A National Difference in Differences Analysis of the Effect of $PM_{2.5}$ on Annual Death Rates

Joel Schwartz, Yaguang Wei, Ma'ayan Yitshak-Sade, Qian Di, Francesca Dominici, Antonella Zanobetti

PII: S0013-9351(20)31546-2

DOI: https://doi.org/10.1016/j.envres.2020.110649

Reference: YENRS 110649

To appear in: *Environmental Research*

Received Date: 10 November 2020

Revised Date: 16 December 2020

Accepted Date: 17 December 2020

Please cite this article as: Schwartz, J., Wei, Y., Yitshak-Sade, M.'a., Di, Q., Dominici, F., Zanobetti, A., A National Difference in Differences Analysis of the Effect of $PM_{2.5}$ on Annual Death Rates, *Environmental Research*, https://doi.org/10.1016/j.envres.2020.110649.

A National Difference in Differences Analysis of the Effect of $PM_{2.5}$ on Annual Death Rates

Joel Schwartz[1,2],Yaguang Wei[1], Ma'ayan Yitshak-Sade[1,3], Qian Di[1,4], Francesca Dominici[5], Antonella Zanobetti[1]

1. Department of Environmental Health, Harvard TH Chan School of Public Health
2. Department of Epidemiology, Harvard TH Chan School of Public Health
3. Department of Environmental Medicine, Mount Sinai School of Medicine
4. Vanke School of Public Health, Tsinghua University, Beijing, China
5. Department of Biostatistics, Harvard TH Chan School of Public Health

Address communications to:

Joel Schwartz joel@hsph.harvard.edu. 401 Park Drive, Suite 404H, Boston MA 02215.

Abstract

Many studies have reported that $PM_{2.5}$ was associated with mortality, but these were criticized for unmeasured confounding, not using causal modeling, and not focusing on changes in exposure and mortality rates. Recent studies have used propensity scores, a causal modeling approach that requires the assumption of no unmeasured confounders.

We used differences in differences, a causal modeling approach that focuses on exposure changes, and controls for unmeasured confounders by design to analyze $PM_{2.5}$ and mortality in the U.S. Medicare population, with 623,036,820 person-years of follow-up, and 29,481,444 deaths. We expanded the approach by clustering ZIP codes into 32 groups based on racial, behavioral and socioeconomic characteristics, and analyzing each cluster separately. We controlled for multiple time varying confounders within each cluster. A separate analysis examined participants whose exposure was always below 12 $\mu g/m^3$.We found an increase of 1 $\mu g/m^3$ in $PM_{2.5}$ produced an increased risk of dying in that year of 3.85×10^{-4} (95% CI 1.95 $\times 10^{-4}$, 5.76 $\times 10^{-4}$). This corresponds to 14,000 early deaths per year per 1 $\mu g/m^3$. When restricted to exposures below 12 $\mu g/m^3$, the increased mortality risk was 4.26×10^{-4} (95% CI 1.43×10^{-4}, 7.09×10^{-4}). Using a causal modeling approach robust to omitted confounders, we found associations of $PM_{2.5}$ with increased death rates, including below U.S. and E.U. standards.

Key words: Air pollution, $PM_{2.5}$, causal, difference in differences, mortality

Authors contributions: JS conceived the approach, developed the methodology, and did the analysis and reviewed and edited the draft manuscript; YW wrote the first draft and prepared the data; MY-W helped with the programming, methodology and commented on the draft; FD obtained funding and the Medicare data; AZ obtained funding.

Funding: This study was funded by NIH grants ES000002 and ES024332, and U.S. EPA grant RD-83615601. The funders had no input into the design or conduct of the study.

Dr. Schwartz declares that he has testified as a health expert for the U.S. Department of Justice in a law suit over clean air act violations. No other authors have anything to declare.

This study was approved by the Human Subjects Committee of the Harvard School of Public Health.

1. Introduction

The Clean Air Act requires the US Environmental Protection Agency (EPA) to set National Ambient Air Quality Standards (NAAQS) to protect vulnerable populations with an adequate margin of safety. Many studies have reported associations of $PM_{2.5}$ and mortality and morbidity following long and short-term exposure(Abu Awad et al., 2019; Beelen et al., 2014; Crouse et al., 2015; Di et al., 2017; Hoek et al., 2013; Pinault et al., 2016; Pope et al., 2019; Vodonos et al., 2018). These were undertaken by many investigators with over 50 cohorts in the most recent $PM_{2.5}$ meta-analysis(Vodonos et al., 2018), and have resulted in EPA sequentially tightening the $PM_{2.5}$ standard. The global burden of disease ranks air pollution among the largest public health risks.

Recent studies have reported associations between $PM_{2.5}$ and mortality at concentrations below the 2012 U.S. EPA NAAQS or World Health Organization air quality guidelines(Di et al., 2017; Wang et al., 2016; Yang et al., 2012). However, some have criticized many of these studies for not using causal modeling approaches.

Causal modeling methods can aid in assessing causality. The general approach is to try to make an observational study closely mimic a randomized trial. In addition, causal methods provide marginal estimates of the effects of exposure, that do not depend on the distribution of the covariates in the study population(Imai and van Dyke, 2004). A common approach is to use propensity score matching or inverse probability weighting to make the exposure independent of all measured confounders(Baccini et al., 2017; Rubin, 1997). Recent studies have used that approach to examine the association of $PM_{2.5}$ with mortality, and provided robust findings(Abu Awad et al., 2019; Schwartz et al., 2018; Wei et al., 2020; Wu et al., 2020; Yitshak-Sade et al., 2019). However, propensity scores only control for measured confounders, and therefore do not address the

argument that there is unmeasured confounding. Hence it is important to complement that approach using methods that can address unmeasured confounders.

Approaches that control for unmeasured confounders by design include difference-in-differences (DID) analyses (Wang et al., 2016; Yitshak-Sade et al., 2019). In a classical DID model, the mean response is calculated for the exposed and non-exposed groups in pre-exposure and post-exposure periods. Since all slowly varying predictors of outcome such as socioeconomic status (SES), smoking, obesity, etc. are the same in each group in both periods, the difference between outcomes in the two periods in the exposed group cannot be confounded by those variables. The difference between pre-exposure and post-exposure periods in the unexposed group is a negative outcome control for the difference in the exposed group. It controls for changes in an outcome due to covariates that can change between periods similarly between the two locations. The difference in these pairs of differences is a causal estimate, assuming that no other exposure has affected the two groups differently over time(Donald and Lang, 2007). Because the DID approach examines the effect of changes in exposure (post vs pre periods) on change in outcome, it is precisely the type of study that EPA's CASAC says it prefers. The method has been generalized to look at more than two locations, more than two time periods, and continuous, time varying exposures(Wang et al., 2016). With multiple locations time-invariant omitted confounders are controlled using an indicator variable for each location. However, it still require the assumption that changes in mortality rates by year due to changing risk factors are common across locations. Here we simultaneously adopt two approaches to relax that assumption, and hence strengthen the evidence for causality. We apply them to assess whether changes in $PM_{2.5}$ are associated with changes in mortality rates in a national cohort of Medicare participants in the U.S. In addition, as

few previous cohort studies have controlled for temperature, we adjusted for mean warm season and mean cold season temperature.

2.0 Data and Methods

2.1 Medicare cohort

We obtained the Medicare beneficiary denominator file, which contains information on all Medicare participants in the U.S., from the Centers for Medicare and Medicaid Services(RESDAC, 2018). We followed all beneficiaries' ≥65 years in the contiguous U.S. from 2000 to 2016. Medicare insurance covers over 95% of the population ≥65 years of age in the United States. Medicare participants alive on January 1 of the year following their enrollment in Medicare entered the open cohort, and follow-up periods were calendar years. For the DID analysis, we computed an annual mortality rate in each ZIP code, in each group stratified by age (>84 or not), sex, race, and Medicaid coverage. This study was approved by the human subjects committee at the Harvard T. H. Chan School of Public Health.

2.2 Covariates

From the Medicare denominator file for each calendar year, we obtained the age, sex, race, ZIP code of residence for that year, eligibility for Medicaid for that year, and date of death (or censoring) of each participant. Age, ZIP code, and Medicaid eligibility were updated annually. This file is publicly available from the Centers for Medicare and Medicaid Services(RESDAC, 2018).

We obtained small area–level social, economic, and housing characteristic variables from the U.S. Census Bureau 2000 and 2010 Census Summary File 3(Bureau, 2010) at the ZIP code tabulation–area level (ZCTA) and the American Household Survey for each year after 2010. These included percent of people ≥ 65 living in poverty, median household income, median house value, percent

of owner occupied homes, percent black, percent Hispanic, population density, and education. We updated these variables for missing years by linearly extrapolating between the measured years. In addition, the county-level percentage of people who ever smoked and their mean body mass index (BMI) were obtained from the CDC Behavioral Risk Factor Surveillance survey(CDC, 2013) , which were then assigned to each ZCTA within the county and updated each year. From the Dartmouth Health Atlas, we obtained percentage of Medicare participants who had a hemoglobin A1c test, a low-density lipoprotein cholesterol (LDLC) test, a mammogram, an eye exam, and a visit to an annual checkup for each year in each hospital catchment area and assigned it to all ZCTAs in that area(Wennberg and Cooper, 1996). We also computed the distance from each ZIP code centroid to the nearest hospital. To capture long-term smoking history of Medicare participants in each ZIP code, we computed their hospitalization rate for lung cancer by ZIP code for each year. This risks over-control because air pollution has been associated with increased risk of lung cancer. To capture year-to-year changes in mortality rates due to temperature, we downloaded daily temperature data on a 12km grid from the NASA NLDAS-2 website (https://ldas.gsfc.nasa.gov/index.php/nldas/v2/models). We averaged all grid cells within the boundaries of a ZIP code, and constructed two measures for each year, the average temperature in the warm months (April-September) and in the cold months (October-March).

2.3 Exposure assessment

We estimated exposure using a validated prediction model calibrated to measurements at almost 2000 monitoring stations using an ensemble of machine learners that provided daily estimates for a 1km grid of the contiguous U.S.(Di et al., 2019; Di et al., 2020). In brief, the model used data from multiple sources including predictions of chemical transport models (GEOS-Chem, CMAQ, and MERRA-2), meteorological data, land-use terms, and satellite-based measures of aerosol

optical depth, surface reflectance, and absorbing aerosol index. We trained a neural network, a random forest, and a gradient boosting machine to monitoring data from the United States Environmental Protection Agency (EPA) Air Quality System to generate daily predictions on a 1×1 km grid. The models were fit using data from all years. The three predictions for $PM_{2.5}$ were combined in a nonlinear geographically weighted regression. The model showed good performance with ten-fold cross validation on held out monitoring sites yielding an out of sample R^2 of 0.89 for annual average predictions of $PM_{2.5}$. Penalized splines showed linear relationships between observed and predicted $PM_{2.5}$ from 0 to 60 μg/m^3. Predictions for all grid cells whose centroids were inside the ZIP code boundary were averaged for each year and assigned to participants in that ZIP code in that year.

2.4 Statistical analysis

The standard DID estimator for a continuous predictor posits that

$$E(Y_{ij}) = \beta_0 + \beta_1 PM_{2.5} + \beta_2 C_i + \beta_3 X_t \quad (1)$$

where Y_{ij} is the mortality rate in ZIP code i in demographic group (by age >84 or not, sex, race, and Medicaid coverage) j, C_i are the time-invariant or slowly changing confounders in ZIP code i, X_t are the time varying confounders that are common across ZIP codes. The C_i are controlled by fitting individual intercepts for each ZIP code. The time varying confounders are removed by fitting a nonlinear time trend; we used a natural spline function of year with 3 degrees of freedom, yielding:

$$E(Y_{ij}) = \beta_0 + \beta_1 PM_{2.5} + \delta_i + ns(year, 3) \quad (2)$$

where δ_i is a dummy variable for each ZIP Code. Since ZIP code is controlled, this model compares year-to-year variations around ZIP code average $PM_{2.5}$ and common time trend to year-

to-year variations of mortality rates about ZIP code average and common time trend. Differences in e.g. SES, smoking, or diabetes between ZIP code are removed by the dummy variable for ZIP code. For a causal interpretation of the DID estimate to hold, we must assume that all the ZIP codes have parallel long-term time trends in mortality rates, other than those caused by different time trends in $PM_{2.5}$.

If covariates producing different time trends in mortality rate by ZIP code are not correlated with ZIP code specific $PM_{2.5}$ trends, the interpretation still holds. It would be preferable to further weaken this assumption.

We added two methods to relax the parallel trends assumption, and combined them in our analysis. We added to equation (2) terms for confounders that we have measured that change over time, possibly differentially by ZIP code. This will control for any temporal trends due to changes in these covariates, which include the SES, race, demographic, behavioral, and health access variables described above. Second, we grouped ZIP codes based on the above covariates and fit separate time trends in each group. We think that ZIP codes that are similar in racial composition, percent living below the poverty level, population density, smoking rates etc. are more likely to have similar time trends in mortality rates than disparate ZIP codes. To accomplish this, we fit a principal component analysis to all the listed potential confounders and took the first 5 principal components. We classified each ZIP code into whether it was higher or lower than average on each of the 5 components, producing 32 categories of ZIP codes. In each of these 32 categories of ZIP codes, we fit separate splines for time trend and separate control for all of the covariates. This controls for time trends in measured covariates such as racial composition, median income, etc, and fits 32 separate time trends to the data to capture any trends unexplained by time trends in the measured covariates. It also allows the effects of the measured covariates to differ by the 32

different groups. In addition by performing analyses stratified by the 32 different groups we are also controlling by matching for the covariate clusters (e.g. SES and race) that characterize each group. Combining these, the final modeling approach is to fit 32 models (k=1:32) and meta-analyze the 32 values of β_1.

$$E(Y_{ij}, k) = \beta_0 + \beta_{1k}PM25 + \delta_i + \delta_{jk} + ns(year, 3, k) + \beta_3 X_{tk} \quad (3)$$

Here δ_i are indicator variables for each ZIP code, X_{tk} are the time varying covariates whose time trends may differ by ZIP code and by which group (k) the ZIP code is in, $ns(year, 3, k)$ is a natural spline for time trend with 3 degrees of freedom for each group k, and δ_{jk} is an indicator variable for each age-race-sex-Medicaid stratum in group k. Results were combined over strata using a random effects meta-analysis.

Finally, equation (3) embodies an additive, rather than multiplicative model for the rate of mortality in each ZIP code-demographic group. This allows us to estimate the additive effect of $PM_{2.5}$ on the probability of dying, provides more interpretative interaction terms, and provides a marginal effect estimate (i.e. not dependent on the distribution of the covariates, as a multiplicative model would be). Additive probability or rate models give unbiased estimates of effect just as the more usual logistic models, but biased estimates of standard errors(Caudill and Jackson, 1989). Therefore have used robust standard errors to estimate the confidence intervals. In a second analysis, we reran the analysis on data restricted to persons whose exposure was always below 12 $\mu g/m^3$, the U.S. standard for $PM_{2.5}$.

3.0 Results

Table 1 shows the characteristics of the Medicare cohort between 2000 and 2016. There were 623,036,820 person-years of follow-up during the study, and 29,481,444 deaths. 85.4% of the

participants were white, and 12.9% were covered by Medicaid, which provides additional benefits to the poor. The mean $PM_{2.5}$ during the study was 10.3 $\mu g/m^3$.

Table 1

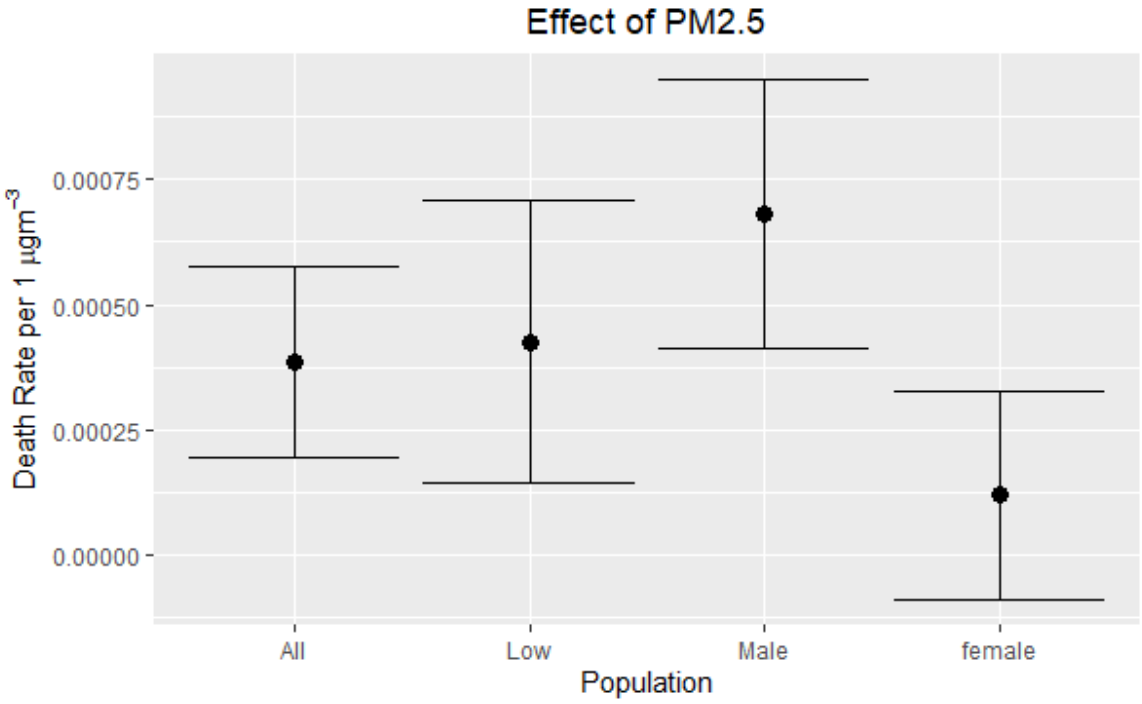
Variable	Values
Year	
Mean (SD)	2009.5 (4.90)
Median [25%, 75%]	2010 [2004, 2013]
Male	42.8 %
Race	
Black	8.4%
Other	6.2%
White	85.4%
Age > 84	13.2%
Medicaid Coverage	12.9%
ZIP Code Covariates	
Median Income	
Mean (SD)	$53,177 ($22,082)
Median [25%, 75%]	$47,998 [$38,030, $63,031]

Median House Value	
Mean (SD)	$200,139 ($159,728)
Median [25%, 75%]	$150,400 [$98,600, $240,300]
Percent ZIP code Black	
Mean (SD)	11% (17.9%)
Median [25%, 75%]	3.7% [1.1%, 12.0%]
Percent ZIP code Hispanic	
Mean (SD)	12.6% (16.4%)
Median [25%, 75%]	5.3% [2.1%, 14.6%]
Percent >65 below poverty	
Mean (SD)	9.5% (6.5%)
Median [25%, 75%]	7.9% [5.3%, 11.8%]
Percent Low Education	
Mean (SD)	25.3% (14.7%)
Median [25%, 75%]	22.6% [14.2%, 33.7%]
Percent with annual Mammogram	
Mean (SD)	63.7% (7.2%)
Median [25%, 75%]	63.9% [59.2%, 68.2%]

Percent with ambulatory Visit	
Mean (SD)	77.8% (6.2%)
Median [25%, 75%]	79.0% [74.4%, 82.1%]
Population Density (persons/mi^2)	
Mean (SD)	3397 (9032)
Median [25%, 75%]	967 [167, 3353]
Percent Owner Occupied	
Mean (SD)	68% (16%)
Median [25%, 75%]	70.8% [59.8%, 79.2%]
Mean BMI (kg/m^2)	
Mean (SD)	27.5 (1.58)
Median [25%, 75%]	27.3 [26.7, 28.0]
Distance to nearest hospital (km)	
Mean (SD)	6.5 (7.4)
Median [25%, 75%]	3.90 [1.98, 8.07]
Ever Smoker	
Mean (SD)	46.2% (6.8%)
Median [25%, 75%]	46.2% [41.8%, 50.4%]

Percent annual HbA1c test	
Mean (SD)	83.1% (4.9%)
Median [25%, 75%]	83.7% [80.5%, 86.3%]
Lung Cancer Rate (x 10^{-4})	
Mean (SD)	3.9 (2.8)
Median [25%, 75%]	3.3 [1.9, 4.9]
Percent annual LDL	
Mean (SD)	79.5% (6.2%)
Median [25%, 75%]	80.1% [76.1%, 83.5%]
Percent Annual Eye Exam	
Mean (SD)	67.4% (6.4%)
Median [25%, 75%]	67.1% [63.9%, 71.0%]
$PM_{2.5}$ ($\mu g/m^3$)	
Mean (SD)	10.3 (3.1)
Median [25%, 75%]	9.8 [7.9, 12.0]

In the meta-analysis of the results of the 32 strata-specific DID analyses, we found that the probability of dying in each year increased by $3.85x10^{-4}$ (95% CI $1.95 x10^{-4}$, $5.76 x10^{-4}$) for each 1 $\mu g/m^3$ increment in $PM_{2.5}$ in that year. The I^2 statistic for heterogeneity was 42%. When we restricted our analysis to persons whose exposure was always below 12 $\mu g/m^3$ during the follow-up period, we found a larger effect size, with the probability of dying in each year increased by 4.26 $x10^{-4}$ (95% CI $1.43x10^{-4}$, $7.09 x10^{-4}$) per 1 $\mu g/m^3$ increase in $PM_{2.5}$. Interaction terms for male gender, age > 84, and race were fit in the full data. Sex was a significant modifier (p for interaction <0.001), with larger effects in males (6.81x10-^{4}, 95% CI $4.14x10^{-4}$, $9.48x10^{-4}$) then females($1.20x10^{-4}$, 95% CI $-8.80x10^{-5}$, $3.29x10^{-4}$). These results are shown in Figure 1. There was no interaction by age. Interaction models for race did not converge because residential segregation in the U.S. resulted in groups with too few Blacks or Asians and other races. Instead, we reran the analysis without separate models for each of the 32 groups. There was no significant interaction by race.



4.0 Discussion

Using a difference in differences design applied to a linear rate model, we found that each 1 $\mu g/m^3$ increment in $PM_{2.5}$ was associated with a 3.85 $x10^{-4}$ increase in the probability of dying in a given year. If the difference in differences assumptions are met, this is a causal increase. We believe they are met for the following reasons. First, since this design controls for each ZIP code, all individual and neighborhood level confounders that change little over time are controlled, whether measured or not. This includes most of the variables (e.g. SES, smoking history, diet) that have be posited as potential confounders. Consequently, only time varying factors can be confounders. Second, we controlled for potential confounding in each ZIP code due to time trends in median household income, median home value, percent owner occupied housing, percent of ZIP code that is Black, percent of households that are Hispanic, percent of persons aged 65 or older living in poverty, smoking rate, BMI, Medicaid eligibility, educational attainment, population density, lung cancer rates, multiple measures of the adequacy of medical care, and summer and winter temperature. Third, we grouped the ZIP codes by these factors, and fit separate nonlinear time trends within each of 32 groups to capture any remaining time trends due to omitted confounders could differ between groups, but would be similar within group. This approach effectively looks at the within ZIP code fluctuations in exposure around the ZIP code mean, trend due to measured time-varying covariates, and common trends by group of ZIP codes. It compares that to the same deviation in mortality. Such an approach, looking at yearly deviations from trend and ZIP code mean in exposure and outcome, is inherently examining the relationship of changes in exposure to changes in outcome. This also addresses the issue of whether previous studies effects are due to primarily recent exposure, or reflect long term exposure, including when pollution concentrations were

higher. The dummy variable for each ZIP code controls for long term exposure at that ZIP code, and the removal of nonlinear time trends during the period under study focuses the exposure variable on the year of the death. EPA Regulatory Impact Analyses spread the estimated mortality effects out over a 20 year period. This study provides an estimate of immediate impact. Because our study incorporates 17 years of follow-up, each year has a new exposure, and a new effect.

In addition this paper adds to the sparse literature controlling for temperature in studies of long-term exposure to air pollution, which some have argued is an important confounder. Further, we estimated the probability of dying in a year, which is more easily interpretable than an instantaneous hazard rate, and by using an additive rather than multiplicative model we estimated the marginal effect if $PM_{2.5}$, not the conditional effect estimated by Cox's proportionate hazard model. This allows one to estimate attributable deaths in health impact assessments without making further, possibly implausible assumptions required when using a conditional estimate.

Some scientists, including the current chair of EPA's Clean Air Scientific Advisory Committee (CASAC), assert that studies using standard epidemiological methods should be given little weight in revising the NAAQS, and propose restricting to studies using causal methods, and particularly ones showing changes in air quality are associated with changes in mortality(CASAC, 2019). The recent meetings of EPA's CASAC highlighted the importance of these issues(CASAC., 2019). Their main criticism is that traditional approaches only show associations that may be confounded, vary depending on modeling approaches, and do not inform causality, which can only be addressed by causal methods. They also emphasize that unmeasured variables, particularly individual characteristics, socioeconomic status, and temperature may confound the published literature.(Cox and Popken, 2015) EPA recently proposed not tightening the NAAQS for $PM_{2.5}$, relying on these arguments. This paper provides an analysis using a causal method, controlling for

temperature and socioeconomic status, and all individual and area level potential confounders, measured or unmeasured, that vary slowly over time.

That EPA proposal also asserted there was insufficient evidence of a causal association at lower levels. When we restricted our analysis to include only persons who never experienced concentrations above 12 $\mu g/m^3$ during 17 years of follow-up we found a somewhat larger effect estimate of 4.26×10^{-4} (95% CI 1.43×10^{-4}, 7.09×10^{-4}) per 1 $\mu g/m^3$ increase in $PM_{2.5}$...5. This indicates that the current U.S. and E.U. standards are not sufficient to protect public health, and that the WHO standard of 10 $\mu g/m^3$ is unlikely to protect public health.

Other studies have applied causal modeling methods to air pollution, primarily propensity score methods(Abu Awad et al., 2019; Schwartz et al., 2018; Wang et al., 2017; Wei et al., 2020; Wu et al., 2020; Yitshak-Sade et al., 2019). These methods use the relationship between exposure and confounders to render the exposure independent of all of the measured confounders, and hence mimic a randomized trial. They have all reported that $PM_{2.5}$ increases mortality rates. The difference in differences approach complements those studies by its ability to deal with unmeasured confounders. All personal and small area time invariant or slowly varying confounders are removed by design, whether measured or not. All confounders whose time trends are due to measured time-varying confounders or similar within groups defined by race, SES, medical access, and behavioral characteristics are controlled whether measured or not. Hence, this paper adds assurance about many possible unmeasured confounders to the large literature of associational studies and smaller literature of propensity score-based models that provide causal estimates. Together, they provide strong evidence for a causal effect of $PM_{2.5}$ on mortality rates.

Since we estimate the probability of dying in each year and not a hazard rate, our effect sizes are not directly comparable to the other causal modeling studies. Compared to the larger literature, a

recent meta-analysis of then extant cohort studies estimated the effect size at 10 μg/m^3 (the mean concentration in this study) as a 1.29% increase in the rate per μg/m^3 increase in PM2.5, (95% CI 1.09%, 1.50%)(Vodonos et al., 2018).The annual mortality rate in the Medicare cohort was $4.7x10^{-2}$. A 1.29% increase in that rate is an additive increase of $6.1x10^{-4}$. The results from a previous Cox regression analysis of the Medicare cohort from 2000-2012 translate to an additive increase of $3.4x10^{-4}$. These are similar to our results. Hence these other studies are unlikely to have been confounded by temperature, or slowly varying SES, racial, and behavioral factors which this study controlled for.

Nor are these effects small. Multiplying our effect estimates by the total person-years in the Medicare cohort, we estimate that had everyone had 1 μg/m^3 lower exposure, 239,900 early deaths would have been avoided during the follow-up period. EPA's National Contingency Plan (40 C.F.R. § 300.430(d)(1)) states that the range of acceptable **lifetime risks** (of developing cancer) for carcinogens should be set between 1 in 10,000 and 1 in a million over a 70-year lifetime. Thus, when EPA considers regulations for carcinogens, it typically regulates if **lifetime risks** exceed 1 in a hundred thousand. In contrast, 1 μg/m^3 of exposure below the current EPA standard for only 1 year results in an increased risk of dying of 4.26 per ten thousand in our study.

Our finding has limitations. First, DID analyses depend on the change over time in other ZIP codes with different changes in $PM_{2.5}$ to serve as controls for changes over time in outcome that may have occurred independent of exposure. If the time trends in the ZIP codes are different, this control will fail. We have dealt with this by controlling for time trends in measured covariates and grouping ZIP codes into 32 groups that are similar on age, sex, race/ethnicity, SES, and access to medical care, and doing the analysis separately within each group, arguing that the time trends in

mortality rates will be similar within group. However, we cannot exclude the possibility that they are not. Second, our exposure estimates are not perfect. While an out of sample R^2 of 0.89 is high, there is still some exposure error, which may bias estimates. In addition, personal exposure within a neighborhood varies around the neighborhood ambient concentration. However, we believe most of that difference is likely to be Berksonian error, and hence not bias coefficients. Moreover, the principle reason for the differences between ambient and personal exposure are behavioral (more driving, more cooking, etc), and incorporating exposure related to those factors would require controlling for other related risk factors (e.g. stress from driving) that are not confounders of the neighborhood ambient concentrations. Hence, this exposure error is beneficial from the point of view of reducing confounding, as has been pointed out previously(Weisskopf and Webster, 2017).

In conclusion, we have found an effect of $PM_{2.5}$ on daily deaths using a causal modeling approach robust to unmeasured confounders. The effect size is similar to those reported in associational studies, suggesting that unmeasured confounders are not an issue with them, and is large enough to indicate that reducing $PM_{2.5}$ concentrations in the U.S. could save tens of thousands of premature deaths each year.

References

Abu Awad, Y., et al., 2019. Change in PM2.5 exposure and mortality among Medicare recipients. Environmental Epidemiology. 3, e054.

Baccini, M., et al., 2017. Assessing the short term impact of air pollution on mortality: a matching approach. Environ Health. 16, 7.

Beelen, R., et al., 2014. 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, 785-95.

Bureau, U. S. C., US. Census 2000. Summary File 3 (SF 3). In: U. S. C. Bureau, (Ed.), 2010.

CASAC, Public Teleconference on Particulate Matter (PM). EPA Chartered Clean Air Scientific Advisory Committee (CASAC), Available at https://yosemite.epa.gov/sab/sabproduct.nsf/MeetingCalCASAC/4F40665AD1DDCEF6852583A000645464?OpenDocument., 2019.

CASAC., CASAC Review of the EPA's Integrated Science Assessment for Particulate Matter. . In: U. S. E. P. Agency, (Ed.), Washington, D.C., Apr. 2019, 2019.

Caudill, S., Jackson, J., 1989. Measuring Marginal Effects in Limited Dependent Variable Models. Journal of the Royal Statistical Society, Series B. 38, 203-6.

CDC, Behavioral Risk Factor Surveillance System. BRFSS 2013 Survey Data and Documentation. . In: C. f. D. Control, (Ed.), 2013.

Cox, L. A., Jr., Popken, D. A., 2015. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? Ann Epidemiol. 25, 162-73.

Crouse, D. L., et al., 2015. Ambient PM2.5, O3, and NO2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). Environ Health Perspect. 123, 1180-6.

Di, Q., et al., 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ Int. 130, 104909.

Di, Q., et al., 2020. Assessing NO2 Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging. Environmental Science & Technology. 54, 1372-1384.

Di, Q., et al., 2017. Air Pollution and Mortality in the Medicare Population. N Engl J Med. 377, 1498-9.

Donald, S., Lang, K., 2007. Inference with difference-in-differences and other panel data. . Rev Econ Stat. 89, 221–233.

Hoek, G., et al., 2013. Long-term air pollution exposure and cardio- respiratory mortality: a review. Environ Health. 12, 43.

Imai, K., van Dyke, D. A., 2004. Causal inference with general treatment regimes: Generalizing the Propensity Score. Journal of the American Statistical Association. 99, 854:866.

Pinault, L., et al., 2016. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. Environmental Health. 15, 18.

Pope, C. A., 3rd, et al., 2019. Mortality Risk and Fine Particulate Air Pollution in a Large, Representative Cohort of U.S. Adults. Environ Health Perspect. 127, 77007.

RESDAC, Denominator File. 2018.

Rubin, D. B., 1997. Estimating causal effects from large data sets using propensity scores. Ann Intern Med. 127, 757-63.
Schwartz, J. D., et al., 2018. Estimating the Effects of PM2.5 on Life Expectancy Using Causal Modeling Methods. Environ Health Perspect. 126, 127002.
Vodonos, A., et al., 2018. The concentration-response between long-term PM2.5 exposure and mortality; A meta-regression approach. Environmental Research.
Wang, Y., et al., 2016. Estimating Causal Effects of Long-Term PM2.5 Exposure on Mortality in New Jersey. Environ Health Perspect. 124, 1182-8.
Wang, Y., et al., 2017. Doubly Robust Additive Hazards Models to Estimate Effects of a Continuous Exposure on Survival. Epidemiology (Cambridge, Mass.). 28, 771-779.
Wei, Y., et al., 2020. Causal effects of air pollution in Massachusetts. American journal of epidemiology. in press.
Weisskopf, M., Webster, T., 2017. Trade-offs of Personal Versus More Proxy Exposure Measures in Environmental Epidemiology. Epidemiology (Cambridge, Mass.). 635-43.
Wennberg, J., Cooper, M., 1996. The Dartmouth atlas of health care: . American Hospital Publishing Chicago, IL;.
Wu, X., et al., 2020. Evaluating the Impact of Long-term Exposure to Fine Particulate Matter on Mortality Among the Elderly. Science Advances.
Yang, C., et al., 2012. Alternative ozone metrics and daily mortality in Suzhou: the China Air Pollution and Health Effects Study (CAPES). Sci Total Environ. 426, 83-9.
Yitshak-Sade, M., et al., 2019. Estimating the causal effect of annual PM2.5 exposure on mortality rates in the Northeastern and mid-Atlantic states. Environmental Epidemiology. Latest Articles.

Figure Legend. Figure 1 shows the effect size estimate (absolute increase in the death rate for each 1 μg/m^3 increase in $PM_{2.5}$ exposure, and 95% Confidence Interval) for the entire Medicare Cohort in 2000-2016 (All), for only persons never exposed to $PM_{2.5}$ concentrations above 12 μg/m^3 (low) during the follow-up period, for males in the entire Medicare Cohort (males) and for females in the entire Medicare Cohort.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Dr. Schwartz declares that he has testified as a health expert for the U.S. Department of Justice in a law suit over clean air act violations. No other authors have anything to declare.