Supplementary Online Content

Di Q, Dai L, Wang Y, et al. Association of short-term exposure to air pollution with mortality in older adults. *JAMA*. doi:10.1001/jama.2017.17923

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Covariates and Spatial Join

In the subgroup analysis, we used 5 variables to define subgroups, including sex (male or female), race/ethnicity (White, non-White, and other groups), age at death (≤69, 70~74, 75~84, and ≥85), Medicaid eligibility (as a proxy for low socioeconomic status [SES]), and quartiles of population density. Sex, race, age at death, and Medicaid eligibility were either retrieved or calculated from Medicare data. Population density was obtained from the 2000 US Census and the 2010 US Census.

Daily air and dew point temperatures data were retrieved from the North American Regional Reanalysis data in approximate 32 km × 32 km grids. We acquired daily 1 km × 1 km gridded air pollution levels ($PM_{2.5}$ and ozone) from previously developed and validated air pollution prediction models.[1,2] The prediction models predict the daily mean of $PM_{2.5}$ and 8-hour maximum ozone. For each individual, we extracted the residential zip code at death and obtained air temperature, dew point temperature, $PM_{2.5}$, and ozone levels by taking the inverse-distance mean of the 4 nearest grid cells to the zip code's centroid.

We used air pollution monitoring data from the United States Environmental Protection Agency's Air Quality System for the Nearest Monitor Analysis.[3] We obtained the daily mean of $PM_{2.5}$ and daily 8-hour maximum ozone. To join monitoring data to each residential zip code, we identified the nearest monitoring site within 50 km of the zip code (based on centroid point) and assigned air pollutant measurements to that zip code. If there was more than one monitoring site, we chose the nearest one; if there were no monitoring sites within 50 km, we treated the monitored exposure level as missing and excluded that zip code from the analysis.

eAppendix 2. Pooled Analysis

We estimated the exposure-response curves between both air pollutants and mortality using penalized splines. Due to computational limitations, running a conditional logistic regression with penalized splines on the whole data set was not possible. Instead, we randomly divided the entire data set into 50 groups with equal probability and estimated exposure-response curves for each group separately. To combine exposure-response curves from group-level analyses, we applied the meta-smoothing approach that was used and modified in previous studies.[5-7] In each group, the predicted relative risk increase(RRI) and its point-wise standard error were computed for each 1-μg/m^3 increment in $PM_{2.5}$ or 1-ppb increment in ozone. These group-level effect estimates ($\hat{\beta}_{ij}$=log RRI) in each group *i* and for exposure level *j*, and corresponding standard error $se(\hat{\beta}_{ij})$ were combined by regressing $\hat{\beta}_{ij}$ against indicator variables for each exposure level, with inverse variance weights. We assumed:

$$\hat{\beta}_{ij} \sim N(\beta_1 d_1 + \beta_2 d_2 + \cdots + \beta_j d_j, V_{ij}),$$

where d_j is the indicator variable for exposure level *j*, and V_{ij} is the estimated variance in group *i* at exposure level *j*.

The meta-analysis was implemented with R package *mvmeta.*[8]

eAppendix 3. Statistical Model

Statistical Model

The case-crossover design can be viewed as a hybrid between a matched case-control design and a traditional crossover design. In this setting, each case subject serves as his/her own control but is from a different time period where the event that defines case status was not experienced. Thus, since the same subject is both the case and control, observed and unobserved time invariant matching factors are controlled for by design. Let the index (case) time for subject i be denoted by ti, the exposure at the index time be denoted by x(ti) and let Wi represent the referent window for subject I (which includes the index and all referent periods).

> The likelihood function in terms of a defined referent selection scheme is constructed as follows. Assume a single event within each matched set i and let Y_{it} be an indicator of whether subject i's index time was on day t. If a localizable and ignorable referent selection scheme is chosen, then the likelihood of the data conditioning on the referent window, exposure series, and number of cases from subject i is:

$$
\begin{aligned}
P(T_i = t_i | \boldsymbol{x}, W_i, \sum_{s=1}^{T} Y_{is} = 1) &= \frac{P(T_i = t_i, \sum_{s=1}^{T} Y_{is} = 1 | \boldsymbol{x}, W_i)}{\sum_{t=1}^{T} P(T_i = t_i, \sum_{s=1}^{T} Y_{is} = 1 | \boldsymbol{x}, W_i)} \\
&= \frac{\lambda_i \exp(\boldsymbol{x}_{t_i} \boldsymbol{\beta})}{\sum_{t \in W_i} \lambda_i \exp(\boldsymbol{x}_t \boldsymbol{\beta})} \\
&= \frac{\exp(\boldsymbol{x}_{t_i} \boldsymbol{\beta})}{\sum_{t \in W_i} \exp(\boldsymbol{x}_t \boldsymbol{\beta})}. \qquad (2.21)
\end{aligned}
$$

Conditional logistic regression takes stratification into consideration.[4] The analysis included 1 case day and 3 or 4 control days in each stratum, denoted as times $t_1, t_2, \ldots, t_M$. The probability that subject *i* dies at time t_k is:$P_{ik} = \frac{exp(\boldsymbol{\beta}^T \boldsymbol{X}_{ik})}{\sum_{j=1}^{M} exp(\boldsymbol{\beta}^T \boldsymbol{X}_{ij})}$

where $\boldsymbol{X}_{ik}$ are the predictors at time t_k for subject i, and included $PM_{2.5}$, ozone, splines of air temperature and dew point temperature; $\boldsymbol{\beta}$ are regression coefficients, and stratum-specific intercepts canceled out.

Estimation of Related Risk

The relative risk increase (RRI) of all-cause mortality for 10-μg/m^3 increase in $PM_{2.5}$ and 10-ppb increase in ozone was given by the conditional logistic regression: $RR_{PM2.5} = exp(10 * \beta_{PM2.5})$ and $RR_{ozone} = exp(10 * \beta_{ozone})$. We also calculated absolute risk difference (ARD), that is, the difference in the daily mortality rate associated with 10-μg/m^3 increase in short-term exposures to $PM_{2.5}$ as following. First, we calculated the baseline daily mortality rate as the daily death rate in the Medicare population during our study period, which we denoted as α. We then calculated the ARD associated with 10-μg/m^3 increase in $PM_{2.5}$ as $ARD_{PM2.5} = \alpha * \frac{RR_{PM2.5}-1}{\mathrm{RR_{PM2.5}}}$ and its standard error as $se(ARD_{PM2.5}) = \alpha * exp(-\beta_{PM2.5} * 10) * se(10 * \beta_{PM2.5})$ according to the delta method. For ozone, the ARD estimate was calculated in a similar way, but using baseline daily mortality rate only for the warm season (from April to September). We calculated subgroup-specific ARD by using subgroup-specific daily mortality rate.

eAppendix 4. Sensitivity Analysis of Warm-Season Results

We restricted our analysis to the warm season (April 1 to September 30) when estimating the effect size of ozone in the main analysis. In a sensitivity analysis, we defined the warm season as May 1 to October 30 and restricted the analysis within this period. Results indicated that every 10-ppb increase in ozone was associated with a 0.59% (95% CI: 0.49%, 0.68%) increase in mortality, compared with a 0.51% (95% CI: 0.41%, 0.61%) increase in the main analysis (Table 2).

We reported the exposure-response relationship for $PM_{2.5}$ in Figure 5; here, we also reported the exposure-response relationship for $PM_{2.5}$ restricted to the warm season (April 1 to September 30) (eFigure 4). We reported the exposure-response relationship for ozone during the warm season (April 1 to September 30) only in Figure 5; here, we also reported the exposure-response relationship for the entire year (eFigure 4).

eAppendix 5. Test for Interaction

To test for statistically significant difference in RRI estimates across categories within subgroups, for example, in males vs females (H_0: $RR_{male} = RR_{female}$), we calculated:

$Z = \frac{RR_{male} - RR_{female}}{\sqrt{se(RR_{male})^2 + se(RR_{female})^2}}$. We tested whether ARD is significantly different in a similar way, using point estimate and standard error of ARD.

We also tested whether RRI estimates are significantly different below and above a certain air pollution threshold. Subgroups were defined in which one category of individuals died with exposure levels above the threshold and the second category died below the threshold. We repeated the above calculation to test whether RRI estimates are significantly different.

eTable. Sensitivity Analysis Using the Same-Day Exposure (lag 0 day) and Previous-Day Exposure (lag 1 day) and Mean of Daily Exposure on the Same Day of Death and One Day Prior (lag 01 day) of $PM_{2.5}$ and Ozone

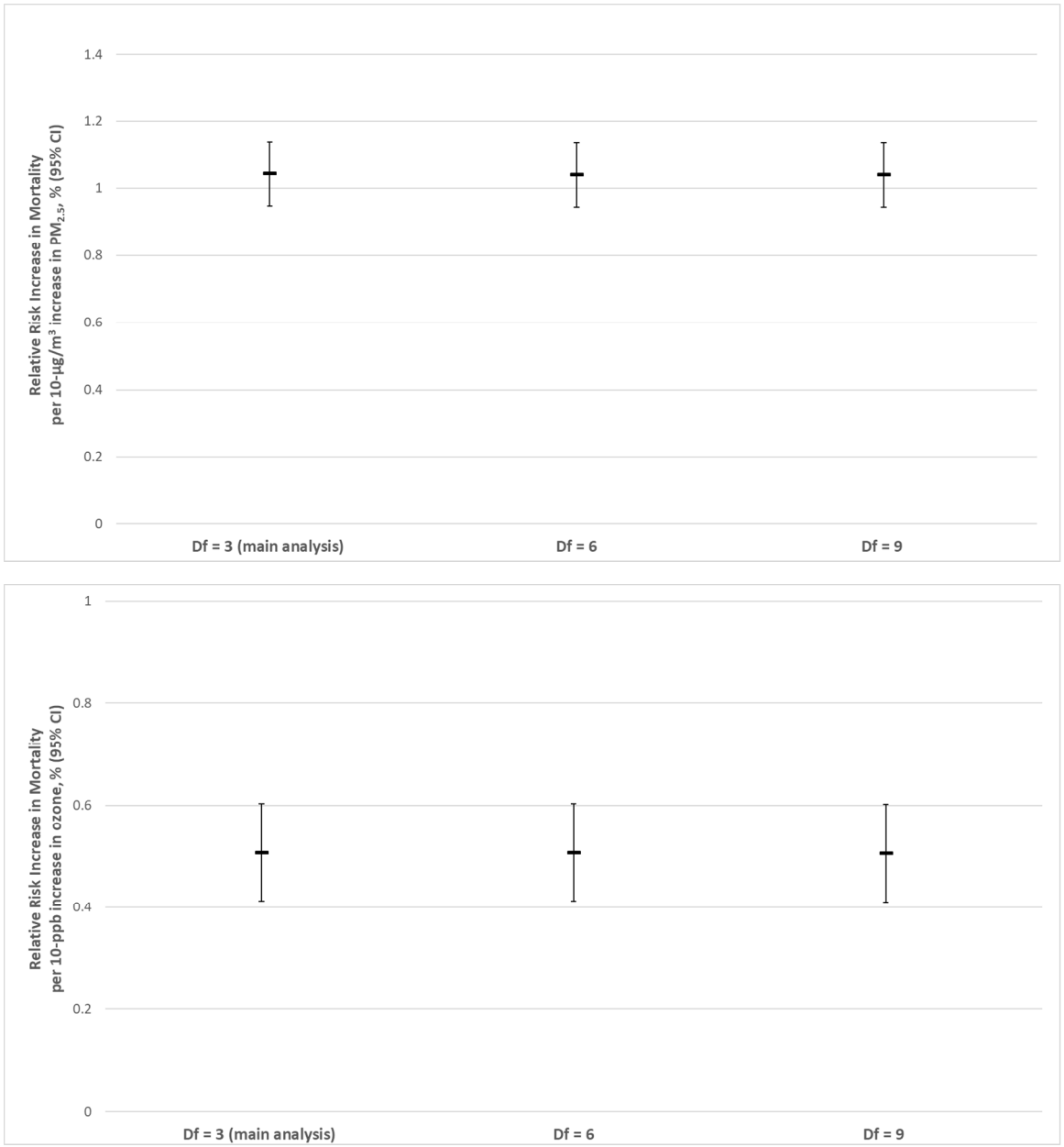
	$PM_{2.5}$[a]		Ozone[a]	
	Relative Risk Increase Estimate (95% CI)	**AIC[b]**	**Relative Risk Increase Estimate (95% CI)**	**AIC[b]**
Lag 01 Day[c]	1.05% (0.95%, 1.15%)	64,646,725	0.51% (0.41%, 0.61%)	30,635,577
Lag 1 Day[d]	0.83% (0.67%, 1.00%)	64,646,901	0.55% (0.38%, 0.72%)	30,635,801
Lag 0 Day[d]	0.79% (0.62%, 0.95%)	64,646,854	0.35% (0.19%, 0.51%)	30,635,663

[a] The analysis estimated $PM_{2.5}$ effect based on case days and control days from the entire year, while the ozone analysis used case days and control days from the warm season only (April 1 to September 30).

[b] Akaike information criterion

[c] The main analysis; results identical to Table 2.

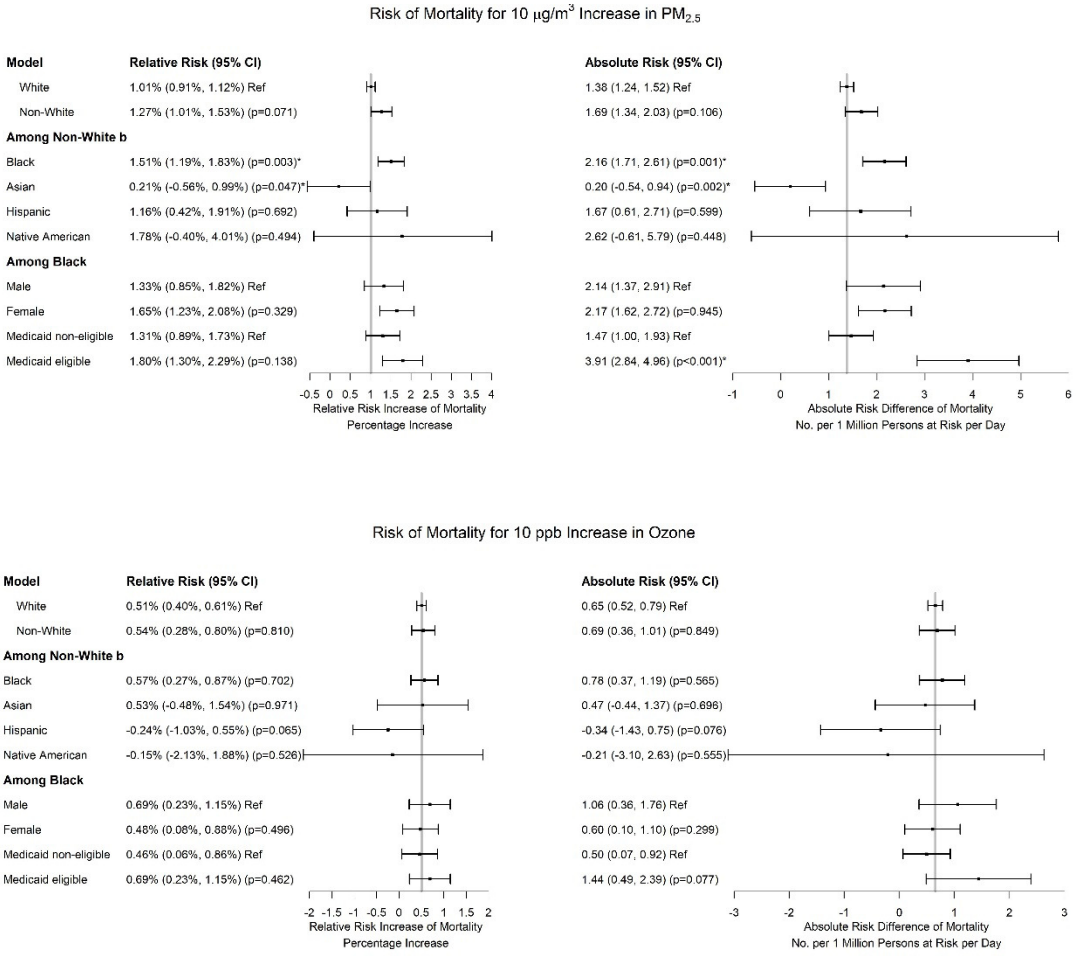
[d] To conduct the sensitivity analysis, we repeated the main analysis, but used the same day exposure (lag 0 day) and previous day exposure (lag 1 day). This model also controlled for natural splines of air and dew point temperatures with 3 degrees of freedom. The 2-pollutant analysis estimated the percentage increase in daily mortality rate associated with each 10-$\mu g/m^3$ increase in $PM_{2.5}$ exposure adjusted for ozone and the percentage increase in the daily mortality rate associated with each 10-ppb increase in ozone exposure adjusted for $PM_{2.5}$.



eFigure 1. Sensitivity Analysis Using Splines on Meteorological Variables With More Degrees of Freedom

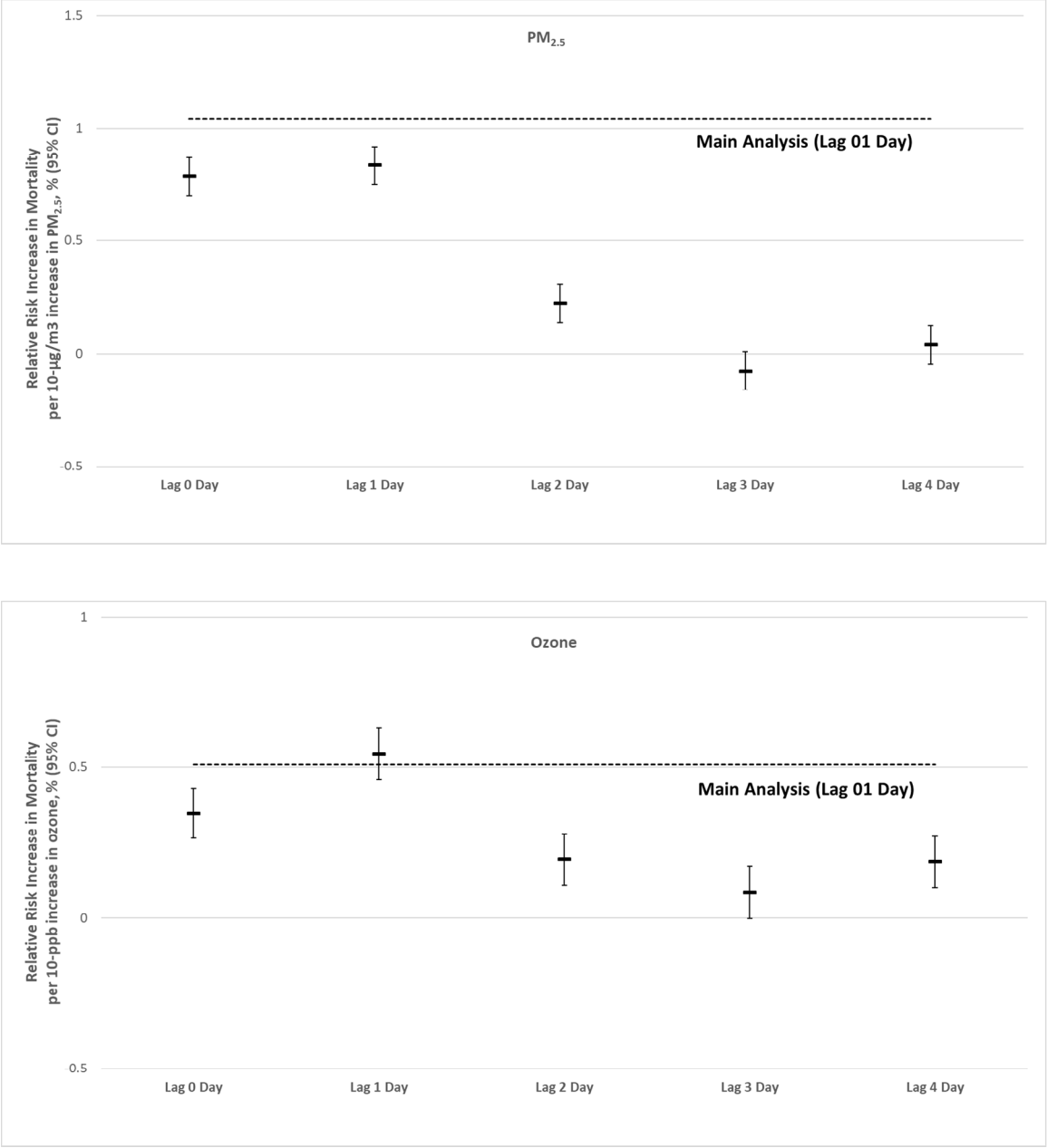
We repeated the main analysis, but changed the natural splines on air temperature and dew point temperature to 6 degrees of freedom and 9 degrees of freedom. The sensitivity analysis estimated the percentage increase of mortality associated with each 10-µg/m^3 increase in $PM_{2.5}$ exposure

adjusted for ozone and the percentage increase of mortality associated with each 10-ppb increase in ozone exposure adjusted for $PM_{2.5}$. The error bar indicates 95% confidence interval.



eFigure 2. Relative Risk Increase and Absolute Risk Difference of Daily Mortality Associated With Each 10-μg/m^3 Increase in $PM_{2.5}$ and 10-ppb Increase in Ozone Among Nonwhites

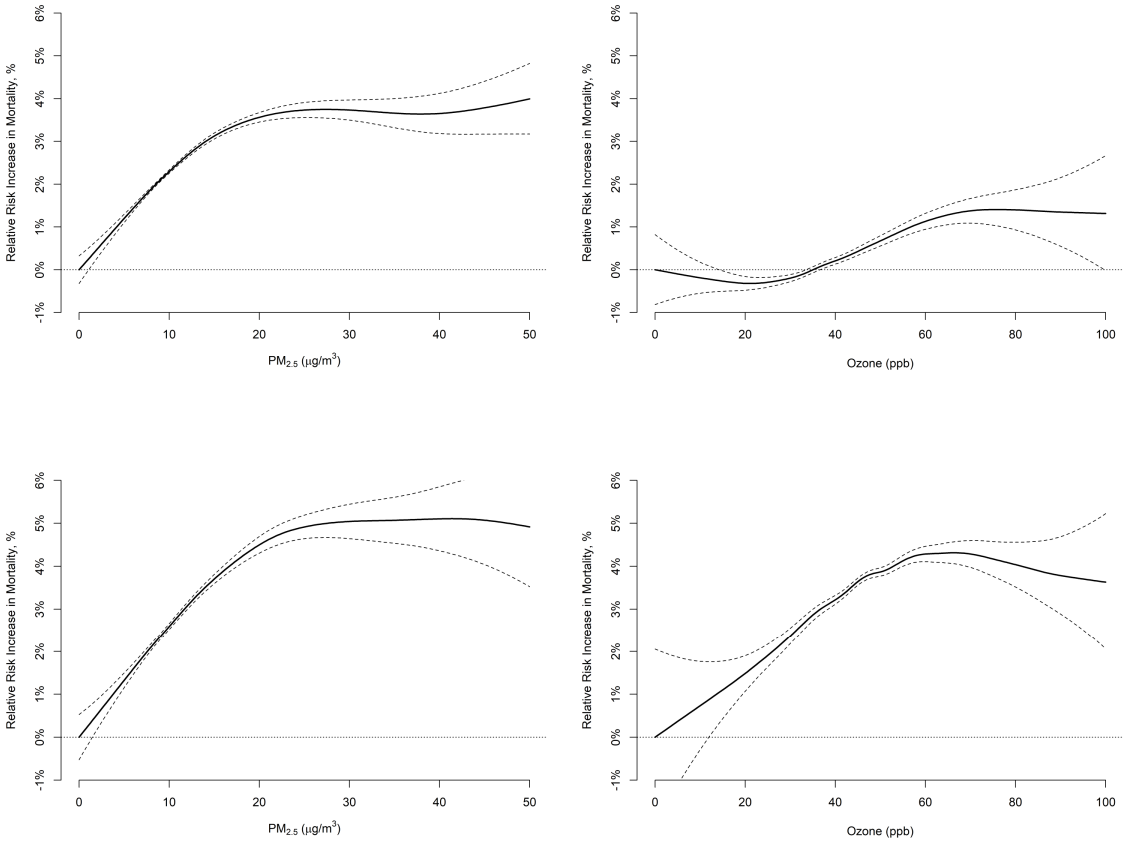
eFigure 2 was created using the same method as that described in Figure 2. The vertical lines were placed at the effect estimate for White individuals.



eFigure 3. Relative Risk Increase Associated With Each 10-µg/m^3 Increase in $PM_{2.5}$ and 10-ppb Increase in Ozone for Single-Lag Models

As a sensitivity analysis of the exposure time window, we used single-lag 2-pollutant models and compared them to our main models that used the mean of daily exposure on the same day of death and one day prior (Lag 01 Day). For example, in the Lag 0 Day Model, we included ozone

and $PM_{2.5}$ levels at day 0 (on the same day of death), temperature, and dew point temperature; in the Lag 1 Day Model, we included ozone and $PM_{2.5}$ levels at day 1 (1 day before the date of death), temperature, and dew point temperature. We fit the single-lag models separately and obtained risk estimates. We considered air pollution from the same day (lag 0 day) to up to 4 days (lag 4 day). eFigure 3 illustrates that lag 0 day and lag 1 day are most relevant to daily mortality. Air pollution concentrations 2 days prior to the date of death were less relevant to daily mortality. Based on the sensitivity analysis results (Table S1), we used the mean of daily exposure on the same day of death and one day prior (lag 01 day) as the exposure metric for both $PM_{2.5}$ and ozone.



eFigure 4. Estimated Exposure-Response Curves for Short-term Exposures to $PM_{2.5}$ and Ozone for the Entire Year and Restricted to the Warm Season

A 2-pollutant analysis with separate penalized splines on $PM_{2.5}$ (left panels) and ozone (right panels) was conducted to assess the percentage increase in daily mortality at various pollution levels. Dashed lines indicate 95% confidence intervals. The mean of daily exposure on the same day of death and one day prior (lag 01 day) were used as metrics of $PM_{2.5}$ and ozone. We plotted the exposure-response relationships for the entire year (upper panels). Analysis for both air pollutants were repeated and restricted to the warm season (April to September) (lower panels).

eReferences

1. Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. *Environ Sci Technol.* 2016;50(9):4712-4721.
2. Di Q, Rowland S, Koutrakis P, Schwartz J. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. *J Air Waste Manag Assoc.* 2017;67(1):39-52.
3. US Environmental Protection Agency. Air Quality System Data Mart (daily summary of PM2.5 FRM/FEM Mass (88101), daily summary of ozone (44201)). 2016.
4. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol.* 1991;133(2):144-153.
5. Schwartz J, Zanobetti A. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. *Epidemiology.* 2000;11(6):666-672.
6. Schwartz J, Ballester F, Saez M, et al. The concentration-response relation between air pollution and daily deaths. *Environmental health perspectives.* 2001;109:1001-1006.
7. Schwartz J, Laden F, Zanobetti A. The concentration-response relation between PM(2.5) and daily deaths. *Environmental health perspectives.* 2002;110(10):1025-1029.
8. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med.* 2012;31(29):3821-3839.