer relative humidity (%)	88.0 (11.7)	86.7 (12.7)
Winter relative humidity (%)	86.2 (7.3)	86.4 (7.6)
PM_{2.5} concentrations ($\mu g/m^3$)	9.8 (3.2)	8.4 (2.3)

12 $\mu g/m^3$, and mortality among Medicare enrollees. Considering (i) the massive study population, (ii) the numerous sensitivity analyses, and (iii) the transparent assessment of covariate balance that indicates the quality of causal inference for recovering randomized experiments, we conclude that long-term $PM_{2.5}$ exposure is causally related to mortality. This conclusion assumes that the causal in-

ference assumptions hold and, more specifically, that we were able to adequately account for confounding bias. We explored various modeling approaches and conducted extensive sensitivity analyses and found that the results were robust across approaches and models. This work relies on publicly available data, and we provide code that allows for reproducibility of our analyses.

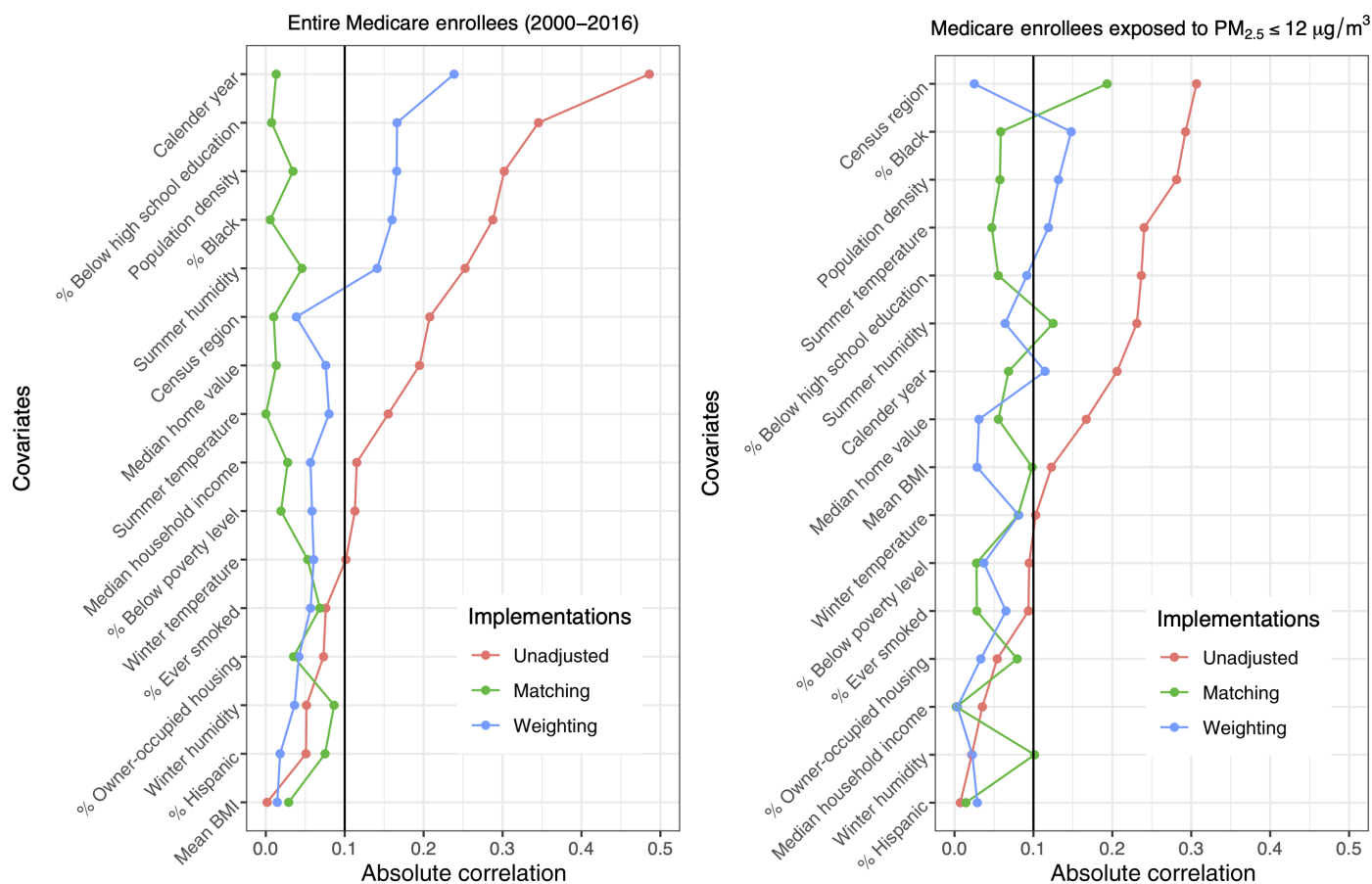


Fig. 2. Mean AC for unadjusted, weighted, and matched populations. Mean AC was smaller than 0.1 using causal inference GPS methods (matching and weighting). AC values of <0.1 indicate good covariate balance, strengthening the interpretability and validity of our analyses as providing evidence of causality. BMI, body mass index.

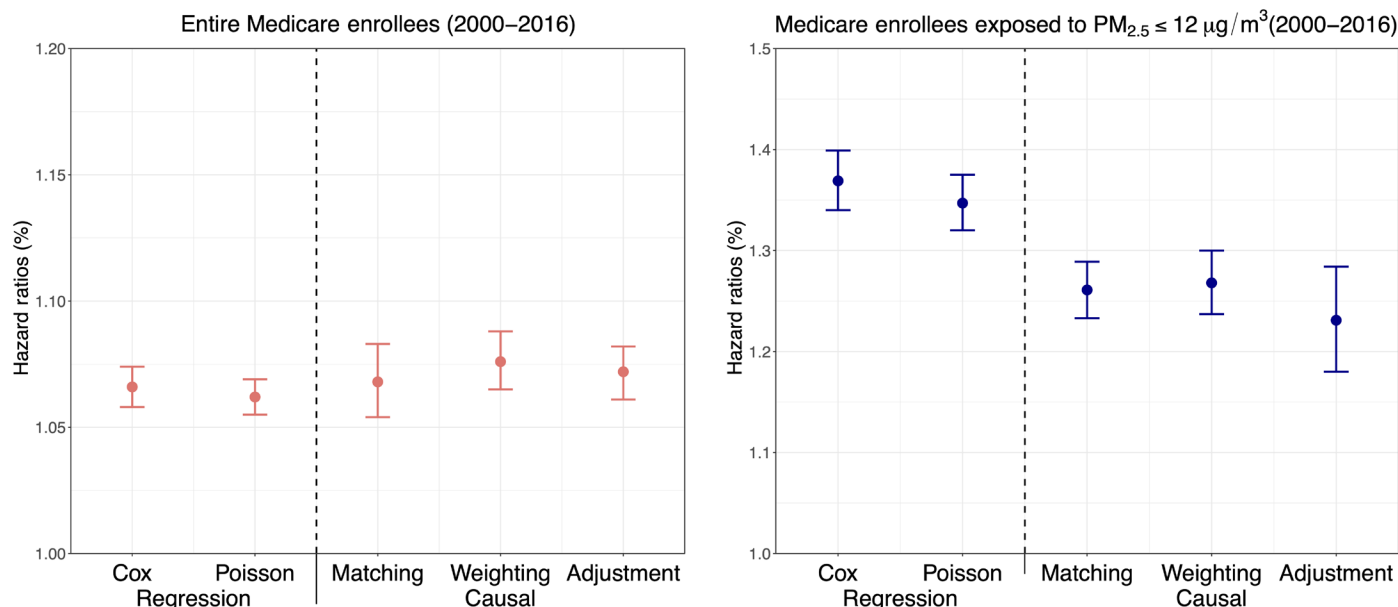


Fig. 3. HR and 95% CIs. The estimated HRs were obtained under five different statistical approaches (two traditional approaches and three causal inference approaches). HRs were adjusted by 10 potential confounders, four meteorological variables, geographic region, and year.

Both traditional and causal inference approaches rely on assumptions. Unless all assumptions are satisfied, regardless of approach, recovery of causal effects is not guaranteed. A critical assumption that guarantees our conclusion's validity is that our statistical analyses account for all confounders. This assumption must always be made in observational studies. We included several publicly available individual- and area-level potential confounders. To mitigate unmeasured confounding bias, we assessed the results' sensitivity by including year as a surrogate for some unmeasured confounders that might have covaried over time with PM_{2.5} and mortality and, thus, confound their association. Even after adjustment for year, the analysis could be affected by confounding bias by unmeasured factors; therefore, we conducted further sensitivity analyses to unmeasured confounding by calculating the E-value and showed that our results are robust to unmeasured confounding bias.

Dominici and Zigler (12) previously discussed three notions of what constitutes evidence of causality in air pollution epidemiology. The first is causality inferred from evidence of biological plausibility (13). The second is consistency of results across many epidemiological studies and adherence to Bradford Hill causal criteria (14). The third is the use of causal inference methods that are more robust to model misspecification compared to traditional approaches and, when assumptions are met, can isolate causal relationships. More specifically, the causal inference approaches considered in this work require the estimation of GPS as the first step. Assuming all causal inference assumptions hold, these approaches are more robust to outcome model misspecification and allow for the transparent evaluation of covariate balance. However, note that if the models are accurately specified and all assumptions are met, then the traditional approaches have the potential to inform causal relationship as well. In particular, we found that a more flexible regression model specification may help adequately adjust for confounding; when implementing these flexible models, we observed similar results compared to the causal inference approaches.

This work estimates the causal relationship using causal inference methods, addressing just one of Dominici and Zigler's three notions of what constitutes scientific evidence of causality. The collective evidence across studies conducted in different populations, using different study designs and methods, is also imperative to inform regulatory action. A recent meta-analysis found robust evidence for an effect on mortality across 52 cohort studies at PM_{2.5} levels below 10 µg/m³ (15).

Exposure to PM_{2.5} was estimated from a prediction model, which, while very good, is not perfect. The PM_{2.5} exposure prediction model developed by Di *et al.* (9) that was used in this analysis indicated excellent model performance, with a 10-fold cross-validated R^2 of 0.89 for annual PM_{2.5} predictions. However, exposure error could have affected all HR estimates. In the original study by Di *et al.* (3), the authors assessed the robustness of the results to the exposure predictions by repeating the analysis based on PM_{2.5} exposure data obtained from 1928 EPA ambient monitors. The additional analysis was restricted to the subpopulation of individuals within 50 km of these monitors. While this subset does not represent the entire population, we found that the analysis based on nearest monitoring site led to an HR estimate that was only slightly lower than the one obtained using the exposure prediction model [i.e., 1.061 (95% CI, 1.059 to 1.063) versus 1.073 (95% CI, 1.071 to 1.075)]. Although these results are reassuring, we recognize that they are not a substitute to a formal analysis that accounts for exposure error.

Accounting for exposure measurement error under a causal inference framework using propensity scores is complex, as the exposure error will affect not only the estimation of the health effects but also the estimation of the propensity score and its implementation to adjust for measured confounders (16). Regression calibration is a common method for measurement error correction (17). Wu *et al.* (18) proposed a regression calibration approach for GPS analysis under categorical exposures. The proposed approach was applied in the context of long-term PM_{2.5} exposure and mortality using the Medicare data in the Northeastern United States. When accounting for exposure error, there was a higher and still statistically significant association between exposure to PM_{2.5} and mortality, although with larger CIs. How to propagate exposure error under a causal inference framework for a continuous exposure is still an area of active research; the presence of exposure measurement error could induce a bias toward the null in all of our estimates (19).

The model parameterization assumes that zip code-specific information is spatially independent, given covariates. Since we adjusted for numerous zip code-level predictors of mortality, including SES and meteorological variables, this assumption is likely to hold. If any residual spatial dependence remains under certain assumptions (e.g., those used in generalized estimating equations), then it would not have affected our point estimates but could have influenced the estimated SEs. However, our bootstrapping procedure partially accounts for this possibility. By randomly sampling zip codes for each bootstrap replicate, we were able to break down spatial dependence given covariates. Therefore, it is unlikely that our results are affected by spatial correlation.

Our study is based on publicly available data sources and have made all code developed for our analyses publicly available. Our approach maximizes reproducibility and transparency. We provide robust evidence that the current U.S. standards for PM_{2.5} concentrations are not protective enough and should be lowered to ensure that vulnerable populations, such as the elderly, are safe.

Our results raise awareness of the continued importance of assessing the impact of air pollution exposure on mortality. There are currently numerous disputes regarding the evidence from previous air pollution epidemiologic studies, with arguments made for only using causal inference methods or only including studies that make participants' information publicly available. We oppose these very strongly. Most epidemiological studies must rely on confidential patient data to provide evidence on adverse health effects of environmental exposures on outcomes and also focus on populations that cannot be studied using administrative data. We hope this work will help researchers and policy makers, particularly as revision discussions of national PM_{2.5} standards are underway.

MATERIALS AND METHODS

Study population

Our study population is composed of more than 68.5 million Medicare enrollees (≥65 years of age) between 2000 and 2016. Medicare claims data, obtained from the Centers for Medicare and Medicaid Services (8), are an open cohort, including demographic information such as age, sex, race/ethnicity, date of death, and residential zip code. A unique patient ID is assigned to each person to allow for tracking over time. Medicare enrollees entered our cohort in 2000 if enrolled before 2000, or upon their enrollment after 2000. After enrollment, each individual was followed annually until the year of their death

or the end of our study period (31 December 2016). This study was conducted under a protocol approved by the Harvard T.H. Chan School of Public Health Human Subjects Committee.

Exposure assessment

We estimated daily $PM_{2.5}$ levels at a high spatiotemporal resolution using a 1-km² grid network across the contiguous United States and a well-validated ensemble-based prediction model (9). This model used ensemble learning approaches to combining three machine learning models: a random forest regression, a gradient boosting machine, and an artificial neural network (20–22). These machine learning algorithms used more than 100 predictor variables from satellite data, land-use information, weather variables, and output from chemical transport model simulations. The ensemble-based model was trained on daily $PM_{2.5}$ concentrations measured at 2156 U.S. EPA monitoring sites, with an average cross-validated R^2 of 0.86 for daily $PM_{2.5}$ predictions and 0.89 for annual predictions, indicating excellent performance that was improved compared to previously developed models (23, 24).

Residential addresses are not available for Medicare enrollees, only residential zip codes. For each standard zip code, we used zonal statistics to calculate the daily average $PM_{2.5}$ concentration based on all 1-km² grid cell predictions within the zip code via aggregations. More specifically, we first overlaid the zip code boundaries to the 1-km² grid cells and then averaged the predictions at 1-km² grid cells whose centroids fall within the boundary of that zip code (25). For P.O. Box–only zip codes, the average $PM_{2.5}$ concentrations were calculated by linking to the predictions from the nearest 1-km² grid cell. Annual zip code averages were estimated by averaging the daily concentrations. We assigned the annual estimated zip code average $PM_{2.5}$ concentration to individuals who resided in that zip code for each calendar year.

Potential confounders

The stratification by individual-level characteristics also adjusted for potential confounding by these variables. Furthermore, to adjust for confounding bias by community-level factors, we used information about multiple zip code–level SES variables collected from the U.S. Census, ACS, and BRFSS. All data [including census data, which are reported for Zip Code Tabulation Area (ZCTA)] were mapped to postal zip codes. Specifically, we included (i) two county-level variables: average body mass index and smoking rate; (ii) eight zip code–level census variables: proportion of Hispanic residents, proportion of Black residents, median household income, median home value, proportion of residents in poverty, proportion of residents with a high school diploma, population density, and proportion of residents that own their house; and (iii) four zip code–level meteorological variables: the summer (June to September) and winter (December to February) averages of maximum daily temperatures and relative humidity. We obtained zip code–level meteorological variables using area-weighted aggregations based on daily temperature and humidity data on 4-km² gridded rasters from Gridmet via Google Earth Engine (26, 27). We also considered two indicator variables indicating (i) the four census geographic regions of the United States (Northeast, South, Midwest, and West) and (ii) calendar years (2000–2016) to adjust for some residual or unmeasured spatial and temporal confounding, respectively. The data used for this study are publicly available and sources are listed in table S1.

Data linkage

Outcome data were available at the postal zip code level, at which we also assigned annual $PM_{2.5}$ exposures. Outcome and exposure information were available for 35,924 zip codes. We then mapped potential confounders at ZCTA to postal zip codes to link the outcome and exposure data to potential confounders obtained from the U.S. Census, ACS, and the BRFSS. The total number of zip codes included in our main analysis with information about all outcome, exposure, and confounder data was 31,337. See section S2 for more details.

Statistical analysis

For all models, we performed stratified outcome model analysis by four individual-level characteristics: (i) a 5-year category of age at entry (65 to 69, 70 to 74, 75 to 79, 80 to 84, 85 to 89, 90 to 94, 95 to 99, and above 100 years of age), (ii) race/ethnicity (White, Black, Asian, Hispanic, Native American, and other), (iii) sex (male or female), and (iv) an indicator variable for Medicaid eligibility (a surrogate for individual-level SES).

We fit five different statistical models to estimate the causal relationship between long-term $PM_{2.5}$ exposure and our outcome of interest, all-cause mortality among the elderly. Below, we describe the five approaches; details on the statistical methods and assumptions are provided in section S1.

Cox proportional hazard approach

We fit stratified Cox proportional hazards models using annual $PM_{2.5}$ as the time-varying exposure and stratifying by four individual-level characteristics. In our main analysis, we adjusted for 14 zip code– or county-level time-varying covariates, as well as a dummy region variable and dummy calendar year variable. The Cox proportional hazards survival model is specified as Survival(follow-up year, death) $\sim PM_{2.5}$ + area-level risk factors + meteorological variables + dummy year + dummy region + strata(age, race, gender, Medicaid eligibility).

Poisson regression approach

We fit the Poisson regression model, using annual $PM_{2.5}$ as the time-varying exposure; the count of deaths at the given follow-up year, calendar year, and zip code as the outcome and the corresponding total person-time as the offset term. To adjust for potential confounding, we included the same 14 zip code– or county-level time-varying covariates, as well as a dummy region variable and dummy calendar year variable, as those included in the Cox proportional hazards models. We used a stratified Poisson regression model formulation to account for the strata-specific baseline risk rates by stratifying on individual-level characteristics. The Poisson regression model is specified as $\log(E[\text{death counts}]) \sim PM_{2.5}$ + area-level risk factors + meteorological variables + dummy year + dummy region + strata(age, race, gender, Medicaid eligibility, follow-up year) + offset(log[person year]).

Causal inference approaches

GPS estimation. The three proposed causal inference approaches required the estimation of GPS as the first step. In our study, we modeled the conditional density of exposure (i.e., annual average $PM_{2.5}$) on the 14 zip code– or county-level time-varying covariates, as well as a dummy region variable and dummy calendar year variable, using gradient boosting machine with normal residuals (28, 29). The gradient boosting machine model is specified as $PM_{2.5} \sim \text{area-level risk factors} + \text{meteorological variables} + \text{dummy year} + \text{dummy region} + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$.

GPS matching approach. We constructed the matched pseudo-population as described in section S1. We first checked the covariate balance in the matched pseudo-population, and if covariate balance was achieved (average AC, <0.1), then we fit a univariate Poisson regression model regressing the death counts with an offset person-time term, on the exposure $PM_{2.5}$, stratifying by four individual-level characteristics and the same follow-up year. The Poisson regression model is specified as $\log(E[\text{death counts}]) \sim PM_{2.5} + \text{strata}(\text{age, race, gender, Medicaid eligibility, follow-up year}) + \text{offset}(\log[\text{person year}])$, on the matched pseudo-population. Additional details on the matching procedure can be found in section S1.

GPS weighting approach. We constructed the weighted pseudo-population as described in section S1. We first checked the covariate balance on the weighted pseudo-population, and if covariate balance was achieved (average AC, <0.1), then we fit a weighted univariate Poisson regression model regressing the death count with offset term the person-time on $PM_{2.5}$ exposure incorporating the assigned weights and stratifying by the four individual-level characteristics and the same follow-up year. The Poisson regression model is specified as $\log(E[\text{death counts}]) \sim PM_{2.5} + \text{strata}(\text{age, race, gender, Medicaid eligibility, follow-up year}) + \text{offset}(\log[\text{person year}])$, $\text{weights} = f(PM_{2.5})/GPS$, where $f(PM_{2.5})$ is the marginal density function of exposure $PM_{2.5}$, which serves as a stabilizing term (30).

GPS adjustment approach. We modeled the conditional expectation of the death counts given the exposure and the estimated GPS as a stratified Poisson regression with flexible formulation of bivariate variables, with the corresponding person-time offset. The Poisson regression model was specified as $\log(E[\text{death counts}]) \sim PM_{2.5} + PM_{2.5} \times GPS + GPS + GPS^2 + \text{strata}(\text{age, race, gender, Medicaid eligibility, follow-up year}) + \text{offset}(\log[\text{person year}])$. In contrast to the GPS matching/weighting approaches, where the analysis is complete after fitting the Poisson regression model, for the GPS adjustment approach, the coefficients from the Poisson regression model do not provide any causal interpretation; instead, the causal outcome analysis is conducted on the counterfactuals predicted by the Poisson model (31). We fit a univariate linear regression model regressing the counterfactual mean hazard rates for each $PM_{2.5}$ level, stratifying by four individual-level characteristics and the same follow-up year. The outcome linear regression model is specified as $E(\text{hazard rates}) \sim PM_{2.5} + \text{strata}(\text{age, race, gender, Medicaid eligibility, follow-up year})$. Additional details are provided in section S1.

Total events avoided

We estimated the total number of deaths that would be avoided among the elderly per decade if all areas were in compliance with the current WHO guidelines (annual $PM_{2.5}$ exposure, $\leq 10 \mu\text{g}/\text{m}^3$) (32). Nethery *et al.* (33) defined and identified a causal quantity named the total events avoided (TEA) under causal assumptions 1 to 3 (see section S1), defined as the difference in the expected number of health events under the counter-factual pollution exposures and the observed number of health events under the factual pollution exposures. Such a causal quantity is particularly related to the health policy that intends to answer the question “How many deaths were avoided in the Medicare population per decade due to the U.S. National Ambient Air Quality Standards (NAAQS) changes in particulate matter ($PM_{2.5}$) in the same time?”

We created the counterfactual $PM_{2.5}$ exposures if all zip codes in the continental United States complied with the current WHO guidelines (annual $PM_{2.5}$, $\leq 10 \mu\text{g}/\text{m}^3$). For zip codes that did not comply with the standard until 2016, their counterfactual was assumed to

be exposure exactly at this hypothesized standard ($10 \mu\text{g}/\text{m}^3$). This is a conservative estimate, as it answers the question of TEA if these zip codes were exactly at $10 \mu\text{g}/\text{m}^3$ and not lower than this concentration. For zip codes already in compliance, we assumed that their concentration was unchanged, which otherwise would result in even higher TEA.

We compared this counterfactual scenario to the factual scenarios during the most recent decade (2007–2016). For zip codes with an annual $PM_{2.5}$ concentration of $> 12 \mu\text{g}/\text{m}^3$, the numbers for the TEA were obtained using the most conservative HR from our main analysis [HR, 1.06 (95% CI, 1.05 to 1.08); see table S3]. For zip codes with an annual $PM_{2.5}$ concentration of 10 to $12 \mu\text{g}/\text{m}^3$, the numbers for the TEA were obtained using the most conservative HR from our low-level analysis [HR, 1.23 (95% CI, 1.18 to 1.28)]. Zip codes with an annual $PM_{2.5}$ concentration of $< 10 \mu\text{g}/\text{m}^3$ did not contribute to the TEA. For the CI calculation, we used the lower and upper bounds of the 95% CIs from the HR estimates (which were obtained by bootstrap).

Evaluation of unmeasured confounding

We conducted a sensitivity analysis to evaluate the robustness of our results to unmeasured confounding by calculating the E-value. The E-value for the point estimate of interest (in our case, the HR) can be defined as the minimal strength of an association, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and outcome, conditional on the covariates already included in the model, to fully explain the observed association under the null. We calculated the E-value for our reported HRs per increase ($10 \mu\text{g}/\text{m}^3$) of long-term exposure to $PM_{2.5}$. The calculation of E-value can be implemented through the E-value calculator by Mathur *et al.* (11), available at <https://www.evalue-calculator.com/>.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/29/eaba5692/DC1>

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. 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 and mortality in six U.S. cities. *N. Engl. J. Med.* **329**, 1753–1759 (1993).
2. R. Beelen, O. Raaschou-Nielsen, M. Stafoggia, Z. J. Andersen, G. Weinmayr, B. Hoffmann, K. Wolf, E. Samoli, P. Fischer, M. Nieuwenhuijsen, P. Vineis, W. W. Xu, K. Katsouyanni, K. Dimakopoulou, A. Oudin, B. Forsberg, L. Modig, A. S. Havulinna, T. Lanki, A. Turunen, B. Oftedal, W. Nystad, P. Nafstad, U. De Faire, N. L. Pedersen, C.-G. Östenson, L. Fratiglioni, J. Penell, M. Korek, G. Pershagen, K. T. Eriksen, K. Overvad, T. Ellermann, M. Eeftens, P. H. Peeters, K. Meliefste, M. Wang, B. Bueno-de-Mesquita, D. Sugiri, U. Kramer, J. Heinrich, K. de Hoogh, T. Key, A. Peters, R. Hampel, H. Concin, G. Nagel, A. Ineichen, E. Schaffner, N. Probst-Hensch, N. Künzli, C. Schindler, T. Schikowski, M. Adam, H. Phuleria, A. Villier, F. Clavel-Chapelon, C. Declercq, S. Grioni, V. Krogh, M.-Y. Tsai, P. Ricceri, C. Sacerdote, C. Galassi, E. Migliore, A. Ranzi, G. Cesaroni, C. Badaloni, F. Forastiere, I. Tamayo, P. Amiano, M. Dorronsoro, A. Katsoulis, A. Trichopoulos, B. Brunekreef, G. Hoek, Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* **383**, 785–795 (2014).
3. 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).
4. C. Liu, R. Chen, F. Sera, A. M. Vicedo-Cabrera, Y. Guo, S. Tong, M. S. Z. S. Coelho, P. H. N. Saldiva, E. Lavigne, P. Matus, N. Valdes Ortega, S. Osorio Garcia, M. Pascal, M. Stafoggia, M. Scortichini, M. Hashizume, Y. Honda, M. Hurtado-Díaz, J. Cruz, B. Nunes, J. P. Teixeira, H. Kim, A. Tobias, C. Iñiguez, B. Forsberg, C. Åström, M. S. Ragettli, Y.-L. Guo, B.-Y. Chen, M. L. Bell, C. Y. Wright, N. Scovronick, R. M. Garland, A. Milojevic, J. Kysely, A. Urban, H. Orru, E. Indermitte, J. J. K. Jaakkola, N. R. I. Rytty, K. Katsouyanni, A. Analitis, A. Zanobetti, J. Schwartz, J. Chen, T. Wu, A. Cohen, A. Gasparini, H. Kan, Ambient

- particulate air pollution and daily mortality in 652 cities. *N. Engl. J. Med.* **381**, 705–715 (2019).
5. A. J. Pappin, T. Christidis, L. L. Pinault, D. L. Crouse, J. R. Brook, A. Erickson, P. Hystad, C. Li, R. V. Martin, J. Meng, S. Weichenthal, A. van Donkelaar, M. Tjepkema, M. Brauer, R. T. Burnett, Examining the shape of the association between low levels of fine particulate matter and mortality across three cycles of the Canadian Census Health and Environment Cohort. *Environ. Health Perspect.* **127**, 107008 (2019).
 6. Environmental Protection Agency (EPA), Chartered Clean Air Scientific Advisory Committee (CASAC) public teleconference on particulate matter (PM) (2019); <https://yosemite.epa.gov/sab/sabproduct.nsf/MeetingCalCASAC/4F40665AD1DDCEF6852583A000645464?OpenDocument> [accessed 13 April 2020].
 7. G. T. Goldman, F. Dominici, Don't abandon evidence and process on air pollution policy. *Science* **363**, 1398–1400 (2019).
 8. Centers for Medicare and Medicaid Services (CMS); www.cms.gov [accessed 1 November 2017].
 9. 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).
 10. T. J. VanderWeele, P. Ding, Sensitivity analysis in observational research: Introducing the E-value. *Ann. Intern. Med.* **167**, 268–274 (2017).
 11. 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).
 12. F. Dominici, C. Zigler, Best practices for gauging evidence of causality in air pollution epidemiology. *Am. J. Epidemiol.* **186**, 1303–1309 (2017).
 13. L. W. Stanek, J. S. Brown, J. Stanek, J. Gift, D. L. Costa, Air pollution toxicology—A brief review of the role of the science in shaping the current understanding of air pollution health risks. *Toxicol. Sci.* **120** Suppl 1, S8–S27 (2011).
 14. A. B. Hill, The environment and disease: Association or causation? *Proc. R. Soc. Med.* **58**, 295–300 (1965).
 15. A. Vodonos, Y. A. Awad, J. Schwartz, The concentration-response between long-term PM_{2.5} exposure and mortality: a meta-regression approach. *Environ. Res.* **166**, 677–689 (2018).
 16. D. Braun, M. Gorfine, G. Parmigiani, N. D. Arvola, F. Dominici, C. Zigler, Propensity scores with misclassified treatment assignment: A likelihood-based adjustment. *Biostatistics* **18**, 695–710 (2017).
 17. R. J. Carroll, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective* (CRC Press, ed. 2, 2006).
 18. X. Wu, D. Braun, M.-A. Kioumourtoglou, C. Choirat, Q. Di, F. Dominici, Causal inference in the context of an error prone exposure: Air pollution and mortality. *Ann. Appl. Stat.* **13**, 520–547 (2019).
 19. M.-A. Kioumourtoglou, D. Spiegelman, A. A. Szpiro, L. Sheppard, J. D. Kaufman, J. D. Yanosky, R. Williams, F. Laden, B. Hong, H. Suh, Exposure measurement error in PM_{2.5} health effects studies: A pooled analysis of eight personal exposure validation studies. *Environ. Health* **13**, 2 (2014).
 20. T. G. Dietterich, Ensemble learning, in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. (MIT Press, ed. 2, 2002), pp. 405–408.
 21. C. Zhang, Y. Ma, Eds., *Ensemble Machine Learning: Methods and Applications* (Springer Science & Business Media, 2012).
 22. R. Polikar, Ensemble learning, in *Ensemble Machine Learning* (Springer, 2012), pp. 1–34.
 23. 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).
 24. Q. Di, P. Koutrakis, J. Schwartz, A hybrid prediction model for PM_{2.5} mass and components using a chemical transport model and land use regression. *Atmospheric Environ.* **131**, 390–399 (2016).
 25. M. Zeiler, *Modeling Our World: The ESRI Guide to Geodatabase Design* (ESRI Inc., 1999).
 26. J. T. Abatzoglou, Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* **33**, 121–131 (2013).
 27. N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, R. Moore, Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
 28. T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), pp. 785–794; <https://doi.org/10.1145/2939672.2939785>.
 29. Y. Zhu, D. L. Coffman, D. Ghosh, A boosting algorithm for estimating generalized propensity scores with continuous treatments. *J. Causal Inference* **3**, 25–40 (2015).
 30. J. M. Robins, M. A. Hernán, B. Brumback, Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560 (2000).
 31. K. Hirano, G. W. Imbens, The propensity score with continuous treatments, in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, A. Gelman, X.-L. Meng, Eds. (John Wiley & Sons Ltd., 2004), pp. 73–84.
 32. World Health Organization, WHO air quality standards (WHO, 2005); [www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](http://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health).
 33. R. C. Nethery, F. Mealli, J. D. Sacks, F. Dominici, Causal inference and machine learning approaches for evaluation of the health impacts of large-scale air quality regulations. [arXiv:1909.09611](https://arxiv.org/abs/1909.09611) [stat.AP] (15 September 2019).
 34. P. K. Andersen, R. D. Gill, Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–1120 (1982).
 35. T. R. Holford, The analysis of rates and of survivorship using log-linear models. *Biometrics* **36**, 299–305 (1980).
 36. N. Laird, D. Olivier, Covariance analysis of censored survival data using log-linear analysis techniques. *J. Am. Statist. Assoc.* **76**, 231–240 (1981).
 37. R. T. Burnett, A. Dewanji, F. Dominici, M. S. Goldberg, A. Cohen, D. Krewski, On the relationship between time-series studies, dynamic population studies, and estimating loss of life due to short-term exposure to environmental risks. *Environ. Health Perspect.* **111**, 1170–1174 (2003).
 38. X. Wu, F. Mealli, M.-A. Kioumourtoglou, F. Dominici, D. Braun, Matching on generalized propensity scores with continuous exposures. [arXiv:1812.06575](https://arxiv.org/abs/1812.06575) [stat.ME] (17 December 2018).
 39. J. Splawa-Neyman, D. M. Dabrowska, T. P. Speed, On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5**, 465–472 (1990).
 40. D. B. Rubin, Causal inference using potential outcomes. *J. Am. Statist. Assoc.* **100**, 322–331 (2005).
 41. A. P. Dawid, Causal inference without counterfactuals. *J. Am. Statist. Assoc.* **95**, 407–424 (2000).
 42. J. M. Robins, S. Greenland, Comment: Causal inference without counterfactuals. *J. Am. Statist. Assoc.* **95**, 431–435 (2000).
 43. D. B. Rubin, Comment: Causal inference without counterfactuals. *J. Am. Statist. Assoc.* **95**, 435–438 (2000).
 44. P. R. Rosenbaum, D. B. Rubin, The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
 45. R. J. Desai, J. M. Franklin, Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: A primer for practitioners. *BMJ* **367**, l5657 (2019).
 46. V. S. Harder, E. A. Stuart, J. C. Anthony, Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychol. Methods* **15**, 234–249 (2010).
 47. G. W. Imbens, D. B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction* (Cambridge Univ. Press, 2015).
 48. M.-A. Bind, Causal modeling in environmental health. *Annu. Rev. Public Health* **40**, 23–43 (2019).
 49. E. H. Kennedy, Z. Ma, M. D. McHugh, D. S. Small, Nonparametric methods for doubly robust estimation of continuous treatment effects. *J. R. Stat. Soc. Series B Stat. Methodol.* **79**, 1229–1245 (2017).
 50. D. R. Cox, *Planning of Experiments* (John Wiley & Sons Inc., 1958).
 51. R. C. Nethery, Y. Yang, A. J. Brown, F. Dominici, A causal inference framework for cancer cluster investigations using publicly available data. [arXiv:1811.05997](https://arxiv.org/abs/1811.05997) [stat.ME] (14 November 2018).
 52. D. B. Rubin, For objective causal inference, design trumps analysis. *Ann. Appl. Statist.* **2**, 808–840 (2008).
 53. E. A. Stuart, D. B. Rubin, Best practices in quasi-experimental designs: Matching methods for causal inference, in *Best Practices in Quantitative Methods*, J. Osborne, Ed. (Sage, 2008), pp. 155–176.
 54. P. R. Rosenbaum, Modern algorithms for matching in observational studies. *Ann. Rev. Statist. Application* **7**, 143–176 (2020).
 55. R. K. Crump, V. J. Hotz, G. W. Imbens, O. A. Mitnik, Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96**, 187–199 (2009).
 56. B. K. Lee, J. Lessler, E. A. Stuart, Weight trimming and propensity score weighting. *PLOS ONE* **6**, e18174 (2011).
 57. A. Agresti, M. Kateri, Categorical data analysis, in *International Encyclopedia of Statistical Science*, M. Lovric, Ed. (Springer Berlin Heidelberg, 2011), pp. 206–208.
 58. P. C. Austin, Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat. Methods Med. Res.* **28**, 1365–1377 (2019).
 59. M. A. Hernán, B. Brumback, J. M. Robins, Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11**, 561–570 (2000).
 60. H. Friedman, Magnitude of experimental effect and a table for its rapid estimation. *Psychol. Bull.* **70**, 245–251 (1968).
 61. M. B. Mathur, T. J. VanderWeele, A simple, interpretable conversion from Pearson's correlation to Cohen's for d continuous exposures. *Epidemiology* **31**, e16–e18 (2020).
 62. P. C. Austin, Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat. Med.* **28**, 3083–3107 (2009).

63. Q. Di, L. Dai, Y. Wang, A. Zanobetti, C. Choirat, J. D. Schwartz, F. Dominici, Association of short-term exposure to air pollution with mortality in older adults. *JAMA* **318**, 2446–2456 (2017).
64. J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge Univ. Press, 2000).
65. M. A. Hernán, J. M. Robins, *Causal Inference: What If* (Chapman & Hall/CRC, 2020).

Acknowledgments: We would like to thank L. Goodwin for editorial assistance in the preparation of this manuscript. The computations in this paper were run on the Odyssey cluster, supported by the FAS Division of Science, Research Computing Group at Harvard University, and the Research Computing Environment, supported by the Institute for Quantitative Social Science in the Faculty of Arts and Sciences at Harvard University. **Funding:** This work was made possible by the support from NIH grants R01 ES024332-01A1, P50 MD010428, R21 ES024012, R01 ES026217, R01 ES028033, MD012769, R01 ES030616, and P30 ES09089; HEI grant 4953-RFA14-3/16-4; and USEPA grants 83587201-0 and RD-83479801. The contents are solely the responsibility of the grantee and do not necessarily represent the official views of the funding agencies. Further, the funding agencies do not endorse the purchase of any commercial products or services related to this publication. **Author**

contributions: X.W.: study design, analysis, and writing. D.B.: study design and writing. J.S.: data preparation. M.A.K.: study design and writing. F.D.: study design and writing. **Competing interests:** J.S. served as an expert witness for the U.S. Department of Justice in a Clean Air Act violation case. The authors declare that they have no other 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. Those interested in the original data and code can contact the corresponding author. Additional data related to this paper is available on https://github.com/wxwx1993/National_Causal.

Submitted 16 December 2019

Accepted 4 June 2020

Published 17 July 2020

10.1126/sciadv.aba5692

Citation: 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).

Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly

X. Wu, D. Braun, J. Schwartz, M. A. Kioumourtoglou and F. Dominici

Sci Adv 6 (29), eaba5692.

DOI: 10.1126/sciadv.aba5692 originally published online June 26, 2020

ARTICLE TOOLS

<http://advances.sciencemag.org/content/6/29/eaba5692>

SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2020/06/26/sciadv.aba5692.DC1>

REFERENCES

This article cites 46 articles, 2 of which you can access for free
<http://advances.sciencemag.org/content/6/29/eaba5692#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

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.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).