Long-term effect of exposure to lower concentrations of air pollution on mortality among US Medicare participants and vulnerable subgroups: a doubly-robust approach

Mahdieh Danesh Yazdi, Yan Wang, Qian Di, Weeberb J Requia, Yaquang Wei, Liuhua Shi, Matthew Benjamin Sabath, Francesca Dominici, Brent Coull, John S Evans, Petros Koutrakis, Joel D Schwartz

Summary

Background Long-term exposure to air pollution has been linked with an increase in risk of mortality. Whether existing US Environmental Protection Agency standards are sufficient to protect health is unclear. Our study aimed to examine the relationship between exposure to lower concentrations of air pollution and the risk of mortality.

Methods Our nationwide cohort study investigated the effect of annual average exposure to air pollutants on all-cause mortality among Medicare enrollees from the beginning of 2000 to the end of 2016. Patients entered the cohort in the month of January following enrolment and were followed up until the end of the study period in 2016 or death. We restricted our analyses to participants who had only been exposed to lower concentrations of pollutants over the study period, specifically particulate matter less than 2.5 µg/m³ in diameter (PM₂.₅) at a concentration of up to 12 µg/m³, nitrogen dioxide (NO₂) at a concentration of up to 53 parts per billion (ppb), and summer ozone (O₃) at concentrations of up to 50 ppb. We adjusted for two types of covariates, which were individual level and postal code-level variables. We used a doubly-robust additive model to estimate the change in risk. We further looked at effect-measure modification by stratification on the basis of demographic and socioeconomic characteristics.

Findings We found an increased risk of mortality with all three pollutants. Each 1 µg/m³ increase in annual PM₂.₅ concentrations increased the absolute annual risk of death by 0.073% (95% CI 0.071–0.076). Each 1 ppb increase in annual NO₂ concentrations increased the annual risk of death by 0.003% (0.003–0.004), and each 1 ppb increase in summer O₃ concentrations increased the annual risk of death by 0.081% (0.080–0.083). This increase translated to approximately 11,540 attributable deaths (95% CI 11,087–11,992) for PM₂.₅, 1,176 attributable deaths (998–1,353) for NO₂, and 15,115 attributable deaths (14,896–15,333) for O₃ per year for each unit increase in pollution concentrations. The effects were higher in certain subgroups, including individuals living in areas of low socioeconomic status. Long-term exposure to permissible concentrations of air pollutants increases the risk of mortality.

Funding The US Environmental Protection Agency, National Institute of Environmental Health Services, and Health Effects Institute.

Copyright © 2021 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.
Research in context

Evidence before this study
We searched PubMed for articles published in English with the following search terms: "air pollution", "particulate matter", "ozone", or "nitrogen dioxide" and "mortality". We limited the range of our search to the past 10 years and studies focusing on human adults. Several large cohort studies have found that the adverse health effects of long-term exposure to air pollution might be persistent even at lower concentrations, such as threshold concentrations set by the US Environmental Protection Agency and WHO. These cohorts are useful in that they include large subsets of individuals exposed to lower concentrations of air pollutants with enough power to detect effects. Studies that modelled dose-response curves did not find a pollution level at which no harm was observed.

Added value of this study
In this study we used a doubly-robust, causal method to estimate the risk of mortality among people who were exposed to low concentrations of air pollutants over the study period.

We used a large dataset that gave us enough power to detect whether an association existed in demographic and socioeconomic subgroups. Furthermore, we assigned exposure on the basis of prediction models that were built on a fine scale with strong validation metrics.

All of the studied pollutants increased the risk of death in our study. We found that thousands of deaths could be attributed to small increases in annual air-pollution concentrations. We also found an increased risk of death for people living in lower-income areas, suggesting that disparities exist in the adverse effects of air pollution.

Methods

Study population
We designed this study as a nationwide cohort study. Our cohort included all patients aged at least 65 years enrolled in Medicare in the USA from beginning 2000 to end 2016. Medicare is a national insurance programme for US residents that primarily covers the insurance of older people (≥65 years of age). Medicare beneficiaries made up between 13% and 16% of the total population between 2000 and 2016 in the USA and 97% of older people in the year 2000. Patients entered the cohort in the month of January of the year following enrolment and were followed up until the end of study period in 2016 or death. Our study was approved by the IRB at Harvard TH Chan School of Public Health.

Exposure assessment
We established the level of exposure to PM$_{2.5}$, NO$_2$, and O$_3$ on the basis of estimates generated from spatiotemporal ensemble models. Briefly, predictors were extracted from satellite-based measurements, land-use data, meteorological data, and chemical-transport models. The predictors were used as input for three machine-learning algorithms, including a random forest, a gradient-boosting machine, and a neural network. We put the predictions generated by these algorithms into a geographically weighted generalised additive model, which in turn produced daily predictions for the contiguous USA, between 2000 and 2016 on a 1 km$^2$ spatial scale. We averaged predictions across days in each calendar year to obtain annual averages, and aggregated grid cells to postal codes and assigned exposures to individuals on the basis of their residential postal code in each year. All models showed strong performance, with ten-times cross-validation $R^2$ values of 0.89 for PM$_{2.5}$, 0.84 for NO$_2$, and 0.86 for O$_3$. We specifically used warm-season O$_3$ concentrations, from April 1 to Sept 30, given that this season is associated with higher average concentrations. All references to O$_3$ in this Article refer to warm-season O$_3$.

We limited our dataset to individuals who were exposed to concentration that were lower than the maximum recommended by US regulations for all years of follow-up. For PM$_{2.5}$, that threshold is 12 µg per m$^3$, although the calculation of the EPA involves taking the average of annual concentrations over 3 years at each monitoring station. The maximum recommended concentration for NO$_2$ is 53 ppb. O$_3$ does not have regulations for long-term exposure, so we chose 50 ppb to have enough observations.

We further dropped the lowest three percentile for each exposure. Such low exposures are rare, and modelling is not as accurate when there are scarce measured values to train the models.
Outcome assessment
We obtained information on all-cause mortality from the Medicare denominator file.

Covariate assessment
We adjusted for two types of covariates in both the propensity-score model and the outcome regression, which were individual-level and postal code-level (ie, ZIP code-level) variables. Individual level variables included age, sex, race, and Medicaid eligibility (supplemental insurance for individuals with low incomes). These data were obtained from the Medicare-denominator file and were updated annually.

We derived postal code-level variables from data generated by the US Decennial Census and the American Community Survey in 2000, 2010, 2011, 2012, 2013, 2014, 2015, and 2016. Values for the remaining years were obtained through interpolation. The variables we included in our models were population density, percentage of the population older than 65 years living below the poverty line, percentage of the housing occupied by the owners, median value of the housing occupied by the owners, median household income, percentage of the population listed as Black, percent of the population listed as Hispanic, and percentage of the elderly population who did not graduate from high school. We further derived information on the percentage of the population who have ever smoked and the mean body-mass index of the respondents at the county level from the Behavioral Risk Factor Surveillance System. We used the Dartmouth Health Atlas to obtain information on the proportion of Medicare beneficiaries with at least one haemoglobin-A1c test in a year, the proportion of beneficiaries with diabetes older than 65 years who had a lipid-panel test in a year, the proportion of beneficiaries who had an eye examination in a year, the proportion of beneficiaries with at least one ambulatory doctor’s visit in a year, and the proportion of female beneficiaries who had a mammogram over a 2 year period. We obtained information on the rate of lung cancer in the population from hospital admission data in the Medicare Provider Analysis and Review (MEDPAR) dataset. We also included information on the division of residence, as defined by the US Census, and distance to the nearest hospital. We calculated the distance to the nearest hospital using the distance from the centroid of the postal code to the nearest facility on the basis of data from ESRI in 2010.

We assigned temperature levels using the gridMET dataset, which provides daily estimated meteorological parameters on a 4 km-by-4 km scale. We aggregated these levels to postal codes and years. We created one variable for the warm season, which was the average temperature from April to the end of September, and one variable for the cold season, which was the average temperature from January to March and October to December of the same calendar year. Observations with missing information were assumed to be missing at random and were excluded from analysis. These represented less than 1% of the data.

Statistical analysis
Several papers have used propensity scores to analyse the association between air pollution and mortality. Briefly, the approach starts by first examining the association between exposure and confounders and uses that relationship to mimic a randomised trial, which makes the exposure independent from the confounders by creating a pseudopopulation in which individuals exposed to different concentrations of exposure are exchangeable with respect to the covariates. Specifically, we accounted for confounding via two mechanisms. First, with inverse-probability weights (IPWs) of exposure, and second, by adjusting for confounders in the outcome regression model. If either of the models is correctly specified, the estimated coefficient is unbiased. In the first step, we calculated

| PM$_{2.5}$ | NO$_2$ | O$_3$ |
|-----------|-------|-------|
| Number of individuals | 40,422,099 | 72,769,408 | 44,430,747 |
| Number of deaths | 10,365,012 (25.6%) | 28,994,819 (39.30%) | 14,589,797 (32.8%) |
| Sex | | | |
| Female | 21,692,819 (53.7%) | 40,310,234 (55.39%) | 24,283,979 (54.7%) |
| Male | 18,729,280 (46.3%) | 32,459,174 (44.61%) | 20,146,768 (45.3%) |
| Race | | | |
| White | 34,270,600 (84.8%) | 61,205,628 (84.11%) | 36,688,655 (82.6%) |
| Black | 2,889,598 (7.1%) | 6,525,454 (8.97%) | 4,924,472 (9.7%) |
| Other | 3,261,901 (8.6%) | 5,038,326 (6.92%) | 3,447,620 (7.8%) |
| Demographic characteristics by observations | | | |
| Number of person-years | 267,111,005 | 604,486,421 | 315,767,917 |
| Medicaid eligibility | | | |
| Yes | 39,548,971 (11.4%) | 76,308,600 (12.62%) | 39,968,480 (12.7%) |
| No | 236,562,034 (88.6%) | 528,177,821 (87.38%) | 275,799,437 (87.3%) |
| Age group | | | |
| Older than 64 years and younger than 75 years | 178,420,757 (66.8%) | 327,234,320 (54.13%) | 189,968,539 (60.2%) |
| 75 years or older and younger than 85 years | 63,724,472 (23.9%) | 198,374,334 (32.82%) | 89,226,756 (28.3%) |
| At least 85 years of age | 24,965,776 (9.3%) | 78,897,967 (13.05%) | 36,572,622 (11.6%) |
| Division | | | |
| Pacific | 40,504,825 (15.2%) | 81,087,764 (13.41%) | 40,028,882 (12.7%) |
| Southwest central | 29,691,495 (11.1%) | 62,146,699 (10.28%) | 31,444,181 (9.9%) |
| Northwest central | 30,581,901 (11.4%) | 43,995,791 (7.28%) | 26,397,005 (8.4%) |
| Northeast central | 21,649,239 (8.1%) | 39,875,500 (6.52%) | 59,337,855 (18.18%) |
| Southeast central | 6,839,987 (2.6%) | 35,755,594 (5.92%) | 12,336,829 (3.9%) |
| Mountains | 26,845,187 (10.1%) | 34,138,704 (5.65%) | 3,131,927 (10.0%) |
| South Atlantic | 58,408,003 (21.9%) | 124,552,351 (20.60%) | 58,080,670 (18.4%) |
| Middle Atlantic | 26,95,2775 (10.1%) | 90,360,494 (14.95%) | 55,836,410 (17.7%) |
| New England | 25,636,593 (9.6%) | 32,575,524 (5.39%) | 29,474,158 (9.3%) |

Data are n (%). PM$_{2.5}$ - particulate matter less than 2.5 µg/m$^3$ in diameter. NO$_2$ - nitrogen dioxide. O$_3$ - ozone.

Table 1: Demographic characteristics of Medicare enrollees
stabilised IPWs for each exposure of interest using the following formula:

$$\frac{f(x)}{SW_{\text{stabilised}}} = \frac{f(x)}{f(x|v)}$$

In which x represents the exposure, v represents the covariates, and SW is the stabilised weight. The numerator was calculated as the probability density function of exposure from an intercept-only linear regression. The denominator was calculated as the probability density function of exposure given the covariates from a linear regression that included quadratic terms for all continuous confounders. To account for outliers, the highest percentile of weights was given the value at the 99th percentile, and the lowest percentile of weights was given the value at the 1st percentile.

In the second step, we ran a linear probability model to estimate the probability of death given the covariates and exposures of interest weighted by the IPWs calculated in the previous step.

$$Pr(\text{death} = 1) = \beta_0 + \beta x + s(v, y)$$

In which x represents the exposure, v represents the vector of covariates, and y represents the parameterisation of the covariates, which in this case included cubic terms for the other pollutants, temperature variables, age, and median household income. The equation was weighted by the weights calculated in the previous steps. For each exposure, we adjusted for the other two pollutants, both in the IPWs and in the outcome regression. A directed acyclic graph of the outcome regression is shown in the appendix (p 2). We used robust standard errors to account for the heteroscedasticity created by the inclusion of multiple observations from the same person and the use of a linear rather than logistic probability model. Because of our use of a linear probability model, the regression models estimate the absolute change in the risk of dying. To estimate the number of annual cases attributable to the exposures, we used the following formula:

Attributable cases = Risk difference per unit change per year × (Person–years of observation / duration of study)

Given that the dataset was large, we divided it randomly into groups and obtained the coefficients and standard errors using a fixed-effects meta-analysis of the group-specific results.

We also did single-pollutant analyses for comparison. We further identified subgroups vulnerable to potential environmental justice issues by looking at effect-measure modification (EMM) through stratification by Medicaid eligibility, race, sex, age group, quartiles of population density (as a measure of urbanicity), proportion of the population who identify as Hispanic, and median household income.

As a sensitivity analysis, we calculated E-values. Evidence-for-causality values identify the magnitude of the strength of the relationship that an unmeasured confounder would need to have with both the exposure and the outcome for its inclusion to change the effect estimate found to the null. A higher E-value is evidence of an analysis that is more robust to unmeasured confounding. Further, we ran the analyses adding cubic terms for all continuous predictors to the propensity-score model and quadratic terms in the outcome-regression model and quadratic terms in the outcome-regression model.
model to see whether this adjustment would alter the effect estimates. We evaluated the balance of covariates (after weighting) for the continuous variables by calculating the average absolute correlation coefficient between the exposures and the continuous variables, and tested whether using cubic terms would be better than quadratic terms.27

All data cleaning and statistical analyses were done in R statistical software version 3.5.1.

Role of the funding source
The funders of this study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results
The baseline characteristics of our population are presented in table 1. We identified approximately 40 million individuals with 267 million person-years of follow-up for PM$_{2.5}$, 73 million individuals with 604 million person-years of follow-up for NO$_2$, and 44 million individuals with 316 million person-years of follow-up for O$_3$. Most Medicare participants were White and there were slightly more women than men.

The distribution of the exposure of interest in each dataset is shown in table 2 and the exposure distribution by demographic characteristics is shown in the appendix (pp 3–6). The correlation matrix for the exposures in each dataset is also presented in the appendix (p 7). The pollutants showed low-to-moderate correlation with each other, with correlation coefficient values of 0·11 to 0·59. Quartile definitions used in the EMM analyses are presented in the appendix (p 8).

The results of the primary analysis can be seen in table 3. Each single-unit increase in pollution concentration was associated with a 0·073% (95% CI 0·071–0·076) increase in the risk of death per year for PM$_{2.5}$, 0·003% (0·002–0·004) for NO$_2$, and 0·081% (0·080–0·083) for O$_3$. This increase translated to approximately 73 million individuals with 604 million person-years of follow-up for NO$_2$, and 44 million individuals with 316 million person-years of follow-up for O$_3$. Most Medicare participants were White and there were slightly more women than men.

The distribution of the exposure of interest in each dataset is shown in table 2 and the exposure distribution by demographic characteristics is shown in the appendix (pp 3–6). The correlation matrix for the exposures in each dataset is also presented in the appendix (p 7). The pollutants showed low-to-moderate correlation with each other, with correlation coefficient values of 0·11 to 0·59. Quartile definitions used in the EMM analyses are presented in the appendix (p 8).

The results of the primary analysis can be seen in table 3. Each single-unit increase in pollution concentration was associated with a 0·073% (95% CI 0·071–0·076) increase in the risk of death per year for PM$_{2.5}$, 0·003% (0·002–0·004) for NO$_2$, and 0·081% (0·080–0·083) for O$_3$. This increase translated to approximately 73 million individuals with 604 million person-years of follow-up for NO$_2$, and 44 million individuals with 316 million person-years of follow-up for O$_3$. Most Medicare participants were White and there were slightly more women than men.

The distribution of the exposure of interest in each dataset is shown in table 2 and the exposure distribution by demographic characteristics is shown in the appendix (pp 3–6). The correlation matrix for the exposures in each dataset is also presented in the appendix (p 7). The pollutants showed low-to-moderate correlation with each other, with correlation coefficient values of 0·11 to 0·59. Quartile definitions used in the EMM analyses are presented in the appendix (p 8).

The results of the primary analysis can be seen in table 3. Each single-unit increase in pollution concentration was associated with a 0·073% (95% CI 0·071–0·076) increase in the risk of death per year for PM$_{2.5}$, 0·003% (0·002–0·004) for NO$_2$, and 0·081% (0·080–0·083) for O$_3$. This increase translated to approximately 73 million individuals with 604 million person-years of follow-up for NO$_2$, and 44 million individuals with 316 million person-years of follow-up for O$_3$. Most Medicare participants were White and there were slightly more women than men.

The distribution of the exposure of interest in each dataset is shown in table 2 and the exposure distribution by demographic characteristics is shown in the appendix (pp 3–6). The correlation matrix for the exposures in each dataset is also presented in the appendix (p 7). The pollutants showed low-to-moderate correlation with each other, with correlation coefficient values of 0·11 to 0·59. Quartile definitions used in the EMM analyses are presented in the appendix (p 8).

The results of the primary analysis can be seen in table 3. Each single-unit increase in pollution concentration was associated with a 0·073% (95% CI 0·071–0·076) increase in the risk of death per year for PM$_{2.5}$, 0·003% (0·002–0·004) for NO$_2$, and 0·081% (0·080–0·083) for O$_3$. This increase translated to approximately 73 million individuals with 604 million person-years of follow-up for NO$_2$, and 44 million individuals with 316 million person-years of follow-up for O$_3$. Most Medicare participants were White and there were slightly more women than men.

The distribution of the exposure of interest in each dataset is shown in table 2 and the exposure distribution by demographic characteristics is shown in the appendix (pp 3–6). The correlation matrix for the exposures in each dataset is also presented in the appendix (p 7). The pollutants showed low-to-moderate correlation with each other, with correlation coefficient values of 0·11 to 0·59. Quartile definitions used in the EMM analyses are presented in the appendix (p 8).

The results of the primary analysis can be seen in table 3. Each single-unit increase in pollution concentration was associated with a 0·073% (95% CI 0·071–0·076) increase in the risk of death per year for PM$_{2.5}$, 0·003% (0·002–0·004) for NO$_2$, and 0·081% (0·080–0·083) for O$_3$. This increase translated to approximately 73 million individuals with 604 million person-years of follow-up for NO$_2$, and 44 million individuals with 316 million person-years of follow-up for O$_3$. Most Medicare participants were White and there were slightly more women than men.

The distribution of the exposure of interest in each dataset is shown in table 2 and the exposure distribution by demographic characteristics is shown in the appendix (pp 3–6). The correlation matrix for the exposures in each dataset is also presented in the appendix (p 7). The pollutants showed low-to-moderate correlation with each other, with correlation coefficient values of 0·11 to 0·59. Quartile definitions used in the EMM analyses are presented in the appendix (p 8).

The results of the primary analysis can be seen in table 3. Each single-unit increase in pollution concentration was associated with a 0·073% (95% CI 0·071–0·076) increase in the risk of death per year for PM$_{2.5}$, 0·003% (0·002–0·004) for NO$_2$, and 0·081% (0·080–0·083) for O$_3$. This increase translated to approximately 73 million individuals with 604 million person-years of follow-up for NO$_2$, and 44 million individuals with 316 million person-years of follow-up for O$_3$. Most Medicare participants were White and there were slightly more women than men.

The distribution of the exposure of interest in each dataset is shown in table 2 and the exposure distribution by demographic characteristics is shown in the appendix (pp 3–6). The correlation matrix for the exposures in each dataset is also presented in the appendix (p 7). The pollutants showed low-to-moderate correlation with each other, with correlation coefficient values of 0·11 to 0·59. Quartile definitions used in the EMM analyses are presented in the appendix (p 8).
Table 3: Main study results, including E values

| Pollutant | CI 95% | Number of cases (CI 95%) | E value |
|-----------|--------|--------------------------|---------|
| PM$_{2.5}$, µg/m³ | 0.062% (0.056–0.068) | 2454 (2209–2699) | 1.057 |
| NO$_2$, ppb | 0.005% (0.004–0.006) | 432 (337–527) | 1.015 |
| O$_3$, ppb | 0.054% (0.051–0.056) | 2495 (2379–2611) | 1.053 |

3rd-quartile Hispanic percentage

| Pollutant | CI 95% | Number of cases (CI 95%) | E value |
|-----------|--------|--------------------------|---------|
| PM$_{2.5}$, µg/m³ | 0.062% (0.056–0.068) | 2454 (2209–2699) | 1.057 |
| NO$_2$, ppb | 0.005% (0.004–0.006) | 432 (337–527) | 1.015 |
| O$_3$, ppb | 0.054% (0.051–0.056) | 2495 (2379–2611) | 1.053 |

4th-quartile Hispanic percentage

| Pollutant | CI 95% | Number of cases (CI 95%) | E value |
|-----------|--------|--------------------------|---------|
| PM$_{2.5}$, µg/m³ | 0.075% (0.070–0.080) | 2950 (2739–3160) | 1.063 |
| NO$_2$, ppb | 0.004% (0.003–0.005) | 313 (222–405) | 1.013 |
| O$_3$, ppb | 0.071% (0.069–0.073) | 3280 (3191–3368) | 1.061 |

PM$_{2.5}$=particulate matter less than 2.5 µg/m³ in diameter. NA=not applicable. NO$_2$=nitrogen dioxide. O$_3$=ozone.

*Per one-unit increase in pollutants per year (1 µg/m³ for PM$_{2.5}$ and 1 ppb for NO$_2$ and O$_3$). †Covariates included other cases were not calculated.‡Negative values on a probability scale are not logical. As such, attributable number of people who did not graduate from high school, mean body-mass index, rate of smoking, Hispanic, population density per square mile, percentage of the population who are Black, percentage of the population who are Hispanic, percentage of the population living below the poverty line, percentage of the population who are White, percentage of people who had a mammogram, percentage of people who had a cholesterol exam, percentage of people who had an eye exam, percentage of people who had an A1c exam, percentage of people who had an ambulatory-care visit, percentage of people who had an eye exam, percentage of women who had a mammogram, percentage of the population who are Black, percentage of the population who are Hispanic, population density per square mile, percentage of the population living below the poverty line, percentage of the housing occupied by the owners, median value of house owned by the owners, median household income, percentage of the population who did not graduate from high school, mean body-mass index, rate of smoking, and rate of lung cancer. Negative values on a probability scale are not logical. As such, attributable number of cases were not calculated.

Table 3: Main study results, including E values

### Figure 1: Effect-measure modification by demographic characteristics

Risk-difference percentage (95% CI) change for each one-unit increase (1 µg per m$^3$ for PM$_{2.5}$ and 1 ppb for NO$_2$ and O$_3$) in annual pollutant concentrations among people who are always exposed to lower concentrations of air pollutants stratified by individual demographic characteristics. Pairwise comparisons of coefficients were done. Statistically significant differences (p<0.05) are indicated using asterisks. Ppb=parts per billion. PM$_{2.5}$=particulate matter less than 2.5 µg/m³ in diameter. NO$_2$=nitrogen dioxide. O$_3$=ozone.

The results of the analyses stratified by socioeconomic status can be seen in figure 2. For all pollutants, people who were eligible for Medicaid were at a greater risk of death than those who were not eligible for Medicaid (0.113%, 0.103–0.124 vs 0.067%, 0.064–0.070 for PM$_{2.5}$; 0.012%, 0.010–0.014 vs 0.004%, 0.004–0.005 for NO$_2$; and 0.117%, 0.113–0.120 vs 0.073%, 0.072–0.075 for O$_3$). This trend persisted when looking at income. People in the higher quartiles of postal code-level median household income had a lower risk of death than those in the lower quartiles of income.

The results of the sensitivity analyses can be seen in figure 3. Although changing the specification of the propensity-score model and the outcome regression altered the magnitude of the effect, the effect estimates were consistently positive. Our main model resulted in conservative estimates compared with several of the alternative specifications.

Finally, we checked the balance of covariates (appendix p 8). For all pollutants, weighting improved the balance of continuous covariates. We also ran the same analysis by adding cubic terms and it did not improve the balance of the continuous covariates for PM$_{2.5}$ and NO$_2$, although for O$_3$, the improvement was minor.
Discussion

Using a doubly-robust method, we found that among older individuals who were only exposed to lower concentrations of air pollutants during the study follow-up period, these pollutants increased the risk of death on an additive scale, after adjusting for other pollutants. This translated to tens of thousands of additional deaths per year per unit difference in exposure. We also confirmed that very old people and people with lower incomes were more vulnerable to air pollution. The results were made directly interpretable by our use of an additive probability model instead of the more usual multiplicative model, which builds in interactions. This finding adds to the literature indicating that the effects of air pollution on mortality are causal, and that standards are inadequate.

The EPA released a decision regarding tightening the standards for ambient PM$_{2.5}$. The EPA argued that “based on the available evidence, the Administrator has concluded that the current primary PM$_{2.5}$ standards are requisite to protect public health, with an adequate margin of safety, from effects of PM$_{2.5}$ in ambient air and should be retained, without revision.” Our results, which were based on current standards and were obtained using causal-modelling methodology, combined with the scientific assessment done by the EPA provide evidence that the US EPA Administrator’s decision for the annual PM$_{2.5}$ standard was unjustified. Our findings suggest that reduction of air-pollution concentrations through stricter regulations would reduce mortality among older people, and given that the pollutants that

---

Figure 2: Effect Measure Modification by Socioeconomic Characteristics

Risk-difference percentage (95% CI) change for each one-unit increase (1 µg/m³ for PM$_{2.5}$ and 1 ppb for NO$_2$ and O$_3$) in annual pollutant concentrations among people who are always exposed to lower concentrations of air pollutants stratified by socioeconomic characteristics. Income refers to postal code-level median household income and population density refers to postal code-level population density. Pairwise comparisons of coefficients were done. Statistically significant differences (p<0.05) are indicated using asterisks. Ppb=parts per billion. PM$_{2.5}$=particulate matter less than 2.5 µg/m³ in diameter. NO$_2$=nitrogen dioxide.

Figure 3: Model specification sensitivity analysis results

Risk-difference percentage (95% CI) change for each one-unit increase (1 µg/m³ for PM$_{2.5}$ and 1 ppb for NO$_2$ and O$_3$) in annual pollutant concentrations among people who are always exposed to lower concentrations of air pollutants. The figure shows a comparison of the main model used in our Article with a model that has cubic terms for the continuous variables in the PS model of the IPWs and a model that uses quadratic terms for age, median household income, temperature variables, and other pollutants in the OR. IPW=inverse probability weight. Ppb=parts per billion. PM$_{2.5}$=particulate matter less than 2.5 µg/m³ in diameter. NO$_2$=nitrogen dioxide. OR=outcome regression. O$_3$=ozone.
we studied have several common sources, reductions in one pollutant would most likely reduce concentrations of the other pollutants as well. There are several areas in which intervention could occur that would affect all the pollutants of interest, including but not limited to stricter controls on industry and fossil-fuel electric-generating units, larger and more efficient catalysts on automobiles, city planning to promote active transport, and improved public transit. For example, the reduction of NO\textsubscript{2} concentrations through higher-efficiency catalytic reduction or decreased vehicle use would likely reduce both PM\textsubscript{2.5} and O\textsubscript{3} as well.

Stratified analyses showed that people in higher age groups were at greater risk of death for all pollutants. Men were at greater risk of death caused by air pollution from PM\textsubscript{2.5} and O\textsubscript{3}, and people who identified as Black had a higher risk of death caused by NO\textsubscript{2} and O\textsubscript{3}. People with lower socioeconomic status tended to have a higher risk of death than those with higher socioeconomic status. The additional risk of mortality that we found in our stratified analyses for racial minorities and individuals with lower incomes raises concerns about environmental justice. These groups are historically disenfranchised and might suffer from exposure to air pollution and its adverse effects in a disproportionate manner.

Our results are not directly comparable to previous research, given that most studies have been done on a multiplicative scale or included higher exposures. A nationwide analysis of mortality in the Medicare cohort found that each increase of 10 µg per m\textsuperscript{3} in PM\textsubscript{2.5} concentration was associated with a hazard ratio of 1.136 (95% CI 1.131–1.141) for observations with exposure concentrations of less than 12 µg per m\textsuperscript{3}. The annual mortality in this cohort was 4.7×10\textsuperscript{-4}, and a 1.28% increase in that mortality for an increment of 1 µg per m\textsuperscript{3} in PM\textsubscript{2.5} would result in an additive increase of 6.0×10\textsuperscript{-4}, or a 0.060% increase in the absolute risk of dying each year, similar to our result of 0.073%. That model, however, did not control for NO\textsubscript{2}. Another study looking at Medicare enrollees in the southeast region of the USA found that each one-unit increase in PM\textsubscript{2.5} concentration was associated with a hazard ratio of 1.033 (95% CI 1.031–1.035) for concentrations of less than 12 µg per m\textsuperscript{3}. The additional risk of mortality that we found in our stratified analyses for racial minorities and individuals with lower incomes raises concerns about environmental justice. These groups are historically disenfranchised and might suffer from exposure to air pollution and its adverse effects in a disproportionate manner.

Our results are not directly comparable to previous research, given that most studies have been done on a multiplicative scale or included higher exposures. A nationwide analysis of mortality in the Medicare cohort found that each increase of 10 µg per m\textsuperscript{3} in PM\textsubscript{2.5} concentration was associated with a hazard ratio of 1.136 (95% CI 1.131–1.141) for observations with exposure concentrations of less than 12 µg per m\textsuperscript{3}. The annual mortality in this cohort was 4.7×10\textsuperscript{-4}, and a 1.28% increase in that mortality for an increment of 1 µg per m\textsuperscript{3} in PM\textsubscript{2.5} would result in an additive increase of 6.0×10\textsuperscript{-4}, or a 0.060% increase in the absolute risk of dying each year, similar to our result of 0.073%. That model, however, did not control for NO\textsubscript{2}. Another study looking at Medicare enrollees in the southeast region of the USA found that each one-unit increase in PM\textsubscript{2.5} concentration was associated with a hazard ratio of 1.033 (95% CI 1.031–1.035) for concentrations of less than 12 µg per m\textsuperscript{3}. A meta-regression looking at the dose-response relationship between long-term exposure to PM\textsubscript{2.5} and mortality found a 2.4% (95% CI 0.8–4.0) increase in the risk of death for each increase of 1 µg per m\textsuperscript{3} in annual concentrations by restricting to studies with mean pollution concentrations of less than 10 µg per m\textsuperscript{3}.

Our study has several strengths. First, we used doubly-robust causal methodology to estimate our effects. Second, our outcome regression model was an additive model. This meant that our effect estimates could be used to directly estimate the number of deaths attributable to air pollution, without regard to the distribution of covariates. Third, our initial study population included everyone enrolled in Medicare. As such, we have sufficient power to detect effects in our analyses and our results have greater generalisability. Fourth, many of the previous papers using causal methods have only examined associations with PM\textsubscript{2.5}, NO\textsubscript{2}, and O\textsubscript{3} simultaneously. Finally, we adjusted for several variables that could serve as proxies for socioeconomic status, which is the main confounder of concern in the air pollution–mortality relationship.

However, our study does have certain limitations. The causal implications of our study rely on the untestable assumption of no unmeasured confounding. We expect that should there be unmeasured confounding, it would most likely be caused by residual confounding through our inability to completely capture area-level socioeconomic status, although we did control for numerous variables in this category. We did try to assess the risk of potential unmeasured confounding by calculating E-values and found robust associations for PM\textsubscript{2.5} and O\textsubscript{3}. Moreover, we cannot know whether the specification of either the propensity score or outcome regression is correct, although balance metrics showed that, at least for the continuous variables, there is minimal correlation with the exposure after weighting. Furthermore, we did not have data on comorbidities, which might have acted as mediators or effect modifiers of the air pollution and mortality relationship. Our pollutants were moderately correlated with one another, which might cause some collinearity, although given the large sample size, this problem is less likely to be of concern. Finally, the exposures were assigned on the basis of estimations from a prediction model, which will result in some measurement error. This measurement error is likely to be higher at extreme values for which there is limited monitoring data. We addressed this issue by restricting to lower concentrations and excluding the last three percentiles of observations. However, there is still potential for error. Further, NO\textsubscript{2} and traffic-particle concentrations tend to decrease to urban background concentrations within 100–200 m of a busy road, and our NO\textsubscript{2} and PM\textsubscript{2.5} models do not capture the high exposure of people living within that distance of such a road. This is a limitation of our analysis.

Our study shows that for older individuals who are only exposed to low air-pollution concentrations during the follow-up period, an increase in the concentration of pollutants might still increase the risk of death. These results persisted in subgroup analyses, particularly for individuals with lower incomes. Regulators should consider setting annual standards for O\textsubscript{3} and tightening regulations for PM\textsubscript{2.5} and NO\textsubscript{2}.

Contributors MDY conceptualised the study, curated the data, and contributed to the formal data analysis, the investigation, methodology, and visualisation of the study, writing of original draft, and reviewing and editing of the manuscript. YWa contributed towards the methodology, and review and editing of the manuscript. QD participated in data curation. WJR, YWe, and LS participated in data curation, and
reviewing and editing of the manuscript. MBS contributed to the data curation, project administration, resources, and software use. FD contributed towards data curation, project administration, resources, software use, funding acquisition, and reviewing and editing of the manuscript. BC conceptualised the study and contributed to its methodology. JSE contributed towards the conceptualisation, methodology, and reviewing and editing of the manuscript. JC conceptualised the study and contributed to its methodology. JDS contributed towards the conceptualisation, methodology, funding acquisition, investigation, supervision, and reviewing and editing of the manuscript.

Declaration of interests

JDS has appeared as an expert witness on behalf of the US Department of Justice in cases involving violations of the Clean Air Act. FD has received consulting and speaking fees from Johnson and Johnson, Colgate, Sanofi, and Visa, but for research topics that are not related to the one of this Article. All other authors declare no competing interests.

Acknowledgments

This publication is made possible by US Environmental Protection Agency (EPA) grant RD-835872. Its contents are solely the responsibility of the grantee and do not necessarily represent the official view of the US EPA. Further, the US EPA does not endorse the purchase of any commercial products or services mentioned in the publication. This work was also supported by National Institute of Environmental Health Sciences P30-ES000002 and Health Effects Institute grant 4953-RFA14-3/16-4.

Data sharing

The data used in this study will not be made available publicly or to other researchers because of restrictions in the data-use agreement with the Centres for Medicare and Services (CMS). However, other investigators can apply to the CMS for their own data-use agreements to access the Medicare data.

References

1. Lepeule J, Laden F, Dockery D, Schwartz J. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard six cities study from 1954 to 2009. Environ Health Perspect 2012; 120: 965–70.

2. Beelen R, Raaschou-Nielsen O, Stafoggia M, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 21 European cohorts within the multicentre ESCAPE project. Lancet 2014; 383: 785–95.

3. Turner MC, Jerrett M, Pope CA, et al. Long-term ozone exposure and mortality in a large prospective study. Am J Respir Crit Care Med 2016; 193: 113–22.

4. Di Q, Wang Y, Zanobetti A, et al. Air pollution and mortality in the Medicare population. N Engl J Med 2017; 376: 2513–22.

5. Jerrett M, Burnett RT, Pope CA, et al. Long-term ozone exposure and mortality. N Engl J Med 2009; 360: 1085–95.

6. Pope CA 3rd, Lefler JS, Ezzati M, et al. Mortality risk and fine particulate air pollution in a large, representative cohort of US adults. Environ Health Perspect 2019; 127: 77007.

7. US Environmental Protection Agency. NAAQS Table. Jan 7, 2021. https://www.epa.gov/criteria-air-pollutants/naaqs-table (accessed Jan 7, 2021)

8. Pinault L, Tieploma M, Crouse DL, et al. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. Environ Health 2016; 15: 18.

9. Crouse DL, Peters PA, van Donkelaar A, et al. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study. Environ Health Perspect 2012; 120: 708–14.

10. Shi L, Zanobetti A, Kloog I, et al. Low-concentration PM2.5 and mortality: estimating acute and chronic effects in a population-based study. Environ Health Perspect 2016; 124: 46–52.

11. Crouse DL, Peters PA, Hystad P, et al. 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 2015; 123: 1180–6.

12. Wang Y, Shi L, Lee M, et al. Long-term exposure to PM2.5 and mortality among older adults in the southeastern US. Epidemiology 2017; 28: 207–14.

13. Wang Y, Kloog I, Coull BA, Koheleza A, Zanobetti A, Schwartz JD. Estimating causal effects of long-term PM2.5 exposure on mortality in New Jersey. Environ Health Perspect 2016; 124: 1182–88.

14. Wei Y, Wang Y, Wu X, et al. Causal effects of air pollution on mortality rate in Massachusetts. Am J Epidemiol 2020; 189: 1136–23.

15. Schwartz J, Wei Y, Yitshak-Sade M, Di Q, Dominici F, Zanobetti A. A national difference in differences analysis of the effect of PM2.5 on annual death rates. Environ Res 2021; 194: 110649.

16. US Census Bureau. Percentage of people covered by Medicare in the United States from 1990 to 2019. Statista. 2020. https://www.statista.com/statistics/200962/percentage-of-americans-covered-by-medicare/ (accessed April 28, 2021).

17. Health Care Financing Administration. Medicare 2000: 35 Years of Improving Americans’ Health and Security. Baltimore: Health Care Financing Administration, 2000.

18. Di Q, Amini H, Shi L, et al. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ Res 2019; 130: 104909.

19. Di Q, Amini H, Shi L, et al. Assessing NO2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. Environ Sci Technol 2019; 54: 1372–84

20. Requia W, Di Q, Silvern R, et al. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. Environ Sci Technol 2020; 54: 11037–47.

21. ESRI Data and Maps, United States Geological Survey. USA Hospitals. ArcGIS. 2010. https://www.arcg.is/home/item.html?id=516f737725a24dd69c2016feaf8875 (accessed July 9, 2020).

22. Abatoglu JT. Development of gridded surface meteorological data for ecological applications and modelling. Int J Climatol 2013; 33: 121–31.

23. Wu X, Braun D, Kosmourtzoglou M-A, Choiart C, Di Q, Dominici F. Causal inference in the context of an error prone exposure: air pollution and mortality. Ann Appl Stat 2019; 13: 520–47.

24. Wang Y, Lee M, Liu P, et al. Doubly robust additive hazards models to estimate effects of a continuous exposure on survival. Epidemiology 2017; 28: 771–79.

25. Hernan M, Robins J. Causal inference: what if. 2020. https://cdnl.sph.harvard.edu/wp-content/uploads/sites/1268/2020/11/ciwhatif_hernanrobins_23nov20.pdf (accessed June 30, 2020).

26. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017; 167: 268.

27. Wu X, Braun D, Schwartz JD, Kosmourtzoglou M-A, Dominici F. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. Sci Adv 2020; 5692: 1–13.

28. US Environmental Protection Agency. Review of the national ambient air quality standards for particulate matter. 2020. https://www.govinfo.gov/content/pkg/FR-2020-12-18/pdf/2020-27125.pdf (accessed June 20, 2020).

29. Vodonaes A, Awad YA, Schwartz J. The concentration-response between long-term PM 2.5 exposure and mortality: a meta-regression approach. Environ Res 2018; 166: 677–89.